PROBABILITY THEORY

1° LECTURE - 26/11/2018

Aim of Probability theory: Provide mathematical description of a random experiment, whose outcome is unknown a priori:

→ Probability models: sample space and probability

→ Observable of the experiment: random variables and distribution

Probability space

1. Sample space $\Omega = \{\text{outcomes of experiment}\}$

   Ex: * choice of 1 card from a deck: $\Omega = \{\text{set of distinct cards}\}$

   * # client arrivals in a service: $\Omega = \mathbb{N}_0 = \mathbb{N} \cup \{0\}$

   * waiting time on a queue: $\Omega = \mathbb{R}^+$

2. σ-algebra of events $\mathcal{F}$:

   events are subsets $A \subset \Omega$ to which we will assign a probability

   Denote by $\mathcal{P}(\Omega) = \{A \subset \Omega\}$ power set of Ω

   * When Ω is uncountable (e.g. $\mathbb{R}$, or $[a,b] \subset \mathbb{R}$) the set $\mathcal{P}(\Omega)$ is too big for concrete applications and is better to consider a "suitable" family of events $\mathcal{F} \subset \mathcal{P}(\Omega)$. To develop a concrete theory, we ask $\mathcal{F}$ to be a σ-algebra:

   1: $\phi \in \mathcal{F}$

   2: if $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ (closed w.r.t. complement)

   3: if $(A_n)_{n \in \mathbb{N}} \in \mathcal{F} \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$ (closed w.r.t. countable union)

When $\Omega = \mathbb{R}$ (or $[a,b]$), as it will be in all the situation we will consider, we choose the σ-algebra of Borel: $\mathcal{B}(\mathbb{R})$ (or $\mathcal{B}([a,b])$).

This is the σ-algebra that includes all <u>open sets</u> (and then <u>closed sets!</u>) and can be identified as the σ-algebra containing all the half lines
$\underline{\{S = (-\infty, t],\ t \in \mathbb{R}\} = \mathcal{S}}$ (useful when we will speak about distribution)

the couple $(\Omega, \mathcal{F})$ is called measurable space.

3. Given $(\Omega, \mathcal{F})$ (when Ω is discrete, $\mathcal{F} \equiv \mathcal{P}(\Omega)$), a function
$$\mathbb{P}: \mathcal{F} \longrightarrow [0,1]$$
$$A \longrightarrow \mathbb{P}(A)$$
is a <u>probability</u> if

i. $\mathbb{P}(\Omega) = 1$

ii. If $(A_n)_{n\in\mathbb{N}}$ is a family of disjoint subsets $\mathcal{F}$,
$$\mathbb{P}\left(\bigsqcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n) \qquad (\sigma\text{-additivity})$$

<u>Remark</u>: From i. and ii. we get (prove as exercise)

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- If $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ (monotonicity)
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leq \mathbb{P}(A) + \mathbb{P}(B)$
- $\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1 < \dots < i_k} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k})$ (inclusion-exclusion formula)

prove by induction

Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

Example 1: Uniform DISCRETE space (sampling from a finite population)

Let Ω finite, $|\Omega| = \#$ of outcomes $< \infty$, $\mathcal{F} = \mathcal{P}(\Omega)$

Assume that all outcomes are equally likely:

$\omega \in \Omega$: $\mathbb{P}(\omega) = c$ $\quad c =$ constant

- $\Longrightarrow 1 = \mathbb{P}\left(\bigsqcup_{\omega \in \Omega} \{\omega\}\right) = \sum_{\omega \in \Omega} \mathbb{P}(\omega) = c|\Omega| \Longleftrightarrow c = \frac{1}{|\Omega|}$
- $\Longrightarrow A \subset \Omega$: $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$

$A = \cup \{\omega\}$

Example 2: Uniform CONTINUOUS space (sampling from an interval)

Let $\Omega = [a,b] \subset \mathbb{R}$

All outcomes are equally likely, meaning that any infinitesimal interval dx has same probability.

NOTICE: In that case $\mathbb{P}(\omega) = 0$ $\forall \omega \in \Omega$. Otherwise we will have the contradiction $1 = \mathbb{P}(\Omega) \geq \sum_{n \in \mathbb{N}} \mathbb{P}(\omega_n) = \infty \neq$

$\hookrightarrow (\omega_n)_{n \in \mathbb{N}}$ a sequence in $[a,b]$

If $I \subset [a,b]$ is an interval, we set

- $\mathbb{P}(I) = \int_I \frac{dx}{\ell([a,b])} = \frac{\ell(I)}{\ell[a,b]}$, $\quad \ell(I) =$ length of I $=$ Lebesgue measure of I

and in general, for $A = \bigcup_{n \in \mathbb{N}} I_n$

- $\mathbb{P}(A) = \int_A \frac{dx}{\ell([a,b])}$ (by σ-additivity)

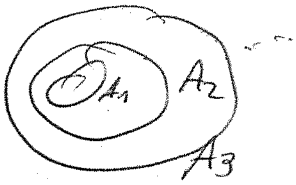
This measure can be assigned to every $E \in \mathcal{B}([a,b])$.

Probability Properties

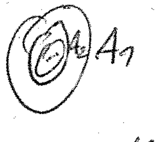
1. Continuity of the probability

Let $(A_n)_{n\in\mathbb{N}}$ events in $\mathcal{F}$. We say that

- $(A_n)_{n\in\mathbb{N}}$ is increasing if $A_1 \subset A_2 \subset \dots \subset A_n \dots$



- $(A_n)_{n\in\mathbb{N}}$ is decreasing if $A_1 \supset A_2 \supset \dots \supset A_n \dots$



<u>Prop</u>: If $(A_n) \in \mathcal{F}$ s.t. is decreasing or increasing, then

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\lim_{n\to\infty} A_n\right)$$

↳ Notice that
- if $(A_n)\nearrow \Rightarrow \lim\limits_{n\to\infty} A_n = \bigcup\limits_{n=1}^{\infty} A_n$ $\quad$ $*$
- if $(A_n)\searrow \Rightarrow \lim\limits_{n\to\infty} A_n = \bigcap\limits_{n=1}^{\infty} A_n$

---

2. Liminf and Limsup of events

<u>Def</u>: Let $(A_n)_{n\in\mathbb{N}}$ sequence of events in $\mathcal{F}$. Define

1. $\limsup\limits_{n\to\infty} A_n := \lim\limits_{n\to\infty} \bigcup\limits_{k\geq n} A_k \overset{①}{=} \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k\geq n} A_k$
2. $\liminf\limits_{n\to\infty} A_n := \lim\limits_{n\to\infty} \bigcap\limits_{k\geq n} A_k \overset{②}{=} \bigcup\limits_{n=1}^{\infty} \bigcap\limits_{k\geq n} A_k$

Notice: If $B_n := \bigcup\limits_{k\geq n} A_k \;\forall n\in\mathbb{N} \Rightarrow (B_n)_{n\in\mathbb{N}} \searrow$ and identity ① follows from $*$.

If $B_n = \bigcap\limits_{k\geq n} A_k \;\forall n\in\mathbb{N} \Rightarrow (B_n)_{n\in\mathbb{N}} \nearrow$ and identity ② follows from $*$.

Interpretation:

1. $\limsup_{n\to\infty} A_n =$ "$\forall n, \exists k \geq n$ s.t. A_k happens"

$= \{A_n \text{ infinitely often}\} = \{A_n \text{ i.o.}\}$

2. $\liminf_{n\to\infty} A_n =$ "$\exists n$ st $\forall k \geq n$ A_k happens"

$= \{A_n \text{ eventually}\} = \{A_n \text{ ev}\}$

Properties:

- $\left(\limsup_{n\to\infty} A_n\right)^c = \left(\bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k\right)^c = \bigcup_{n=1}^{\infty} \bigcap_{k\geq n} A_k^c = \liminf_{n\to\infty} A_n^c$

and similarly: $\left(\liminf_{n\to\infty} A_n\right)^c = \limsup_{n\to\infty} A_n^c$

- $\liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$ (verify as an exercise)

Lemma di Borel-Cantelli

1. If $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}$ s.t. $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < +\infty$

$\Longrightarrow \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0$

2. If $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}$ st. A_n are .... and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty$

$\Longrightarrow \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1$

Example: Infinite toss coin, with Head $= 1$, Tail $= 0$

Then $\Omega = \{0,1\}^{\mathbb{N}} \ni \omega = (\omega_n)_{n\in\mathbb{N}}$ with $\mathbb{P}(\omega_j = 1) = p$

Set $A_n = \{\omega : \omega_n = 1\}$ so that $\mathbb{P}(A_n) = p$

$\rightarrow$

then $\sum_{n=1} \mathbb{P}(A_n) = \infty$ and also $\sum_{n=1} \mathbb{P}(A_n^c) = \infty$

Since the events $A_n, n \in \mathbb{N}$ are independent, by the lemma of BC

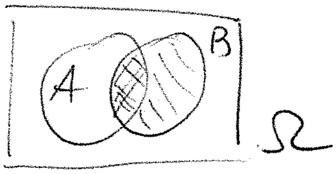
$\mathbb{P}(\limsup_{n \to \infty} A_n) = 1$ and $\mathbb{P}(\liminf_{n\to\infty} A_n) = 1 - \mathbb{P}(\limsup_{n\to\infty} A_n^c) = 0$ #

3. Conditional probability

For $A, B \in \mathcal{F}$ s.t. $\mathbb{P}(B) \neq 0$, define

$$(*) \quad \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

conditional probability (probability of A given B)



From $(*)$ we derive useful rules:

- Bayes formula

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

- $\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$ for $\Omega = B \sqcup B^c$
- If $\Omega = \bigsqcup_{i=1}^{n} B_i$: $\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$

4. Independence of events

$A, B \subset \Omega$ are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \iff \begin{cases} \mathbb{P}(A|B) = \mathbb{P}(A) \\ \mathbb{P}(B|A) = \mathbb{P}(B) \end{cases}$$

Note: If A, B independent, then A^c, B indep, A, B^c indep., and A^c, B^c indep. (prove as exercise!)

e.g. $\mathbb{P}(A) = \mathbb{P}(B \cap A) + \mathbb{P}(B^c \cap A) = \mathbb{P}(B)\mathbb{P}(A) + \mathbb{P}(B^c \cap A)$

$\Rightarrow \mathbb{P}(B^c \cap A) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(B^c)$ #

Bernoully trials (Example of probability space constructed from independence)

$$\Omega_n = \{\underline{\omega} = (\omega_1, \dots, \omega_n),\ \omega_j \in \{0,1\},\ \forall j = 1, \dots n\}$$

= sequence of 0,1, corresponding (respect.) to fail or success of some experiment repeated n times.

Assume that $\mathbb{P}(\omega_j) = \begin{cases} p & \text{if } \omega_j = 1 \\ 1-p & \text{if } \omega_j = 0 \end{cases}$, independently $\forall j = 1, \dots n$

or in short $\mathbb{P}(\omega_j) = p^{\omega_j}(1-p)^{1-\omega_j}$

then $$\mathbb{P}(\underline{\omega}) = \prod_{j=1}^{n} \mathbb{P}(\omega_j) = \prod_{j=1}^{n} p^{\omega_j}(1-p)^{1-\omega_j} = p^{\sum_{j=1}^{n}\omega_j} \cdot (1-p)^{n - \sum_{j=1}^{n}\omega_j}$$

Random variables: quantities (usually numerical) depending on the experiment

$$X : \Omega \longrightarrow \mathcal{X}, \quad \omega \longrightarrow X(\omega)$$

($\mathcal{X}$ is the sample space of X, usually is $\mathbb{N}, \mathbb{R}, \mathbb{R}^d, [a,b], \dots$)

$$\mathcal{X} = \{x \text{ s.t. } X(\omega) = x;\ \omega \in \Omega\}$$

Ex: Indicator functions

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

Note: Any r.v. X s.t. $\mathcal{X} = \{0,1\}$ can be written as an indicator fct:

set $A = \{\omega : X(\omega) = 1\} \Longrightarrow X(\omega) = \mathbb{1}_A(\omega)$

this r.v.'s are called Bernoulli.

Consider the measurable space $(\mathcal{X}, \mathcal{E})$, where $\mathcal{E}$ is σ-algebra on $\mathcal{X}$ (e.g. $\mathcal{P}(\mathcal{X})$ or $\mathcal{B}(\mathcal{X})$):

* For $A \in \mathcal{E}$, let

$$X^{-1}(A) \equiv \{X \in A\} := \{\omega \in \Omega : X(\omega) \in A\} \subset \Omega$$

e.g. $\{X = a\} = \{X \in \{a\}\} = \{\omega : X(\omega) = a\}$

$\{X > a\} = \{X \in (a, +\infty)\} = \{\omega : X(\omega) > a\}$

These are the <u>events generated by X</u> :

$$\begin{array}{ccc} \Omega & \xrightarrow{X} & \mathcal{X} \\ \cup & & \cup \\ \{X \in A\} & \xleftarrow{X^{-1}} & A \end{array}$$

$\Downarrow$

* <u>Law (or distribution) of X</u> on $(\mathcal{X}, \mathcal{E})$

Define $\mathbb{P}_X : \mathcal{E} \to [0,1]$

$$A \to \mathbb{P}_X(A) := \mathbb{P}(X \in A)$$

$\mathbb{P}_X$ is <u>the law of X</u> (and is a probability on $(\mathcal{X}, \mathcal{E})$)

$\hookrightarrow (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathcal{X}, \mathcal{E}, \mathbb{P}_X)$

. <u>DISCRETE R.V.</u> :

If X is a r.v. on $\mathcal{X}$ discrete (finite or countable)

we let $(\mathbb{P}_X(a), \ a \in \mathcal{X})$ be the <u>density of X</u>.

<u>The density identifies the law of X</u> through

$$\mathbb{P}_X(A) = \sum_{a \in A} \mathbb{P}_X(a) \qquad \left(\begin{array}{l} \text{as } A = \bigsqcup_{a \in A} \{a\} \\ \text{and } \mathbb{P}_X \text{ is a prob.} \end{array} \right)$$

<u>Examples</u> $\hookrightarrow \sum_{a \in \Omega} \mathbb{P}_X(a) = 1$

1. $X \sim Be(p)$, $p \in [0,1]$ <u>Bernoulli law</u>

if $\mathcal{X} = \{0, 1\}$, $\mathbb{P}_X(1) = \mathbb{P}(X = 1) = p$, $\mathbb{P}_X(0) = \mathbb{P}(X = 0) = 1 - p$

2. Binomial (n, p), $n \in \mathbb{N}$, $p \in [0,1]$

$X \sim Bi(n,p)$ if $\mathcal{X} = \{0, 1, \ldots n\}$ and

$\mathbb{P}_X(k) = \mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ if $k \in \{0, \ldots n\}$

3. Geometric p, $p \in [0,1]$

$X \sim Geo(p)$ if $\mathcal{X} = \mathbb{N}$ and

$\mathbb{P}_X(k) = (1-p)^{k-1} \cdot p$

4. Uniforme (DISCRETE) on $\mathcal{X}$

$X \sim \mathcal{U}(\mathcal{X})$ if $\mathbb{P}_X(k) = \frac{1}{|\mathcal{X}|}$ if $k \in \mathcal{X}$

5. Poisson λ, $\lambda \in \mathbb{R}^+$

$X \sim P_0(\lambda)$ if $\mathcal{X} = \mathbb{N}_0$

$\mathbb{P}_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$

(Verify as an exercise that $\sum_{a \in \Omega} \mathbb{P}_X(a) = 1$)

2. CONTINUOUS R.V.

If $\mathcal{X}$ is a continuous space (say $\mathbb{R}$), to assign a law $\mathbb{P}_X$ is equivalent to provide the distribution function of X.

Def: Let X r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$ on value on $\mathcal{X}$.

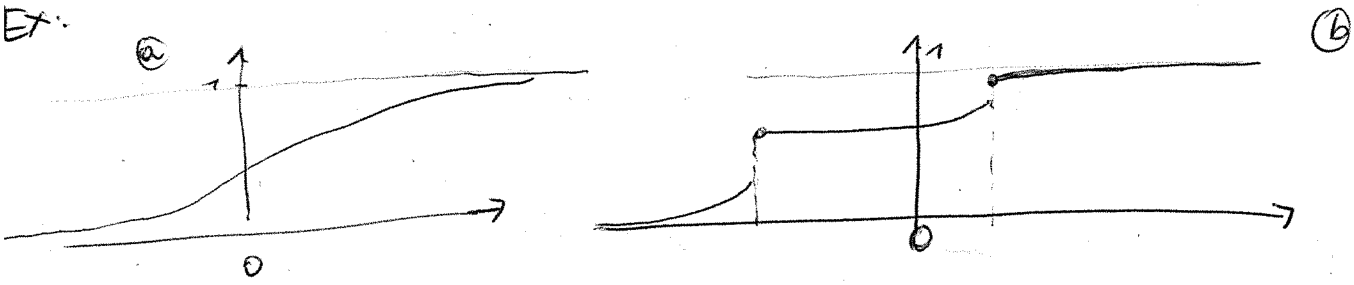
The distribution function

$$F_X : \mathbb{R} \longrightarrow [0,1], \quad x \in \mathbb{R}$$
$$x \longrightarrow \mathbb{P}(X \leq x)$$

Properties of F:

1) F_X is monotone increasing
2) $F_X(x) \xrightarrow[X\to+\infty]{} 1$, $F_X(x) \xrightarrow[X\to-\infty]{} 0$
3) F_X is right continuous: $\forall x_0 \in \mathbb{R}$, $\lim_{x\to x_0^+} F_X(x) = F_X(x_0)$

Ex:



Notice that it may be

$$F_X(x_0^-) := \lim_{X\to X_0^-} F_X(x) \neq \lim_{X\to X_0^+} F_X(x) \overset{(3)}{=} F_X(x_0) \quad \text{(as in (b))}$$

In that case $\underline{P(X=x_0) = F_X(x_0^-) - F_X(x_0)}$ (*)

Through the distr. function we can compute the probability of all events of interest.

Ex: Let X r.v. with $F_X(x) = \begin{cases} 0 & X < -1 \\ \frac{1}{2}\left(1+\frac{X}{2}\right) & X \in [-1,1) \\ 1 & x \geq 1 \end{cases}$

Compute:

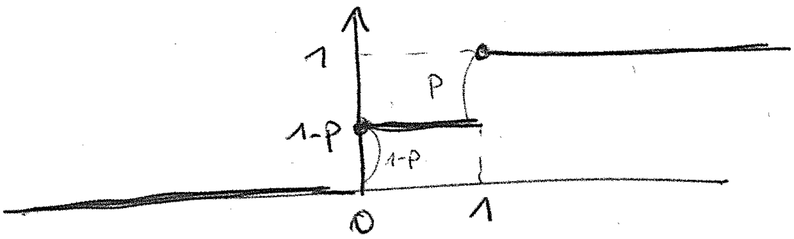
a. $P(X \leq 0)$ b. $P(X > \frac{1}{2})$ c. $P(0 < X < \frac{1}{2})$ d. $P(X=-1$

Discrete RV: If X is a discrete r.v., then

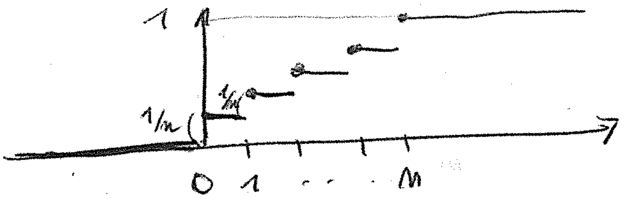
$$F_X(x) = P(X \leq x) = \sum_{\substack{k \in \mathcal{X} \\ k \leq x}} P(X=k)$$

$\Rightarrow$ F_X is a constant step function with jumps given by (*).

Ex 1 $X \sim Be(p)$ $\quad \mathbb{P}(X=0) = 1-p$, $\mathbb{P}(X=1) = p$



Ex 2 $X \sim \mathcal{U}\{0, 1, \dots, n-1\}$ $\quad \mathbb{P}(X=k) = \frac{1}{n}$ $\quad \forall k = \{0, \dots n-1\}$



Notice that $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \in (-\infty, x])$, thus, assigning F_X is equivalent to assign a probability to all the half-lines $\mathcal{S} = \{S = (-\infty, x], x \in \mathbb{R}\}$, and thus to all the element of $\mathcal{B}(\mathbb{R})$, that is the σ-algebra containing $\mathcal{S}$.

<u>CONTINUOUS R.V.</u>

If $\mathcal{X}$ is an uncountable space (X a cont. r.v.) is not anymore true that $\mathbb{P}(X \in A) = \sum_{a \in A} \mathbb{P}(X=a)$

Indeed, $A = \bigsqcup_{a \in A} \{a\}$ may be uncountable and then

$$\mathbb{P}(A) \neq \sum_{a \in A} \mathbb{P}_X(a) \quad \text{(true only for countable union)}$$

In particular, it holds that $\underline{\mathbb{P}(X=a) = 0}$ $\forall a \in \mathcal{X}$ but a finite or countable set

---

The most studied case is when F_X is a continuous function obtained as integral of a density fct.

Def: X is an absolutely continuous r.v. if $\exists f_X : \mathbb{R} \to \mathbb{R}^+$ s.t. $F_X(t) = \int_0^t f_X(s)\,ds$

f_X is called density of X and satisfies the following:

1. $\int_{\mathbb{R}} f_X(x)\,dx = 1$ 2. $\lim_{x \to \pm\infty} f_X(x) = 0$ 3. $f_X(x) = F_X'(x)$ $\forall x$ s.t. f_X is continuous

Then we have: $\forall A \in \mathcal{F}$, $\mathbb{P}_X(A) = \int_A f_X(x)\,dx$

Average (or mean)

- Discrete case:

$$\mathbb{E}(X) = \sum_{x \in \mathcal{X}} x \cdot \mathbb{P}(X = x)$$

Moreover if $g : \mathcal{X} \to \mathbb{R}$, then $\underbrace{g \circ X}_{= g(X)} : \Omega \to \mathbb{R}$ and we have

$$\mathbb{E}(g(X)) = \sum_{x \in \mathcal{X}} g(x)\,\mathbb{P}(X = x)$$

- Continuous case (with density)

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \cdot f_X(x)\,dx$$

Moreover, if $g : \mathcal{X} \to \mathbb{R}$ then $g \circ X : \Omega \to \mathbb{R}$ and we have

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

Properties:

1. Monotonicity: if $X \leq Y \Rightarrow \mathbb{E}(X) \leq \mathbb{E}(Y)$
2. Linearity: $\mathbb{E}(aX + b) = a\,\mathbb{E}(X) + b$

Example

- If $\mu = \mathbb{E}(X)$ and $g(x) = (x-\mu)^2$, then

$$\mathbb{E}((X-\mu)^2) =: \mathrm{Var}(X)$$

$$\overset{\shortparallel}{\int_{\mathbb{R}} (x-\mu)^2 f_X(x)\,dx} \qquad \left(\text{or } \sum_{x \in \mathcal{X}} (x-\mu)^2\, \mathbb{P}(X=x) \text{ in discrete case} \right)$$

and it turns out from monotonicity that $\mathrm{Var}(X) \geq 0$

Mixed distribution: If X is continuous but not absolutely continuous, that is F_X has some jumps, then proceed as follows:

- Set $H_X(x) = \sum_{y \leq x} F_X(y) - F_X(y^-)$ $\quad \forall x \in \mathbb{R}$
- Set $G_X(x) = F_X(x) - H_X(x)$ $\quad \forall x \in \mathbb{R}$

and by construction notice that G_X is continuous

- Define $g_X(x) = G_X'(x)$ (for points $x \in \mathbb{R}$ when this is well defined)

Then $\mathbb{E}(X) = \int_{\mathbb{R}} x \cdot g_X(x)\,dx + \sum_{x \in \mathbb{R}} x \cdot \mathbb{P}(X=x)$

and $\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)\, g_X(x)\,dx + \sum_{x \in \mathbb{R}} h(x) \cdot \mathbb{P}(X=x)$.