Lezione 1 - Intro to Probability

Engineering Statistics.

Intro to Probability → Start

All statistic analysis begin by analyzing a sample.

We are not interested in how the data in generated and collected.

Governments and engineers and others collect data.

Data can be numeric or categorical (e.g smoker or non-smoker)

Example using celiac disease, we used two groups and measured the count, when $\bar{x}_y$ is the antibody count.

Exploratory Data Analysis
↳ the part which is consemed with describing summarising data, like calculating mean.

Suppose $\bar{x}_2$ (PLACEBO) is higher than $\bar{x}_1$ (NEWDRAG)

Is the drug effective?

We look at the data to see if the data is reliable, we need to measure the plausibility.

Inferensial Statistics
↳ Drawing conclusions from data

Data value are measured value, that are not deterministic but random.

Since measuring the whole population is not possible, we extract at random a sample of the italian population, and measure from that, we understand that there will be some uncertainty, intrinsic to the fact that we are not measuring the whole population.

From a random sample we will have variability between the results of the data.

$$X = \mu + \varepsilon$$

μ ↳ constant

ε → "noise" → variation due to various factors.

How do we know that the results are not by chance?

Probability Theory
↳ Part of mathematics that studies a random experiment

DEF: A random number is a numeric variable whose values are determined by the outcome of the random experiment.

- The number can be random if it is associated with a random experiment, even if the experiment is not numerical in nature
  - It is standard to call it a random variable.
  - Denoted with capital letter

$X \rightarrow$ random variable (r.v)

$x \rightarrow$ a possible value for this random variable.

Examples:

- Number of heads in 10 coins $\rightarrow$ can only be a limited set

  $X \in \{0, 1, 2, 3, \dots, 9, 10\}$

- Number of defective parts produced daily
- Time to repair a television. $\rightarrow$ can be any number > 0

$(x_1, x_2, \dots, x_{10}) = (50,1, \dots, 51,2, 49,9) \rightarrow$ value measured

Probability: the change a particular event will occur.

Eg. $P(X \in [9{,}8, 10{,}15])$ is between 0 and 1.

<u>Notation:</u>

E: any interval of $\mathbb{R}$, with E^c being it's complement.

E_i and E_j are pairwise disjoint if $E_i \cap E_j = \emptyset$ for any $i \neq j$

$\mathcal{J}$ is the collection of all real intervals, and of all sets which are obtained from intervals, doing at most an infinite (numerable) set of 3 operations (union, intersect, complement)

Definition:

Let X be a random variable. $\underset{\in \mathcal{J}}{E} \longrightarrow \underset{\in \mathbb{R}}{P(X \in E)}$

is called probability distribution of X if these assumptions are met:

1) $P(X \in E) \geq 0 \;\; \forall E \in \mathcal{J}$

2) $P(X \in \mathbb{R}) = 1$

Sigma Additivity 3) $E_1, E_2, E_3, \dots, \in \mathcal{J}$ are pairwise disjoint

$\rightarrow P(X \in (E_1 \cup E_2 \cup E_3 \cup \dots)) = P(X \in E_1) + P(X \in E_2) + \dots +$

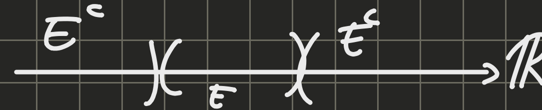
$\Rightarrow$ 3.1) $\underbrace{E_1, E_2, \dots, E_{\underset{\text{finite number}}{n}}}_{\text{a finite set}} \in \mathcal{J}$ are pairwise disjoint:

$$P(X \in (E_1 \cup E_2 \cup \dots \cup E_n)) = P(X \in E_1) + P(X \in E_2) + \dots + P(X \in E_n)$$

If we say E_{n+1} is the emptyset then we get 3.

<u>Properties</u> of the probability distribution of X $\rightarrow$ the probability function

4) $P(X \in E^c) = 1 - P(X \in E) \rightarrow$ Proof:



$\mathbb{R} = \underbrace{E \cup E^c}_{\text{disjoint}}$

↳ Since $P(X \in \mathbb{R}) = 1$; (2)

$= P(X \in E) + P(X \in E^c)$ (3)

$\Rightarrow P(X \in E) = 1 - P(X \in E^c)$

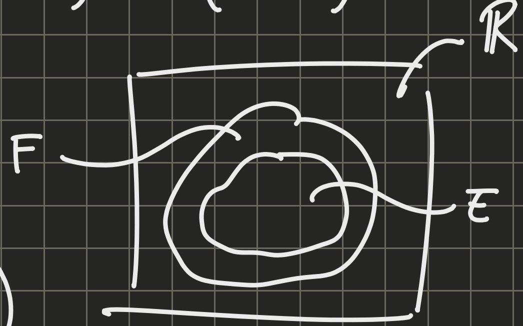
Example: $\left.\begin{array}{l} P(X > x) = 1 - P(X \le x) \\ E = (x, \infty) \Rightarrow E^c = (0, x) \end{array}\right\}$ Very useful

5) If $S \in \mathcal{F}$ and $P(X \in S) = 1 \Rightarrow P(X \in S^c) = 0$

↳ 5.1) $P(X \in \emptyset) = 0$

6) $E, F \in \mathcal{F}$, if $E \subset F \Rightarrow P(X \in E) \le P(X \in F)$

7) $E, F \in \mathcal{F} \Rightarrow P(X \in (E \cup F)) =$



$= P(X \in E) + P(X \in F) - P(X \in (E \cap F))$

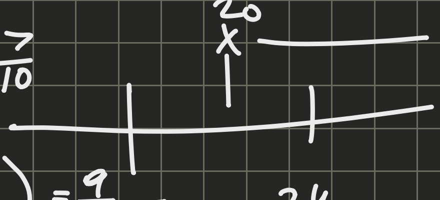
↳ Without knowing whether they intersect or not

Comment: $0 \overset{(2)}{\le} P(X \in E) \overset{(6)}{\le} 1 \quad \forall E \in \mathcal{F}$

Example

$P(X \le 15) = \frac{3}{10}$, $P(15 < X \le 24) = \frac{3}{5}$, $P(X > 20) = \frac{1}{20}$

1) $P(X > 15) = 1 - P(X \le 15) = 1 - \frac{3}{10} = \frac{7}{10}$



2) $P(X \le 24) = P(X \le 15) + P(15 < X \le 24) = \frac{9}{10}$

different since they have to be disjoined

3) $P(15 < X \leq 20) =$

Remember

$\forall a < b \quad P(a < X \leq b) = P(X \leq b) - P(X \leq a)$

$(-\infty, b] = (-\infty, a] \cup (a, b] \Rightarrow P(X \leq b) = P(X \leq a) + P(a < X \leq b)$

$\Rightarrow = P(X \leq 20) - P(X \leq 15) = 1 - \underbrace{P(X > 20)}_{= \frac{2}{10}} - P(X \leq 15)$

Homework

Compute

4) $P(X \leq 18)$ knowing that $P(18 < X \leq 24) = \frac{2}{5}$