Inferential Statistics

25/01/24
14/02/24
06/24
06/24
09/24

1 pagina con appunti, no libri

Voto esame + pti bonus

- partecipazione alle lezioni → [0,3]
- validi per tutto l'anno

Valuta il procedimento e non il risultato

"Mathematical Statistics, Hogg McKean Craig"

"Exercises and Solutions in Statistical theory"

Utilizziamo "R" (simile a Python)

Statistics is useful in Data Science: we want to provide:

- evidence about some theory
- support for a business decision problem

An algorithm is just a mean to reach a solution.

Statistics is a way to better understand a problem and helps making an algorithm

3 steps in inferential statistics:

- translate the problem in terms of a statistical model
- fit the model with the data in my model
- translate the output of the model in terms of the original problem

2 key points:
- choice of the statistical model
  ↳ many models are right for each problem
- fit the model to the data

Choosing the model is hard → every model we choose is wrong → always an aproximation
↳ We never have "the" right model
↳ carefully chosen modeles can still be usefule

ESTIMATION → Estimate the most probable outcome by testing a product in a lab.
↳ Lots of samples: can't mesure all of it

HYPOTESIS TESTING → Use the data collected to validate or invalidate an hypotesis

CONFIDENCE SET → How confident with the result

Based on the type of problems we have different methods

	Parametric	Non-param.
Frequentist		
Bayesian		

probability models with a finite number of parameters

min 40

Sample space

S: set of all realisations s of a rnd experiment

es: # faces of die (6)

tosses of a coin (discrete but infinite)

sample point s: an element of S
event E: subset of S

es: two dice experiment

$$S = \{(1,1), \dots, (6,6)\}$$

$s = (1,1)$ → "two dice show 1"

E = "both numbers ≤ 2": $E = \{(1,1), (1,2), (2,1), (2,2)\}$

there are two special events: $E = S$, $E = S^c = \emptyset$

let's call $\mathcal{A} = P(S)$ all possible events, we have $\begin{cases} E \subset S \\ E \in \mathcal{A} \end{cases}$
↳ Power Set

It works on finite set, in continuous S we have to be more carefule

Probability is a function $P: \mathcal{A} \to [0,1] \iff$
- $P(S) = 1$
- $0 \le P(E) \le 1$
- if $E_1, \dots, E_n \mid E_i \cap E_j = \emptyset \ \forall i,j$, $P(\bigcup_i E_i) = \sum_i P(E_i)$

$$1 = P(S) = \underbrace{P(S \cup \emptyset)}_{} = P(S) + P(\emptyset) \Rightarrow P(\emptyset) = 0$$

- $S \cup \emptyset = S$:
  - $S \cup \emptyset \subseteq S$
    $x \in S \cup \emptyset \to x \in S$ since $\emptyset$ is empty
  - $S \subseteq S \cup \emptyset$
    $x \in S \to x \in S \cup$ (any set)
- $S \cap \emptyset = \emptyset$:
  - $\emptyset \subseteq S \cap \emptyset$
    ↳ if it weren't true, $\emptyset$ would have an element inside → impossible
  - $S \cap \emptyset \subseteq \emptyset$
    ↳ to not be emtpy, $\emptyset$ must have an element that's in S, which is impossible

Conditional probability

$P(E|A) = P(E \cap A) P(A) \Rightarrow E$ is indipendent from $A \iff P(E|A) = P(E)$
↳ A makes no effect on E