BIOSTATISTICA

LEZIONE 1 03-03-19

Cos'è la statistica?
Essendo che si occupa dello studio di fenomeni collettivi che presentano una certa variabilità. I metodi statistici presuppongono lo studio di una variabile e la sua descrizione attraverso metodi matematici.

Noi studieremo i fenomeni ALEATORI che si distinguono da quelli DETERMINISTICI perché questi ultimi li possiamo descrivere con esattezza attraverso una formula matematica. Per quelli aleatori, cioè che hanno random, possiamo solo stimare una soluzione, ragioneremo sempre in termini di stima probabilistica.

Qualsiasi variabile che noi osserviamo la possiamo pensare come l'uscita di un sistema dinamico cioè qualcosa che proviene dall'elaborazione di un certo ingresso.
Un sistema è deterministico se si può determinare in modo univoco il valore delle variabili in uscita in funzione di quelle in ingresso.
Un sistema caotico è descritto da equazioni non lineari e possono avere grandi variazioni in uscita per una condizione infinitesima in ingresso.

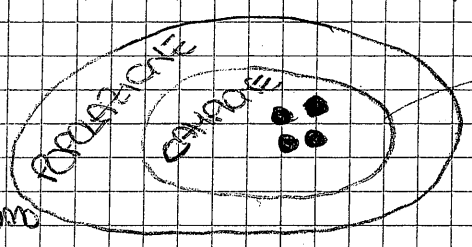
SISTEMI STOCASTICI / PROBABILISTICI
Non possiamo descriverli tramite un'equazione perché il sistema ha un infinito numero di gradi di libertà.
Esempi principali sono i fenomeni fisiologici.

$$H = -\sum_i p_i \log_2(p_i) \rightarrow \text{ENTROPIA DI SHANNON} \} \text{CURIOSITÀ}$$

Per studiare un fenomeno statisticamente esso deve avere una VARIABILITÀ altrimenti non ha senso.
La statistica DESCRITTIVA ci fornisce i metodi per sintetizzare una mole di dati grandi e piacere con pochi numeri. Per comodità si riducono i dati eseguendo in tabelle di frequenze. Poi si passa alla rappresentazione dell'istogramma delle frequenze.

Partiamo da un CAMPIONE e lo caratterizziamo attraverso la statistica descrittiva ma il nostro obiettivo è poter dire qualcosa sull'intera popolazione partendo da esso.

- POPOLAZIONE insieme più grande possibile di elementi
- CAMPIONE sottoinsieme della popolazione che studiamo



Questo processo è detto INFERENZA.

La statistica descrittiva (come detto sopra) è l'input della statistica inferenziale (fare affermazioni sulle caratteristiche di tutta la popolazione). Il risultato finale dell'inferenza è il calcolo della probabilità.

PROBABILITÀ
Numero dei casi favorevoli fratto il numero dei casi totali

$$P[i] = \lim_{N\to\infty} \frac{n_i}{N}$$

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$ FATTORE BINOMIALE

Il problema del compleanno
Probabilità che in un gruppo di n persone prese a caso due di esse siano nate lo stesso giorno.
Per n = 100 P(n) = 99,99997% anche se inizialmente sembrano molto basse. Vedrai risalto dopo.

- Formalmente:
Dato lo terno Ω (insieme), A (σ-algebra di parti di Ω), P (probabilità su A) allora la probabilità è un insieme di funzioni $P: A \to \mathbb{R}^+$ tale che $P(\Omega) = 1$ e se $\{A_n\}_n$ è una successione di elementi di A a due a due disgiunti allora
$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$

A è l'insieme di tutti i sottoinsiemi di Ω, tutte le combinazioni degli elementi di Ω
A è una sigma algebra perché posso fare due proprietà:
$\bigcup_{n=1} A_n \in A$ e $\bigcap_{n=1} A_n \in A$

- Es: LANCIO DEL DADO

$\Omega = \{1,2,3,4,5,6\}$ $A = P(\Omega)$
$P\{1\} = P\{2\} = \dots = P\{6\} = p$

$P(\Omega) = P[\{1\} \cup \{2\} \cup \dots \{6\}] = P\{1\} + P\{2\} + \dots + P\{6\} = 6p$

$P(\Omega) = 1 \Rightarrow 6p = 1 \Rightarrow p = \frac{1}{6}$

Assiomi della probabilità
- $P(A) \geq 0$
- $P(\Omega) = 1$
- $P(A \cup B) = P(A) + P(B)$ se $A \cap B = \emptyset$

Se $A \cap B \neq \emptyset \Rightarrow P[A \cup B] = P[A \cup B] = P[A] + P[B] - P[A \cap B]$

$P[A \cap B]$

UNIONE INSIEMISTICA = OR LOGICO
INTERSEZIONE INSIEMISTICA = AND LOGICO

PROBABILITÀ CONDIZIONATA
Ho già osservato eventi precedenti.

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

A evento da descrivere
B evento osservato

ANALISI
- semplice (studio di un caso, confronto tra due dati)
- dati appaiati (due campioni con medesimi soggetti)
- multivariata (n campioni)

LEZIONE 2 07-03-17

Seguiremo in generale il METODO SCIENTIFICO
Ipotesi → Esperimento → Misure → Risultato
- Ipotesi nulla H_0
- Ipotesi alternativa H_1

Se H_0 ha una probabilità troppo bassa, accetto H_1.
Ad ogni modo non possiamo eliminare l'incertezza, ci aiuta solo quantificando il sistema aleatorio.

TEORIA DELLA PROBABILITÀ
Formalizzazione matematica della probabilità

$$Pr\{A\} = \lim_{n \to \infty} \frac{n_A}{n}$$

- $Pr\{A\}$: evento da studiare
- n_A: numero di volte in cui si è verificato il valore A
- n: numero di prove effettuate
- frequenza relativa dell'evento

Questo vale nel caso di eventi EQUIPROBABILI (lancio dei dadi)

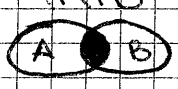
SPAZIO CAMPIONE
Insieme di tutte le possibili uscite considerando le combinazioni tra le variabili
EVENTO
Ciò a cui siamo interessati

TEORIA DEGLI INSIEMI
Gli eventi sono sottoinsieme dello spazio campione

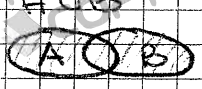
COMPLEMENTO



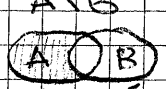
INTERSEZIONE $A \cap B$



UNIONE $A \cup B$



DIFFERENZA $A \setminus B$



$A \cap B^c$

$\cup$ e $\cap$ sono associative

DEFINIZIONE FORMALE (già vista)
CHIARIMENTI
- Ω spazio dei campioni, $\mathcal{A}$ famiglia di parti di Ω
- σ-algebra ha proprietà particolari quali essere chiusa rispetto alle operazioni che facciamo sui sottoinsiemi
- $P(\Omega)$ probabilità totale
- An successione di elementi disgiunti

PROPRIETÀ DEGLI EVENTI

- Mutua esclusività $A \cap B = \emptyset$ (INCOMPATIBILITÀ)
- Esaustività collettiva $A \cup B =$ SPAZIO CAMPIONE

Se gli eventi hanno entrambe le proprietà allora sono una partizione dello spazio campione

ISTOGRAMMA

In ascissa divido la variabile continua in intervalli
In ordinata identifico la frequenza relativa con cui si osservano gli eventi

$$f_E = \frac{n_E}{n} \qquad 0 \le f_E \le 1 \qquad f_\Omega = 1$$



Altri assiomi della probabilità

- $P(A^c) = 1 - P(A)$
- $P(A \cap B) = P(A) + P(B) - P(A \cup B)$ $\qquad P(A \cap B) \equiv P(AB)$
- $A \subset B \Rightarrow P(A) \le P(B)$

COEFFICIENTE BINOMIALE

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \quad \text{for } 0 \le K \le n$$

Quantifica tutti i possibili sottoinsiemi di k elementi quando ho un insieme più grande di n elementi

Ci serve per generalizzare una formula:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) + \dots + (-1)^{(r+1)} \sum_{i_1<i_2<\dots<i_r} P(E_{i_1}E_{i_2}\dots E_{i_r}) + \dots$$

$$\dots = (-1)^{(n+1)} P(E_1 E_2 \dots E_n)$$

Se due eventi sono INDIPENDENTI, cioè non si influenzano tra loro,

$$P[A \cap B] = P[A \in B] = P[A] \cdot P[B]$$

ESEMPI

① LANCIO DEL DADO: qual è la probabilità che esca un numero inferiore o uguale a 2 oppure un numero pari?

$A = \{1,2\}$ $\quad P[A] = P[1] + P[2] = \frac{1}{3}$

$B = \{2,4,6\}$ $\quad P[B] = P[2] + P[4] + P[6] = \frac{1}{2}$

$C = A \cap B = \{2\}$ $\quad P[A \cap B] = P[2] = \frac{1}{6}$

$P[A \cup B] = P[A] + P[B] - P[A \cap B] = \frac{2}{3}$

② LANCIO DADI: qual è la probabilità che esca 2 da uno e 3 dall'altro?

Sono eventi indipendenti quindi

$$P[2 \text{ e } 3] = P[2] \cdot P[3] = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

③ LANCIO DADO: qual è la probabilità che lanciando due volte un dado esca almeno un 6?

$$P[6] = \frac{1}{6} \qquad P[\text{not } 6] = \frac{5}{6}$$

2° LANCIO:

$$P[6 \text{ e } 6] = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

$$P[6 \text{ e not } 6] = P[\text{not } 6 \text{ e } 6] = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{6} \cdot \frac{1}{6} = \frac{5}{36}$$

$$P[\text{not } 6 \text{ e not } 6] = \frac{5}{6} \cdot \frac{5}{6} = \frac{25}{36}$$

$$P[\text{almeno un } 6] = \frac{1}{36} + \frac{5}{36} + \frac{5}{36} = 1 - \frac{25}{36} = \frac{11}{36}$$

PROBABILITÀ CONDIZIONATA DI A DA B

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

se due eventi non sono INDIPENDENTI

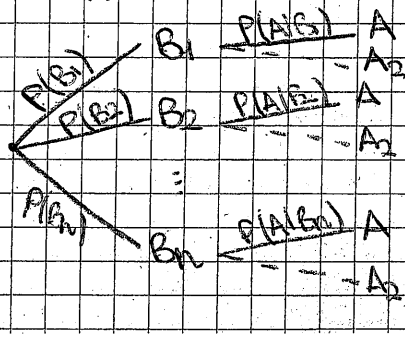
$$P[A \cap B] = P[A \text{ e } B] = P[A] \cdot P[B|A]$$

TEOREMA DELLA PROBABILITÀ TOTALE

Nello spazio campione abbiamo n insiemi di eventi B che sono mutuamente esclusivi e collettivamente esaustivi cioè non si sovrappongono e la loro unione ci dà tutto lo spazio campione. Allora se ho un nuovo evento A

$$P[A] = \sum_i P[A \cap B_i] = \sum_i P[B_i] P[A|B_i]$$

Rappresentazione ad albero



$$P[A] = P[B_1] P[A|B_1] + \dots + P[B_n] P[A|B_n]$$

$$= P[A|B] \cdot P[B] + P[A|\bar{B}] \cdot P[\bar{B}]$$

nel caso più semplice