STATISTICA

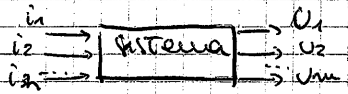
disciplina che si occupa dell'elaborazione dei risultati dell'osservazione di uno o più caratteri posseduti dagli elementi di un insieme determinato, con l'intento di esprimere un giudizio su alcuni aspetti della realtà di interesse che, in quanto riferita ad un insieme e non ai singoli elementi che lo compongono, viene chiamata FENOMENO COLLETTIVO.

FENOMENI
- deterministici → dalle leggi di evoluzione di un sistema e note le cond. iniziali riesco con buona approssimazione a prevedere il futuro del sistema
- aleatori → non possono essere previsti con precisione → variabili aleatorie (stocastiche - casuali)

Concezione frequentistica: probabilità viene assegnata sulla base di risultati di un esperimento ripetuto molte volte nelle stesse condizioni

SISTEMI DINAMICI

↳ caratterizzati da:



- variabili di ingresso I
- variabili di uscita U
- variabili di stato (che definiscono configurazione interna)

↓ in base al comportamento →

Sistema deterministico: è possibile determinare (in modo univoco) il valore delle variabili di uscita in funzione dei valori delle variabili in ingresso

determinismo ≠ predicibilità

! molte volte, uscite fortemente sensibili a variazioni ingressi ("butterfly effect") (sopra piccoli)

Caratteristico dei sistemi con legge NON LINEARE

ES. $x_{n+1} = r x_n (1 - x_n)$ $\quad \varepsilon(t) = \varepsilon_0 e^{\lambda t}$ diverge (t.i.t. casi in cui $\lambda > 0$) converge (s.i.t. deterministici $\lambda < 0$)

FENOMENO MOSTRA UN REGIME DI CAOS DETERMINISTICO quando:

- dipendenza molto sensibile dalle condizioni iniziali
- incertezza iniziale che cresce esponenzialmente con il tempo $\varepsilon(t) = \varepsilon_0 e^{\lambda t}$
- ⇒ IMPREDICIBILITÀ a lungo termine della sua evoluzione futura

SISTEMA MECCANICO E PROBABILISTICO

mass è individuabile corrispondente univoca tra sollecitazione e comportamento, è necessaria descrizione PROBABILISTICA del sistema

la ragion d'essere della STATISTICA è la presenza di un certo livello di variabilità nei dati

→ al variare dell'unità STATISTICA U entro una certa popolazione P = {U} variano certe caratteristiche misurate su di esse

OBBIETTIVO: quantificare la variabilità e incertezza → con lo scopo di associarci probabilità

POPOLAZIONE: ENTITÀ associato al fenomeno aleatorio, comprende TUTTI i casi di esistenza al mondo; TUTTI quelli che hanno una certa caratteristica.

NO interessarli tutti sarebbe impossibile → uso un SOTTOINSIEME, ovvero un CAMPIONE

! Per effettuare analisi → VARIABILITÀ del campione che riflette variabilità nella POPOLAZIONE !

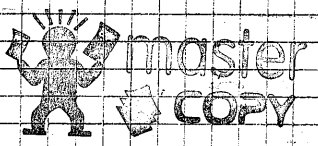
STATISTICA DESCRITTIVA → descrive campione

metodi per sintetizzare, con opportuna grafica, le caratteristiche salienti dei fenomeni collettivi

la descrizione passa attraverso le fasi di formazione del dato statistico e del trattamento matematico dello stesso

Solitamente uso TABELLA DELLE FREQUENZE (se dominio discreto o continuo)

GRAFICA molto utilizzo ISTOGRAMMA

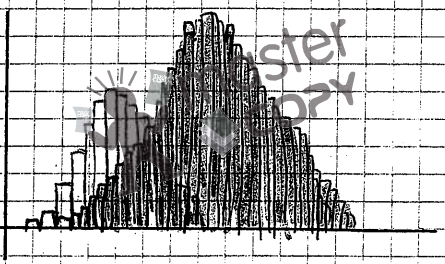


utilizzo una serie di QUANTIFICATORI:

- DI POSIZIONE:

MEDIA: $\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$

MODA: PUNTO DI MAX della distribuzione



MEDIANA: valore sotto al quale cadono la metà dei valori campionati. Si dispongono in ordine crescente, e si prende quello che occupa la posizione centrale (N dispari) o la media dei due valori in posizione centrale (N pari)

MISURE DI DISPERSIONE

- VARIANZA $s^2 = \frac{\sum_i (x_i - \bar{x})^2}{N-1}$

- DEVIAZIONE STANDARD $s \quad (+\sqrt{s^2})$

- RANGE $R = x_{max} - x_{min}$

MISURE DI FORMA

- SKEWNESS (coefficiente di asimmetria) $\sum_i \left(\frac{x_i - \bar{x}}{\sigma}\right)^3 \frac{1}{N}$
  - >0: coda a dx
  - <0: coda a sx
  - $=0$ SIMMETRIA

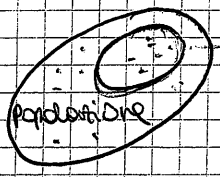
- CURTOSI misura quanto distribuzione è appuntita
  $\sum_i \left(\frac{x_i - \bar{x}}{\sigma}\right)^4$
  - >3 molto appuntita
  - <3 molto appiattita

FENOMENO COLLETTIVO: riguarda una grande collezione di elementi (POPOLAZIONE)

elementi della popolazione: UNITÀ STATISTICHE

Popolazione: variabile aleatoria

Campione: realizzazione di una variabile aleatoria



STATISTICA INFERENZIALE

"inferenza" = argomentazione, ragionare
- da particolare → generale

Sapendo qualcosa dal CAMPIONE cerca di dire qualcosa sulla POPOLAZIONE

→ COMPIERE inferenza su tutta la popolazione a partire dal campione

Quindi:

→ il campione viene visto come descritto (st. descrittiva), dopodiché le osservazioni ottenute sul campione vengono usate per fare affermazioni sulle caratteristiche di tutta la popolazione (st. inferenziale)

Inferenza è quindi un processo d'astrazione, e l'incertezza viene misurata in termini probabilistici

La PROBABILITÀ è il fondamento logico per fare riferimento alla popolazione

↳ PROBABILITÀ

NON definizione:

SE (E solo SE) eventi equiprobabili,

$$Pr\{A\} = \lim_{N \to +\infty} \frac{n_a}{N}$$

Es. Probabilità di fare full-house a poker?

5 carte da un mazzo da 52 (13 carte ♥ semi)

→ campionamento senza reinserimento (ordine non conta)

Coeff. binomiale → $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ con $0 \leq k \leq n$

Assumo ogni elemento EQUIPROBABILE

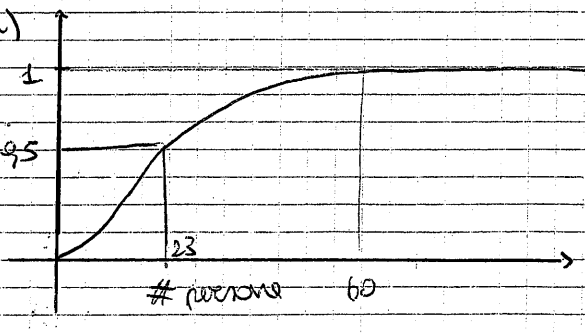
$$\rightarrow P = \frac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}} = 0{,}001441$$

Es. The birthday problem

→ calcolare la prob. che n persone scelte random ci sia almeno una coppia di persone nate lo stesso giorno

$$p(n) = 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \dots \times \left(1 - \frac{n-1}{365}\right) = \frac{365 \times 364 \times \dots \times (365 - n + 1)}{365^n}$$

$$\bar{p}(n) = 1 - p(n)$$



ES. Monty Hall (problema delle tre porte)

→ se cambi porta $\frac{1}{3}$ prob di vincere in +

ES. Giocatore d'azzardo (gambler) che inizia con 1€ soldi ex ogni giocata successiva vince (o perde) 1€

→ obiettivo: arrivare a N€

$$P_i = \begin{cases} \frac{1-(\frac{q}{p})^i}{1-(\frac{q}{p})^N} & \text{se } p \neq q \\ \frac{i}{N} & \text{se } p = q = 0{,}5 \end{cases}$$

p = prob del banco

master COPY

Probabilità è calcolo primitivo, per definirla devo introdurre alcuni ingredienti:

- esperimento casuale
- evento
- spazio degli eventi B

master COPY

! probabilità è una funzione matematica su B

il codominio è $[0,1]$

Def

esperimento: serie di atti sperimentali in cui osserviamo variabile che rappresenta fenomeno collettivo, e rappresenta, con un certo verificabilità, un FENOMENO ALEATORIO

uscita: risultato di ciò che osserviamo

spazio campione: unione di tutte le possibili uscite che possono essere osservate durante esperimento

evento: sottospazio di spazio campione

! in teoria degli insiemi, $A \setminus B$ (sottrazione) è $A \cap \bar{B}$ (complementare)

$U \in \cap$ sono omologhe

master COPY

PREMESSO:

per definire prob. ho bisogno di 3 elementi:

- dominio: è σ-algebra, ovvero algebra con qualche proprietà in più
- spazio campione:
- funzione di probabilità su σ-algebra (ne compongono tutti i possibili sottoinsiemi (inclusi $\emptyset$ e Ω))