Probabilità e Statistica

Alex Narder

October 7, 2022

Contents

1 Introduzione

Per estrarre informazioni dai dati dobbiamo processarli adeguatamente;
La statistica ci fornisce quegli stumenti per estrarre le informazioni dai dati, uno statistico sa:

- combinare informazioni di diverso tipo
- valutare l'affidabilità
- sintetizzare e presentare molti dati
- costruire modelli
- calcolare previsioni e formulare ipotesi di decisione

La statistica non è l'unico strumento usato per analizzare i dati, ma è quello più adatto in presenza di incertezze.
Anche l'informatica svolge un ruolo fondamentale nel salvataggio e nella gestione dei dati.

2 Termini della statistica

- La **popolazione di riferimento**, è un insieme di elementi chiamati **unità statistiche**.
- I **dati** sono il risultato di rilevazioni o misurazioni.
- La **statistica** ci permette di estrarre le informazioni dai dati, generando nuove conoscenze o ipotesi di decisione.
- Ogni caratteristica rilevata sulle unità statistiche si chiama **variabile** e i dati corrispondenti a ogni variabile sono le **realizzazioni**.
- Se le variabili non sono rilevate su tutte le statistiche, il sottoinsieme della popolazione oggetto della rilevazione è chiamato il **campione**.

3 Tipi di variabili

Una variabile è **qualitativa** o **categorica** quando i suoi possibili valori o modalità prendono la forma di aggettivi o di altre espressioni verbali.
Le variabili qualitative possono essere:

- Sconnesse

- Ordinali

Le variabili categoriche possono essere:

- Discrete o intere

- Continue o reali

4 Calcolo della Probabilità

Perché l'inferenza porti a risultati sensati, bisogna che sia noto il legame fra popolazione e campione.

• Il calcolo delle probabilità fornisce i modelli matematici utili per descrivere la relazione fra campione e popolazione.
• Il calcolo delle probabilità è lo strumento necessario per l'inferenza. Permette di quantificare gli errori che commettiamo nel passaggio dal particolare (campione) al generale (popolazione).

5 Esempi di probabilità

- Lancio di 1 dado (normale), qual è la probabilità che esca 6? 1/6

- Stessa domanda ma il dado ha 20 facce. 1/20

Qual è la probabilità che il risultato sia 2 o 6, sempre in riferimento della prima domanda? La probabilità che esca è 2/6.

- Pensiamo di avere delle antenne in linea, se due antenne di fila non funzionano allora il sistema non funziona. Se una antenna è rotta ma quella successiva no il sistema funziona lo stesso perchè il segnale passa alla successiva.
Sapendo che n antenne sono difettose, qual è la probabilità che il sistema funzioni?
$m <= n$

Se n = 4 e m = 2:
1 $\rightarrow$ se l'antenna funziona
0 $\rightarrow$ se l'antenna non funziona

n rappresenta il numero delle antenne, mentre m è il nunero delle antenne rotte, quindi nel momento in cui ho un gruppo di antenne 1100 sò che le prime 2 funzionano, mentre le ultime 2 non funzionano e sono consecutive, quindi il sistema non funziona. Nel secondo caso invece le antenne non funzionanti non sono consecitive, 0110 quindi è un sistema che funziona. Vado a prendere solo i casi in cui so che solo 2 non funzionano, perciò 1111, 0000 oppure 1110 ecc ecc non vado a considerarli. $\rightarrow$

1100 non funziona
0110 funziona
1010 funziona
0011 non funziona
0101 funziona
1001 non funziona

probabilità richiesta: 3/6 = 1/2

totale: 6, quindi sono 3 funzionanti

6 Contare le probabilità

6.1 Principio fondamentale del calcolo combinatorio:

Se una scelta può essere fatta in m1 modi diversi e un'altra scelta può essere fatta in m2 modi diversi, allora esistono in totale

m1 * m2 possibilità di scelta

esempio:
10 cavalieri e 12 dame partecipano a un ballo. Ci sono $10 \times 12 = 120$ possibili coppie danzanti.

6.2 Principio fondamentale generalizzato:

Se ciascuna delle r scelte successive può essere fatta in m_i modi, allora esistono rispettivamente in totale:

$$\prod_{i=1}^{r} m_i = m_1 * ... * m_r$$

possibilità di scelta esempio:
Una commissione parlamentare deve essere composta da un membro del partito A, che conta 10 rappresentanti, da un membro del partito B, che conta 15 rappresentanti, e da un membro del partito C, che conta 2 rappresentanti. Ci sono in totale $10 \times 15 \times 2 = 300$ possibili commissioni parlamentari.

7 Disposizioni

Consideriamo un insieme di n elementi. Una **disposizione** di r di essi è una scelta ordinata di r elementi tra quegli n.

• Si distinguono le disposizioni con **ripetizione** da quelle **semplici** (o **senza ripetizione**), a seconda che uno stesso elemento possa essere scelto più di una volta:

• Le disposizioni con **ripetizione** di n elementi presi r alla volta sono in numero di

$$\prod_{i=1}^{r} n = n^r$$

esempio:
le parole lunghe due lettere che si possono comporre con le lettere I, L, A sono $3^2 = 9$: II, IL, IA, LI, LL, LA, AI, AL, AA.

• Le **disposizioni semplici o senza ripetizione** di n elementi presi r alla volta sono in numero di

$$n * (n-1) \ ... \ *(n-r+1)$$

esempio:
le parole di due lettere diverse che si possono comporre con le lettere I, L, A sono $3 \times 2 = 6$: IL, IA, LI, LA, AI, AL.

8 Campionamento da un'urna

Il **campionamento casuale da un'urna** è un'estrazione di palle da un'urna. Può essere fatto **con o senza reintroduzione**.

- Per **casuale** si intende dire che prima di ogni estrazione l'urna viene mescolata appropriatamente per essere riportata a una condizione di irriconoscibilità delle palle. Un'operazione del genere viene fatta, ad esempio, per le estrazioni del lotto.

- La **reintroduzione** fa invece riferimento al fatto di rimettere nell'urna ciascuna palla subito dopo averla estratta e averne registrate le caratteristiche di interesse, per esempio il suo numero o il suo colore.

Quindi:
- Se un'urna contiene n palle distinguibili e r palle vengono estratte con reintroduzione, ci sono n^r possibilità di estrazione.
- Se un'urna contiene n palle distinguibili e r palle vengono estratte senza reintroduzione, le estrazioni possibili sono $n * (n-1) \ldots *(n-r+1)$.

9 Permutazioni

Le disposizioni semplici di n elementi presi n alla volta si chiamano anche **permutazioni** perché rappresentano tutti i modi in cui n elementi possono essere ordinati in fila. Esse sono in numero di

$$n * (n-1) \ldots * 2 * 1 =: n!$$

Il simbolo speciale n! che rappresenta questa quantità si legge n **fattoriale**.

Esempio 1:
Le permutazioni delle lettere I, L, A sono $3 \times 2 \times 1 = 3! = 6$:
ILA, IAL, LIA, LAI, AIL, ALI.

Esempio 2:
Supponiamo di fare due file, i 6 maschietti a destra e le 4 femminucce a sinistra. Ci sono 6! possibili file di maschietti e 4! file di femminucce possibili. Quindi, dal principio fondamentale del calcolo combinatorio, in tutto ci sono $6! \times 4! = 17280$ possibili file.

10 Combinazioni

In generale; Un sottoinsieme di numerosità r scelto da un insieme con n elementi si chiama **combinazione** di n elementi presi r alla volta.
Il numero di combinazioni di n elementi r alla volta è

$$\frac{n * (n-1) \ldots (n-r+1)}{r!} =: \binom{n}{r}$$

esempio:
La professoressa di matematica interroga ogni lunedì 10 studenti da una

classe di 25. Esistono per lei $\binom{25}{10} = 3.268.760$ possibilità di scelta.

$\rightarrow$ Il nome coefficente binomiale deriva dalla seguente espressione:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$$

detta formula del **Binomio di Newton**.

Esempio:

$$(a+b)^2 = \binom{2}{0} a^2 b^0 + \binom{2}{1} ab + \binom{2}{2} a^0 b^2 = a^2 + 2ab + b^2$$

11 Fenomeni Aleatori

- La logica del **certo** è la logica della teoria degli insiemi e del calcolo su proposizioni (o eventi) che possono assumere il valore di vero o falso in modo **deterministico**.
- Il **calcolo delle probabilità** è invece la logica dell'**incerto**. Si usa per ragionare sui possibili risultati di un **fenomeno aleatorio** o **casuale**, del quale cioè non si può prevedere con certezza l'esito.
- Si può pensare al termine **aleatorio** come l'opposto di **deterministico**.

11.1 Esempi di fenomeni aleatori:

- Il lancio di un dado
- Il lancio di una stessa moneta 4 volte
- La classificazione di 10 pezzi prodotti da una macchina in conformi o non conformi alle specifiche di progetto
- L'estrazione di una mano di poker, cioè un insieme di cinque carte, da un mazzo di 52
- L'osservazione del tempo di guasto [min] di un circuito elettrico formato da tre resistenze in serie

11.2 Esempi dell'incertezza in informatica:

- Il tempo o lo spazio su un disco richiesti per l'installazione di un software
- Il numero di difetti di un nuovo software
- La memoria richiesta per processare un programma
- Il tempo richiesto per una stampa o il numero di lavori in coda di stampa prima di questo
- Il momento nel quale un virus attacca un sistema o il numero di file e directory infetti

12 Spazio campionario, risultati ed eventi

12.1 Spazio campionario

Lo **spazio campionario** è l'insieme di tutti i possibili risultati di un fenomeno aleatorio. Che viene rappresentato con Ω. Un generico **risultato** è un elemento dello spazio campionario, $\omega \in \Omega$.

Esempi di spazio campionario:

- Lancio di un dado: $\Omega = 1, 2, 3, 4, 5, 6$
- Lancio di una stessa moneta quattro volte: $\Omega =$ le sedici possibili sequenze di quattro dei simboli T e C, dove T indica "testa" e C indica "croce"
- La classificazione di 10 pezzi con 2 possibili risultati C e N , dove C indica 'conforme' e N indica 'non conforme': $\Omega =$ le 210 possibili sequenze di dieci dei simboli C e N
- Una mano di poker: $\Omega =$ i (525) possibili sottoinsiemi delle 52 carte.
- Il tempo di guasto del circuito elettrico:$\Omega = R+ := [0, \infty)$, cioè tutti i numeri non negativi, visto che il tempo di guasto è un numero non negativo

- I livelli massimi giornalieri di polveri nel Gennaio 2015: Ω = tutte le possibili sequenze di 31 numeri non negativi (la maggior parte contenuti tra 10 e 350)

12.2 Eventi

Un **evento** è un sottoinsieme A $\subset \Omega$. Una volta che il fenomeno aleatorio di interesse è stato osservato si può dire se un qualsiasi evento A sia vero o falso. Quando un evento è vero, si dice che si è **realizzato o verificato**. I possibili risultati ω, visti come singoletti, cioè insiemi contenenti un solo elemento, sono anch'essi eventi, detti eventi **elementari**. Ω viene anche chiamato l'**evento certo**, perché sicuramente si verificherà.
Eventi impossibili $\rightarrow \emptyset$

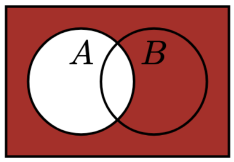
Esempi di eventi

- Il dado dà un punteggio superiore a quattro: $A = 5, 6$
- Otteniamo almeno tre teste sui quattro lanci:
$$A = TTTC, TTCT, TCTT, CTTT, TTTT$$
- Tutti i pezzi sono conformi:
$$A = CCCCCCCCCC$$
(questo è anche un singoletto)
- Si ottiene un poker: l'evento A di interesse è dato da tutte le possibili mani contenenti un poker, che sono in numero di $13 * 48$ perché 13 sono i possibili poker e 48 sono, per ogni dato poker, i modi di scegliere la quinta carta.
- Il circuito ha una durata di meno di 50 ore: $A = [0, 50)$

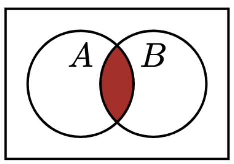
13 operazioni sugli eventi e Diagrammi di Venn - partizioni

• La negazione o **complementare** di un evento A, viene indicato con $\hat{A}$, è l'evento che è vero quando A è falso ed è falso quando A è vero.

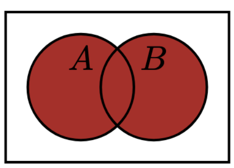
La negazione dell'evento certo è l'**evento impossibile**: $\hat{\Omega} = \emptyset$ (evento impossibile = insieme vuoto).



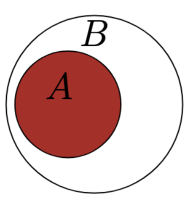
• L'**intersezione** di due eventi A e B, indicata con $A \cap B$, è l'evento che è vero quando sia A che B sono veri.



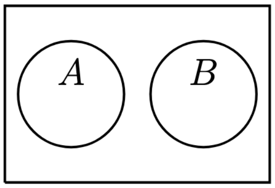
• L'**unione** di due eventi A e B, indicata con $A \cup B$, è l'evento che è vero quando o A o B o entrambi sono veri.



• L'evento A è **incluso** nell'evento B, in simboli $A \subset B$, se il verificarsi di A implica il verificarsi di B.

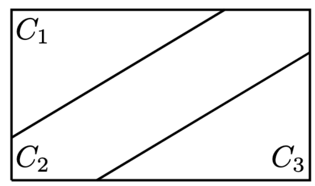


• Due eventi A e B, si dicono **incompatibili o disgiunti**, se non è possibile che siano entrambi veri, cioè se $A \cap B = \emptyset$



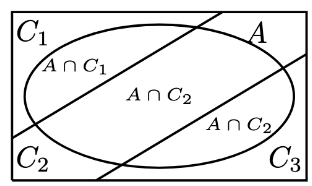
• Una famiglia di eventi si dice una **partizione** dello spazio campionario se ogni coppia di insiemi della famiglia ha intersezione vuota e l'unione di tutti i componenti della famiglia è Ω stesso

- Una partizione può essere **finita**, ad esempio, $\{C_1, C_2, C_3\}$ è una partizione di 3 elementi se $C_1 \cap C_1 = C_1 \cap C_3 = C_2 \cap C_3 = \emptyset$ e $C_1 \cup C_2 \cup C_3 = \Omega$.



- Un qualsiasi evento A si può scrivere come unione delle sue intersezioni con gli elementi di una partizione:

$$A = (A \cap C_1) \cup (A \cap C_2) \cup (A \cap C_3)$$



- In generale, si può pensare a una partizione **numerabile** $C_1, C_2, \ldots$:

$$C_i \cap C_j = \emptyset \quad \forall i, j; \qquad \bigcup_{i=1}^{\infty} C_i = \Omega$$

Scrivendo un qualsiasi evento A come unione numerabile delle sue intersezioni con gli elementi di una partizione:

$$A = \bigcup_{i=1}^{\infty} (A \cap C_i)$$

13.1 esempio:

fenomeno aleatorio: lancio di un dado.
eventi: Ne consideriamo i seguenti:

$A = \{5, 6\} \rightarrow$ il risultatodel lancio è superiore a 4
$B = \{2, 4, 6\} \rightarrow$ il risultato del lancio è pari

allora:

$A \cap B = \{6\} \rightarrow$ il risultato del lancio è pari e superiore a 4
$A \cup B = \{2, 4, 5, 6\} \rightarrow$ il risultato del lancio è pari oppure superiore a 4

partizione: numeri divisibili per 3 e non:

$C_1 = \{3, 6\}$
$C_2 = \{1, 2, 4, 5\}$

allora:

$A = (A \cap C_1) \cup (A \cap C_2) = \{6\} \cup \{5\}$
$B = (B \cap C_1) \cup (B \cap C_2) = \{6\} \cup \{2, 4\}$

14 Definizione assiomatica di probabilità

Formalmente, la **probabilità** è una funzione che assegna ad ogni evento di uno spazio campionario un valore in R+, ossia un numero non negativo, e deve sodisfare i seguenti assiomi:

- **Positività**: $0 <= P[A] <= 1$

- **Normalizzazione**: $P[\Omega] = 1$

- **Aditività**: Se $A_1, A_2, ...$ è una sequenza di eventi incompatibili, cioè se $A_i \cap A_j = \emptyset$ $\forall i \neq j$, allora:

$$\mathbb{P}\left[\bigcup_{n=1}^{\infty} A_n\right] = \sum_{n=1}^{\infty} \mathbb{P}[A_n]$$