Probabilità e Statistica[1]

Isadora Antoniano-Villalobos
isadora.antoniano@unive.it

Laurea in Informatica
(Data science/ Tecnologie e scienze dell'informazione)
Università Ca' Foscari di Venezia

Anno academico 2023/2024

[1] Materiale didattico redatto da: Isadora Antoniano-Villalobos & Federica Giummolè

Probabilità condizionata e indipendenza

Probabilità condizionata

Probabilità condizionata

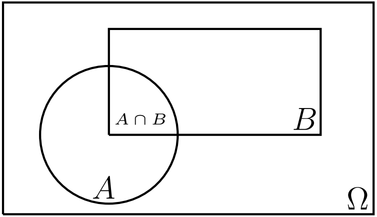
Sia B un evento di probabilità positiva. La **probabilità condizionata** dell'evento A dato l'evento B è

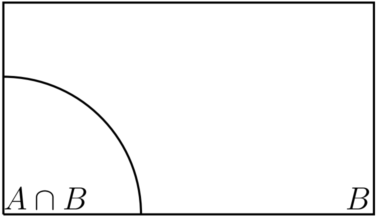
$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}, \quad \mathbb{P}[B] > 0.$$

- $\mathbb{P}[A|B]$ è anche chiamata la **probabilità subordinata** o **condizionale** di A subordinatamente a B.
  ☞ Da notare l'uso della sbarra verticale $|$ che collega l'evento condizionato A e l'evento condizionante B.
- $\mathbb{P}[A|B]$ rappresenta la probabilità di A valutata in presenza dell'informazione aggiuntiva che B si verifichi.

Probabilità condizionata

- Intuitivamente, si restringe il campo delle possibilità non alla totalità dei possibili risultati Ω ma ad un suo sottoinsieme proprio $B \subset \Omega$.



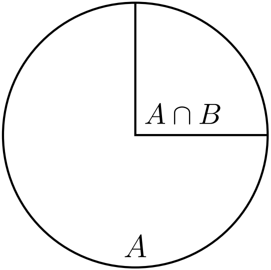


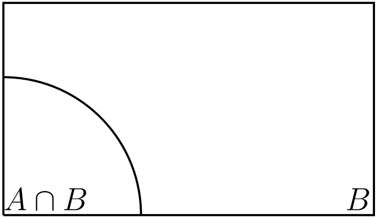
Prima del esperimento, si sa soltanto che B potrebbe avverarsi: $0 \leq \mathbb{P}[B] \leq 1$.
Per qualsiasi evento $A \subset \Omega$, $0 \leq \mathbb{P}[A \cap B] \leq \mathbb{P}[B]$.

Dopo l'esperimento, si sa con certezza che B è vero: $\mathbb{P}[B|B] = \frac{\mathbb{P}[B]}{\mathbb{P}[B]} = 1$.
Per qualsiasi evento $A \subset \Omega$, $0 \leq \mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \leq 1$.

Probabilità condizionata

Attenzione! $\mathbb{P}[A|B]$ e $\mathbb{P}[B|A]$ significano cose molto diverse:





$\mathbb{P}[B|A] = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[A]}$ si riferisce all'incertezza residua su B dopo il resultato del esperimento che ha reso A vero.

$\mathbb{P}[A|B] = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]}$ si riferisce all'incertezza residua su A dopo il resultato del esperimento che ha reso B vero.

Probabilità condizionata

- $\mathbb{P}[A|B]$ e $\mathbb{P}\left[\bar{A}|B\right]$ sono in relazione diretta:

$$\mathbb{P}\left[\bar{A}|B\right] = \frac{\mathbb{P}\left[\bar{A} \cap B\right]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B] - \mathbb{P}[A \cap B]}{\mathbb{P}[B]} = 1 - \mathbb{P}[A|B],$$

Le probabilità **condizionate allo stesso evento** obbediscono alle leggi (assiomi) della probabilità!

- $\mathbb{P}[A|B]$ e $\mathbb{P}\left[A|\bar{B}\right]$ non sono in relazione diretta: $\mathbb{P}[A|B]$ potrebbe essere uguale a $1 - \mathbb{P}\left[A|\bar{B}\right]$ in qualche esempio, ma in generale non è così.

Probabilità condizionata

➔ **Esempio:** Si ricordi l'esempio dell'urna composta da quattro palle bianche numerate da 1 a 4 e tre palle nere numerate da 1 a 3, dalla quale si campiona casualmente una palla, con eventi:

$$\begin{aligned} B &= \text{"viene estratta una palla bianca"}; \\ N &= \text{"viene estratta una palla nera"} \\ C_i &= \text{"viene estratto il numero } i\text{"}, i = 1, 2, 3, 4. \end{aligned}$$

Si valutino le probabilità condizionate che la palla estratta abbia il numero 1 dato che è bianca, che abbia il numero 1 dato che è nera e che la palla estratta sia nera dato che ha il numero 1.

Probabilità condizionata

Formalmente:

$$\mathbb{P}[C_1|B] = \frac{\mathbb{P}[C_1 \cap B]}{\mathbb{P}[B]} = \frac{1/7}{4/7} = 1/4$$
$$\mathbb{P}[C_1|N] = \frac{\mathbb{P}[C_1 \cap N]}{\mathbb{P}[N]} = \frac{1/7}{3/7} = 1/3$$
$$\mathbb{P}[N|C_1] = \frac{\mathbb{P}[N \cap C_1]}{\mathbb{P}[C_1]} = \frac{1/7}{2/7} = 1/2.$$

- $\mathbb{P}[N|C_1] \neq \mathbb{P}[C_1|N]$.
- Invece, per esempio, $\mathbb{P}[N|C_1] = 1 - \mathbb{P}[B|C_1] = 1/2$, $(N = \bar{B})$.

La formula delle probabilità composte

- La definizione di probabilità condizionata si può anche usare come formula pratica per la fattorizzazione della probabilità di un'intersezione, sempre che $\mathbb{P}[A|B]$ sia ben definita:

$$\mathbb{P}[A \cap B] = \mathbb{P}[A|B]\,\mathbb{P}[B],$$

- Questa formula si generalizza ad un qualsiasi numero di eventi $A_1, \ldots, A_n$ e viene anche chiamata la **formula delle probabilità composte**.

Formula delle probabilità composte

$$\begin{aligned}\mathbb{P}[A_1 \cap A_2 \cap \cdots \cap A_n] =& \mathbb{P}[A_1]\,\mathbb{P}[A_2|A_1]\,\mathbb{P}[A_3|A_1 \cap A_2]\ldots \\ & \ldots \mathbb{P}[A_n|A_1 \cap \cdots \cap A_{n-1}].\end{aligned}$$

La formula delle probabilità composste

➔ **Esempio:** Si consideri ora l'esperimento che consiste nell'estrazione di 3 palline senza reinserimento dalla solita urna. Qual è la probabilità che le prime due siano bianche e la terza nera?

Risulta utile considerare i seguenti eventi:

$B_i =$ "pallina bianca all'i-esima estrazione"
$N_i =$ "pallina nera all'i-esima estrazione"

Applicando la forula delle probabilità composte:

$$\begin{aligned}\mathbb{P}\left[B_1 \cap B_2 \cap N_3\right] &= \mathbb{P}\left[B_1\right]\mathbb{P}\left[B_2|B_1\right]\mathbb{P}\left[N_3|B_1 \cap B_2\right] \\ &= \frac{4}{7}\frac{3}{6}\frac{3}{5} = \frac{6}{35}.\end{aligned}$$

☞ Qual è la probabilità che siano tutte tre nere?

Eventi indipendenti

2 eventi indipendenti

Si dice che A e B sono **indipendenti** nel caso particolare in cui

$$\mathbb{P}\left[A|B\right] = \mathbb{P}\left[A\right].$$

In questo caso, vale anche $\mathbb{P}\left[A|\bar{B}\right] = \mathbb{P}\left[A\right]$

- Se A e B sono indipendenti, si ha allora

$$\mathbb{P}\left[A \cap B\right] = \mathbb{P}\left[A\right]\mathbb{P}\left[B\right]$$

che può anche essere presa come **definizione di eventi indipendenti**.

Eventi indipendenti

- La definizione si può estendere a più di 2 eventi.

n eventi (reciprocamente) indipendenti

Gli eventi $A_1, \ldots, A_n$, si dicono (reciprocamente) **indipendenti** se comunque si prendono $k > 1$ di essi, si ha

$$\mathbb{P}\left[A_{i_1} \cap \cdots \cap A_{i_k}\right] = \mathbb{P}\left[A_{i_1}\right] \ldots \mathbb{P}\left[A_{i_k}\right].$$

- La definizione si può estendere ulteriormente a una collezione infinita di eventi $A_1, \ldots, A_n, \ldots$, che saranno indipendenti se ogni sottoinsieme finito di essi è formato da eventi indipendenti.

➔ **Esempio:** Tornando alla urna, se le estrazioni dell'esempio al lucido precedente si effettuano con reinserimento, allora i tre eventi B_1, B_2 e N_3 sono indipendenti e si ha:

$$\mathbb{P}\left[B_1 \cap B_2 \cap N_3\right] = \mathbb{P}\left[B_1\right] \mathbb{P}\left[B_2\right] \mathbb{P}\left[N_3\right] = \frac{4}{7}\,\frac{4}{7}\,\frac{3}{7} = \frac{48}{147}.$$

Eventi indipendenti

- Eventi indipendenti e eventi disgiunti sono cose molto diverse. Due eventi sono disgiunti o meno a prescindere dalle loro probabilità. Due eventi disgiunti possono essere indipendenti soltanto se uno di essi ha probabilità zero.
- **Esempio:** Si consideri l'esperimento di lanciare un dado equo due volte. Si definiscano i seguenti eventi:

  $A =$ "la somma dei dadi è 6"
  $B =$"la somma dei dadi è 7"
  $C =$"il primo dado dà 4"

  Si ha
  $$\mathbb{P}[A] = \frac{5}{36}, \quad \mathbb{P}[B] = \frac{1}{6}, \quad \mathbb{P}[C] = \frac{1}{6}.$$

Eventi indipendenti

- Poiché
$$\mathbb{P}[A \cap C] = \mathbb{P}[(4,2)] = \frac{1}{36} \ ,$$
allora A e C non sono indipendenti.
- B e C sono invece indipendenti, ma non disgiunti. Infatti,
$$\mathbb{P}[B \cap C] = \frac{1}{36} = \mathbb{P}[B]\,\mathbb{P}[C] \, .$$
- Infine, A e B sono disgiunti ma non indipendenti:
$$\mathbb{P}[A \cap B] = \mathbb{P}[\emptyset] = 0.$$

Esempio: Sistemi in serie

Un sistema formato da n componenti separati si dice **in serie** se funziona quando tutti gli n componenti funzionano.

$$\boxed{in} \longrightarrow \boxed{C_1} \longrightarrow \boxed{C_2} \longrightarrow \boxed{\cdots} \longrightarrow \boxed{C_n} \longrightarrow \boxed{out}$$

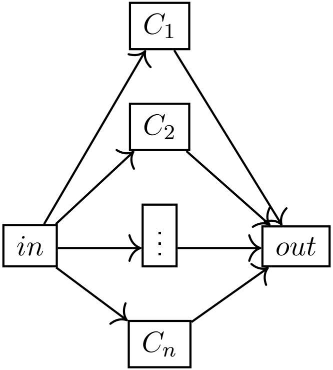
- Si supponga che i componenti del sistema si guastino in modo **indipendente** e che la probabilità di guasto del componente i-esimo sia p_i.
- Sia inoltre A_i l'evento in cui il componente i-esimo funziona e A l'evento in cui l'intero sistema funziona (quindi $A = \cap_{i=1}^{n} A_i$)

☞ Allora, usando l'ipotesi di indipendenza dei componenti:

$$\mathbb{P}[A] = \mathbb{P}\left[\bigcap_{i=1}^{n} A_i\right] = \prod_{i=1}^{n} \mathbb{P}[A_i] = \prod_{i=1}^{n}(1 - p_i),$$

Esempio: Sistemi in parallelo

Un sistema formato da n componenti separati si dice **in parallelo** se funziona quando almeno uno degli n componenti funziona.



Esempio: Sistemi in parallelo

- Si supponga che i componenti del sistema si guastino in modo **indipendente** e che la probabilità di guasto del componente i-esimo sia p_i.
- Sia inoltre A_i l'evento in cui il componente i-esimo funziona e B l'evento in cui l'intero sistema funziona (quindi $B = \cup_{i=1}^{n} A_i$)

☞ Allora, usando l'ipotesi di indipendenza dei componenti:

$$\mathbb{P}[B] = 1 - \mathbb{P}\left[\bar{B}\right] = 1 - \mathbb{P}\left[\bigcap_{i=1}^{n} \bar{A}_i\right] = 1 - \prod_{i=1}^{n} \mathbb{P}\left[\bar{A}_i\right] = 1 - \prod_{i=1}^{n} p_i,$$

Esempio: Test diagnostici

- La frazione dei soggetti affetti da una certa malattia (per esempio, la sieropositività HIV oppure la tubercolosi) in una popolazione si chiama **prevalenza**.
- Si consideri un test diagnostico per la malattia. La **sensitività** di un test è la probabilità che il test, somministrato a un malato, sia positivo.
- La **specificità** di un test è la probabilità che il test, somministrato a un non malato, sia negativo.

☞ **Situazione ideale:** sensitività $=$ specificità $= 1$. Purtroppo, la situazione ideale non è raggiungibile, e i test reali sono imperfetti, cioè con sensitività < 1 e specificità < 1. Questo da origine a **Falsi positivi e falsi negativi**

Esempio: Test diagnostici

- Si immagini di somministrare un test diagnostico non perfetto a una persona estratta a caso dalla popolazione e si considerino gli eventi:

$$\begin{aligned} M &= \text{la persona estratta è malata} \\ + &= \text{il test dà risultato positivo} \\ - &= \text{il test dà risultato negativo} \\ \bar{M} \cap + &= \text{il test dà un falso positivo} \\ M \cap - &= \text{il test dà un falso negativo} \end{aligned}$$

- Si ha allora

$$\mathbb{P}[M] = \text{prevalenza} \quad \mathbb{P}[+|M] = \text{sensitività} \quad \mathbb{P}\left[-|\bar{M}\right] = \text{specificità}$$

Esempio: Test diagnostici

- Probabilità di un falso positivo:

$$\mathbb{P}\left[\bar{M} \cap +\right] = \mathbb{P}\left[\bar{M}\right] \mathbb{P}\left[+|\bar{M}\right]$$
$$= (1 - \text{prevalenza}) \times (1 - \text{specificità}).$$

- Probabilità di un falso negativo:

$$\mathbb{P}\left[M \cap -\right] = \mathbb{P}\left[M\right] \mathbb{P}\left[-|M\right]$$
$$= \text{prevalenza} \times (1 - \text{sensitività}).$$

- Si studi un test per l'HIV:

$$\text{prevalenza} = \mathbb{P}\left[HIV\right] = 0.001$$
$$\text{sensitività} = \mathbb{P}\left[+|HIV\right] = .95$$
$$\text{specificità} = \mathbb{P}\left[-|\overline{HIV}\right] = .98$$

Esempio: Test diagnostici

- La probabilità di falso positivo è

$$\begin{aligned}\mathbb{P}\left[\overline{HIV} \cap +\right] &= \mathbb{P}\left[\overline{HIV}\right]\mathbb{P}\left[+|\overline{HIV}\right] \\ &= (1-0.001)(1-0.98) = 0.01998.\end{aligned}$$

- La probabilità di falso negativo è

$$\begin{aligned}\mathbb{P}\left[HIV \cap -\right] &= \mathbb{P}\left[HIV\right]\mathbb{P}\left[-|HIV\right] \\ &= 0.001(1-0.95) = 0.00005.\end{aligned}$$

La legge della probabilità totale

Legge della probabilità totale

Se $C_1, C_2, \ldots$ sono una partizione dell'evento certo, la probabilità di un qualsiasi evento A può essere scritta come

$$\mathbb{P}[A] = \sum_i \mathbb{P}[A \cap C_i] = \sum_i \mathbb{P}[C_i]\,\mathbb{P}[A|C_i]$$

che si dice **legge della probabilità totale**.

- La prima uguaglianza viene dal fatto che

  $$A = \bigcup_i (A \cap C_i),$$

  che sono eventi a due a due disgiunti.
- La seconda uguaglianza segue dalla definizione di probabilità condizionata, sempre che $\mathbb{P}[C_i] > 0$, $\forall i = 1, 2, \ldots$.

La legge della probabilità totale

- **Esempio HIV:** Nell'esempio di prima, HIV e $\overline{HIV}$ costituiscono una semplice partizione formata da due blocchi.
- Calcoliamo la probabilità che il test, somministrato ad una persona campionata a caso dalla popolazione, sia positivo:

$$\begin{aligned}\mathbb{P}[+] &= \mathbb{P}[+|HIV]\,\mathbb{P}[HIV] + \mathbb{P}\left[+|\overline{HIV}\right]\mathbb{P}\left[\overline{HIV}\right]\\ &= 0.95 \times 0.001 + (1-0.98) \times (1-0.001)\\ &= 0.02093\end{aligned}$$

cioè in pratica avremo, a lungo andare, il 2 per cento di positivi, siano essi veri positivi o falsi positivi.

La formula di Bayes

Formula o teorema di Bayes

Sia data la partizione $C_1, C_2, \dots$ e tutti i suoi elementi abbiano probabilità positiva. Sia A un ulteriore evento, anch'esso con probabilità positiva. Fissiamo l'attenzione su uno specifico elemento C_m della partizione. Abbiamo allora

$$\mathbb{P}\left[C_m|A\right] == \frac{\mathbb{P}\left[A|C_m\right]\mathbb{P}\left[C_m\right]}{\sum_i \mathbb{P}\left[A|C_i\right]\mathbb{P}\left[C_i\right]}$$

- **Thomas Bayes (1702–1761):** matematico e ministro presbiteriano britannico. Deve la sua fama ai suoi studi nel campo della matematica e della filosofia; è noto soprattutto nella statistica per il teorema di Bayes, vertente sulla probabilità condizionata, pubblicato postumo nel 1763 (Wikipedia)

Thomas Bayes?

La formula di Bayes

La formula si può derivare da:

1. La definizione di probabilità condizionata:

$$\mathbb{P}\left[C_m | A\right] = \frac{\mathbb{P}\left[C_m \cap A\right]}{\mathbb{P}\left[A\right]}$$

2. La formula delle probabilità composte (per il numeratore):

$$\mathbb{P}\left[C_m \cap A\right] = \mathbb{P}\left[C_m\right] \mathbb{P}\left[A | C_m\right]$$

3. La legge della probabilità totale (per il denominatore):

$$\mathbb{P}\left[A\right] = \sum_i \mathbb{P}\left[A | C_i\right] \mathbb{P}\left[C_i\right]$$

La formula di Bayes

Perché dovremmo essere interessati alla probabilità condizionata di un singolo elemento della partizione, C_m?

➔ **Esempio:** Riconsideriamo l'esempio HIV. È di primario interesse l'evento che una persona risultata positiva sia effettivamente malata.

- Usando i numeri di prima:

$$\begin{aligned}\mathbb{P}\left[HIV|+\right] &= \frac{\mathbb{P}\left[HIV\right]\mathbb{P}\left[+|HIV\right]}{\mathbb{P}\left[HIV\right]\mathbb{P}\left[+|HIV\right]+\mathbb{P}\left[\overline{HIV}\right]\mathbb{P}\left[+|\overline{HIV}\right]} \\ &= \frac{0.001\cdot 0.95}{0.001\cdot 0.95+(1-0.001)(1-0.98)}\approx 0.045\end{aligned}$$

- Nel campo della diagnostica, $\mathbb{P}\left[M|+\right]$ viene chiamata **valore predittivo positivo**, mentre $\mathbb{P}\left[\overline{M}|-\right]$ è il **valore predittivo negativo**.

La formula di Bayes

- **Interpretazione del teorema di Bayes:**
  - Permette di aggiornare l'assegnazione di probabilità data **a priori** a certi eventi C_m, alla luce di nuova informazione (A si è verificato).
  - Il risultato di questa operazione di aggiornamento sono le nuove probabilità $\mathbb{P}[C_m|A]$, dette anche **a posteriori**.
  - Le probabilità a posteriori costituiscono una sintesi dell'informazione disponibile a priori su un certo fenomeno e dell'informazione empirica.

➔ **Esempio:** Consideriamo un filtro automatico che blocchi i messaggi *spam* in arrivo alla casella di posta elettronica.

- Il metodo per selezionare i messaggi spam si basa sulla presenza di alcune parole chiave all'interno del messaggio stesso (ad esempio `viagra`, `vi@gra`, `v1@gr@`, `V.iagra`, ecc.).
- Consideriamo gli eventi:
  S = "il messaggio è spam"
  W = "il messaggio contiene determinate parole chiave"

La formula di Bayes

- Per il teorema di Bayes possiamo fissare a priori la probabilità che un messaggio sia spam e poi aggiornarla nel caso in cui il messaggio contenga le parole chiave considerate:

$$\mathbb{P}\left[S|W\right] = \frac{\mathbb{P}\left[W|S\right]\mathbb{P}\left[S\right]}{\mathbb{P}\left[W|S\right]\mathbb{P}\left[S\right] + \mathbb{P}\left[W|\bar{S}\right]\mathbb{P}\left[\bar{S}\right]},$$

  - In mancanza di ulteriore informazione, possiamo fissare $\mathbb{P}\left[S\right] = 1/2$, cioè a priori possiamo immaginare di avere la stessa probabilità di ricevere un messaggio spam o uno buono;
  - $\mathbb{P}\left[W|S\right]$ e $\mathbb{P}\left[W|\bar{S}\right]$ possiamo **stimarle** utilizzando l'esperienza, cioè i dati. Ad esempio, immaginiamo di aver analizzato un campione di 2000 messaggi spam e di aver trovato le parole chiave in 250 di questi. Mentre su 1000 messaggi non spam solo 5 contenevano le parole chiave. Allora, possiamo fissare $\mathbb{P}\left[W|S\right] \simeq 250/2000$ e $\mathbb{P}\left[W|\bar{S}\right] \simeq 5/1000$.

La formula di Bayes

- Allora possiamo calcolare

$$\mathbb{P}[S|W] = \frac{0.125 \times 0.5}{0.125 \times 0.5 + 0.005 \times 0.5} = 0.962.$$

- Il filtro automatico confronta questa probabilità con una soglia prefissata. Se $\mathbb{P}[S|W]$ supera la soglia allora il messaggio viene classificato come spam e dunque bloccato. Se invece $\mathbb{P}[S|W]$ sta sotto la soglia, allora il messaggio viene consegnato nella casella di posta.
- Il valore della soglia dipende da quanto sono infastidito dai messaggi spam (soglia bassa) e quanto invece mi preoccupa perdere un messaggio importante (soglia alta).
- Se nell'esempio fissiamo una soglia pari a 0.9, allora il messaggio considerato viene bloccato come spam: $\mathbb{P}[S|W] = 0.962 > 0.9$.