

PROBABILITÀ CONDIZIONATE

PROPRIETÀ di BASE

delle PROBABILITÀ - seconda parte

Giovanna Nappo

Università La Sapienza di Roma

ottobre 2024

- **Probabilità condizionata di A dato B :** $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ ($\mathbb{P}(B) > 0$)

- **Formula delle probabilità composte:** $\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B)$

- **Generalizzazione:** dati n eventi $E_1, \dots, E_n$,

$$\begin{aligned}\mathbb{P}(E_1 \cap E_2 \cap \dots \cap E_n) &= \\ &= \mathbb{P}(E_1) \cdot \mathbb{P}(E_2|E_1) \cdot \mathbb{P}(E_3|E_1 \cap E_2) \cdots \mathbb{P}(E_n|E_1 \cap E_2 \cap \dots \cap E_{n-1}).\end{aligned}$$

- **Formula delle probabilità totali** $\mathbb{P}(A) = \mathbb{P}(B)\mathbb{P}(A|B) + \mathbb{P}(\bar{B})\mathbb{P}(A|\bar{B})$

- **Generalizzazione:** data una partizione $\mathcal{H} = \{H_1, H_2, \dots, H_m\}$, $\mathbb{P}(H_j) > 0$,

$$\mathbb{P}(A) = \sum_{j=1}^m \mathbb{P}(H_j)\mathbb{P}(A|H_j)$$

- **Formula di Bayes o delle probabilità inverse:** $\mathbb{P}(B|A) = \frac{\mathbb{P}(B)\mathbb{P}(A|B)}{\mathbb{P}(A)}$

- **Generalizzazione:** data una partizione $\mathcal{H} = \{H_1, H_2, \dots, H_m\}$, con $\mathbb{P}(H_j) > 0$,

$$\mathbb{P}(H_i|A) = \frac{\mathbb{P}(H_i)\mathbb{P}(A|H_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(H_i)\mathbb{P}(A|H_i)}{\sum_{j=1}^m \mathbb{P}(H_j)\mathbb{P}(A|H_j)}$$

Definizione (Probabilità condizionata)

Siano E ed H due eventi, con $\mathbb{P}(H) > 0$. Viene detta **probabilità condizionata di E dato H** , ed indicata con il simbolo $\mathbb{P}(E|H)$, la quantità

$$\mathbb{P}(E|H) = \frac{\mathbb{P}(E \cap H)}{\mathbb{P}(H)}.$$

interpretazione di $\mathbb{P}(E|H)$

All'interno di ciascuna delle interpretazioni della probabilità (classica, frequentista, soggettivista, ...) il numero $\mathbb{P}(E|H)$ rappresenta la probabilità che, coerentemente con tale interpretazione, dovremmo attribuire al verificarsi di E , se sapessimo che si è verificato H .

Esempio

Iniziamo ricordando che un mazzo di carte napoletane è composto da 40 carte, con carte di 4 semi:

10 carte di denari, 10 carte di coppe, 10 carte di bastoni, 10 carte di spade;

le carte di ogni seme sono numerate da 1 a 10, ma si usa anche dire Asso per l'1, Fante per l'8, Cavallo per il 9 e Re per il 10.

Dopo aver mischiato bene un mazzo di carte napoletane prendiamo la carta in cima al mazzo e, **senza guardarla**, la mettiamo da parte con il dorso verso l'alto, così che non si possa sapere che carta sia.

Se vi chiedessi quanto vale probabilità che la carta, non ancora vista, sia l'Asso di bastoni, sono sicura che chiunque di voi risponderebbe (e giustamente) che tale probabilità vale $1/40$, ossia, posto

$$A_b = \{ \text{la carta in cima al mazzo è l'Asso di bastoni} \},$$

la risposta sarebbe $\mathbb{P}(A_b) = \frac{1}{40}$.

Similmente se vi chiedessi la probabilità dell'evento $A_d = \{ \text{la prima carta del mazzo è l'Asso di denari} \}$, rispondereste che $\mathbb{P}(A_d) = 1/40$.

In effetti in cima al mazzo è possibile che ci sia una qualunque tra le 40 carte e con uguale probabilità, quindi 40 casi possibili ed equiprobabili e c'è un solo caso favorevole ad A.

Abbiamo cioè preso come

$\Omega = \{1d, 2d, \dots, 10d, 1c, 2c, \dots, 10c, 1b, 2b, \dots, 10b, 1s, 2s, \dots, 10s\}$, $A_b = \{1b\}$
e $\mathbb{P}$ come la probabilità classica, QUINDI

$$\mathbb{P}(A_b) = \frac{|A_b|}{|\Omega|} = \frac{1}{40}$$

Supponiamo ora che io guardi la carta e che, pur non facendovela vedere, **vi informi che si tratta di una carta di bastoni**. A questo punto, se vi rifacessi la stessa domanda sono certa che non direste più che la probabilità che *la prima carta del mazzo sia l'Asso di bastoni* vale $1/40$, ma mi rispondereste che vale $1/10$, ossia $\mathbb{P}_B(A_b) = 1/10$ mentre se vi chiedessi la probabilità che *la prima carta del mazzo sia l'Asso di denari*, rispondereste che vale 0 , ossia $\mathbb{P}_B(A_d) = 0$. **Cosa è successo?** l'informazione che vi ho dato, ossia che si è verificato l'evento $B = \{\text{la prima carta del mazzo è una carta di bastoni}\} = \{1b, 2b, \dots, 10b\}$, vi ha fatto cambiare la valutazione della probabilità di A_b e anche quella di A_d .

In pratica, **grazie all'informazione che si è verificato l'evento B** , ora sapete che sono possibili solo 10 casi, corrispondenti alle 10 carte di bastoni, mentre il caso favorevole rimane sempre uno solo, (e ovviamente date probabilità zero a tutti i casi relativi all'estrazione di una carta non di bastoni e quindi anche all'estrazione dell'Asso di denari). Inoltre prima di avere l'informazione, usando la probabilità classica, consideravate **equiprobabili** tutti gli eventi elementari/casi, compresi quelli di B , **dopo aver saputo che si è verificato B** ritenete i casi di B **equiprobabili**, ma ritenete che quelli del complementare di B abbiano probabilità nulla.

Controlliamo ora che $1/10$ è lo stesso valore che si ottiene per $\mathbb{P}(A_b|B)$ usando la definizione

$$\mathbb{P}_B(A_b) = \mathbb{P}(A_b|B) = \frac{\mathbb{P}(A_b \cap B)}{\mathbb{P}(B)} = \frac{\frac{|A_b \cap B|}{|\Omega|}}{\frac{|B|}{|\Omega|}} = \frac{|A_b \cap B|}{|B|} = \frac{1}{10}$$

e similmente

$$\mathbb{P}_B(A_d) = \mathbb{P}(A_d|B) = \frac{\mathbb{P}(A_d \cap B)}{\mathbb{P}(B)} = \frac{\frac{|A_d \cap B|}{|\Omega|}}{\frac{|B|}{|\Omega|}} = \frac{|A_d \cap B|}{|B|} = \frac{0}{10} = 0$$

È chiaro come la precedente verifica si può ripetere in ogni caso in cui si usi la probabilità classica.

In questo esempio, invece di usare la Probabilità classica, useremo la frequenza. Nella seguente tabella sono riportati i dati veri ottenuti da un noto sito WEB di Meteo e si riferiscono alle temperature Massime registrare il 25 settembre dei venti anni dal 2000 al 2019 compresi.

SPIEGAZIONE DELLA TABELLA

Abbiamo posto, per brevità , Temperatura Massima = T_M . Nella prima colonna, ci sono le date esaminate, nella seconda colonna ci sono le temperature massime registrate, nella rimanenti colonne abbiamo segnato 1 oppure 0 a seconda se nella data corrispondente si è verificata la condizione in cima alla colonna.

data	T_M	$T_M \leq 23^0 C$	$T_M \geq 28^0 C$	$T_M \leq 25^0 C$	$T_M > 23^0 C$
25/09/2019	$26^0 C$	0	0	0	1
25/09/2018	$25^0 C$	0	0	1	1
25/09/2017	$24^0 C$	0	0	1	1
25/09/2016	$26^0 C$	0	0	0	1
25/09/2015	$24^0 C$	0	0	1	1
25/09/2014	$22^0 C$	1	0	1	0
25/09/2013	$27^0 C$	0	0	0	1
25/09/2012	$26^0 C$	0	0	0	1
25/09/2011	$28^0 C$	0	1	0	1
25/09/2010	$25^0 C$	0	0	1	1
25/09/2009	$29^0 C$	0	1	0	1
25/09/2008	$22^0 C$	1	0	1	0
25/09/2007	$23^0 C$	1	0	1	0
25/09/2006	$21.1^0 C$	1	0	1	0
25/09/2005	$26.3^0 C$	0	0	0	1
25/09/2004	$22.1^0 C$	1	0	1	0
25/09/2003	$27^0 C$	0	0	0	1
25/09/2002	$20^0 C$	1	0	1	0
25/09/2001	$24.1^0 C$	0	0	1	1
25/09/2000	$28.4^0 C$	0	1	0	1
frequenza relativa		6/20	3/20	11/20	14/20

Passiamo ora a chiederci come cambia la frequenza relativa degli eventi $\{T_M \leq 23^0C\}$, $\{T_M \geq 28^0C\}$ e $\{T_M \leq 25^0C\}$ nel caso in cui si sappia che si è verificato l'evento $\{T_M > 23^0C\}$.

Ovvero vogliamo calcolare, sempre usando la frequenza, le probabilità condizionate

$$\mathbb{P}(\{T_M \leq 23^0C\}|\{T_M > 23^0C\}), \quad \mathbb{P}(\{T_M \geq 28^0C\}|\{T_M > 23^0C\}), \\ \mathbb{P}(\{T_M \leq 25^0C\}|\{T_M > 23^0C\}), \quad \mathbb{P}(\{T_M > 23^0C\}|\{T_M > 23^0C\}).$$

Si tratta quindi di contare il numero $n_{>23}$ dei giorni nei quali si verifica $T_M > 23^0C$, e **tra questi** contare il numero m di quelli in cui si verificano **anche** gli eventi che ci interessano (ovviamente uno per volta) e poi considerare il rapporto $\frac{m}{n_{>23}}$.

Nella tabella seguente abbiamo colorato di rosso i casi che ci interessano, ossia quelli in cui nell'ultima colonna, che è relativa a $T_M > 23^0C$, compare un 1.

anno	T_M	$T_M \leq 23^0 C$	$T_M \geq 28^0 C$	$T_M \leq 25^0 C$	$T_M > 23^0 C$
25/09/2019	$26^0 C$	0	0	0	1
25/09/2018	$25^0 C$	0	0	1	1
25/09/2017	$24^0 C$	0	0	1	1
25/09/2016	$26^0 C$	0	0	0	1
25/09/2015	$24^0 C$	0	0	1	1
25/09/2014	$22^0 C$	1	0	1	0
25/09/2013	$27^0 C$	0	0	0	1
25/09/2012	$26^0 C$	0	0	0	1
25/09/2011	$28^0 C$	0	1	0	1
25/09/2010	$25^0 C$	0	0	1	1
25/09/2009	$29^0 C$	0	1	0	1
25/09/2008	$22^0 C$	1	0	1	0
25/09/2007	$23^0 C$	1	0	1	0
25/09/2006	$21.1^0 C$	1	0	1	0
25/09/2005	$26.3^0 C$	0	0	0	1
25/09/2004	$22.1^0 C$	1	0	1	0
25/09/2003	$27^0 C$	0	0	0	1
25/09/2002	$20^0 C$	1	0	1	0
25/09/2001	$24.1^0 C$	0	0	1	1
25/09/2000	$28.4^0 C$	0	1	0	1
frequenza relativa	condizionata a $T_M > 23^0 C$	0/14	3/14	5/14	14/14

Possiamo quindi affermare che con l'approccio frequentista (usando i dati dei 20 anni dal 2000 al 2019) si ha

$$\begin{aligned}\mathbb{P}(\{T_M \leq 23^0 C\} | \{T_M > 23^0 C\}) &= 0/14, \quad \mathbb{P}(\{T_M \geq 28^0 C\} | \{T_M > 23^0 C\}) = 3/14, \\ \mathbb{P}(\{T_M \leq 25^0 C\} | \{T_M > 23^0 C\}) &= 5/14, \quad \mathbb{P}(\{T_M > 23^0 C\} | \{T_M > 23^0 C\}) = 14/14 = 1.\end{aligned}$$

Anche in questo esempio controlliamo che valutare tali probabilità in questo modo equivale a calcolarle usando la definizione formale: consideriamo solo un caso, gli altri si ottengono in modo del tutto equivalente

$$\mathbb{P}(\{T_M \leq 25^0 C\} | \{T_M > 23^0 C\}) = \frac{\mathbb{P}(\{T_M \leq 25^0 C\} \cap \{T_M > 23^0 C\})}{\mathbb{P}(\{T_M > 23^0 C\})} = \frac{\frac{5}{20}}{\frac{14}{20}} = \frac{5}{14}$$

Inoltre, come è naturale che sia, **sapendo che la temperature massima è (strettamente) maggiore di 23^0** , vale 0 la probabilità (condizionata) che la temperature massima sia minore o uguale a 23^0 , mentre vale 1 la probabilità (sempre condizionata) che la temperature massima sia (strettamente) maggiore di 23^0 , in simboli

$$\mathbb{P}(\{T_M \leq 23^0 C\} | \{T_M > 23^0 C\}) = 0 \text{ e } \mathbb{P}(\{T_M > 23^0 C\} | \{T_M > 23^0 C\}) = 1$$

IMPORTANTE

Indicheremo con $E \mapsto \mathbb{P}_H(E)$ la probabilità che dovremmo adottare nel caso venissimo a sapere che si è verificato l'evento H . Vedremo ora una giustificazione del fatto che $\mathbb{P}_H(E)$ coincide con $\mathbb{P}(E|H) = \frac{\mathbb{P}(E \cap H)}{\mathbb{P}(H)}$

L'idea della definizione di probabilità condizionata è la seguente: da una parte, **se sapessimo che si è verificato l'evento H** , daremmo probabilità nulla a tutti gli eventi contenuti in $\overline{H}$, QUINDI per ogni evento E , poiché $E \cap \overline{H} \subset \overline{H}$, dovrebbe valere, come è naturale,

$$\mathbb{P}_H(E \cap \overline{H}) = 0$$

e si potrebbero essere verificati sono solo gli elementi ω di E che appartengono **anche** ad H , ossia a $E \cap H$, inoltre appare del tutto naturale che la probabilità di E , **condizionata al verificarsi di H** sia **proporzionale** alla probabilità di $E \cap H$, ossia

$$\mathbb{P}_H(E) \propto \mathbb{P}(E \cap H)$$

$\Leftrightarrow$

$$\exists \text{ un } K \text{ t.c. } \forall E, \mathbb{P}_H(E) = K\mathbb{P}(E \cap H)$$

DOMANDA: Perché la costante K di proporzionalità

in $\mathbb{P}_H(E) = K\mathbb{P}(E \cap H)$ vale

$$K = \frac{1}{\mathbb{P}(H)} ?$$

Una semplice risposta è la seguente:

necessariamente la probabilità dell'evento certo, **anche se condizionata ad H** , deve sempre valere 1, ossia

$$\mathbb{P}_H(\Omega) = 1$$

$$(e \text{ anche } \mathbb{P}(H|H) = 1)$$

QUINDI

$$1 = \mathbb{P}_H(\Omega) = K\mathbb{P}(\Omega \cap H) = K\mathbb{P}(H)$$

$\Rightarrow$

$$K = \frac{1}{\mathbb{P}(H)}.$$

Quindi necessariamente

$$\mathbb{P}_H(E) = \frac{\mathbb{P}(E \cap H)}{\mathbb{P}(H)} = \mathbb{P}(E|H)$$

$E \mapsto \mathbb{P}_H(E) := \mathbb{P}(E|H)$ è una probabilità

Dobbiamo però controllare SE la funzione $\mathbb{P}_H$, definita da $E \mapsto \mathbb{P}_H(E) = \mathbb{P}(E|H) := \frac{\mathbb{P}(E \cap H)}{\mathbb{P}(H)}$, sia effettivamente una probabilità ossia dobbiamo controllare SE verifica gli assiomi *i*), *ii*), *iii*) :

i) $\boxed{\mathbb{P}_H(E) := \mathbb{P}(E|H) \in [0, 1]}$ infatti

$$E \cap H \subseteq H, \Rightarrow 0 \leq \underbrace{\mathbb{P}(E \cap H)}_{\text{proprietà di monotonia di } \mathbb{P}} \leq \mathbb{P}(H) .$$

quindi
$$0 \leq \mathbb{P}_H(E) = \frac{\mathbb{P}(E \cap H)}{\mathbb{P}(H)} \leq 1,$$

ii) $\boxed{\mathbb{P}_H(\Omega) := \mathbb{P}(\Omega|H) = 1}$ **(normalizzazione)**
(sostanzialmente già verificata)

$$\Omega \cap H = H \Rightarrow \mathbb{P}(\Omega|H) = \frac{\mathbb{P}(\Omega \cap H)}{\mathbb{P}(H)} = \frac{\mathbb{P}(H)}{\mathbb{P}(H)} = 1$$

$E \mapsto \mathbb{P}_H(E) := \mathbb{P}(E|H)$ è una probabilità

iii) **addittività finita:** dobbiamo verificare che

$$E_1, E_2, \text{ con } E_1 \cap E_2 = \emptyset, \Rightarrow \mathbb{P}_H(E_1 \cup E_2) = \mathbb{P}_H(E_1) + \mathbb{P}_H(E_2)$$

infatti $(E_1 \cup E_2) \cap H = (E_1 \cap H) \cup (E_2 \cap H)$ e inoltre

$$(E_1 \cap H) \cap (E_2 \cap H) = E_1 \cap E_2 \cap H \subseteq E_1 \cap E_2 = \emptyset,$$

$$\Rightarrow \mathbb{P}((E_1 \cup E_2) \cap H) = \mathbb{P}((E_1 \cap H) \cup (E_2 \cap H)) = \mathbb{P}(E_1 \cap H) + \mathbb{P}(E_2 \cap H),$$

da cui:

$$\begin{aligned} \mathbb{P}_H(E_1 \cup E_2) &= \mathbb{P}(E_1 \cup E_2 | H) = \frac{\mathbb{P}((E_1 \cup E_2) \cap H)}{\mathbb{P}(H)} = \frac{\mathbb{P}(E_1 \cap H) + \mathbb{P}(E_2 \cap H)}{\mathbb{P}(H)} \\ &= \frac{\mathbb{P}(E_1 \cap H)}{\mathbb{P}(H)} + \frac{\mathbb{P}(E_2 \cap H)}{\mathbb{P}(H)} = \mathbb{P}(E_1 | H) + \mathbb{P}(E_2 | H) = \mathbb{P}_H(E_1) + \mathbb{P}_H(E_2) \end{aligned}$$

$E \mapsto \mathbb{P}_H(E) := \mathbb{P}(E|H)$ è una probabilità

iii)' **additività numerabile:**

$$E_n, n \geq 1, \text{ con } E_i \cap E_j = \emptyset, \text{ per } i \neq j \Rightarrow \mathbb{P}_H\left(\bigcup_{n \geq 1} E_n\right) = \sum_{n \geq 1} \mathbb{P}_H(E_n)$$

infatti $\left(\bigcup_{n \geq 1} E_n\right) \cap H = \bigcup_{n \geq 1} (E_n \cap H)$ e inoltre $E_n \cap H, n \geq 1$, sono tali che $(E_i \cap H) \cap (E_j \cap H) \subseteq E_i \cap E_j = \emptyset$, per $i \neq j$

$$\Rightarrow \mathbb{P}\left(\bigcup_{n \geq 1} E_n \cap H\right) = \sum_{n \geq 1} \mathbb{P}(E_n \cap H), \quad \text{da cui:}$$

$$\begin{aligned} \mathbb{P}_H\left(\bigcup_{n \geq 1} E_n\right) &= \mathbb{P}\left(\bigcup_{n \geq 1} E_n | H\right) = \frac{\mathbb{P}\left(\left(\bigcup_{n \geq 1} E_n\right) \cap H\right)}{\mathbb{P}(H)} = \frac{\sum_{n \geq 1} \mathbb{P}(E_n \cap H)}{\mathbb{P}(H)} \\ &= \sum_{n \geq 1} \frac{\mathbb{P}(E_n \cap H)}{\mathbb{P}(H)} = \sum_{n \geq 1} \mathbb{P}(E_n | H) = \sum_{n \geq 1} \mathbb{P}_H(E_n) \end{aligned}$$

Il fatto che, fissato H con $\mathbb{P}(H) > 0$, la probabilità condizionata $E \mapsto \mathbb{P}_H(E)$ sia una probabilità, ci permette di affermare che valgono tutte le proprietà delle probabilità, quindi, ad esempio,

la probabilità condizionata del complementare di E si calcola attraverso la formula $\mathbb{P}_H(\bar{E}) = 1 - \mathbb{P}_H(E)$, che però di solito si scrive

$$\mathbb{P}(\bar{E}|H) = 1 - \mathbb{P}(E|H) \quad \Leftrightarrow \quad \mathbb{P}(E|H) + \mathbb{P}(\bar{E}|H) = 1;$$

più in generale, se $\mathcal{K} = \{K_1, K_2, \dots, K_r\}$ è una partizione, allora

$$\sum_{\ell=1}^r \mathbb{P}(K_\ell|H) = 1;$$

e anche, ad esempio,

$$A \subseteq B \quad \Rightarrow \quad \mathbb{P}(A|H) \leq \mathbb{P}(B|H).$$

Probabilità composte per due eventi

Se $\mathbb{P}(H) > 0$, allora

$$\mathbb{P}(E \cap H) = \mathbb{P}(H) \cdot \mathbb{P}(E|H)$$

verifica: basta osservare che

$$\mathbb{P}(E|H) = \frac{\mathbb{P}(E \cap H)}{\mathbb{P}(H)} \quad \Leftrightarrow \quad \mathbb{P}(H) \cdot \mathbb{P}(E|H) = \cancel{\mathbb{P}(H)} \cdot \frac{\mathbb{P}(E \cap H)}{\cancel{\mathbb{P}(H)}}$$

Chiaramente, se A e B sono due eventi, con $\mathbb{P}(A) > 0$ e $\mathbb{P}(B) > 0$, possiamo scrivere sia $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A)$ sia $\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B)$

Uso pratico della formula delle probabilità composte

si usa quando è facile/comodo calcolare/valutare $\mathbb{P}(H)$ e $\mathbb{P}_H(E) (= \mathbb{P}(E|H))$ separatamente, rispetto a calcolare direttamente $\mathbb{P}(E \cap H)$.

ESEMPIO

Una prima urna contiene $b_1 = 3$ palline bianche ed $r_1 = 4$ rosse, una seconda urna ne contiene $b_2 = 2$ ed $r_2 = 5$.

Viene scelta la prima urna con probabilità $p_1 = 1/3$ e la seconda con probabilità $p_2 = 1 - p_1 = 2/3$,

successivamente si estrae una pallina dall'urna scelta.

Posto $H_1 = \{\text{viene scelta l'urna 1}\}$, $H_2 = \overline{H_1} = \{\text{viene scelta l'urna 2}\}$ e $B = \{\text{viene estratta una pallina bianca}\}$,

è chiaro che

$\mathbb{P}_{H_1}(B) = \mathbb{P}(B|H_1) = \frac{b_1}{b_1+r_1} = \frac{3}{7}$ e inoltre sappiamo che $\mathbb{P}(H_1) = p_1 = \frac{1}{3}$ e possiamo quindi calcolare la probabilità dell'evento

$H_1 \cap B = \{\text{viene scelta l'urna 1 e viene estratta una pallina bianca}\}$

$$\mathbb{P}(H_1 \cap B) = \mathbb{P}(H_1)\mathbb{P}(B|H_1) = p_1 \frac{b_1}{b_1 + r_1} = \frac{1}{3} \cdot \frac{3}{7} = \frac{1}{7}.$$

Similmente, per $H_2 = \overline{H_1}$, si ha $\mathbb{P}_{H_2}(B) = \mathbb{P}(B|H_2) = \frac{b_2}{b_2+r_2} = \frac{2}{7}$

$$\mathbb{P}(H_2 \cap B) = \mathbb{P}(H_2)\mathbb{P}(B|H_2) = p_2 \frac{b_2}{b_2 + r_2} = \frac{2}{3} \cdot \frac{2}{7} = \frac{4}{21}.$$

Formula delle Probabilità totali (versione con una partizione H e $\bar{H}$)

Primo esempio di calcolo di $\mathbb{P}(E)$: Supponiamo di conoscere $\mathbb{P}(H)$, la **probabilità a priori** di H , (e quindi anche la probabilità a priori del suo complementare $\mathbb{P}(\bar{H})$), e supponiamo di conoscere, oltre a $\mathbb{P}(E|H)$, la **verosimiglianza** di E dato H , anche $\mathbb{P}(E|\bar{H})$, la verosimiglianza di E dato $\bar{H}$. Allora, ricordando la proprietà base $E = (H \cap E) \cup (\bar{H} \cap E)$, otteniamo

$$\mathbb{P}(E) = \mathbb{P}(H \cap E) + \mathbb{P}(\bar{H} \cap E) = \mathbb{P}(H)\mathbb{P}(E|H) + \mathbb{P}(\bar{H})\mathbb{P}(E|\bar{H}).$$

si tratta del primo esempio di **formula delle probabilità totali**

nel gergo statistico/probabilistico

$\mathbb{P}(H)$ è anche detta **probabilità a priori** di H

$\mathbb{P}(\bar{H})$ è anche detta **probabilità a priori** di $\bar{H}$

$\mathbb{P}(E|H)$ è anche detta **verosimiglianza** di E (rispetto a/dato H)

$\mathbb{P}(E|\bar{H})$ è anche detta **verosimiglianza** di E (rispetto a/dato $\bar{H}$)

ESEMPIO: una sola estrazione da un'urna con composizione incognita

Supponiamo di avere due urne e che l'urna i contiene b_i palline bianche ed r_i rosse, $i = 1, 2$. Viene scelta la prima urna con probabilità p_1 e la seconda con probabilità $p_2 = 1 - p_1$, e si estrae una pallina dall'urna scelta.

Per fissare le idee possiamo pensare che, come prima, $b_1 = 3$ ed $r_1 = 4$, $b_2 = 2$ ed $r_2 = 5$, e che $p_1 = 1/3$ e quindi $p_2 = 1 - p_1 = 2/3$.

calcolare la probabilità che la prima pallina sia bianca

NOTA BENE: va sottolineato che stiamo supponendo che chi valuta la probabilità sappia solo le modalità di scelta delle urne, la composizione delle due urne, ma (ovviamente) non sappia quale urna è stata scelta, in questo senso la composizione dell'urna da cui viene estratta la pallina è incognita

soluzione

Posto

$B = \{\text{viene estratta una pallina bianca}\},$

$H_i = \{\text{viene scelta l'urna } i\}, \quad \text{per } i = 1, 2,$

e si ha

$$\mathbb{P}(B|H_i) = \frac{b_i}{b_i + r_i}, \quad \text{per } i = 1, 2,$$

Si tratta di un modo rapido per scrivere

$H_1 = \{\text{viene scelta l'urna } 1\}, H_2 = \{\text{viene scelta l'urna } 2\},$

$$\text{e } \mathbb{P}(B|H_1) = \frac{b_1}{b_1+r_1} = \frac{3}{7}, \mathbb{P}(B|H_2) = \frac{b_2}{b_2+r_2} = \frac{2}{7},$$

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(H_1 \cap B) + \mathbb{P}(H_2 \cap B) = \overbrace{\mathbb{P}(H_1)}^{p_1=1/3} \overbrace{\mathbb{P}(B|H_1)}^{\frac{b_1}{b_1+r_1}=3/7} + \overbrace{\mathbb{P}(H_2)}^{p_2=2/3} \overbrace{\mathbb{P}(B|H_2)}^{\frac{b_2}{b_2+r_2}=2/7} \\ &= p_1 \cdot \frac{b_1}{b_1+r_1} + p_2 \cdot \frac{b_2}{b_2+r_2} \\ &= \frac{1}{3} \cdot \frac{3}{7} + \frac{2}{3} \cdot \frac{2}{7} = \frac{7}{21} \end{aligned}$$

Probabilità inversa e formula di Bayes

Se anche $\mathbb{P}(E) > 0$, ha senso calcolare anche $\mathbb{P}(H|E)$. Nel caso in cui si interpreta E come un “effetto” ed H come una possibile “causa” di E , allora $\mathbb{P}(H|E)$ è detta **probabilità inversa** (o anche probabilità della “causa” H dato l’“effetto” E) e vale la seguente

FORMULA DI BAYES (prima versione)

$$\mathbb{P}(H|E) = \frac{\mathbb{P}(H) \cdot \mathbb{P}(E|H)}{\mathbb{P}(E)}$$

infatti, per definizione della probabilità condizionata e per la formula delle probabilità composte

$$\mathbb{P}(H|E) = \frac{\mathbb{P}(E \cap H)}{\mathbb{P}(E)} = \frac{\mathbb{P}(H) \cdot \mathbb{P}(E|H)}{\mathbb{P}(E)}$$

ricordiamo che, nel gergo statistico/probabilistico

$\mathbb{P}(H)$ è anche detta **probabilità a priori** di H

$\mathbb{P}(E|H)$ è anche detta **verosimiglianza** di E (rispetto ad H)

$\mathbb{P}(H|E)$ è anche detta **probabilità a posteriori** di H (dato l’effetto E)

ESEMPIO: una sola estrazione da un'urna con composizione incognita

Supponiamo di avere due urne e che l'urna i contiene b_i palline bianche ed r_i rosse, $i = 1, 2$. Viene scelta la prima urna con probabilità p_1 e la seconda con probabilità $p_2 = 1 - p_1$, e si estrae una pallina dall'urna scelta.

Per fissare le idee possiamo pensare che, come prima, $b_1 = 3$ ed $r_1 = 4$, $b_2 = 2$ ed $r_2 = 5$, e che $p_1 = 1/3$ e quindi $p_2 = 1 - p_1 = 2/3$.

sapendo che la prima pallina estratta è bianca, come va valutata la probabilità che sia stata scelta la prima urna? e la probabilità che sia stata scelta la seconda urna?

NOTA BENE: va sottolineato che stiamo supponendo che chi valuta la probabilità sappia solo le modalità di scelta delle urne, la composizione delle due urne, e **inoltre sappia che sia stata estratta una pallina bianca**, ma (ovviamente) non sappia quale urna è stata scelta, in questo senso la composizione dell'urna da cui viene estratta la pallina è incognita

Posto

$$B = \{\text{viene estratta una pallina bianca}\},$$

$$H_i = \{\text{viene scelta l'urna } i\}, \quad \text{per } i = 1, 2,$$

e si ha $\mathbb{P}(B|H_i) = \frac{b_i}{b_i + r_i}$, per $i = 1, 2$, $b_1 = 3, r_1 = 4, b_2 = 2, r_2 = 5$,

$$\begin{aligned} \mathbb{P}(H_1|B) &= \frac{\mathbb{P}(H_1)\mathbb{P}(B|H_1)}{\mathbb{P}(H_1)\mathbb{P}(B|H_1) + \mathbb{P}(H_2)\mathbb{P}(B|H_2)} \\ &= \frac{\frac{1}{3} \frac{3}{7}}{\frac{1}{3} \frac{3}{7} + \frac{2}{3} \frac{2}{7}} = \frac{\frac{1}{3} \frac{3}{7}}{\frac{1}{3} \frac{3}{7} + \frac{2}{3} \frac{2}{7}} = \frac{3}{3 + 4} = \frac{3}{7} \end{aligned}$$

$$\begin{aligned} \mathbb{P}(H_2|B) &= \frac{\mathbb{P}(H_2)\mathbb{P}(B|H_2)}{\mathbb{P}(H_1)\mathbb{P}(B|H_1) + \mathbb{P}(H_2)\mathbb{P}(B|H_2)} \\ &= \frac{\frac{2}{3} \frac{2}{7}}{\frac{1}{3} \frac{3}{7} + \frac{2}{3} \frac{2}{7}} = \frac{\frac{2}{3} \frac{2}{7}}{\frac{1}{3} \frac{3}{7} + \frac{2}{3} \frac{2}{7}} = \frac{4}{3 + 4} = \frac{4}{7} = \overbrace{1 - \frac{3}{7} = \frac{4}{7}}^{(=1 - \frac{3}{7} = \frac{4}{7})} \end{aligned}$$

ESEMPIO: continuazione

Nel caso generale in cui si hanno due urne e che l'urna i contiene b_i palline bianche ed r_i rosse, $i = 1, 2$ e in cui viene scelta la prima urna con probabilità p_1 e la seconda con probabilità $p_2 = 1 - p_1$, e si estrae una pallina dall'urna scelta:

$$\mathbb{P}(H_1|B) = \frac{\mathbb{P}(H_1)\mathbb{P}(B|H_1)}{\mathbb{P}(H_1)\mathbb{P}(B|H_1) + \mathbb{P}(H_2)\mathbb{P}(B|H_2)} = \frac{p_1 \frac{b_1}{b_1+r_1}}{p_1 \frac{b_1}{b_1+r_1} + p_2 \frac{b_2}{b_2+r_2}}$$

$$\mathbb{P}(H_2|B) = \frac{\mathbb{P}(H_1)\mathbb{P}(B|H_1)}{\mathbb{P}(H_1)\mathbb{P}(B|H_1) + \mathbb{P}(H_2)\mathbb{P}(B|H_2)} = \frac{p_2 \frac{b_2}{b_2+r_2}}{p_1 \frac{b_1}{b_1+r_1} + p_2 \frac{b_2}{b_2+r_2}} = 1 - \mathbb{P}(H_1|B)$$

OSSERVAZIONE: Nel caso particolare in cui le due urne contengano lo stesso numero di palline, ossia SE $b_1 + r_1 = b_2 + r_2$, ALLORA $\mathbb{P}(H_i|B) \propto p_i b_i$, ossia

$$\mathbb{P}(H_i|B) = \frac{p_i b_i}{p_1 b_1 + p_2 b_2}$$

NOTA BENE: Stiamo supponendo che la nostra informazione consista solo in

- (i) la **composizione delle due urne**, quindi le **due verosimiglianze** $\mathbb{P}(B|H_i)$, $i = 1, 2$,
- (ii) le **due probabilità priori** di scelta dell'urna $\mathbb{P}(H_i)$, per $i = 1, 2$,
- (iii) il fatto che si è verificato l'evento $B = \{\text{è stata estratta una pallina bianca}\}$,
MA NON SAPPIAMO QUALE URNA È STATA SCELTA.

Questo tipo di problematica si presenta in molte situazioni, tra le quali una riguarda i test medici, ad esempio i test per la gravidanza, per il corona virus COVID-19, per l'HIV, etc.

CHIARIAMO il fatto che alcuni test sono economici, ma non perfettamente sicuri, e possono dare luogo a falsi positivi (ossia si può risultare positivi al test anche se NON si è affetti dalla malattia, o non si aspetta un bambino) o a falsi negativi (ossia si può risultare negativi al test, ma si abbia la malattia o si sia incinta). Ovviamente RIPETO si deve trattare di test economici, e non del tutto affidabili.

L'idea è che dopo aver ottenuto un risultato positivo al test economico, per avere la certezza è necessario fare un test più accurato (e quindi più costoso) ma che è sicuro al 100%.

Nel caso del **COVID-19** il test economico è il test antigenico rapido e **l'analisi accurata** (ma più costosa e lunga) è **il tampone molecolare**. Segnalo l'esempio del test per l'HIV negli appunti [SN], ma si potrebbe pensare anche alla situazione più recente del Test per il Covid-19.

SE H rappresenta l'evento **essere malato**
e quindi $\bar{H}$ rappresenta l'evento **non essere malato**
ed E_- rappresenta l'evento **risultare negativi al test**,
ed $E_+ = \bar{E}_-$ rappresenta l'evento **risultare positivi al test**
ALLORA

falso negativo è l'evento $E_- \cap H$, (test negativo, ma malato) $\rightarrow$ ERRORE

vero negativo è l'evento $E_- \cap \bar{H}$,

falso positivo è l'evento $E_+ \cap \bar{H}$, (test positivo, ma sano) $\rightarrow$ ERRORE

vero positivo è l'evento $E_+ \cap H$,

Dal punto di vista delle analisi è importante che le probabilità $\mathbb{P}(\text{falso negativo})$ e $\mathbb{P}(\text{falso positivo})$ siano “piccole”,
in modo che la **probabilità di errore** sia “piccola”, ossia che sia piccola
$$\mathbb{P}((E_- \cap \bar{H}) \cup (E_+ \cap H)) = \mathbb{P}(E_- \cap \bar{H}) + \mathbb{P}(E_+ \cap H)$$

Le probabilità (a priori) di falso negativo e di falso positivo, sono importanti.

Tuttavia, **dato/noto l'esito del test**, è importante che sia piccola la probabilità di errore condizionata a questa informazione:

dopo che abbiamo ottenuto l'esito del test, ad esempio, se sappiamo che si è verificato E_- (siamo risultati negativi al test) ci interessa anche che sia piccola la probabilità di errore, condizionata a questa informazione, ossia, ci interessa che sia piccola

$$\mathbb{P}(H|E_-)$$

cioè **la probabilità (condizionata) di essere malato, dato l'esito negativo del test**.

Ovviamente è importante anche che sia piccola $\mathbb{P}(\bar{H}|E_+)$, la probabilità di non essere malati, dato l'esito positivo del test.

Dal punto di vista medico è più importante che sia piccola $\mathbb{P}(H|E_-)$: la prassi usuale è che se si risulta negativi, allora non si continuano le analisi: MA SE CI FOSSE UN ERRORE si potrebbe perdere traccia dei malati non individuati dal test economico, mentre, se si risulta positivi, allora si viene poi sottoposti all'analisi accurata: E SE CI FOSSE UN ERRORE si farebbe un'analisi costosa inutilmente Dovrebbe ora essere chiaro il motivo per cui è importante **dal punto di vista medico** che $\mathbb{P}(H|E_-)$ sia molto piccola. Ovviamente anche $\mathbb{P}(\bar{H}|E_+)$ deve essere piccola, ma, più che dal punto di vista medico, è importante dal punto di vista economico: serve per evitare di fare inutilmente analisi costose.

Dopo aver fatto una sperimentazione su un campione “grande”, scelto a caso di una popolazione prefissata, e dopo aver sottoposto tutti sia all’analisi accurata sia al test economico, si divide la popolazione esaminata in 4 gruppi:

i pazienti sicuramente malati e positivi al test economico,

i pazienti sicuramente malati e negativi al test economico,

i pazienti sicuramente sani e positivi al test economico,

i pazienti sicuramente sani e negativi al test economico,

possiamo pensare che siano note tutte le probabilità che ci interessano.

sia la probabilità/frequenza relativa della malattia, ossia $\mathbb{P}(H)$, che le probabilità/frequenze relative condizionate $\mathbb{P}(E_+|H)$ e $\mathbb{P}(E_+|\bar{H})$,

e quindi sono note anche

$\mathbb{P}(\bar{H}) = 1 - \mathbb{P}(H)$, $\mathbb{P}(E_-|H) = 1 - \mathbb{P}(E_+|H)$ e $\mathbb{P}(E_-|\bar{H}) = 1 - \mathbb{P}(E_+|\bar{H})$

GENERALIZZAZIONI: Probabilità composte

Probabilità composte per due eventi

$$\mathbb{P}(E_1) > 0 \Rightarrow \mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1) \cdot \mathbb{P}(E_2|E_1).$$

(già visto: porre $H = E_1$ ed $E = E_2$)

Probabilità composte per tre eventi

$$\mathbb{P}(E_1 \cap E_2) > 0 \Rightarrow \mathbb{P}(E_1 \cap E_2 \cap E_3) = \mathbb{P}(E_1) \cdot \mathbb{P}(E_2|E_1) \cdot \mathbb{P}(E_3|E_1 \cap E_2).$$

infatti, posto $H = E_1 \cap E_2$ ed $E = E_3$ si ha

$$\mathbb{P}(E_1 \cap E_2 \cap E_3) = \mathbb{P}(E_1 \cap E_2) \cdot \mathbb{P}(E_3|E_1 \cap E_2) = \mathbb{P}(E_1) \cdot \mathbb{P}(E_2|E_1) \cdot \mathbb{P}(E_3|E_1 \cap E_2)$$

Probabilità composte per n eventi

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_{n-1}) > 0$$



$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) = \mathbb{P}(E_1) \cdot \mathbb{P}(E_2|E_1) \cdot \mathbb{P}(E_3|E_1 \cap E_2) \cdots \mathbb{P}(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1}).$$

si dimostra per induzione

ESEMPIO

Una prima urna contiene $b_1 = 3$ palline bianche ed $r_1 = 4$ rosse, una seconda urna ne contiene $b_2 = 2$ ed $r_2 = 5$. Viene scelta la prima urna con probabilità $p_1 = 1/3$ e la seconda con probabilità $p_2 = 1 - p_1 = 2/3$, e si estraggono una dopo l'altra, **SENZA REISERIMENTO**, due palline dall'urna scelta. Posto $H_1 = \{\text{viene scelta l'urna 1}\}$ e $B_k = \{\text{alla } k\text{-sima estrazione esce una pallina bianca}\}$, per $k = 1, 2$, si ha che $H_1 \cap B_1 \cap B_2 = \{\text{viene scelta l'urna 1 e vengono estratte due palline bianche}\}$ e

$$\begin{aligned}\mathbb{P}(H_1 \cap B_1 \cap B_2) &= \mathbb{P}(H_1)\mathbb{P}(B_1|H_1)\mathbb{P}(B_2|B_1 \cap H_1) \\ &= p_1 \cdot \frac{b_1}{b_1 + r_1} \cdot \frac{b_1 - 1}{b_1 + r_1 - 1} = \frac{1}{3} \cdot \frac{3}{7} \cdot \frac{2}{6} = \frac{1}{21},\end{aligned}$$

infatti $\mathbb{P}(H_1) = p_1 = \frac{1}{3}$, $\mathbb{P}(B_1|H_1) = \frac{b_1}{b_1 + r_1} = \frac{3}{7}$ e $\mathbb{P}(B_2|B_1 \cap H_1) = \frac{b_1 - 1}{b_1 + r_1 - 1}$ in quanto, se si è verificato $B_1 \cap H_1$, dopo la prima estrazione nell'urna 1 sono rimaste $b_1 + r_1 - 1 = 7 - 1 = 6$ palline, di cui $b_1 - 1 = 3 - 1 = 2$ bianche.

Similmente

$$\mathbb{P}(\overline{H}_1 \cap B_1 \cap B_2) = \mathbb{P}(\overline{H}_1)\mathbb{P}(B_1|\overline{H}_1)\mathbb{P}(B_2|\overline{H}_1 \cap B_1) = p_2 \frac{b_2}{b_2 + r_2} \frac{b_2 - 1}{b_2 + r_2 - 1} = \frac{2}{3} \cdot \frac{2}{7} \cdot \frac{1}{6}$$

partizioni finite

Sia $\mathcal{H} = \{H_1, H_2, \dots, H_m\}$ una partizione finita (ossia $\bigcup_{i=1}^m H_i = \Omega$ e $H_i \cap H_j = \emptyset$ per ogni $i \neq j$) con $\mathbb{P}(H_i) > 0$, per ogni $i = 1, 2, \dots, m$.
per ogni evento E si ha

$$\mathbb{P}(E) = \mathbb{P}(E \cap H_1) + \mathbb{P}(E \cap H_2) + \dots + \mathbb{P}(E \cap H_m) = \sum_{i=1}^m \mathbb{P}(E \cap H_i)$$

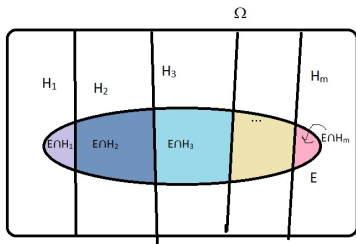
e quindi

$$\mathbb{P}(E) = \mathbb{P}(H_1)\mathbb{P}(E|H_1) + \mathbb{P}(H_2)\mathbb{P}(E|H_2) + \dots + \mathbb{P}(H_m)\mathbb{P}(E|H_m)$$

o sinteticamente

$$\mathbb{P}(E) = \sum_{i=1}^m \mathbb{P}(H_i)\mathbb{P}(E|H_i)$$

(in alcuni testi quest'ultima forma della formula delle probabilità totali viene chiamata formula di disintegrazione)



$$E = (E \cap H_1) \cup (E \cap H_2) \cup \dots \cup (E \cap H_m)$$

$H_1, H_2, \dots, H_m$

formano una partizione:

$$\bigcup_{i=1}^m H_i = \Omega$$

e $H_i \cap H_j = \emptyset$ per ogni $i \neq j$

verifica

$$E = E \cap \Omega = E \cap \bigcup_{i=1}^m H_i = \bigcup_{i=1}^m (E \cap H_i)$$

per la proprietà distributiva dell'unione rispetto all'intersezione,

inoltre $(E \cap H_i) \cap (E \cap H_j) \subseteq H_i \cap H_j = \emptyset$, $\forall i \neq j$, e quindi per l'additività finita e la formula delle probabilità composte

$$\mathbb{P}(E) = \mathbb{P}\left(\bigcup_{i=1}^m (E \cap H_i)\right) = \sum_{i=1}^m \mathbb{P}(E \cap H_i) = \sum_{i=1}^m \mathbb{P}(H_i) \mathbb{P}(E|H_i)$$

GENERALIZZAZIONI: Probabilità totali

partizioni numerabili

$\mathcal{H} = \{H_n, n \geq 1\}$ una partizione numerabile (ossia $\bigcup_{n \geq 1} H_n = \Omega$ e $H_i \cap H_j = \emptyset$ per ogni $i \neq j$) con $\mathbb{P}(H_i) > 0$, per ogni $n \geq 1$.

$$\mathbb{P}(E) = \sum_{n \geq 1} \mathbb{P}(H_n) \mathbb{P}(E|H_n)$$

verifica

$$E = E \cap \Omega = E \cap \bigcup_{i=1}^{\infty} H_i = \bigcup_{i=1}^{\infty} (E \cap H_i)$$

per la proprietà distributiva dell'unione rispetto all'intersezione,
inoltre $(E \cap H_i) \cap (E \cap H_j) \subseteq H_i \cap H_j = \emptyset \forall i \neq j$, e quindi per l'additività numerabile e
per la formula delle probabilità composte

$$\mathbb{P}(E) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} (E \cap H_i)\right) = \sum_{i=1}^{\infty} \mathbb{P}(E \cap H_i) = \sum_{i=1}^{\infty} \mathbb{P}(H_i) \mathbb{P}(E|H_i)$$

GENERALIZZAZIONI: Formula di Bayes per partizioni

formula di Bayes per partizioni finite

Sia $\mathcal{H} = \{H_1, H_2, \dots, H_m\}$ una partizione finita con $\mathbb{P}(H_i) > 0$, per ogni $i = 1, 2, \dots, m$.

Per ogni $j = 1, 2, \dots, m$ e per ogni evento E con $\mathbb{P}(E) > 0$ si ha

$$\mathbb{P}(H_j|E) = \frac{\mathbb{P}(H_j)\mathbb{P}(E|H_j)}{\sum_{i=1}^m \mathbb{P}(H_i)\mathbb{P}(E|H_i)}$$

verifica

basta usare formula di Bayes per E ed H_j e poi la formula delle probabilità totali:

$$\mathbb{P}(H_j|E) = \frac{\mathbb{P}(H_j)\mathbb{P}(E|H_j)}{\mathbb{P}(E)} = \frac{\mathbb{P}(H_j)\mathbb{P}(E|H_j)}{\sum_{i=1}^m \mathbb{P}(H_i)\mathbb{P}(E|H_i)}$$

GENERALIZZAZIONI: Formula di Bayes per partizioni

formula di Bayes per partizioni numerabili

Sia $\mathcal{H} = \{H_n, n \geq 1\}$ una partizione finita con $\mathbb{P}(H_i) > 0$, per ogni $i \geq 1$.
Per ogni $j \geq 1$ e per ogni evento E con $\mathbb{P}(E) > 0$ si ha

$$\mathbb{P}(H_j|E) = \frac{\mathbb{P}(H_j)\mathbb{P}(E|H_j)}{\sum_{i=1}^{\infty} \mathbb{P}(H_i)\mathbb{P}(E|H_i)}$$