

Formula di Bayes

G. M. Marchetti

Aprile 2018

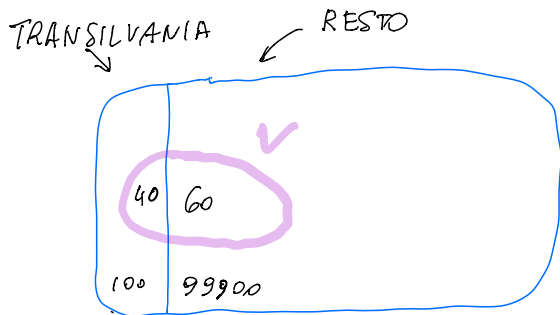
Riassunto

- È utile pensare la probabilità in modo classico
- Probabilità = proporzione = $\frac{\text{casi favorevoli}}{\text{casi possibili}}$

Probabilità condizionata

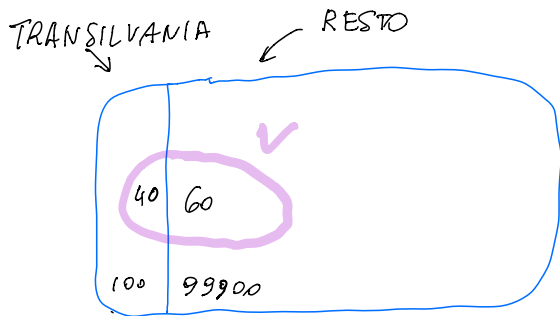
- Probabilità NON condizionata: la proporzione di vampiri nella popolazione è lo $0.1\% = \frac{100}{100.000}$
- Probabilità condizionata: la proporzione di vampiri in Transilvania è il $40\% = \frac{40}{100}$

Venn



$$\Pr(V | T) = \frac{\Pr(V \cap T)}{\Pr(T)} = \frac{40}{100} = 0.4$$

Venn



$$\Pr(V) = \frac{40 + 60}{1000 + 99000} = \frac{100}{100.000} = 0.001$$

Identità: probabilità totale

$$\begin{aligned}\Pr(V) &= \frac{40 + 60}{1000 + 99000} \\ &= \frac{40}{1000} \cdot \frac{1000}{100.000} + \frac{60}{99000} \cdot \frac{99000}{100.000} \\ &= \Pr(V | T) \Pr(T) + \Pr(V | \bar{T}) \Pr(\bar{T}) = 0.001\end{aligned}$$

La probabilità di V è la media ponderata di $\Pr(V | T)$ e $\Pr(V | \bar{T})$.

Esercizio

- La macchina A produce l'80% dei pezzi
- La macchina B produce il 20% dei pezzi
- La probabilità di difettoso è il 5% per la macchina A
- La probabilità di difettoso è l'1% per la macchina B

Qual è la probabilità di difettoso?

Soluzione intuitiva

- A produce 800 pezzi di cui 40 difettosi (il 5%)
- B produce 200 pezzi di cui 2 difettosi (l'1%)

in totale sono 42 su 1000 difettosi (il 4.2%)

Soluzione formale

$$\Pr(D | A) = 0.05, \quad \Pr(D | B) = 0.01$$

$$\Pr(A) = 0.80, \quad \Pr(B) = 0.20$$

$$\Pr(D \cap A) = 0.05 \cdot 0.8 = 0.04, \quad \Pr(D \cap B) = 0.01 \cdot 0.2 = 0.002$$

Quindi

$$\Pr(D) = \Pr(D \cap A) + \Pr(D \cap B) = 0.04 + 0.002 = 0.042$$

Formula di Bayes



Il rev. Thomas Bayes (1702-1761)

Test diagnostico

- Supponiamo che un test diagnostico possa scoprire da un prelievo di sangue se uno/a è un vampiro nel 95% dei casi. È un test che funziona molto bene.
- Il test però può anche sbagliare perché ci sono dei **falsi positivi**. Infatti, nell'1% dei casi il test risulta positivo per una persona normale.
- Sappiamo che i vampiri sono molto rari: solo lo 0.1% della popolazione è un vampiro.

Problema

Uno (o una) va a fare il test e il test risulta positivo! Qual è la probabilità che lui (o lei) sia davvero un vampiro? (da McElrath, 2016)

Soluzione semplice: ragionare con i numeri interi

Supponiamo che la popolazione sia di 100.000 persone.

- Allora 100 sono vampiri (ricorda: i vampiri sono lo 0.1%)
- Di questi 100 vampiri 95 sono positivi al test
- Dei restanti 99.900 mortali 999 risultano positivi al test

Ora se testiamo tutti i 100.000 e prendiamo i

$$positivi = 95 + 999 = 1094$$

quanti di loro sono VERAMENTE vampiri? Ovviamente solo 95 e quindi

$$\Pr(\text{vampiri} \mid +) = \frac{95}{1094} \simeq 0.087$$

Soluzione formale

- Primo problema: la **traduzione**

$$0.95 = \Pr(+ \mid \textit{vampiro})$$

$$0.01 = \Pr(+ \mid \textit{normale})$$

$$0.001 = \Pr(\textit{vampiro})$$

- Secondo problema: cosa debbo cercare

$$\Pr(\textit{vampiro} \mid +) = ???$$

Primo metodo

Fai una tabella doppia e riempi la usando le informazioni.

- 1 sai che i vampiri sono lo 0.1%

<i>Risultato</i>	<i>Vampiro</i>	<i>Mortale</i>	<i>Totale</i>
+			
-			
<i>Totale</i>	100	99.900	100.000

- ② Sai che tra i vampiri il 95% sono positivi al test

<i>Risultato</i>	<i>Vampiro</i>	<i>Mortale</i>	<i>Totale</i>
+	95		
-			
<i>Totale</i>	100	99.900	100.000

- ③ sai che tra i mortali l'1% sono positivi al test

<i>Risultato</i>	<i>Vampiro</i>	<i>Mortale</i>	<i>Totale</i>
+	95	999	
-			
<i>Totale</i>	100	99.900	100.000

4 Calcola l'altro margine della tabella

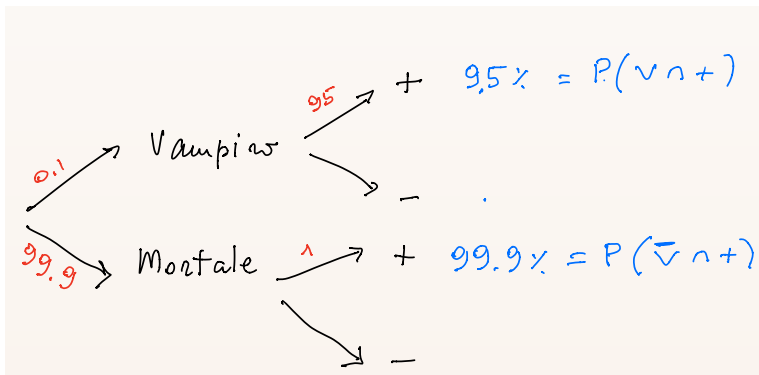
<i>Risultato</i>	<i>Vampiro</i>	<i>Mortale</i>	<i>Totale</i>
+	95	999	1094
-			
<i>Totale</i>	100	99.900	100.000

- 5 Calcola la probabilità condizionata nell'altro verso

$$\Pr(\text{vampiro} \mid +) = \frac{95}{1094}$$

Metodo 2

Usa un albero di probabilità



$$\Pr(\text{vampiro} | +) = \frac{\Pr(\text{vampiro} \cap +)}{\Pr(+)} = \frac{9.5}{9.5 + 99.9} = \frac{9.5}{109.4}$$

Test della mammografia

- Circa l'1% delle donne tra 40 e 50 anni ha un tumore al seno.
- Una donna con il tumore al seno ha il 90% di probabilità di un test positivo alla mammografia
- Una donna senza tumore al seno ha il 5% di probabilità di un falso positivo alla mammografia

Qual'è la probabilità di avere un tumore al seno se una donna è risultata positiva al test?

Soluzione semplice

- Su 10000 donne tra 40 e 50 anni, 100 hanno un tumore al seno e 9900 no
- Tra le 100 col tumore, 90 risultano positive (il 90%)
- Tra le 9900 senza, 495 risultano positive (il 5%)

In totale le positive sono $90 + 495 = 585$. Di queste quante hanno il tumore? Solo 90.

Quindi

$$\Pr(T | +) = \frac{90}{585} = 0.1538.$$

Soluzione con la formula

$$\begin{aligned}\Pr(T | +) &= \frac{\Pr(+ | T) \Pr(T)}{\Pr(+ | T) \Pr(T) + \Pr(+ | \bar{T}) \Pr(\bar{T})} \\ &= \frac{0.90 \cdot 0.01}{0.90 \cdot 0.01 + 0.05 \cdot 0.99} \simeq 0.1538.\end{aligned}$$

Email e spam

- Arriva un email e può essere 'spam' o normale ('ham').

$$S = \textit{spam}, \quad \bar{S} = \textit{non spam}$$

Hanno diverse **probabilità a priori**

$$\Pr(S) = 0.10, \quad \Pr(\bar{S}) = 0.90$$

- L'evento E può essere causato da entrambi:

$$E = \textit{il messaggio contiene la parola 'gratis'}$$

ma la probabilità è diversa sotto le due ipotesi, es.

$$\Pr(E | S) = 0.8, \quad \Pr(E | \bar{S}) = 0.05$$

- Se arriva un messaggio che contiene la parola 'gratis', qual è la probabilità che il messaggio sia spam?

Soluzione

- La probabilità cercata è

$$\Pr(S | E) = \Pr(\textit{spam} | \textit{il messaggio contiene la parola gratis})$$

che viene detta **probabilità a posteriori**

$$\begin{aligned}\Pr(S | E) &= \frac{\Pr(E | S) \Pr(S)}{\Pr(E | S) \Pr(S) + \Pr(E | \bar{S}) \Pr(\bar{S})} \\ &= \frac{0.8 \cdot 0.1}{0.8 \cdot 0.1 + 0.05 \cdot 0.9} \simeq 0.64\end{aligned}$$

- Il controllo anti spam segnala lo spam perché la probabilità a posteriori è > 0.5

Esercizi dagli HW

- Un'associazione studentesca ha distribuito materiale informativo sull'iscrizione alle matricole.
- Tra coloro che hanno ricevuto il materiale, il 40% erano maschi ed il 60% erano femmine.
- In seguito si è rilevato che il 6% dei maschi ed l' 8% delle femmine che avevano ricevuto il materiale si è iscritto all'associazione.

Domanda (a)

Trova la probabilità che uno studente, scelto a caso tra coloro che hanno ricevuto il materiale informativo, si sia iscritto all'associazione

Soluzione semplice

- Supponiamo che ci siano 1000 studenti di cui 400 maschi e 600 femmine
- Dei 400 maschi se ne sono iscritti 24 (il 6%)
- Delle 600 femmine se ne sono iscritte 48 (l'8%)

La probabilità che uno studente si iscriva al corso è $\frac{24+48}{1000} = 0.0720$

Usando la formula

$$\begin{aligned}\Pr(I) &= \Pr(I | M) \Pr(M) + \Pr(I | F) \Pr(F) \\ &= 0.06 \cdot 0.4 + 0.08 \cdot 0.6 = 0.0720.\end{aligned}$$

Domanda (b)

Trova la probabilità che uno studente, scelto a caso tra coloro che si sono iscritti all'associazione dopo aver ricevuto il materiale informativo, sia una femmina.

Soluzione semplice

In totale si sono iscritti all'associazione $24 + 48 = 72$ studenti. Di questi quante sono femmine?

Sono 48. Pertanto

$$\Pr(F | I) = \frac{48}{72} = \frac{2}{3} = 0.6667$$

Usando la formula

$$\Pr(F | I) = \frac{\Pr(I | F) \Pr(F)}{\Pr(I)} = \frac{0.08 \cdot 0.6}{0.072} = 0.6667$$