



Probabilità derivate da frequenze relative

Certo, certissimo, anzi probabile.
Ennio Flaiano



Dai dati alle probabilità

- Le ***frequenze relative*** di una distribuzione osservata sono sempre interpretabili come ***probabilità*** se pensiamo di estrarre a caso una unità statistica da quella stessa popolazione
- Esempio:
 - su 200 studenti che questo anno frequentano il corso di Statistica, 80 provengono da un liceo
 - se estraiamo uno studente a caso tra quei 200, la probabilità che provenga da un liceo è pari alla frequenza relativa, ovvero $80/200 = 0.4$



Dai dati alle probabilità – esempio

Probabilità = frequenza relativa associata a ciascuna modalità (evento elementare) della variabile di interesse

Modalità di acquisto di un televisore	Frequenza assoluta	Frequenza relativa
Grande magazzino (A)	183	$0.61 = P(A)$
Internet (B)	87	$0.29 = P(B)$
Posta (C)	30	$0.10 = P(C)$
Totale (S)	300	$1 = P(S)$

Gli eventi elementari (punti campione) costituiscono una partizione: $A \cap B \cap C = \emptyset$ e $A \cup B \cup C = S$

Domanda: 0.29 è una buona stima della probabilità che un consumatore italiano acquisti oggi un televisore via internet?

Risposta: dipende da dove e quando sono stati rilevati i dati in tabella; siccome si tratta di dati USA di qualche anno fa e il fenomeno degli acquisti via Internet è in fase di rapida crescita e varia molto da paese a paese, l'applicazione ad un caso italiano oggi appare azzardata.



Stimare la probabilità

- Spesso la probabilità viene calcolata su una certa popolazione con l'intenzione di prevedere quello che potrebbe accadere in un'altra popolazione, ad esempio
 - l'anno scorso in Italia 1000 studenti sono stati vaccinati contro la goliardite, e 15 di loro hanno avuto effetti collaterali (frequenza relativa = $15/1000 = 0.015$)
 - se facciamo una estrazione casuale dal registro dei vaccinati la probabilità di estrarre uno studente che ha avuto effetti collaterali è senz'altro 0.015
 - ma la domanda interessante è: qual è la probabilità che il prossimo studente che si sottopone a vaccinazione abbia effetti collaterali? Tale probabilità è 0.015?



Stimare la probabilità /cont.

- Usare la frequenza relativa osservata su una certa popolazione per stimare la probabilità per un nuovo soggetto rappresenta un'estrapolazione basata ipotesi forti:
 - Ipotesi 1: il fenomeno è stabile nel tempo, per cui la probabilità calcolata in base a ciò che è accaduto ieri (la distribuzione di frequenza) è valida anche per ciò che accadrà domani (nell'esempio precedente, se la composizione del vaccino è stata radicalmente modificata i dati relativi all'anno scorso sono di poca utilità)
 - Ipotesi 2: il nuovo soggetto è rappresentativo della popolazione (la frequenza osservata sugli studenti italiani è una buona stima anche per i francesi o per i maori? E anche per uno studente di 80 anni?)
 - Ipotesi 3: la popolazione è omogenea (supponiamo che tra i 1000 studenti vaccinati 500 siano maschi e tra i 15 studenti con effetti collaterali 14 siano maschi... la probabilità per maschi e femmine è piuttosto diversa!)



Stimare probabilità congiunte

Nel caso in cui gli eventi di interesse siano relativi a due o più caratteristiche, la situazione è rappresentata da una distribuzione multipla di frequenze (doppia, se le caratteristiche sono solo due)

Acquisto pianificato	Acquisto effettivo		Totale
	Si	No	
Si	200	50	250
No	100	650	750
Totale	300	700	1000

Frequenza *congiunta* assoluta

Frequenza *marginale* assoluta



Stimare probabilità congiunte \cont.

Considerando la distribuzione delle frequenze relative, si ottengono le probabilità di interesse

Acquisto pianificato	Acquisto effettivo		Totale
	Sì (B)	No (B')	
Sì (A)	$0.20 = P(A \cap B)$	$0.05 = P(A \cap B')$	$0.25 = P(A)$
No (A')	$0.10 = P(A' \cap B)$	$0.65 = P(A' \cap B')$	$0.75 = P(A')$
Totale	$0.30 = P(B)$	$0.70 = P(B')$	$1 = P(S)$

probabilità *marginale*:
acquisto effettivo

probabilità *marginale*:
acquisto non pianificato

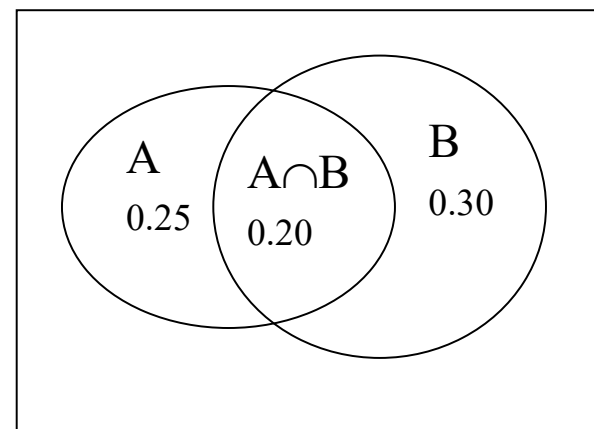
probabilità *congiunta*:
acquisto non pianificato e acquisto effettivo
(probabilità dell'intersezione $A' \cap B$)



Esempio: $P(A \cup B)$

Supponiamo di essere interessati alla probabilità dell'evento d'interesse:

$C = \{\text{acquisto pianificato o acquisto effettivo}\}$:



Acquisto pianificato	Acquisto effettivo		Totale
	Sì (B)	No (B')	
Sì (A)	$0.20 = P(A \cap B)$	$0.05 = P(A \cap B')$	$0.25 = P(A)$
No (A')	$0.10 = P(A' \cap B)$	$0.65 = P(A' \cap B')$	$0.75 = P(A')$
Totale	$0.30 = P(B)$	$0.70 = P(B')$	$1 = P(S)$

Due modi per calcolare $P(C) = P(A \cup B)$

$$P(A \cup B) = \begin{cases} P(A \cap B) + P(A \cap B') + P(A' \cap B) = 0.20 + 0.05 + 0.10 = 0.35 \\ P(A) + P(B) - P(A \cap B) = 0.25 + 0.30 - 0.20 = 0.35 \end{cases}$$



Esempio: probabilità condizionata

Qual è la probabilità che un consumatore che ha pianificato l'acquisto lo abbia poi effettuato?

Acquisto pianificato	Acquisto effettivo		Totale
	Si	No	
Si	200	50	250
No	100	650	750
Totale	300	700	1000

Facendo riferimento alla prima riga della tabella (i 250 consumatori che hanno pianificato l'acquisto) la risposta è $200/250 = 0.80$



Esempio: $P(B|A)$

Partendo dalla tabella delle frequenze relative (viste come probabilità) la domanda può essere posta come: qual è la probabilità di B (acquisto effettivo) condizionatamente a A (acquisto pianificato)?

Acquisto pianificato	Acquisto effettivo		Totale
	Sì (B)	No (B')	
Sì (A)	$0.20 = P(A \cap B)$	$0.05 = P(A \cap B')$	$0.25 = P(A)$
No (A')	$0.10 = P(A' \cap B)$	$0.65 = P(A' \cap B')$	$0.75 = P(A')$
Totale	$0.30 = P(B)$	$0.70 = P(B')$	$1 = P(S)$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.20}{0.25} = 0.80$$



Esempio: $P(A|B)$

Un'altra domanda è: qual è la probabilità che un cliente che ha effettivamente acquistato avesse pianificato l'acquisto? Ovvero: qual è la probabilità di A (acquisto pianificato) condizionatamente a B (acquisto effettivo)?

Acquisto pianificato	Acquisto effettivo		Totale
	Sì (B)	No (B')	
Sì (A)	$0.20 = P(A \cap B)$	$0.05 = P(A \cap B')$	$0.25 = P(A)$
No (A')	$0.10 = P(A' \cap B)$	$0.65 = P(A' \cap B')$	$0.75 = P(A')$
Totale	$0.30 = P(B)$	$0.70 = P(B')$	$1 = P(S)$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.20}{0.30} = 0.67$$



Esempio: A e B indipendenti?

Acquisto pianificato	Acquisto effettivo		Totale
	Sì (B)	No (B')	
Sì (A)	$0.20 = P(A \cap B)$	$0.05 = P(A \cap B')$	$0.25 = P(A)$
No (A')	$0.10 = P(A' \cap B)$	$0.65 = P(A' \cap B')$	$0.75 = P(A')$
Totale	$0.30 = P(B)$	$0.70 = P(B')$	$1 = P(S)$

In precedenza abbiamo calcolato

$$P(A | B) = 0.67 \quad \text{e} \quad P(B | A) = 0.80$$

Per stabilire che A e B non sono indipendenti basta fare una delle seguenti verifiche (è sufficiente una sola):

- $P(A | B) \neq P(A)$ poiché $0.67 \neq 0.25$
- $P(B | A) \neq P(B)$ poiché $0.80 \neq 0.30$
- $P(A \cap B) \neq P(A) P(B)$ poiché $0.20 \neq 0.25 \times 0.30$