Probabilità condizionale e indipendenza

Sia (Ω, P) uno spazio di probabilità discreto.

Def.: Sia B un evento tale che $P(B) > 0$.

Per un evento A definiamo

la <u>probabilità condizionale</u> di A <u>dato</u> B

mediante
$$P(A \mid B) \doteq \frac{P(A \cap B)}{P(B)} .$$

notazione

Interpretazione:

$P(A \mid B)$ è il grado di fiducia che si assegna all'evento A supponendo che si sia verificato B.

(dipende dalla misura di partenza P !).

Esempio: Nella situazione delle estrazioni di n palline di prima

$\tilde{\Omega} \doteq \{\omega \in \Omega : \omega_i \neq \omega_j \text{ se } i \neq j\}$ evento con $P(\tilde{\Omega}) > 0$

$$\tilde{P}(A) \doteq \frac{P(A \cap \tilde{\Omega})}{P(\tilde{\Omega})} = P(A \mid \tilde{\Omega})$$

probabilità condizionale di A dato $\tilde{\Omega}$

Nota: $A \mapsto \tilde{P}(A) = P(A \mid \tilde{\Omega})$ definisce una nuova misura di probabilità.

Proprietà delle probabilità condizionali

Sia (Ω, P) uno spazio di probabilità discreto.

1) Sia B un evento con $P(B) > 0$. Allora

$\mathcal{P}(\Omega) \ni A \mapsto P(A|B)$ definisce una

misura di probabilità su $\mathcal{P}(\Omega)$.

2) Siano $A_1, \ldots, A_n$ eventi tali che $P(\bigcap_{i=1}^{n} A_i) > 0$,

regola della catena

$n \geq 2$. Allora

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot \prod_{i=2}^{n} P(A_i | A_1 \cap \ldots \cap A_{i-1})$$

3) Sia $(B_i)_{i \in I}$ una famiglia al più numerabile

probabilità totali

di eventi disgiunti a due a due e tali che

$$\Omega = \bigcup_{i \in I} B_i \quad \text{e} \quad P(B_i) > 0 \quad \forall i \in I.$$

Allora, per ogni ~~un~~ evento A,

$$P(A) = \sum_{i \in I} P(A|B_i) \cdot P(B_i).$$

In particolare, ~~se~~ se B è un evento tale che $P(B) \in (0,1)$,

!

allora $$P(A) = P(A|B) \cdot P(B) + P(A|B^c) \cdot (1 - P(B)).$$

Teorema : formula di Bayes

Siano A, B eventi in uno spazio di probabilità discreto (Ω, P). Se $P(A) > 0$ e $P(B) > 0$, allora

$$P(A \mid B) = \frac{P(A)}{P(B)} \cdot P(B \mid A)$$

Dim. : dalla def. //

La formula di Bayes permette di "invertire" il condizionamento

Corollario :

1) Se $(B_i)_{i \in I}$ è una partizione al più numerabile di Ω tale che $P(B_i) > 0 \ \forall i \in I$, allora $\forall i \in I$

$$P(B_i \mid A) = \frac{P(B_i) \cdot P(A \mid B_i)}{\sum_{j \in I} P(A \mid B_j) \cdot P(B_j)}$$

! 2) Se B è un evento tale che $P(B) \in (0,1)$,

allora $$P(B \mid A) = \frac{P(B) \cdot P(A \mid B)}{P(A \mid B) \cdot P(B) + P(A \mid B^c) \cdot (1 - P(B))}.$$

Esempio: test clinico

Test per determinare la presenza di un virus negli individui di una popolazione.

Dati sulla qualità del test:

a) sensibilità del test: probabilità che il test dia risultato positivo quando il virus è presente $\doteq p_V$

b) specificità del test: probabilità che il test dia risultato positivo quando il virus <u>non</u> è presente $\doteq p_F$

Ad esempio: $p_V = 99\%$ ("veri positivi")
$p_F = 0{,}5\%$ ("falsi positivi")

Domanda: Scegliamo a caso un individuo per il test. Qual è la probabilità che l'individuo è portatore del virus se il test risulta positivo?

Modello senza specificare lo spazio di probabilità:

Eventi di interesse:

A = "l'individuo è portatore del virus"

B = "il test risulta positivo"

Domanda corrisponde a calcolare $P(A|B)$, la probabilità condizionale di A dato B.

Dai dati noti sul test:

$$p_V = P(B|A), \qquad p_F = P(B|A^c).$$

formula di Bayes (+ Corollario)

$$P(A|B) = \frac{P(A)}{P(B)} \cdot P(B|A)$$

e $$P(B) = P(B|A) \cdot P(A) + P(B|A^c) \cdot (1-P(A))$$

$$\rightsquigarrow \quad P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot (1-P(A))}$$

! Manca un dato: $P(A)$ l'incidenza del virus nella popolazione.

Ad esempio,

incidenza $\quad P(A) = q = \dfrac{40}{100.000} = 4 \cdot 10^{-4}$

(cioè, in media 40 individui su 100000 portatori del virus

Con questi dati $\quad (p_V = 0{,}99\%, \quad p_F = 0{,}5\%)$:

$$P(A|B) = \frac{q \cdot p_V}{q \cdot p_V + p_F \cdot (1-q)}$$

| se q piccolo (risp. a p_F)

$$\approx \frac{q \cdot p_V}{p_F}$$

$$= 4 \cdot 10^{-4} \cdot \frac{0{,}99}{\frac{1}{200}}$$

$$\approx 8 \cdot 10^{-2} = 8\%$$

Nonostante ~~la~~ l'alta qualità del test,
se l'incidenza del virus è bassa, allora
la probabilità di avere un falso positivo per l'individuo
è alta (qui: > 90%)!