CdP - TdD - XII - 15/10/2008

Titolo nota [redacted]

Un altro tipo di problema: calcolare la probabilità che una certa causa abbia generato l'evidenza che osserviamo

Esempio 3a (II parte). Supponiamo che il nuovo assicurato abbia un incidente entro un anno dall'acquisto della polizza. Qual è la probabilità che si tratti di una persona propensa agli incidenti?

evidenza

I = l'assicurato ha avuto un incidente durante l'anno

A quale classe appartiene l'assicurato

Vogliamo calcolare $P(R \mid I)$

$P(R|I)$ = prob. che l'assicurato sia spericolato sapendo che ha avuto un incidente.

$$P(R|I) = \frac{P(RI)}{P(I)}$$

$P(I) = 0.26$ (già calcolato prima)

$$P(RI) \underset{\uparrow}{=} P(R)P(I|R) = (0.3)(0.4) = 0.12$$

reg. prodotto

$$P(R|I) = \frac{P(RI)}{P(I)} = \frac{0.12}{0.26} = \frac{\cancel{12}^{\,6}}{\cancel{26}_{\,13}} \simeq 0.46$$

confrontando le probabilità

$$P(R) = 0.3$$

in seguito alla nuova informazione
(la persona ha avuto un incidente)

$$P(R|I) \triangleq 0.66$$

la probabilità è aumentata,
perché l'evidente I (ha avuto un incidente) favorisce questa ipotesi.

In maniera simile calcolo.

$$P(D|I) = \frac{P(DI)}{P(I)}$$

$$P(I) = 0.26$$

$$P(DI) = P(D)P(I|D) = (0.7)(0.2)$$

prob. totali

$$= 0.14$$

$$P(D|I) = \frac{P(DI)}{P(I)} = \frac{0.14}{0.26} \simeq 0.54$$

Prima dell'evidenza $P(D) = 0.7$

Dopo $P(D|I) \simeq 0.54$ ↓

l'evidenza I rende l'ipotesi D meno probabile

Osservazione

senza l'informazione I

$$P(D) + P(S) = 0.7 + 0.3 = 1$$

con l'informazione I

$$P(D|I) + P(S|I) = \frac{6}{13} + \frac{7}{13} = \frac{13}{13} = 1$$

I valori sono stati aggiornati ma la somma rimane 1.

Esempio 3d. In un laboratorio di analisi l'esame del sangue è efficace al 95% nell'individuare una certa malattia quando essa è presente nell'organismo. L'esame tuttavia rileva anche dei "falsi positivi" nell'1% delle persone sane che si sottopongono all'esame. (Cioè, se l'esame è effettuato da una persona sana, allora, con probabilità 0.01, l'esame rivela che la persona è malata.) Se lo 0.5% della popolazione soffre della malattia, qual è la probabilità che una persona risultata positiva all'esame abbia la malattia?

D = la persona ha la malattia

D^c = la persona è "sana".

E = la persona è positiva al test

il test dice

$P(E|D) = 0.95$

$P(E|D^c) = 0.01$

$P(D) = 0.005$

$P(D|E) = ?$

$$P(D|E) = \frac{P(DE)}{P(E)}$$

$$P(DE) \overset{\text{r. prodotto}}{=} P(D)P(E|D) = (0.005)(0.95) = 0.00475$$

$P(E) = ?$ $\quad E \subset (D \cup D^c) = S$

Applico legge Prob. Total. $\quad$ disgiunti

$$P(E) = P(D)P(E|D) + P(D^c)P(E|D^c)$$

$$\left(P(D^c) = 1 - P(D) = 1 - 0.005 = 0.995\right)$$

$$P(E) = (0.005)(0.95) + (0.995)(0.01) = = 0.00475 + 0.00975 = 0.147$$

$$P(D|E) = \frac{P(ED)}{P(E)} = \frac{0.00475}{0.0147} \simeq 0.323$$

La prob. è sorprendentemente bassa

SPIEGAZIONE INTUITIVA

POPOLAZIONE 10000 AB.

- 10000
  - 50 malati
    - 2.5 negativi
    - 47.5 positivi (veri positivi)
  - 9950 non malati
    - 99.5 positivi (falsi p.)
    - 9850.5 negativi

Il test rileva un numero di persone positive dato da: $50(0.95) = 47.5$ nei malati

e $(9950)(0.01) = 99.5$ nei non malati

Quando risulta l'evidenza E l'individuo può stare fra i veri positivi e i falsi positivi.
La prob. che stia fra i veri positivi data l'evidenza E è

$$\frac{\#\text{ veri positivi}}{\#\text{ positivi}} = \frac{47.5}{47.5+99.5} = \frac{47.5}{147} \simeq 0.323$$

Riassunto:

- Due cause A, B incompatibili.
  (può essere $B = A^c$)
- Un evidenza $E \subset (A \cup B)$

Obiettivo: risalire dall'evidenza alle cause che hanno generato l'evidenza.

Si vuole quindi calcolare
$P(A|E)$ e $P(B|E)$

Conoscendo
- la rilevanza delle cause
  $P(A)$ $P(B)$
- la probabilità con la quale le cause generano l'evidenza
  $P(E|A)$ e $P(E|B)$

Il procedimento utilizzato è

$$P(A|E) = \frac{P(AE)}{P(E)} = \frac{\text{Reg. prodotto}}{\text{legge prob totali}}$$

$$= \frac{P(A)\,P(E|A)}{P(A)\,P(E|A) + P(B)\,P(E|B)}$$

Teorema di BAYES

Similmente

$$P(B|E) = \frac{P(BE)}{P(E)} =$$

$$= \frac{P(B)\,P(E|B)}{P(A)\,P(E|A) + P(B)\,P(E|B)}$$

Osservazioni:

1) Frequentemente $P(E)$ è già calcolato. Non c'è bisogno di applicare la legge della prob totale:

2) $P(B|E) = 1 - P(A|E)$

Il teorema è generalizzabile al caso di più di 2 cause

$\underbrace{F_1, F_2 \dots F_n}_{\text{cause}}$ incompatibili

$E \subset (F_1 \cup F_2 \cup \dots \cup F_n)$

evidente.

$$P(F_i|E) \overset{\text{def}}{=} \frac{P(E F_i)}{P(E)} =$$

$$= \frac{\text{regola del prodotto}}{\text{legge prob. totale}}$$

$$= \frac{P(F_i)\,P(E|F_i)}{P(F_1)P(E|F_1) + P(F_2)P(E|F_2) + \dots + P(F_n)P(E|F_n)}$$

$$= \frac{P(F_i)\,P(E|F_i)}{\sum_{j=1}^{n} P(F_j)\,P(E|F_j)}$$ T. Bayes. (forma generale)

Esempio 3l. Una scatola contiene varie lampadine di 3 tipi diversi. La probabilità che una lampadina di tipo 1 duri più di 1000 ore è pari a 0.7, mentre essa vale 0.4 per le lampadine di tipo 2 e 0.3 per quelle di tipo 3. Si supponga che il 20% delle lampadine della scatola sia di tipo 1, il 30% sia di tipo 2 e il restante 50% di tipo 3.

Si sceglie una lampadina a caso.

(b) Sapendo che la lampadina prescelta dura più di 1000 ore, qual è la probabilità condizionata che si tratti di una lampadina di tipo $j = 1, 2, 3$?

E evidenza

E = la lampadina dura più di 1000 ore

Cause

F_1 = la lampadina è di tipo 1

F_2 = " " " " " 2

F_3 = " " " " " 3

Dati del problema

$P(F_1) = 0.2 \quad P(F_2) = 0.3 \quad P(F_3) = 0.5$

$P(E|F_1) = 0.7 \quad P(E|F_2) = 0.6 \quad P(E|F_3) = 0.3$

Dobbiamo trovare

$P(F_1|E) \quad P(F_2|E) \quad P(F_3|E)$

$$P(F_1|E) \overset{\text{T. Bayes}}{=} \frac{P(F_1 E)}{P(E)} =$$

$$= \frac{P(F_1)P(E|F_1)}{P(F_1)P(E|F_1) + P(F_2)P(E|F_2) + P(F_3)P(E|F_3)}$$

$$= \frac{(0.2)(0.7)}{(0.2)(0.7)+(0.3)(0.4)+(0.5)(0.3)} =$$

$$= \frac{0.14}{0.14+0.12+0.15} = \frac{0.14}{\underbrace{0.41}_{P(E)}} \simeq 0.342$$

$$P(F_2|E) = \frac{P(EF_2)}{P(E)} = \frac{P(F_2)\,P(E|F_2)}{P(E)}$$

fie calculat $\rightarrow P(E)$

$$= \frac{(0.3)(0.4)}{0.41} = \frac{0.12}{0.41} = \frac{12}{41} \simeq 0.293$$

$$P(F_3|E) = \frac{P(EF_3)}{P(E)} = \frac{P(F_3)\,P(E|F_3)}{P(E)}$$

$$= \frac{(0.5)(0.3)}{0.41} = \frac{0.15}{0.41} = \frac{15}{41} \simeq 0.365$$

In alternativa

$$P(F_3|E) = 1 - P(F_1|E) - P(F_2|E)$$

$$= 1 - \frac{14}{41} - \frac{12}{41} = \frac{41-14-12}{41} = \frac{15}{41}$$

$P(F_1) = 0.2$

$P(F_2) = 0.3$

$P(F_3) = 0.5$

DISTRIBUZIONE A PRIORI

è successo E

la lampadina dura + di 1000 ore

$P(F_1|E) = 0.342$

$P(F_2|E) = 0.293$

$P(F_3|E) = 0.365$

DISTRIBUZIONE A POSTERIORI