14 - TEOREMA DI BAYES (SULLA "PROBABILITÀ DELLE CAUSE")

14.1 - La "probabilità delle cause": formula di Bayes

- **In un paese scandinavo il 70% delle ragazze ha i capelli Biondi, il 20% li ha Rossi, il 10% Mori. Risulta poi che ha gli occhi Scuri il 10% delle Bionde, il 25% delle Rosse, il 50% delle More.**
  Se la ragazza con cui ho fatto amicizia tramite Internet mi fa sapere che ha gli occhi Scuri, che probabilità c'è che sia Bionda?
- **In un bar ci sono due macchinette mangiasoldi A e B.**
  Effettuando una singola giocata su A si vince con probabilità 1/2 (in altre parole: si vince mediamente 1 volta su 2, o, se preferisci, all'incirca 500 volte su 1000), mentre giocando su B si vince con probabilità 1/4.
  Supponiamo di non sapere quale sia la macchinetta A e quale la B; se ne scegliamo una a caso, giochiamo una sola volta, e vinciamo, che probabilità c'è che la macchinetta scelta sia stata A?

Ecco **due tipici problemi di "probabilità delle cause".**

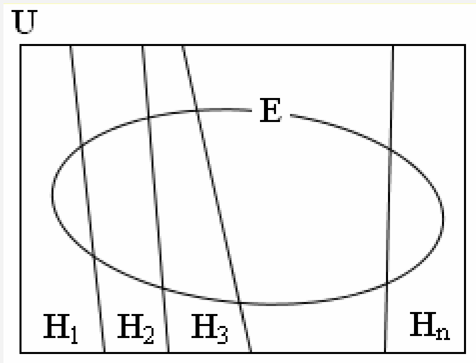
Per tali problemi, esistono più tecniche di risoluzione; ad esempio, sono molto belle ed efficaci quelle che si basano su di una "visione frequentista", o sull'idea delle "fette di certezza".

Comunque, in un problema di "probabilità delle cause" la risorsa più utile è senz'altro la *formula di Bayes*.

TEOREMA DI BAYES (sulla "probabilità delle cause")

Supponiamo che in una singola prova possa verificarsi uno e uno solo fra più possibili eventi $H_1, H_2, \dots, H_n$ (indichiamo con $p(H_i)$ la probabilità che si verifichi H_i), e che, qualora si verifichi l'evento H_i, ci sia una ben determinata probabilità $p(E/H_i)$ che si verifichi un dato evento E.

Insomma, gli eventi $H_1, H_2, \dots, H_n$ costituiscono le possibili CAUSE dell'evento E; tali cause sono:



- **fra loro INCOMPATIBILI**
  (= non è possibile che si verifichino contemporaneamente due eventi H_i, H_j, se $i \neq j$)
- **ed "ESAUSTIVE"** (= nessuna altra causa, al di fuori delle $H_1, \dots, H_n$, può generare l'evento E).

Allora, se si verifica l'evento E, la probabilità che esso sia stato provocato dalla causa H_i è data dalla formula

$$p(H_i/E) = \frac{p(H_i)\cdot p(E/H_i)}{p(H_1)\cdot p(E/H_1) + p(H_2)\cdot p(E/H_2) + \dots + p(H_n)\cdot p(E/H_n)} = \frac{p(H_i)\cdot p(E/H_i)}{\sum_{i=1}^{n} p(H_i)\cdot p(E/H_i)}$$

OSSERVAZIONE CHE AIUTA MOLTO A RICORDARE LA FORMULA
Il denominatore si ottiene riscrivendo il numeratore, e poi scrivendo gli altri addendi analoghi, che si ottengono "facendo variare le cause".

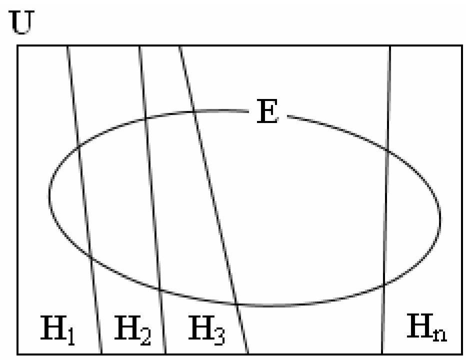
Dimostrazione (con riferimento alla figura):

$$p(H_i/E) = \frac{p(H_i \cap E)}{p(E)} =$$

$$= \frac{p(H_i \cap E)}{p((H_1 \cap E) \cup (H_2 \cap E) \cup \dots \cup (H_n \cap E))}$$

$$= \frac{p(H_i)\cdot p(E/H_i)}{p(H_1 \cap E) + p(H_2 \cap E) + \dots + p(H_n \cap E)} =$$

$$= \frac{p(H_i)\cdot p(E/H_i)}{p(H_1)\cdot p(E/H_1) + p(H_2)\cdot p(E/H_2) + \dots + p(H_n)\cdot p(E/H_n)}$$



Giustificazioni dei passaggi nella dimostrazione:

- nel primo passaggio abbiamo applicato una nota formula ricavata dal Teorema delle Probabilità Composte:

$$p(\mathrm{A}/\mathrm{B}) = \frac{p(\mathrm{A} \cap \mathrm{B})}{p(\mathrm{B})}$$

- nel secondo passaggio, un'ovvia relazione insiemistica
- nel terzo passaggio,
  - il Teorema delle Probabilità Composte a numeratore
  - e il Teorema delle Probabilità Totali per eventi incompatibili a denominatore
- nel quarto passaggio, nuovamente il Teorema delle Probabilità Composte

OSSERVAZIONI

- **La dimostrazione data si riferisce a situazioni in cui possiamo porci**
  in un insieme universo di casi equipossibili,
  quindi si adatterebbe perfettamente al primo dei due esempi da cui abbiamo preso le mosse
  (le ragazze scandinave),
  in quanto il secondo esempio
  (le macchinette mangiasoldi)
  è piuttosto una "prova a due fasi", nella quale i casi non sono equipossibili,
  a meno di passare ad una opportuna
  "prova modificata, probabilisticamente equivalente a quella di partenza".
  Bene!
  Si può tuttavia dimostrare che
  ♥ **IL TEOREMA DI BAYES VALE ANCHE CON RIFERIMENTO AGLI "EVENTI A DUE FASI".**
  Basterà, a tale scopo, semplicemente sostituire, nei passaggi formali della nostra dimostrazione,
  il simbolo di $\cap$ con una congiunzione "$\wedge$" da intendersi come indicante
  successione temporale o comunque "accostamento, abbinamento" di eventi;
  oppure, si potrà ricorrere ad una opportuna
  "prova modificata, probabilisticamente equivalente a quella data",
  analogamente a quanto già fatto nel paragrafo 8.2.
- **Si comprende poi facilmente che**
  ♥ **LA FORMULA DEL TEOREMA DI BAYES RIMANE VALIDA**
  PURE SE GLI EVENTI $\mathrm{H}_1, \mathrm{H}_2, \ldots, \mathrm{H}_n$
  NON VENGONO INTERPRETATI COME "CAUSE" DI E, MA SEMPLICEMENTE
  COME EVENTI CHE POSSONO ESSERE "CONCOMITANTI" CON E.

ESEMPIO

- **In una certa facoltà universitaria, è obbligatorio sostenere un esame di Lingua Straniera.**
  Ogni studente può scegliere fra:
  Inglese, Francese, Spagnolo, Tedesco.
  Le statistiche dicono che le probabilità di scelta sono rispettivamente:
  0,4 0,3 0,2 0,1
  D'altra parte, per la diversa difficoltà dei corsi e severità degli insegnanti,
  le probabilità di riportare la massima votazione (30 trentesimi)
  variano da lingua a lingua e sono rispettivamente:
  0,1 0,2 0,3 0,9
  Supponiamo di sapere che un certo studente ha riportato 30 trentesimi nell'esame di Lingua.
  Che probabilità c'è che la materia d'esame sia stata Inglese?

  Risoluzione

$$p(\text{Inglese}/\text{"30"}) =$$
$$= \frac{p(\mathrm{I}) \cdot p(\text{"30"}/\mathrm{I})}{p(\mathrm{I}) \cdot p(\text{"30"}/\mathrm{I}) + p(\mathrm{F}) \cdot p(\text{"30"}/\mathrm{F}) + p(\mathrm{S}) \cdot p(\text{"30"}/\mathrm{S}) + p(\mathrm{T}) \cdot p(\text{"30"}/\mathrm{T})} =$$
$$= \frac{0,4 \cdot 0,1}{0,4 \cdot 0,1 + 0,3 \cdot 0,2 + 0,2 \cdot 0,3 + 0,1 \cdot 0,9} = \frac{0,04}{0,04 + 0,06 + 0,06 + 0,09} = \frac{0,04}{0,25} = \frac{4}{25} = 0,16$$