IL CALCOLO DELLE PROBABILITA’

MISURA DELLA PROBABILITA’

Per probabilità di un **evento** si intende quel numero che esprime una misura della **“possibilità”** del verificarsi dell’evento stesso. In sostanza è una misura del “**grado di fiducia”** sul verificarsi dell’evento.

DEFINIZIONI

ESPERIMENTO CASUALE = operazione il cui risultato non può essere previsto.

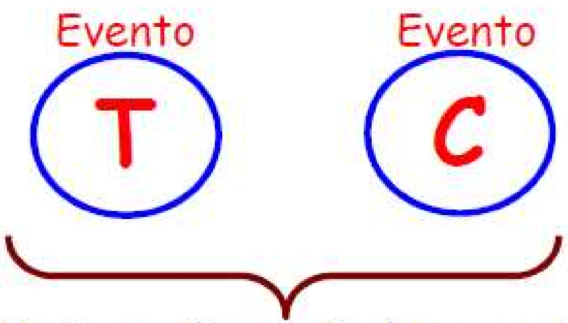
EVENTO = ogni risultato possibile di un esperimento.

SPAZIO CAMPIONARIO = insieme di tutti i risultati possibili per l'esperimento.

Facciamo un esempio di esperimento casuale: LANCIO UNA MONETA

A questo esperimento casuale si possono associare 2 EVENTI: Testa e Croce

Esperimento Casuale



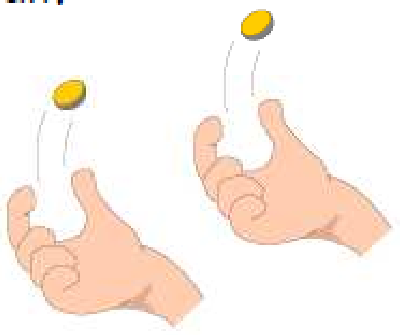
L'insieme di questi due eventi forma lo Spazio Campionario dell'esperimento.

DEFINIZIONI DI PROBABILITA'

Esistono varie definizioni di probabilità:

1. **CLASSICA:** la probabilità di un evento è il rapporto tra tra il numero dei casi favorevoli all'evento ed il totale dei casi possibili, purché tutti i casi siano ugualmente possibili. (Bernoulli, Laplace)

ESEMPIO: se lancio 2 monete, qual è la probabilità che escano 2 facce uguali?



Si hanno dunque 4 casi possibili, e tutti sono ugualmente possibili.

Lo Spazio Campionario associato a questo esperimento casuale è composto da quattro eventi:

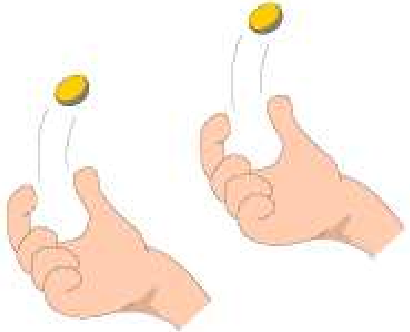
T T
T C
C T
C C

Ma di questi 4 casi possibili, quelli a noi "favorevoli" (2 facce uguali) sono solamente 2

DEFINIZIONI DI PROBABILITA'

Se indichiamo con E il nostro evento abbiamo che la sua probabilità è:

$$p(E) = \frac{2}{4} = 0.5$$



Lo Spazio Campionario associato a questo esperimento casuale è composto da quattro eventi:

T T
T C
C T
C C

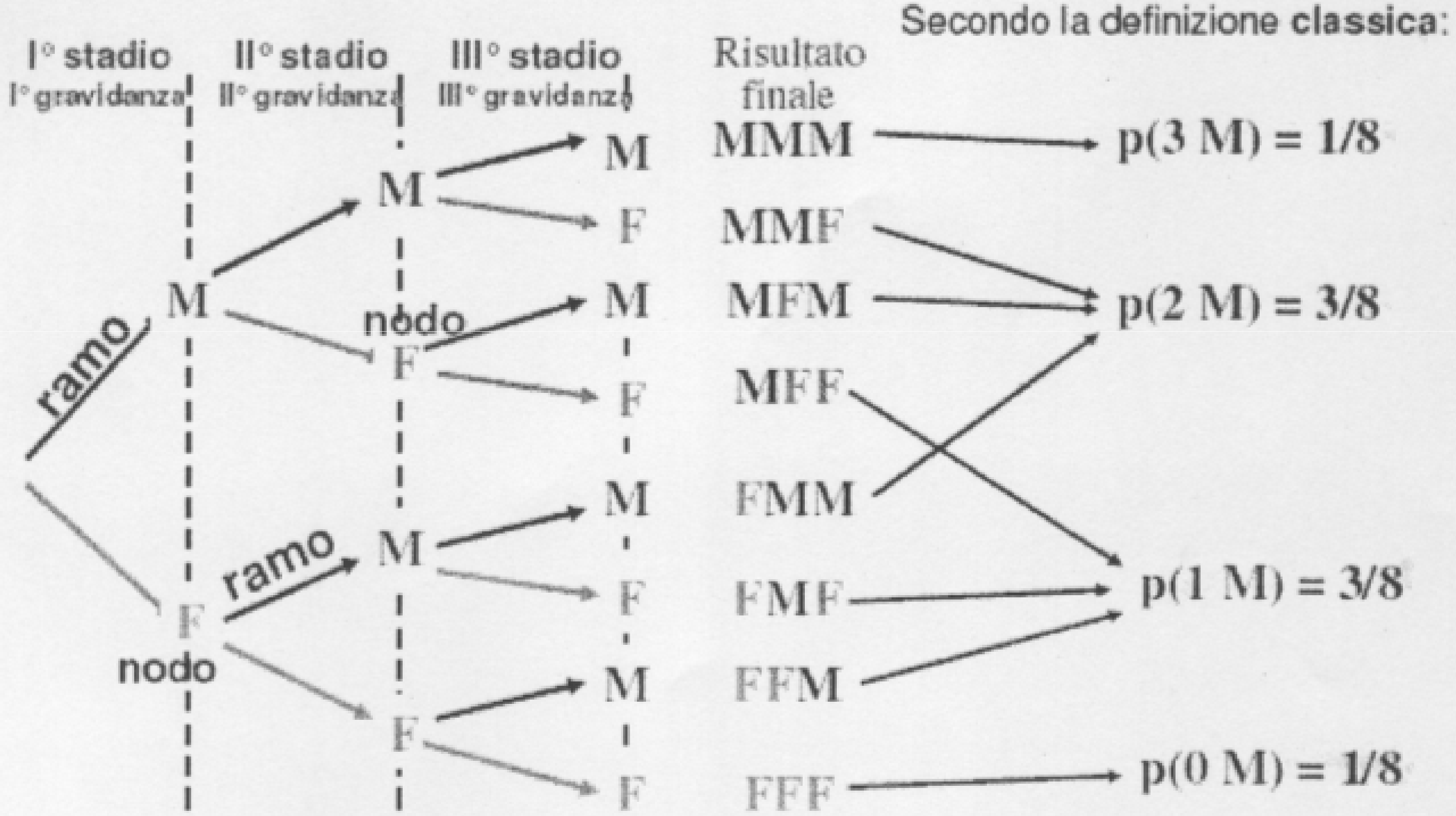
Ma di questi 4 casi possibili, quelli a noi "favorevoli" (2 facce uguali) sono solamente 2

Si hanno dunque 4 casi possibili, e tutti sono ugualmente possibili.

Diagramma ad albero

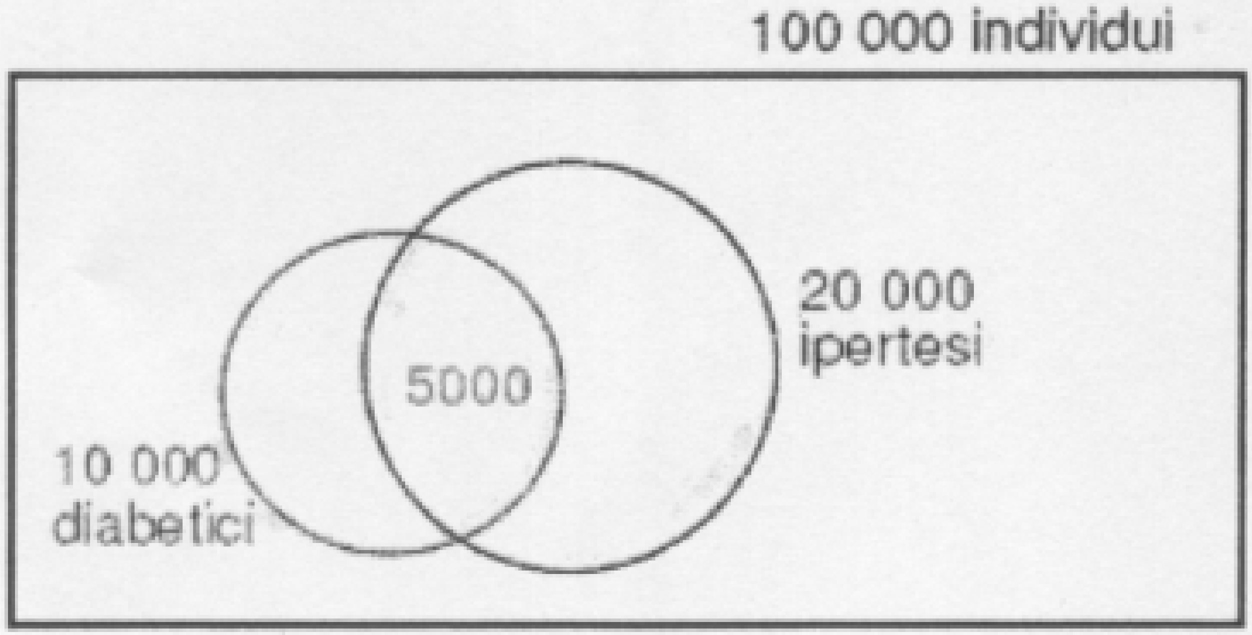
Se un esperimento è a più stadi, il problema di descrivere i possibili risultati può essere semplificato mediante l'uso di diagrammi ad albero.

Esempio: *Quanti figli maschi possono nascere su 3 gravidanze?*

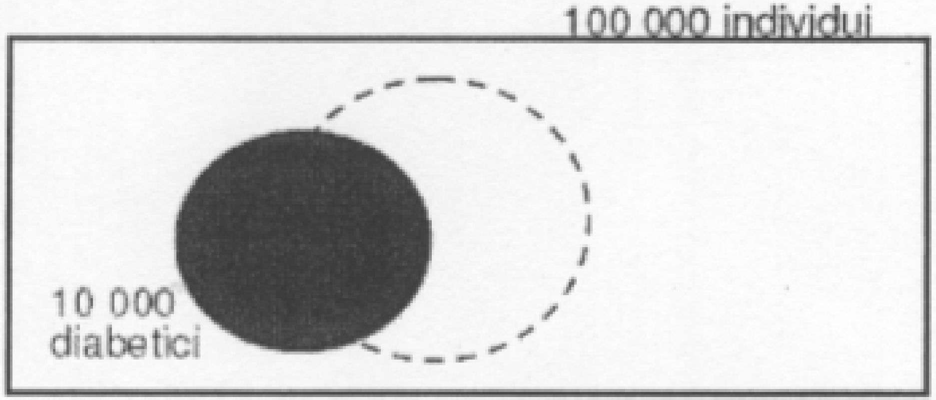


ESERCIZIO: CALCOLO DELLE PROBABILITA'

In una popolazione di 100 000 individui vi sono:
10 000 diabetici (e 90 000 non-diabetici)
20 000 ipertesi (e 80 000 non-ipertesi).
5000 persone che hanno sia il diabete che l'ipertensione.

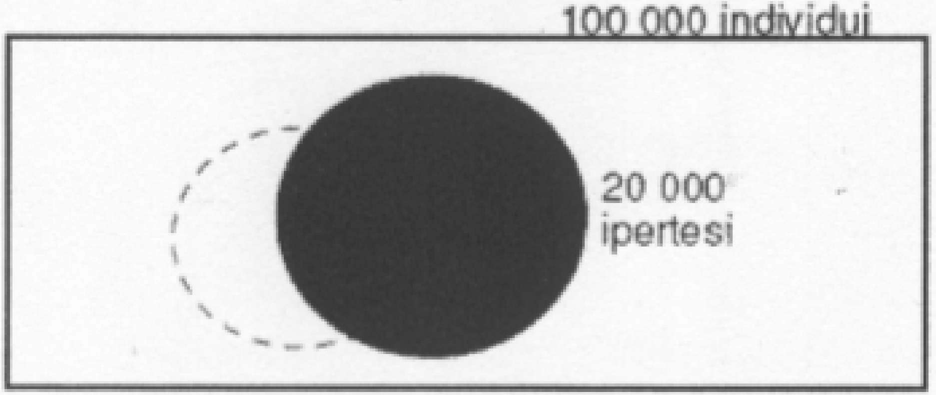


Qual è la probabilità di avere il diabete in quella popolazione?



p (diabete) = 10 000 / 100 000 = 0,1 = 10%

Qual è la probabilità di avere l'ipertensione in quella popolazione?



p (ipertensione) = 20 000 / 100 000 = 0,2 = 20%

N.B. E' stato usato l'approccio frequentista: la probabilità è stata stimata dalla frequenza relativa

DEFINIZIONI DI PROBABILITA'

2. FREQUENTISTA: la probabilità di un evento è il limite, al crescere del numero di esperienze, della serie delle frequenze dell'evento ottenute da esperienze fatte in condizioni uguali. (Von Mises, Neyman)

E' come dire che le frequenze relative osservate in un numero molto elevato di prove possono essere considerate come un'approssimazione della probabilità.

ESEMPIO: abbiamo simulato al computer un esperimento con due soli eventi, A e B, di cui uno favorevole (A) e l'altro contrario (B). I due eventi sono equiprobabili, cioè P(A)=P(B)=0,50. Eseguendo 10, 100, 1000 e 10000 prove abbiamo registrato i seguenti risultati.

N° prove	Fr. B	Fr. A	Fr. rel. A
10	4	6	0,60
100	43	57	0,57
1000	452	548	0,54
10000	4890	5110	0,51

Si nota chiaramente che al crescere del numero delle prove la frequenza relativa tende al valore teorico della probabilità: Fr. Rel. → 0,50

DEFINIZIONI DI PROBABILITA'

3. LOGICA: misura dell'aspettativa di un evento, sulla quale misura vi è concordanza della maggioranza degli individui. (Keynes, Boole)

4. SOGGETTIVISTA: stima personale di ciascun individuo del grado di aspettativa di un evento. (De Finetti, Savage)

In questa definizione la probabilità dipende dal grado di fiducia che un individuo ripone sul verificarsi di un dato fenomeno, fiducia che a sua volta dipende dalle informazioni che l'individuo ha a disposizione.

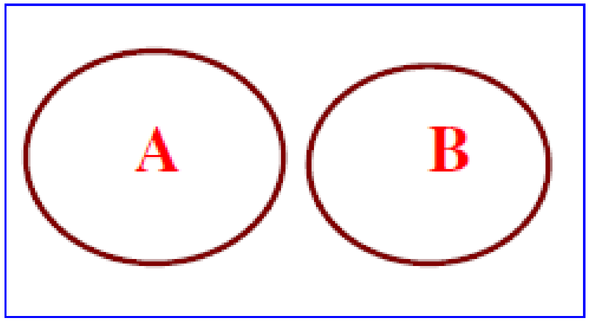
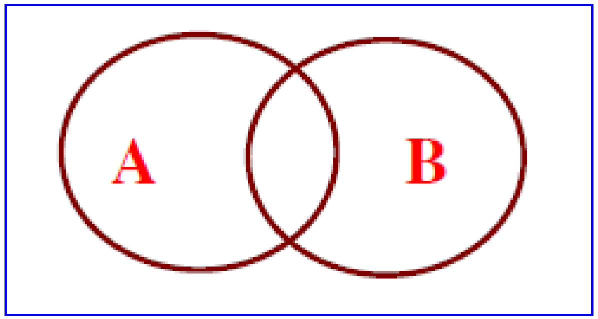
Per cui è possibile che persone diverse possano differire nel loro grado di fiducia di fronte alla stessa evidenza dei fatti.

ESEMPIO: un economista di fronte a un aumento del prezzo della benzina stima una probabilità di aumento del costo dei trasporti di 10 a 1. Si dirà quindi che la probabilità che tali costi aumentino è di 10/11 contro una probabilità di 1/11 che gli stessi restino invariati.

POSTULATI SULLE PROBABILITA'

POSTULATO DELLE PROBABILITA' TOTALI

Si distinguono 2 casi

EVENTI INCOMPATIBILI	EVENTI COMPATIBILI
	
$P(A \cup B) = P(A) + P(B)$	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

POSTULATI SULLE PROBABILITA'

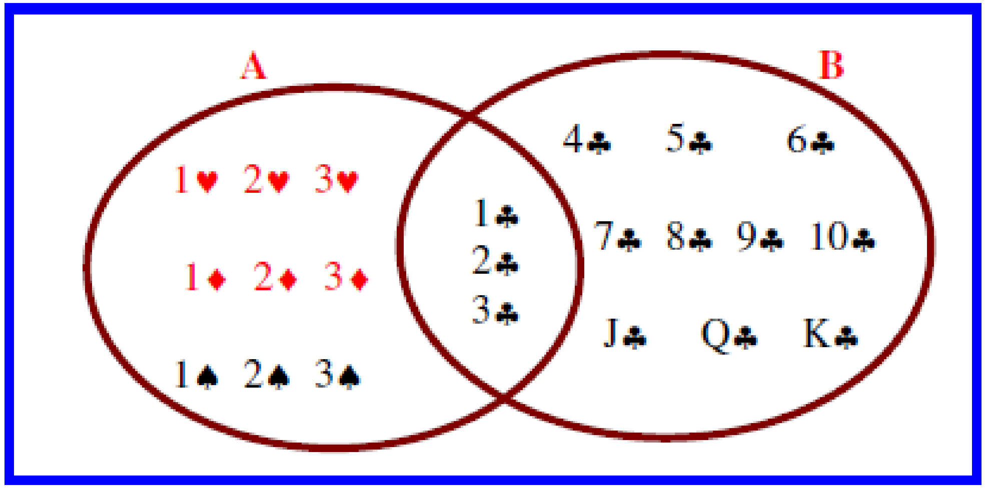
ESEMPIO: dato un mazzo di 52 carte definiamo i seguenti eventi:

A = (carta $\leq$3)

B = (carta di Fiori)

Si noti che questi 2 eventi sono chiaramente COMPATIBILI

Definiamo lo spazio campionario in riferimento agli eventi A e B e calcoliamo le probabilità P(A) e P(B)



Con il rapporto fra casi favorevoli e casi possibili otteniamo le probabilità seguenti:

$P(A) = 12 / 52$

$P(B) = 13 / 52$

PROBABILITA' CONDIZIONATA

Si definisce probabilità di A condizionato B, o A dato B, la seguente

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

Essa esprime la probabilità che si verifichi l'evento A (condizionato), sapendo che si è verificato l'evento B (condizionante)

ESEMPIO: si supponga di lanciare 3 monete e si definiscano i seguenti eventi:

A → escono almeno 2 Teste

B → esce C sulla prima moneta

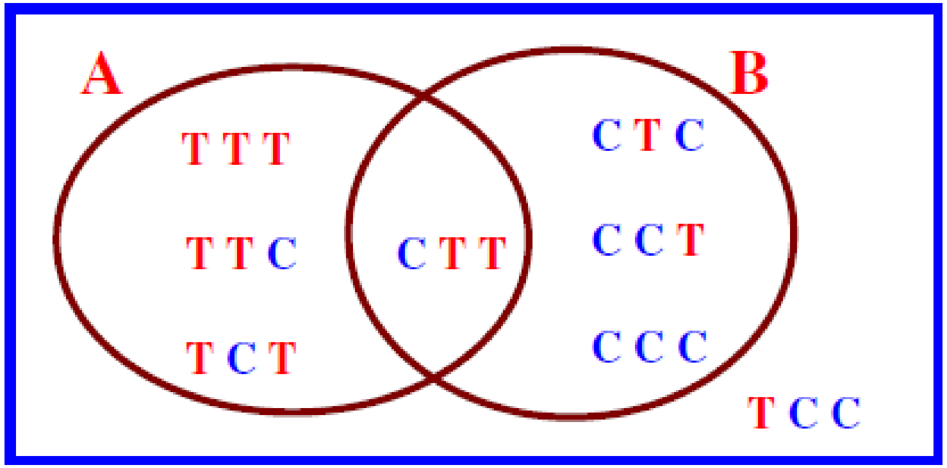
PROBABILITA' CONDIZIONATA

Definiamo lo spazio campionario:

A: almeno 2 Teste

B: C sulla prima moneta

Ci chiediamo qual'è la probabilità che escano 2T sapendo che si è verificata una C sulla prima moneta?



In altre parole ci stiamo chiedendo la probabilità condizionata P(A|B)

Siccome la probabilità dell'intersezione e la probabilità dell'evento B sono rispettivamente:

$$P(A \cap B) = \frac{1}{8}$$

$$P(B) = \frac{4}{8}$$

la probabilità cercata è uguale a:

$$P(A \mid B) = \frac{1/8}{4/8} = \frac{1}{4}$$

POSTULATI SULLE PROBABILITA'

POSTULATO DELLE PROBABILITA' COMPOSTE

Si distinguono 2 casi

EVENTI INDIPENDENTI

Il verificarsi di B **non** influenza **la probabilità** del verificarsi di A

$$P(A \cap B) = P(A) \times P(B)$$

EVENTI DIPENDENTI

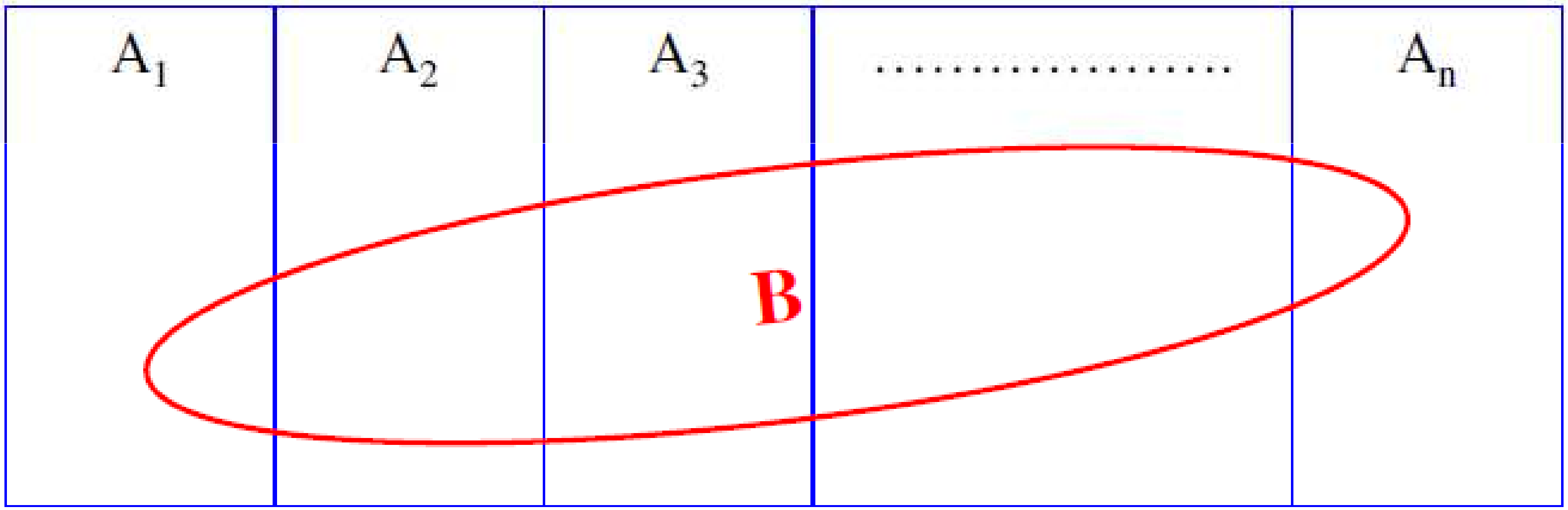
Il verificarsi di B influenza **la probabilità** del verificarsi di A

$$P(A \cap B) = P(A) \times P(B|A) = P(B) \times P(A|B)$$

TEOREMA DI BAYES

Il teorema di Bayes consente di calcolare la probabilità che, essendosi verificato un dato evento B, abbia agito la causa A_i in un gruppo di *n* cause A_1, A_2,…, A_n, incompatibili ed esaustive.

La situazione può essere schematizzata come segue:



Come si vede gli eventi A_i (le cause) formano una partizione completa dello spazio campionario S.

TEOREMA DI BAYES

Il teorema di Bayes si riassume con la seguente formula nella quale intervengono le probabilità a posteriori dell'ipotesi A_i dato l'effetto B, le probabilità a priori che hanno le singole ipotesi di verificarsi e le verosimiglianze (probabilità probative) che l'effetto sia stato causato da una data ipotesi:

Verosimiglianze: rappresentano le probabilità con cui le singole cause A_i generano l'evento B

Probabilità a priori: non dipendono dal risultato empirico dell'evento B.

$$P(A_i / B) = \frac{P(A_i)P(B / A_i)}{\sum_{i=1}^{n} P(A_i)P(B / A_i)}$$

Probabilità a posteriori: sapendo che l'evento B si è verificato, esse dicono con quale probabilità la causa i-esima ha agito nel determinare B.

TEOREMA DI BAYES

ESEMPIO: una fabbrica produce i suoi prodotti in 3 differenti stabilimenti, nelle seguenti percentuali: 50%, 10%, 40%.

Inoltre sappiamo che ogni stabilimento produce unità difettose nelle percentuali: 1%, 5%, 2%.

Dall'intera produzione si estrae 1 unità difettosa.

Quali sono le probabilità che provenga dal primo, secondo o terzo stabilimento?

Gli eventi si possono formalizzare come segue:

A_1 = {unità prodotta nello stabilimento 1}

A_2 = {unità prodotta nello stabilimento 2}

A_3 = {unità prodotta nello stabilimento 3}

B = {unità difettosa}

TEOREMA DI BAYES

I dati che abbiamo sono:

$$P(A_1) = 0.50 \qquad P(B/A_1) = 0.01$$

$$P(A_2) = 0.10 \qquad P(B/A_2) = 0.05$$

$$P(A_3) = 0.40 \qquad P(B/A_3) = 0.02$$

Per il postulato delle PROBABILITA' TOTALI risulta che:

$$P(B) = P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + P(A_3)P(B/A_3) =$$

$$= 0{,}005 + 0{,}005 + 0{,}008 = 0.018$$

TEOREMA DI BAYES

Applicando il teorema di Bayes la probabilità che l'unità difettosa provenga dal primo stabilimento è:

$$P(A_1 / B) = \frac{P(A_1)P(B / A_1)}{P(B)} = \frac{0.50 * 0.01}{0.018} \cong 0.278$$

la probabilità che l'unità difettosa provenga dal secondo stabilimento è:

$$P(A_2 / B) = \frac{P(A_2)P(B / A_2)}{P(B)} = \frac{0.10 * 0.05}{0.018} \cong 0.278$$

la probabilità che l'unità difettosa provenga dal terzo stabilimento è:

$$P(A_3 / B) = \frac{P(A_3)P(B / A_3)}{P(B)} = \frac{0.40 * 0.02}{0.018} \cong 0.444$$

Il teorema di Bayes.

- Il Teorema di Bayes permette di ricavare la probabilità di un dato evento (H), avendo osservato un fatto (evento E) eventualmente “originato” da H.
- L’evento E non è necessariamente la conseguenza diretta ed esaustiva di H, potrebbero esistere, infatti, altre "cause" che avrebbero potuto generare lo stesso evento.

Il teorema di Bayes.

- Esempio:
- Grossa vincita ad un gioco (H) → il vincitore diventa miliardario (E)
- Ma H non è una causa esclusiva di E. Se si osserva un soggetto miliardario (E) si può solo dire che è PROBABILE che H (la vincita) ne sia stata la causa.
- Inoltre data una causa H possono seguire diversi effetti (E_i).
- In generale, di fronte a diverse cause (H_i) e diversi effetti (E_i) non si può risalire, con certezza, dagli E_i alle H_i generatrici.

Il teorema di Bayes.

- Esempio 1:
- Da una coppia nascono 5 figlie femmine (E) la “causa” potrebbe essere attribuita alla tendenza della coppia a generare figlie femmine (H).
- Ipotizzando $P(F) \cong P(M) \cong 1/2$ l'evento verificatosi ha probabilità:
- $P(E)=(1/2)^5=1/32=0.03$ che risulta molto bassa.
- Il numero di figlie porterebbe a ritenere che quei genitori tendano a generare femmine, ma la figliolanza di 5 femmine è un evento raro (P(E)=0.03) dovuto al “CASO”. L’erroneità della precedente deduzione sta nel confondere la PROBABILITA’ che si verifichi un E con la PROBABILITA’ INVERSA (ignota) che, essendosi verificato un fatto (E), esso dipende dalla causa H (tendenza a generare figlie femmine), invece si è

Il teorema di Bayes.

- Tutta questa problematica è stata formalizzata da Bayes circa 2 secoli fa (1763).
- Il **teorema di Bayes** permette di determinare, data l'osservazione di un dato evento E, la probabilità degli eventi-cause H_1, H_2, …, H_i,…, H_n necessari ed incompatibili.
- Le **probabilità** $P(H_1)$, P(H2), …, $P(H_i)$, …, $P(H_n)$, dette **a priori,** sono stimabili dalla letteratura (ad esempio epidemiologica).
- Le **probabilità** $P(E\backslash H_1)$,…, $P(E\backslash H_i)$, …, $P(E\backslash H_n)$, dette **probative**, sono date dalla letteratura scientifica e stimano la probabilità di E condizionata al verificarsi di H_i
- La PROBABILITA' A POSTERIORI risponde alla formula (Bayes)

$$P(Hi/E) = \frac{P\,(Hi)\;P\,(E/Hi)}{P(Hi)P(E(/Hi) + \ldots + \ldots P(Hi)P(E/Hi) + \ldots + P(Hn)P(E/Hn)}$$

- probabilità che sia H_i la "causa" della realizzazione di E. Calcolando $P(H_i/E)$, per ciascun H_i, viene presa quale "causa" di E l'evento H_i per il quale $P(H_i/E)$ è più alta.

Il teorema di Bayes.

- <u>Esempio 2:</u>
- I gemelli possono essere veri gemelli (VG) (omozigoti) e sono dello stesso sesso, pseudo-gemelli (PG) (eterozigoti) e, in tal caso, è 1/2 la probabilità che siano dello stesso sesso.
- Consideriamo gli eventi:

  E = stesso sesso ("i gemelli sono dello stesso sesso",)

  E1 = omozigoti/e (VG = "i gemelli sono veri gemelli"),

  E2 = eterozigoti/e (PG = "i gemelli sono pseudo-gemelli").

- Probabilità a priori P(VG)=0.35 P(PG)=0.65
- Probabilità probative P(E/VG)=1 P(E/PG)=0.5

Il teorema di Bayes.

- Probabilità a posteriori: che dati due gemelli dello stesso sesso, siano omozigoti

$$P(VG/E) = \frac{P(VG)\ P(E/VG)}{P(VG)\ P(E/VG) + P(PG)\ P(E/PG)} = \frac{0.35 \cdot 1}{0.35 \cdot 1 + 0.65 \cdot 0.5} = 0.52;$$

$$P(PG/E) = \frac{P(PG)\ P(E/PG)}{P(VG)\ P(E/VG) + P(PG)\ P(E/PG)} = \frac{0.65 \cdot 0.5}{0.35 \cdot 1 + 0.65 \cdot 0.5} = 0.48$$

- 0.52 è la probabilità che avendo riscontrato due gemelli dello stesso sesso questi siano omozigoti, e 0.48 che siano eterozigoti.

Il teorema di Bayes.

- <u>Esempio 3:</u>
- E: l'infarto del miocardio
- C1 fumo da sigaretta, C2 ipertensione, C3 il caso (le altre cause)
- Prob a priori P(C1)=0.20, P(C2)=0.40, P(C3)=0.40
- Prob probative P(E/C1)=0.30 , P(E/C2)=0.80 , P(E/C3)=0.50
- Se E si è verificato con quale probabilità ha agito Ci?
- PROBABILITA' A POSTERIORI (Bayes)

$$P(C_1/E) = \frac{0.20 \cdot 0.30}{0.20 \cdot 0.30 + 0.40 \cdot 0.80 + 0.40 \cdot 0.50} = 0.10$$

- La probabilità di essere fumatore, se ha avuto l'infarto, è data da:

$$P(C_2/E) = \frac{0.40 \cdot 0.80}{0.20 \cdot 0.30 + 0.40 \cdot 0.80 + 0.40 \cdot 0.50} = 0.55$$

- La probabilità di essere iperteso, se ha avuto l'infarto, è

Il teorema di Bayes.

- Preso a caso un paziente infartuato, la probabilità che egli non sia né iperteso né fumatore (che abbiano agito altre cause, il caso) è data da:

$$P(C_3 / E) = \frac{0.40 \cdot 0.50}{0.20 \cdot 0.30 + 0.40 \cdot 0.80 + 0.40 \cdot 0.50} = 0.34$$

- allora la causa più probabile è C2 (l'ipertensione).

Il teorema di Bayes.

- **Esempio 4**
- Supponiamo che un paziente abbia l'ulcera duodenale (E).
- Con quale probabilità hanno agito le seguenti cause?
- C1: presenza di Helicobacter pilory (Hp^+)
- C2: Abitudini di vita non razionali (tipo di alimentazione, fumo, etc)
- C3: Il caso (le altre cause)
- Prob. a priori: P(C1)=0.50, P(C2)=0.30, P(C3)=0.20.
- Prob. probative: P(E/C1)=0.70, P(E/C2)=0.20, P(E/C3)=0.10
- La probabilità che un paziente avendo l'ulcera essa dipenda dalla presenza dell'Hp^+ è data da:

$$P(C_1/E) = \frac{0.50 \cdot 0.70}{0.50 \cdot 0.70 + 0.30 \cdot 0.20 + 0.20 \cdot 0.10} = 0.81$$

Il teorema di Bayes.

- La probabilità che un paziente avendo l'ulcera essa dipenda dall'igiene di vita scadente è data da:

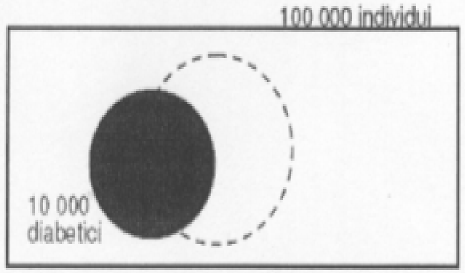
$$\mathrm{P}(\mathrm{C}_2 / E) = \frac{0.30 \cdot 0.20}{0.50 \cdot 0.70 + 0.30 \cdot 0.20 + 0.20 \cdot 0.10} = 0.14$$

- la probabilità che un paziente avendo l'ulcera essa dipenda da altre cause è data da:

$$\mathrm{P}(\mathrm{C}_3 / E) = \frac{0.20 \cdot 0.10}{0.50 \cdot 0.70 + 0.30 \cdot 0.20 + 0.20 \cdot 0.10} = 0.05$$

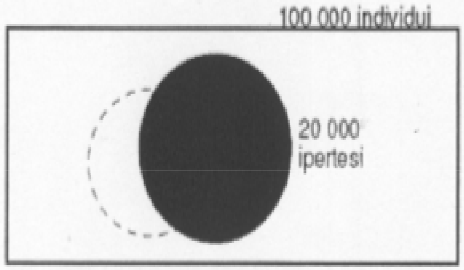
- E' molto probabile (81%), quindi, che l'ulcera sia stata causata dall'Hp^+.

Qual è la probabilità di avere il diabete in quella popolazione?



p (diabete) = 10 000 / 100 000 = 0,1 = 10%

Qual è la probabilità di avere l'ipertensione in quella popolazione?



p (ipertensione) = 20 000 / 100 000 = 0,2 = 20%

N.B. E' stato usato l'approccio frequentista: la probabilità è stata stimata dalla frequenza relativa