Probabilità

Cicchitelli Cap. 12

Alcune definizioni

- **Esperimento aleatorio** – un processo che porta ad un risultato incerto
- **Evento elementare** – un possibile risultato di un esperimento aleatorio
- **Spazio campionario** – l'insieme di tutti i possibili risultati di un esperimento aleatorio
- **Evento** – qualsiasi sottoinsieme di eventi elementari di uno spazio campionario

Esperimento aleatorio

- Un esperimento aleatorio descrive una situazione il cui esito è incerto
  - **giochi di sorte** (come il lancio di una moneta, l'estrazione di un numero al lotto, l'estrazione di un numero alla roulette),
  - **esperimenti di laboratorio** (come il test di durata di un pneumatico, la somministrazione di un principio attivo ad una cavia)
  - **misurazioni fisiche** (come la temperatura minima di domani in una certa stazione meteorologica)
  - **fenomeni economici e sociali** (come il numero di computer prodotti da un'impresa del settore, il PIL italiano fra 5 anni o il ROE di un impresa nel prossimo esercizio)
  - in generale tutte le prove, operazioni, attività o fenomeni il cui esito non è prevedibile con certezza.

Spazio campionario

- Dato un esperimento aleatorio, si dice **spazio campionario** l'insieme S di tutti i *possibili risultati*, esaustivi e mutualmente esclusivi, dell'esperimento stesso. Tali possibili risultati sono detti **punti campionari** o **eventi elementari**
  - Es. lanciando una moneta $S = \{T, C\}$
- Se l'esperimento aleatorio viene ripetuto *k* volte, lo spazio campionario complessivo è dato dal prodotto cartesiano $S \times S \times \ldots \times S$ *k* volte
  - Es. lanciando due volte una moneta lo spazio campionario complessivo è $\{T, C\} \times \{T, C\}$, i cui punti campionari sono *TT, TC, CT, CC*

Spazio campionario - esempi

- Un soggetto chiede un finanziamento ad una banca
  → S = {*concesso*, *rifiutato*}
- Estrazione di un numero al lotto
  → S = {1, 2, …, 90} (cardinalità finita)
- Numero di casi di influenza nel prossimo anno
  → S = {0, 1, 2, …} (cardinalità infinita numerabile)
- Tempo di attesa per essere serviti ad un sportello bancario
  → S = [0, +∞) (cardinalità infinita non numerabile)

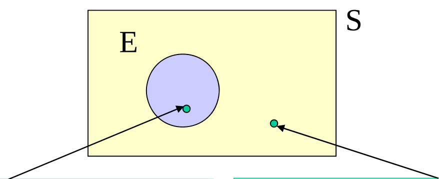
Eventi

- Dato uno spazio campionario S, un evento è un sottoinsieme di S, quindi è costituito da uno o più punti campionari (a parte il caso dell'*evento impossibile*, denotato con il simbolo dell'insieme vuoto ∅)
- Un evento *E* **si verifica** (si realizza) quando il risultato dell'esperimento casuale è un qualsiasi punto campionario di *E*; in caso contrario *E* **non si verifica**
- Esempio: lanciando un dado S = {1, 2, …,6}, alcuni dei possibili eventi sono:
  - A = {Numero pari} = {2,4,6}
  - B = {Numero minore o uguale a 3} = {1,2,3}

  se ad es. esce il 4 → A si verifica, mentre B non si verifica

Diagramma di Venn

- Logica delle proposizioni (eventi) ⇔ Operazioni sugli insiemi
- Lo spazio campione S è rappresentato da un rettangolo e un evento E è rappresentato da figura ivi contenuta



Se il risultato dell'esperimento aleatorio è un punto incluso in E → E si verifica

Se il risultato dell'esperimento aleatorio è un punto non incluso in E → E non si verifica

Diagramma di Venn - esempio

- Es. Lancio di un dado

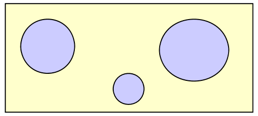
B = {Numero minore o uguale a 3}

1	3	5
2	4	6

S

Tipi di eventi

- Due o più eventi si dicono ***disgiunti*** o ***incompatibili*** o ***mutuamente esclusivi*** quando la realizzazione di uno esclude la realizzazione dell'altro\i. (due eventi elementari sono sempre incompatibili)

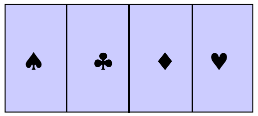


S

Es. Nell'estrazione di una carta, A={picche} e B={fiori} sono incompatibili

- Due o più eventi si dicono ***collettivamente esaustivi*** quando almeno uno di loro si verifica sicuramente

- Due o più eventi formano una ***partizione*** quando sono contemporaneamente disgiunti e collettivamente esaustivi

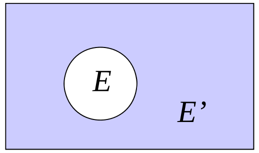


S

Es. Nell'estrazione di una carta, A={picche}, B={fiori}, C={quadri}, D={cuori} formano una partizione

Evento complementare

- Il ***complementare*** di un evento *E* è rappresentato dall'insieme di tutti gli altri elementi dello spazio campionario, e viene indicato con *E'* (oppure con $\bar{E}$)



S

Es. Nel lancio di un dado, posto *E* ={numero pari} si ha *E'* ={numero dispari}

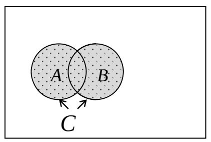
Un evento e il suo complementare formano una partizione

Unione ed intersezione

Le operazioni di unione e intersezione producono nuovi eventi

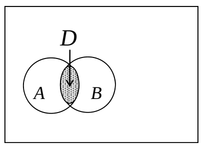
Unione di eventi:

- Produce un nuovo evento che si verifica quando *almeno* uno degli eventi uniti si verifica
- È indicata dalla simbologia $C=A\cup B$
- Lettura logica: "o"



Intersezione di eventi:

- Produce un nuovo evento che è vero quando *tutti* gli eventi in considerazione si verificano contemporaneamente
- È indicata dalla simbologia $D=A\cap B$
- Lettura logica: "e"



Unione ed intersezione - esempio

Es. nel lancio di un dado, poniamo

A = {Numero pari} = {2,4,6}

B = {Numero minore o uguale a 3} = {1,2,3}

Unione di eventi:

$C = A\cup B$

$= \{1,2,3,4,6\}$

1	3	5
2	4	6

Intersezione di eventi:

$D = A\cap B$

$= \{2\}$

1	3	5
2	4	6

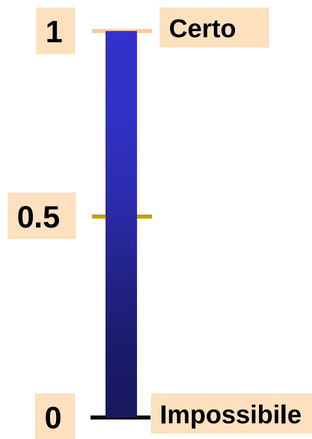
Eventi disgiunti, eventi esaustivi

- Usando le nozioni di unione e intersezione due eventi A e B sono
  - Disgiunti (incompatibili) quando $A \cap B = \varnothing$
  - Collettivamente esaustivi quando $A \cup B = S$

Probabilità

- Probabilità – la possibilità che un evento incerto si manifesti (sempre tra 0 e 1)

$0 \leq P(A) \leq 1$ Per qualsiasi evento A



Probabilità: definizioni

CLASSICA:

Probabilità come rapporto tra casi favorevoli su casi possibili

È basata su una conoscenza delle caratteristiche dell'esperimento indipendentemente dalla sua effettiva realizzazione

FREQUENTISTA:

È basato sulla sola osservazione dei dati, in assenza di informazioni preesistenti sulle modalità dell'esperimento

SOGGETTIVA:

Probabilità è il grado di fiducia associato al verificarsi di un certo evento di interesse espresso dal soggetto che esprime la valutazione probabilità

Probabilità classica

Assumendo che *S* abbia un numero finito di punti campionari *con uguale probabilità*, la probabilità di *E* è calcolata come

$$P(E) = \frac{\text{Numero di punti campionari in } E}{\text{Numero di punti campionari in } S} = \frac{\text{casi favorevoli}}{\text{casi possibili}}$$

Es. nel lancio di un dado, posto A = {Numero pari} = {2,4,6}

Approccio classico a priori: assumendo che tutti i punti campione abbiano la stessa probabilità (cioè che il dado sia bilanciato)

$$P(A) = \frac{3}{6} = 0.5$$

L'assunzione che tutti i punti campione abbiano la stessa probabilità (cioè che il dado sia bilanciato) è cruciale: se è vera il valore 0.5 è ben calcolato, se è falsa il valore 0.5 non va bene. Tuttavia, per verificare la plausibilità di tale assunzione bisogna ripetere più volte l'esperimento

Probabilità frequentista/1

Quando si osserva una serie di prove e si assume che le prove siano ripetizioni *indipendenti* e *in identiche condizioni* di un certo esperimento aleatorio → la probabilità di *E* è calcolata come

$$P(E) = \frac{\text{Numero di prove in cui si è verificato } E}{\text{Numero totale di prove}}$$

Lancio di un dado, posto A = {Numero pari} = {2,4,6}

- l'esperimento aleatorio "lancio del dado" viene ripetuto 50 volte, in 23 delle quali è uscito un numero pari e quindi si è verificato A.
- ➔ P(*A*) = 23/50=0.46.

Probabilità frequentista/2

- Osservando altre prove la stima si modifica
  - Per esempio, lanciando il dado altre 50 volte l'evento A si potrebbe verificare 26 volte
  - quindi la nuova stima sarebbe (23+26)/(50+50)=0.49.
- Per fortuna al crescere del numero di prove la stima diventa sempre più precisa e converge ad un valore
  - in questo esempio, se il dado è bilanciato converge a 0.5

Formalmente P(*E*) è definita come il limite a cui tende la proporzione di prove in cui si verifica *E* quando il numero di prove tende a infinito

Verifica empirica della probabilità /1

- L'approccio classico è *solo apparentemente oggettivo*, perché si basa sull'**ipotesi di equiprobabilità** degli eventi elementari: credere in tale ipotesi è *soggettivo*, perché dipende dalla conoscenza a priori del fenomeno (dado bilanciato, lanciatore onesto …)
- Nel caso in cui l'esperimento sia ripetibile più volte nelle stesse identiche condizioni si può stimare la probabilità come frequenza, per cui l'adeguatezza delle ipotesi può essere *valutata empiricamente* (nel senso che si possono raccogliere prove contro di essa).
- Se ad es. dopo un grande numero di ripetizioni in identiche condizioni la frequenza relativa con cui si è presentata testa è lontana da 0.5, allora l'ipotesi di equiprobabilità non tiene, per cui va sostituita con un'altra.
  - grande e lontana sono termini vaghi ma possono (e devono!) essere resi precisi, cioè sostituiti con numeri (come vedremo più avanti)

Verifica empirica della probabilità /2

- Indicando con *n* il numero di ripetizioni dell'esperimento aleatorio, se il valore ipotizzato della probabilità è corretto, al crescere di *n* la frequenza relativa osservata converge alla probabilità.
- Attenzione: per *n* piccolo la discrepanza può essere notevole anche se il modello è corretto.
  - Esempio: ipotizziamo che una moneta sia bilanciata, cioè ipotizziamo che P(T)=0.5, e per verificare tale modello probabilistico lanciamo la moneta *n* volte.
  - Vogliamo valutare l'evidenza empirica contro tale modello nel caso in cui esca sempre testa, cioè la frequenza relativa osservata di T è 1.
  - Chiaramente 1 è molto lontano da 0.5, tuttavia se *n* è molto piccolo l'evidenza empirica contro il modello ipotizzato è comunque debole. Infatti, se il modello ipotizzato è vero, la probabilità di osservare *n* teste su *n* lanci è $1/2^n$ e quindi, ad es., è 1/4 per *n*=2, 1/16 per *n*=4, 1/256 per *n*=8.

Probabilità soggettiva

La probabilità di un evento E è definita come il **grado di fiducia** che un individuo razionale attribuisce al verificarsi di un evento

- La misura (soggettiva) di probabilità si deriva ponendo l'individuo (razionale) di fronte ad un'operazione di scommessa chiedendo quanto è disposto a puntare per ricevere 1 nel caso in cui l'evento in questione si realizzi

Probabilità soggettiva/2

- Questo approccio **si può usare sempre**, in tutte le situazioni, ma è davvero importante quando gli approcci classico e frequentista non sono utilizzabili perché
  - Lo spazio campione S non è costituito da un insieme finito di punti equiprobabili → l'approccio classico è inutilizzabile
  - Non si dispone di osservazioni indipendenti e in identiche condizioni → l'approccio frequentista è inutilizzabile

es. Qual è la probabilità che il governo adotti un certo provvedimento?
Qual è la probabilità che la Fiorentina vinca la prossima partita?

Definizione assiomatica della probabilità

La probabilità è una *funzione d'insieme*, $P(\cdot)$, definita nello spazio campione S, con le seguenti proprietà:

1. $P(S) = 1$
2. $P(A) \geq 0$ per ogni evento A
3. $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$ per ogni successione di eventi di S a due a due incompatibili

Conseguenze degli assiomi

- $P(\emptyset) = 0$
- $P(A) \leq 1$
- Se $A \cup B = S$ allora $P(A \cup B) = 1$
- Se $A \cap B = \emptyset$ allora $P(A \cup B) = P(A) + P(B)$
- Se S è formato da N eventi elementari ugualmente possibili $E_1, E_2, \ldots, E_N$, allora $P(E_i) = 1/N$
- Se S è formato da N eventi elementari ugualmente possibili e l'evento A è costituito da N_A di questi elementi, allora $P(A) = N_A / N$

Regole della probabilità

- La Regola dell'evento complementare:

$$\boxed{P(\bar{A}) = 1 - P(A)} \qquad \text{ovvero} \quad P(A) + P(\bar{A}) = 1$$

- La Regola della somma (o additiva):
  – La probabilità dell'unione di due eventi è

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

  – Se A e B sono incompatibili $P(A \cap B) = P(\varnothing) = 0$, quindi l'espressione si semplifica in

$$\boxed{P(A \cup B) = P(A) + P(B)}$$

Tabella delle probabilità /1

Probabilità **congiunte** per due eventi A e B:

	B	$\bar{B}$	
A	$P(A \cap B)$	$P(A \cap \bar{B})$	**?**
$\bar{A}$	$P(\bar{A} \cap B)$	$P(\bar{A} \cap \bar{B})$	**? ?**

$$P(A \cap B) + P(A \cap \bar{B}) = P\big((A \cap B) \cup (A \cap \bar{B})\big) = P(A)$$

$$P(\bar{A} \cap B) + P(\bar{A} \cap \bar{B}) = P\big((\bar{A} \cap B) \cup (\bar{A} \cap \bar{B})\big) = P(\bar{A})$$

Tabella delle probabilità /2

Probabilità **congiunte** e **marginali** per due eventi A e B:

	B	$\bar{B}$	
A	$P(A \cap B)$	$P(A \cap \bar{B})$	$P(A)$
$\bar{A}$	$P(\bar{A} \cap B)$	$P(\bar{A} \cap \bar{B})$	$P(\bar{A})$
	$P(B)$	$P(\bar{B})$	$P(S) = 1$

$$P(A) + P(\bar{A}) = P(S) = 1$$

Regola additiva – esempio

Considera un mazzo di 52 carte, con i quattro semi: ♥ ♣ ♦ ♠

Evento A = la carta è un asso   Evento B = la carta è rossa

P(Rossa ∪ Asso) = P(Rossa) + P(Asso) - P(Rossa ∩ Asso)

= 26/52 + 4/52 - 2/52 = 28/52

Non contare due volte i due assi rossi!

Tipo	Colore		Totale
	Rossa	Nera	
Asso	2	2	4
Non-Asso	24	24	48
Totale	26	26	52

(probabilità classica: casi favorevoli su casi possibili)

Contare i possibili risultati

- Probabilità classica: casi favorevoli su casi possibili
- Talvolta è difficile contare i casi perché sono molti e non è pratico elencarli tutti, uno ad uno
- Soluzione: usare il **calcolo combinatorio**

Calcolo combinatorio

- Si considerano *n* oggetti presi *k* alla volta
- Es. abbiamo le lettere A, B e C e ne prendiamo 2 alla volta (quindi *n*=3 e *k*=2)

AB	BA	AA
AC	CA	BB
BC	CB	CC

Combinazioni $\underbrace{\text{AB, AC, BC}}_{C_{3,2}=3}$

Disposizioni $\underbrace{\text{AB, AC, BC, BA, CA, CB}}_{D_{3,2}=6}$

Disposizioni con ripetizione $\underbrace{\text{AB, AC, BC, BA, CA, CB, AA, BB, CC}}_{D^{(r)}_{3,2}=9}$

Due aspetti: Conta l'ordine? Contano le ripetizioni?

Disposizioni con ripetizione

- Disposizioni con ripetizione di *n* oggetti di classe *k* (cioè presi a *k* a *k*)

$$D^{(r)}_{n,k} = n \cdot n \ldots \cdot n = n^k$$

Esempio: quante password di 5 caratteri si possono creare con: i) un alfabeto di 26 lettere; ii) un alfabeto di 26 lettere dove maiuscole minuscole sono caratteri diversi; iii) alfabeto + numeri; iv) con il vincolo che i primi 3 caratteri sono alfabetici e gli altri 2 numerici

Con riferimento all'ultimo caso: estraendo una password a caso, qual è la probabilità che i numeri siano tutti inferiori a 5?

Disposizioni con ripetizione - esempio

Il poeta francese Raymond Queneau pubblicò un libro intitolato *Cent mille milliards de poèmes*, consistente in un sonetto su ognuna delle 10 pagine. Le pagine erano tagliate in modo che le 14 righe di ciascun sonetto potessero essere girate separatamente.

1 … … … … …
2 … … … …
3 … … … … …
4 … … … … …
5 … … … …
6 … … … … … …
7 … … … …
8 … … … … …
9 … … … …
10 … … … … …
11 … … … … …
12 … … … … … …
13 … … … …
14 … … … … …

Queneau affermò che tutti i risultanti 10^{14} sonetti (centomila miliardi, appunto) avevano un senso, sebbene sia ragionevole pensare che tale affermazione non sarà mai controllata!

Disposizioni semplici e permutazioni

- Disposizioni di *n* oggetti di classe *k*

$$D_{n,k} = n(n-1)(n-2)\ldots(n-k+1) = \frac{n!}{(n-k)!}$$

$$n! = n(n-1)(n-2)\ldots 3\cdot 2\cdot 1$$
$$0! = 1 \quad \text{(per definizione)}$$

- Caso speciale *k*=*n* → **permutazioni**: $P_n = D_{n,n} = n!$

Esempio del compleanno: qual è la probabilità che in un insieme di *k* persone almeno due compiano gli anni lo stesso giorno?

Combinazioni

- Combinazioni di *n* oggetti di classe *k*

$$\boxed{C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}}$$

Binomio di Newton

- Dimostrazione:

$$C_{n,k} = \frac{D_{n,k}}{P_k^k} = \frac{\frac{n!}{(n-k)!}}{k!} = \frac{n!}{k!(n-k)!}$$

Oss: $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}$ in particolare, $\binom{n}{0} = \binom{n}{n} = 1$

Combinazioni - esempi

- Esempio della tavola di Galton
- Estrazione senza reimbussolamento da un'urna con *N* palline di cui *R* rosse: qual è la probabilità che estraendo *n* palline ve ne siano *r* rosse? (*n*<*N*, *r*≤*R*)

$$\frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}$$

  - Probabilità di fare cinquina al gioco del lotto $\frac{1}{\binom{90}{5}}$
  - Probabilità di fare terno al gioco del lotto $\frac{\binom{5}{3}\binom{85}{2}}{\binom{90}{5}}$

Probabilità condizionata /1

- Consideriamo due eventi *A* e *B* e supponiamo di sapere che l'evento *B* si è verificato (quindi su *B* non vi è più incertezza)
  - In generale questa conoscenza modifica la probabilità dell'evento *A*
- Nell'approccio classico condizionarsi a *B* significa che i punti campione (casi possibili) da considerare al denominatore della probabilità non sono tutti quelli dello spazio campionario S, ma solo quelli contenuti in *B*
  - in altri termini, lo spazio campionario va modificato alla luce delle informazioni sopraggiunte → *B* è il nuovo spazio campionario dell'esperimento

Probabilità condizionata /2

- La probabilità di *A* condizionatamente a *B*, detta anche <u>probabilità di *A* dato *B*</u> e scritta *P*(*A* | *B*), consiste nella valutazione della probabilità di un evento *A* valutato subordinatamente allo spazio campionario generato dall'evento *B*
- Approccio classico: *P*(*A* | *B*) è il rapporto tra il numero di casi favorevoli (punti campione per cui si verificano *A* e *B*) ed il numero di casi possibili (punti campione per cui si verifica *B*):

Probabilità congiunta

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

Probabilità marginale

Probabilità condizionata

[la definizione richiede che P(B)>0]

Probabilità condizionata – es. 1

Riprendiamo l'esempio del lancio di un dado, ponendo

A = {Numero pari} = {2,4,6} → *P*(*A*)=1/2

B = {Numero minore o uguale a 3} = {1,2,3} → *P*(*B*)=1/2

Si noti che $A \cap B$ = {2} → $P(A \cap B)$=1/6

1	3	5
2	4	6

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3} \quad \left(= \frac{\text{punti campionari in } A \cap B}{\text{punti campionari in } B} \right)$$

P(*A* | *B*) ≠ *P*(*A*) → l'informazione che *B* si è verificato cambia la probabilità di *A*, in questo esempio *diminuisce*

Probabilità condizionata – es. 2

Continuiamo l'esempio del lancio di un dado, ponendo

A = {Numero pari} = {2,4,6} → *P*(*A*)=1/2

C = {Numero tra 2 e 4} = {2,3,4} → *P*(*C*)=1/2

Si noti che $A \cap C$ = {2,4} → $P(A \cap C)$=2/6=1/3

1	3	5
2	4	6

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{2/6}{3/6} = \frac{2}{3} \quad \left(= \frac{\text{punti campionari in } A \cap C}{\text{punti campionari in } C} \right)$$

P(*A* | *C*) ≠ *P*(*A*) → l'informazione che *C* si è verificato cambia la probabilità di *A*, in questo esempio *aumenta*

Probabilità condizionata – es. 3

Continuiamo l'esempio del lancio di un dado, ponendo

A = {Numero pari} = {2,4,6} → *P*(*A*)=1/2

D = {Numero minore o uguale a 2} = {1,2} → *P*(*C*)=1/3

Si noti che $A \cap D$ = {2} → $P(A \cap D)$=1/6

1	3	5
2	4	6

$$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{1/6}{2/6} = \frac{1}{2} \quad \left(= \frac{\text{punti campionari in } A \cap D}{\text{punti campionari in } D} \right)$$

P(*A* | *D*) = *P*(*A*) → l'informazione che *D* si è verificato *non* cambia la probabilità di *A* → <u>*A* è indipendente da *D*</u>

Probabilità condizionata – es. 4

Negli esempi precedenti

$P(A) = 1/2 = 0.50$

$P(A \mid B) = 1/3 = 0.33$ diminuisce

$P(A \mid C) = 2/3 = 0.67$ aumenta

$P(A \mid D) = 1/2 = 0.50$ invariata (indipendenza)

Osservazione: con riferimento alla probabilità di *A*, il condizionamento a *B* ha un effetto opposto al condizionamento a *C* anche se $P(B)=P(C) \rightarrow$ il valore della probabilità dell'evento condizionante non ha niente a che fare con l'effetto del condizionamento

A dato B oppure B dato A?

In generale $\boxed{P(A \mid B) \neq P(B \mid A)}$

Ad es. $P(A) = 0.8 \quad P(B) = 0.4 \quad P(A \cap B) = 0.2$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.2}{0.4} = 0.5$$

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{0.2}{0.8} = 0.25$$

L'errore di pensare che le due condizionate siano uguali è noto in ambito giudiziario come *fallacia dell'accusatore*

Es. A = "il DNA di Tizio è compatibile con quello dell'assassino", B = "Tizio è innocente", P(A | B) = 1/100000000, P(B | A) = ?

Regola moltiplicativa

- Applicando la definizione di probabilità condizionata, è anche possibile formulare la probabilità congiunta tramite le condizionate:

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

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

Indipendenza statistica /1

- Si parla di *indipendenza statistica* (o stocastica) quando la conoscenza dell'evento *B* **non** modifica la probabilità che si verifichi l'evento *A*, cioè

$$P(A \mid B) = P(A)$$

Probabilità condizionata Probabilità marginale

- La realizzazione dell'evento *B* è *ininfluente* per determinare la probabilità dell'evento *A*

Indipendenza statistica /2

Dalle relazioni $\boxed{P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)}$

viste in precedenza segue che l'*indipendenza statistica* può essere espressa in 3 modi equivalenti (si assuma che A e B siano eventi di probabilità non nulla):

$$\boxed{P(A|B) = P(A)} \longleftrightarrow \boxed{P(B|A) = P(B)}$$

$$\boxed{P(A \cap B) = P(A)P(B)}$$

Moltiplicare le probabilità?

- La probabilità congiunta di due eventi (= probabilità dell'intersezione) è uguale al prodotto delle probabilità marginali <u>se e solo se</u> i due eventi sono indipendenti
- Moltiplicare le probabilità quando non vi è indipendenza è un errore comune

Esempio: un testimone ha visto una persona sul luogo del delitto, e ne ricorda alcune caratteristiche: capelli neri, occhi neri, barba. Qual è la probabilità che una persona presa a caso dalla popolazione abbia le caratteristiche indicate? Supponiamo che le probabilità siano le seguenti:

A = (capelli neri)	P(A) = 5/10	
B = (occhi neri)	P(B) = 3/10	P(B|A) = 6/10
C = (barba)	P(C) = 1/10	P(C|A,B) = 5/10

La probabilità corretta è $P(A \cap B \cap C) = P(A)\, P(B|A)\, P(C|A,B) = 150/1000 = 0.15$

Mentre con la regola del prodotto si ottiene $P(A)\, P(B)\, P(C) = 15/1000 = 0.015$

Dai dati alle probabilità

- Seguendo l'approccio frequentista, una **distribuzione di frequenza** può fornire stime di probabilità di alcuni eventi
- Assunzione fondamentale: le probabilità devono essere stabili nel **tempo** e nello **spazio**, per cui la stima fatta in base a ciò che è accaduto ieri a alcuni soggetti (la distribuzione di frequenza) è valida anche per ciò che accadrà domani ad altri soggetti

- 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: $A \cap B \cap C = \varnothing$ e $A \cup B \cup C = S$

Domanda: 0.29 è una buona stima della probabilità che un consumatore <u>italiano</u> acquisti <u>oggi</u> 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.

- 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 Si	No	Totale
Si	200	50	250
No	100	650	750
Totale	300	700	1000

Frequenza *congiunta* assoluta   Frequenza *marginale* assoluta

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

Acquisto pianificato	Acquisto effettivo Sì (B)	No (B')	Totale
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 = $P(S)$

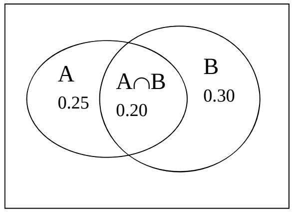
probabilità *marginale*:
acquisto effettivo

probabilità *marginale*:
acquisto non pianificato

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

Esempio: P(A∪B)

- Supponiamo di essere interessati alla probabilità dell'evento d'interesse: *C* = {acquisto pianificato *o* acquisto effettivo}:



Acquisto pianificato	Acquisto effettivo Sì (B)	No (B')	Totale
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 Si	No	Totale
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 riposta 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: Sì (B)	Acquisto effettivo: No (B')	Totale
Sì (A)	0.20 = P(A∩B)	0.05 = P (A∩B')	0.25 = P (A)
No (A')	0.10 = P (A'∩B)	0.65 = P (A'∩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: Sì (B)	Acquisto effettivo: No (B')	Totale
Sì (A)	0.20 = P(A∩B)	0.05 = P (A∩B')	0.25 = P (A)
No (A')	0.10 = P (A'∩B)	0.65 = P (A'∩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: Sì (B)	Acquisto effettivo: No (B')	Totale
Sì (A)	0.20 = P(A∩B)	0.05 = P (A∩B')	0.25 = P (A)
No (A')	0.10 = P (A'∩B)	0.65 = P (A'∩B')	0.75 = P (A')
Totale	0.30 = P (B)	0.70= P (B')	1 = P(S)

In precedenza abbiamo calcolato

$P(A \mid B) = 0.67$ e $P(B \mid A) = 0.80$

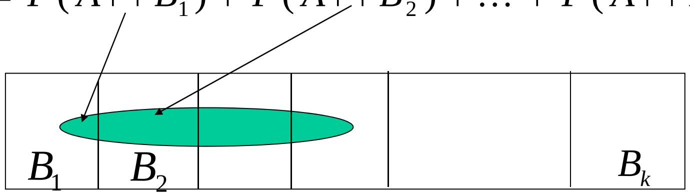
Per stabilire che *A* e *B* non sono indipendenti basta fare una delle seguenti verifiche (è sufficiente una sola):

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

Probabilità marginale

Data una partizione $B_1, B_2, \ldots, B_k$, e dato un evento *A* di cui si conoscono le probabilità delle intersezioni con gli elementi della partizione, la probabilità di *A* è

$$\begin{aligned} P(A) &= P(A\cap S) = P\big(A\cap(B_1\cup B_2\cup\ldots\cup B_k)\big) \\ &= P\big((A\cap B_1)\cup(A\cap B_2)\cup\ldots\cup(A\cap B_k)\big) \\ &= P(A\cap B_1)+P(A\cap B_2)+\ldots+P(A\cap B_k) \end{aligned}$$



Questo metodo si chiama marginalizzazione perché è proprio quello che si usa nelle tabelle doppie per calcolare le frequenze marginali

Formula delle probabilità totali

Riprendendo la formula della probabilità marginale

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_k)$$

e applicando la regola moltiplicativa si ottiene

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_k)P(B_k)$$
$$= \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

La prob di A si ottiene come *media pesata* delle prob di A dato B_i, con pesi pari alle prob di B_i

Esempio: probabilità totali

Posto

A = {acquisto pianificato}

$B_1 = B$ = {acquisto effettuato}

$B_2 = B'$ = {acquisto *non* effettuato}

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$$
$$= \left(\frac{200}{300}\right)\left(\frac{300}{1000}\right) + \left(\frac{50}{700}\right)\left(\frac{700}{1000}\right)$$
$$= 0.20 + 0.05 = 0.25$$

Ragionamento per scenari

- La formula delle probabilità totali corrisponde ad un *ragionamento per scenari* usato molto spesso
- Es. A={l'anno prossimo il fatturato della mia azienda aumenta} e B_1, B_2, B_3 sono 3 scenari relativi all'andamento del settore economico in cui opera l'azienda (crescita, stazionario, recessione). Ragionare per scenari significa attribuire la probabilità non direttamente ad A, ma ad A dato B_1, A dato B_2 ecc. (perché l'attribuzione è più facile) e poi derivare la probabilità di A tramite la formula delle probabilità totali
  - Supponiamo di attribuire le seguenti probabilità: $P(A|B_1)=0.8$, $P(A|B_2)=0.5$, $P(A|B_3)=0.1$
  - E' necessario attribuire una probabilità anche agli scenari, supponiamo $P(B_1)=0.1$, $P(B_2)=0.7$, $P(B_3)=0.2$ (ovviamente la somma è 1)
  - Pertanto $P(A)= 0.8\times0.1+0.5\times0.7+0.1\times0.2 = 0.45$

Formula di Bayes /1

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)}$$
$$= \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_k)P(B_k)}$$

dove B_i è l'i-mo di k eventi costituenti una partizione

Interpretazione: gli eventi della partizione $B_1,\ldots,B_k$ sono i possibili *stati di natura*, mentre A è un nuovo fatto (*evidenza empirica*)

Prima di osservare il fatto A, gli stati di natura hanno certe probabilità $P(B_i)$, dette ***a priori***. La formula di Bayes consente di *aggiornare* tali probabilità alla luce del fatto A, ottenendo le probabilità $P(B_i | A)$, dette ***a posteriori***

Formula di Bayes /2

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \ldots + P(A \mid B_k)P(B_k)}$$

La formula è utile quando si dispone delle

- probabilità marginali $P(B_i)$, dette ***probabilità a priori***, e
- probabilità condizionate $P(A \mid B_i)$, dette ***verosimiglianze***.

La formula è detta anche "della probabilità inversa" perché consente di invertire il verso del condizionamento nei casi in cui si dispone di $P(A \mid B_i)$ ma siamo invece interessati a $P(B_i \mid A)$

Esempio: chi è il colpevole?

- In un problema di investigazione per omicidio
  - B_1 = {il colpevole è Tizio}
  - B_2 = {il colpevole è Caio}
  - B_3 = {il colpevole non è né Tizio, né Caio}
  - A = {uno specifico indizio}
- B_1, B_2, B_3 costituiscono una partizione
- L'investigatore attribuisce a B_1, B_2, B_3 delle probabilità a priori (cioè prima di osservare l'indizio A)
- Se l'indizio è rilevante le probabilità si modificano, cioè $P(B_i \mid A) \neq P(B_i)$, e la formula di Bayes consente di calcolarle

Esempio: test diagnostico /1

- Tipica applicazione in medicina
  - D = soggetto affetto da una certa malattia (Disease)
  - D'= soggetto non affetto da quella malattia
  - T = Test positivo (cioè segnala la malattia)
  - T' = Test negativo
- Si vuole determinare $P(D \mid T)$ = probabilità che un soggetto per il quale il test dà esito positivo sia effettivamente affetto dalla malattia in questione

Esempio: test diagnostico /2

- Fino a che non si conosce l'esito del test, il soggetto ha una probabilità $P(D)$ di avere la malattia in questione. Tale probabilità viene stimata tramite la ***prevalenza*** nella popolazione
  - Supponiamo che la prevalenza sia di 3 persone ogni 100: pertanto la probabilità che il soggetto abbia la malattia è $P(D) = 0.03$, mentre la probabilità che sia sano (nel senso che non ha quella malattia) è $P(D') = 0.97$

Esempio: test diagnostico /3

- Il secondo elemento necessario per il calcolo è costituito dalla capacità del test di segnalare correttamente chi è sano e chi è malato:
  - Sensitività: probabilità di segnalare correttamente che un soggetto ha quella malattia: $P(T \mid D) = 0.90$
    - da ciò segue $P(T' \mid D) = 0.10$ (nel 10% dei soggetti malati il test sbaglia perché non rivela la malattia – *falso negativo*)
  - Specificità: probabilità di segnalare correttamente che un soggetto non ha quella malattia: $P(T' \mid D') = 0.98$
    - da ciò segue $P(T \mid D') = 0.02$ (nel 2% dei soggetti sani il test sbaglia perché segnala la malattia – *falso positivo*)

Esempio: test diagnostico /4

Tipi di errore:

Realtà \ **Esito test**	***Sano*** (T')	***Malato*** (T)
Sano (D')	OK P(T'|D')=0.98	Falso positivo P(T|D')=0.02
Malato (D)	Falso negativo P(T'|D)=0.10	OK P(T|D)=0.90

I due tipi di errore sono ben diversi, sia concettualmente che per le conseguenze!

Esempio: test diagnostico /5

Applicando la formula di Bayes …

$$P(D \mid T) = \frac{P(T \mid D)\,P(D)}{P(T \mid D)\,P(D) + P(T \mid D')\,P(D')}$$

$$= \frac{(0.90)(0.03)}{(0.90)(0.03) + (0.02)(0.97)}$$

$$= \frac{0.0270}{0.0270 + 0.0194} = \frac{0.0270}{0.0464} = 0.582$$

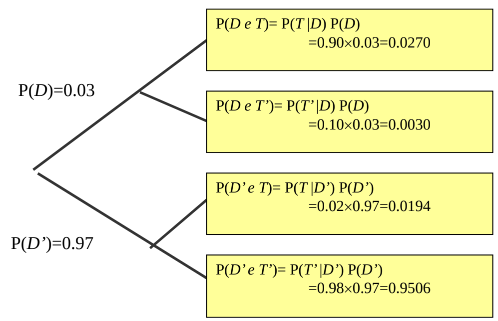
che è molto più grande di $P(D)=0.03$

ma molto più piccolo di $P(T \mid D) = 0.90$

Talvolta i medici confondono $P(D \mid T)$ con $P(T \mid D)$, affermando che il paziente positivo al test ha una prob. 0.90 di avere la malattia in questione

Esempio: test diagnostico /6

Evento D_i	Probabilità a priori $P(D_i)$	Probabilità condizionata $P(T \mid D_i)$	Probabilità congiunta $P(T \mid D_i)\,P(D_i)$	Probabilità aggiornata $P(D_i \mid T)$
D = soggetto malato	0.03	0.90	0.0270	0.582
D' = soggetto sano	0.97	0.02	0.0194	0.418



Esempio: test diagnostico /7

Cosa succederebbe se lo stesso test fosse usato per una malattia piuttosto rara, con prevalenza di 3 su 1000?

Applicando la formula di Bayes con $P(D)$=0.003:

$$\mathrm{P}(D \mid T) = \frac{(0.90)(0.003)}{(0.90)(0.003) + (0.02)(0.997)} = 0.119$$

Il test produrrebbe una quantità enorme di falsi positivi, cioè soggetti sani per i quali il test segnala la malattia (questo fa capire perché gli screening di massa siano problematici)

A chi è incuriosito dalla probabilità e vuole capire il suo ruolo nella vita quotidiana, suggerisco due eccellenti libri divulgativi