

# Sistemi Intelligenti Introduzione al calcolo delle probabilità - I

Alberto Borghese

Università degli Studi di Milano  
Laboratory of Applied Intelligent Systems (AIS-Lab)  
Dipartimento di Informatica  
[borghese@di.unimi.it](mailto:borghese@di.unimi.it)



A.A. 2025-2026

1/56

<http://borghese.di.unimi.it>

1



## Overview



Probabilità semplice e condizionata

Teorema di Bayes

A.A. 2025-2026

2/56

<http://borghese.di.unimi.it>

2



## Lo statistico



## Incertezza



- Le azioni “intelligenti” vengono fatte verso un ambiente che presenta una certa dose di **incertezza nel futuro**.

E.g. Dobbiamo andare a Malpensa. Quanto tempo prima dobbiamo partire?

Dalla nostra esperienza deriviamo che 60 minuti sono sufficienti se.....

Rimane un po’ di incertezza. Se partiamo 120 minuti prima ci teniamo un margine, ma passeremo facilmente tanto tempo in aeroporto senza fare nulla.

Quando prendiamo una decisione, teniamo conto in modo più o meno esplicito di questi elementi di incertezza legati al risultato delle azioni. Questi elementi hanno a che fare con la statistica.



## Probabilità (visione frequentista)



$$P(A = a_i) = \lim_{N \rightarrow \infty} \frac{n_{A=a_i}}{N} = \lim_{N \rightarrow \infty} \frac{n_i}{N}$$

Per il teorema del limite centrale la frequenza di un evento su infinite realizzazioni è uguale alla sua probabilità.

La probabilità che si verifichi uno tra i vari casi possibili è sempre 1. Ovverosia la somma delle probabilità di tutti gli eventi (se mutuamente esclusivi) somma 1.

Supponiamo  $A = \{a_1, a_2\}$

$$P(A = a_1) \cup P(A = a_2) = \lim_{N \rightarrow \infty} \frac{n_{A=a_1}}{N} + \lim_{N \rightarrow \infty} \frac{n_{A=a_2}}{N} = \lim_{N \rightarrow \infty} \frac{n_{A=a_1} + n_{A=a_2}}{N} = 1$$

$$P(A) = P(A = a_1) + P(A = a_2) = 1$$

**Probabilità totale**

$$P(A) = \sum_{k=1}^N \left( \lim_{N \rightarrow \infty} \left( \frac{n_{A=a_k}}{N} \right) \right) = 1$$



## Altri modi di definire la probabilità



- **Visione oggettivista.** Tendenza di un fenomeno ad accadere. Se lanciamo una moneta in aria, possiamo affermare che avremo 50% di probabilità che esca testa e 50% che esca croce. Ci aspettiamo che questa affermazione venga supportata quando effettuiamo infiniti esperimenti.
- **Visione soggettivista.** La probabilità viene espressa come credenza del soggetto. “Secondo me la probabilità di avere una carie è del 10%”. Non dipendono da un ragionamento fisico e rappresentano una probabilità a-priori. Deve potere essere corretta quando arrivano evidenze sperimentali.



## Semantica della probabilità



Possiamo ottenere un **grado di credenza (belief)** nell'affermazione. Questa potrà rivelarsi vera o falsa con una certa probabilità (e.g. «c'è una carie?»).

La probabilità è basata sulla conoscenza (a-priori) non sull'evento che si è già verificato!! La conoscenza a-priori è ricavata dall'analisi di tutti gli altri casi già osservati o su una conoscenza oggettiva o soggettiva dell'evento.



**La probabilità ci consente di trattare le diverse possibilità di un evento.**

**E ci consente di associare agli eventi un grado di credenza, prima che si verifichino.**



## Problematiche associate alla probabilità



Problema della visione frequentista:

- **Omogeneità del campione** (classe di riferimento). Come posso effettuare la media di eventi in modo “sicuro”?
- **Limitatezza del campione**

**Laziness** (svogliatezza). Non si riescono ad elencare tutte le situazioni associate al mal di denti

**Ignoranza teorica**. Non abbiamo una conoscenza che spieghi tutto nel dominio di interesse.

**Ignoranza pratica**. Anche se avessimo una conoscenza completa, non riusciamo a conoscere le condizioni esatte in cui si verifica l'evento (mal di denti del paziente).





## Probabilità semplice



Qual'è la probabilità che la proposizione: "E' uscito 6" si avveri?



$P(N=6) = 1/6$  secondo la visione oggettivista.

E' una probabilità semplice o assoluta perché non è legata al verificarsi di altri eventi.

Estendiamo a più eventi.



## Connettivi logici nel dominio della probabilità



- La probabilità che si verifichi uno tra più eventi indipendenti è data dalla somma delle probabilità
- La probabilità che si verifichino tutti gli eventi indipendenti è data dal prodotto delle probabilità.
- *Nel caso di eventi con una dipendenza tra di loro, appartenenti a insiemi non disgiunti, occorre correggere la somma e il prodotto.*



## Combinazione di probabilità con i connettivi logici (probabilità congiunta)



Qual'è la probabilità che la proposizione: "E' uscito 12" tirando due dadi si avveri?



Probabilità congiunta di due eventi indipendenti

$$P(N=12) = P(\text{dado}_1 = 6 \text{ AND dado}_2 = 6) = P(\text{dado}_1 = 6, \text{dado}_2 = 6) = P(\text{dado}_1 = 6) P(\text{dado}_2 = 6) = 1/6 * 1/6 = 1/36.$$

Nel caso di **indipendenza**, la probabilità di A e B è data dal **prodotto** delle probabilità:  
 $P(X = A \text{ AND } Y = B) = P(X=A) P(Y=B) = P(X=6) P(Y=6) = 1/6 * 1/6 = 1/36$

E' la frequenza con cui uscirebbe la configurazione «doppio 6» su un numero infinito di lanci.



## Combinazione di probabilità con i connettivi logici (probabilità congiunta)



Qual'è la probabilità che la proposizione: "E' uscito 11" tirando due dadi si avveri?

Nel caso di eventi indipendenti

$$P(N=11) = P(\text{dado}_1 = 5 \text{ AND dado}_2 = 6) \text{ OR } P(\text{dado}_1 = 6 \text{ AND dado}_2 = 5) = 1/6 * 1/6 + 1/6 * 1/6 = 2/36$$



Probabilità congiunta di due eventi indipendenti

Probabilità Totale



$$P(N=11) = P(X=5 \text{ AND } Y = 6) \text{ OR } P(X=6 \text{ AND } Y = 5) = 1/6 * 1/6 + 1/6 * 1/6 = 1/36 + 1/36 = 2/36$$



## Combinazione di probabilità con i connettivi logici (probabilità congiunta)



Qual'è la probabilità che la proposizione: "E' uscito almeno un 6" tirando due dadi si avveri?

$$P(N=6) = P( (\text{dado}_1 == 6 \text{ AND } \text{dado}_2 \neq 6) \text{ OR } (\text{dado}_1 \neq 6 \text{ AND } \text{dado}_2 == 6) \text{ OR } ((\text{dado}_1 == 6 \text{ AND } \text{dado}_2 == 6)) = 1/6*5/6 + 1/6*5/6 + 1/6*1/6 = 11/36$$



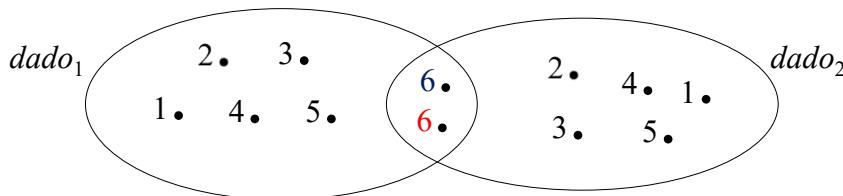
## Combinazione di 2 variabili con dipendenze con gli insiemi



Qual'è la probabilità che la proposizione: «E' uscito almeno un 6» si avveri lanciando 2 dadi?

$$P(N=6) = P(\text{dado}_1 = 6 \text{ OR } \text{dado}_2 = 6) - P(\text{dado}_1 = 6 \text{ AND } \text{dado}_2 = 6) = 1/6 + 1/6 - 1/6*1/6 = 11/36$$

Non voglio contare due volte la probabilità di ottenere 6 (intersezione tra i due insiemi).



In generale:  $P(X = A \text{ OR } Y = B) = P(X=A) + P(Y=B) - P(X=A \text{ AND } Y=B)$ .  
Unione dei due insiemi.

Se gli eventi sono disgiunti, appartengono a due insiemi diversi:  
 $P(X = A \text{ OR } Y = B) = P(X=A) + P(Y=B) - 0$ .



## Probabilità di 2 variabili con dipendenza



OR:

$$(X = A \text{ OR } Y = B) = P(X=A) + P(Y=B) - P(X=A \text{ AND } Y=B) \Rightarrow \\ \text{se sono indipendenti} = P(X=A) + P(X=B)$$

Conto due volte gli elementi nell'intersezione dei 2 insiemi, tengo 1 occorrenza.

AND:

$$P(X = A \text{ AND } Y = B) = P(X=A) + P(Y=B) - P(X=A \text{ OR } Y=B) \Rightarrow \\ \text{se sono indipendenti} = P(X=A) + P(X=B) - (P(X=A) + P(Y=B) - P(X=A \text{ AND } Y=B)) = \\ P(X=A) P(Y=B)$$

Conto due volte gli elementi nell'intersezione dei 2 insiemi, tolgo tutti gli elementi 1 volta, rimane 1 occorrenza di un elemento nell'intersezione.



## Paradosso del compleanno



Qual è la probabilità che due compagni di classe siano nati lo stesso giorno?

Si ragione per negazione.

Si considera la data di nascita del primo compagno.

La probabilità che un secondo compagno non sia nato lo stesso giorno è  $P_2 = 364/365$ .

La probabilità che un terzo compagno non sia nato lo stesso giorno dei primi due è  $P_3 = 363/365$ .

.....

La probabilità che due compagni NON siano nati lo stesso giorno sarà:  $P = P_1 \times P_2 \times P_3 \times \dots$

La probabilità che due compagni siano nati lo stesso giorno sarà,  $P_{\text{true}} = 1 - P$

Per un numero di studenti pari a 25  $P_{\text{true}}$  è superiore al 50%....



# Dipendenza



- Cosa succede se X e Y non sono indipendenti, ma **Y dipende da X** e X è indipendente?  $P(X) \Rightarrow P(Y)$ ?
- Cosa succede se esiste una relazione funzionale tra X e Y, dove X è indipendente, mentre Y dipende da quello che fa la variabile X?
  - “ $Y = f(X)$ ” nel dominio dell’analisi funzionale - matematica.
  - “ $P(Y|X)$ ” nel dominio statistico (probabilità condizionata).



# Probabilità congiunta e condizionata



## Probabilità congiunta:

Qual’è la probabilità che la proposizione: “E’ uscito 11” tirando due dadi si avveri?

$$P(N=11) = P(\text{dado}_1 = 5 \text{ AND } \text{dado}_2 = 6) \text{ OR } P(\text{dado}_1 = 6 \text{ AND } \text{dado}_2 = 5) = 1/6 * 1/6 + 1/6 * 1/6 = 2/36$$

$$P(\text{Dado}_1 = 5, \text{Dado}_2 = 6) = 1/36$$

$$P(\text{Dado}_1 = 6, \text{Dado}_2 = 5) = 1/36$$

## Probabilità condizionata:

Qual è la probabilità che la proposizione «E’ uscito 11» con due dadi si avveri, quando ho già tirato un dado che abbia mostrato 5.

$$P(N=11 | \text{Dado}_1 = 5) = 1/6 > 2/36 = P(N=11) \text{ lanciando 2 dadi.} \quad \text{Abbiamo un’incertezza minore.}$$

*Un agente cerca di raccogliere più informazioni possibili per diradare l’incertezza e formulare quindi una soluzione più certa. I problemi che hanno una natura statistica e che prevedono una risposta in funzione dei dati in ingressi sono descrivibili con probabilità condizionate.*

La probabilità condizionata stabilisce una precedenza, una corrispondenza, una dipendenza funzionale tra X e Y. Dato X, determino Y. Lo stesso meccanismo che definisce una relazione funzionale nell’analisi matematica.



## Relazione tra probabilità condizionata e congiunta



*Nel caso dei dadi, quando c'è dipendenza:  $P(N = 11 | Dado_1 = 5) = 1/6$*

$$P(a \text{ AND } b) = P(a | b) * P(b)$$

$P(a \text{ AND } b)$  è probabilità congiunta

$P(a | b)$  è probabilità condizionata

$$P(N=11 \text{ AND } Dado_1 = 5) = P(N = 11 | Dado_1 = 5) * P(Dado_1 = 5) = (1/6) * (1/6) = 1/36$$

*congiunta* *condizionata* *semplice*

Consideriamo prima b e poi a.

$b = Dado_1 = 5$ , restringe le possibili configurazioni. Ne scarta 5/6.

*Possiamo anche scrivere:*

$$P(N = 11 | Dado_1 = 5) = P(N=11 \text{ AND } Dado_1 = 5) / P(Dado_1 = 5) = (1/36) / (1/6) = 1/6$$

*Potremmo anche scrivere tenendo conto che N e Dado<sub>1</sub> non sono indipendenti:*

$$P(N=11 \text{ AND } Dado_1 = 5) = P(X=11) + P(Y=5) - P(X=11 \text{ OR } Y=5) = 2/36 + 1/6 - 2/36 = 1/6$$

*perdendo la relazione tra causa ed effetto*



## Probabilità condizionata e semplice



Consideriamo un mazzo di 40 carte:

- vogliamo valutare quale sia la probabilità che una carta estratta a caso sia un re (probabilità semplice)
- vogliamo valutare quale sia la probabilità che una carta estratta a caso sia un re, sapendo di avere estratto una figura (probabilità condizionata)

$P(Y)$  = probabilità che sia un re

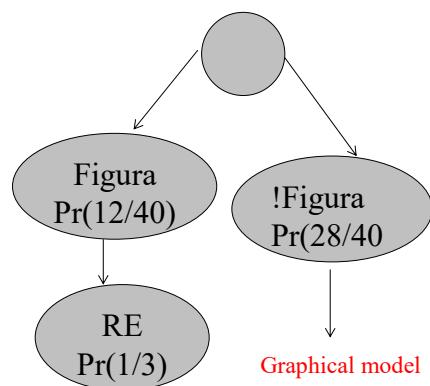
$P(X)$  = probabilità che sia una figura

Le figure sono 12 in un mazzo di 40 carte

$$P(Y | X) = P(\text{re} | \text{figura}) = \\ 1/12 + 1/12 + 1/12 + 1/12 = 1/3$$

$$P(\text{re}) = P(Y) = 1/10$$

$$P(Y) = P(Y/X) P(X) = 1/3 \cdot 12/40 = 4/40$$





## Probabilità composta di variabili dipendenti



Probabilità di avere un re di cuori o un re di quadri

$$P((Y = \text{re\_cuori}) \text{ OR } (Y = \text{re\_quadri})) = P(Y = \text{re\_quadri}) + P(Y = \text{re\_cuori}) = 2/40$$

Probabilità di avere un re di cuori e una figura (non sono indipendenti: il re è una figura!)

$$P((Y = \text{re\_cuori}) \text{ AND } (Y = \text{figura})) \neq P(Y = \text{re\_cuori}) P(Y = \text{figura}) = 1/40 \times 12/40 = 12/1600 - \text{sbagliato}$$

$$P((Y = \text{re\_cuori}) \text{ AND } (Y = \text{figura})) = P(Y = \text{re\_cuori}) + P(\text{figura})$$

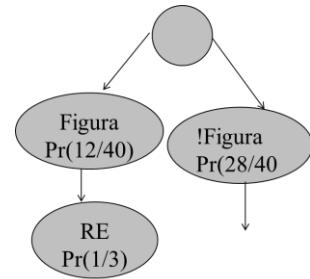
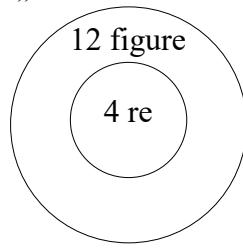
$$- P(Y = \text{re\_cuori}) \text{ OR } (Y = \text{figura})) = 1/40 + 12/40 - 12/40 = 1/40$$

Probabilità di avere un re di cuori o una figura (non sono indipendenti: il re è una figura!)

$$P((Y = \text{re\_cuori}) \text{ OR } (Y = \text{figura})) \neq P(Y = \text{re\_cuori}) + P(Y = \text{figura}) = 1/40 + 12/40 = 13/40 - \text{sbagliato}$$

$$P((Y = \text{re\_cuori}) \text{ OR } (Y = \text{figura})) = P(Y = \text{re\_cuori}) + P(\text{figura})$$

$$- P(Y = \text{re\_cuori}) \text{ AND } (Y = \text{figura})) = 1/40 + 12/40 - 1/40 = 12/40$$



A.A. 2025-2026

<http://borghese.di.unimi.it/>

21



## Inferenza statistica



- Calcolo della probabilità di un evento, a partire dall'informazione sugli elementi che costituiscono l'evento, collezionata sperimentalmente (visione frequentista).
- Consideriamo tre variabili binarie: Mal di denti, Carie, Cavità in dente, e le probabilità congiunte (stimate dal dentista in base alla sua esperienza passata):



	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008
!carie	0,016	0,064	0,144	0,576

$$\sum P(a_i, b_j, c_k) = 1$$

La nostra "funzione" misura il mal di denti e se c'è una cavità (effetto) in dipendenza o meno della presenza di carie (la causa)

A.A. 2025-2026

22/56

<http://borghese.di.unimi.it/>

22



## Marginalizzazione



	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008
!carie	0,016	0,064	0,144	0,576

$$P(\text{carie}) = 0,108 + 0,012 + 0,072 + 0,008 = 0,2$$

$$P(Y) = \sum_{z \in Z} P(Y, z)$$

Marginalizzazione rispetto a “carie” = Y (summing out): tutte le variabili diverse da “carie”, collassano nella sommatoria.



	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008
!carie	0,016	0,064	0,144	0,576

$$P(\text{mal di denti}) = 0,108 + 0,012 + 0,016 + 0,064 = 0,2$$

<http://borghese.di.unimi.it/>

23



## Esempi di ragionamento statistico



	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008
!carie	0,016	0,064	0,144	0,576

$$P(\text{carie OR mal di denti}) = P(\text{carie}) + P(\text{mal di denti}) - P(\text{carie AND mal di denti}) = 0,108 + 0,012 + 0,072 + 0,008 + 0,108 + 0,012 + 0,016 + 0,064 - (0,108 + 0,012) = 0,2 + 0,2 - 0,12 = 0,28$$

	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008
!carie	0,016	0,064	0,144	0,576

$$P(\text{carie AND mal di denti}) = P(\text{carie}) + P(\text{mal di denti}) - P(\text{carie OR mal di denti}) = 0,108 + 0,012 + 0,072 + 0,008 + 0,108 + 0,012 + 0,016 + 0,064 - (0,28) = 0,2 + 0,2 - 0,28 = 0,12$$

A.A. 2025-2026

24/56

<http://borghese.di.unimi.it/>

24



## Condizionamento statistico



	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008
!carie	0,016	0,064	0,144	0,576

Probabilità semplici (assolute):

$$P(\text{mal di denti}) = 0,108 + 0,012 + 0,016 + 0,064 = 0,2$$

$$P(\text{!mal di denti}) = 0,8$$

Probabilità condizionate:  $P(a | b) = P(a \text{ AND } b) / P(b)$

	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
P(carie   mal di denti)	0,54	0,06	0,09	0,01
P(carie   !mal di denti)	0,08	0,32	0,18	0,72

A.A. 2025-2026

25/56

<http://borghese.di.unimi.it/>

Probabilità condizionate al mal di denti:

$$P(\text{caries} | \text{mal di denti}) = \frac{P(\text{caries AND mal di denti})}{P(\text{mal di denti})} \text{ - Divido per 0.2}$$

$$P(\text{caries} | \text{!mal di denti}) \text{ - Divido per 0.8}$$



## Inferenza statistica: probabilità condizionata



	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,54	0,06	0,09	0,01
!carie	0,08	0,32	0,18	0,72

	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008
!carie	0,016	0,064	0,144	0,576

Probabilità condizionata  
 $P(\text{caries} | \text{mal di denti})$

Probabilità congiunta  
 $P(\text{caries AND mal di denti})$

$$P(\text{mal di denti}) = 0,108 + 0,012 + 0,016 + 0,064 = 0,2$$

$$P(\text{!mal di denti}) = 0,8$$

$$\begin{aligned} P(\text{caries}) &= P(Y) = \sum_{z \in Z} P(Y | z)P(z) \\ &= P(\text{Caries} | \text{mal di denti})P(\text{mal di denti}) + P(\text{Caries} | \text{!mal di denti})P(\text{!mal di denti}) \\ &= (0,54 + 0,06) * 0,2 + (0,09 + 0,01) * 0,8 = 0,2 \end{aligned}$$

A.A. 2025-2026

26/56

<http://borghese.di.unimi.it/>

26



## Overview



Probabilità semplice e condizionata

Teorema di Bayes



## Teorema di Bayes (1701-1761)



$$P(X, Y) = P(Y|X)P(X) = P(X|Y)P(Y)$$

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

X = causa      Y = effetto



Thomas Bayes

$$P(\text{causa}|\text{effetto}) = \frac{P(\text{Effetto}|\text{Causa})P(\text{Causa})}{P(\text{Effetto})}$$

We usually do not know the statistics of the cause, but we can measure the effect and, through frequency, build the statistics of the effect or we know it in advance.

A doctor knows  $P(\text{Symptoms}|\text{Causa})$  and wants to determine  $P(\text{Causa}|\text{Symptoms})$



## Esempio A - I



Software in Scilab available!

In una città lavorano due compagnie di taxi:  
blu e verde:  $X = \{T_{\text{blu}}, T_{\text{verde}}\}$



con una Distribuzione di 85% di taxi verdi e 15% di taxi blu.  
 $P(\text{Taxi} = \text{verde}) = 0.85$   
 $P(\text{Taxi} = \text{blu}) = 0.15$

Succede un incidente in cui è coinvolto un taxi.

Un testimone dichiara che il taxi era blu  $P(Y=\text{blu})$ . Era sera e l'affidabilità del testimone è stata valutata dell'80% (è una probabilità condizionata):  
 $P(Y=\text{blu} | X = \text{blu}) = P(Y=\text{verde} | X = \text{verde}) = 0.8$

Qual è la probabilità che il taxi fosse effettivamente blu?

**Non è l'80%!**

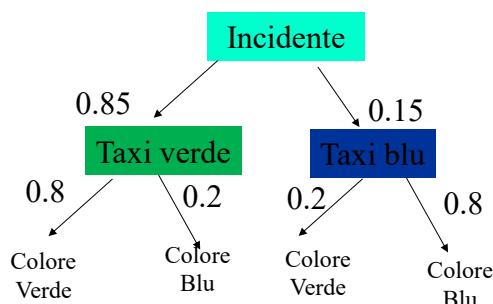


## Esempio A - II



$X = \{\text{Taxi} = \text{blu}, \text{Taxi} = \text{verde}\}$   
“Causa”

$Y = \{\text{Colore} = \text{blu}, \text{Colore} = \text{verde}\}$   
“Effetto”



- $P(\text{Taxi} = \text{blu}) = \text{Probabilità a-priori} = 0.15$
- $P(\text{Colore} = \text{blu} | \text{Taxi} = \text{blu}) = P(\text{Colore} = \text{verde} | \text{Taxi} = \text{verde}) - \text{Affidabilità del testimone} - \text{Probabilità condizionata} = 0.8$

Come combino queste informazioni per ottenere una stima sulla probabilità che il taxi dell'incidente sia effettivamente blu?



## Esempio A - III

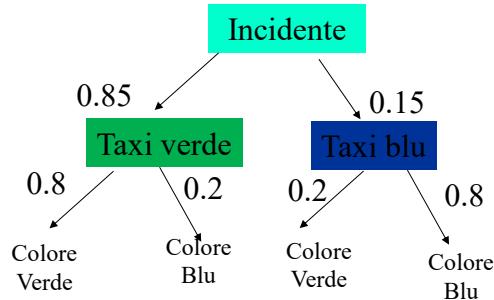


$X = \{\text{Taxi} = \text{blu}, \text{Taxi} = \text{verde}\}$   
"Causa"

$Y = \{\text{Colore} = \text{blu}, \text{Colore} = \text{verde}\}$   
"Effetto"

Invertendo la relazione tra causa ed effetto  
applicando Bayes:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$



$$P(\text{Taxi} = \text{blu} | \text{Colore} = \text{blu}) = \\ P(\text{Colore} = \text{blu} | \text{Taxi} = \text{blu})P(\text{Taxi} = \text{blu}) / P(\text{Colore} = \text{blu})$$

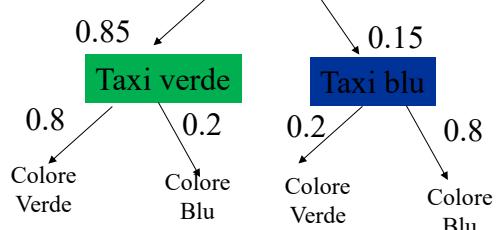


## Esempio A- IV



$X = \{\text{Taxi} = \text{blu}, \text{Taxi} = \text{verde}\}$   
"Causa"

$Y = \{\text{Colore} = \text{blu}, \text{Colore} = \text{verde}\}$   
"Effetto"



$$P(\text{Colore} = \text{blu}) = \{\text{Probabilità marginale di } Y \text{ (probabilità semplice)}\} = \\ P(\text{Colore} = \text{blu} | \text{Taxi} = \text{blu})P(\text{Taxi} = \text{blu}) + \\ P(\text{Colore} = \text{blu} | \text{Taxi} = \text{verde})P(\text{Taxi} = \text{verde}) = 0.8*0.15 + 0.2*0.85 = 0.29$$

$$P(\text{Colore} = \text{blu} | \text{Taxi} = \text{blu}) = 0.8 \quad \{\text{Affidabilità del testimone}\} \\ P(\text{Colore} = \text{verde} | \text{Taxi} = \text{verde}) = 0.8 \quad \{\text{Affidabilità del testimone}\}$$

$$P(\text{Taxi} = \text{blu}) = 0.15 \quad \{\text{Probabilità a-priori che si incontri un taxi blu}\} \\ P(\text{Taxi} = \text{verde}) = 0.85 \quad \{\text{Probabilità a-priori che si incontri un taxi verde}\}$$

$$P(\text{Taxi} = \text{blu} | \text{Colore} = \text{blu}) = \{\text{Teorema di Bayes}\} \\ P(\text{Colore} = \text{blu} | \text{Taxi} = \text{blu})P(\text{Taxi} = \text{blu}) / P(\text{Colore} = \text{blu}) = 0.8*0.15 / 0.29 = 0.41 \quad 0.15 < 0.41 << 0.8!! \\ \text{Pesano anche gli "errori" commessi quando il testimone vede un blu ma era verde!}$$



## Esempio A - V



$$X = \{\text{Taxi} = \text{blu}, \text{Taxi} = \text{verde}\}$$

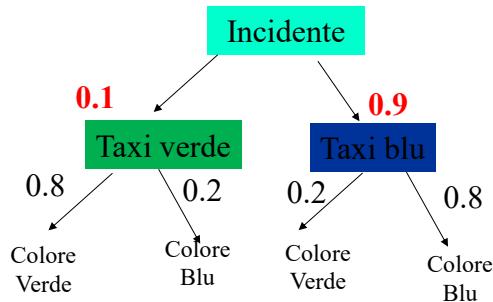
“Causa” – Molti più taxi blu.

$$P(\text{Taxi} = \text{verde}) = 0.1$$

$$P(\text{Taxi} = \text{blu}) = 0.9$$

$$Y = \{\text{Colore} = \text{blu}, \text{Colore} = \text{verde}\}$$

“Effetto”



$$P(\text{Colore} = \text{blu} | \text{Taxi} = \text{blu}) = P(\text{Colore} = \text{blu} | \text{Taxi} = \text{blu})P(\text{Taxi} = \text{blu}) / P(\text{Colore} = \text{blu})$$

$$P(\text{Colore} = \text{blu}) = \{\text{Probabilità marginale di } Y \text{ (probabilità semplice)}\} =$$

$$P(\text{Colore} = \text{blu} | \text{Taxi} = \text{blu})P(\text{Taxi} = \text{blu}) + P(\text{Colore} = \text{blu} | \text{Taxi} = \text{verde})P(\text{Taxi} = \text{verde}) = \\ 0.8 * 0.9 + 0.2 * 0.1 = 0.74$$

$$P(\text{Taxi} = \text{blu} | \text{Colore} = \text{blu}) = P(\text{Colore} = \text{blu} | \text{Taxi} = \text{blu})P(\text{Taxi} = \text{blu}) / P(\text{Colore} = \text{blu}) = \\ 0.8 * 0.9 / 0.74 = 0.97$$

Testimonianza molto affidabile in questo caso!



## Esempio B - I



Lo strumento principe per lo screening per il tumore al seno è la radiografia (mammografia).



Definiamo  $X$  la situazione della donna:  $X = \{\text{sana}, \text{malata}\}$

Definiamo  $Y$  l'esito della mammografia:  $Y = \{\text{positiva}, \text{negativa}\}$

La percentuale di donne malate sulla popolazione è dell'1%.  
 $P(X) = 0.01$  – probabilità a-priori.

La sensitività della mammografia è intorno al 90%:

$$\text{sensitività} = \frac{n_{\text{positive}}}{N_{\text{ill}}} \Rightarrow P(Y = \text{positive} | X = \text{ill (positive)})$$

La specificità della mammografia è anch'essa intorno al 90%:

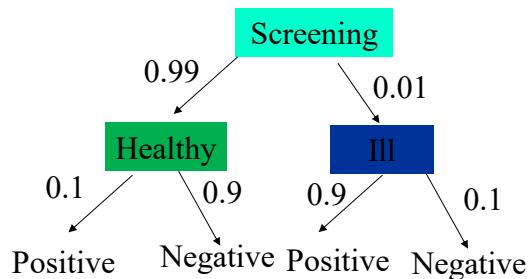
$$\text{specificità} = \frac{n_{\text{negative}}}{N_{\text{healthy}}} \Rightarrow P(Y = \text{negative} | X = \text{healthy (negative)})$$



## Esempio B - II - Probabilità marginali



$X = \{\text{Healthy, Ill}\}$   
 $Y = \{\text{Positive, Negative}\}$



$$P(Y=\text{Positive} | X=\text{Ill}) * P(X=\text{Ill}) = 0.9 * 0.01 = 0.009$$

$$P(Y=\text{Positive} | X=\text{Healthy}) * P(X=\text{Healthy}) = 0.1 * 0.99 = 0.099$$

$$P(Y=\text{Positive}) = P(Y=\text{Positive} | X=\text{Ill}) * P(X=\text{Ill}) + P(Y=\text{Positive} | X=\text{Healthy}) * P(X=\text{Healthy}) = 0.009 + 0.099 = 0.108 \text{ (probabilità marginale di } Y=\text{Positive})$$

10.8% di probabilità di avere un esame positivo a fronte di uno 0.01% di donne malate!  
 Solo lo 0,9% proviene da donne effettivamente malate, le altre sono false positive!



## Esempio B - III - Donne malate



10.8% di probabilità di avere un esame positivo a fronte di uno 0.01% di donne malate!  
 Solo lo 0,9% circa proviene da donne effettivamente malate, le altre sono false positive!

Qual'è la probabilità che una donna sia veramente malata se il test risulta positivo?

Applichiamo Bayes  $P(X=\text{Ill} | Y=\text{Positive})$  - PPV (Positive Predictive Value)

$$P(X=\text{Ill} | Y=\text{Positive}) = P(Y=\text{Positive} | X=\text{Ill})P(X=\text{Ill}) / P(Y=\text{Positive}) = 0.09 / 0.108 = 0.083 \text{ (8.3%)}$$

**Solo 8.3% delle donne con mammografia positiva sono effettivamente ammalate.**

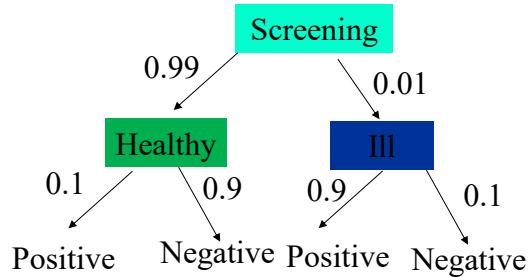
Analizzando la formula del teorema di Bayes, dove ha senso investire per ottenere un rendimento delle screening maggiore?



## Esempio B - IV - Screening negativo



$X = \{\text{Healthy, Ill}\}$   
 $Y = \{\text{Positive, Negative}\}$



$$P(Y=\text{Negative} | X=\text{Ill}) * P(X=\text{Ill}) = 0.1 * 0.01 = 0.001$$

$$P(Y=\text{Negative} | X=\text{Healthy}) * P(X=\text{Healthy}) = 0.9 * 0.99 = 0.891$$

$$P(Y=\text{Negative}) = P(Y=\text{Negative} | X=\text{Ill}) * P(X=\text{Ill}) + P(Y=\text{Negative} | X=\text{Healthy}) * P(X=\text{Healthy}) = 0.001 + 0.891 = 0.902 \text{ (probabilità marginale di } Y=\text{Negative}) - \text{ più di 90\% esami negativi}$$

Di questi 90% di esami negativi, il 10% circa sono relative a donne in realtà malate.



## Esempio B - V



$$P(Y=\text{Positive}) = P(Y=\text{Positive} | X=\text{Ill}) * P(X=\text{Ill}) + P(Y=\text{Positive} | X=\text{Healthy}) * P(X=\text{Healthy}) = 0.001 + 0.0891 = 0.902 \text{ (probabilità marginale di } Y=\text{Positive}) - \text{ più di 90\% esami positivi}$$

Qual'è la probabilità che una donna sia in realtà malata se il test risulta negativo?

Applichiamo Bayes  $P(X=\text{Ill} | Y=\text{Negative})$

$$P(X=\text{Ill} | Y=\text{Negative}) = P(Y=\text{Negative} | X=\text{Ill})P(X=\text{Ill}) / P(Y=\text{Negative}) = 0.1 * 0.01 / 0.891 = 0.00112 (0,11\%)$$

**Lo 0,11% delle donne malate ricevono una mammografia negativa.**

Analizzando la formula del teorema di Bayes, dove ha senso investire per ottenere un rendimento delle screening maggiore?



## Valutazione delle prestazioni dei test binari



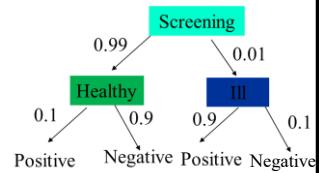
	Malati (positivi)	Sani (negativi)
Classifico Positivi	Veri +	Falsi +
Classifico Negativi	Falsi -	Veri -

**Sensitività:** 
$$\frac{Veri+}{N_{pos}} = \frac{Veri+}{(Veri+) + (Falsi-)} = \frac{Veri+}{Num\ malati}$$

**Specificità:** 
$$\frac{Veri-}{N_{neg}} = \frac{Veri-}{(Veri-) + (Falsi+)} = \frac{Veri-}{Num\ sani}$$

**Positive predictive value:** 
$$\frac{Veri+}{N_{class\ pos}} = \frac{Veri+}{(Veri+) + (Falsi+)}$$

**Negative predictive value:** 
$$\frac{Veri-}{N_{class\ neg}} = \frac{Veri-}{(Veri-) + (Falsi-)}$$



Dove conviene investire?

A.A. 2025-2026

39/56

<http://borghese.di.unimi.it/>

39

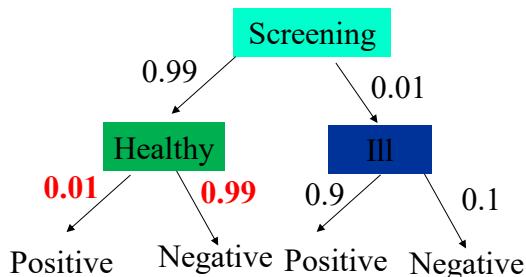


## Investiamo sulla specificità del test



$X = \{\text{Healthy, Ill}\}$   
 $Y = \{\text{Positive, Negative}\}$

$P(Y=\text{Negative} | X=\text{Healthy}) = 0.99$



$P(Y=\text{Positive} | X=\text{Ill}) = 0.9 * 0.01 = 0.009$

$P(Y=\text{Positive}) = P(Y=\text{Positive} | X=\text{Ill}) + P(Y=\text{Positive} | X=\text{Healthy}) = 0.009 + 0.99 * 0.01 = 0.0189$

$P(X=\text{Ill} | Y=\text{Positive}) = P(Y=\text{Positive} | X=\text{Ill})P(X=\text{Ill}) / P(Y=\text{Positive}) = 0.009 / 0.0189 = 0.476 = 47,6\% >> 8.3\%$

A.A. 2025-2026

40/56

<http://borghese.di.unimi.it/>

40

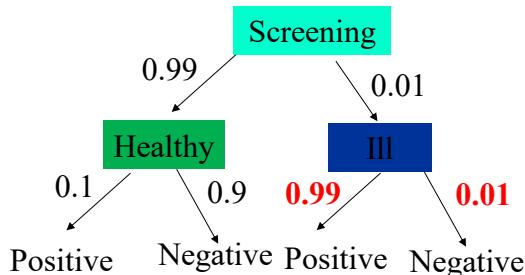


## Investiamo sulla Sensitività del test



$X = \{\text{Healthy, Ill}\}$   
 $Y = \{\text{Positive, Negative}\}$

$$P(Y=\text{Positive} | X=\text{Ill}) = 0.99$$



$$P(Y=\text{Positive} | X=\text{Ill}) = 0.99 * 0.01 = 0.0099$$

$$P(Y=\text{Positive}) = P(Y=\text{Positive} | X=\text{Ill}) + P(Y=\text{Positive} | X=\text{Healthy}) = 0.0099 + 0.99 * 0.1 = 0.1098$$

$$P(X=\text{Ill} | Y=\text{Positive}) = P(Y=\text{Positive} | X=\text{Ill})P(X=\text{Ill}) / P(Y=\text{Positive}) = 0.0099 / 0.1098 = 0.09 = 9\% > 8.3\%.$$

A.A. 2025-2026

41/56

<http://borghese.di.unimi.it/>

41

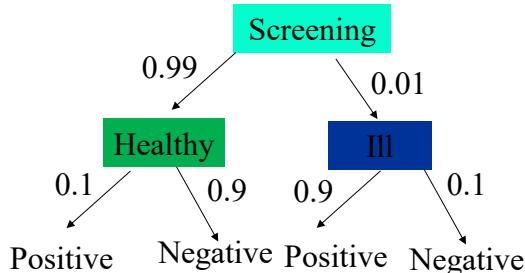


## Esempio B - VI



$X = \{\text{Healthy, Ill}\}$   
 $Y = \{\text{Positive, Negative}\}$

Falsi negativi?  
 $P(X = \text{Ill} | Y = \text{Negative})?$



$$P(Y=\text{Negative} | X=\text{Ill}) = 0.1 * 0.01 = 0.001$$

$$P(Y=\text{Negative}) = P(Y=\text{Negative} | X=\text{Ill}) + P(Y=\text{Positive} | X=\text{Healthy}) = 0.001 + 0.99 * 0.9 = 0.891$$

$$P(X=\text{Ill} | Y=\text{Negative}) = P(Y=\text{Negative} | X=\text{Ill})P(X=\text{Ill}) / P(Y=\text{Negative}) = 0.001 / 0.891 = 0.11\% \\ \text{Una donna ogni mille non viene diagnosticata!}$$

A.A. 2025-2026

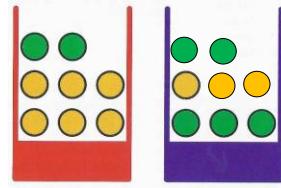
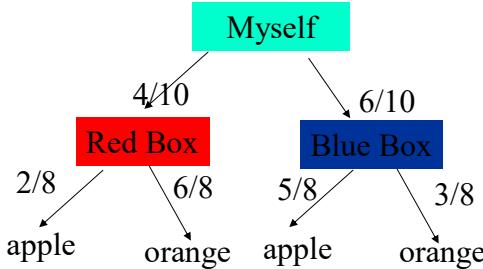
42/56

<http://borghese.di.unimi.it/>

42



## Esempio C - I



1) Supponiamo di conoscere  $P(X)$ , probabilità di scelta del box,

$$- P(\text{blu}) = 0.6$$

$$- P(\text{rosso}) = 0.4$$

e la  $P(Y|X)$ , probabilità di avere una mela (arancia) se scegliamo un certo box, possiamo determinare la probabilità assoluta (semplice) di scegliere un certo frutto,  $P(Y)$ ?

2) Supponiamo di non conoscere  $P(X)$ , probabilità di scelta del box, conosciamo la probabilità  $P(Y|X)$  e  $P(Y)$ .

Possiamo determinare  $P(X)$ ?



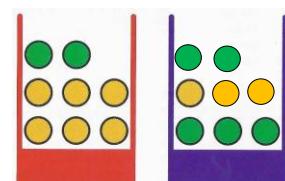
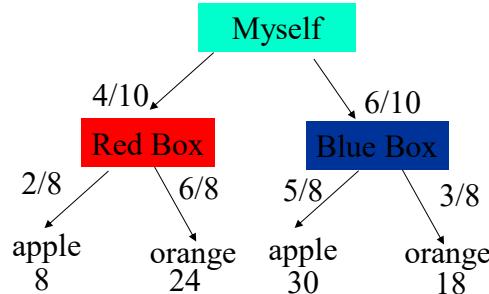
## 1) Determino $P(Y = \text{Apple})$



$$P(Y=\text{apple} | X = \text{blue}) = 5/8 - \text{nota}$$

$$P(Y=\text{apple} | X = \text{red}) = 2/8 - \text{nota}$$

$$P(Y=\text{apple}) \neq (2+5) / 8 = 7/8$$



$$P(Y=\text{apple} | X = \text{blue}) + P(Y=\text{orange} | X = \text{blue}) = 1$$

$$P(Y=\text{apple} | X = \text{red}) + P(Y=\text{orange} | X = \text{red}) = 1$$

$$P(Y=\text{apple}) = P(Y=\text{apple} | X = \text{blue}) P(X=\text{blue}) + \\ P(Y=\text{apple} | X = \text{red}) P(X=\text{red}) = 5/8 * 6/10 + 2/8 * 4/10 = 38/80 \neq 7/8$$



# 1) Determino $P(Y = \text{orange})$



$$P(Y=\text{apple} | X = \text{blue}) = 5/8 - \text{nota}$$

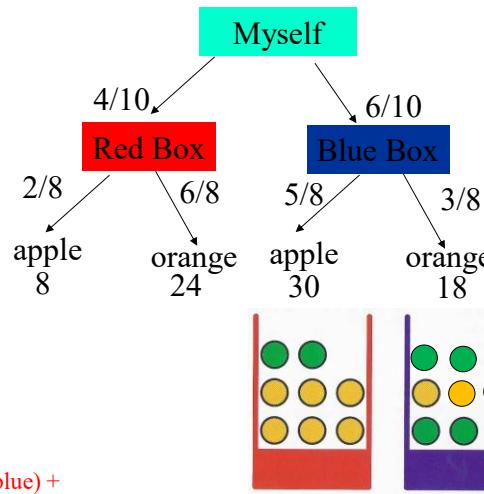
$$P(Y=\text{orange} | X = \text{blue}) = 3/8 - \text{nota}$$

$$P(Y=\text{apple} | X = \text{red}) = 2/8 - \text{nota}$$

$$P(Y=\text{orange} | X = \text{red}) = 6/8 - \text{nota}$$

$$P(Y=\text{apple}) \neq (2+5)/8 = 7/8$$

$$P(Y=\text{orange}) \neq (6+3)/8 = 9/8$$



$$P(Y=\text{orange}) = P(Y=\text{orange} | X = \text{blue}) P(X=\text{blue}) + P(Y=\text{orange} | X = \text{red}) P(X=\text{red}) = 3/8 * 6/10 + 6/8 * 4/10 = 42/80 \neq 9/8$$

$$P(Y=\text{apple}) = 38/80$$

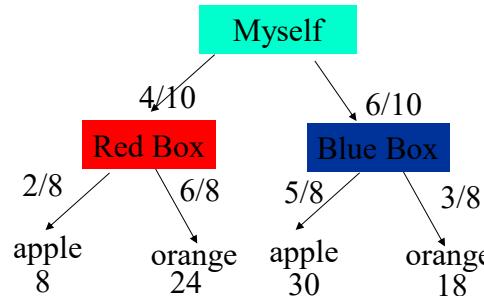
$$P(Y=\text{apple}) + P(Y=\text{orange}) = 1$$



# Determino $P(X=\text{red} | Y)$



$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$



$$P(X=\text{red} | Y=\text{orange}) = P(Y=\text{orange}|X=\text{red}) P(X=\text{red}) / P(Y=\text{orange}) = (6/8 * 4/10) / (21/40) = 24/42 > 4/10$$

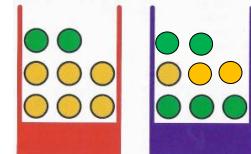
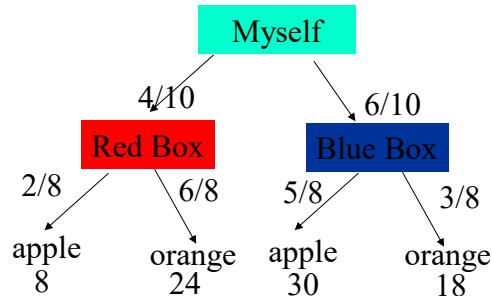
$$P(X=\text{red} | Y=\text{apple}) = P(Y=\text{apple} | X=\text{red}) P(X=\text{red}) / P(Y=\text{apple}) = (2/8 * 4/10) / (19/40) = 8/38 << 4/10$$



## Determino $P(X | Y)$



$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$



$$P(X=\text{red} | Y=\text{orange}) = 24/42 > 4/10$$

$$P(X=\text{blue} | Y=\text{orange}) = P(Y=\text{orange} | X=\text{blue}) P(X=\text{blue}) / P(Y=\text{orange}) = (3/8 * 6/10) / (21/40) = 18/42 < 6/10$$

$$P(X=\text{red} | Y=\text{apple}) = 8/38 << 4/10$$

$$P(X=\text{blue} | Y=\text{apple}) = P(Y=\text{apple} | X=\text{blue}) P(X=\text{blue}) / P(Y=\text{apple}) = (5/8 * 6/10) / (19/40) = 30/38 >> 6/10$$



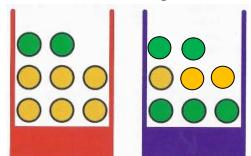
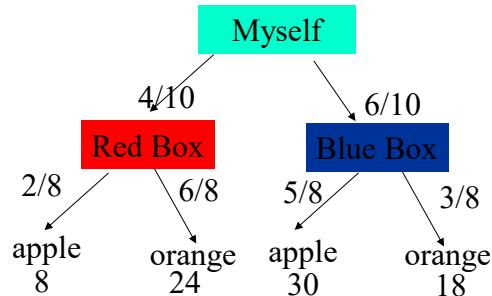
## Interpretazione



$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

$$P(X=\text{red} | Y=\text{apple}) = 8/38 << 4/10$$

$$P(X=\text{blue} | Y=\text{apple}) = 30/38 >> 6/10$$



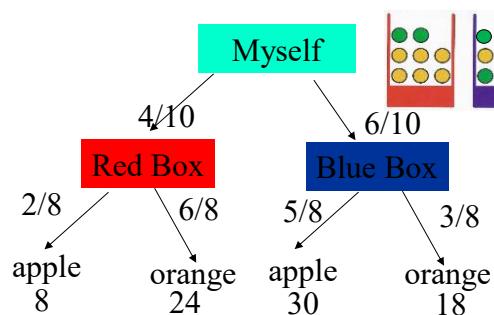
Correggo la probabilità a-priori,  $P(X)$  con le informazioni raccolte,  $P(Y)$ , e ottengo una nuova valutazione della probabilità di  $X$  (che dipende da  $Y$ ),  $P(X | Y)$  detta probabilità a-posteriori.

	Red	Blue	
Apple	$4/10 * 2/8 = 8/80 - 8$	$6/10 * 5/8 = 30/80 - 30$	$y_j$
Orange	$4/10 * 6/8 = 24/80 - 24$	$6/10 * 3/8 = 18/80 - 18$	$x_i$



$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

## Importanza



Questo è un tipico esempio di problema **inverso**.

Raccogliamo delle misure Y e vogliamo determinare da quale sistema (modello probabilistico) possono essere state generate.

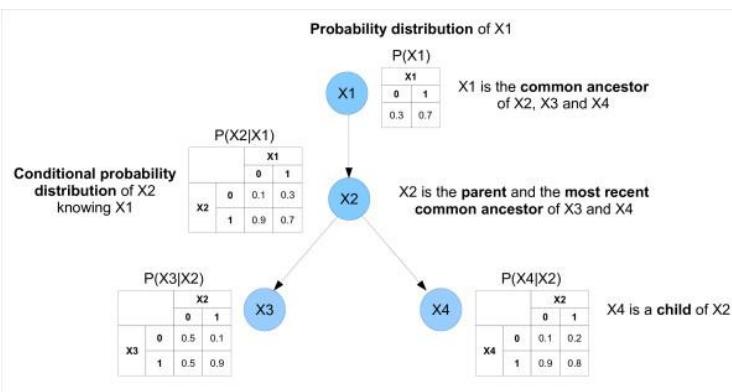
Possiamo inserire delle informazioni statistiche (a-priori) su X, cioè sulla forma del modello (e.g. smoothness)



## Graphical models



A **graphical model** o **modello probabilistico su grafo (PGM)** è un modello probabilistico che evidenzia le dipendenze tra le variabili randomiche (può evolvere eventualmente in un albero). Viene utilizzato nell'inferenza statistica.





## Estensione a più variabili



Sostituisco un'espressione logica a una variabile:

$P(X|Y_1, Y_2)$  if ( $P(Y_1) = y_1$  and  $P(Y_2) = y_2$ ) then  $P(X) = X$

$Z = Y_1 \text{ and } Y_2$



		mal di denti	!mal di denti		
		cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008	
	0,016	0,064	0,144	0,576	

Dalla tabella delle probabilità congiunte ricaviamo ( $Z = \text{mal di denti AND cavità}$ ):

$$P(\text{carie}; Z) = P(\text{carie AND} (\text{mal di denti AND cavità})) = 0,108$$

$$P(\text{carie} | Z) = P(\text{carie} | (\text{mal di denti AND cavità})) =$$

$$P(\text{carie AND} (\text{mal di denti AND cavità})) / P(\text{mal di denti AND cavità}) = \\ 0,108 / (0,016 + 0,108) = 0,108 / 0,124 = 0,871 - \text{complessa}$$

A.A. **Carie, mal di denti e cavità non sono indipendenti, come trattiamo l'AND?**

<http://borghese.di.unimi.it/>

51



## Estensione a più variabili



$P(X|Y_1, Y_2)$  if ( $P(Y_1) = y_1$  and  $P(Y_2) = y_2$ ) then  $P(X) = X$

$Z = Y_1 \text{ and } Y_2$



		mal di denti	!mal di denti		
		cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008	
	0,016	0,064	0,144	0,576	

$$P(\text{carie} | Z) = P(\text{carie} | (\text{mal di denti AND cavità})) =$$

$$P(\text{carie AND} (\text{mal di denti AND cavità})) / P(\text{mal di denti AND cavità}) = \\ 0,108 / (0,016 + 0,108) = 0,108 / 0,124 = 0,871$$

$$P(\text{mal di denti AND cavità} | \text{carie}) = P((\text{mal di denti AND cavità AND carie}) / P(\text{carie}) = \\ 0,108 / 0,2 = 0,54$$

A.A. 2025-2026

52/56

<http://borghese.di.unimi.it/>

52



## Estensione a più variabili



$P(X|Y_1, Y_2)$  if ( $P(Y_1) = y_1$  and  $P(Y_2) = y_2$ ) then  $P(X) = x$

$Z = Y_1 \text{ and } Y_2$



		mal di denti	!mal di denti		
		cavità	!cavità	cavità	!cavità
carie	mal di denti	0,108	0,012	0,072	0,008
	!mal di denti	0,016	0,064	0,144	0,576

Carie, mal di denti e cavità non sono indipendenti, come trattiamo l'AND?

$$P(\text{carie AND mal di denti}) = P(\text{carie}) + P(\text{mal di denti}) - P(\text{carie OR mal di denti}) = 0,108 + 0,012 + 0,072 + 0,008 + 0,108 + 0,012 + 0,016 + 0,064 - (0,28) = 0,2 + 0,2 - 0,28 = 0,12$$

Come lo possiamo trattare invece se sono indipendenti?

$$P(X \text{ AND } Y) = P(X)P(Y)$$



## Naïve Bayes



Sostituiamo un'espressione a una variabile logica:

$$P(X | (Y_1 \text{ AND } Y_2)) * P(Y_1 \text{ AND } Y_2) = P(Y_1 \text{ AND } Y_2 | X) * P(X) \quad \text{Bayes}$$

Introduciamo un'altra ipotesi. Cosa succede se  $Y_1$  e  $Y_2$  sono indipendenti? **Dipendono entrambe da X ma non dipendono tra di loro.**

Sono cioè **condizionatamente indipendenti**, cioè porta a:

$$P((Y_1 \text{ AND } Y_2) | X) * P(X) = (P(Y_1 | X) * P(Y_2 | X)) * P(X)$$

In questo caso viene semplificato il calcolo dell'AND, che viene calcolato come prodotto delle probabilità.

**Modello Naive Bayes** **Gli effetti sono supposti indipendenti tra loro e dipendono da una stessa causa**

$$\text{In generale: } P(\text{Causa} | \text{Effetto}_1 \text{ and } \text{Effetto}_2 \text{ and } \dots \text{ Effetto}_N) = \prod_{i=1}^N P(\text{Effetto}_i | \text{Causa})$$



## Riepilogo



$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} = \frac{P(X,Y)}{P(Y)}$$

Teorema di Bayes

Lega probabilità condizionate, congiunte, semplici (marginali)

Consente di inferire la probabilità di un evento causa, X, a partire dalla probabilità associata alla frequenza di una certa misura, effetto, P(Y), dalla frequenza relativa dell'evento associato alla misura, P(Y), e dalla probabilità nota a-priori, P(X), della causa.

La probabilità P(X|Y) viene per questo detta probabilità a-posteriori ed è una probabilità condizionata.

Viene utilizzata nei problemi inversi.



## Overview



Probabilità semplice e condizionata

Teorema di Bayes