

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



Overview



Probabilità semplice e condizionata

Esempio



Incertezza



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

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. 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.

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

La probabilità che si verifichi uno tra tutti i casi possibili è sempre 1.

Oververosia la somma delle probabilità di tutti gli eventi (se mutuamente esclusivi) somma 1.

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

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



Altri aspetti della probabilità

Problema della visione frequentista: **omogeneità del campione** (classe di riferimento).
Come posso effettuare la media di eventi in modo “sicuro”?

- **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.



Esempio di incertezza

Mal di denti => Carie Quando possiamo essere sicuri che questa proposizione (evento) sia vera?
(notice that proposition is used in logic, it is an event in AI terminology).

In realtà non è sempre vera: il mal di denti può avere diverse **cause**

Mal di denti => Carie OR Problemi gengive OR ascessi OR

Vale il viceversa? Carie => Mal di denti

Quali sono i problemi con l'approccio puramente logico?

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 (paziente).

Possiamo ottenere un **grado di credenza (belief)** nell'affermazione. Questa potrà rivelarsi vera o falsa con una certa probabilità.

La probabilità è basata sulla conoscenza (a-priori) non sull'evento che si è già verificato!!





Formulazione analitica

Gli eventi sono mutuamente esclusivi e costituiscono lo spazio delle probabilità: Ω , dove ciascun evento viene indicato con ω .

Questo porta a definire gli assiomi di base della teoria della probabilità:

Quindi:

1. $0 \leq P(\omega) \leq 1$ per ogni ω

$$\sum_{\omega \in \Omega} P(\omega) = 1$$

2. $P(a) = \text{True} \iff P(!a) = \text{False}$

3. $P(a \text{ OR } b) = P(a) + P(b) - P(a \text{ AND } b) =$

$$\frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$$

Principio di inclusione-esclusione

$P(a \text{ AND } b) = P(a) P(b)$ **se sono indipendenti**

$$\frac{1}{2} \frac{1}{2} = \frac{1}{4}$$

P(a) P(b)



a b a OR b

0 0 0

0 1 1

1 0 1

1 1 1

← a AND b



Teorema di de Finetti (1931)

- Se un Agente 1 esprime un insieme di credenze (probabilità a-priori) inconsistenti (che violano gli assiomi della teoria delle probabilità), allora ci sarà una combinazione di credenze di un agente 2 che **garantisce** che l'agente 1 perderà.
- Il comportamento di un agente è più efficace quando tutta la conoscenza statistica viene sfruttata in modo congruo (supponendo la conoscenza statistica un-biased).

Questo teorema è stato tradotto in inglese solo nel 1993!



Probabilità a-priori e a-posteriori



Qual'è la probabilità che la proposizione: "E' uscito 12" tirando due dadi si avveri?
 $P(N=12) = \{6 \text{ AND } 6\} = \{6,6\} = P(6,6) = 1/6 * 1/6 = 1/36$

Qual'è la probabilità che la proposizione: "E' uscito 11" tirando due dadi si avveri?
 $P(N=11) = \{5,6\} \text{ OR } \{6,5\} = P((5,6) + P(6,5) = 1/6 * 1/6 + 1/6 * 1/6 = 1/18$

Probabilità **incondizionata** o **a-priori**. Non richiede o dipende da altre informazioni.

Supponiamo ora che il primo dado abbia mostrato 5. Abbiamo un'informazione. Perché $N = 11$, occorre che il secondo dato mostri 6.

$P(N=11 | \text{Dado}_1 = 5) = 1/6 > P(N=11)$ Abbiamo un'incertezza minore.

Probabilità **condizionata** o **a-posteriori**.

Un agente cerca di raccogliere più informazioni possibili per diradare l'incertezza e formulare quindi dei problemi descrivibili con probabilità condizionate.



Probabilità condizionata e congiunta



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

Nel caso dei dadi:

$P(N = 11 | \text{Dado}_1 = 5) = 1/6$

Possiamo riscrivere la probabilità condizionata come: $P(a \text{ AND } b) = P(a | b) P(b)$

$P(N = 11 | \text{Dado}_1 = 5) = P(N=11 \text{ AND } \text{Dado}_1 = 5) / P(\text{Dado}_1 = 5) = (1/36) / (1/6) = 1/6$

Ovverosia:

$P(N=11 \text{ AND } \text{Dado}_1 = 5) = P(N = 11 | \text{Dado}_1 = 5) * P(\text{Dado}_1 = 5) = (1/6) * (1/6) = 1/36.$

$b = \text{Dado}_1 = 5$, restringe le possibili configurazioni. Ne scarta 5/6.

Si può vedere la probabilità condizionata come una funzione $a = f(b)$



Inferenza statistica

- Calcolo della probabilità a-posteriori di un evento, a partire dall'informazione collezionata sperimentalmente.
- Consideriamo tre variabili binarie: Mal di denti, Carie, Cavità in dente, e le probabilità congiunte:



	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)



Esempi di inferenza statistica

	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

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

$P(\text{carie AND mal di denti}) \neq P(\text{carie})P(\text{mal di denti})$ - Non sono indipendenti!!

	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.



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 a-priori:

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

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

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

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

Probabilità condizionate
P(carie | mal di denti)



Condizionamento statistico

	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

Probabilità condizionate
P(carie | mal di denti)

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

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

$$P(\text{carie}) = P(Y) = \sum_{z \in Z} P(Y | z)P(z) = P(\text{Carie} | \text{mal di denti})P(\text{mal di denti}) +$$

$$P(\text{Carie} | \text{!mal di denti})P(\text{!mal di denti}) = (0,54+0,06) * 0,2 + (0,09+0,01) * 0,8 = 0,2$$



Come utilizziamo la stima a-posteriori



Vogliamo determinare la probabilità di un evento partendo dalla conoscenza sperimentale di altre.

	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

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

$$= (0.108 + 0.012) / (0.108 + 0.012 + 0.016 + 0.064) = 0.6$$

$$P(!\text{carie} | \text{mal di denti}) = P(!\text{carie AND mal di denti}) / P(\text{mal di denti}) = 0.4$$

P(mal di denti) è la probabilità marginale relativa al mal di denti $P(\text{mal di denti}) = 0,2$. Ha una funzione di normalizzazione. Il rapporto tra P(carie) e P(!carie) non dipende da P(mal di denti).

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



Inferenza statistica nel caso generale



- Consideriamo una funzione con incertezza: $P(X | E, Y)$ dove e sono le variabili osservate e y quelle non osservate.

- La $P(X | e) = P(X, e) = \sum_{y \in Y} (P(X, e, y) / P(y))$

Quando si vuole una valutazione comparativa, il termine $P(y)$ non cambia la funzione $P(X|e)$ e può essere omissso:

$$P(X | e) = P(X, e) = \alpha \sum_{y \in Y} P(X, e, y)$$

Non si riesce a rappresentare graficamente comunque quando il numero di variabili cresce.

Indipendenza statistica

$P(X | Y) = P(X)$ se X non dipende da Y

$P(X, Y) = P(X) P(Y)$ se X ed Y sono indipendenti.



Overview

Probabilità semplice e condizionata

Esempio

A.A. 2017-2018

17/27

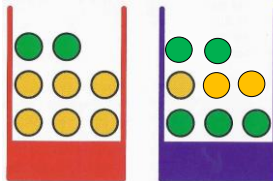
<http://borghese.di.unimi.it/>



Un altro esempio

Box = {r(ed), b(lue)}

Fruit = {a(ple), o(range)} (apple sono mele verdi)



Supponiamo di prendere un frutto alla volta e di sostituirlo con un frutto dello stesso tipo (sampling with replacement)

Supponiamo di selezionare $\text{Box} = r$ 40% delle volte e $\text{Box} = b$ il 60% delle volte.

$$P(\text{Box} = r) = 0,4$$

$$P(\text{Box} = b) = 0,6$$

$$P(\text{Box}) = 1$$

A.A. 2017-2018

18/27

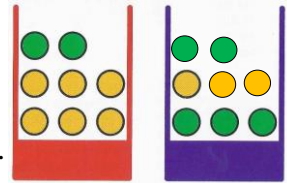
<http://borghese.di.unimi.it/>



Formulazione statistica

	Red	Blue	
Apple		n_{ij}/n_{TOT}	y_j
Orange			
	x_i		

$X = \{x_i\} = \{\text{blue, red box}\}$ box colour
 $Y = \{y_j\} = \{\text{apple, orange}\}$ fruit type



N trials – z_{ij} is the sample observed in each trial (x_{ij}, y_{ij}).

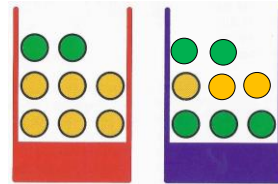
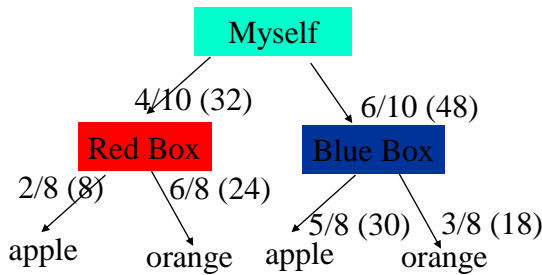
n_{ij} is the number of trials in which x_i AND y_j were observed.

PROBABILITA' CONGIUNTA

$$\sum_i \sum_j \frac{n_{i,j}}{N_{TOT}} = 1$$



Esempio (N=80)



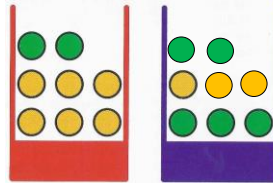
	Red	Blue	
Apple	8/80	30/80	y_j
Orange	24/80	18/80	
	x_i		

$$\sum_{i,j} \frac{n_{i,j}}{N} = 1 = \frac{1}{80} (8 + 24 + 30 + 18) = 1$$



Some questions

- What is the probability that the selection procedure will produce an apple?
- What is the probability that the box selected was the blue one if we pick up an orange?



A.A. 2017-2018

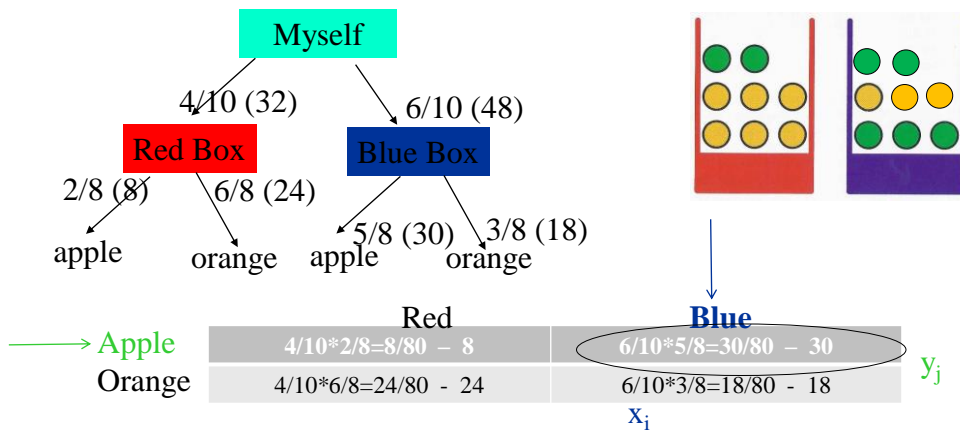
21/27

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



Probabilità congiunta

- What is the probability that the **blue** box is chosen **and** an **apple** is selected?



$$P(X=x_i; Y=y_j) = \lim_{N \rightarrow +\infty} \frac{n_{i,j}}{N}$$

$$P(X=b; Y=a) = 30/80 = 3/8$$

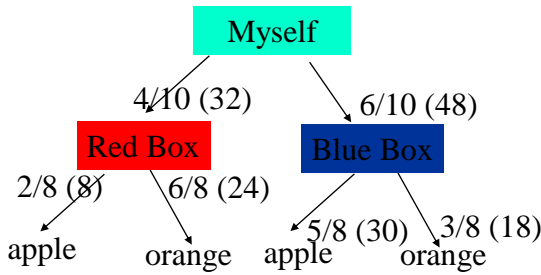
A.A. 2017-2018

22/27

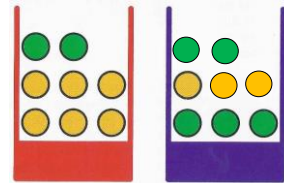
<http://borghese.di.unimi.it/>



Conditional (posterior) probability



Apple when the blue box is chosen.



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

y_j

x_i

$$P(Y=y_j | X = x_i) = \lim_{N \rightarrow +\infty} \frac{n_{ij}}{\sum_j n_{ij}}$$

$$P(Y=a | X = b) = \frac{n_{a,b}}{n_{a,b} + n_{o,b}} = \frac{30}{30 + 18} = 5/8$$

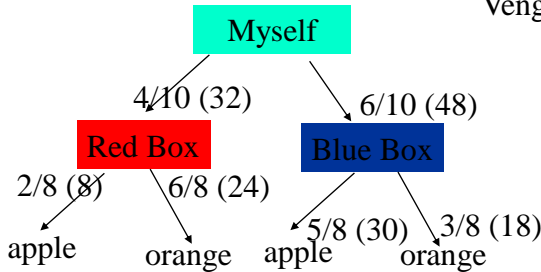
A.A. 2017-2018

23/27

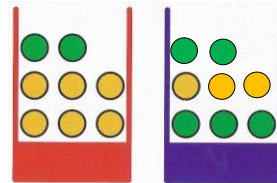
<http://borghese.di.unimi.it>



Probabilità semplice (marginale)



Vengono eliminate le altre variabili



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

y_j

x_i

$$P(X = r) = \lim_{N \rightarrow +\infty} \frac{\sum_j n_{rj}}{N} = \frac{8}{80} + \frac{24}{80} = \frac{32}{80} = 4/10$$

$$P(Y = o) = \lim_{N \rightarrow +\infty} \frac{\sum_i n_{i,o}}{N} = \frac{24}{80} + \frac{18}{80} = \frac{42}{80}$$

$$P(Y = a) = \lim_{N \rightarrow +\infty} \frac{\sum_i n_{i,a}}{N} = \frac{8}{80} + \frac{30}{80} = \frac{38}{80}$$

A.A. 2017-2018

24/27

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



Probabilità marginale

$$P(X = r) = \lim_{N \rightarrow +\infty} \frac{\sum_j n_{r,j}}{N} = \frac{8}{80} + \frac{24}{80} = \frac{32}{80} = 4/10$$

E' la probabilità ottenuta sommando le probabilità delle altre variabili (mettendo ai margini), in questo caso la variabile Y (frutto)

$$P(X = r) = \lim_{N \rightarrow +\infty} \frac{\sum_j n_{r,j}}{N} = \sum_j P(X = r, Y = y_j)$$

$$P(Y = o) = \lim_{N \rightarrow +\infty} \frac{\sum_i n_{i,o}}{N} = \sum_i P(X = x_i, Y = o)$$

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

y_j

x_i



Riassumendo

$$P(X=x_i \text{ AND } Y=y_j) = \lim_{N \rightarrow +\infty} \frac{n_{i,j}}{N} \quad \text{Probabilità congiunta}$$

$$= P(Y=y_j | X = x_i) P(X=x_i)$$

$$P(X=x_i) = \lim_{N \rightarrow +\infty} \frac{\sum_j n_{ij}}{N} \quad \text{Probabilità semplice (marginale)}$$

$$P(Y=y_j | X = x_i) = \lim_{N \rightarrow +\infty} \frac{n_{ij}}{\sum_j n_{ij}} \quad \text{Probabilità condizionata}$$

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



Overview



Probabilità semplice e condizionata

Esempio