Probabilità e Processi Stocastici (455AA)

Lezione 2

Dario Trevisan – *https://web.dm.unipi.it/trevisan*

24/09/2025

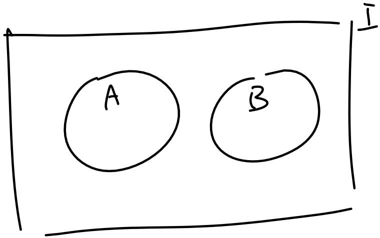
Richiami

Regole principali

- $0 \leq P(A|I) \leq 1$

Regole principali

- $0 \leq P(A|I) \leq 1$
- **Somma** $P(A \text{ oppure } B) = P(A) + P(B)$ purché A, B *incompatibili*,
$$P(A \text{ e } B) = 0.$$



Regole principali

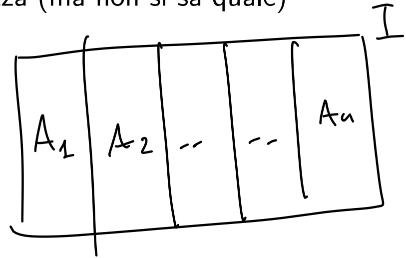
- $0 \leq P(A|I) \leq 1$
- **Somma** $P(A \text{ oppure } B) = P(A) + P(B)$ purché A, B *incompatibili*,
$$P(A \text{ e } B) = 0.$$
- Negazione $P(\text{ non } A) = 1 - P(A)$

Regole principali

- $0 \leq P(A|I) \leq 1$
- **Somma** $P(A \text{ oppure } B) = P(A) + P(B)$ purché A, B *incompatibili*,
$$P(A \text{ e } B) = 0.$$
- Negazione $P(\text{ non } A) = 1 - P(A)$
- **Prodotto** $P(A \text{ e } B) = P(A)P(B|A)$.

Regole principali

- $0 \leq P(A|I) \leq 1$
- **Somma** $P(A \text{ oppure } B) = P(A) + P(B)$ purché A, B *incompatibili*,
$$P(A \text{ e } B) = 0.$$
- Negazione $P(\text{ non } A) = 1 - P(A)$
- **Prodotto** $P(A \text{ e } B) = P(A)P(B|A)$.
- Sistema di alternative $(A_i)_{i=1}^n$ se *esattamente una* delle A_i si realizza (ma non si sa quale)



Regole principali

- $0 \leq P(A|I) \leq 1$
- **Somma** $P(A \text{ oppure } B) = P(A) + P(B)$ purché A, B *incompatibili*,
$$P(A \text{ e } B) = 0.$$
- Negazione $P(\text{ non } A) = 1 - P(A)$
- **Prodotto** $P(A \text{ e } B) = P(A)P(B|A)$.
- Sistema di alternative $(A_i)_{i=1}^n$ se *esattamente una* delle A_i si realizza (ma non si sa quale)
- Densità discreta associata a un sistema $(A_i)_{i=1}^n$ e all'informazione I:
$$i \mapsto P(A_i|I)$$

Regole principali

- $0 \leq P(A|I) \leq 1$
- **Somma** $P(A \text{ oppure } B) = P(A) + P(B)$ purché A, B *incompatibili*,
$$P(A \text{ e } B) = 0.$$
- Negazione $P(\text{ non } A) = 1 - P(A)$
- **Prodotto** $P(A \text{ e } B) = P(A)P(B|A)$.
- Sistema di alternative $(A_i)_{i=1}^n$ se *esattamente una* delle A_i si realizza (ma non si sa quale)
- Densità discreta associata a un sistema $(A_i)_{i=1}^n$ e all'informazione I:
$$i \mapsto P(A_i|I)$$
- Densità Bernoulli, densità uniforme

Diagrammi ad albero: generalità

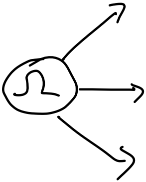
- Le regole di somma e prodotto (e le conseguenti formule di disintegrazione) forniscono stumenti utili per *analizzare* problemi di probabilità elementare, riducendoli a sotto-problemi più semplici tramite l'introduzione (anche ripetuta) di **sistemi di alternative**.

Diagrammi ad albero: generalità

- Le regole di somma e prodotto (e le conseguenti formule di disintegrazione) forniscono stumenti utili per *analizzare* problemi di probabilità elementare, riducendoli a sotto-problemi più semplici tramite l'introduzione (anche ripetuta) di **sistemi di alternative**.
- È utile rappresentare l'analisi tramite **diagrammi ad albero** costruiti con il seguente algoritmo.

Costruzione dell'albero

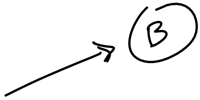
- Si introduce un nodo **radice** (informazione iniziale, ad esempio indicata con Ω). Si iterano poi i seguenti passi:



Costruzione dell'albero

- Si introduce un nodo **radice** (informazione iniziale, ad esempio indicata con Ω). Si iterano poi i seguenti passi:

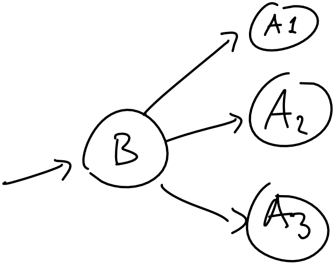
1. si considera un nodo del grafo che sia una "foglia", ossia senza archi uscenti, etichettato da una affermazione B,



Costruzione dell'albero

- Si introduce un nodo **radice** (informazione iniziale, ad esempio indicata con Ω). Si iterano poi i seguenti passi:

1. si considera un nodo del grafo che sia una "foglia", ossia senza archi uscenti, etichettato da una affermazione B,
2. si sceglie un sistema di alternative $(A_i)_{i=1}^n$, e si introducono tanti nodi quante le alternative, etichettate appunto da esse,



Costruzione dell'albero

- Si introduce un nodo **radice** (informazione iniziale, ad esempio indicata con Ω). Si iterano poi i seguenti passi:

1. si considera un nodo del grafo che sia una "foglia", ossia senza archi uscenti, etichettato da una affermazione B,
2. si sceglie un sistema di alternative $(A_i)_{i=1}^n$, e si introducono tanti nodi quante le alternative, etichettate appunto da esse,
3. si introducono archi uscenti dalla foglia (B) verso il nodo corrispondente a ciascuna alternativa (A_i),

Costruzione dell'albero

- Si introduce un nodo **radice** (informazione iniziale, ad esempio indicata con Ω). Si iterano poi i seguenti passi:

1. si considera un nodo del grafo che sia una "foglia", ossia senza archi uscenti, etichettato da una affermazione B,
2. si sceglie un sistema di alternative $(A_i)_{i=1}^n$, e si introducono tanti nodi quante le alternative, etichettate appunto da esse,
3. si introducono archi uscenti dalla foglia (B) verso il nodo corrispondente a ciascuna alternativa (A_i),
4. si pesa ciascun arco introdotto sopra con la probabilità

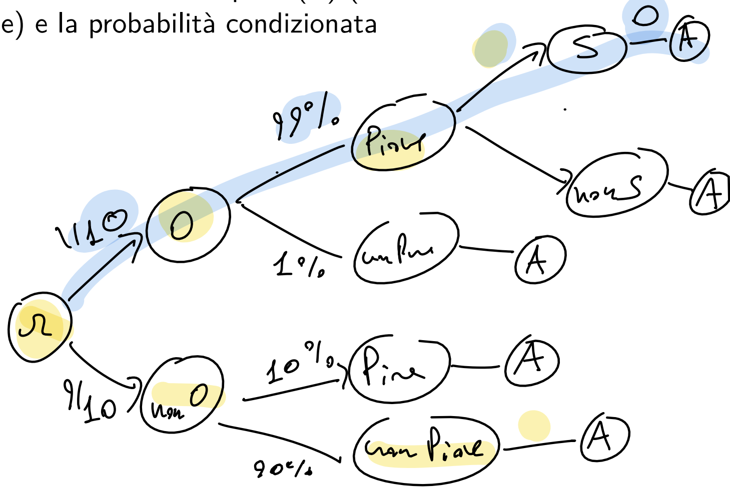
$$P(A_i|B, I),$$

   dove I consiste della congiunzione di tutte le affermazioni nell'unico cammino (orientato) che collega l'informazione iniziale Ω a B.

Supponendo di aver costruito un albero, come calcolare $P(A|\Omega)$?

- Si effettuano ancora una volta i passaggi 1-4 aggiungendo ad ogni foglia l'alternativa semplice (A) (ed eventualmente la sua negazione) e la probabilità condizionata

Note:



Calcolo delle probabilità

Supponendo di aver costruito un albero, come calcolare $P(A|\Omega)$?

- Si effettuano ancora una volta i passaggi 1-4 aggiungendo ad ogni foglia l'alternativa semplice (A) (ed eventualmente la sua negazione) e la probabilità condizionata
- Si calcola il peso di ciascun cammino che porta da una foglia (A) verso la radice, **moltiplicando** le probabilità sugli archi (è una applicazione della regola del prodotto)

Note:

Calcolo delle probabilità

Supponendo di aver costruito un albero, come calcolare $P(A|\Omega)$?

- Si effettuano ancora una volta i passaggi 1-4 aggiungendo ad ogni foglia l'alternativa semplice (A) (ed eventualmente la sua negazione) e la probabilità condizionata
- Si calcola il peso di ciascun cammino che porta da una foglia (A) verso la radice, **moltiplicando** le probabilità sugli archi (è una applicazione della regola del prodotto)
- Si **sommano** i pesi di tutti i cammini così ottenuti (è una applicazione della regola della somma)

Note:

Calcolo delle probabilità

Supponendo di aver costruito un albero, come calcolare $P(A|\Omega)$?

- Si effettuano ancora una volta i passaggi 1-4 aggiungendo ad ogni foglia l'alternativa semplice (A) (ed eventualmente la sua negazione) e la probabilità condizionata
- Si calcola il peso di ciascun cammino che porta da una foglia (A) verso la radice, **moltiplicando** le probabilità sugli archi (è una applicazione della regola del prodotto)
- Si **sommano** i pesi di tutti i cammini così ottenuti (è una applicazione della regola della somma)

Note:

1. se un arco ha probabilità 0 allora i pesi dei cammini che lo percorrono sono 0, non serve contarli.

Calcolo delle probabilità

Supponendo di aver costruito un albero, come calcolare $P(A|\Omega)$?

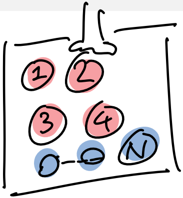
- Si effettuano ancora una volta i passaggi 1-4 aggiungendo ad ogni foglia l'alternativa semplice (A) (ed eventualmente la sua negazione) e la probabilità condizionata
- Si calcola il peso di ciascun cammino che porta da una foglia (A) verso la radice, **moltiplicando** le probabilità sugli archi (è una applicazione della regola del prodotto)
- Si **sommano** i pesi di tutti i cammini così ottenuti (è una applicazione della regola della somma)

Note:

1. se un arco ha probabilità 0 allora i pesi dei cammini che lo percorrono sono 0, non serve contarli.
2. se è richiesta la probabilità $P(A|I)$ è l'informazione (cumulata) I è un nodo dell'albero, basta considerare solo l'albero con radice I

Attenzione!

Estrazioni da un'urna (senza rimpiazzo)



Scatola con $N = 10$ palline

$3 = R$ rosse $\{1, 2, 3, \dots, R\}$

$B = N - R = 7$ blu $\{R+1, R+2, \dots, N\}$

Ω

R1 = "prima pallina estratta è rossa"

B1 = " " " " è blu "

$$P(R1 \mid \Omega) = \frac{R}{N}$$

$$P(B1 \mid \Omega) = \frac{B}{N} = 1 - \frac{R}{N}$$

$$P(\text{"estrae pallina } i\text{"} \mid \Omega) = \frac{1}{N}$$

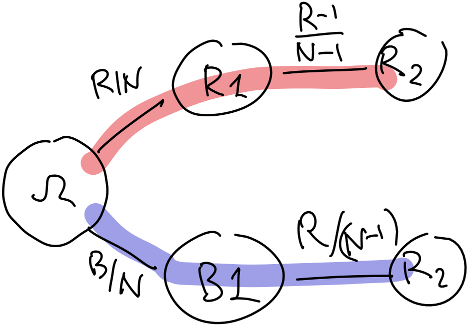
R2 = "seconda pallina rossa"
B2 = " " " blu "

$N \geq 2$

$$\boxed{P(R2|\Omega) = \frac{R}{N}}$$

$$P(R2|R1,\Omega) \overset{?}{=}$$

$$P(R2|B1,\Omega) \overset{?}{=}$$



$$\frac{R}{N} \cdot \frac{R-1}{N-1}$$

$$\frac{B}{N} \cdot \frac{R}{N-1}$$

$+$

$$P(R_2|\Omega) = \frac{R}{N} \cdot \frac{R-1}{N-1} + \frac{B}{N} \cdot \frac{R}{N-1}$$

$$\left[P(\underbrace{R1, B2, R3, B4}\,|\,\Omega) \overset{?}{=} \frac{R}{N} \cdot \frac{B}{N-1} \cdot \frac{R-1}{N-2} \cdot \frac{B-1}{N-3} \right.$$

- **Probabilità di estrarre una precisa sequenza ordinata** di $n \leq N$ palline colorate, di cui $r \leq R$ sono rosse e le rimanenti $b \leq B$ sono blu:

$$\frac{\overbrace{R(R-1)\cdot\ldots\cdot(R-r+1)}^{r \text{ fattori}} \cdot \overbrace{B(B-1)\cdot\ldots\cdot(B-b+1)}^{b \text{ fattori}}}{\underbrace{N(N-1)\cdot\ldots(N-n+1)}_{n \text{ fattori}}}.$$

$$\parallel$$

$$\underline{P(}\, R_1, R_2, R_3, \ldots, R_r, B_{r+1}, B_{r+2}, \ldots, B_n \,|\, \Omega)$$

$$P(R_2 | \Omega) = \frac{R}{N}$$

- **Probabilità di estrarre una precisa sequenza ordinata** di $n \leq N$ palline colorate, di cui $r \leq R$ sono rosse e le rimanenti $b \leq B$ sono blu:

$$\frac{R(R-1)\cdot \ldots \cdot (R-r+1)\cdot B(B-1)\cdot \ldots \cdot (B-b+1)}{N(N-1)\cdot \ldots (N-n+1)}.$$

Densità ipergeometrica

$$P(\text{"estrarre 2 rosse su 3 estrazioni"} \mid \Omega)$$

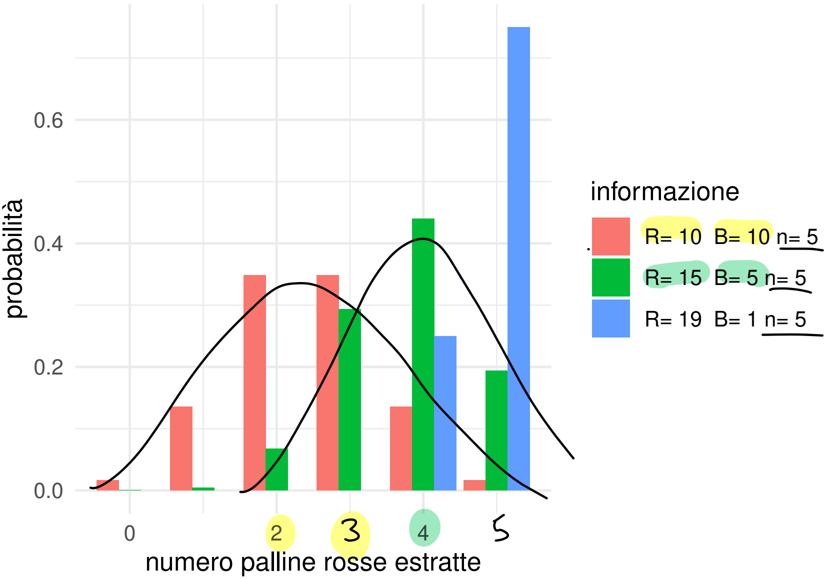
$$\stackrel{?}{=} \text{esercizio}$$

- **Probabilità di estrarre una qualsiasi sequenza ordinata** di $n \leq N$ palline colorate, di cui $r \leq R$ sono rosse e le rimanenti $b \leq B$ sono blu:

$$\boxed{\frac{\binom{R}{r}\binom{B}{b}}{\binom{N}{n}}.}$$

$$\binom{n}{k} = \frac{n(n-1)\cdot \ldots \cdot (n-k+1)}{k(k-1)\cdot \ldots \cdot 1} \qquad \text{coeff. binomiale}$$

$$r \longmapsto \frac{\binom{R}{r}\binom{B}{b}}{\binom{N}{n}} = P(A_r \mid \Omega) \qquad r = 0, 1, 2, \ldots, n$$


probabilità
0.0
0.2
0.4
0.6
0
2
3
4
5
numero palline rosse estratte
informazione
R= 10 B= 10 n= 5
R= 15 B= 5 n= 5
R= 19 B= 1 n= 5

Formula di Bayes e applicazioni

Formula di Bayes

A, B affermazioni

$$P(A \text{ e } B) = P(A)\,P(B|A)$$

$$\shortparallel$$

$$P(B \text{ e } A) = P(B)\,\boxed{P(A|B)}$$

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

Date affermazioni A, B e l'informazione nota I, vale

$$\overbrace{P(A|B,I)}^{\text{“a posteriori”}} = \underbrace{P(A|I)}_{\text{“a priori”}} \overbrace{\frac{P(B|A,I)}{P(B|I)}}^{\text{verosimiglianza}}$$

(purché $P(B|I) > 0$).

Verosimiglianza (*Likelihood*) di A rispetto a B,

$$L(A; B) = P(B|A).$$

$$P(A|B) = \frac{P(A) \cdot L(A; B)}{P(B)}$$

Statistica bayesiana

Dato un sistema di alternative $(A_i)_{i=1}^n$ e una qualsiasi affermazione B, otteniamo

densità discreta

$$P(A_i|B) = P(A_i)P(B|A_i) \cdot \frac{1}{P(B)}. \tag{1}$$

densità discreta

Notazione compatta (più facile da ricordare)

$$P(A_i|B) \propto P(A_i)P(B|A_i) = P(A_i)L(A_i; B). \tag{2}$$

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B|A_i) \leftarrow$$

formula di disintegrazione

$$(A_i)_{i} = (A_0, A_1) \qquad B$$

$B =$ "ombrello" $\qquad A_0 =$ piove

$\qquad\qquad\qquad\qquad\qquad A_1 =$ non piove

$$P(A_0 \mid B) \overset{?}{<} P(A_1 \mid B)$$

$$\frac{P(A_0)\,P(B \mid A_0)}{P(B)} \qquad\qquad \frac{P(A_1)\,P(B \mid A_1)}{P(B)}$$

moda $\quad i_{max} \in \arg\max \{ P(A_i | B) \}$

Massimo a posteriori stima

$$i_{\mathrm{MAP}} \in \arg\max\{P(A_i)L(A_i;B) \; : \; i \in \{1,\dots,n\}\},$$

Se $P(A_i) = \frac{1}{n}$ uniforme

Stima di **massima verosimiglianza**

$$i_{\text{MLE}} \in \arg\max_{i=1,\dots,n} L(A_i; B) \cdot \frac{1}{n}$$

$\underline{i_{MAP} = i_{MLE}}$ se a priori (A_i) uniformi

Esempio: estrazioni da un'urna non nota

Il soggetto *non* è informato sul numero di palline rosse R, ma solamente sul totale $N = 3$. Dopo un'estrazione, si osserva una pallina rossa è stata estratta. Cosa può dedurre circa il contenuto dell'urna?

Introduciamo un sistema di alternative relativo al contenuto dell'urna:

$$A_i = \text{l'urna contiene } R = i \text{ palline rosse},$$

per $i \in \{0, 1, 2, 3\}$.

Probabilità a priori uniforme

$$P(A_i|\Omega) = \frac{1}{4}.$$

sapendo che ci sono
i rosse

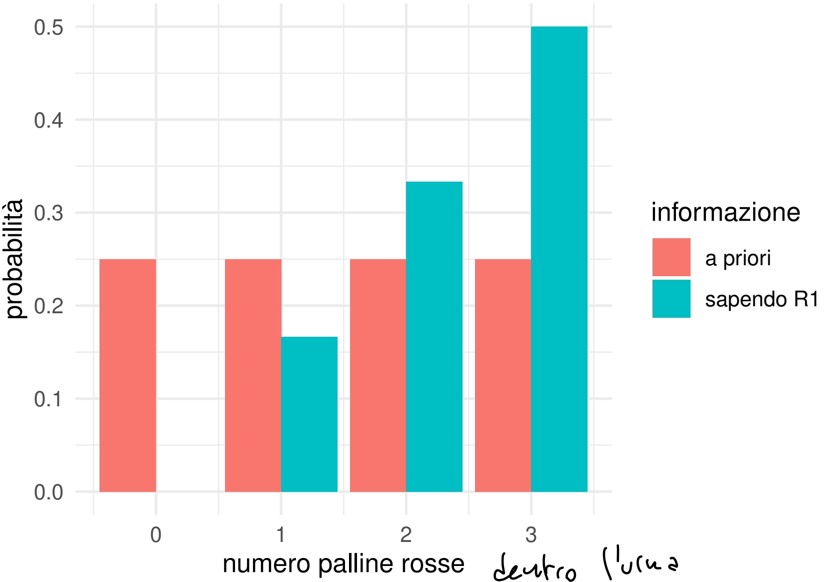
Verosimiglianza

$$L(A_i; R^1) = P(R^1|A_i) = \frac{i}{3},$$

e formula di Bayes:

$$P(A_i|R^1) \propto \frac{1}{4} \cdot \frac{i}{3} \propto i$$

Possiamo confrontare le densità discrete:



dentro l'urna

Decisioni e test statistici

Obiettivo: scartare (rifiutare) alcune alternative A_i sulla base dell'osservazione di B. **Intuizione**: eliminare quelle con probabilità a posteriori nulla (ovvio) o bassa.

Nella pratica ci si appoggia alla verosimiglianza invece delle probabilità a posteriori (simile se a priori sono uniformi).

Ipotesi nulla $\mathcal{H}_0$ (da rifiutare)) e alternativa $\mathcal{H}_1$. **Valore** p (*p-value*) di Fisher:

$$p = \max_{A_i \in \mathcal{H}_0} P(B|A_i) = \max_{A_i \in \mathcal{H}_0} L(A_i; B).$$

Più piccolo è il valore p, minore è la probabilità che B sia vero rispetto a l'ipotesi $\mathcal{H}_0$, e quindi, invocando Bayes, anche che A_i sia vero sapendo B. Quindi possiamo rifiutare $\mathcal{H}_0$ con maggiore sicurezza.

Torniamo all'esempio dell'urna: seconda estrazione

Cosa deduce il soggetto se viene informato che alla seconda estrazione (senza rimpiazzo) la pallina estratta è blu (B^2)?

Non serve tornare alla densità iniziale (uniforme), ma basta usare come nuova densità *a priori* la densità discreta rispetto all'informazione R^1. Bayes:

$$P(A_i|R^1, B^2) = P(A_i|R^1)P(B^2|R^1, A_i) \cdot \frac{1}{P(B^2|R^1)}.$$

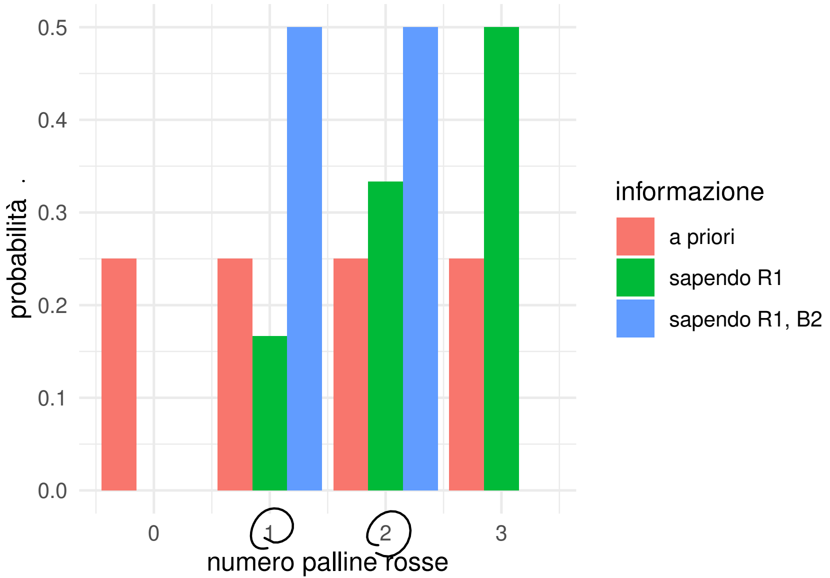
Otteniamo

$$P(B^2|R^1, A_i) = \frac{3-i}{2}.$$

e

$$P(A_i|R^1, B^2) \propto i(3-i)$$

.



0.5
0.4
0.3
0.2
0.1
0.0
probabilità
0
1
2
3
numero palline rosse
informazione
a priori
sapendo R1
sapendo R1, B2

Cenni agli assiomi di Kolmogorov

Problemi dell'approccio elmentare

1. Come attribuire le **probabilità iniziali** (quelle che abbiamo chiamato *a priori*)?

Problemi dell'approccio elmentare

1. Come attribuire le **probabilità iniziali** (quelle che abbiamo chiamato *a priori*)?
2. Come garantire la **consistenza** del calcolo, ossia che $P(A|I)$ sia ben definita?

Problemi dell'approccio elmentare

1. Come attribuire le **probabilità iniziali** (quelle che abbiamo chiamato *a priori*)?
2. Come garantire la **consistenza** del calcolo, ossia che $P(A|I)$ sia ben definita?
3. Come trattare i **passaggi al limite**, in particolare, nel caso di infinite affermazioni?

Alcune risposte sono fornite dalla **descrizione assiomatica** della probabilità proposta da Kolmogorov.

L'idea: formalizzare i diagrammi di Eulero-Venn, identificando

- le affermazioni (eventi) A, I ecc. con *sottoinsiemi* di un insieme "universo" Ω (l'informazione iniziale)

Alcune risposte sono fornite dalla **descrizione assiomatica** della probabilità proposta da Kolmogorov.

L'idea: formalizzare i diagrammi di Eulero-Venn, identificando

- le affermazioni (eventi) A, I ecc. con *sottoinsiemi* di un insieme "universo" Ω (l'informazione iniziale)
- la probabilità $P(A|I)$ con una nozione astratta di *area* (relativa), definendo prima la probabilità rispetto all'informazione inizale Ω.

a) insieme Ω

Si fissa un insieme “universo” Ω che rappresenta tutte le possibili situazioni (scenari) che si potrebbero presentare nel problema che si sta considerando.

Ad esempio, nel caso di un lancio di dado,

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

che corrisponde ai possibili esiti (ovviamente altre scelte sono ragionevoli).

b) Eventi

Si identificano quali affermazioni $A \subseteq \Omega$ sono d'interesse per il problema che si vuole affrontare.

Insieme $\mathcal{A}$ i cui elementi $A \in \mathcal{A}$ sono sottoinsiemi di Ω, detto la σ*-algebra degli eventi*:

- $\Omega \in \mathcal{A}$,

b) Eventi

Si identificano quali affermazioni $A \subseteq \Omega$ sono d'interesse per il problema che si vuole affrontare.

Insieme $\mathcal{A}$ i cui elementi $A \in \mathcal{A}$ sono sottoinsiemi di Ω, detto la *σ-algebra degli eventi*:

- $\Omega \in \mathcal{A}$,
- se A, $B \in \mathcal{A}$, anche A^c ("non A"), $A \cap B$ ("A e B") e $A \cup B$ ("A oppure B") sono eventi in $\mathcal{A}$.

b) Eventi

Si identificano quali affermazioni $A \subseteq \Omega$ sono d'interesse per il problema che si vuole affrontare.

Insieme $\mathcal{A}$ i cui elementi $A \in \mathcal{A}$ sono sottoinsiemi di Ω, detto la *σ-algebra degli eventi*:

- $\Omega \in \mathcal{A}$,
- se A, $B \in \mathcal{A}$, anche A^c ("non A"), $A \cap B$ ("A e B") e $A \cup B$ ("A oppure B") sono eventi in $\mathcal{A}$.
- Inoltre, per permettere di passare al limite, si richede che valga lo stesso per l'unione infinita di eventi: dati $A_n \in \mathcal{A}$, pure $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$.

c) funzione di probabilità (iniziale)

Si introduce una $P : \mathcal{A} \to [0,1]$ tale che $P(\Omega) = 1$ e, per ogni A, $B \in \mathcal{A}$ con $A \cap B = \emptyset$ valga

$$P(A \cup B) = P(A) + P(B).$$

- $P(A)$ corrisponde alla probabilità $P(A|\Omega)$ rispetto all'informazione iniziale, di cui si richiede valga la regola della somma (per eventi incompatibili).

c) funzione di probabilità (iniziale)

Si introduce una $P : \mathcal{A} \to [0,1]$ tale che $P(\Omega) = 1$ e, per ogni A, $B \in \mathcal{A}$ con $A \cap B = \emptyset$ valga

$$P(A \cup B) = P(A) + P(B).$$

- $P(A)$ corrisponde alla probabilità $P(A|\Omega)$ rispetto all'informazione iniziale, di cui si richiede valga la regola della somma (per eventi incompatibili).
- Per passare al limite, si richiede in più che la regola della somma si estenda ad infiniti $A_n \in \mathcal{A}$ tali che $A_n \cap A_m = \emptyset$ per $n \neq m$:
$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

d) Probabilità condizionata ad I

Per ogni $A, I \in \mathcal{A}$ tale che $P(I) > 0$, si pone

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

$(\Omega, \mathcal{A}, P)$ è uno **spazio di probabilità** secondo Kolmogorov.