

Probabilità e Processi Stocastici (455AA)

Lezione 2

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

24/09/2025

Richiami

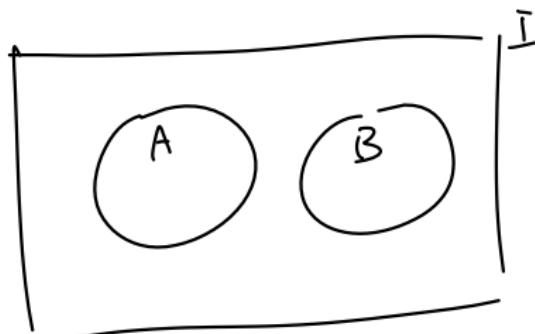
Regole principali

- ▶ $0 \leq P(A|I) \leq 1$

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*,

$$P(A \text{ e } B) = 0.$$

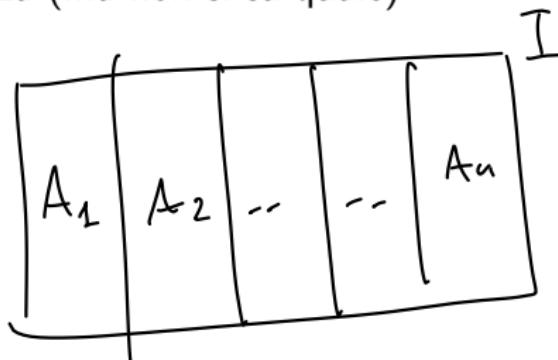
- ▶ Negazione $P(\text{ non } A) = 1 - P(A)$
- ▶ **Prodotto** $P(A \text{ e } B) = P(A)P(B|A).$

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

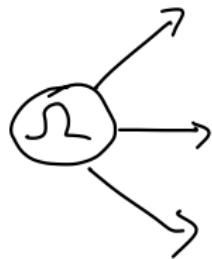
- ▶ Le regole di somma e prodotto (e le conseguenti formule di disintegrazione) forniscono strumenti utili per *analizzare* problemi di probabilità elementare, riducendoli a sotto-problemi più semplici tramite l'introduzione (anche ripetuta) di **sistemi di alternative**.

Diagrammi ad albero: generalità

- ▶ Le regole di somma e prodotto (e le conseguenti formule di disintegrazione) forniscono strumenti 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:



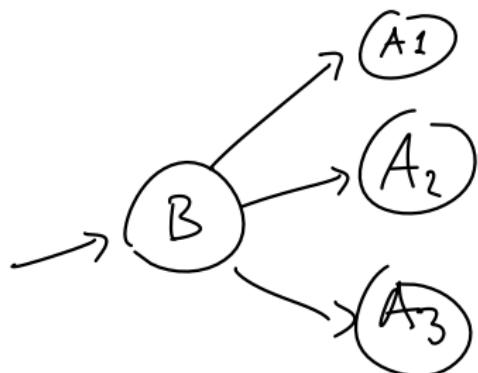
Costruzione dell'albero

- ▶ 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

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

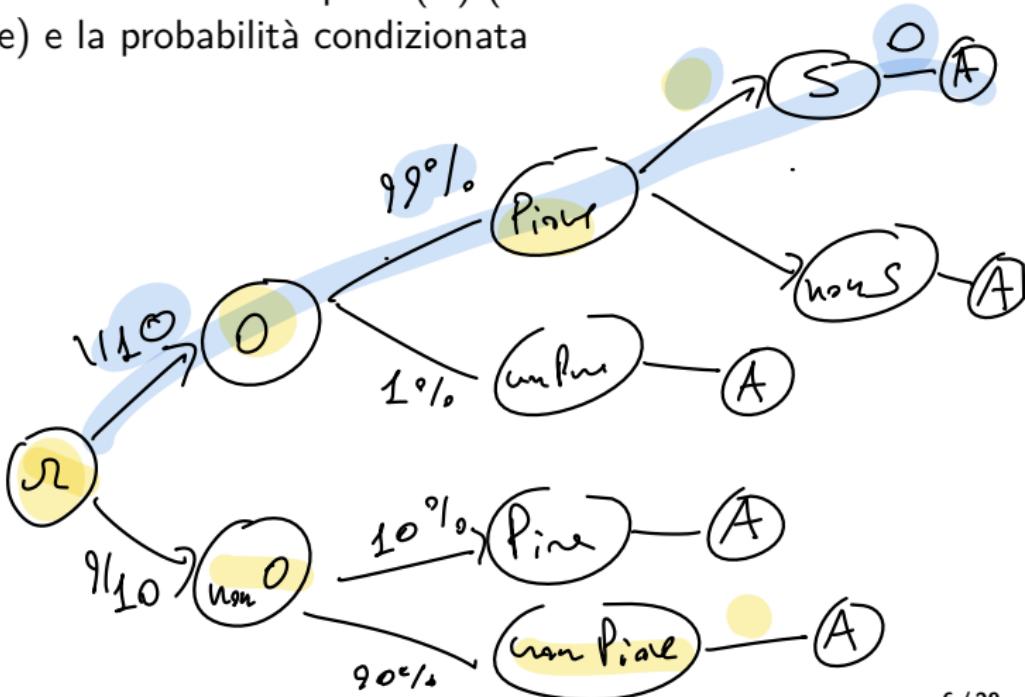
dove I consiste della congiunzione di tutte le affermazioni nell'unico cammino (orientato) che collega l'informazione iniziale Ω a B .

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

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 **sommiano** 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 **sommiano** 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 $P(A|\Omega)$?

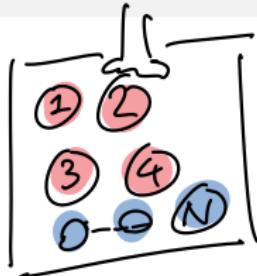
- ▶ Si effettuano ancora una volta i passaggi 1-4 aggiungendo ad ogni foglia l'alternativa semplice (A) (ed eventualmente la sua negazione) e la probabilità condizionata
- ▶ Si calcola il peso di ciascun cammino che porta da una foglia (A) verso la radice, **moltiplicando** le probabilità sugli archi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

Attenzione! / è un nodo dell'albero, basta considerare solo l'albero con radice /

Estrazioni da un'urna (senza rimpiazzo)



Scatola con $N=10$ palline

$R = 4$ rosse $\{1, 2, 3, \dots, R\}$

$B = N - R = 6$ blu $\{R+1, R+2, \dots, N\}$

$R_1 =$ "prima pallina estratta è rossa"

$B_1 =$ " " " " è blu"

$$P(R_1 | \mathcal{S}_2) = \frac{R}{N}$$

$$P(\text{"estrae pallina i"} | \mathcal{S}_2) = \frac{1}{N}$$

$$P(B_1 | \mathcal{S}_2) = \frac{B}{N} = 1 - \frac{R}{N}$$

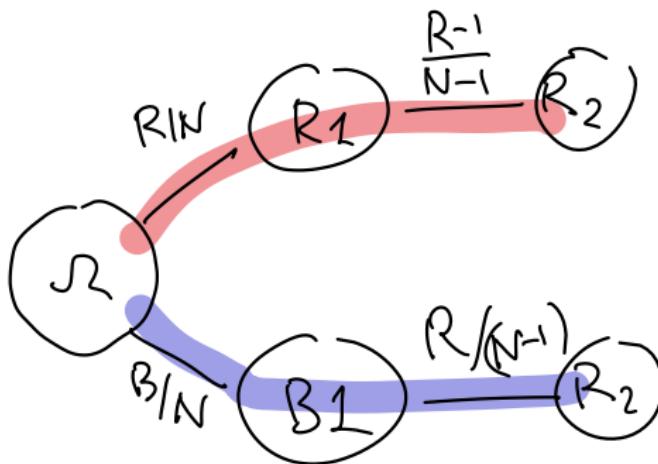
R_2 = "seconda pallina rossa"
 B_2 = " " " blu "

$N \geq 2$

$$P(R_2 | \mathcal{S}_2) = \frac{R}{N}$$

$$P(R_2 | R_1, \mathcal{S}_2) = ?$$

$$P(R_2 | B_1, \mathcal{S}_2) = ?$$



• $\frac{R}{N} \cdot \frac{R-1}{N-1}$

• $\frac{B}{N} \cdot \frac{R}{N-1}$

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

$$P(\underbrace{R_1, B_2, R_3, B_4}_{\text{?}} | \Omega) = \frac{R}{N} \cdot \frac{B}{N-1} \cdot \frac{R-1}{N-2} \cdot \frac{B-1}{N-3}$$

- ▶ 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: $\underbrace{(r \text{ fattori})}_{\text{n fattori}}$ $\underbrace{(b \text{ fattori})}_{\text{n fattori}}$

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

$\underbrace{\hspace{10em}}_{n \text{ fattori}}$

$$P(R_1, R_2, R_3, \dots, R_c, B_{c+1}, B_{c+2}, \dots, B_n | \Omega)$$

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

- ▶ **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:

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

Densità ipergeometrica

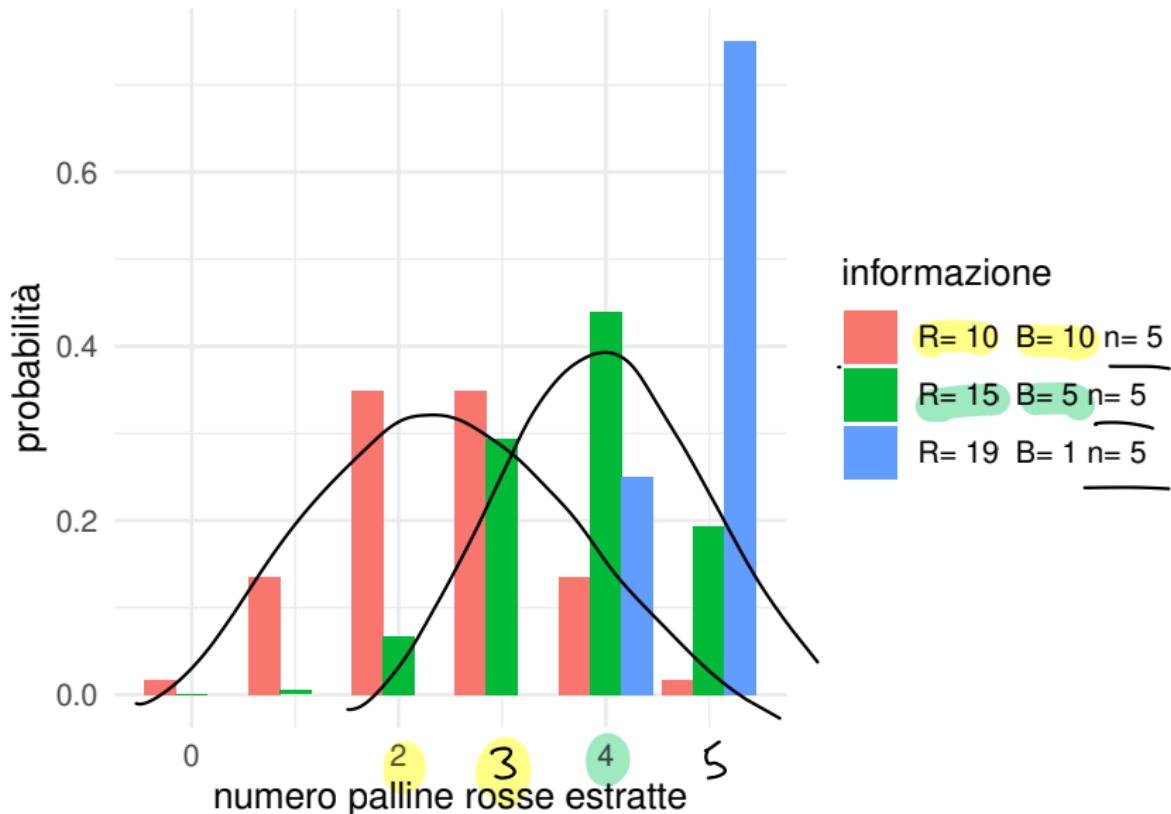
P ("estrarre 2 rosse su 3 estratti" (S))
=?
= esercizio

- ▶ **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:

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

$$\binom{n}{k} = \frac{n(n-1) \cdots (n-k+1)}{k(k-1) \cdots 1} \quad \text{coeff. binomiale}$$

$$c \mapsto \frac{\binom{R}{c} \binom{B}{b}}{\binom{N}{n}} = P(A_c | S) \quad (c=0,1,2,\dots,n)$$



Formula di Bayes e applicazioni

Formula di Bayes A, B affermazioni

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

$$\text{"} \quad P(B \text{ e } A) = P(B)\overline{P(A|B)}$$

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

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

$$\text{"a posteriori"} \longrightarrow P(A|B, I) = P(A|I) \underbrace{\frac{P(B|A, I)}{P(B|I)}} \longrightarrow \text{verso i dati}$$

(purché $P(B|I) > 0$).

"a priori"

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

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

densità discreta

(1)

Notazione compatta (più facile da ricordare)

$$P(A_i|B) \propto P(A_i)P(B|A_i) = P(A_i)L(A_i; B). \quad (2)$$

$$P(B) = \sum_{i=1}^n P(A_i) \cdot P(B|A_i) \leftarrow \begin{matrix} \text{formula} \\ \text{discretizzazione} \end{matrix}$$

$$(A_i)_{i \in \cdot} = (A_0, A_1) \quad B$$

B = "ombrello" A_0 = pire

A_1 = um pire

$$P(A_0 | B) \stackrel{?}{<} P(A_1 | B)$$

$$\frac{P(A_0) P(B | A_0)}{P(B)}$$

$$\frac{P(A_1) P(B | A_1)}{P(B)}$$

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

Massimo a posteriori stima

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

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

Stima di **massima verosimiglianza**

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

$i_{MAP} = i_{MLE}$ se \geq priori(A_i) uniforme

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 = l'urna contiene $R = i$ palline rosse,
per $i \in \{0, 1, 2, 3\}$.

Probabilità a priori uniforme

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

Verosimiglianza

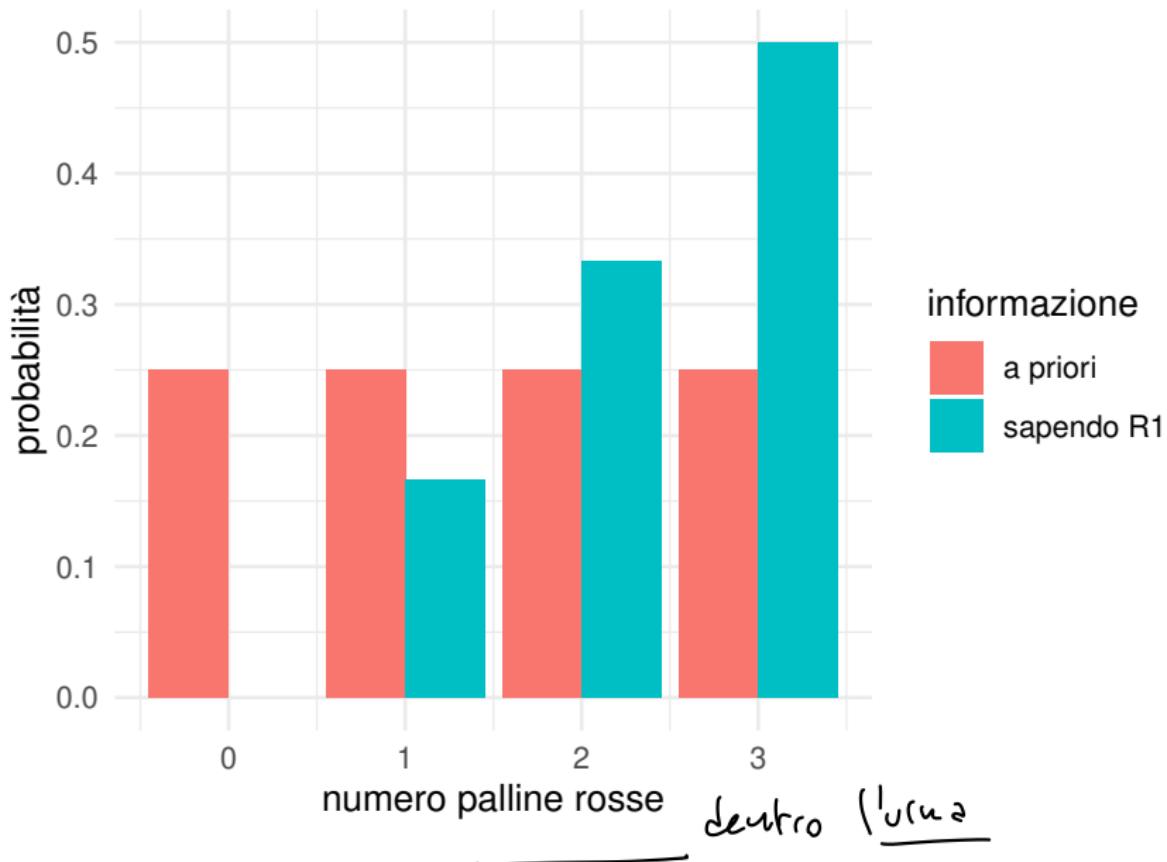
*Saperdo che ci sono
i cosse*

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

e formula di Bayes:

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

Possiamo confrontare le densità discrete:



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

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

