Probabilità e Processi Stocastici (455AA)

Lezione 2

Dario Trevisan

26/09/2024

Section 1

Richiami

Regole principali

Nota

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

"$P(I \to A)$" $\simeq P(A|I)$

Regole principali

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

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

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

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

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:

$$i \mapsto P(A_i|I)$$

<u>moda</u> $\quad i_{max} \in \arg\max \{ P(A_i|I) \}$

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

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

Estrazioni da un'urna (senza rimpiazzo)

$$\begin{cases} R^1 = \text{prima estratta è rossa} \\ B^1 = \text{non } R^1 \end{cases}$$

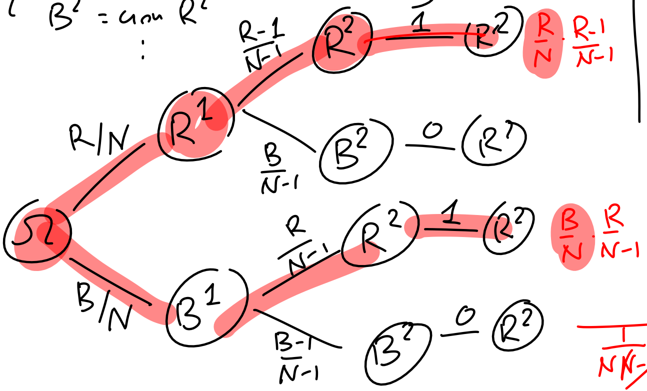
$$\begin{cases} R^2 = \text{seconda estratta rossa} \\ B^2 = \text{non } R^2 \end{cases}$$

$\vdots$

N palline

R rosse

$B = N - R$ blu

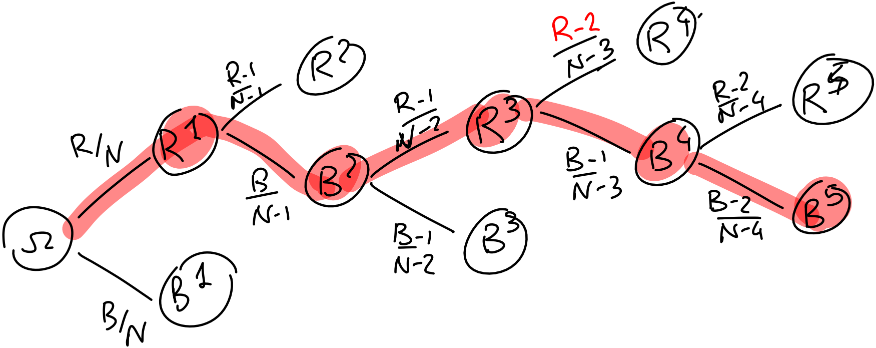


$$P(R^2 | \Omega) \overset{?}{=} \frac{R}{N}$$

$$P(B^2 | \Omega) = \frac{B}{N}$$

$$\frac{1}{N\cancel{(N-1)}} \cdot R(\cancel{R-1} + B)$$

$$P(R^1, B^2, R^3, B^4, B^5) \overset{?}{=} \frac{R\, B\, (R-1)(B-1)(B-2)}{N(N-1)(N-2)(N-3)(N-4)}$$



- **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:

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

Se invece non è importante l'ordine con cui compaiono i colori

- le varie sequenze ordinate lunghe n con r rosse e b blu sono tra loro incompatibili

$$P(\text{osservare } r \text{ rosse, } b \text{ blu in } n \text{ estrazioni}) =$$

$$= \sum_{\text{sequenze}} \frac{R(R-1)\cdots \quad B\cdot(B-1)\cdots}{N(N-1)\cdots} \qquad \binom{n}{r}$$

- **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:

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

è una densità discreta $\boxed{\text{Ipergeometrica}}$

del sistema $(A_r)_{r=0,\dots,n}$

<u>Esercizi</u>

- Calcolare $P(R^1 \text{ e } R^3 \mid R^2)$
- Calcolare $P(R^3 \mid \Omega)$
- " $P(R^3 \mid R^1)$

Section 2

Formula di Bayes e applicazioni

Formula di Bayes / Laplace

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

$$\|$$

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

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

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

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

(purché $P(B|I) > 0$).

"a priori"

$P(B|A,I)$

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$
$$= L(A;B)P(A) + L(A^c;B)P(A^c)$$

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

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

$$\underset{\text{"a posteriori"}}{P(A|B)} = \frac{P(A) \cdot L(A;B)}{P(B)}$$

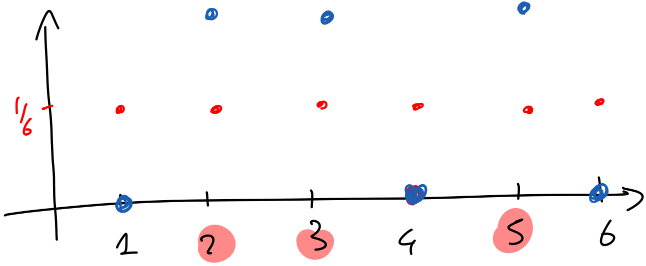
Statistica bayesiana

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

$$\underbrace{P(A_i|B)}_{\text{densità discreta}} = P(A_i)P(B|A_i) \cdot \frac{1}{P(B)}. \tag{1}$$

Notazione compatta (più facile da ricordare)

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



A_i = "esito del lancio è i"

B = "esito del lancio è un numero primo"

$$P(A_3 \mid B) = P(A_3) \cdot \underbrace{\overset{1}{\overset{\shortparallel}{P(B \mid A_3)}}}_{L(A_3;\, B)} \frac{1}{P(B)}$$

$$i \mapsto P(A_i \mid B) = P(A_i) L(A_i ; B) \frac{1}{P(B)}$$

Massimo a posteriori stima

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

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

Stima di **massima verosimiglianza**

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

$$L(A_i;B) = P(B|A_i)$$

Esempio: estrazioni da un'urna non nota

Il robot *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?

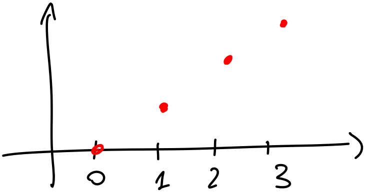
Introduciamo un sistema di alternative relativo al contenuto dell'urna:

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

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

Probabilità a priori uniforme

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



Verosimiglianza

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

e formula di Bayes:

$$P(A_i|R^1) \propto P(A_i) L(A_i; R^1)$$

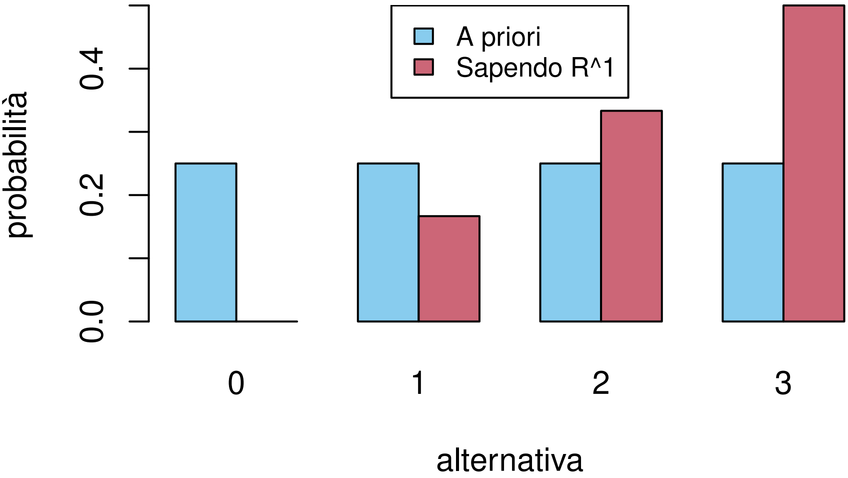
$$= \frac{1}{4} \cdot \frac{i}{3} = \frac{i}{12} \qquad i: 0 \dots 4$$

$$\propto i$$

Possiamo confrontare le densità discrete mediante un grafico a barre:

```
## [1] 0.25 0.25 0.25 0.25
```

```
## [1] 0.0000000 0.1666667 0.3333333 0.5000000
```



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:

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

$\mathcal{H}_1 = ?$

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 robot 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

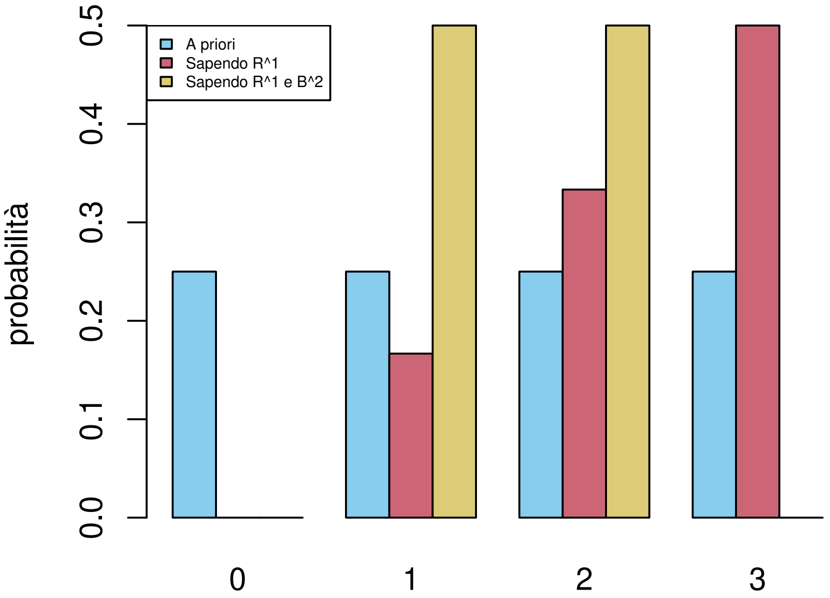
$$\boxed{P(B^2|R^1, A_i) = \frac{3-i}{2}.}$$

e

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

.

```
## [1] 0.0 0.5 0.5 0.0
```


A priori
Sapendo R^1
Sapendo R^1 e B^2
probabilità
0.0
0.1
0.2
0.3
0.4
0.5
0
1
2
3

- Nell'esempio dell'urna con R non noto cambiate la probabilità a priori

$$P(A_i \mid \Omega) \propto i^2 \qquad i \in \{0,1,2,3\}$$

- Provate a calcolare le prob. a posteriori riconducendosi a quella iniziale

Section 3

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?

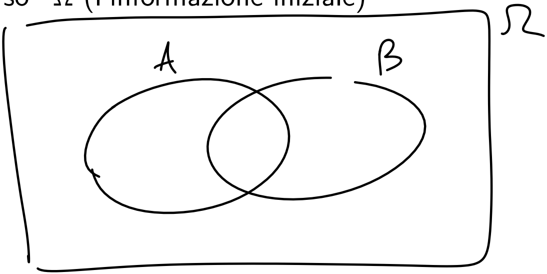
Problemi dell'approccio elmentare

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.

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

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

$\{1,2,3\} \subseteq \{1,2,3,4,5,6\}$

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

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

- $P(A)$ 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.