Cenni al Calcolo delle Probabilità
e
all'Inferenza Statistica

Francesco Palumbo

Università degli Studi di Napoli Federico II

a.a. 2016-2017

Riferimenti storici fondamentali

La *nascita* teoria della probabilità viene fatta coincidere con la pubblicazione del *Traité du Triangle Arithmétique* che parla del Triangolo di Tartaglia di Pascal nel 1654

1657 Huygens pubblica il *de ratiociniis in ludo alae* e nel 1666 Leibniz la sua *Dissertatio de arte combinatorica*.

1713 Bernoulli enuncia *La legge dei grandi numeri* nel *Ars conjectandi*.

1718 De Moivre risolve il problema centrale della teoria della probabilità

1763 Escono dei lavori di Bayes sulla concetto di Probabilità condizionata e alcune formule inerenti ad essa

1809 Durante il suo studio degli errori di osservazione in astronomia, Gauss ritrova la curva che poi in futuro prenderà il suo nome.

1809 Laplace dimostra il *teorema centrale limite*.

Gli studi sulla probabilità vengono assiomatizzati solo intorno agli anni venti e trenta del 1900 con la pubblicazione di Kolmogorov. Si diffonderanno solo a partire dal 1933 con la pubblicazione della traduzione in tedesco dei *Concetti fondamentali del Calcolo delle Probabilità* (Grundbegriffe der Wahrscheinlichkeitsrechnung)

1. Il calcolo delle probabilità è stata una delle più grandi scoperte della matematica. La conoscenza del calcolo delle probabilità influenza la nostra vita quotidiana molto di più di quanto si possa immaginare.
2. Ecco alcuni semplici affermazioni dietro le quali si scorge il contributo del calcolo delle probabilità
   - *"Deve fare un esame di screening"*
   - *"Quanto spendi per assicurare la tua automobile?"*
   - *"Ti è arrivata la mia email?"*
   - *"Lo vedo in forma. Vedrai che vince"*
   - ...
3. In tutte queste circostanze la domanda o l'affermazione presuppone la valutazione della possibilità che **un evento aleatorio si verifichi** in un tempo futuro

1. Il calcolo delle probabilità è stata una delle più grandi scoperte della matematica. La conoscenza del calcolo delle probabilità influenza la nostra vita quotidiana molto di più di quanto si possa immaginare.
2. Ecco alcuni semplici affermazioni dietro le quali si scorge il contributo del calcolo delle probabilità
   - *"Deve fare un esame di screening"*
   - *"Quanto spendi per assicurare la tua automobile?"*
   - *"Ti è arrivata la mia email?"*
   - *"Lo vedo in forma. Vedrai che vince"*
   - ...
3. In tutte queste circostanze la domanda o l'affermazione presuppone la valutazione della possibilità che **un evento aleatorio si verifichi** in un tempo futuro

1. Il calcolo delle probabilità è stata una delle più grandi scoperte della matematica. La conoscenza del calcolo delle probabilità influenza la nostra vita quotidiana molto di più di quanto si possa immaginare.
2. Ecco alcuni semplici affermazioni dietro le quali si scorge il contributo del calcolo delle probabilità
   - *"Deve fare un esame di screening"*
   - *"Quanto spendi per assicurare la tua automobile?"*
   - *"Ti è arrivata la mia email?"*
   - *"Lo vedo in forma. Vedrai che vince"*
   - ...
3. In tutte queste circostanze la domanda o l'affermazione presuppone la valutazione della possibilità che **un evento aleatorio si verifichi** in un tempo futuro

"*La teoria della probabilità è, al fondo, semplice buon senso tradotto in calcolo... È degno di nota che questa scienza, nata al servizio dei giochi d'azzardo, sia diventata il più importante oggetto della conoscenza umana... Le più importanti questioni della vita sono, per la maggior parte, solo dei problemi di probabilità*"

Laplace

Prova

Esperimento aleatorio dall'*esito* incerto

Esito

Risultato di una prova, l'esito corrisponde all'evento che si è verificato

Probabilità

Misura dell'attitudine di un evento a manifestarsi.

Concetti primitivi

Prova

Esperimento aleatorio dall'*esito* incerto

Esito

Risultato di una prova, l'esito corrisponde all'evento che si è verificato

Probabilità

Misura dell'attitudine di un evento a manifestarsi.

Prova

Esperimento aleatorio dall'*esito* incerto

Esito

Risultato di una prova, l'esito corrisponde all'evento che si è verificato

Probabilità

Misura dell'attitudine di un evento a manifestarsi.

Spazio campionario

Consideriamo un esperimento aleatorio, dall'esito incerto. Assumiamo che il risultato del singolo esperimento non possa essere previsto con esattezza, ma si conosce l'insieme degli eventi, o dei possibili esiti, $\{E_1, E_2, \ldots, E_n\}$, allora l'insieme di tutti i gli eventi possibili prende il nome di **spazio campionario** e lo indicheremo con Ω, se gli eventi sono necessari e incompatibili.

Definizioni:

necessari: gli eventi $\{E_1, E_2, \ldots, E_n\}$ si dicono necessari se rappresentano l'insieme esaustivo di tutti gli esiti possibili della prova. In altre parole uno di essi si dovrà verificare *necessariamente*.

incompatibili: il verificarsi di un evento esclude la possibilità che se ne verifichi anche un altro.

Esempio

Consideriamo un l'esperimento che prevede come prova il lancio di un dato (non truccato o comunque ben bilanciato), l'insieme $\Omega = \{1,2,3,4,5,6\}$ è uno spazio campionario poiché:

- il presentarsi di una faccia esclude la possibilità che se ne presenti un'altra;
- è impossibile che venga estratto un numero diverso da quelli indicati sulle sei facce del dado.

Esiti ed eventi

- Abbiamo indicato con esito l'evento che si è verificato fra gli eventi possibili. L'insieme degli eventi possibili costituisce l'insieme Ω.
- Riprendiamo l'esempio del lancio del dado. È possibile che la nostra attenzione sia rivolta non al singolo evento, ma a un insieme di eventi, per esempio siamo interessati a valutare come risultato della prova che esca un numero pari. Combinazioni di eventi determinano *partizioni dello spazio campionario* e definiscono eventi. L'evento numero pari sarà definito dalla unione degli eventi $\mathsf{PARI} = \{2 \cup 4 \cup 6\}$.
- Il simbolo $\cup$ si legge unione.
- Analogamente possiamo definire l'evento numero minore o uguale a 3: $\leq 3 = \{1 \cup 2 \cup 3\}$

Esiti ed eventi

evento favorevole l'evento a cui rivolgiamo la nostra attenzione prende il nome di evento favorevole.

evento contrario l'insieme degli eventi che non appartengono all'insieme dell'evento favorevole costituiscono l'evento contrario.

- Indicheremo con E il generico evento favorevole e con $\overline{E}$ il generico evento contrario.
- Si ricava che:

$$\begin{aligned} \{E \cup \overline{E}\} &= \Omega \\ \{E \cap \overline{E}\} &= \emptyset \end{aligned}$$

- dove l'operatore $\cap$ si legge "e" e indica l'intersezione, cioè indica l'insieme che contiene E e $\overline{E}$, cioè sia E che $\overline{E}$
- ciò è impossibile e quindi la intersezione dell'evento favorevole e del contrario è l'insieme vuoto $\emptyset$, che non contiene alcun evento.

Notazioni

- Ω Spazio campionario o spazio degli eventi.
- E_i Evento generico (sottoinsieme dello spazio campionario). Un esito può corrispondere ad un singolo evento o ad una unione di eventi.
- $\overline{E}$ Evento contrario. L'evento contrario include tutti gli eventi che non costituiscono l'evento favorevole E.
- $E_1 \cup E_2$ Unione di eventi. L'evento $E_{1,2} = (E_1 \cup E_2)$ contiene è un evento definito dalla unione die E_1 e E_2 (in un unico evento). L'esito che corrisponde al verificarsi di $E_{1,2}$ è che si verifichi E_1 o E_2. (cioè uno qualsiasi degli eventi che costituiscono $E_{1,2}$. Se ne deduce da ciò che: $\{E \cup \overline{E}\} = \Omega$.
- $E_1 \cap E_2$ Intersezione di eventi. Il verificarsi dell'evento $(E_1 \cap E_2)$, invece, presuppone il verificarsi di E_1 e E_2.
- $\emptyset$ Insieme vuoto. È vuoto l'insieme costituito dall'intersezione $(E \cap \overline{E})$.

Assiomi fondamentali della Probabilità

Una funzione P che assegna un numero reale $P(E_i)$ a ciascun evento $\{E_1, E_2, \ldots, E_n\}$ di Ω è una distribuzione di probabilità se soddisfa i tre seguenti assiomi (assiomi fondamentali del calcolo delle probabilità):

Assioma 1: $P(E_i) \geq 0$ per qualsiasi E_i che appartiene ad Ω.

Assioma 2: $P(\Omega) = \max$. Ω rappresenta l'evento certo e quindi la sua probabilità è massima. Per convenzione si ha che $\max(P) = 1$ e quindi $P(\Omega) = 1$.

Assioma 3: Siano $\{E_1, E_2, \ldots, E_n\}$ eventi disgiunti, non hanno esiti in comune, allora:

$$P(E_i \cup E_j) = P(E_i) + P(E_j), \qquad \text{con } i \neq j \text{ e}$$

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$$

Teoremi fondamentali del Calcolo delle Probabilità

Teorema 1 $P(E_c) = 1 - P(\overline{E}_c)$

Prova: se E_c e $\overline{E}_c$ sono eventi disgiunti tali che $E_c \cup \overline{E}_c = \Omega$ (E_c definisce una partizione di Ω), per gli assiomi della probabilità 2 e 3, abbiamo $1 = P(\Omega) = P(E_c \cup \overline{E}_c) = P(E_c) + P(\overline{E}_c)$.

Teorema 2 $P(E_a \cup E_b) = P(E_a) + P(E_b) - P(E_a \cap E_b)$

Prova: Rappresentiamo $E_a \cup E_b$ come l'unione di due eventi disgiunti: E_a e $(\overline{E}_a \cap E_b)$, da cui $E_a \cup E_b = E_a \cup (\overline{E}_a \cap E_b)$. L'evento E_b può essere rappresentato come $E_b = (E_a \cap E_b) \cup (\overline{E}_a \cap E_b)$.
Da cui:

$$\begin{aligned} P(E_a \cup E_b) &= P(E_a) + P(\overline{E}_a \cap E_b) \\ P(E_b) &= P(E_a \cap E_b) + P(\overline{E}_a \cap E_b). \end{aligned}$$

Teoremi fondamentali del Calcolo delle Probabilità

Teorema 1 $P(E_c) = 1 - P(\overline{E}_c)$

Prova: se E_c e $\overline{E}_c$ sono eventi disgiunti tali che $E_c \cup \overline{E}_c = \Omega$ (E_c definisce una partizione di Ω), per gli assiomi della probabilità 2 e 3, abbiamo $1 = P(\Omega) = P(E_c \cup \overline{E}_c) = P(E_c) + P(\overline{E}_c)$.■

Teorema 2 $P(E_a \cup E_b) = P(E_a) + P(E_b) - P(E_a \cap E_b)$

Prova: Rappresentiamo $E_a \cup E_b$ come l'unione di due eventi disgiunti: E_a e $(\overline{E}_a \cap E_b)$, da cui $E_a \cup E_b = E_a \cup (\overline{E}_a \cap E_b)$. L'evento E_b può essere rappresentato come $E_b = (E_a \cap E_b) \cup (\overline{E}_a \cap E_b)$.
Da cui:

$$\begin{aligned} P(E_a \cup E_b) &= P(E_a) + P(\overline{E}_a \cap E_b) \\ P(E_b) &= P(E_a \cap E_b) + P(\overline{E}_a \cap E_b), \end{aligned}$$

pertanto $P(E_a \cup E_b) = P(E_a) + P(E_b) - P(E_a \cap E_b)$.□

Teoremi fondamentali del Calcolo delle Probabilità

Teorema 1 $P(E_c) = 1 - P(\overline{E}_c)$

Prova: se E_c e $\overline{E}_c$ sono eventi disgiunti tali che $E_c \cup \overline{E}_c = \Omega$ (E_c definisce una partizione di Ω), per gli assiomi della probabilità 2 e 3, abbiamo $1 = P(\Omega) = P(E_c \cup \overline{E}_c) = P(E_c) + P(\overline{E}_c)$.■

Teorema 2 $P(E_a \cup E_b) = P(E_a) + P(E_b) - P(E_a \cap E_b)$

Prova: Rappresentiamo $E_a \cup E_b$ come l'unione di due eventi disgiunti: E_a e $(\overline{E}_a \cap E_b)$, da cui $E_a \cup E_b = E_a \cup (\overline{E}_a \cap E_b)$. L'evento E_b può essere rappresentato come $E_b = (E_a \cap E_b) \cup (\overline{E}_a \cap E_b)$.
Da cui:

$$\begin{aligned} P(E_a \cup E_b) &= P(E_a) + P(\overline{E}_a \cap E_b) \\ P(E_b) &= P(E_a \cap E_b) + P(\overline{E}_a \cap E_b), \end{aligned}$$

pertanto $P(E_a \cup E_b) = P(E_a) + P(E_b) - P(E_a \cap E_b)$■

$P(\emptyset) = 0$

Prova: ...

Teoremi fondamentali del Calcolo delle Probabilità

Teorema 1 $P(E_c) = 1 - P(\overline{E}_c)$

Prova: se E_c e $\overline{E}_c$ sono eventi disgiunti tali che $E_c \cup \overline{E}_c = \Omega$ (E_c definisce una partizione di Ω), per gli assiomi della probabilità 2 e 3, abbiamo $1 = P(\Omega) = P(E_c \cup \overline{E}_c) = P(E_c) + P(\overline{E}_c)$.■

Teorema 2 $P(E_a \cup E_b) = P(E_a) + P(E_b) - P(E_a \cap E_b)$

Prova: Rappresentiamo $E_a \cup E_b$ come l'unione di due eventi disgiunti: E_a e $(\overline{E}_a \cap E_b)$, da cui $E_a \cup E_b = E_a \cup (\overline{E}_a \cap E_b)$. L'evento E_b può essere rappresentato come $E_b = (E_a \cap E_b) \cup (\overline{E}_a \cap E_b)$.
Da cui:

$$\begin{aligned} P(E_a \cup E_b) &= P(E_a) + P(\overline{E}_a \cap E_b) \\ P(\overline{E}_a \cap E_b) &= P(E_b) - P(E_a \cap E_b), \end{aligned}$$

pertanto $P(E_a \cup E_b) = P(E_a) + P(E_b) - P(E_a \cap E_b)$■

Teorema 3 $P(\varnothing) = 0$

Prova: ...

Se $E_b \subset E_a \longrightarrow P(E_b) \leq P(E_a)$

Prova: ...

Teoremi fondamentali del Calcolo delle Probabilità

Teorema 1 $P(E_c) = 1 - P(\overline{E}_c)$

Prova: se E_c e $\overline{E}_c$ sono eventi disgiunti tali che $E_c \cup \overline{E}_c = \Omega$ (E_c definisce una partizione di Ω), per gli assiomi della probabilità 2 e 3, abbiamo $1 = P(\Omega) = P(E_c \cup \overline{E}_c) = P(E_c) + P(\overline{E}_c)$.■

Teorema 2 $P(E_a \cup E_b) = P(E_a) + P(E_b) - P(E_a \cap E_b)$

Prova: Rappresentiamo $E_a \cup E_b$ come l'unione di due eventi disgiunti: E_a e $(\overline{E}_a \cap E_b)$, da cui $E_a \cup E_b = E_a \cup (\overline{E}_a \cap E_b)$. L'evento E_b può essere rappresentato come $E_b = (E_a \cap E_b) \cup (\overline{E}_a \cap E_b)$.
Da cui:

$$\begin{aligned} P(E_a \cup E_b) &= P(E_a) + P(\overline{E}_a \cap E_b) \\ P(E_b) &= P(E_a \cap E_b) + P(\overline{E}_a \cap E_b), \end{aligned}$$

pertanto $P(E_a \cup E_b) = P(E_a) + P(E_b) - P(E_a \cap E_b)$■

Teorema 3 $P(\varnothing) = 0$

Prova: ...

Teorema 4 Se $E_b \subset E_a \longrightarrow P(E_b) \leq P(E_a)$

Prova: ...

Teoremi fondamentali del Calcolo delle Probabilità

Teorema 1 $P(E_c) = 1 - P(\overline{E}_c)$

Prova: se E_c e $\overline{E}_c$ sono eventi disgiunti tali che $E_c \cup \overline{E}_c = \Omega$ (E_c definisce una partizione di Ω), per gli assiomi della probabilità 2 e 3, abbiamo $1 = P(\Omega) = P(E_c \cup \overline{E}_c) = P(E_c) + P(\overline{E}_c)$.■

Teorema 2 $P(E_a \cup E_b) = P(E_a) + P(E_b) - P(E_a \cap E_b)$

Prova: Rappresentiamo $E_a \cup E_b$ come l'unione di due eventi disgiunti: E_a e $(\overline{E}_a \cap E_b)$, da cui $E_a \cup E_b = E_a \cup (\overline{E}_a \cap E_b)$. L'evento E_b può essere rappresentato come $E_b = (E_a \cap E_b) \cup (\overline{E}_a \cap E_b)$.
Da cui:

$$\begin{aligned} P(E_a \cup E_b) &= P(E_a) + P(\overline{E}_a \cap E_b) \\ P(E_b) &= P(E_a \cap E_b) + P(\overline{E}_a \cap E_b), \end{aligned}$$

pertanto $P(E_a \cup E_b) = P(E_a) + P(E_b) - P(E_a \cap E_b)$■

Teorema 3 $P(\varnothing) = 0$

Prova: ...

Teorema 4 Se $E_b \subset E_a \longrightarrow P(E_b) \leq P(E_a)$

Prova: ...

Teoremi fondamentali del Calcolo delle Probabilità

Teorema 1 $P(E_c) = 1 - P(\overline{E}_c)$

Prova: se E_c e $\overline{E}_c$ sono eventi disgiunti tali che $E_c \cup \overline{E}_c = \Omega$ (E_c definisce una partizione di Ω), per gli assiomi della probabilità 2 e 3, abbiamo $1 = P(\Omega) = P(E_c \cup \overline{E}_c) = P(E_c) + P(\overline{E}_c)$.■

Teorema 2 $P(E_a \cup E_b) = P(E_a) + P(E_b) - P(E_a \cap E_b)$

Prova: Rappresentiamo $E_a \cup E_b$ come l'unione di due eventi disgiunti: E_a e $(\overline{E}_a \cap E_b)$, da cui $E_a \cup E_b = E_a \cup (\overline{E}_a \cap E_b)$. L'evento E_b può essere rappresentato come $E_b = (E_a \cap E_b) \cup (\overline{E}_a \cap E_b)$.
Da cui:

$$\begin{aligned} P(E_a \cup E_b) &= P(E_a) + P(\overline{E}_a \cap E_b) \\ P(E_b) &= P(E_a \cap E_b) + P(\overline{E}_a \cap E_b), \end{aligned}$$

pertanto $P(E_a \cup E_b) = P(E_a) + P(E_b) - P(E_a \cap E_b)$■

Teorema 3 $P(\varnothing) = 0$

Prova: ...

Teorema 4 Se $E_b \subset E_a \longrightarrow P(E_b) \leq P(E_a)$

Prova: ...

Soggettivisti e Frequentisti

Esiste un lungo dibattito sulla valutazione della quantità $P(E_i)$. In una estrema semplificazione (ma veramente estrema!!) la questione può essere banalizzata in questi termini.

- I soggettivisti (detti anche Bayesiani, in quanto ispirati al pensiero di Thomas Bayes) sostengono che è possibile utilizzare tutta la nostra conoscenza relativa all'esperimento aleatorio per determinare una regola che possa valutare $P(E_i)$.
- I frequentasti sostengono che la nostra informazione riguardo $P(E_i)$ si ricava dalla osservazione empirica. Solo ripetendo un numero (infinito) molto grande di volte l'esperimento, il rapporto fra i numero di casi favorevoli (casi in cui si è osservato l'evento E_i) e il numero totale delle prove si potrà determinare $P(E_i)$.

Soggettivisti e Frequentisti

Riprendiamo l'esempio del dado (ben bilanciato),

Soggettivisti: Non è necessario ripetere l'esperimento un gran (infinito) numero di volte per stabilire che $P(⚂) = \frac{1}{6}$. In un dado ben bilanciato tutte le facce hanno la medesima attitudine a presentarsi.

Frequentisti: Solo ripetendo numerose volte l'esperimento sarà possibile determinare la frequenza con cui ciascuna faccia si presenta. La probabilità si determina quando il rapporto fra casi favorevoli e numero di prove si stabilizza attorno ad un valore.

Eventi Indipendenti

Definizione

Due eventi A e B si dicono indipendenti se possiamo scrivere che

$$P(A \cap B) = P(A) * P(B)$$

La condizione di indipendenza fra eventi può essere determinata in diversi modi:

- Si assume esplicitamente che due eventi sono indipendenti. Lanciando due dadi non c'è ragione di ritenere che l'esito di uno possa influenzare l'esito dell'altro. In tal caso è possibile assumere esplicitamente che gli eventi sono indipendenti.
- Due eventi disgiunti A e B non sono indipendenti se $P(A) * P(B) > 0$ e allo stesso tempo se $P(A \cap B) = P(\emptyset)$ allora $P(A \cap B) = 0$.

Probabilità condizionata

Si assume che $P(B) > 0$ la quantità $P(A \mid B)$ indica la probabilità che si verifichi l'evento A sapendo che (dato che) si è già verificato l'evento B.

Definizione

Se $P(B) > 0$ allora la probabilità condizionata di A dato B è definita come:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

ovvero la probabilità dell'evento intersezione (A e B) diviso la probabilità dell'evento condizionante.

Probabilità condizionata

Esempio

Consideriamo l'esperimento lancio di un dado e definiamo di due seguenti spazi campionari:

- $\Omega_1 = (\text{PARI} \cup \text{DISPARI}) = ((\text{⚁} \cup \text{⚃} \cup \text{⚅}) \bigcup (\text{⚀} \cup \text{⚂} \cup \text{⚄}))$
- $\Omega_2 = (\leq 3 \cup > 3) = ((\text{⚀} \cup \text{⚁} \cup \text{⚂}) \bigcup (\text{⚃} \cup \text{⚄} \cup \text{⚅}))$
- e consideriamo l'evento $P(\text{PARI} \mid > 3)$, i numeri pari e >3 sono il 4 e il 6, cioè 2 su sei, poiché assumiamo che è già noto che il numero è >3, allora lo spazio campionario si riduce solo a $(> 3) = (\text{⚃} \cup \text{⚄} \cup \text{⚅})$ e quindi il 4 e il 6 sono 2 su tre possibili esiti. Si ricava quindi che:

$$P(\text{PARI} \mid > 3) = \frac{P(\text{PARI} \cap > 3)}{P(> 3)} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{\cancel{6}} * \frac{\cancel{6}}{3} = \frac{2}{3}$$

Probabilità condizionata

Se A e B sono indipendenti il condizionamento non ha alcun effetto.

Lancio di due dadi

Assumiamo che la prova consiste nel lancio di due dadi. Il lancio di due dadi ha come spazio campionario:

$$\Omega = \{(\epsdice{1};\epsdice{1}), (\epsdice{1};\epsdice{2}), \cdots, (\epsdice{5};\epsdice{6}), (\epsdice{6};\epsdice{6})\}$$

e i due eventi sono chiaramente indipendenti, pertanto la probabilità che si verifichi una qualsiasi coppia è pari a

$$P(\epsdice{2} \cap \epsdice{3}) = P(\epsdice{2}) * P(\epsdice{3}).$$

Consideriamo il seguente condizionamento: $P(\epsdice{2} \mid \epsdice{3})$

$$P(\epsdice{2} \mid \epsdice{3}) = \frac{P(\epsdice{2} \cap \epsdice{3})}{P(\epsdice{3})} = \frac{P(\epsdice{2}) * \cancel{P(\epsdice{3})}}{\cancel{P(\epsdice{3})}} = P(\epsdice{2})$$

Teorema di Bayes

Legge della probabilità totale

Siano gli eventi $A_1, A_2, \ldots . A_n$ una partizione di Ω_A quindi

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = 1,$$

e sia B un generico evento dello spazio campionario Ω_B tale che $P(B) > 0$, si dimostra facilmente che

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i)$$

infatti basta sostituire il generico termine $P(B \mid A_i)$ con $\frac{P(B \cap A_i)}{P(A_i)}$ ■

Teorema di Bayes

Definizione

Siano gli eventi $A_1, A_2, \ldots . A_n$ una partizione di Ω_A, tali che $P(A_i) > 0 \; \forall i$, se $P(B) > 0$ allora si ha

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}$$

Anche in questo caso la dimostrazione è immediata, basta considerare che:

- $P(B \mid A_i)P(A_i) = P(B \cap A_i)$

  e che
- $P(B) = \sum_{i=1}^{n} P(B \mid A_i)P(A_i)$ per la legge della probabilità totale.

In uno studio sulla relazione fra il grado di istruzione genitori/figli, sono stati estratti indipendentemente due gruppi di 100 soggetti di nati fra il 1980 e il 1982: laureati e non laureati. Hanno dichiarato di avere il padre laureato, rispettivamente, 38 laureati e 15 non laureati.
Sulla base dei dati ISTAT sappiamo che fra i nati negli anni '80-82 il tasso di laureati è del 24,5%. Determinare la probabilità che il figlio sia laureato, posto che il padre (non) sia laureato.

	Padre Lau	**Padre non Lau**	**Totale**
Lau	38	62	100
non Lau	15	85	100

Passiamo alla tabella delle probabilità empiriche e a margine indichiamo l'informazione aggiuntiva: $P(\text{laurea}) = 0,245$.

	Padre Lau	Padre non Lau	P(laurea)
Lau	0,38	0,62	0,245
non Lau	0,15	0,85	0,755

La probabilità $P(\text{Padre Lau} \mid \text{Lau}) = 0,38$ si legge come: probabilità di aver un padre laureato dato che si possiede la laurea.
Allo stesso modo, $P(\text{Padre non Lau} \mid \text{non Lau}) = 0,85$ e si legge come la probabilità di avere un padre non laureato dato che non si possiede la laurea.

Utilizzando l'informazione aggiuntiva, relativa alla probabilità di essere laureati nell'intera popolazione di riferimento, e ricordando che:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

possiamo ricavare le probabilità relative alle intersezioni come: $P(A \cap B) = P(A \mid B) * P(B)$, quindi:

$$P(\text{Padre Lau} \cap \text{Lau}) = P(\text{Padre Lau} \mid \text{Lau}) * P(\text{Lau})$$
$$P(\text{Padre Lau} \cap \text{Lau}) = 0,38 * 0,245 = 0,093$$

La distribuzione congiunta sarà quindi:

	Padre Lau	Padre non Lau	Totale
Lau	0,093	0,152	0,245
non Lau	0,113	0,642	0,755
	0,206	0,794	1,000

A questo punto punto è possibile calcolare la distribuzione marginale relativa alla Laurea Padre.
La probabilità di estrarre un soggetto che ha la laurea, dato che il padre sia laureato, si determina attraverso il seguente condizionamento:

$$P(\text{Lau} \mid \text{Padre Lau}) = \frac{0,093}{0,206} = 0,451.$$

E' interessante notare come questa probabilità sia molto più grande di $P(\text{Lau} \mid \text{Padre non Lau}) = 0,191$

Variabili casuali (v.c.)

Una variabile casuale è una applicazione di Ω in $\mathbb{R}$

$$X : \Omega \mapsto \mathbb{R}$$

Dove $\mathbb{R}$ indica l'insieme dei numeri reali.

- Una variabile casuale non è altro che una regola che assegna a ciascun elemento dello spazio campionario un elemento di $\mathbb{R}$.
- Se Ω è un insieme (al più) numerabile allora $X(E_i)$ è una v.c. discreta (o categorica), per $E_i \in \Omega$.
- Se Ω è un insieme non numerabile allora $X(E_i)$ è una v.c. continua, per $E_i \in \Omega$ e $E_i \subset \mathbb{R}$.
- Poiché l'insieme $\mathbb{R}$ è per definizione un insieme ordinato, la funzione delle probabilità cumulate (o di ripartizione) associata alla v.c. X è una funzione definita in $\mathbb{R}$ tale che $F(x) \to [0;1]$, $\forall x \in X$.

Definizione

Sia X una variabile casuale discreta, si definisce funzione di distribuzione delle probabilità per X la funzione che assegna a ciascun valore della v.c. X un probabilità: $f(x_i) = \mathrm{P}(X = x_i)$.

Dalla precedente definizione si ricava che:

- $f(x_i) \geq 0$ per un insieme al più numerabile di $x_1, x_2, \ldots, x_n$;
- $f(x_i) = 0$ altrove;
- $f(x_i) = P(X = x_i) \equiv P(E_i)$, con $i = 1, 2, \ldots, n$;
- $\sum_{i=1}^{n} f(x_i) = 1$.

Definizione

Sia X una variabile casuale continua, si definisce funzione di densità delle probabilità la funzione $f(x)$ tale che:

- $f(x) \geq 0 \ \forall x$
- $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$
- $\mathrm{P}(a < X < b) = \int_{a}^{b} f(x)\mathrm{d}x$, con $a \leq b$

Dalla precedente definizione si ricava che:

- $F(X = a) = \int_{-\infty}^{a} f(x)\mathrm{d}x = \mathrm{P}(X \leq a)$;
- $f(x) = F'(X)$ in tutti i punti in cui $F(x)$ è differenziabile.

  la funzione di densità di probabilità equivale alla derivata prima della funzione di ripartizione $F(x)$

Valore atteso e Varianza di una *v.c.*

Il *valore atteso*[1] di una variabile casuale può essere interpretato come una sorta di media della v.c., ed in certi casi lo è. Nel caso di v.c. discrete, il valore atteso non è altro che la sommatoria dei prodotti di ciascuna $X = x_i$ ponderata per la probabilità corrispondente $P(X = x_i)$. Nel caso continuo, non potendo utilizzare l'operatore di sommatoria dobbiamo ricorrere al calcolo integrale, ma il concetto resta lo stesso.

Caso continuo

$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$$

Caso discreto

$$\mu = E[X] = \sum_{j=1}^{N} x_j P(X = x_j)$$

[1] Il valore atteso è detto anche *Speranza Matematica*

Il *valore atteso*[1] di una variabile casuale può essere interpretato come una sorta di media della v.c., ed in certi casi lo è. Nel caso di v.c. discrete, il valore atteso non è altro che la sommatoria dei prodotti di ciascuna $X = x_i$ ponderata per la probabilità corrispondente $P(X = x_i)$. Nel caso continuo, non potendo utilizzare l'operatore di sommatoria dobbiamo ricorrere al calcolo integrale, ma il concetto resta lo stesso.

Caso continuo

$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$$

Caso discreto

$$\mu = E[X] = \sum_{j=1}^{N} x_j P(X = x_j)$$

[1] Il valore atteso è detto anche *Speranza Matematica*

Il *valore atteso*[1] di una variabile casuale può essere interpretato come una sorta di media della v.c., ed in certi casi lo è. Nel caso di v.c. discrete, il valore atteso non è altro che la sommatoria dei prodotti di ciascuna $X = x_i$ ponderata per la probabilità corrispondente $P(X = x_i)$. Nel caso continuo, non potendo utilizzare l'operatore di sommatoria dobbiamo ricorrere al calcolo integrale, ma il concetto resta lo stesso.

Caso continuo

$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$$

Caso discreto

$$\mu = E[X] = \sum_{j=1}^{N} x_j P(X = x_j)$$

[1] Il valore atteso è detto anche *Speranza Matematica*

La varianza è una misura di variabilità definita come la *media* degli scarti al quadrato rispetto alla media:

Caso continuo

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu^2) f(x) \mathrm{d}x$$

Con poca difficoltà si può dimostrare che $E[(X-\mu)^2] = E[X^2] - \mu^2$

Caso discreto

$$\sigma^2 = E[(X-\mu)^2] = \sum (x_i-\mu)^2 P(X=x_i)$$

La varianza è una misura di variabilità definita come la *media* degli scarti al quadrato rispetto alla media:

Caso continuo

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu^2) f(x) \mathrm{d}x$$

Con poca difficoltà si può dimostrare che $E[(X-\mu)^2] = E[X^2] - \mu^2$

Caso discreto

$$\sigma^2 = E[(X-\mu)^2] = \sum_{j=1}^{N} (x_j-\mu)^2 \mathrm{P}(X = x_j)$$

Valore atteso e Varianza di una *v.c.*

La varianza è una misura di variabilità definita come la *media* degli scarti al quadrato rispetto alla media:

Caso continuo

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu^2) f(x) \mathrm{d}x$$

Con poca difficoltà si può dimostrare che $E[(X-\mu)^2] = E[X^2] - \mu^2$

Caso discreto

$$\sigma^2 = E[(X-\mu)^2] = \sum_{j=1}^{N} (x_j - \mu)^2 \mathrm{P}(X = x_j)$$

Valore atteso e Varianza di una *v.c.*

La varianza è una misura di variabilità definita come la *media* degli scarti al quadrato rispetto alla media:

Caso continuo

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu^2) f(x) \mathrm{d}x$$

Con poca difficoltà si può dimostrare che $E[(X-\mu)^2] = E[X^2] - \mu^2$

Caso discreto

$$\sigma^2 = E[(X-\mu)^2] = \sum_{j=1}^{N} (x_j-\mu)^2 \mathrm{P}(X = x_j)$$

La varianza è una misura di variabilità definita come la *media* degli scarti al quadrato rispetto alla media:

Caso continuo

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu^2) f(x) \mathrm{d}x$$

Con poca difficoltà si può dimostrare che $E[(X-\mu)^2] = E[X^2] - \mu^2$

Caso discreto

$$\sigma^2 = E[(X-\mu)^2] = \sum_{j=1}^{N} (x_j-\mu)^2 \mathrm{P}(X = x_j)$$

Momenti *v.c.*

Sia X una generica variabile casuale, la quantità

$$m^r = \int_{-\infty}^{\infty} x^r f(x)\mathrm{d}x$$

si definisce momento di ordine r. Analogamente, la quantità $m^r = \int_{-\infty}^{\infty}(x-\mu)^r f(x)\mathrm{d}x$ si definisce momento centrato di ordine r.

Se ne deduce che:

- la media è il momento di ordine 1
- la varianza è il momento centrato di ordine 2 e che può essere definita come la differenza fra il momento di ordine 2 ($\int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x$) ed il quadrato del momento di ordine 1:

  $$\sigma^2 = \int_{-\infty}^{\infty} (x)^2 f(x)\mathrm{d}x - \mu^2$$

  o analogamente

  $$\sigma^2 = E[X^2] - \mu^2$$

Sia Ω uno spazio campionario finito ed *equiprobabile* (tutti gli eventi hanno eguale probabilità), la v.c. X associata ad Ω prende il nome di variabile casuale Uniforme discreta e sarà indicata con la notazione $X \sim U\{\min, \max\}$.

Esempio

L'esempio più tipico di v.c. Uniforme discreta è quello relativo all'esperimento del lancio di un dado.
La v.c. X associata all'esperimento lancio di un dado è definita come $X \sim U\{1,6\}$ e la corrispondente funzione di distribuzione delle probabilità è:

$$\begin{aligned} f(X = x_i) &= \frac{1}{6}, \qquad \forall x_i \in \{1,6\} \\ f(X) &= 0 \qquad \text{altrove.} \end{aligned} \tag{1}$$

Definizione:

La funzione di ripartizione è la funzione $F_X : \mathbb{R} \to [0,1]$ definita come:

$$F(X = x_i) = P(X \leq x_i)$$

È una funzione monotona e non decrescente tale che:
$\lim_{x \to -\infty} F_X(x) = 0$ e
$\lim_{x \to \infty} F_X(x) = 1$

F_X per $X \sim U\{1,6\}$

$$\begin{aligned} F(X = x_i) &= i * 1/6, \qquad \forall x_i \in \{1,6\} \\ F_X &= 0 \text{ if } X < 1, && (2) \\ F_X &= 1 \text{ if } X \geq 6. && (3) \end{aligned}$$

v.c. di Bernulli

Si consideri un esperimento aleatorio che prevede solo due possibili esiti: favorevole (F) e contrario (C). L'esempio più classico, ma anche il più astratto, è quello del lancio della moneta, che prevede solo due possibili esiti: Testa/Croce.

In realtà, a pensarci bene, durante una qualsiasi giornata si verificano un'enorme quantità di esperimenti bernulliani. Eccone una breve lista: *mi sveglio con/senza il raffreddore*, *arrivo in ufficio in orario/ritardo*, *nasce un bambino maschio/femmina* ecc..

$X \sim \mathrm{B}(\pi)$

Se assegnamo il valore 1 all'evento F e il valore 0 all'evento C e attribuiamo la probabilità π ad F, diremo che $X \sim \mathrm{B}(\pi)$, che si legge: *X si distribuisce secondo v.c. di Bernulli a parametro π* (il simbolo $\sim$ si legge "*si distribuisce*".

$$P(X = 1) = \pi$$
$$P(X = 0) = 1 - \pi,$$

che può essere scritta nella forma

$$f(x) = \pi^x (1-\pi)^{1-x}, \qquad \text{con } x \in \{0, 1\}.$$

Le distribuzioni delle probabilità si possono interpretare utilizzando gli stessi criteri delle distribuzioni delle frequenze relative. In fondo la frequenza relativa può essere considerata una probabilità empirica se n, il numero delle prove, è grande abbastanza. Secondo la concezione frequentista, "*grande abbastanza*" indica che un incremento del numero di prove influenza in modo del tutto trascurabile le probabilità empiriche.

Assumiamo aver eseguito 1000 lanci di una moneta e di aver ottenuto 482 volte testa. Indichiamo con $\hat{p}$ la probabilità empirica associata a testa, pertanto $\hat{p} = 0,4820$.
Se eseguiamo un ulteriore lancio e osserviamo testa (ma sarebbe indifferente se assumessimo croce), la probabilità empirica associata diventa: $(482+1)/1001 = 0,4825$.
Man mano che il denominatore aumenta (cioè aumenta il numero delle prove eseguite) il quoziente fra eventi favorevole e prove si modificherà in misura sempre minore.

Valore atteso e varianza della v.c. di Bernulli

Sia X una generica v.c. di Bernulli a parametro π, indichiamo con $E[X]$ il suo valore atteso che sarà definito come la somma del valore di X moltiplicato per la probabilità dell'evento corrispondente.
Abbiamo detto che:

- $X = 1$ se l'evento è F, mentre $X = 0$ nel caso di C;
- $P(X = 1) = \pi$ e $P(X = 0) = (1 - \pi)$

Valore atteso di X

$$E[X] = 1 * \pi + 0 * (1 - \pi) = \pi$$

Varianza di X

$$\begin{aligned} E[(X-\pi)^2] &= (1-\pi)^2 * \pi + (0-\pi)^2 * (1-\pi) = \\ &= (1 - 2 * \pi + \pi^2) * \pi + \pi^2 * (1-\pi) = \\ &= \pi - 2 * \pi^2 + \cancel{\pi^3} + \pi^2 - \cancel{\pi^3} = \pi * (1-\pi). \end{aligned}$$

Valore atteso e varianza della v.c. di Bernulli

Sia X una generica v.c. di Bernulli a parametro π, indichiamo con $E[X]$ il suo valore atteso che sarà definito come la somma del valore di X moltiplicato per la probabilità dell'evento corrispondente.
Abbiamo detto che:

- $X = 1$ se l'evento è F, mentre $X = 0$ nel caso di C;
- $P(X = 1) = \pi$ e $P(X = 0) = (1 - \pi)$

Valore atteso di X

$$E[X] = 1 * \pi + 0 * (1 - \pi) = \pi$$

Varianza di X

Analogamente, ricordando che la varianza è definibile anche come la differenza fra il momento secondo e il quadrato del momento primo si ha:

$$\begin{aligned} E[(X - \pi)^2] &= (1^2 * \pi + 0^2 * (1 - \pi)) - \pi^2 \\ &= \pi - \pi^2 \\ &= \pi * (1 - \pi). \end{aligned}$$

v.c. Binomiale

Si consideri una serie di n prove bernulliane *identiche* (π costante) e *indipendenti*, per cui possiamo scrivere che la probabilità di avere n eventi favorevoli in n prove sarà

$$P(F \cap F \cap \cdots \cap F) = \pi^n.$$

Se, assumiamo che in n prove bernulliane si possono verificare x eventi favorevoli e $(n-x)$ eventi contrari bisogna considerare tutte le possibili combinazioni degli x eventi favorevoli su n posizioni. Se non si vuol tenere conto dell'ordine, ma solo del numero di eventi favorevoli sulle n prove bisogna valutare il numero di combinazioni "*equivalenti*"

$n = 5$ e $x = 3$

Dodici combinazioni "equivalenti":

11100, 11010, 11001, 10110, 10101, 10011, 01110, 01010, 01011, 00111

Definizione

Si consideri un esperimento aleatorio che consiste in n prove bernulliane *i.i.* (identiche e indipendenti) a parametro π, si dirà che $X \sim Bin(n\pi)$ (X si distribuisce secondo una v.c. Binomiale con parametri n e π) se lo spazio campionario è costituito dal numero di successi che si possono verificare in n prove, senza tenere conto dell'ordine della sequenza.

Il coefficiente binomiale

Per determinare il numero di sequenze bisogna fare ricorso al calcolo combinatorio. Il numero di combinazioni di n elementi presi 'a x a x' è definito dalla seguente espressione

$$\binom{n}{x} = \frac{n!}{(n-x)! * x!}$$

dove $n!$ (si legge: "*n fattoriale*") è dato da

$$n! = n * (n-1) * (n-2) \cdots * 1$$

È molto semplice capire la logica del coefficiente binomiale.

Assumiamo che $n = 5$ e $x = 3$ e indichiamo i tre eventi favorevoli con F_1, F_2, F_3 e i due eventi contrari con C_1 e C_2, ricordiamo che non teniamo conto dell'ordine.

Se le prove sono $n = 5$, allora le combinazioni senza ripetizione sono $5 * 4 * 3 * 2 * 1 = 120$, ma poiché consideriamo *uguali* gli eventi che si differenziano solo per permutazioni fra gli F e i C, *es.* F_1, F_2, F_3, C_1, C_2 e F_2, F_3, F_1, C_2, C_1, bisogna dividere 120 per 12.
Si osservi, infatti, che i tre esiti F si possono combinare fra loro in 6 modi diversi $(3 * 2)$ e i due esiti C in due modi, complessivamente in 12 modi diversi, infatti $(3 * 2 * 1) * (2 * 1) = 12$.

Definizione

La v.c. $X \sim \mathrm{Bin}(n, \pi)$ se $P(X = x)$ esprime la probabilità che su n prove bernulliane si verifichino x eventi favorevoli, indipendentemente dall'ordine con cui si presentano, la sua funzione di distribuzione di probabilità è

$$f(x) = \frac{n!}{x! * (n-x)!} * \pi^x * (1-\pi)^{(n-x)}$$

Valore atteso e varianza

Avendo definito la v.c. Binomiale come una somma di v.c. di Bernulli possiamo ricavare il valore atteso e la varianza come somma del valore atteso (proprietà della media) e della varianza (proprietà della varianza[2]) della Bernulli, pertanto

- $E[X] = n * \pi$
- $E[(X - E[X])^2] = n * \pi * (1-\pi)$

[2] Se le prove sono eseguite sotto la condizione di indipendenza e il parametro π è costante, la varianza totale si può esprimere come somma delle varianze.