Calcolo delle Probabilità

Il calcolo delle probabilità è presupposto essenziale per il processo di inferenza statistica. In realtà il calcolo delle probabilità è una disciplina a sé stante:

- inizialmente sviluppata per lo studio dei giochi d'azzardo
- con applicazioni in numerosi campi della scienza (fisica, genetica, ...)

Definizioni

Esperimento: Insieme di procedure volte a produrre un certo risultato

Esperimento aleatorio o casuale: esperimento il cui esito non può essere predetto con certezza

Spazio campionario o spazio degli eventi: insieme dei risultati possibili di un esperimento casuale. Si indica spesso con S o Ω.

Gli eventi

Un evento è un qualunque sottoinsieme dello spazio campionario.

Alcuni esempi:

- lancio di un dado $S = \{1, 2, 3, 4, 5, 6\}$.
  Alcuni eventi sono $A = \{1, 3\}$, $B = \{5\}$, $\emptyset$, S
- lancio di una moneta $S = \{T, C\}$
- partita di calcio $S = \{1, \times, 2\}$
- nel caso di due lanci successivi di una moneta allora

$$S = \{(T, T), (T, C), (C, T), (C, C)\} = \{T, C\}^2$$

  cioè i risultati possibili sono coppie; se i lanci sono tre saranno terne e così via

Cos'è la probabilità?

Le definizioni di probabilità sono molteplici. Le più rilevanti sono:

- definizione classica
- definizione frequentista
- definizione soggettiva
- definizione assiomatica

Definizione classica

La probabilità di un evento E è data dal rapporto tra:

- numero dei casi favorevoli al verificarsi dell'evento
- numero di casi possibili, purché *ugualmente possibili*

$$P(E) = \frac{\#\text{ casi favorevoli}}{\#\text{ casi possibili}} = \frac{\#F}{\#S}.$$

Conseguenze:

- $0 \leq P(E) \leq 1$
- $P(\varnothing) = 0,\ P(S) = 1$

Esempio

Supponiamo vi siano 3 diverse strade per andare dalla città A alla città B e 5 diverse strade per andare dalla città B alla città C; quante strade diverse si possono percorrere per andare da A a C passando per B?

Esempio

Supponiamo vi siano 3 diverse strade per andare dalla città A alla città B e 5 diverse strade per andare dalla città B alla città C; quante strade diverse si possono percorrere per andare da A a C passando per B?

Indichiamo con

- S_3 *l'insieme delle strade che vanno da* A *a* B,
- S_5 *l'insieme delle strade che vanno da* B *a* C.

Allora i risultati possibili sono gli elementi di

$$S = S_3 \times S_5$$

e si ha

$$\#S = \#S_3 \cdot \#S_5 = 3 \cdot 5 = 15.$$

Esempio

Se viene lanciata una moneta per 7 volte, quanti sono i possibili risultati?

Casi possibili

Esempio

Se viene lanciata una moneta per 7 volte, quanti sono i possibili risultati?

L'insieme dei risultati possibili è

$$S = \{T, C\}^7$$

Quindi

$$\#S = 2^7$$

Per fare i conti è comodo utilizzare il

modello dell'urna

Si considera un insieme U (detto "urna") contenente n elementi e si fanno k etrazioni successive.

Modalità di estrazione

Le estrazioni successive possono essere fatte con le seguenti modalità:

- **senza ripetizione**, cioè senza rimettere nell'urna l'elemento estratto,
- **con ripetizione**, cioè rimettendo ogni volta nell'urna l'elemento estratto.

Disposizioni e combinazioni

Le estrazioni successive vengono dette

- **disposizioni** se gli elementi estratti sono "disposti" nell'ordine in cui vengono estratti
- **combinazioni** se l'ordine è irrilevante

Attenzione: le "combinazioni" delle casseforti sono in realtà disposizioni

Contiamo, nei vari casi, quanti sono i risultati possibili.

Disposizioni senza ripetizione di k elementi su un insieme di n

Effettuiamo k estrazioni senza rimettere ogni volta nell'urna l'elemento estratto.
Osserviamo che i possibili risultati sono

- n alla prima estrazione
- $n-1$ alla seconda
- ...
- $n-k+1$ alla k-esima

I risultati possibili sono gli elementi del prodotto cartesiano

$$S_n \times S_{n-1} \times \cdots \times S_{n-k+1} \text{ con } \#S_i = i.$$

e il loro numero è quindi

$$D_{k,n} := n(n-1)(n-2)\cdots(n-k+1)$$

Nel caso in cui $\mathbf{k = n}$ si parla di **permutazioni** ed il numero corrispondente è

$$D_{n,n} = n(n-1)(n-2)\cdots 3\cdot 2\cdot 1 = n!$$

È inoltre facile verificare che

$$D_{k,n} = \frac{n!}{(n-k)!}.$$

Esempio

Supponiamo di dover scegliere un presidente ed un segretario di una commissione di 10 membri. Quante sono le possibili scelte?

Esempio

Supponiamo di dover scegliere un presidente ed un segretario di una commissione di 10 membri. Quante sono le possibili scelte?

Si tratta di disposizioni senza ripetizione di 2 *oggetti su* 10

$$D_{2,10} = 10 \cdot 9 = 90.$$

Esempio

5 persone si dispongono allineate per fare un fotografia. Quante diverse fotografie possono essere fatte?

Esempio

5 persone si dispongono allineate per fare un fotografia. Quante diverse fotografie possono essere fatte?

Si tratta di contare le permutazioni di un insieme di 5 elementi, che sono $5! = 120$.

Disposizioni con ripetizione

Effettuiamo k estrazioni **rimettendo ogni volta** nell'urna l'elemento estratto. Osserviamo che i possibili risultati sono

- n alla prima estrazione
- n alla seconda
- ...
- n alla k-esima

Allora i k oggetti possono essere scelti in

$$D^r_{k,n} := \# S^k_n = n^k \quad \text{modi}$$

Esempio

Quante parole di 5 lettere si possono scrivere con le 21 lettere dell'alfabeto, indipendentemente dal loro significato?

Esempio

Quante parole di 5 lettere si possono scrivere con le 21 lettere dell'alfabeto, indipendentemente dal loro significato?

Si tratta di disposizioni con ripetizione di 5 *elementi su un insieme di* 21. *Sono quindi* 21^5.

Combinazioni senza ripetizione

Le **combinazioni senza ripetizione**, in cui il risultato non dipende dall'ordine di estrazione sono date da

$$C_{k,n} = \frac{D_{k,n}}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

Combinazioni senza ripetizione

Esempio

In quanti modi è possibile pescare 2 carte da un mazzo di 27?

Esempio

In quanti modi è possibile pescare 2 carte da un mazzo di 27?

Siccome l'ordine in cui le carte vengono pescate non ha importanza, e non può esserci ripetizione, si tratta di combinazioni senza ripetizione di 2 *elementi di un insieme di* 27 *che sono*

$$\binom{27}{2} = \frac{27!}{2!25!} = \frac{27 \cdot 26}{2} = 351.$$

Combinazioni con ripetizione

Sono date da

$$C^r_{k,n} := \binom{n+k-1}{k} = \binom{n+k-1}{n-1}.$$

Dimostrare per esercizio che vale la seconda uguaglianza.

Critiche alla definizione classica

- **di ordine teorico:** la definizione è circolare (ugualmente possibili significa ugualmente probabili)
- **di ordine pratico:** non sempre è possibile enumerare tutti i casi possibili, oppure i casi possibili non sono ugualmente possibili

Definizione assiomatica

Gli assiomi del calcolo delle probabilità sono i seguenti.

La probabilità è una funzione $P : \wp(S) \to [0,1]$ tale che

1. $P(S) = 1$ (cioè la probabilità dell'evento certo è pari a 1);
2. $A \cap B = \varnothing \implies P(A \cup B) = P(A) + P(B)$ (cioè la probabilità è una funzione additiva).

Definizione assiomatica

La definizione assiomatica

- stabilisce alcune regole (di carattere logico-formale) alle quali la probabilità deve sottostare
- è (quasi) universalmente accettata e condivisa
- non dà indicazioni su come assegnare probabilità agli eventi (vediamo con un esempio come questo dipenda dal contesto)

Esempio - lancio di una moneta

Spazio degli eventi: $S = \{T, C\}$.

Assegniamo la probabilità che esca testa: $P(\{T\}) := p$, $p \in [0, 1]$
($p = 1/2$ se la moneta non è truccata)

Osserviamo che
1. $\implies P(\{T, C\}) = 1$
2. $\implies P(\{C\}) = P(\{T, C\}) - P(\{T\}) = 1 - p$
e quindi la funzione P risulta completamente determinata.

Esistono quindi infinite probabilità che soddifano gli assiomi *1.* e *2.*, una per ciascun valore di $p \in [0, 1]$.

Teorema delle probabilità totali

Dati due eventi A e B comunque scelti

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Attenzione: il teorema si differenzia dal terzo assioma in quanto gli eventi non sono necessariamente disgiunti.

DIMOSTRAZIONE. Osservato che $A \cup B$ è unione disgiunta degli insiemi $A \setminus B$, $A \cap B$ e $B \setminus A$, per la additività si ha

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

Sempre per l'additività si ha

$$P(A) = P(A \setminus B) + P(A \cap B), \quad P(B) = P(B \setminus A) + P(A \cap B)$$

e la tesi si ottiene sostituendo. $\square$

Condizionamento

$A|B$ si legge "A condizionato (o dato) B"

Si suppone di aver osservato il verificarsi di B e ci si chiede se ed in quale misura questa informazione modifichi la valutazione di probabilità su A.

In generale

$$P(A|B) \neq P(A)$$

Esempio - doppio lancio di una moneta

Supponiamo di lanciare una moneta due volte, e che in ciascun lancio testa e croce abbiano la stessa probabilità di uscire. Lo spazio degli eventi è

$$S = \{(T,T),(T,C),(C,T),(C,C)\}$$

La probabilità che esca testa in entrambi i lanci è

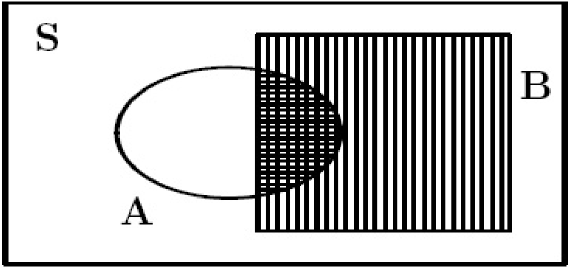
$$P(\{(T,T)\}) = \frac{\#\{(T,T)\}}{\#\{(T,T),(T,C),(C,T),(C,C)\}} = \frac{1}{4}$$

Se sappiamo che nel primo lancio esce testa allora la probabilità che esca testa in entrambi i lanci diventa

$$P(\{(T,T)\}) = \frac{\#\{(T,T)\}}{\#\{(T,T),(T,C)\}} = \frac{1}{2}$$

Condizionamento

Il condizionamento consiste in una ridefinizione dello spazio campionario che si riduce da S a B (nell'esempio del lancio della moneta si ha $B = \{\{(T, T), (T, C)\}\}$)



Condizionamento

Definizione di probabilità condizionata

Sia $P(B) > 0$. La probabilità di un evento A condizionata al verificarsi di B si definisce nel modo seguente

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

Si riproporziona la probabilità di A in funzione della riduzione dello spazio campionario. Si osservi che $P(B|B) = 1$ e inoltre se $A \cap C = \varnothing$ allora $P(A \cup C|B) = P(A|B) + P(C|B)$. Dunque $P(\cdot|B)$ è una probabilità su B.

Indipendenza

Definizione

Due eventi A e B si dicono indipendenti se e solo se

$$P(A|B) = P(A)$$

o, equivalentemente, $P(A \cap B) = P(A)P(B)$, ossia se il verificarsi dell'evento B non modifica la valutazione di probabilità su A.

Esempio - doppio lancio di una moneta

Mostriamo che nel lancio doppio di una moneta i risultati di ciascun lancio sono tra loro indipendenti.

Supponiamo che T e C siano equiprobabili.
Spazio degli eventi $S = \{(T, T), (T, C), (C, T), (C, C)\}$
Sia A l'evento "esce testa al primo lancio" e
sia B l'evento "esce testa al secondo lancio", cioè

$$A = \{(T, T), (T, C)\}, \quad B = \{(T, T), (C, T)\}$$

Si ha

$$P(A \cap B) = P(\{(T, T)\}) = \frac{1}{4}$$

$$P(A)P(B) = P(\{(T, T), (T, C)\}) \cdot P(\{(T, T), (C, T)\}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Esempio - Famiglie con 4 figli

In una famiglia di 4 figli, ci si chiede qual'è la probabilità che 1, 2, 3 o tutti i figli siano maschi (considerando equiprobabile la nascita di maschi e femmine).

La popolazione S in tal caso è costituita dall'insieme delle famiglie con 4 figli.

Introduciamo una funzione X definita su S che conta i figli maschi. Data una famiglia x si avrà

$$X(x) = \#\ \text{figli maschi di } x$$

Matematicamente...

Quando l'esito di un esperimento si può rappresentare con un numero X e ad ogni realizzazione dell'esperimento questo numero può assumere valori diversi, allora X prende il nome di *variabile aleatoria o casuale*.

Cioè, una variabile casuale o aleatoria X è una funzione

$$X : S \to \mathbb{R}$$

Sono analoghe alle variabili statistiche e anch'esse si classificano in **discrete** (possono assumere solo un numero finito o una infinità numerabile di valori) e **continue** (possono assumere tutti i valori all'interno di un intervallo).

Funzione di distribuzione di probabilità

Definizione

Sia X una v.c. discreta che può assumere i valori x_1, x_2, ...
Si chiama funzione di distribuzione (o massa) di probabilità la funzione

$$f(x_i) = P(X = x_i) := P(\{s \in S \;:\; X(s) = x_i\}) = P(X^{-1}(\{x_i\}))$$

che ad ogni valore x_i associa la probabilità che la v.c. X assuma il valore x_i.

Probabilità e frequenza relativa

La distribuzione delle frequenze relative p_i di una variabile statistica corrisponde alla distribuzione di probabilità f nel caso in cui la popolazione abbia un numero finito N di elementi e le modalità siano equiprobabili. Infatti in tal caso

$$f(x_i) = P(X^{-1}(\{x_i\})) = \frac{\#X^{-1}(\{x_i\})}{N} = \frac{n_i}{N} = p_i$$

Esempio - Famiglie con 4 figli

Nel caso della v.c.

$$X : S \to \{0, 1, 2, 3, 4\}, \quad X(x) = \# \text{ figli maschi di } x$$

dove

$$S = \{\text{possibili famiglie con 4 figli}\}$$

si ha che

$$\#S = 2^4 = 16 \quad \text{disposizioni con ripetizione con } n = 2 \text{ e } k = 4$$

Tra queste, quelle che hanno un numero i di figli maschi sono pari al numero di sottoinsiemi di $\{1, 2, 3, 4\}$ che hanno i elementi, (combinazioni con ripetione di i elementi su 4) cioè $\binom{4}{i}$.

Famiglie con 4 figli

Pertanto

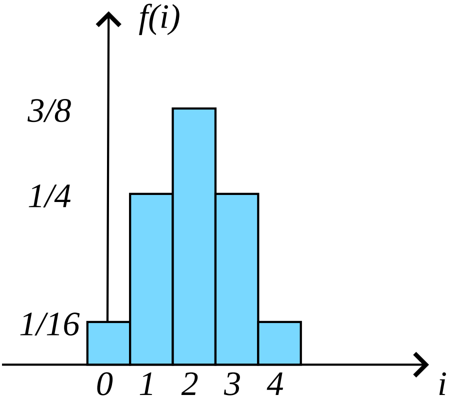
$$f(i) = P(X = i) = \binom{4}{i}\frac{1}{16}$$

ovvero

$$\begin{aligned}
f(0) &= P(X = 0) = \tfrac{1}{16};\\
f(1) &= P(X = 1) = 4 \cdot \tfrac{1}{16} = \tfrac{1}{4};\\
f(2) &= P(X = 2) = \tbinom{4}{2} \cdot \tfrac{1}{16} = \tfrac{3}{8};\\
f(3) &= P(X = 3) = \tbinom{4}{3} \cdot \tfrac{1}{16} = \tfrac{1}{4};\\
f(4) &= P(X = 4) = \tfrac{1}{16}
\end{aligned}$$

Famiglie con 4 figli

Distribuzione di probabilità di X



Media e varianza di una v.c. discreta

In completa analogia con le variabili statistiche discrete, la media o valore atteso e la varianza di una v.c. discreta sono date da

$$E(X) := \sum_i x_i f(x_i), \qquad \mathrm{Var}(X) := \sum_i [x_i - E(X)]^2 f(x_i).$$

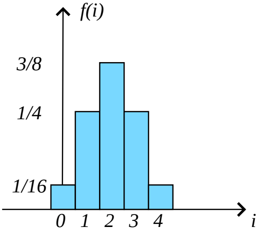
Osservazioni:

- dipendono solamente dalla distribuzione di probabilità
- v.c. identicamente distribuite hanno la stessa media e la stessa varianza
- è quindi naturale parlare di media e di varianza di una distribuzione di probabilità.

Vale la formula alternativa

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \sum_i x_i^2 f(x_i) - E(X)^2.$$

Esempio - Famiglie con 4 figli



$$E(X) = \sum_{i=0}^{4} i f(i) = f(1)+2f(2)+3f(3)+4f(4) = \frac{1}{4}+2\frac{3}{8}+3\frac{1}{4}+4\frac{1}{16} = 2,$$

$$\mathrm{Var}(X) = \sum_{i=0}^{4} [i - E(X)]^2 f(i) = \sum_{i=0}^{4} [i-2]^2 f(i) = 1$$

Variabile casuale binomiale o di Bernoulli

Definizione

Si chiamano v.c. di Bernoulli quelle del tipo

$$X : S \rightarrow \{0 \text{ (insuccesso)}, 1 \text{ (successo)}\}$$

e tali che

$$P(\{1\}) = p, \quad P(\{0\}) = 1 - p$$

Il **parametro** $p \in]0, 1[$, pari alla probabilità di osservare un successo, rappresenta una caratteristica (generalmente incognita) del fenomeno rappresentato mediante la v.c. di Bernoulli (per es. la probabilità di sopravvivenza o che esca testa).

È indicata per descrivere fenomeni che si manifestano con due sole modalità possibili (per es. la sopravvivenza o il lancio di una moneta).

Distribuzione binomiale

La distribuzione di probabilità di qualunque v.c. di Bernoulli è

$$\boxed{f(x;p) = p^x(1-p)^{1-x}} \quad \text{dove } x \in \{0,1\}$$

ed è detta distribuzione binomiale di parametro p.
Per indicare che una v.c. X ha questa distribuzione si scrive

$$X \sim BI(1,p)$$

Si ha $f(0;p) = P(X=0) = 1-p,\ f(1;p) = P(X=1) = p$. e quindi risulta

$$E(X) = \sum_{i=0}^{1} if(i;p) = 0 \cdot (1-p) + 1 \cdot p = p,$$

e si può facilmente verificare che

$$Var(X) = p(1-p).$$

Distribuzione binomiale su n prove

Se invece X è la v.c. che conta il numero di successi ottenuti in n prove indipendenti (esempio: n lanci di una moneta), allora

$$X : S^n \to \{0, 1, 2, \ldots, n\}$$

e la sua distribuzione risulta

$$\boxed{f_n(i; p) := P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}} \quad i = 0, 1, \ldots, n$$

e si scrive che

$$X \sim BI(n; p).$$

Esempio: la variabile che conta i figli maschi delle famiglie con 4 figli ha distribuzione $BI(4, 1/2)$.

Si noti che dalla formula del binomio di Newton segue che

$$\sum_{k=0}^{n} P(X=i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = (p+1-p)^n = 1$$

in accordo col fatto che la probabilità totale deve valere 1.

Media e varianza di BI(n; p)

Media e varianza della distribuzione $BI(n,p)$ sono date da

$$\boxed{E(X) = np; \quad \mathrm{Var}(X) = np(1-p)}$$

Esercizio - Marcatura

Di una popolazione di 15 lupi, 5 vengono catturati, marcati con un collare e rilasciati nel loro ambiente.

Successivamente, 3 lupi vengono catturati sperando che tra essi ve ne siano alcuni di quelli marcati, in modo da osservare le differenze con l'analisi precedente.

Qual'è la probabilità che esattamente 2 tra i 3 animali catturati siano già marcati?

Marcatura e distribuzione ipergeometrica

Spazio campionario S: tutti i sottoinsiemi di 3 lupi che sono

$$\#S = \binom{15}{3} = 455$$

Casi favorevoli: terne in cui almeno due lupi sono marcati, che sono in totale

$$\#(\text{sottoinsiemi di 2 lupi tra i 5 marcati})$$
$$\times$$
$$\#(\text{modi di scegliere un lupo tra i 10 rimanenti})$$
$$=$$
$$\binom{5}{2} \cdot \binom{10}{1}$$

Indicata con X la v.c. che conta i lupi marcati si ha

$$P(X=2) = \frac{\binom{5}{2} \cdot \binom{10}{1}}{\binom{15}{3}} = \frac{20}{91} \simeq 0.22$$

Marcatura e distribuzione ipergeometrica

Procedendo in maniera analoga si trova che

$$P(X=i)=\frac{\binom{5}{i}\cdot\binom{10}{3-i}}{\binom{15}{3}}$$

da cui si calcola facilmente l'intera distribuzione di probabilità di X

$$P(X=0)=\frac{\binom{5}{0}\cdot\binom{10}{3}}{\binom{15}{3}}=\frac{10!12!}{7!15!}=\frac{8\cdot 9\cdot 10}{13\cdot 14\cdot 15}=\frac{24}{91}\simeq 0.26$$

$$P(X=1)=\frac{\binom{5}{1}\cdot\binom{10}{2}}{\binom{15}{3}}=5\frac{10!}{2!8!}\frac{3!12!}{15!}=\frac{45}{91}\simeq 0.49$$

$$P(X=3)=\frac{\binom{5}{3}\cdot\binom{10}{0}}{\binom{15}{3}}=\frac{5!}{3!2!}\frac{3!12!}{15!}=\frac{2}{91}\simeq 0.02$$

Distribuzione ipergeometrica

Generalizzando al caso di una popolazione di N elementi di cui $0 \leq K \leq N$ marcati e supponendo di pescarne a caso n, la v.c. che conta gli esemplari marcati ha la seguente distribuzione:

$$\boxed{P(X = i) = \frac{\binom{K}{i} \cdot \binom{N-K}{n-i}}{\binom{N}{n}}}$$

detta distribuzione ipergeometrica. Media e varianza di X valgono

$$E(X) = \frac{K}{N} n, \quad \mathrm{Var}(X) = \frac{K}{N}\left(1 - \frac{K}{N}\right)$$

Esercizio

In una popolazione di 20 lupi ne vengono marcati 4. Determinare il numero minimo di animali da ricatturare per essere sicuri al 90% di prenderne almeno uno marcato.

Esercizio

In una popolazione di 20 lupi ne vengono marcati 4. Determinare il numero minimo di animali da ricatturare per essere sicuri al 90% di prenderne almeno uno marcato.

La probabilità che su n lupi ve ne sia almeno uno marcato è

$$P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$$

ma è meno calcoloso osservare che

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0).$$

Basta quindi calcolare $P(X = 0)$.

Marcatura e distribuzione ipergeometrica

Sapendo che $K = 4$ e $N = 20$ si ha

$$P(X=0) = \frac{\binom{K}{i} \cdot \binom{N-K}{n-i}}{\binom{N}{n}}$$

$$= \frac{\binom{4}{0} \cdot \binom{16}{n}}{\binom{20}{n}} = \frac{(20-n)(19-n)(18-n)(17-n)}{20 \cdot 19 \cdot 18 \cdot 17}.$$

Dobbiamo ora imporre che

$$P(X \geq 1) = 1 - P(X=0) \geq 0.9$$

cioè

$$1 - \frac{(20-n)(19-n)(18-n)(17-n)}{20 \cdot 19 \cdot 18 \cdot 17} \geq 0.9$$

Calcolando il primo membro per i diversi valori di n si trova che il minimo n per cui vale la disuguaglianza è $n = 9$.

Esempio - Gioco del Lotto

Si eseguono 5 estrazioni contemporanee da un'urna contenente sfere identiche numerate da 1 a 90.
Le estrazioni vengono eseguite 11 volte, una per ogni "ruota".

La probabilità di uscita di un singolo numero giocando su di una sola ruota è

$$p = 5/90 = 1/18 \simeq 0.05$$

Qual'è la probabilità di realizzare un ambo giocando 3 numeri su una ruota sola?

Esempio - Gioco del Lotto

Si eseguono 5 estrazioni contemporanee da un'urna contenente sfere identiche numerate da 1 a 90.
Le estrazioni vengono eseguite 11 volte, una per ogni "ruota".

La probabilità di uscita di un singolo numero giocando su di una sola ruota è

$$p = 5/90 = 1/18 \simeq 0.05$$

Qual'è la probabilità di realizzare un ambo giocando 3 numeri su una ruota sola?
Pensando di marcare i numeri su cui giochiamo, possiamo ricorrere alla distribuzione ipergeometrica con $K = 3$, $n = 5$, $N = 90$. Si ha

$$P(X = 2) = \frac{\binom{K}{i} \cdot \binom{N-K}{n-i}}{\binom{N}{n}} = \frac{\binom{3}{2} \cdot \binom{87}{3}}{\binom{90}{5}} \simeq 0.007$$

Esempio - Numeri ritardatari

Calcoliamo la probabilità di ritardo di un numero nel gioco del Lotto.

Sia X la v.c. che conta a quale estrazione esce il numero.

La probabilità che il numero esca alla prima estrazione (ritardo 0) è p. Di conseguenza quella che non esca è $1-p$.

Allora la probabilità che esca alla seconda (evento A) e non alla prima (evento B) è

$$P(X=2) = P(A \cap B) = P(A)P(B) = p(1-p)$$

Numeri ritardatari e distribuzione geometrica

In generale, se si verificano $k-1$ insuccessi e il numero esce alla k-esima estrazione, si ha

$$\boxed{P(X=k)=p(1-p)^{k-1}}$$

detta **distribuzione geometrica** di parametro p.
In particolare la probabilità che un numero esca esattamente alla 101-esima estrazione è

$$P(X=101)=\frac{1}{18}(1-\frac{1}{18})^{100}\simeq 2\cdot 10^{-5}$$

Numeri ritardatari e distribuzione geometrica

La probabilità che il numero ritardi almeno di 100 estrazioni (ma non esca necessariamente alla 101-esima) è

$$\begin{aligned} P(X \geq 101) &= 1 - P(X < 101) = 1 - \sum_{k=1}^{100} P(X = k) \\ &= 1 - \sum_{k=1}^{100} p(1-p)^{k-1} = 1 - p \sum_{k=0}^{99} (1-p)^k \\ &= (1-p)^{100} = (\frac{17}{18})^{100} \simeq 0.003 \end{aligned}$$

Siccome le ruote sono 11 e i numeri 90, vi saranno, in media, $90 \cdot 11 \cdot 0.003 \simeq 3$ numeri con ritardi superiori alle 100 estrazioni.

Esempio - Distribuzione di alberi su un territorio

In una zona pianeggiante di $10\ km^2$ *sono distribuite* 40.000 *querce. Con quale probabilità analizzando una zona limitata, per esempio di* $1000\ m^2$*, possiamo trovare "i" querce?*

Esempio - Distribuzione di alberi su un territorio

In una zona pianeggiante di 10 km^2 *sono distribuite* 40.000 *querce. Con quale probabilità analizzando una zona limitata, per esempio di* 1000 m^2, *possiamo trovare "i" querce?*

Suddividiamo la zona grande di 10 km^2 in quadrati di area pari a 1000 m^2. Siccome

$$10\ km^2 = 10 \cdot (10^3 m)^2 = 10^7\ m^2$$

allora il numero di quadrati della suddivisione è $10^7/10^3 = 10^4$.

Sia X la v.c. che conta le querce che cadono in uno dei quadrati.

Una singola quercia avrà probabilità $p = 1/10^4$ di appartenere al quadrato (successo) e $1 - p$ di non appartenere (insuccesso).

Quindi la probabilità che una quercia appartenga al quadrato ha distribuzione binomiale $BI(1, p)$.

Ripetendo l'esperimento per ogni quercia, cioè 40000 volte, si ha che

$$X \sim BI(n; p) \text{ con } n = 4 \cdot 10^4 \text{ e } p = 10^{-4}$$

ovvero

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} = \binom{4 \cdot 10^4}{i} 10^{-4i} (1 - 10^{-4})^{4 \cdot 10^4 - i}.$$

Si ha dunque

$$\begin{aligned}
P(X = 0) &= \tbinom{4 \cdot 10^4}{0} (1 - 10^{-4})^{4 \cdot 10^4} = 0.9999^{40000} \simeq 0.018, \\
P(X = 1) &= \tbinom{4 \cdot 10^4}{1} 10^{-4} (1 - 10^{-4})^{4 \cdot 10^4 - 1} = 40.9999^{39999} \simeq 0.073, \\
P(X = 2) &\simeq 0.146, \\
P(X = 3) &\simeq 0.195, \\
P(X = 4) &\simeq 0.195, \\
P(X = 5) &\simeq 0.156 \ldots
\end{aligned}$$

Il conto è disagevole a causa degli alti valori di n. Ma c'e' un modo per semplificare i conti... vediamo come fare.

Anzitutto è facile calcolare

$$E(X) = np = 4 \cdot 10^4 \cdot 10^{-4} = 4$$

$$\mathrm{Var}(X) = np(1-p) = 4(1-10^{-4}) \simeq 4$$

Eventi rari e distribuzione di Poisson

Il conto è disagevole a causa degli alti valori di n. Ma c'e' un modo per semplificare i conti... vediamo come fare.

Anzitutto è facile calcolare

$$E(X) = np = 4 \cdot 10^4 \cdot 10^{-4} = 4$$

$$\mathrm{Var}(X) = np(1-p) = 4(1-10^{-4}) \simeq 4$$

Si pone dunque $np = m$ (es. $m = 4$)., da cui $p = m/n$, e si sostituisce p nell'espressione di $P(X = i)$ ottenendo

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} = \binom{n}{i} (\frac{m}{n})^i (1 - \frac{m}{n})^{n-i}$$

A questo punto si passa al limite per $n \to \infty$ ottenendo...

Eventi rari e distribuzione di Poisson

$$\lim_{n\to\infty} P(X=i) = \lim_{n\to\infty}\left[\binom{n}{i}\left(\frac{m}{n}\right)^i\left(1-\frac{m}{n}\right)^{n-i}\right] = \frac{m^i}{i!}\,\mathrm{e}^{-m}$$

Utilizzando la formula con $m = 4$

$$P(X=i) \simeq \frac{4^i}{i!}\,\mathrm{e}^{-4}\,.$$

Nel caso dell'esempio e approssimando alla terza cifra decimale si riottengono, con meno fatica, gli stessi valori calcolati in precedenza per $i = 1, 2, 3, 4, 5$.

Eventi rari e distribuzione di Poisson

definizione

Si chiama distribuzione di Poisson di media m, la distribuzione di probabilità

$$\boxed{P(X = i) = \frac{m^i}{i!}\,\mathrm{e}^{-m}}$$

Caratteristiche della legge di Poisson:

- si usa nel caso di ripetizione di molti eventi che hanno singolarmente una piccola probabilità di realizzarsi; viene anche detta **legge degli eventi rari**
- descrive la distribuzione di probabilità di una v.c. X che può assumere un numero infinito di valori interi (i), con $E(X) = \mathrm{Var}(X) = m$
- nella pratica, a partire dai dati sperimentali si possono determinare il valore atteso e la varianza di X. Solo se questi valori sono simili, si può ipotizzare che i dati siano distribuiti con legge di Poisson.

Esercizi consigliati...

Esercizi su variabili discrete: Testo da 11.1 a 11.8.

Quale distribuzione usare?

Dipende dal contesto:

- Nella **ripetizione di eventi indipendenti** la variabile X **che conta il numero di successi** è distribuita con legge **binomiale**. Al crescere del numero di ripetizioni il calcolo si complica; se però la probabilità di successo è piccola (evento raro) e il numero di ripetizioni è grande allora X è distribuita con legge di **Poisson** (media e varianza sono uguali).
- **Se in un insieme di N elementi, k di essi posseggono una caratteristica che li distingue dagli altri**, allora la distribuzione di probabilità della v.c. X **che conta gli elementi con quella caratteristica in un campione casuale di cardinalità** n è distribuita con legge **ipergeometrica**.
- Nel caso di **prove ripetute e indipendenti, la v.c. che conta a quale ripetizione un dato evento si verifica per la prima volta** è distribuita con legge **geometrica**.

Esercizio - Esempio 11.27 Testo

In un dipartimento si usano vari microscopi elettronici prodotti da una stessa ditta. In 10 anni, ogni microscopio ha avuto in media 6 guasti, con una deviazione standard $\sigma = 2.5$*. Determinare la probabilità che in 10 anni si abbiano più di 10 guasti.*

Esercizio - Esempio 11.27 Testo

In un dipartimento si usano vari microscopi elettronici prodotti da una stessa ditta. In 10 anni, ogni microscopio ha avuto in media 6 guasti, con una deviazione standard $\sigma = 2.5$. Determinare la probabilità che in 10 anni si abbiano più di 10 guasti.

Indichiamo con X il numero di rotture.

Possiamo escludere di usare le leggi geometrica e ipergeometrica. Anche la binomiale semplice è da escludere perchè X non assume solo due valori.

Chiediamoci se possiamo ipotizzare che X sia distribuita con legge di Poisson.

Esercizi

L'ipotesi non è irragionevole poiché

$$\mathrm{Var}(X) = \sigma^2 = 2.5^2 = 6.25$$

quasi uguale alla media (6).

Dunque

$$P(X \geq 10) = 1 - \sum_{i=0}^{9} P(X = i) = 1 - \sum_{i=0}^{9} \frac{6^i}{i!} \mathrm{e}^{-6} \simeq 1 - 0.96 = 0.04$$

Variabili casuali continue

Ricordiamo che le **v.c. discrete** sono caratterizzate dalla **funzione di distribuzione (o massa) di probabilità** definita da

$$f(x_i) = P(X = x_i) := P(\{s \in S \ : \ X(s) = x_i\}) = P(X^{-1}(\{x_i\}))$$

che ad ogni valore x_i associa la probabilità che la v.c. X assuma il valore x_i.

Osservazione: per le **v.c. continue** è poco utile assegnare probabilità ai singoli valori come si vede dal seguente esempio.

Esempio - La distribuzione uniforme in $[0,1[$

Sia X la v.c. che sceglie "a caso" un numero $x \in [0,1[$. Si ha

$$X : S \to [0,1[$$

Osserviamo che

- dividendo l'intervallo $[0,1[$ in n intervalli di ugual ampiezza, la probabilità che x cada in uno di essi è $1/n$
- in generale, dati x_1, $x_2 \in [0,1[$ si ha $P(X \in [x_1, x_2[) = x_2 - x_1$
- i singoli valori hanno probabilità nulla, infatti

$$0 \leq P(X = x_0) \leq P(X \in [x_0, x_0 + \frac{1}{n}[) = \frac{1}{n} \to 0$$

quindi $P(X = x_0) = 0$.

Ha più senso invece assegnare probabilità agli intervalli. Perciò si definisce la **funzione di ripartizione**

definizione

La funzione

$$F(x) := P(X \leq x)$$

si chiama funzione di ripartizione (o di distribuzione cumulativa) della variabile casuale X rispetto alla probabilità P.

Esempio - distribuzione uniforme in $[0,1[$

Nel caso della variabile X dell'esempio precedente (che sceglie "a caso" un numero $x \in [0,1[$) si ha

$$F(x) = P(X \leq x) = x \quad \text{per ogni } x \in [0,1[$$

detta distribuzione uniforme nell'intervallo $[0,1[$.

Variabili casuali continue

definizione

Se la funzione di ripartizione F è derivabile, la sua derivata

$$f(x) := F'(x)$$

è detta densità di probabilità di X ripetto a P.

Densità di probabilità

Per definizione di derivata

$$\begin{aligned} f(x) &= \lim_{h\to 0} \frac{F(x+h)-F(x)}{h} \\ &= \lim_{h\to 0} \frac{P(X\leq x+h)-P(X\leq x)}{h} \\ &= \lim_{h\to 0^+} \frac{P(X\in]x,x+h])}{h} \end{aligned}$$

e ciò si può interpretare dicendo che
la probabilità che X assuma valori in un intervallo di ampiezza h piccola intorno ad x è approssimativamente uguale a $f(x)h$.

In questo senso la densità di probabilità è l'analogo probabilistico della densità di frequenza di una variabile statistica.

Valore atteso e varianza di una v.c. continua

Anche per le v.c. continue $X : S \to \mathbb{R}$ possiamo definire il valore atteso (a fianco la definizione nel caso discreto)

$$E(X) := \int_{-\infty}^{+\infty} x\, f(x)\, dx \qquad \left[E(X) := \sum_i x_i f(x_i)\right]$$

e la varianza

$$\mathrm{Var}(X) := \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x)\, dx \qquad \left[\mathrm{Var}(X) := \sum_i [x_i - E(X)]^2 f(x_i\right.$$

Valgono le proprietà

- $E(aX + bY) = aE(X) + bE(Y)$, $a, b \in \mathbb{R}$;
- $\mathrm{Var}(a + X) = a + \mathrm{Var}(X)$, $a \in \mathbb{R}$.

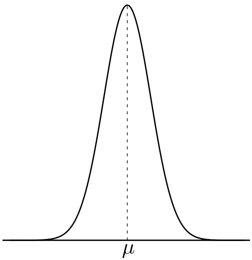
La v.c. Normale

La v.c. continua più importante è la v.c. Normale o Gaussiana

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

$$\boxed{f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\, \mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2}}$$

Grafico della funzione di densità



Importanza della v.c. Normale

La v.c. Normale riveste un ruolo fondamentale perché

- descrive bene il manifestarsi di molti fenomeni, per esempio:
  - errori di misura (genesi della Normale)
  - caratteristiche morfologiche (altezza, lunghezza)
- gode di importanti proprietà (aspetto tecnico rilevante)

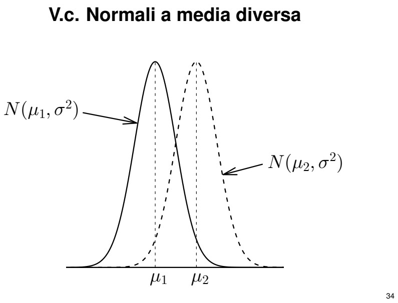
I parametri

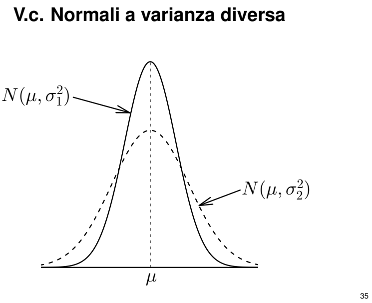
La funzione di densità di probabilità della v.c. Normale dipende da due parametri:

- μ rappresenta il centro (valore atteso o media) della distribuzione; si ha infatti $E[X] = \mu$
- σ^2 modula il grado di dispersione dei valori; si ha infatti $\mathrm{Var}[X] = \sigma^2$

Anche in questo caso i parametri rappresentano caratteristiche incognite del fenomeno studiato.

Variabile casuale Normale





$$\sigma_1 < \sigma_2$$

Errori di misura e genesi della Normale

Gli errori di misurazione di una grandezza si dividono in

- sistematici, dovuti ad esempio ad imperfezioni o starature dello strumento di misura; sono **eliminabili**;
- aleatori, dovuti alle condizioni in cui si replica l'esperimento (es. temperatura, pressione, umidità, umore dello sperimentatore, ecc.); sono variabili da misura a misura e quindi **non eliminabili**.

Siccome gli errori aleatori non sono eliminabili, è importante conoscerne le proprietà statisticamente rilevanti in modo da tenerne conto negli esperimenti.

Errori di misura e genesi della Normale

Misurando una grandezza X otteniamo N misure X_1, X_2,..., X_N.

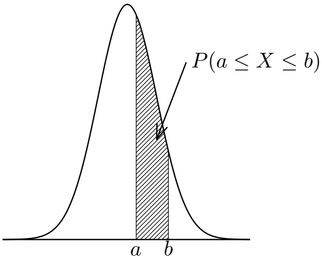
Questi valori non sono generalmente identici, ma oscillano attorno alla loro media m_N.
Queste medie, a loro volta, al crescere di N oscillano attorno ad un valore limite μ.

Molti studiosi, tra cui Gauss, indagando a fondo il comportamento di queste oscillazioni, hanno concluso (Teorema del limite centrale) che m_N è distribuita con una legge gaussiana con valore atteso uguale a μ (da considerarsi il vero valore della grandezza).

Calcolo della probabilità per un intervallo

La probabilità è rappresentata dall'area sotto la curva di densità all'interno dell'intervallo.

Area=Probabilità



36

L'area si determina mediante l'operazione di integrazione. Infatti, per il Teorema Fondamentale del Calcolo Integrale si ha

$$\begin{aligned}\int_a^b f(x)\,dx &= \int_a^b \frac{d}{dx}\big[P(X \le x)\big]\,dx \\ &= P(X \le b) - P(X \le a) = P(a < x \le b).\end{aligned}$$

Densità di probabilità

Osserviamo che affinché f sia una densità di probabilità occorre dunque che

- $f \geq 0$
- $\int_{-\infty}^{+\infty} f(x)\, dx = 1$

Caso della distribuzione Normale

Consideriamo ad esempio una variabile Z gaussiana con media 0 e varianza 1, detta distribuzione Normale standard

$$Z \sim \mathcal{N}(0, 1).$$

Calcoliamo, ad esempio,

$$P(-1 < Z \leq 1) = \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} \mathrm{e}^{-\frac{x^2}{2}}\, dx,$$

ma non è possibile procedere nel calcolo dell'integrale cercando una primitiva della funzione integranda. Infatti non esiste una primitiva esprimibile in termini finiti come somma, prodotto o composizione di funzioni elementari.

Si può effettuare un calcolo approssimato dell'integrale sostituendo l'esponenziale con funzioni più semplici da integrare, ad esempio con opportuni polinomi.
Questa idea ha consentito ai matematici di redarre delle **tavole numeriche** e di predisporre **programmi** per il calcolo numerico degli integrali.

Nella pratica si può quindi ricorrere a

- **computer**
- **tavole numeriche**

Normale standardizzata e tavole

Le tavole della distribuzione Normale (o i computer; ad esempio la funzione DISTRIB.NORM.ST di OpenOffice) consentono di risolvere il seguente problema

$$P(Z \leq a) = ?$$

dove Z indica la Normale standard.

Altri problemi

Qualora si debba calcolare la probabilità per un intervallo di forma diversa, si applicano le seguenti regole

$$P(a < Z \leq b) = P(Z \leq b) - P(Z < a), \quad P(Z > a) = 1 - P(Z \leq a)$$

Gli estremi dell'intervallo sono irrilevanti (la probabilità di un punto è pari a zero).

Esempio:

$$P(-1 < Z \leq 1) = P(Z \leq 1) - P(Z < -1) \simeq 0.841 - 0.159 = 0.682$$

Per le altre distribuzioni normali ...

Per una v.c.

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

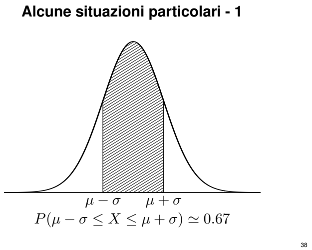
si applica l'operazione di standardizzazione:

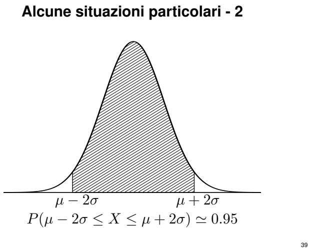
$$Z = \frac{X - \mu}{\sigma}$$

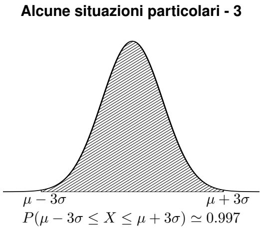
Si ha $Z \sim \mathcal{N}(0, 1)$ e

$$P(X \leq \alpha) = P(Z \leq \frac{\alpha - \mu}{\sigma})$$

Le situazioni seguenti si riferiscono ad una distribuzione $\mathcal{N}(\mu, \sigma^2)$

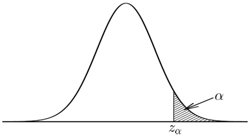






Il problema inverso

Consiste nel calcolare il valore che lascia alla sua destra (o sinistra) un'area prefissata α



Sulle tavole il problema si trova risolto per alcuni valori tipici di α

$$
\begin{array}{lcl}
\alpha = 0.05 & \Longrightarrow & z_\alpha = 1.6449 \\
\alpha = 0.025 & \Longrightarrow & z_\alpha = 1.9600 \\
\alpha = 0.01 & \Longrightarrow & z_\alpha = 2.3263 \\
\alpha = 0.005 & \Longrightarrow & z_\alpha = 2.5758
\end{array}
$$

oppure si può usare la funzione INV.NORM.ST di OpenOffice.

Esempio - Distribuzione delle altezze

L'altezza media dei maschi adulti di una certa popolazione è $m = 175\,\mathrm{cm}$. *Supponendo che le altezze siano distribuite con legge Normale e che lo scarto dalla media sia* $\sigma = 10\,\mathrm{cm}$, *calcolare l'intervallo, centrato intorno alla media, in cui, con probabilità del 99.7% sono distribuite le altezze. Inoltre, preso a caso un individuo della popolazione, determinare con quale probabilità esso avrà un'altezza compresa tra 175 e 195 cm.*

Esempio - Distribuzione delle altezze

L'altezza media dei maschi adulti di una certa popolazione è $m = 175\,\mathrm{cm}$. *Supponendo che le altezze siano distribuite con legge Normale e che lo scarto dalla media sia* $\sigma = 10\,\mathrm{cm}$, *calcolare l'intervallo, centrato intorno alla media, in cui, con probabilità del 99.7% sono distribuite le altezze. Inoltre, preso a caso un individuo della popolazione, determinare con quale probabilità esso avrà un'altezza compresa tra 175 e 195 cm.*

Dobbiamo determinare α tale che

$$P(m - \alpha < X \leq m + \alpha) = 0.997$$

In base a quanto osservato nella "situazione particolare 3" si ha $\alpha = 3\sigma = 30$. Dunque l'intervallo cercato è $[145, 205]$.

Esempio - Distribuzione delle altezze

L'altezza media dei maschi adulti di una certa popolazione è $m = 175\,\mathrm{cm}$. *Supponendo che le altezze siano distribuite con legge Normale e che lo scarto dalla media sia* $\sigma = 10\,\mathrm{cm}$, *calcolare l'intervallo, centrato intorno alla media, in cui, con probabilità del 99.7% sono distribuite le altezze. Inoltre, preso a caso un individuo della popolazione, determinare con quale probabilità esso avrà un'altezza compresa tra 175 e 195 cm.*

Esempio - Distribuzione delle altezze

L'altezza media dei maschi adulti di una certa popolazione è $m = 175\,\mathrm{cm}$. *Supponendo che le altezze siano distribuite con legge Normale e che lo scarto dalla media sia* $\sigma = 10\,\mathrm{cm}$, *calcolare l'intervallo, centrato intorno alla media, in cui, con probabilità del 99.7% sono distribuite le altezze. Inoltre, preso a caso un individuo della popolazione, determinare con quale probabilità esso avrà un'altezza compresa tra 175 e 195 cm.*

Infine

$$\begin{aligned} P(175 < X \leq 195) &= P(m \leq X \leq m + 2\sigma) \\ &= \frac{1}{2} P(m - 2\sigma < X \leq m + 2\sigma) = \frac{1}{2} 0.95 = 0.475 \end{aligned}$$

Riassumendo ...

- le v.c. sono utilizzate come modelli teorici per rappresentare fenomeni reali
- la distribuzione di probabilità di una v.c. dipende da parametri, generalmente incogniti
- i parametri rappresentano caratteristiche intrinseche del fenomeno studiato

Esercizi

Esercizi consigliati: da 11.10 a 11.14