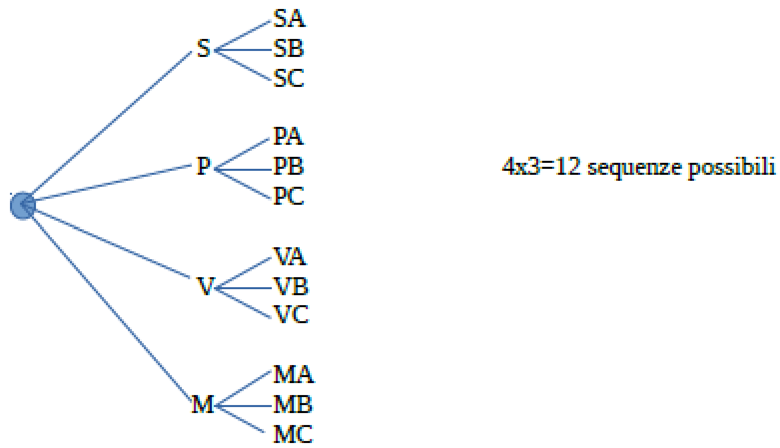
STATISTICA

D'ANGELO CARMINE

CALCOLO COMBINATORIO

Problema: un'azienda produce antenne che possono essere di 4 tipi (S, P, V, M). Ogni modello è disponibile in 3 diverse altezze(A, B, C). In quanti modi si può segliere un'antenna specificando modello ed altezza?



Principio fondametale del calcolo combinatorio

Si realizzano due esperimenti. Sappiamo che il primo espereimento ha *m* esiti possibili e che per ognuno di essi il secondo ne abbia *n*. Allora se sequenza distinte di esiti dei due esperimenti producono esiti finali differenti, in totale avremo $m * n$ esiti possibili.

Esercizi:

- Quanti sono i risultati possibili del lancio di due dadi? 6 * 6 = 36.

- Uno studente può inserire nel piano di studi 2 insegnamenti a sceltaa, di questo 1 deve essere scelto da un elenco di 9 insegnamenti mentre l'altro da un elenco distinto di 10. In quanti nodi diversi può completare il piano di studi? 9 * 10 = 90.

- Si vuole costituire un comitato di 2 persone, partendo da un gruppo di un uomo U e due donne A, B. Vogliamo contare quanti sono i gruppi possibili.

1) Base scelgo una donna – 2 volte.
2) Scelgo una delle due persone rimanenti – 2 volte.

Gruppi possibili = 2 * 2 = 4 **ERRORE**
perchè AB e BA sono uguali, calcolo combinatorio non applicabile.

Problema: Quante sono le funzioni booleane definite su {0,1}.

x	f(x) = x	f(x) = 0	f(x) = 1	f(x) = 1 - x
0	0	0	1	1
1	1	0	1	0

Principio generalizzato del calcolo combinatorio

Si realizzano *r* esperimenti. Supponiamo che il primo esperimento produca n_1 esiti possibili, e che per ognuno di questi esiti, il secondo n_i produca n_2 esiti e che per ognuno degli esiti dei primi due il terzo esperimento abbia n_3 possibili esiti, ... n_r esiti possibili.
Se equenza distinte di esiti degli *r* esperimenti producono esiti finali distinti, in totale hanno n_1, n_2, n_3, ... n_r esiti possibili.

Esercizi:

- Quante sono le possibili targhe formate da 7 caratteri di cui 3 sono numeri e 4 lettere (tra 26 disponibili). *26 * 26 * 26 * 26 * 10 * 10 * 10 =* $26^4 * 10^3$ *= 456 976 000.*
Come cambia se non c'è ripetizione tra numeri e lettere? *26 * 25 * 24 * 23 * 10 * 9 * 8 = 258 336 000*

- Lancio n volte una moneta, quanti sono i possibili esiti? 2^n esiti possibili.
- Quante sono le possibili password di lunghezza pari a 8 che posso costruire partendo da un alfabeto formato da 10 cifre numeriche, 26 lettere maiuscole e 26 lettere minuscole?
$(10 + 26 + 26)^8 = 218 * 10^{12}$.

Permutazioni

Problema: In quanti modi posso ordinare le tre lettere a, b, c?

a b c
a c b
b a c
b c a
c a b
c b a

Sono 6.

Il numero di permutazione di n oggetti è dato da n!. n! = n * (n – 1) * (n – 2) * ... 1.
0! = 1, 1! = 1, 2! = 2 * 1 = 2.

Esercizi:

- Nella prima fila di un'aula devono sedersi 6 studenti: 3 ragazzi e 3 ragazze.
1) In quanti modi si possono sedere gli studenti ? *6!*
2) In quanti modi si possono sedere se sia i maschi che le femmine devono sedersi vicino fra loro?
*Si distinguono 2 gruppi quindi: 2 * 3! * 3!.*
3) In quanti modi si possono sedere se solo i maschi devono sedere vicini? 4! * 3!
4) In quanti modi si possono sedere se due studenti dello stesso sesso non devono stare vicini?
*2 * 3! * 3!*

- Si eseguono 10 programmi di cui 4 scritti in C e 6 in JAVA.
Viene stilato l'elenco dei programmi nell'ordine di terminazione delle esecuzioni.
1) Quanti sono i possibili elenchi dei programmi? *10!*
2) Quanti sono i possibili elenchi se i programmi in C e JAVA compaiono in due liste separate?
*2 * 4! * 6!*

Problema: Quanti sono gli anagrammi della parola casa?
4! = 24 **ERRORE,** problema con due oggetti uguali, si divide per il nummero delle possibili permutazioni degli oggetti uguali. A = 2 quindi 2!.

$$\frac{4!}{2!}$$

Permutazioni di oggetti non tutti distinti

Se ci sono *n* oggetti, di cui n_1 uguali tra di loro, n_2 uguali tra loro e distinti dai precedenti,..., n_r ugali tra loro e distinti da tutti i precedenti, allora il numero delle per mutazioni è n! $\frac{n!}{n1!n2!...nr!}$.

Esercizi:

- Quanti sono i vettori booleani di dimensione *n* che sono formati da *k* bit pari a 1 ed *n-k* pari a 0?

$\frac{n!}{k!(n-k)!}$ = coefficiente binominale $\binom{n}{k}$.

- Un gioco consiste in 12 pezzetti di legno colorati dei quali 6 neri, 4 rossi, 1 bianco e 1 blu. In quanti modi si possono allineare i pezzi? $\frac{12!}{6!4!1!1!}$.

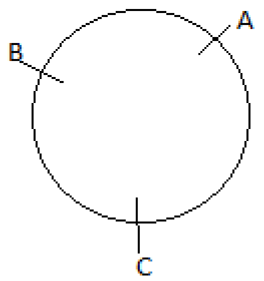
- In una giornata di calcio ci sono 10 partite ed ognuna può dar luogo a 3 tipi di risulatato (1, X, 2).

1) Quanti sono i risultati possibili? 3^{10}.

2) Quante sono le seguenze dei risultati possibili se vi sono 5 vincitori in casa, e 3 pareggi?

$\frac{10!}{5!3!2!}$

- Qunate configurazioni si possono realizzare disponendo 3 oggetti in un allineamento circolare?



3! = 6 possibili permutazioni NO!!!!!!!!!!!!!!!!!!!! **ERRORE**

Le possibili configurazioni sono 2

$\left\{\begin{array}{l}\text{A B C}\\ \text{B C A}\\ \text{C B A}\end{array}\right.$ $\left\{\begin{array}{l}\text{A C B}\\ \text{C B A}\\ \text{B A C}\end{array}\right.$

Se ha n oggetti e indico con x il numero di allineamenti circolari, ottengo: $n * x = n!$

$$N = \frac{n!}{n} = (n-1)!$$

Disposizioni

Problema: Quanti sono i sottoinsiemi di dimensione k di cui si possono formare a partire da un insieme di n oggetti distinti, nell'ipotesi in cui l'ordine è rilevante?

Senza ripetizioni

n = 4	a, b, c, d				
k = 2	a,b	b,a	c,a	d,a	disposizione semplice = $D_{4,2}$
	a,c	b,c	c,b	d,b	
	a,d	b,d	c,d	d,c	

1° elemento = 4 scelte
2° elemento = 3 scelte quindi 4 * 3 = 12 scelte

Se ammettiamo le ripetizioni

n = 4	a, b, c, d				
k = 2	a,a	b,b	c,c	d,d	disposizione con ripetizione = $D'_{4,2}$
	a,b	b,a	c,a	d,a	
	a,c	b,c	c,b	d,b	
	a,d	b,d	c,d	d,c	

1° elemento = 4 scelte
2° elemento = 4 scelte quindi 4 * 3 = 16 scelte

Nota:
Disposizione semplice: disposizione di n oggetti in k classi senza ripetizioni.
Disposizione con ripetizione: disposizione di n oggetti in k classi con ripetizioni.

Problema: Quanti sono i sottoinsiemi ordinati di dimensione k che si possono formare a partire da un insieme di n oggetti distinti?

Abbiamo k posizioni - - - - - - - senza ripetizione

1° 2° 3° ... n°

$$n * n-1 * n-2 * \ldots * (n-k+1) = D_{n,k} = \frac{n!}{(n-k)!} = (n)_k = \text{fattoriale discendente}$$

Con ripetizione : $n * n * n * \ldots * n = D'_{n,4} = n^k$.

3 casi in cui compaiono le disposizioni:

- Sottoinsiemi ordinati di dimensione k a partire da n oggetti.
- Estrazioni di biglie da un'urna nell'ipotesi che è rilevante l'ordine di apparizione.
  ES*:* Da un'urna che contiene n biglie numerate da 1 a n. Si effettuano k estrazioni a caso.
  1. Quante sono le sequenze distinte ottenibili se le estrazioni sono senza inserimento?
  2. Quante sono le sequenze distinte ottenibili se le estrazioni sono senza inserimento?

  Entrambe nell'ipotesi in cui è rilevante anche l'ordine di apparizione.
  1. $n * (n-1) * (n-2) \ldots (n-k+1) = D_{n,k}$.
  2. $n^k = D'_{n,k}$.
- Distribuzione in n celle di k contrassegni.
  ES: Si vogliono distribuire 7 diversi regali a 10 bambini. Quante sono le possibili distribuzioni se nessun bambino può ricevere più di un regalo?

A	1	4					
B	2	2					
C	3	3					

D		1					
E	4						
F	5	5					
G	6						
H	7	6					
I		7					
L							

10 * 9 * 8* 7 *6 * 5 * 4 = 604 800 = $D_{10,\ 7}$

Esercizi:

- Voglio contare in quanti modi 10 studenti possono sedersi in una classe con 15 sedie?

15 * 14 * 13 * 12 * 11 * 10 * 9 * 8 * 7 * 6 = $D_{15,\ 10}$ = 1,09 *10^{10}

Combinazioni

3 casi in cui compaiono le combinazioni:

- Sottoinsiemi non ordinati di dimensione k a partire da n oggetti.
  Combinazioni semplici di n oggetti su k classi:
  $C_{n,k} = \frac{D_{n,k}}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$
  Combinazione con ripetizione:
  $C'_{n,k} = \binom{n+k-1}{k}$
  es.
  n=4 a,b,c,d
  k=2

  insieme ordinato

ab	bc	ca	dc
ac	ba	cb	db
ad	bd	cd	dc

  12 sequenze

  insieme non ordinato

ab	bc	~~ca~~	~~dc~~
ac	ba	~~cb~~	~~db~~
ad	~~bd~~	~~cd~~	dc

  6 sequenze = $\frac{12}{2!}$

$$\frac{12}{2!} = 6 \text{ sequenzae} = \frac{D_{4,2}}{2!} = \frac{4!}{2!} = \binom{4}{2} = \frac{D_{n,k}}{k!}$$

- Estrazioni di k biglie da un'urna nell'ipotesi che l'ordinamento non è rilevante.
- Distribuzione in n celle di 4 contrassegni senza tenet conto di quale oggetto metto nella cella.
  Problema: ci sono 4 amici che in una partita di calcio eseguono complessivamente 3 reti. Quante sono le possibili distribuzioni di queste reti se ciascuno ha segnato al più un goal?

A	X	X	X	
B	X		X	X
C	X	X		X
D		X	X	X

$C_{4,\ 3} = \binom{4}{3} = 4$

Esercizi:

- Una classe di ballo è composta da 22 studenti, di questi 22 sono 10 donne e 12 uomini. In quanti modi si possono formare 5 coppie?

$$\binom{10}{5} * \binom{12}{5} * 5! = 23\,950\,080$$

- Quanti sono i comitati composti da 2 donne e 3 uomini si possono formare da un gruppo di 5 donne e 7 uomini? Quanti se due uomini rifiutano di far parte dello stesso comitato?

1. $\binom{5}{2} * \binom{7}{3}$
2. Abbiamo due casi il primo in cui non viene scelto nessuno dei due: $\binom{5}{2} * \binom{5}{3}$, il secondo in cui segliamo uno dei due : $\binom{5}{2} * \binom{2}{1} * \binom{5}{2}$, li sommiamo e abbiamo 300 comitati.

- Un software antivirus riporta che 3 cartelle su 10 sono infette. Quanti sono le possibili sequenza?
$\binom{10}{3} = 120.$

- In quanti modi si possono disporre in fila 5 donne e 4 uomini in modo che 2 uomini non siano mai consecutivi? ***- x – x – x – x – x -*** $= \binom{6}{4} * 5! * 4! = 43\,200.$

Coefficienti binominali

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

$$\binom{n}{0} = 1 = \frac{n!}{0!(n-0)!} = \frac{n!}{n!} = 1$$

$$\binom{n}{1} = \frac{n!}{1!(n-1)!} = \frac{n\cancel{(n-1)!}}{1!\cancel{(n-1)!}} = n$$

$$\binom{n}{k} = 0 \; k > n \text{ e } k < 0$$

Dal coefficiente binominale possiamo ricavare la **TAVOLA DI TARTAGLIA**

$k \backslash n$	0	1	2	3	4	5
0	1	1	1	1	1	1
1	0	1	2	3	4	5
2	0	0	1	3	6	10
3	0	0	0	1	4	10
4	0	0	0	0	1	5
5	0	0	0	0	0	1

Se sommiamo il risultato di goni colonna abbiamo delle potenze di due, e possiamo esprimerle nella seguente formula: $\sum_{k=0}^{n} \binom{n}{k} = 2^n$ perchè vale?

1. I sottoinsiemi di oggetti sono $2^{n.}$
2. Formula del binomio di Newton: $(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$ con a = 1 e b = 1 abbiamo che vale appunto 2^n.

Esercizi:

- Un'urna contiene n biglie nere e m biglie bianche. In quanti modi posso formare un sacchetto di k biglie?

$$\binom{(n+m)}{k} = \binom{n}{0} * \binom{m}{k} + \binom{n}{1} * \binom{m}{(k-1)} + \binom{n}{2} * \binom{m}{(k-2)} +\ldots+ \binom{n}{(k-1)} * \binom{m}{1} + \binom{n}{k} * \binom{m}{0} = \sum_{k=0}^{n} \binom{m}{j} \binom{m}{(k-j)}$$ ***Principio di Vandermonde***

- Calcolare quanti sono i vettori $(x_1,\ldots, x_k)$:
  1. Ogni x_i è un intero positivo $1 \le x_i \le n$.
  2. Ogni x_i è un intero positivo $1 \le x_i \le n$ e ogni x_i è diverso da tutti i precedenti.
  3. Ogni x_i è un intero positivo $1 \le x_i \le n$ e in più la sequenza è strettamente ordinata $(x_1 < x_2 < \ldots < x_k)$.
  4. Ogni x_i è un intero positivo $1 \le x_i \le n$ e non è strettamente ordinato $(x_1 \le x_2 \le\ldots \le x_k)$.

Soluzione:

1. $n * n * n * \ldots * n = n^k = D'_{n, k}$.
2. S $n * n - 1 * n – 2 * \ldots* n – k +1 = D_{n, k}$.
3. $\binom{n}{k}$ perchè per ogni vettore esiste un solo modo di ordinarli = $C_{n, k}$.
4. $\binom{(n+k-1)}{k} = C'_{n, k}$.

ESPERIMENTO: fenomeno di cui non è possibile prevedere l'esito.
SPAZIO CAMPIONARIO: l'insieme di tutti i possibili esiti dell'esperimento (S).

Esercizi:

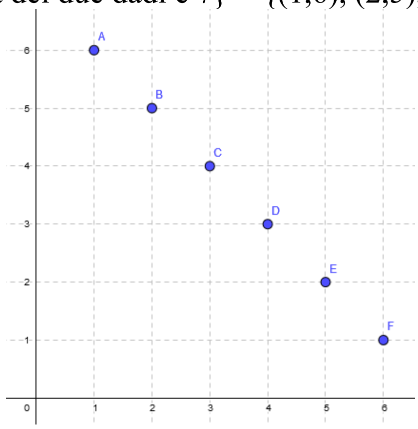
- Lancio 2 dati. S = {(i,j) , i,j = 1, ... , 6}
  |S| = 6^2 = 36
- Lancio n volte una moneta $S = \{(\omega_1, \omega_2, \ldots , \omega_n\}, \omega_i \in (t, c) \}$
  |S| = 2^n
- Lancio ripetutamente una moneta finchè non esce testa. $S = \{ n \in N, n \ge 1\}$ *infinito numerabile*
- L' esperimento consiste nel misurare il tempo di vita di un'apparecchio elettronico.
  $S = \{ x \in R, x \ge 0\}$ infinito non numerabile.
- Una scatola contiene 3 biglie. 1 rossa, 1 verde e una blu. L'esperimento consiste nell'estrarre una biglia una biglia reinserendola e poi estrarne un'altra.
  S = {(i,j), i,j ∈ {rossa, verde, blu}}
  |S| = 3^2 = 9

Che succede se la prima biglia non è reinserita?
S = {(i,j), i,j ∈ {rossa, verde, blu}, i ≠ j}
|S| = 3 * 2

EVENTO: è un qualsiasi sottoinsieme dello spazio campionario. Diremo quindi che l'evento si verifica se l'esito dell'esperimento appartiene all'evento.

Esercizi:

- Lanciamo 2 dadi A = {la somma dei due dadi è 7} = {(1,6), (2,5), (3,4), (4,3) , (5,2), (6,1)}



- Lancio 3 monete.

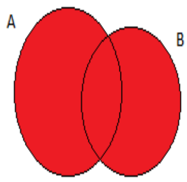
A = { al primo lancio esce croce}

S = { (t t t), (t t c), (t c t), (t c c), (c t t), (c t c), (c c t), (c c c) }

Operazioni tra gli eventi

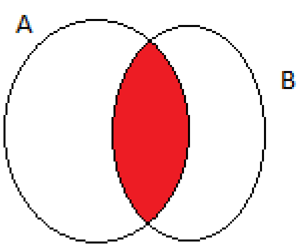
- A U B =

Si genera un terzo evento che contiene l'unione degli eventi A e B



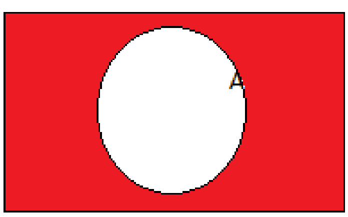
- A ∩ B =

Si genera un terzo evento che contiene l'intersezione degli eventi A e B



- Complementare: Ā =

Evento che contiene gli esiti che non appartengono ad A



Esercizi:

- Il datore di un ospedale classifica i pazienti con:

- 1 se hanno assicurazione
- 0 se non hanno assicurazione

e con:

1. b se lo stato di salute è buono
2. m se lo stato di salute è medio
3. s se lo stato di salute è serio

L'esperimento consiste nel classificare questi pazienti nel seguente modo:

1. Descrivere lo spazio campionario.
2. Sia A = {il paziente è in condizione serie}.
3. Sia B = {il paziente non è assicurato}.
4. **Esiti del complementare di** $\bar{B} \cup A$.

1] S = { (0 b), (0 m), (0 s) (1 b) (1 m) (1 s)}
2] A = { (0 s) (1 s)}
3] B = { (0 b) (0 m) (0 s)}
4] $\bar{B} \cup A$ = {(1 b) (1 m) (1 s) (0 s)}

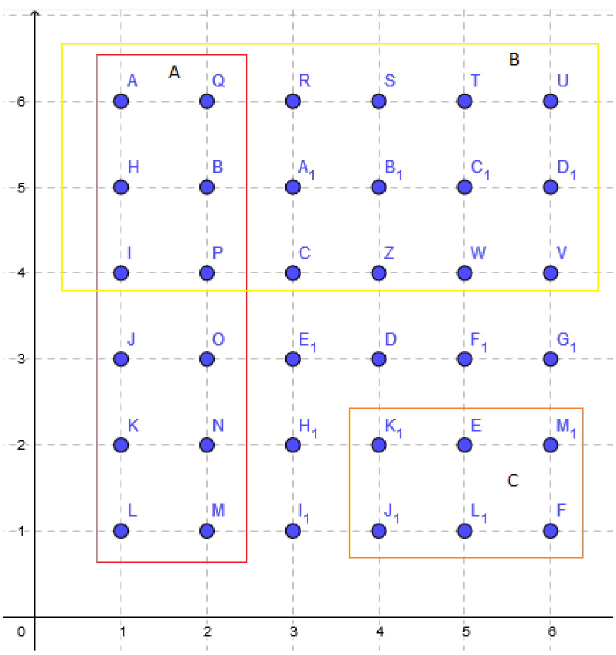
Alcuni notazioni:

Lo spazio campionario S è chiamato anche evento certo.
Ø = evento impossibile.
A, B sono incompatibili se A ∩ B = Ø
A, B sono necessari se AUB = S

Esercizi:

- Lancio 2 dadi

A = {al primo lancio esce 1 o 2}
B = {al secondo lancio esce 4 o 5 o 6}
C = {al primo lancio esce 4 o 5 o 6 mentre al secondo 1 o 2}



A ∩ C = Ø
B ∩ C = Ø
S = { (i, j), i,j= 1, ..., 6}
A ∩ B = {(1 4), (1 5), (1 6), (2 4), (2 5), (2 6)} ≠ Ø

- Scelgo a caso una carta tra le 52 carte francesi. Voglio considerare 4 eventi:

A_1 = {la carta estratta è di picche}
A_2 = {la carta estratta è di cuori}
A_3 = {la carta estratta è di fiori}
A_4 = {la carta estratta è di quadri}

Sono eventi incompatibili? SI
Sono eventi necessari? SI

- A_i = {lo studente all'esame riceve il voto 17 + i} i = 1, ... 13
Sono eventi incompatibili? SI
Sono eventi necessari? No, non sono necessari perchè manca l'insieme dei bocciati

Queste proprietà si possono applicare anche all'infinito:

$\bigcup_{i=1}^{+\infty} A_i$ = evento i cui esiti sono contenuti in almeno un evento A_i

$\bigcap_{i=1}^{+\infty} A_i$ = evento i cui esiti sono compresi in tutti gli A_i

$A \subset B$ diremo che A implica B $\leftrightarrow$ $\bar{B} \subset \bar{A}$

PROPRIETÀ

1. Commutativa: $A \cup B = B \cup A \; e \; A \cap B = B \cap A$
2. Associativa: $A \cup (B \cup C) = (A \cup B) \cup C \; e \; A \cap (B \cap C) = (A \cap B) \cap C$
3. Distributiva: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \; e \; (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
4. Formule di De Morgan:
$\overline{A \cup B} = \bar{A} \cap \bar{B} \; e \; \overline{A \cap B} = \bar{A} \cup \bar{B}$ per un numero infinito di elementi diventa

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \bar{A} \; ; \; \overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \bar{A}_i$$

Esercizi:
- Lancio 3 monete.

A = {esce sempre testa} = {(t t t)}
B = {esce almeno una volta testa} = {(t t t), (t t c), (t c t), (t c c), (c t t), (c t c), (c c t)}
$A \subset B$ = SI

Con i complementari abbiamo:

$\overline{A}$ = esce almeno una volta croce
$\overline{B}$ = esce sempre croce
$\overline{B} \subset \overline{A}$: SI

PROBABILITÀ

Definizione frequentista: ripetiamo un'esperimento nella medisima condizione.
Sia S lo spazio campionario. Sia poi E un'evento.
Frequenza assoluta di E = n(E) = numero di volte che si realizza E nelle prime n prove.

$0 \leqslant n(E) \leqslant n$

$\frac{n(E)}{n}$ = frequenza relativa dell'evento E.

Esercizi:

- Lancio una moneta E = {Esce testa}

n	ω	n(E)	$\frac{n(E)}{n}$
1	c	0	0
2	c	0	0
3	c	0	0
4	t	1	$\frac{1}{4}$ = 0,25
5	t	2	$\frac{2}{5}$ = 0,4
6	c	2	$\frac{2}{6}$ = 0,33
7	t	3	$\frac{3}{7}$ = 0,43
8	t	4	$\frac{4}{8} = \frac{1}{2}$ = 0,5

Quindi possiamo dire che:

$\lim_{n \to +\infty} \frac{n(E)}{n} \simeq 0{,}5$

probabilità dell'evento E: P(E) = $\lim_{n \to +\infty} \frac{n(E)}{n}$. Se prendiamo $0 \leqslant n(E) \leqslant n$ e lo dividiamo per n, abbiamo che 0 = $\frac{0}{n} \leqslant \frac{(n(E))}{n} \leqslant \frac{n}{n}$ = 1 quando $0 \leqslant P(E) \leqslant n$.

Se consideriamo P(S) = $\lim_{n \to +\infty} \frac{n(S)}{n}$ = $\frac{n}{n}$ = 1 quindi P(S) = 1.

Se abbiamo due eventi A e B che sono incompatibili:

$A \cap B = \emptyset$; $\frac{n(A \cup B)}{n} = \frac{(n(A))}{n} + \frac{(n(B))}{n}$ ed otteniamo $P(A \cup B) = P(A) + P(B)$

Definizione soggettiva: La probabilità di un evento A è il prezzo che un individuo ritiene equo pagare per ricevere 1 se si verifica A, 0 altrimenti. Scommetto P(A) e ho un guadagno:

$$\begin{cases} 1 - P(A) \text{ se si verifica A} \\ - P(A) \text{ se a non si verifica} \end{cases}$$

condizione di coerenza: le probabilità vanno fissate in modo che non sia possibile ottenere con un insieme di scommesse una vincita certa o una perdita certa.
Infatti :

se abbiamo P(A) < 0 ? non va bene perchè otteniamo un guadagno certo.
se abbiamo P(A) > 1 ? non va bene perchè otteniamo un perdita certa.

Quindi $0 \leq P(A) \leq n$, questa regola si applica anche con la probabilità sullo spazio campionario, infatti:

se abbiamo P(S) < 1 ? non va bene perchè otteniamo un guadagno certo (- P(S)).
se abbiamo P() > 0 ? non va bene perchè otteniamo un perdita certa (P(S) = 1).

$A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$

CLASSE DEGLI EVENTI

$\mathcal{F}$ = famiglia di sottoinsiemi dello spazio campionario S :

1. S $\in \mathcal{F}$. Questo implica che $\emptyset \in \mathcal{F}$
2. A $\in \mathcal{F}$-> $\bar{A} \in \mathcal{F}$.
3. $A_1, A_2, \ldots, \in \mathcal{F}$ -> $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{F}$. Questo implica $\bigcap_{i=1}^{+\infty}$ che $= A_i \in \mathcal{F}$

Esercizi:

- $\mathcal{F} = \{ \emptyset , S\}$ Si, è la classe degli eventi
- $\mathcal{F} = \{ \emptyset , A, S\}$ No perchè manca il complementare di A
- $\mathcal{F} =$ P(s) si, insieme delle parti.

DEFINIZIONE ASSIOMATICA

P : $\mathcal{F}$-> R : NOTA: P = probabilità

Assiomi:

1. P(S) = 1

2. $\forall\ A \in \mathcal{F},\ 0 \leqslant P(A) \leqslant 1$
3. Additività numerabile:

$A_1, A_2, \ldots \in \mathcal{F},\ A_i \cap A_j = \emptyset, i \neq j,\ P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

Le sue proprietà sono le seguenti:

1. $P(\emptyset)=0$

Dimostrazione. Consideriamo una successione di eventi $A_1, A_2, \ldots$, dove $A_1 = S$, $A_2 = A_3 = \ldots = \emptyset$; allora, essendo gli eventi $A_1, A_2, \ldots$ a due a due incompatibili ed essendo

$$S = \bigcup_{i=1}^{\infty} A_i$$

dall'Assioma 3 ricaviamo

$$P(S) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = P(S) + \sum_{i=2}^{\infty} P(\emptyset)$$

il che implica $\sum_{i=2}^{\infty} P(\emptyset) = 0$ e quindi

$$P(\emptyset) = 0$$

perciò l'evento impossibile ha probabilità 0 di verificarsi.

2. Additività finita : presi $A_i, A_2, \ldots, A_n \in \mathcal{F}$ $A_i \cap A_j = \emptyset, i \neq j \rightarrow$ $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$

Esercizi:

- Lancio una moneta. S = {t, c} P_1 = P{t} P_2 = P{c}

$0 \leqslant P_1 \leqslant 1$ e $0 \leqslant P_2 \leqslant 1$

1 = P(S) = P({t, c}) = P({t}) + P({c}) = $P_1 + P_2$. Quindi abbiamo che $P_1 + P_2 = 1$;

$0 \leqslant P_1 \leqslant 1$; $0 \leqslant P_2 \leqslant 1$ che implica $P_1 = P_2 = \frac{1}{2}$ ma solo se la moneta non è truccata.

Se la moneta è truccata abbiamo $P_1 \neq P_2\ quindi \begin{Bmatrix} P_1 = 2P_2 \\ P_1 + P_2 = 1 \end{Bmatrix} \rightarrow \begin{Bmatrix} P_1 = \frac{2}{3} \\ P_2 = \frac{1}{3} \end{Bmatrix}$

3. $P(\bar{A}) = 1 - P(A) \forall\ \in \mathcal{F}$

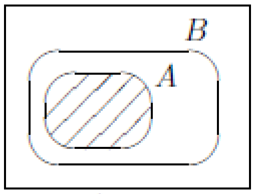
per il primo assioma : $1 = P(S) = P(A \cup \bar{A})$

per l'additività finita :

$$1 = P(S) = P(A \cup \bar{A}) = P(A) + P(\bar{A}) \rightarrow P(A) + P(\bar{A}) = 1 \rightarrow P(\bar{A}) = 1 - P(A)$$

4. $A \in B \rightarrow P(A) \leqslant P(B)$

Dimostrazione. Essendo $A \subset B$, abbiamo che B può essere espresso come $B = A \cup (\overline{A} \cap B)$, con $A \cap (\overline{A} \cap B) = \emptyset$.

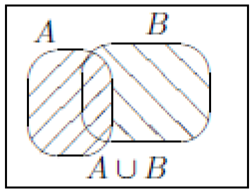


$A \subset B$

Dalla proprietà di additività finita segue
$P(B) = P(A \cup (\overline{A} \cap B)) = P(A) + P(\overline{A} \cap B)$,
da cui si ha $P(B) \geq P(A)$, essendo $P(\overline{A} \cap B) \geq 0$.

5.
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

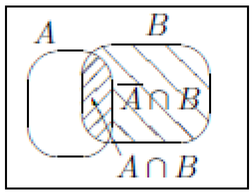
Dimostrazione. Notiamo che $A \cup B$ può essere espresso come unione di due eventi incompatibili A e $\overline{A} \cap B$. Grazie alla proprietà di additività finita otteniamo



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

Inoltre, essendo $B = (A \cap B) \cup (\overline{A} \cap B)$, con $A \cap B$ e $\overline{A} \cap B$ eventi incompatibili, applicando nuovamente la proprietà di additività finita abbiamo

$$P(B) = P(A \cap B) + P(\overline{A} \cap B)$$



o, equivalentemente,

$$P(\overline{A} \cap B) = P(B) - P(A \cap B)$$

che completa la dimostrazione.

Esercizi:

- Uno studente deve sottoporsi a 2 test.
  B_i = {lo studente supera il test i-esimo} i = 1 o 2
  $P(B_1) = 0,5$; $P(B_2) = 0,4$; $P(B_1 \cap B_2) = 0,3$
  Quanto vale la probabilità che non superi nessuno dei due test?

$$P(\bar{B}_1 \cap \bar{B}_2) = P(\overline{(B_1 \cup B_2)}) = 1 - P(B_1 \cup B_2) = 1 - [P(B_1) + P(B_2) - P(B_1 \cap B_2)] = 1 - 0,6 = 0,4$$

- Il 28% degli uomini statunitensi fuma sigarette, il 7% sigari, il 5% entrambi:

a) La probabilità che un uomo non fumi nè sigarette nè sigari.
b) La probabilità che un uomo scelto a caso fumi sigari e non sigarette.
P(uomo fuma sigarette) = 0,28
P(uomo fuma sigari) = 0,07
P(uomo fuma sia sigari che sigarette) = 0,05

Soluzioni:
a)
$$P(\overline{sigari} \cap \overline{sigarette}) = P(\overline{(sigari \cup sigarette)}) = 1 - P(sigari \cup sigarette) = 1 - [0,28 + 0,07 - 0,05]$$

b) $P(sigari \cap (\overline{sigarette})) = P(sigari) - P(sigari \cap sigarette) = 0,07 - 0,05 = 0,02$

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

Esercizi:

- S = {programma presenta errori di sintassi}
  I = {programma presenta errori di I/O}
  O = {programma presenta errori dinon classificati}
  Quindi abbiamo che:

Evento	S	I	O	$S\cap I$	$S\cap O$	$I\cap O$	$S\cap I\cap O$
Probabilità	20/100	10/100	5/100	6/100	3/100	2/100	1/100

A = {programma scelto a caso tra i 100 che presenti tutti gli erroi}:

$$P(A)=P(S)+P(I)+P(O)-P(S\cap I)-P(S\cap O)-P(I\cap O)+-P(S\cap I\cap O)=\frac{25}{100}=\frac{1}{4}$$

FORMULA GENERALIZZATA: PRINCIPIO DI INCLUSIONE ESCLUSIONE:

La probabilità dell'unione di n eventi $A_1, A_2, \ldots, A_n$ può esprimersi al seguente modo:

per $n=2$: $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$;

per $n=3$:
$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) = {} & P(A_1) + P(A_2) + P(A_3) \\ & - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\ & + P(A_1 \cap A_2 \cap A_3); \end{aligned}$$

in generale:

$$\begin{aligned} P(A_1 \cup A_2 \cup \ldots \cup A_n) = {} & \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) \\ & + \ldots + (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots \cap A_n). \end{aligned}$$

SPAZI CAMPIONARI CON ESITI EQUIPROBABILI

In molti esperimenti è naturale assumere che tutti gli esiti dello spazio campionario siano equiprobabili, con S insieme finito: $S = \{1, 2, \ldots, N\}$. Allora si ipotizza che

$$P(\{1\}) = P(\{2\}) = \ldots = P(\{N\})$$

il che implica

$$P(\{i\}) = \frac{1}{N}, \qquad i = 1, 2, \ldots, N,$$

essendo $1 = P(S) = P(\{1\} \cup \{2\} \cup \ldots \cup \{N\}) = P(\{1\}) + P(\{2\}) + \ldots + P(\{N\})$.
Per la proprietà di additività avremo perciò che per ogni evento A

$$P(A) = \frac{|A|}{|S|} = \frac{\text{numero di elementi di } A}{\text{numero di elementi di } S} \qquad \text{(definizione classica di probabilità)}$$

Se assumiamo che tutti gli esiti di un esperimento siano equiprobabili, allora la probabilità di ogni evento A è uguale alla proporzione degli esiti dello spazio campionario contenuti in A (come rapporto di casi favorevoli su casi possibili).

Definizione: si può fare solo se lo spazio campionario è finito con esiti ugualmente probabili.
Esercizi:
- Lancio 2 dadi, qual'è la probabilità che la somma sia 7?
  S = {(i,j) : i,j = 1, ..., 6} |S| = 36 casi possibili
  A = {la somma dei due dadi è 7} = {(1,6), (2,5), (3,4), (4,3) , (5,2), (6,1) }
  |A| = 6 casi favorevoli

$$P(A)=\frac{6}{36}=\frac{1}{6}$$

- Una carta è scelta a caso da un mazzo di 52 carte. Calcolare la pribabilità che la carta sia di fiori.

Casi favorevoli = 13
Casi possibili = 52

$$P=\frac{13}{52}=\frac{1}{4}$$

- Una giovane coppia decide di avere 2 figli. Qual'è la probabilità che siano entrambe femmine?

Casi possibili = S = {(M,M), (M,F), (F,M), (F,F)} = 4 casi, casi favorevoli = 1

$$P=\frac{1}{4}$$

- Estraiamo 3 biglie da un'urna che contiene 6 bianche e 5 nere. Calcolare la probabilità in cui esca 1 bianca e 2 nere.

Abbiamo due possibili casi per operare:

1. Se teniamo conto dell'ordine di estrazione e supponiamo che ogni possibile esito dello spazio campionario sia equiprobabile, la probabilità richiesta è
   A = {1 è bianca e 2 nere} = { (b n n), (n b n), (n n b)} quindi :
   $$\frac{(6 \cdot 5 \cdot 4) + (5 \cdot 6 \cdot 4) + (5 \cdot 4 \cdot 6)}{11 \cdot 10 \cdot 9} = \frac{3 \cdot 120}{990} = \frac{4}{11} = 0{,}3636$$

2. Se non teniamo conto dell'ordine di estrazione, la probabilità richiesta è
   $$\frac{\binom{6}{1}\binom{5}{2}}{\binom{11}{3}} = \frac{6 \cdot 10}{165} = \frac{4}{11} = 0{,}3636$$

- Una commissione di 5 persone viene estratta da un gruppo composto di 6 uomini e 9 donne. Se la selezione avviene in modo casuale, qual `e la probabilit`a che la commissione consti di 3 uomini e 2 donne?

Ognuna delle $\binom{15}{3}$ combinazione è estratta in modo equiprobabile, quindi:

$$\frac{(\binom{6}{3}+\binom{9}{2})}{\binom{15}{5}}=\frac{240}{1001}\simeq 0{,}24$$

- 20 famiglie vivono in un quartiere.
  4: hanno solo un figlio
  8: hanno due figli
  5: hanno tre figli
  2: hanno quattro figli

1: hanno cinque figli

1) Scelgo a caso una famiglia, qual è la probabilità che abbiamo *i* figli? (i = 1, ... , n).
2) Sceltgo a caso uno dei figli, bisogna calcolare la probabilità che appartenga ad una famiglia con figli. (i = 1, ..., n).

Soluzioni:

1) $con \quad i=1 \quad abbiamo: \ \frac{4}{20} \qquad con \quad i=2 \quad abbiamo: \ \frac{8}{20} \qquad con \quad i=3 \quad abbiamo: \ \frac{5}{20}$

$con \quad i=4 \quad abbiamo: \ \frac{2}{20} \qquad con \quad i=5 \quad abbiamo: \ \frac{1}{20}$

2) $con \quad i=1 \quad abbiamo: \ \frac{4}{48} \qquad con \quad i=2 \quad abbiamo: \ \frac{8}{48} \qquad con \quad i=3 \quad abbiamo: \ \frac{5}{48}$

$con \quad i=2 \quad abbiamo: \ \frac{4}{48} \qquad con \quad i=5 \quad abbiamo: \ \frac{1}{48}$

- Un'urna contiene n biglie, di cui una \`e bianca. Se estraiamo k biglie una alla volta, in modo tale che a ogni estrazione la probabilità di estrarre una qualunque delle biglie rimanenti sia la stessa, qual \`e la probabilit\`a che la biglia bianca sia estratta?

Ognuno dei $\binom{n}{k}$ possibili insiemi \`e estratto in modo equiprobabile, quindi:

$$P(\text{si estrae la biglia bianca}) = \frac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{\frac{(n-1)!}{(k-1)!\,(n-k)!}}{\frac{n!}{k!\,(n-k)!}} = \frac{k}{n}$$

- Esempio. Nel gioco del lotto si estraggono a caso 5 numeri da un'urna contenente 90 numeri compresi tra 1 e 90. (Le cinquine sono equiprobabili). Se si fissano k numeri distinti compresi tra 1 e 90 qual \`e la probabilit\`a che questi saranno tra i 5 estratti?

Poiché l'ordine di estrazione qui non è rilevante, lo spazio campionario è l'insieme delle $\binom{90}{5}$ combinazioni semplici di 90 oggetti in 5 classi.

L'evento $A_{k} =$ {i k numeri fissati sono tra i 5 estratti} è costituito dalle sequenze di *S* che presentano all'interno i k numeri fissati; pertanto la sua cardinalità è pari al numero di combinazioni semplici di *90 − k* oggetti in *5 − k* classi: $\binom{90-k}{5-k}$. Si ha quindi:

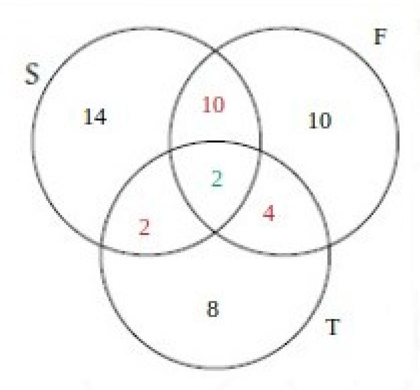
$$\mathbf{P(A_k)} = \frac{(\binom{k}{k}*\binom{90-k}{5-k})}{\binom{90}{5}} \qquad \mathbf{P(A_1)} = \frac{(\binom{1}{1}*\binom{89}{4})}{\binom{90}{5}} = \frac{1}{18} \simeq 0{,}05$$

$$\mathbf{P(A_2)} = \frac{(\binom{2}{2}*\binom{88}{3})}{\binom{90}{5}} = \frac{2}{801} \simeq 0{,}0025 \qquad \mathbf{P(A_3)} = \frac{(\binom{3}{3}*\binom{87}{2})}{\binom{90}{5}} = \frac{1}{11748} \simeq 0{,}000085$$

P(A$_4$) = $\frac{\left(\binom{4}{4}*\binom{86}{1}\right)}{\binom{90}{5}}=\frac{1}{511038}\simeq 0{,}000001957$ **P(A$_5$) =** $\frac{\left(\binom{5}{5}*\binom{85}{0}\right)}{\binom{90}{5}}=\frac{1}{43949268}\simeq 2{,}2*10^{-8}$

- Una scuola offre 3 corsi di lungue {Francese(F), Spagnolo(S), Tedesco(T)}, ed è aperta a 100 studenti della scuola.
Sono organizzati nel seguente modo:

- S = 28
- F = 26
- T = 16
- S ∩ F = 12
- T ∩ F = 6
- T ∩ S = 4
- T ∩ S ∩ F = 2



Scelto uno studente qual è la probabilità che non frequenti nessun corso?

$$P(\overline{T}\cap\overline{F}\cap\overline{S}) \;=\; P(\overline{(T\cup F\cup S)}) \;=\; 1-(T\cup F\cup S) \;=$$
$$= 1-\{P(T)+P(F)+P(S)-P(T\cap F)-P(T\cap S)-P(F\cap S)+P(T\cap F\cap S)\}$$
$$= 1-\left\{\frac{16}{100}+\frac{26}{100}+\frac{28}{100}-\frac{6}{100}-\frac{4}{100}-\frac{12}{100}+\frac{2}{100}\right\} \;=\; 1-\frac{50}{100} \;=\; \frac{1}{2}$$

Calcola la probabilità che lo studente frequenti un solo corso?

$$P(S\cap\overline{T}\cap\overline{F})+P(F\cap\overline{S}\cap\overline{T})+P(T\cap\overline{S}\cap\overline{F})=$$
$$=\;\frac{14}{100}+\frac{10}{100}+\frac{8}{100}=\frac{32}{100}=0{,}32$$

- Gioco del poker :

1. Probabilità di fare colore (5 carte dello stesso seme): $P(colore)=\frac{\binom{13}{5}*\binom{4}{1}}{\binom{52}{5}}\simeq 0{,}002$

2. Probabilità di fare coppia: P(coppia) = P(a, a, b, c, d) = $\frac{\binom{4}{2}*\binom{13}{1}*\binom{4}{1}*\binom{4}{1}*\binom{4}{1}*\binom{12}{3}}{\binom{52}{5}} \simeq 0{,}4226$

3. Probabilità di fare doppia coppia: P(doppia coppia) = P(a, a, b, b, c) = $\frac{\binom{4}{2}*\binom{4}{2}*\binom{13}{2}*\binom{4}{1}*\binom{11}{1}}{\binom{52}{5}} \simeq 0{,}0475$

4. Probabilità di fare poker: P(poker) = P(a, a, a, a, b) = $\frac{\binom{4}{4}*\binom{13}{1}*\binom{4}{1}*\binom{12}{1}}{\binom{52}{5}} \simeq 0{,}00024$

PROBLEMA DELLE CONCORDANZE

Da un'urna che contiene biglie numerate da *1 a n* si estraggono casualmente in sequenza le n biglie. Diciamo che si ha una concordanza nell'i-esima estrazione se in tale estrazione viene estratta la biglia avente numero i, per i = 1, 2, . . . , n. Calcolare la probabilit`a che non si abbia concordanza in nessuna estrazione.

Per i = 1, 2, . . . , n sia Ei = {si ha concordanza nell'i-esima estrazione}. Pertanto, la probabilità che si abbia almeno una concordanza è:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{r=1}^{n}(-1)^{r+1} \sum_{i_1<i_2<\ldots<i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r}).$$

Rappresentiamo l'esito dell'esperimento come un vettore ($\omega_1, \omega_2, \ldots, \omega_n$), dove ω_i de- nota il numero estratto all'estrazione i-esima. Lo spazio campionario `e quindi l'insieme delle n! Permutazioni:

$$S = \{(\omega_1, \omega_2, \ldots, \omega_n) : \omega_i \in \{1, 2, \ldots, n\},\ \omega_i \neq \omega_j \text{ per } i \neq j\}.$$

L'evento E_i si verifica se nell'estrazione i-esima si ha concordanza (e non `e escluso che si possa avere concordanza in altre estrazioni). Ci`o pu`o accadere in (n − 1)! modi, perchè nell'estrazione i-esima l'esito deve essere il numero i (per avere concordanza), mentre nelle altre estrazioni ci sono (n − 1) (n − 2) . . . 2 · 1 modi per individuare gli esiti dell'esperimento. Si ha pertanto:

$$P(E_i) = \frac{(n-1)!}{n!} = \frac{1}{n} \quad \text{e analogamente} \quad P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r}) = \frac{(n-r)!}{n!}.$$

Poiché vi sono $\binom{n}{r}$ termini in $\sum_{i_1<i_2<\ldots<i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r})$, risulta quindi

$$\sum_{i_1<i_2<\ldots<i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r}) = \binom{n}{r}\frac{(n-r)!}{n!} = \frac{n!}{r!\,(n-r)!} \cdot \frac{(n-r)!}{n!} = \frac{1}{r!}$$

da cui:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{r=1}^{n}(-1)^{r+1} \sum_{i_1<i_2<\ldots<i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r}) = \sum_{r=1}^{n}(-1)^{r+1}\frac{1}{r!}.$$

La probabilità che non si abbia concordanza in nessuna estrazione è quindi:

$$p_n = P\left(\overline{\bigcup_{i=1}^{n} E_i}\right) = 1 - P\left(\bigcup_{i=1}^{n} E_i\right) = 1 - \sum_{r=1}^{n} \frac{(-1)^{r+1}}{r!} = 1 + \sum_{r=1}^{n} \frac{(-1)^r}{r!} = \sum_{r=0}^{n} \frac{(-1)^r}{r!}$$

Notiamo che per n grande la probabilità

$$p_n = \sum_{r=0}^{n} \frac{(-1)^r}{r!}$$

può essere approssimata da

$$e^{-1} = \sum_{r=0}^{+\infty} \frac{(-1)^r}{r!} \approx 0{,}367879$$

(Ricordiamo che $\boxed{e^x = \sum_{r=0}^{+\infty} \frac{x^r}{r!}}$ per $x \in \mathbb{R}$)

n	p_n
1	0
2	0,5
3	0,333333
4	0,375
5	0,366667
6	0,368056
7	0,367857
8	0,367882
9	0,367879
10	0,367879

Esercizi spazi campionari con esiti equiprobabili:

- Lancio n volte una moneta

$S=\{(\omega_1,\ldots,\omega_2),\ \omega_i \in \{t,c\},\ i=1\ldots n\}$ $\qquad |S| = 2^n$

T_k = {al k-esimo lancio esce testa} $\qquad 1 \le k \le n$

A_k ={nei primi k lanci esce testa}

A^i_k ={nei primi k lanci esce testa}

B_k ={esce testa esattamente k volte}

$$P(T_k) = \frac{2^{n-1}}{2^n} = \frac{1}{2}$$

$$P(A_k) = \frac{2^{n-k}}{2^n} = \frac{1}{2^n}$$

$$\underbrace{t,t,t,\ldots,t}_{k},\underbrace{x,x,x,x,x}_{n-k}$$

$$P(A^i_k) = \frac{2^{n-k}}{2^n} = \frac{1}{2^k}$$

$$P(B_k) = \frac{\binom{n}{k}}{2^n}$$

T_1 si verifica in A_k quindi abbiamo $A_k \subset T_1$, quindi si prende la probabilità del più piccolo

$$P(T_1 \cup A_k) = P(T_1) + P(A_k) - P(T_1 \cap A_k)$$

$$\cancel{\frac{1}{2}} + \frac{1}{2^k} - P(A_k) = \frac{1}{2} \qquad T_1$$

A_k è più piccolo di T_1 perchè in T_1 abbiamo sia teste che croci e solo al k-esimo lancio esce testa

A_k

$P(T_1 \cup B_k) = P(T_1) + P(B_K) - P(T_1 \cap B_k)$

$$\frac{1}{2} + \frac{\binom{n}{k}}{2^n} + \frac{\binom{n-1}{k-1}}{2^n} = \frac{1}{2} + \frac{1}{2^n}\binom{n-1}{k}$$

(t,……..)

n-1 possibilià

N-k volte croce

{k-1 esce testa}, il (-1) perchè una testa già la tengo fissata

$P(A_k \cup A^i_k)$ = siccome sono incompatibili la probabilità è pari alla somma delle due probabilità

$P(A_k)+P(A^i_k) = 2* \frac{1}{2^n}$

$P(A_k \cup B_k) = P(A_k) + P(B_k) - P(A_k \cap B_k)$

$$\frac{1}{2} + \frac{\binom{n}{k}}{2^n} + \frac{1}{2^n}$$

$\{\underbrace{t,t,t,\dots,t}_{\text{K Volte}},\underbrace{c,c,\dots,c}_{\text{n-k volte}}\}$

$P(A^i_{k-1} \cup T_k) = P(A^i_{k-1}) + P(T_k) - P(A^i_k \cap T_k)$

$$\frac{1}{2^{k-1}} + \frac{1}{2} - \frac{2^{n-k}}{2^n}$$

$\{\underbrace{c,c,c,c,\dots,c}_{\text{k-1 volte}},\underset{k}{t},x,x,x,\dots,x\}$

- La moneta non truccata viene lanciata un numero indefinito di volte.
  E^p = { esce testa per la prima volta in un lancio pari}
  E^d = { esce testa per la prima volta in un lancio dispari}
  E^o = { non esce mai testa}
  F_k = { esce testa per la prima volta al lancio k-esimo} $k \geq 1$

$$E^p = F_2 \cup F_4 \cup F_6 \cup \dots = \bigcup_{k=1}^{+\infty} F_{2k}$$

$$E^d = F_1 \cup F_3 \cup F_5 \cup \dots = \bigcup_{k=0}^{+\infty} F_{2k+1}$$

$P(E^p) = P\left(\bigcup_{k=1}^{+\infty} F_{2k}\right)$ = Per il terzo assioma (additività numerabili) abbiamo che:

$$\sum_{k=1}^{+\infty} P(F_{2k})=P(F_k)=P(A'_{(k-1)}\cap T_k)=\frac{1}{2^k}=\sum_{k=1}^{+\infty}(\frac{1}{2^{2k}})=\frac{1}{3}$$

$P(E^d) =$ $$\sum_{k=0}^{+\infty} P(F_{(2k+1)})=\sum_{k=0}^{+\infty}(\frac{1}{2^{2k}}+1)=\frac{2}{3}$$

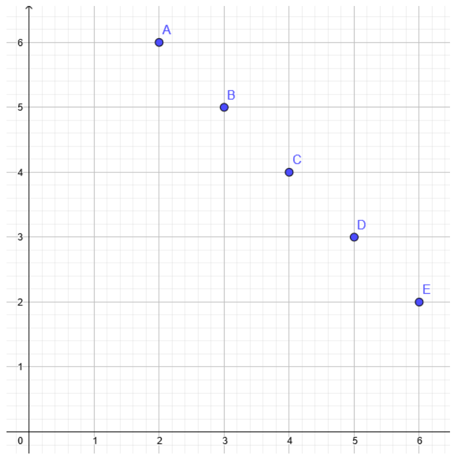
$E^0 = E^p$ **U** Ed **U** $E^0 = S$ sono icompatibili quindi:

$$1=P(S)=P(E^d)+P(E^p)+P(E^o)=\frac{1}{3}+\frac{2}{3}+p(E^o)\rightarrow P(E^o)=0$$

E^0 è quasi impossibile

PROBABILITÀ CONDIZIONATA

Lancio due dati.



A = { la somma dei due dadi è 8} $P(A)=\frac{5}{36}$

Supponiamo di osservare che il primo ha dato come esito 3.

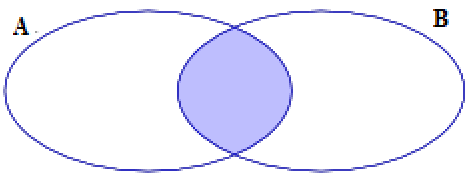
Quanto vale la probabilità di A? $P(A)=\frac{1}{6}$ = ***probabilità condizionata.***

$$P(A|B)=\frac{1}{6}$$

che in questo caso è B = {il primo dato ha esito 3}

la formula è: $P(A|B)=\frac{(P(A\cap B))}{(P(B))}=\frac{(\frac{1}{36})}{(\frac{6}{36})}=\frac{1}{6}$.

Più in generale presi due eventi E, F con P(F) > 0, allora la probabilità di P(E | F) = $\frac{(P(E\cap F))}{(P(F))}$



Esercizi:

- Lancio 3 monete.

E = {primo esito è testa}

F = {esattamente 2 esiti su 3 sono testa}

$$P(E\cap F)=\frac{2}{8}$$

S={ttt, ttc, tct, tcc, ctt, ctc, cct, ccc} = 2^3

$P(F)=\frac{3}{8}$; $P(E|F)=\frac{2}{4}=\frac{1}{2}$; più in generale possiamo dire:

$$P(E|F)=\frac{(P(F\cap E))}{(P(E))}=\frac{(\frac{2}{8})}{(\frac{4}{8})}=\frac{2}{4}=\frac{1}{2}$$

- Un signore proviene da una famiglia con due figli. Voglio calcolare la probabilità che l'altro sia una sorella.

A= {la famiglia è composta da 1M e una 1F}
B= {almeno 1 figlio è maschio}

$$P(A|B)=\frac{(P(A\cap B))}{(P(B))}=\frac{(\frac{2}{4})}{(1-\frac{1}{4})}=\frac{2}{3}$$ questo caso è applicabile se usiamo il negato di P(B)

OPPURE $$P(A|B)=\frac{(P(A\cap B))}{(P(B))}=\frac{(\frac{2}{4})}{(\frac{3}{4})}=\frac{1}{2}*(\frac{4}{3})=\frac{2}{3}$$

- Una coppia ha 2 bambini. Calcolare la probabilità che siano entrambe femmine sapendo che il primo genito è femmina.
A= {i bambini sono entrambe femmine}
B= {il primo genito è femmina}

$$P(A)=\frac{1}{4} \qquad P(B)=\frac{2}{4}$$

$$P(A|B)=\frac{(P(A\cap B))}{(P(B))}=\frac{(\frac{1}{4})}{(\frac{2}{4})}=\frac{1}{4}*(\frac{4}{2})=\frac{1}{2}$$

- Lancio n volte una moneta.
E = {Esce testa in ogni lancio}
F = {Esce testa al primo lancio}
A = {Esce testa in almeno un lancio}

P(E | F) , P(E | A)

$$P(E|F)=\frac{(P(E\cap F))}{(P(F))}=\frac{(\frac{1}{(2^n)})}{(\frac{1}{2})}=\frac{1}{(2^n-1)}$$

e $P(E|A)=\frac{(P(E\cap A))}{(P(A))}=\frac{(P(E\cap A))}{(1-P(\overline{A}))}=P\frac{(E)}{(1-(\frac{1}{2^n}))}=\frac{(\frac{1}{2^n})}{(1-(\frac{1}{2^n}))}=\frac{1}{2^n}-1$

CONTINUO PROBABILITÀ CONDIZIONATA

A,B con P(B) > 0

$$P(A|B)=P\frac{(A\cap B)}{(P(B))}\rightarrow P(A\cap B)=P(A|B)*P(B)$$

Esercizi

- Uno studente sta svolgendo un test da consegnare entro 1 ora. La probabilità che lo studente termini l'esame in meno di x ore è data $\frac{x}{2}$ con $x\in[0,1]$. Sapendo che lo studente è ancora al lavoro dopo $\frac{3}{4}$ d'ore, calcolare la probabilità che usufruisce dell'intera ora?

L_x = { lo studente conclude l'esame in meno di x ore}

F = { lo studente utilizza l'intera ora}

$$P(F|\overline{L_{\frac{3}{4}}})=P\frac{(F\cap\overline{L_{\frac{3}{4}}})}{P(\overline{L_{\frac{3}{4}}})}=P\frac{(F)}{1-P(L_{\frac{3}{4}})}=P\frac{(\overline{L_1})}{1-\frac{3}{8}} \qquad F\subset\overline{L_{\frac{3}{4}}} \qquad F=\overline{L_1}$$

$$P(L_x)=\frac{x}{2}=\frac{\frac{3}{4}}{2}=\frac{3}{8}=1-\frac{P(L_1)}{\frac{5}{8}}=\frac{4}{5}$$

- Da un'urna contenente r biglie (di cui b blu e r-b rosse), si estraggono n biglie a caso ($n\leq r$) senza reinserimento.

1) Calcolare la probabilità che la prima biglia esratta sia blu.

2) Sapendo che k delle biglie estratte sono blu, calcolare la probabilità che la prima estratta sia blu.

Soluzioni:

1) E = {la prima estratta è blu}

B_k = {vengono estratte k biglie blu} $0\leq k\leq n$

P(E) = $\frac{b\text{ \# biglie blue}}{r\text{ \# tutte le biglie}}$

2) $P(E|B_k)=\frac{\binom{n-k}{k-1}}{\binom{n}{k}}=\frac{k}{n}$

REGOLA DEL PRODOTTO

$$P(E_1\cap E_2\cap E_3\cap\ldots\cap E_n)=P(E_1)*P(E_2|E_1)*P(E_3|E_{¿}E_2)*\ldots*P(E_n|E_1\cap\ldots E_{n-1})$$
$$se\ P(E_1\cap E_2\cap E_3\cap ¿\ldots E_{n-1})>0$$

DIM:

$$P(E_1)*P\frac{(E_1\cap E_2)}{P}(E_1)*P\frac{(E_1\cap E_2\cap E_3)}{P(E_1\cap E_2)}*P\frac{(E_1\cap E_2\cap E_3\cap E_4)}{P(E_1\cap E_2\cap E_4)}*\ldots*P\frac{(E_1\cap E_2\cap E_n)}{P(E_1\cap E_2\cap E_{n-1})}=$$
$$=P(E_1\cap E_2\cap E_n)$$

Esercizi

- 52 carte con 4 assi. Sono suddivise in 4 mani da 13 carti ciascuna. Calcolare la probabilità che ogni mano contenga un asso.

E_i = {la mano i-esima ha esattamente un asso} con i = 1, 2, 3, 4

$$P(E_1 \cap E_2 \cap E_3 \cap E_4) = P(E_1) * P(E_2 | E_1) * P(E_3 | E_1 \cap E_2) * P(E_4 | E_1 \cap E_2 \cap E_3) =$$
$$= \frac{\binom{4}{1} * \binom{42}{12}}{\binom{52}{13}} * \frac{\binom{3}{1} * \binom{36}{12}}{\binom{39}{13}} * \frac{\binom{2}{1} * \binom{24}{12}}{\binom{26}{13}} * 1 \simeq 0{,}105$$

- Un'unrna contiene inizialmente 5 biglie bianche e 7 nere. Ad ogni estrazione si prenda nota del colore della biglia che viene poi rimessa nell'urna insieme ad altre due biglie ddello stesso colore.
1) Calcolare la probabilità che le prime due selezionate saranno nere e le due successive bianche.
2) Calcolare la probabilità che esattamente due delle prime 4 estratte saranno nere

N_i = { si estrae biglia nera all'estrazione i -esima } con i = 1, 2 , 3, 4

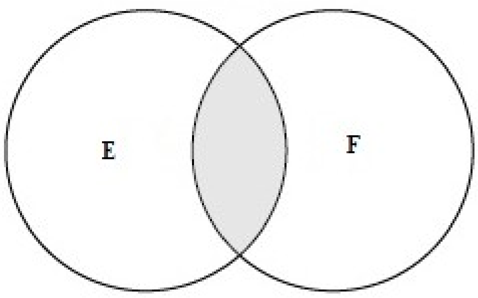
Soluzioni:

1) $$P(N_1 \cap N_2 \cap \overline{N_3} \cap \overline{N_4}) = P(N_1) * P(N_2 | N_1) * P(\overline{N_3} | N_1 \cap N_2) * P(\overline{N_4} | N_1 \cap N_2 ¿ N_3) =$$
$$\frac{7}{12} * \frac{9}{14} * \frac{5}{16} * \frac{7}{18} = \frac{365}{768}$$

2) $\binom{4}{2} * \frac{365}{768} \simeq 0{,}27$

LEGGE DELLE ALTERNATIVA

Avendo due eventi E, F con probabilità : 0 < P(F) < 1



$$E = (E \cap F) \cup (E \cap \overline{F})$$
$$P(E) = P(E \cap F) + P(E \cap \overline{F}) =$$
$$= P(E | F) * P(F) + P(E | \overline{F}) * P(F)$$

Questi sono eventi alternativi, cioè non possono verificarsi entrambi

Esercizi:

- In un laboratorio di analisi, l'esame del sangue è efficace nel 95% dei casi, nell'individuare una certa malattia. Con probabilità 0,01 l'esame indica come malata una persona sana.
Se lo 0,5% della popolazione è affetto dalla malattia, qual'è la probabilità che una persona positiva all'esame sia effettivamente malata?

F = {la persona ha la malattia} P(F) = 0,005 P($\overline{F}$) =0,995 P(E| F) = 0,95
E = {l'esame da esito positivo} P(E | $\overline{F}$) = 0,01

$$P(F | E) = P\frac{(F \cap E)}{P}(E) = \frac{P(F \cap E) * P(F)}{P(E|F) * P(F) * P(E|\overline{F}) * P(\overline{F})} = \frac{0{,}95 * 0{,}005}{0{,}95 * 0{,}005 + 0{,}01 * 0{,}995} \simeq 0{,}323$$

DILEMMA DEI PRIGIONIERI

3 prigionieri A, B, C fanno richieste per uscire dal carcere. I giudici decidono di rilasciare solo 2 dei 3 uomini e comunicano i nomi alla guardia.

Il prigioniero A ritiene che la sua probabilità di essere rilasciato è $\frac{2}{3}$.

Tuttavia, afferma che domandando alla guardia il nome di uno dei due progionieri (diverso da lui) che verrà rilasciato, allora la sua probabilità di uscire si ridurrà a $p=\frac{1}{3}$

È corretto il ragionamento di A?

Inizialmente: dopo:

(A, B) $p=\frac{1}{2}$ *? no perchè*:

(A, C)

(B, C)

$$P(A\,rilasciato \mid la\,guardia\,fa\,il\,nome\,di\,B)$$

$$P(A\,rilasciato \mid la\,indica\,B)=\frac{P((A,B)\,rilasciata\,e\,la\,guardia\,indica\,B)}{P((A,B)\,rilasciati\,e\,la\,guardia\,indica\,B)+P((B,C)\,rilasciati\,e\,la\,guardia\,indica\,B)}$$

$$(A,B)=\frac{1}{3}\xrightarrow{probabilità=1}(A,B)\,rilasciati\,e\,la\,guardia\,dice\,B\rightarrow P=\frac{1}{3}$$

$$(A,C)=\frac{1}{3}\xrightarrow{probabilità=1}(A,C)\,rilasciati\,e\,la\,guardia\,dice\,C\rightarrow P=\frac{1}{3}$$

$$(B,C)=\frac{1}{3}\xrightarrow{probabilità=1}\begin{pmatrix}(B,C)\,rilasciati\,e\,la\,guardia\,dice\,B\rightarrow P=\frac{1}{6}\\(B,C)\,rilasciati\,e\,la\,guardia\,dice\,C\rightarrow P=\frac{1}{6}\end{pmatrix}$$

$$=\frac{\frac{1}{3}}{\frac{1}{3}+\frac{1}{6}}=\frac{\frac{1}{3}}{\frac{3}{6}}=\frac{2}{3}\rightarrow la\,probabilità\,non\,è\in influenzata\,da\,quello\,che\,dice\,la\,guardia\,(antintuitivo).$$

LEGGE DELLE ALTERNATIVE (n alternative)

$F_1, F_2, \dots, F_n$ insieme completo di alternative su più eventi se:

1. $F_i \cap F_j = \emptyset \quad con\, i \neq j$
2. $F_1 \cup F_2 \cup \dots \cup F_n = S$
3. $P(F_i) > 0$

Siano $F_1, \dots, F_n$ un'insieme completo di alternativa, allora :

$$P(E)=P(E|F_1)P(F_1)+P(E|F_2)P(F_2)+\dots+P(E|F_n)P(F_n)=\sum_{i=1}^{n} P(E|F_i)P(F_i)$$

dim:

$$E=E\cap S=E\cap \bigcup_{i=1}^{n} \mathbf{F}_i = \bigcup_{i=1}^{n} E\cap F_i$$

$$P(E) = P\left(\bigcup_{i=1}^{n} E\cap F_i\right) = \sum_{i=1}^{n} P(E\cap F_i)=\sum_{i=1}^{n} P(E|F_i)P(F_i)$$

Esercizi

- Un'urna contiene tre monete.

1. La prima è equa.
2. La seconda mostra testa con probabilità *p*
3. La terza mostra testa con probabilità *1-p* dove o < p < 1

Si seglie a caso una moneta, qual' è la probabilità che mostri testa?

$F_j=\{ \text{ si seglie la moneta } j-esima \} \quad con\, j=1,2,3$

$T=\{ \text{ esce testa} \}$

$P(T)=P(T|F_1)P(F_1)+P(T|F_2)P(F_2)+P(T|F_3)P(F_3)=$

$=\frac{1}{2}*\frac{1}{3}+p*\frac{1}{3}+(1-p)*\frac{1}{3}=\frac{1}{3}*\{\frac{1}{2}+\cancel{p}+(1-\cancel{p})\}=\frac{1}{3}*\frac{3}{2}=\frac{1}{2}$

- Un algoritmo genera sequenze (c_1, ..., c_n) secondo le seguenti regole:

R1] c_i è un intero scelto a caso nell'inventario {1, 2, ..., n}

R2] c_i è diverso da c_{n-1} da i= 1, 2, ..., n

Calcolare le probabilità che una sequenza generata dall'algoritmo risulti:

1) c_1 = n 2) c_2= n 3) $c_1 \geq c_2 \geq ... \geq c_n$

Soluzioni:

$$P(c_1=n)=\frac{1}{n}$$

$$P(c_2=n)=\sum_{n=1}^{n} P(c_2=n|c_1=x)P(c_1=x)=\cancel{P(c_2=n|c_1=n)P(c_1=n)}+P(c_2=n|c_1=n-1)P(c_1=n-1)+...$$

$$+P(c_2=n|c_1=1)P(c_1=1)=\frac{1}{n-1}*\frac{1}{n}+\frac{1}{n-1}+\frac{1}{n}+...=\frac{(n-1)*1}{n-1}*\frac{1}{n}=\frac{1}{n}$$

$$P(c_1 \geq c_2 \geq ... \geq c_n)=\frac{1}{n*(n-1)^{n-1}}$$

TEOREMA DI BEYES

F_1, F_2, ..., F_n insieme completo di alternative con E, P(E) > 0

$$P(F_j|E)=\frac{P(E|F_j)P(F_j)}{\sum_{i=j}^{n} P(E|F_i)P(F_i)}$$

dim:

$$P(F_j|E)=\frac{P(E\cap F_j)}{P(E)}=\frac{P(E|F_j)P(F_j)}{\sum_{i=j}^{n} P(E|F_i)P(F_i)}=\sum_{j=1}^{n} P(F_j|E)=$$

$$=\frac{\sum_{j=1}^{n} P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}=1$$

Esercizi

- SPAM FILTER

w = {il messaggio contiene la parola lotteria}

S = {la mail è spam}

P(S | w) = $\frac{P(w|S)P(S)}{P(w|S)P(S)+P(w|\overline{S})P(\overline{S})}=\underbrace{\frac{P(w|S)}{P(w|S)+P(w|\overline{S})}}_{spamminess}$

IL PROBLEMA DELLE TRE CARTE

Il mazziere di una casa da gioco mette tre carte coperte sul tavolo. Una è asso e le altre due figure. Un giocatore sceglie una delle tre carte e scommette che sia un asso (p = 1/3).
Il mazziere guarda le carte e volta una delle due carte non scelte mostrando la figura. Calcolare la probabilità che ora la carte scelta dal giocatore sia l'asso.

A_k = {la k-esima carta è l'asso}   k = 1, 2 , 3
B = {il mazziere mostra che la terza carta è una figura dopo che il giocatore ha scelto la sceonda}

$$P(A_1|B)=\frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1)+P(B|A_2)P(A_2)+P(B|A_3)P(A_3)}=\frac{1*\frac{1}{3}}{1+\frac{1}{3}+\frac{1}{2}*\frac{1}{3}+0*\frac{1}{3}}=\frac{2}{3}$$

$$P(A_2|B)=\frac{1}{3}$$

$$P(A_3|B)=0$$

Esercizi:

- Uno studente deve sostenere un'esame. Se studia molto lo supera con probabilità 0,9. Se studia poco 0,5. Se non studia per niente 0,1.

- Lo studente decide di studiare molto con probabilità 2p.
- Lo studente decide di studiare poco con probabilità p.
- Lo studente decide di studiare per niente con probabilità 1-3p.

1. Determinare i valori ammissibili per p.
2. Calcolare la probabilità che supero l'esame prendendo p= 1/4.
3. Per p= 1/4, se lo studente supera l'esame qual'è la probabilità che abbia studiato poco.

S = {studente supera l'esame}   M = {Studente studia molto} P = {Studente studia poco}
N = {Studente studia per niente}

P(S|M) = 0,9   P(S|N) = 0,1   P(S|P) = 0,95 P(M) = 2p   P(P) = p   P(N) = 1 – 3p

1) $\left\{\begin{matrix} 0\leq 2p\leq 1 \\ 0\leq p\leq 1 \\ 0\leq 1-3p\leq 1 \end{matrix}\right\} \rightarrow 0\leq p\leq\frac{1}{3}$

2) ***P(M) = 1/2   P(P) = 1/4   P(N) = 1/4***

$$P(S)=P(S|M)P(M)+P(S|P)P(P)+P(S|N)P(N)=0,9*\frac{1}{2}+0,5*\frac{1}{4}+0,1*\frac{1}{4}=0,6$$

3) $P(P|S)=\frac{P(S|P)P(P)}{P(S)}=\frac{0,5*\frac{1}{4}}{0,6}\simeq 0,208\overline{3}$

- Si lanciano n monete non truccate
Per ogni monete che mostra testa si inserisce una biglia nera in un'urna e per ogni croce si inserisce una biglia bianca. Si estrae a caso una biglia bianca. Si estrae a caso una biglia dall'urna:
1) Calcolare la probabilità che sia nera
2) Se la biglia estratta è nera bisogna calcolare la probabilità che nell'urna vi erano k biglie nere

A={la biglia estratta è nera}

E_k={esce testa k volte}

1. **P(A)=** $\sum_{K=0}^{n} P=(A|E_k)P(E_k) = \sum_{K=0}^{n} \frac{k}{n} * \frac{\binom{n}{k}}{2^n} = \sum_{K=0}^{n} \frac{k}{n} * \frac{1}{2^n} * \frac{(n!)}{k!(n-k)!} =$
$\frac{1}{2^n} \sum_{K=0}^{n} \frac{1}{(k-1)!} * \frac{(n-1)!}{(n-k)!} = \frac{1}{2^n} \sum_{k=1}^{n} \binom{n-1}{k-1} = \frac{1}{2^n} \sum_{j=0}^{n-1} \binom{n-1}{j} = \frac{1}{2^n} * 2^{(n-1)} = \frac{1}{2}$

2. $P(E_k | A)$=per il teorema di Bayes = $\frac{P(AE_k)*P(E_K)}{P(a)} = \frac{\frac{\frac{K}{n}*\binom{n}{k}}{2^n}}{\frac{1}{2}} = \left\{ \begin{array}{c} 0 \, se \, k=0 \\ \frac{\binom{n-1}{k-1}}{2^{(n-1)}} se \, K \geq 1 \end{array} \right\}$

- Un sistema di gestione della posta elettronica riceve un messaggio, che si suppone sia spam con probabilità 0,7 e non spam con probabilità 0,3. Il sistema effettua un controllo su ogni messaggio ricevuto; se riceve un messaggio spam lo valuta come tale con probabilità 0,9 (e lo valuta come non spam con probabilità 0,1) mentre se riceve un messaggio non spam lo valuta come tale con probabilità0,8 (e lo valuta come spam con probabilità 0,2).
(i) Calcolare la probabilità che il sistema valuti come spam il messaggio ricevuto.
(ii) Se il sistema ha valutato come spam il messaggio ricevuto, qual `e la probabilità che invece sia non spam? (iii) Se il sistema ha valutato come non spam il messaggio ricevuto, qual `e la probabilità che invece sia spam?
Per le ipotesi sull'evento
F = {il messaggio `e spam} si ha P(F) = 0,7 $P(\overline{F}) = 0,3$.

Inoltre, posto E = {il sistema valuta come spam il messaggio ricevuto}
P(E | F) = 0,9 $\quad P(\overline{E} | F) = 0,1 \quad P(\overline{E} | \overline{F}) = 0,8 \quad P(E | \overline{F}) = 0,2$

(i) $P(E) = P(E | F) P(F) + P(E | \overline{F}) P(\overline{F}) = 0,9 \cdot 0,7 + 0,2 \cdot 0,3 = 0,69.$

(ii) $P(\overline{F} | E) = \frac{P(E|\overline{F})P(F)}{P(E)} = \frac{0,2*0,3}{0,69} = 0,0870$

(iii) $P(F | \overline{E}) = \frac{P(\overline{E}|F)P(\overline{F})}{P(\overline{E})} = \frac{0,1*0,7}{1-P(E)} = \frac{0,07}{0,31} = 0,2258$

EVENTI INDIPENDENTI

A, B si dicono indipendenti se:

1. $P(A \cap B) = P(A) * P(B)$
2. $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) * \cancel{P(B)}}{\cancel{P(B)}} = P(A)$
3. $P(B|A) = P(B)$ perchè $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} = \frac{\cancel{P(A)}P(B)}{\cancel{P(A)}} = P(B)$

Sono tutti equivalenti per richiedere l'indipendenza di due eventi.

ESERCIZI

- Lancio n monete non truccate.

  T_1 = {testa al primo lancio}
  U = {esce lo stesso risultato negli n lanci}
  A = {esce testa almeno 1 volta}

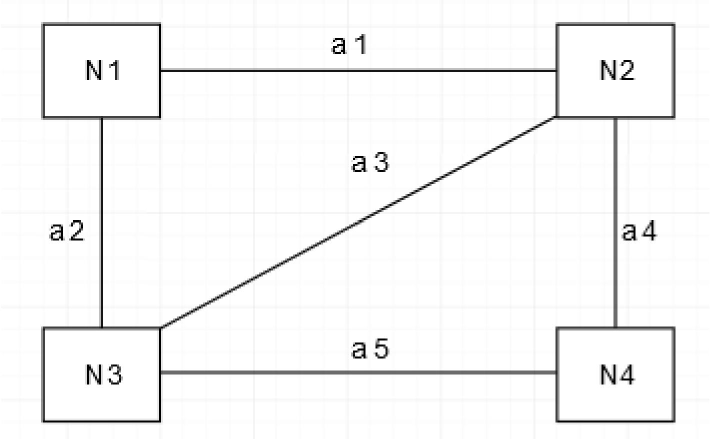
Mostrare che:

1. T_1 e U sono indipendenti
2. T_1 e A sono indipendenti
3. A e U sono indipendenti

Soluzioni:

1. $P(T_1 \cap U) = \frac{1}{2^n} = \frac{1}{2} * \frac{1}{2^{n-1}} = P(T_1) * P(U) \rightarrow T_1$ *e U sono indipendenti*
2. $P(A \cap T_1) = P(T_1) = \frac{1}{2} \neq 1 - \frac{1}{2^n}$ *non sono indipendenti*
3. $(A \cap U) = \frac{1}{2^n} \neq (1 - \frac{1}{2^n}) * \frac{1}{2^{n-1}}$ *sono indipendenti per* $n > 1$

- L'esperimento consiste nello segliere a caso indipendentemente un'arco ed un nodo del grafo.



1. Definire lo spazio campionario e la sua cardinalità.
2. A_1 = { l'arco scelto a caso è connesso ad N_1}
   B = {l 'arco e il nodo scelto a caso sono connessi}
   Studiare l'indipendenza di A e B.

Soluzioni:

1. $S=\{(N_{1,} j):1\leq i\leq 4, 1\leq j\leq 5\}$ $\quad |S|=4*5=20$
2. $A_1=\{(N_{1,} a_1)(N_{1,} a_2)(N_{2,} a_1)(N_{2,} a_2)(N_{3,} a_1)(N_{3,} a_2)(N_{4,} a_1)(N_{4,} a_2)\}$ $\quad P(A_1)=\frac{8}{20}=\frac{2}{5}$

   $B=\{(N_{1,} a_1)(N_{1,} a_2)(N_{2,} a_1)(N_{2,} a_2)(N_{2,} a_3)(N_{3,} a_2)(N_{3,} a_3)(N_{3,} a_5)(N_{4,} a_4)(N_{4,} a_5)\}$

   $$P(A_1)=\frac{10}{20}=\frac{1}{2}$$

   $A_1\cap B=\{\ (N_{1,} a_1)(N_{1,} a_2)(N_{2,} a_1)(N_{3,} a_2)\}=P(A_1\cap B)=\frac{4}{20}=\frac{1}{5}=$

   $=P(A)*P(B)=\frac{1}{5}\, sono\, indipendenti$

Proprietà
Se E ed F sono eventi indipendenti, allora E ed $\overline{F}$ sono indipendenti.

DIM:

$$F=(F\cap E)\cup(F\cap\overline{E})$$

$$P(E)=P(E\cap F)+P(E\cap\overline{F})=P(E)=P(E)P(F)+P(E\cap\overline{F})=$$

$$=P(E\cap\overline{F})=P(E)-P(E)P(F)=P(E)[1-P(F)]=P(E)P(\overline{F})=E\, e\, \overline{F}\, sono\, indipendenti$$

ESERCIZI:

Consideriamo i seguenti eventi nel lancio di due dadi non truccati: $E=\{$la somma dei dadi è 7$\}$, $F=\{$il primo dado dà 4$\}$, $G=\{$il secondo dado dà 3$\}$. Esaminare l'indipendenza delle coppie di eventi E ed F, E e G, E ed $F\cap G$.

E ={ (1, 6) (2, 5) (3, 4) (4, 3) (5, 2) (6, 1)} = $\frac{6}{36}=\frac{1}{6}$

F ={(4, 1) (4, 2) (4, 3) (4, 4) (4, 5) (4, 6)} = $\frac{1}{6}$

G ={(1, 3) (2, 3) (3, 3) (4, 3) (5, 3) (6, 3)} = $\frac{1}{6}$

Soluzioni:

1. $P(E\cap F)=\frac{1}{36}=P(E)P(F)=\frac{1}{6}*\frac{1}{6}=sono\, indipendeti$
2. $P(E\cap G)=\frac{1}{36}=P(E)P(G)=\frac{1}{6}*\frac{1}{6}=sono\, indipendeti$
3. $P(E\cap(F\cap G))=\frac{1}{36}\neq P(E)P(F\cap G)=\frac{1}{6}*\frac{1}{36}=non\, sono\, indipendeti$

   $$P(F\cap G)=\frac{1}{36}$$

A,B,C si dicono indipendenti se:

1. P($A\cap B$)=P(A) * P(B)
2. P($A\cap C$)=P(A) * P(B)
3. P($B\cap B$)=P(A) * P(B)
4. P($A\cap B\cap C$)=P(A) * P(B) * P(C)

Esercizi

-In un esperimento si genera una sequenza di 3 bit
A={i tre bit sono uguali}
B={almeno un bit è pari a 1}
C={almeno due bit sono 0}
Studiare l'indipendenza.

$$S=\begin{pmatrix} 111 \\ 110 \\ 101 \\ 100 \\ 011 \\ 010 \\ 001 \\ 000 \end{pmatrix} \qquad S=\{(w_1,w_2,w_3),\ w_i \in \{0,1\}\ \}$$

$|S|=2^3$

$P(A)=\frac{2}{8}=\frac{1}{4}$ $\qquad P(A\cap B)=\frac{1}{8}\overset{?}{=}P(A)*P(B)=\frac{1}{8}*\frac{7}{8}=\frac{7}{32}$ quindi non sono indipendenti

$P(B)=\frac{7}{8}$ $\qquad P(A\cap C)=\frac{1}{8}\overset{?}{=}P(A)*P(C)=\frac{1}{4}*\frac{1}{2}=\frac{1}{8}$ ✓

$P(C)=\frac{4}{8}=\frac{1}{2}$ $\qquad P(B\cap C)=\frac{3}{8}$

$P(A\cap B\cap C)=0$

-Lanciamo 2 monete eque
T_1={testa al primo lancio}
T_2={testa al secondo lancio}
U={stesso esito nei due lanci}
Studiare l'indipendenza.

S={(t,t) , (t,c), (c,t), (c,c)}

|S|= 4

T_1={(t,t), (t,c)} $\qquad P(T_1)=\frac{2}{4}=\frac{1}{2}$

T_2={(t,t), (c,t)} $\qquad P(T_2)=\frac{1}{8}=\frac{1}{2}$

U={(t,t), (c,c)} $\qquad P(U)=\frac{2}{4}=\frac{1}{2}$

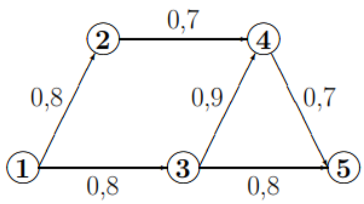
$P(T_1 \cap T_2) = \frac{1}{4} = P(T_1) * P(T_2) = \frac{1}{2} * \frac{1}{2}$ ✓

$P(T_1 \cap U) = \frac{1}{4} = P(T_1) * P(U) = \frac{1}{2} * \frac{1}{2}$ ✓

$P(T_2 \cap U) = \frac{1}{4} = P(T_2) * P(U) = \frac{1}{2} * \frac{1}{2}$ ✓

$P(T_1 \cap T_2 \cap U) = \frac{1}{2} * \frac{1}{2} * \frac{1}{2} = \frac{1}{8}$ ✘ quindi non sono indipendenti

-Si consideri un sistema costituito da 5 unità collegate mediante archi orientati.
Su ogni arco c'è la probabilità che l'arco sia attivo indipendentemente dagli altri. Calcola la probabilità che vi sia almeno un percorso da 1 a 5 con tutti gli archi attivi.



A={gli archi del percorso 1234 sono attivi}
B={gli archi del percorso 1345 sono attivi}
C={gli archi del percorso 135 sono attivi}

$P(A \cup B \cup C) = P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)$
con $P(A) = 0{,}8\cdot 0{,}7\cdot 0{,}7 = 0{,}392$; $P(B) = 0{,}8\cdot 0{,}9\cdot 0{,}7 = 0{,}504$; $P(C) = 0{,}8\cdot 0{,}8 = 0{,}64$;
$P(A\cap B) = 0{,}8\cdot 0{,}7\cdot 0{,}7\cdot 0{,}8\cdot 0{,}9 = 0{,}2822$; $P(A\cap C) = 0{,}8\cdot 0{,}7\cdot 0{,}7\cdot 0{,}8\cdot 0{,}8 = 0{,}2509$;
$P(B\cap C) = 0{,}8\cdot 0{,}9\cdot 0{,}7\cdot 0{,}8 = 0{,}4032$; $P(A\cap B\cap C) = 0{,}8\cdot 0{,}7\cdot 0{,}7\cdot 0{,}8\cdot 0{,}9\cdot 0{,}8 = 0{,}2258$;
ne segue: $P(A \cup B \cup C) = 0{,}8255$.
$E_1,\dots,E_n$ si dicono indipendenti se per ogni sottoinsieme $E_{i1}, E_{i2},\dots,E_{ir}$ $2\le r \le n$ è costituito da eventi indipendenti

es.
E_1,E_2,E_3,E_4:

- verificare indipendenza a coppie $P(E_i \cap E_j) = P(E_i)*P(E_j)$ $1\le i,j \le 4$ $i\neq j$
- verificare indipenza a tre $P(E_i \cap E_j \cap E_k) = P(E_i) * P(E_j) * P(E_k)$ $1\le i,j,k \le 4$ $i\neq j\neq k$
- verificare $P(E_1 \cap E_2 \cap E_3 \cap E_4) = P(E_1) * P(E_2) * P(E_3) * P(E_4)$

in totale bisogna fare $\binom{4}{2} + \binom{4}{3} + \binom{4}{4} = 6+4+1 = 11$ controlli.

PROVE INDIPENDENTI

Un singolo esperimento è uguale ad una successione di sotto-esperimenti. Quando quest'ultimi hanno lo stesso spazio e stessa funzione di probabilità si dicono PROVE. Se qualsiasi esiti di un qualsiasi gruppo di sottoesperimenti non hanno effetti sugli esiti degli altri eìsottoesperimenti diciamo che questi sottoesperimenti sono indipenenti.
Se i sottoesperimenti sono indipendenti allora gli eventi E_1, E_2, ... E_i sono indipendenti dove E_i è un evento che dipende solo dall'esito dell'i-esimo esperimento.

ESERCIZI

-Nel lancio di 3 monete non truccate, consideriamo i seguenti eventi:
A ={testa al terzo lancio}
B = {croce al secondo lancio}
Studiare l'indipendenza di A, B.

$$S=\{\ (i,i), i\in\{t,c\}\ \}=8$$

$$P(A)=\frac{4}{8}=\frac{1}{2} \qquad P(B)=\frac{4}{8}=\frac{1}{4} \qquad P(A\cap B)=\frac{2}{8}=\frac{1}{4} \qquad A\, e\, B\, sono\, indipendenti$$

Usando le prove abbiamo:

Visto che la probabilità di A dipende solo dal terzo lancio e quella di B dal secondo, abbiamo che A B sono indiepndeti perchè non sono influite dallo stesso evento.

-Estraiamo con reinserimento 2 biglie da un'urna che contiene :

- 1 biglia azzurra
- 2 biglie bianche
- 3 biglie rosse

A = {le due biglie estratte non sono rosse}
B = {almeno 2 delle 2 è bianca}

$$P(A)=\frac{3}{6}*\frac{3}{6}=\frac{1}{4} \qquad P(B)=1-(\frac{4}{6}*\frac{4}{6})=\frac{5}{9}$$

$$P(A\cap B)=2*\frac{1}{6}*\frac{2}{6}+\frac{2}{6}*\frac{2}{6}=\frac{2}{9}\neq P(A)*P(B) \qquad A\, e\, B\, non\, sono\, indipendenti$$

Usando le prove capiamo che non sono indipendenti perchè A e B dipendono dalla stesse prove.

-Un'urna contien 4 biglie numerate da 1 a 4. Si estraggono 3 biglie senza reinserimento.
A = {esattamente 1 delle 3 estratte ha il numero 1 o numero 2}
B = {le tre biglie estratte hanno i numeri 1, 2, 3 oppure 2, 3, 4}

$$P(A)=\frac{\binom{2}{1}*\binom{2}{2}}{\binom{4}{3}}=\frac{2}{4}=\frac{1}{2} \qquad P(B)=\frac{2}{\binom{4}{3}}=\frac{2}{4}=\frac{1}{2} \qquad P(A\cap B)=\frac{1}{4}$$

A, B sono indipendenti anche se le loro prove non sono indipendenti.

- In una sequenza infinita di prove indipendenti ogni prova ha 2 esiti: successo con probabilità **p** e insuccesso con probabilità **1 − p**. Qual `e la probabilità che

a) vi sia almeno un successo nelle prime n prove;
b) vi siano esattamente k successi nelle prime n prove;
c) tutte le prove abbiano successo?

Soluzioni:

(a) Ponendo $E_i = \{\text{si ha successo alla prova } i\text{-esima}\}$, $i = 1, 2, \ldots$, con $P(E_i) = p$ e $P(\overline{E}_i) = 1 - p$, la probabilità richiesta è

$$P\left(\bigcup_{i=1}^{n} E_i\right) = P\left(\overline{\bigcap_{i=1}^{n} \overline{E}_i}\right) = 1 - P\left(\bigcap_{i=1}^{n} \overline{E}_i\right) = 1 - \prod_{i=1}^{n} P(\overline{E}_i) = 1 - (1-p)^n.$$

(b) Consideriamo le sequenze di prove che nei primi n esiti hanno k successi e $n-k$ insuccessi. Ognuna di esse si realizza, per l'indipendenza, con probabilità $p^k(1-p)^{n-k}$. Ad esempio

$$P(E_1 \ldots E_k \overline{E}_{k+1} \ldots \overline{E}_n) = P(E_1) \ldots P(E_k) P(\overline{E}_{k+1}) \ldots P(\overline{E}_n) = p^k (1-p)^{n-k}.$$

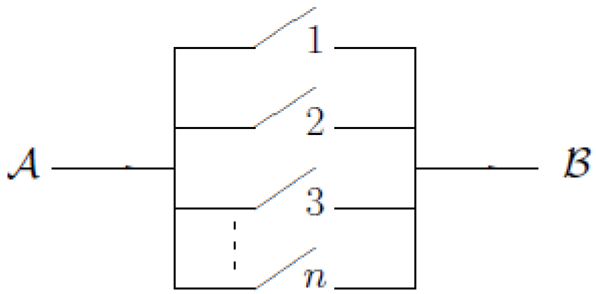
Vi sono $\binom{n}{k}$ sequenze di questo tipo poiché vi sono $\frac{n!}{k!(n-k)!}$ permutazioni distinte di k successi e $n-k$ insuccessi, quindi la probabilità richiesta è

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

(c) La probabilità di avere n successi nelle prime n prove è, per l'indipendenza,

$$P\left(\bigcap_{i=1}^{n} E_i\right) = \prod_{i=1}^{n} P(E_i) = p^n.$$

-Un sistema formato da n componenti `e detto in parallelo se esso funziona quando almeno uno dei suoi componenti funziona. Il componente i-esimo, indipendentemente dagli altri, funziona con probabilit`a pi, i = 1, 2, . . . , n. Qual `e la probabilit`a che il sistema funzioni?



Soluzione:

Sia $E_i = \{\text{il componente } i\text{-esimo funziona}\}$. Allora

$$P(F_p) = P(\text{il sistema funziona}) = P\left(\bigcup_{i=1}^{n} E_i\right) = 1 - P\left(\bigcap_{i=1}^{n} \overline{E}_i\right) = 1 - \prod_{i=1}^{n} P\left(\overline{E}_i\right) = 1 - \prod_{i=1}^{n} (1 - p_i)$$

per l'indipendenza.

- Lancio n volte una moneta truccata.

T_k = {al k-esimo lancio esce testa} con $1 \leq k \leq n$

$P(T_k) = p$

Se i lanci sono indipendenti, calcolare le probabilità degli eventi :

A_k = {nei primi k lanci esce testa}

A'_k = {nei primi k lanci esce croce}

B_k = {esce testa k volte}.

$$P(A_k) = P(T_1 \cap T_2 \cap \ldots \cap T_k) = P(T_1)\, P(T_2) \ldots P(T_k) = p^k$$

$$P(A'_k) = (1-p)^k$$

$$P(B_k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Assiomi applicati sulla probabilità condizionata

1)

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad \text{se } P(B)>0$$

2)

$$P(S \mid B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

3)

$0 \leq P(A \mid B) \leq 1 \rightarrow \dfrac{P(A \cap B)}{P(B)} \geq 0$ perche $P(B)>0$

$\dfrac{P(A|B)}{P(B)} \leq 1$ perché $(A \cap B) \subseteq B \rightarrow P(A \cap B) \leqslant P(B)$

4)

$A_1, A_2, A_3, \ldots, A_j = 0 \quad i \neq j$

$$P\left(\bigcup_{j=1}^{\infty} A_j \mid B\right) = \frac{P\left(\bigcup_{j=1}^{\infty} A_j \mid B\right) \cap B}{P(B)} = \frac{P\left(\bigcup_{j=1}^{\infty} A \cap B\right)}{P(B)}$$ $\rightarrow$ per la proprietà dell'additività numerabile

$$= \frac{\sum_{J=1}^{\infty} P(A_j \cap B)}{P(B)} = \sum_{J=1}^{\infty} \frac{P(A_j \cap B)}{P(B)} = \sum_{J=1}^{\infty} P(A_j|B)$$

5) $P(\overline{A}|B) = 1 - P(A|B)$

6) $P(A \cup C|B) = P(A \mid B) + P(C \mid B) - P(A \cap C|B)$

7) $P(A \cup C) = P(A) + P(C) - P(A \cap C)$

ESERCIZIO

Uno studente ha intensione di dare un esame sostenendo 3 prove intermedie. Se una di queste non è superata, non può sostenere la successiva

P(prima prova)=0,9

P(seconda prova)= 0,8

P(terza prova)= 0,7

1) Calcola la probabilità che superi tutte e 3 le prova

2) Se lo studente non ha superato una delle prove, calcola la probabilità che abbia fallito la seconda.

A_i ={lo studente supera la prova i-esima} i=1,2,3

$P(A_1)= 0,9$

$P(A_2 \mid A_1)=0,8$

$P(A_3 \mid (A_2 \cap A_1))=0,7$

1)

$P(A_1 \cap A_2 \cap A_3)$= per la regola del prodotto= $P(A_1)*P(A_2 \mid A_1)* P(A_3 \mid A_2 \cap A_1) = 0,9 * 0,8 * 0,7 = 0,504$

2)

$$P(\overline{A_2}|\overline{A_1}\cup\overline{A_2}\cup\overline{A_3})=\frac{P(\overline{A_2}\cap(\overline{A_1}\cup\overline{A_2}\cup\overline{A_3}))}{P(\overline{A_1}\cup\overline{A_2}\cup\overline{A_3})}=\frac{P(\overline{A_2}\cap(\overline{A_1}\cup\overline{A_2}\cup\overline{A_3}))}{1-P(A_1\cup A_2\cup A_3)}=\frac{P(\overline{A_2})}{1-P(A_1\cup A_2\cup A_3)}=$$

$$\frac{P(\overline{A_2}|A_1)*P(A_1)*P(\overline{A_2}|\overline{A_1})*P(A_1)}{1-P(A_1\cup A_2\cup A_3)}=\frac{1-P(A_2|A_1)P(A_1)}{1-0{,}504}=\frac{(1-0{,}8)*0{,}9}{0{,}496}=0{,}362$$

DEFINIZIONE

Due eventi sono incondizionatamente indipendenti dato un evento F se:

- $P((A\cap B)|F)$ = P(A| B)* P(B | F)

ESERCIZIO

Un’urna contiene 3 monete. La prima mostra testa con probabilità $\frac{1}{2}$;

La seconda con probabilità $\frac{1}{4}$; La terza con probabilità $\frac{3}{4}$;

(i) Se si lancia una moneta scelta a caso, qualè la probabilità che esca testa?

(ii) Se la moneta scelta dà testa al primo lancio, qual `e la probabilità che la stessa dia testa anche al secondo lancio?

T_i = {esce testa al lancio i-esimo}, i = 1, 2.

(F1)={scelgo la prima moneta};

(F2)={scelgo la seconda moneta};

(F3)={scelgo la terza moneta};

(i) Per la formula delle alternative si ha: $P(T_1) = \sum_{j=1}^{3} P(T_1|F_j)P(F_j)=\frac{1}{2}\cdot\frac{1}{3}+\frac{1}{4}\cdot\frac{1}{3}+\frac{3}{4}\cdot\frac{1}{3}=\frac{1}{2}$

(ii) Se la moneta scelta dà testa al primo lancio, la probabilità che la stessa dia testa anche al secondo lancio è data da P(T2|T1) = P(T1 ∩ T2) =

$$P(T_1\cap T_2|F_1)*P(F_1)+P(T_1\cap T_2|F_2)*P(F_2)+P(T_1\cap T_2|F_3)*P(F_3)=$$
$$=P(T_1|F_1)*P(T_2|F_1)*P(F_1)+P(T_1|F_2)*P(T_2|F_2)*P(F_2)+P(T_1|F_3)*P(T_2|F_3)*P(F3)=$$
$$=\left(\frac{1}{2}*\frac{1}{2}*\frac{1}{3}\right)*\left(\frac{1}{4}*\frac{1}{4}*\frac{1}{3}\right)+\left(\frac{3}{4}*\frac{3}{4}*\frac{1}{3}\right)=\frac{7}{24}\rightarrow P(T_2|T_1)=\frac{\left(\frac{7}{24}\right)}{\left(\frac{1}{2}\right)}=\frac{7}{12}\simeq 0{,}583$$

VARIABILI ALEATORIE

S, F, P

$X : S \rightarrow R$

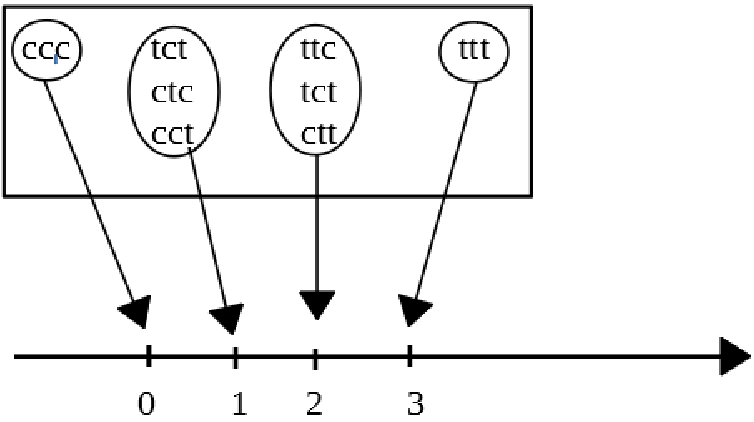
$\quad \omega \rightarrow X(\omega)$

t.c. $\{\omega \in S : X(\omega) \leq x\} \in F \quad \forall x \in R$

ESERCIZIO

Lancio 3 monete eque

y= numero di volte che esce testa



$P(Y=0)= P(\{ccc\})= \frac{1}{8}$

$P(Y=1)=P(\{(tcc),(ctc),(cct)\})=\frac{3}{8}$

$P(Y=2)= P(\{(ttc),(tct),(ctt)\})=\frac{3}{8}$

$P(Y=3)=P(\{ttt\})=\frac{1}{8}$

$$P\left(\bigcup_{i=0}^{3} (Y=i)\right)=1$$

$P(1\leq Y\leq 2)=P(Y=1)+P(Y=2)=\frac{6}{8}$

ESERCIZIO

Si estraggono 2 biglie da un’urna che ne contiene 8 bianche, 4 nere e 2 gialle. Si vincono 2 euro per ogni per ogni biglia nera estratta e se ne perdono 1 ogni bianca estratta

X= vincita

$(b,b) \rightarrow X=-2 \qquad (n,g) \rightarrow X=2$

$(b,g) \rightarrow X=-1 \qquad (n,n) \rightarrow X=4$

$(b,n) \rightarrow X= 1 \qquad (g,g) \rightarrow X=0$

$X\in\{-2,-1,0,1,2,4\}$

$$P(X=-2) = P(b,b) = \frac{\binom{8}{2}}{\binom{14}{2}} = \frac{28}{91} \qquad P(X=1) = P(b,n) = \frac{\binom{8}{2}*\binom{4}{1}}{\binom{14}{2}} = \frac{32}{91}$$

$P(X=-0) = P(b,g) = \dfrac{\binom{8}{1} * \binom{2}{1}}{\binom{14}{2}} = \dfrac{16}{91}$

$P(X=2) = P(n,g) = \dfrac{\binom{4}{1} * \binom{2}{1}}{\binom{14}{2}} = \dfrac{8}{91}$

$P(X=0) = P(g,g) = \dfrac{\binom{2}{2}}{\binom{14}{2}} = \dfrac{1}{91}$

$P(X=4) = P(n,n) = \dfrac{\binom{4}{2}}{\binom{14}{2}} = \dfrac{6}{91}$

ESERCIZIO

Si estraggono 3 biglie a caso senza reinserimento da un'urna contenente 20 biglie numerate da 1 a 20. Calcolare la probabilità che almeno una tra le biglie estratte abbia un numero maggiore o uguale 17.

X= massimo tra i 3 numeri estratti

$P(x\geq 17)$ $P(x=17)+P(x=18)+P(x=19)+P(x=20)$

$P(x=17)$ $\dfrac{\binom{16}{2}}{\binom{20}{3}}+\dfrac{\binom{17}{2}}{\binom{20}{3}}+\dfrac{\binom{18}{2}}{\binom{20}{3}}+\dfrac{\binom{19}{2}}{\binom{20}{3}}$

ESERCIZIO

Si lancia ripetutamente una moneta truccata.

In un singolo lancio da testa con probabilità p

Si lancia finché non appare testa per la prima volta oppure si siano fatti n lanci

X= numero di lanci effettuati

$X\in\{1,2,\dots,n\}$

$P(x=1)=P(T_1)=p$

$P(x=2)= P(\overline{T}_1\cap \overline{T}_2)=(1-P_1)*p$

$P(X=3)= P(\overline{T}_1\cap\overline{T}_2\cap\overline{T}_3)=(1-P_1)^2*p$

se $1\leq k \leq n-1$ $P(x=k)=(1-p)^{k-1} *p$

se k=n $P(x=n)=$

$$\underbrace{P(\overline{T}_1\cap\overline{T}_2\cap\dots\cap T_n)}_{caso\ in\ cui\ è\ uscita\ testa\ all'ultimo\ lancio} + \underbrace{P(\overline{T}_1\cap\overline{T}_2\cap\dots\cap\overline{T}_n)}_{caso\ in\ cui\ non\ esce\ testa\ ma\ mi\ fermo\ perchè\ ho\ fatto\ n\ lanci} =(1-p)^{(n-1)}*p=$$

$$=(1-p)^n=(1-p)^{(n-1)}*[\cancel{p}+1-\cancel{p}]=(1-p)^{(n-1)}\ .$$

FUNZIONE DI DISTRIBUZIONE

$F_x(x)=P(\{\omega\in S: X(\omega)\leq x\})=P(X\leq x), x\in R$

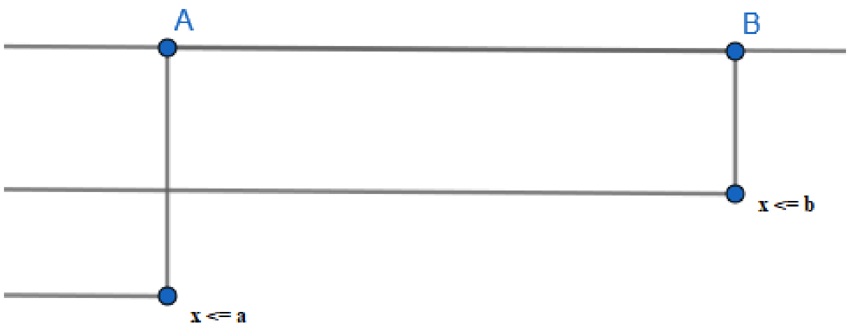
$F_x: R\rightarrow[0,1]$

Proprietà della funzione di distribuzione:

1) $F_x(x)$ è non decrescente, presi $a, b \in R, a < b \Rightarrow F_x(a) \le F_x(b)$

Dim:

$F_x(b) = P(X \le b)$



$(X \le b) = (X \le a) \cup (a < X \le b)$ sono due eventi disgiunti quindi
$P(X \le b) = P(X \le a) + P(a < X \le b) = F_x(b) = F_x(a) + P(a < X \le b) \ge F_x(a)$
perchè $P(a < X \le b)$ è sempre positiva, ricaviamo anche che:

$$P(a < X \le b) = F_x(b) - F_x(a)$$

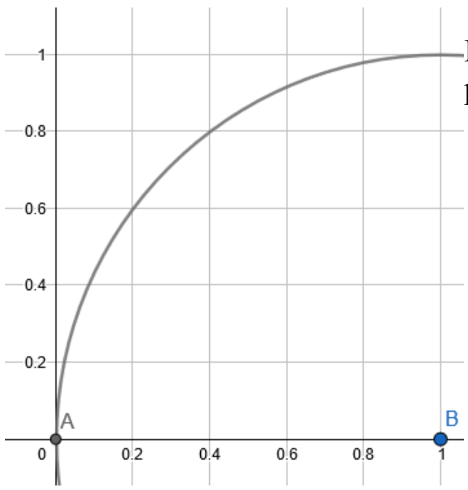
2) $\lim_{x \to +\infty} F_x(x+h) = F_x(x) = 1$ invece $\lim_{x \to -\infty} F_x(x+h) = F_x(x) = 0$

Dim:

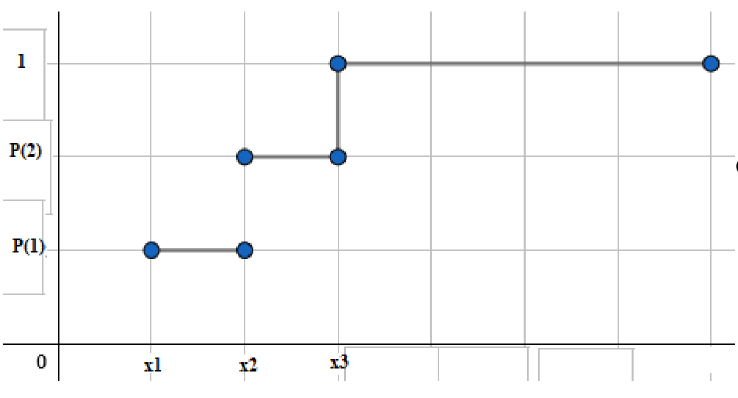
$\lim_{x \to +\infty} F_x(x) = \lim_{x \to +\infty} P(\{\omega \in S : X(\omega) \le x\})$ intendiamo trovare un omega che sia un generico numero reale e quindi ila soluzione è P(S) = 1.
$\lim_{x \to -\infty} F_x(x) = \lim_{x \to -\infty} P(\{\omega \in S : X(\omega) \le x\})$ si riduce all'evento impossibile perché richiediamo un valore di omega più piccolo di $-\infty$ quindi $P(\varnothing) = 0$.

3) $\lim_{h \to 0^+} F_x(x+h) = F_x(x)$ $F_x(x)$ è continua da destra.

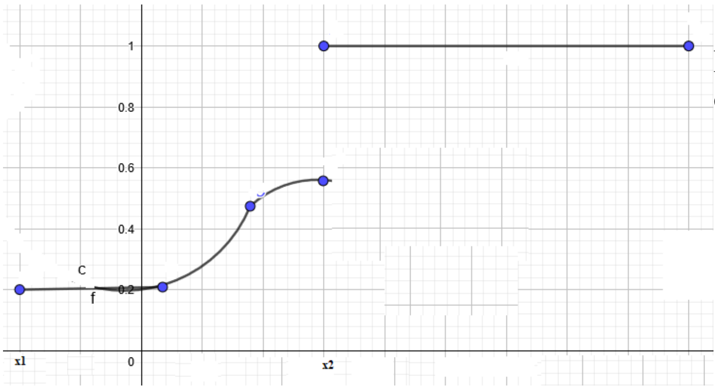
Esercizi:

- 

Funzione di distribuzione perché è continua e ha tutte e tre le proprietà.

- 

Funzione di distribuzione perché è continua e ha tutte e tre le proprietà.

- 

Funzione di distribuzione perché è continua e ha tutte e tre le proprietà.

Fino ad ora abbiamo visto che:

- $P(X \leq x) = F_x(x)$
- $P(X < x) = \lim_{h \to o^-} F_z(x+h) \leq \lim_{h \to o^+} F_z(x+h)$ per la terza proprietà, vediamo quindi che il limite è $F_x(x)$ possiamo dire che $P(X < x) = F_x(x^-) \leq F_x(x)$ quindi $P(X = x) = P(X \leq x) - P(X < x) = F_x(x) - F_x(x^-)$

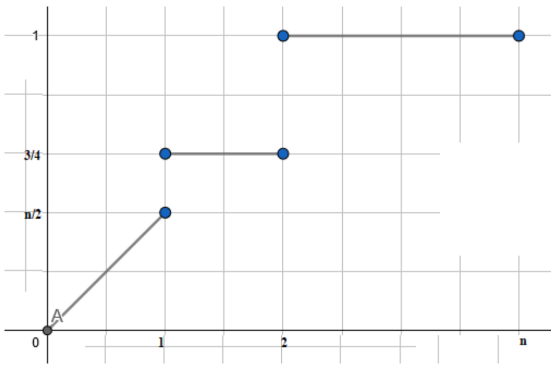
Per le variabili continue val 0 per le discontinue vale maggiore di 0.

Altre proprietà della funzione di distribuzione

1. $P(X = x) = F_x(x) - F_x(x^-)$
2. $P(a < X \leq b) = F_c(b) - F_x(a)$ è anche uguale a $P(x \leq b) - P(x \leq a)$
3. $P(a < X < b) = F_x(b^-) - F_x(a)$
4. $P(a \leq X < b) = F_x(b^-) - F_x(a^-)$
5. $P(a \leq X \leq b) = F_x(b) - F_x(a^-)$
6. $P(x > a) = 1 - P(x \leq a) = 1 - F_x(a)$
7. $P(x \geq a) = 1 - P(x < a) = 1 - F_x(a^-)$

Esercizi:

- $F_x(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2} & 0 \leq x < 1 \\ \frac{3}{4} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$



Calcolare:

a) $P(X < 2) = F_x(2^-) = \frac{3}{4}$

b) $P(X = 2) = F_x(1) - F_x(1^-) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$

c) $P(X>1)=1-F_x(\frac{1}{2})=1-\frac{\frac{1}{2}}{2}=1-\frac{1}{4}=\frac{3}{4}$

d) $P(1<X\leq 3)=F_x(3)-F_x(1)=1-\frac{3}{4}=\frac{1}{2}$

e) $P(1\leq X\leq 2)=F_x(2)-F_x(1^-)=1+\frac{1}{2}=\frac{1}{2}$

f) $P(X>1|X>\frac{1}{2})=\frac{P((x>1)\cap(x>\frac{1}{2}))}{P(x>\frac{1}{2})}=\frac{P(x>1)}{P(x>\frac{1}{2})}=\frac{1-F_x(1)}{1-F_x(\frac{1}{2})}=\frac{1-\frac{3}{4}}{1-\frac{1}{4}}=\frac{1}{3}$

VARIABILI ALEATORIE DISCRETE

X si dice discreta se assume un numero finito di valori oppure al più un infinita numerabile.
Per variabili discrete abbiamo:

$$P(X=k)=\underbrace{P_x(k)}_{\text{densità discreta di } X}\quad ,k\text{ è un intero}$$

$$S_x=\{\ x_1,x_2,x_3,\dots\}\ \textit{insieme dei valori che assume } X$$

$P_x(x_i)\geq 0$ mentre $P_x(x)=0\ \forall\ x\neq x_1,x_2,\dots$

Esempio:

X = esito del lancio del dado
S_x = {1, 2, 3, 4, 5, 6}

$P_x(1)=P(x=1)=\frac{1}{6}$ $\quad$ $P_x(2)=P(x=2)=\frac{1}{6}$

$P_x(3)=P(x=3)=\frac{1}{6}$ $\quad$ $P_x(4)=P(x=4)=\frac{1}{6}$

$P_x(5)=P(x=5)=\frac{1}{6}$ $\quad$ $P_x(6)=P(x=6)=\frac{1}{6}$

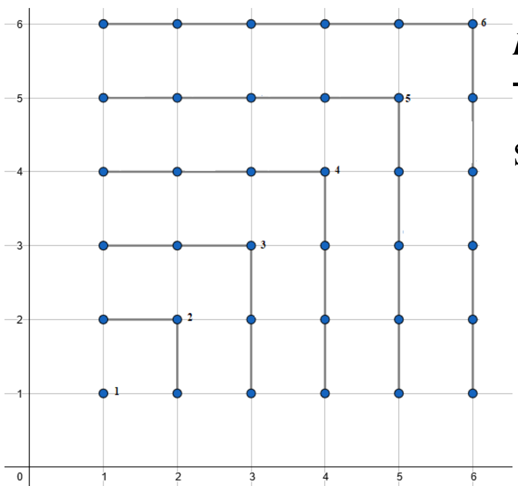
$P_x(8)=P(x=8)=0$ $\quad$ quindi $\quad$ $P_x(x)=0\ \forall\ x\notin S_x$

Se sommiamo tutte le probabili abbiamo che la loro somma ci da 1 per: $\sum_{i=1}^{6}P_x(i)$.

La $\sum_{i=1}P_x(i)=1$ è chiamata condizione di normalizzazione.

Per le variabili aleatorie discrete possiamo calcolare anche $P_x(x_i)$ sugli intervalli:

- $P(a\leq X\leq b)=\sum_{k:a\leq x_k\leq b}P_x(x_k)$
- $P(x\in b)=\sum_{k:x_k b}P_x(x_k)$



Esercizi:
-X = massimo che si ottiene lanciando 2 dadi

S_x = {1, 2, 3, 4, 5, 6}

$$P(x=1)=\frac{1}{36}$$
$$P(x=2)=\frac{3}{36}$$
$$P(x=3)=\frac{5}{36}$$
$$P(x=4)=\frac{7}{36}$$
$$P(x=5)=\frac{9}{36}$$
$$P(x=6)=\frac{11}{36}$$

$$P_x(k)=\frac{2k-1}{36} \qquad k=1,\dots,6$$

Calcolare anche $P(4\leq X\leq 6)=P_x(4)+P_x(5)+P_x(6)=\frac{7+9+11}{36}=\frac{27}{36}$

- X è una variabile discreta con densità discreta $P_x(k)=\frac{\lambda^k}{k!}e^{-\lambda} \quad con\,\lambda>0\, e\, k=0,1,2,3,\dots$

Calcolare:

1. $P(x=0)=P_x(0)=\frac{\lambda^0}{0!}e^{-\lambda}=e^{-\lambda}$
2. $P(x>2)=\sum_{k=2}^{+\infty}P_x(k)=1-P(x\leq 2)=1-\{P(x=0)+P(x=1)+P(x=2)\}=$
$=1-\{\,P_x(0)+P_x(1)+P_x(2)\}=1-\{\,e^{-\lambda}+\lambda e^{-\lambda}+\frac{\lambda^2}{2}e^{-\lambda}\}$

FUNZIONE DI DISTRIBUZIONE PER VARIABILI ALEATORIE DISCRETE

X discreta, $S_X=\{x_1, x_2, x_3, \dots\}$ con $x_1<x_2<x_3<\dots$ indichiamo con $P_x(x)$ la sua densità discreta:

$$F_x(x)=P(x\leq x)=\sum_{k:x_k\leq x}P_x(x_k)$$

$$F_x(x)=\left\{\begin{array}{ll} 0 & x<x_1 \\ P_x(x_1) & x_1\leq x<x_2 \\ P_x(x_1)+P_x(x_2) & x_2\leq x<x_3 \\ \quad\vdots & \end{array}\right\}$$

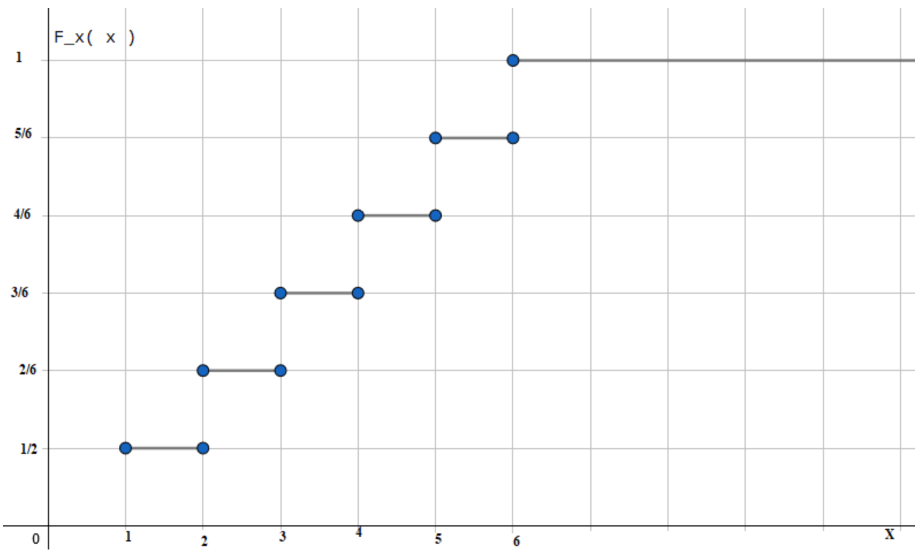
abbiamo 2 possibili schemi finali:

1) Se $S_x=\{x_1, x_2, x_3, \dots, x_n\}$ allora l'ultimo valore sarà $1 \qquad x\geq x_n$
2) altrimenti la sequenza è infinita

Esercizio:
X = esito del lancio del dado

$S_x = \{1, 2, 3, 4, 5, 6\}$ $\quad P_x(x)=\frac{1}{6} \quad \forall x \in S_x$

$$F_x(x)=\begin{cases} 0 & x<1 \\ P_x(1)=\frac{1}{6} & 1 \le x < 2 \\ P_x(1)+P_x(2)=\frac{2}{6} & 2 \le x < 3 \\ \frac{3}{6} & 3 \le x < 4 \\ \frac{4}{6} & 4 \le x < 5 \\ \frac{5}{6} & 5 \le x < 6 \\ \frac{6}{6}=1 & x \ge 6 \end{cases}$$



VALORE ATTESO
X discreta, con densità discreta $P_x(X)$, allora è il valore atteso e si indica con E(X).

$$E(X)=\sum_{x:P_x(X)>0} x * P_x(X)$$

Esempio:
Abbiamo una variabile x, $S_x = \{0, 1\}$ con $P_x(0)= 1 / 2$ e $P_x(1)= 1/ 2$

$$E(X)=0*P_{(X)}+1*P_{(X)}=0*\frac{1}{2}+1*\frac{1}{2}=\frac{1}{2}$$

$$-x \in S_x=\{0,1\} \quad \text{con } P_x(0)=1/3 \quad \text{e} \quad P_x(1)=2/3$$

$$E(X)=0*\frac{1}{3}+1*\frac{2}{3}=\frac{2}{3}$$

Esercizi:

- Calcolare il valore atteso di X = esito del lancio del dado

S_x={1, 2, , 4, 5, 6}

$P_x(1)$= 1/6 = $P_x(i)$ con i = 2, 3, 4, 5, 6

$$E(X)=1*\frac{1}{6}+2*\frac{1}{6}+3*\frac{1}{6}+4*\frac{1}{6}+5*\frac{1}{6}+6*\frac{1}{6}=\frac{1}{6}*[1+2+3+4+5+6]=3,5$$

-A: evento

$$I_a(u)=\begin{cases}1 & se\, u\in A\\ 0 & altrimenti\end{cases}$$

$$E(I_A)=1*P(I_A=1)+0*P(I_A=0)=1*P(A)+0*P(\overline{A})=P(A)$$

- Quattro autobus che portono 148 studenti allo stadio. Gli autobus portano rispettivamente 40, 33, 25 e 50 studenti.

- Si seglie a caso uno studente e indichiamo con x = numero degli studenti che hanno viaggiato sull'autobus dello studente scelto a caso.
- Scelto a caso un conducente, Y= numero degli studenti che hanno viaggiato sull'autobus del conducente scelto a caso.

E(X) = ? E(Y) = ?

S_x= {40, 33, 25, 50 }

$$P_x(40)=\frac{40}{148} \qquad P_x(33)=\frac{33}{148} \qquad P_x(25)=\frac{25}{148} \qquad P_x(50)=\frac{50}{148}$$

$$E(X)=40*\frac{40}{148}+33*\frac{33}{148}+25*\frac{25}{148}+50*\frac{50}{148}\simeq 39,28$$

S_y = { 40, 33, 25, 50 }

$$P_x(40)=\frac{1}{4} \qquad P_x(33)=\frac{1}{4} \qquad P_x(25)=\frac{1}{4} \qquad P_x(50)=\frac{1}{4}$$

$$E(Y)=(40+33+25+50)*\frac{1}{4}=37$$

- C'è un libro di scommesse che suggerice il seguente modo di vincere:

- Si scommette 1 € sul rosso.
- Se esce il colore rosso (P=18/38) il gioco prende la sua vincita di 1 € e lascia il gioco.
- Se perde la prima giocata (1-p) fa un'altra giocata di 1 € per i successivi due giri e poi lascia il gioco.

X= guadagno del giocatore

1. La probabilità che sia > 0 : P(X > 0)
2. E(X)

1 ° partita	**2 ° partita**	**2 ° partita**	**X**	**P**
Vince	/	/	+ 1	p
Perde	Vince	Vince	+ 1	$p^2(1-p)$
Perde	Perde	Perde	- 1	$p(1-p)^2$
Perde	Perde	Vince	- 1	$p(1-p)^2$
Perde	Perde	Perde	- 3	$p(1-p)^3$

$P(X>0)=P(X=+1)=p+p^2(1-p)=0{,}5918$
$E(X)=(-3)*(1-p)^3+(-1)*[2p*(1-p)^2]+1*[p+p^2(1-p)]=-0{,}108\Rightarrow$
$\Rightarrow la\ strategia\ non\ è\ vincente$

- X discreta $S_x = \{-1, 0, 1\}$
$P(x=-1)= 0{,}2$ $P(x=0)= 0{,}5$ $P(x=1)= 0{,}3$
$E(x^2)=?$ $y=x^2$
$S_y = \{0, 1\}$

$$S(y=0)=P(x=0)=0{,}5$$

$$S(y=1)=\left\{\begin{matrix} P(x=-1) & = & 0{,}3 \\ & + & \\ P(x=1) & = & 0{,}3 \end{matrix}\right\}=0{,}5$$

$$E(Y)=0*0{,}5+1*0{,}5=0{,}5$$

$$y=g(X)$$

Teorema
Se X è una variabile aleatoria discreta e assume valore $\{x_1, x_2, ...\}$ con probabilità $\{P_x(x_1), P_x(x_2), \ldots\}$ e se *g* è una funzione a valori reali allora: $E(g(X))=\sum_i g(X)*P_x(X)$

$$E(X^2)=(-1)^2*0{,}2+(0)^2*0{,}5+(1)^2*0{,}3=0{,}5$$

PROPRIETÀ DI LINEARITÀ
X discreta, e sia $a,b \in R$.
$E(aX+b) = a* E(X)+b$.
Dim:
$g(x) = ax+b$ per il teorema $E(aX+b)=\sum(ax+b)*P_x(x_i)=a_i\sum_i x_i*P_x(x_i)+b*\sum_i P_x(x_i)=$
$a*E(x)+b*1$

Momento di ordine n di X $(x\geq 1): E(X^n)=\sum_{x:P_x(x)>0} x^n*P_x(x)$, in particolare per n = 1 si ritrova E(X).

Esercizi:
- X discreta.
$P_x(x) = c(1 +|x|)$ con $x = -1, 0, 1, 2$
1. Determina c.
2. Ricavare $F_x(X)$.
3. Valore atteso e momento del 2° ordine.

Soluzioni:

1. $P_x(-1)=c(1+1)=2c$ $P_x(0)=c(10)=c$
$P_x(1)=c(1+1)=2c$ $P_x(2)=c(1+2)=3c$

$$\left\{\begin{matrix} 0<2c<1 \\ 0<3c<1 \\ 0<c<1 \end{matrix}\right\} e\sum P_x(x)=1=2c+3c+2c+c=1\Rightarrow 8c=1 drarrow c=\frac{1}{8}$$

$$P_x(x)=\frac{1}{8}(1+|x|)$$

2. $P_x(-1)=\frac{2}{8}$ $P_x(0)=c(10)=\frac{1}{8}$
$P_x(1)=c(1+1)=\frac{2}{8}$ $P_x(2)=c(1+2)=\frac{3}{8}$

$$F_x(x)=\begin{cases} 0 & x<(-1) \\ \frac{2}{8} & -1<x<0 \\ \frac{3}{8} & 0<x<1 \\ \frac{5}{8} & 1<x<2 \\ 1 & x\geq 2 \end{cases}$$

3. $E(X)=(-1)*\frac{2}{8}+0*\frac{1}{8}+1*\frac{2}{8}+2*\frac{3}{8}=\frac{3}{4}$

$$E(X^2)=(-1)^2*\frac{2}{8}+(0)^2*\frac{1}{8}+(1)^2*\frac{2}{8}+(2)^2*\frac{3}{8}=2$$

VARIANZA

Ci sono 2 utenti.

Il primo riceve ogni giorno 48 o 52 email con uguale probabilità. Il secondo riceve ogni giorno 100 o 0 email con uguale probabilità.

X_1 = numero mail ricevute giornalmente dal 1° utente

X_2 = numero mail ricevute giornalmente dal 2° utente

$$E(x_1)=48*\frac{1}{2}+52*\frac{1}{2}=50 \qquad \textit{bassa variabilità}$$

$$E(x_2)=100*\frac{1}{2}+0*\frac{1}{2}=50 \qquad \textit{altà variabilità}$$

$$Var(x)=E[(X-E(x))^2]$$

Esercizi:

- X = esito del lancio del dado

S_x = {1, 2, 3, 4, 5, 6} $\quad P(i)=\frac{1}{6} \qquad i\in\{1,2,3,4,5,6\}$

$$E(X)=\frac{7}{2}$$

$$Var(X)=E\{(x-\frac{7}{2})^2\}=(1-\frac{7}{2})^2+\frac{1}{6}+(2-\frac{7}{2})^2*\frac{1}{6}+(3-\frac{7}{2})^2*\frac{1}{6}+(4-\frac{7}{2})^2*\frac{1}{6}+(5-\frac{7}{2})^2*\frac{1}{6}+(6-\frac{7}{2})^2*\frac{1}{6}\simeq$$

$$\simeq 2{,}91$$

FORMULA ALTERNATIVA

$$Var(X)=E(x^2)-\{E(X)\}^2$$

Dim:

$$E(x)=\mu$$

$$Var(X)=E[(x-\mu)^2]=E[x^2+\mu^2-2\mu x]=E(x^2)+E(\mu^2)-E(2\mu x)=E(x^2)+\mu^2-2\mu E(x)=$$

$$=E(x^2)+\mu^2-2\mu^2=E(x^2)-\mu^2=E(x^2)-(E(x))^2$$

Proposizione: Var(aX+b)=a^2 * var(x) non è lineare

Dim:

$$Var(aX+b)=E\{[(aX+b)-E(aX+b)]^2\}=E\{[aX+\cancel{b}-aE(x)-\cancel{b}]^2\}=E\{[a(X-E(x))]^2\}=$$

$$=E\{a^2*[X-E(x)]^2\}=a^2*E[(X-E(x))^2]=a^2*Var(X)$$

$\mathrm{Var}(X) \geq 0$

Deviazione standard: $\sqrt{\mathrm{Var}(X)}$

Esercizi:

- Un esperimento consiste nel lanciare 6 volte una moneta equa:

$$X = \begin{cases} 1 & \textit{se si realizza lo stesso numero di testa e croce} \\ 2 & \textit{se esce testa un numero di volte maggiore del numero di croce} \\ 0 & \textit{altrimenti} \end{cases}$$

Calcolare:

1. $P_x(X) = P(X=x)$ → densità distribuzione
2. $F_x(X)$ → funzione di distribuzione
3. $F(X)$, $\mathrm{var}(X)$ → evento atteso e varianza

Soluzioni:

$$1)\ P(X=1) = \frac{\binom{6}{3}}{2^6} = \frac{15}{16} \qquad P(X=2) = \frac{\binom{6}{4}}{2^6} + \frac{\binom{6}{5}}{2^6} + \frac{\binom{6}{6}}{2^6} = \frac{11}{32} \qquad P(X=0) = P(X=2) = \frac{11}{32}$$

$$2)\ F(x) = \begin{cases} 0 & x<0 \\ \frac{11}{32} & 0 \leq x < 1 \\ \frac{21}{32} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

$$3)\ E(x) = 0 * \frac{11}{32} + 1 * \frac{5}{16} + 2 * \frac{11}{32} = 1$$

$$var(x) = E(x^2) - \{E(x)\}^2 = 0^2 * \frac{11}{32} + 1^2 * \frac{5}{16} + 2^2 * \frac{11}{32} - 1 = \frac{27}{16} - 1 = \frac{11}{16}$$

VARIABILE DEGENERE

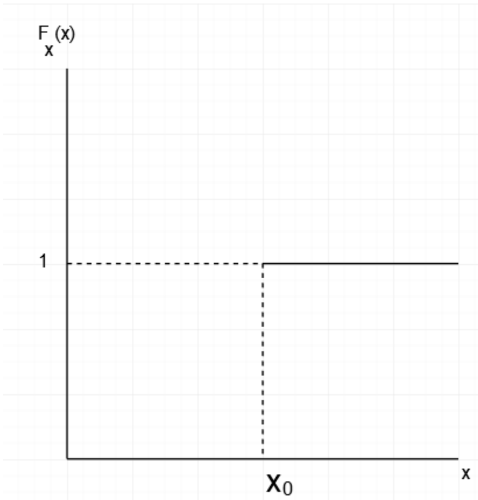
$P(X=x_0) = 1$ quindi se $S_{x=}\{x_0\}$

$$F_x(x) = \begin{cases} 0 & x < x_0 \\ 1 & x \geq x_0 \end{cases}$$

$$E(X) = x_0 * P(X=x_0) = x_0 * 1 = x_0$$

$$E(X^2) = x_0^2 * P(X=x_0) = x_0^2 * 1 = x_0^2$$

$$var(X) = E(x^2) - \{E(X)\}^2 = x_0^2 - x_0^2 = 0$$



VARIABILE DI BERNOULLI

Supponiamo di eseguire un'esperimento con 2 esiti possibili: successo e insuccesso. Indichiamo con *p* la probabilità di successo e con *1-p* di insuccesso, dove $0 \leq p \leq 1$.

Sia X=0 se l'esito dell'esperimento è un insuccesso ed X=1 se l'esito è successo.

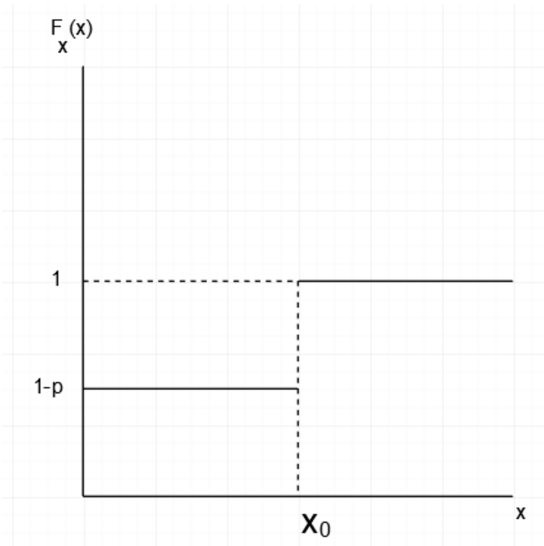
$P(X=0) = 1-p$ $\qquad$ $P(X=1) = p$

X è chiamata variabile aleatoria di bernoulli.

Si può esprimere anche come $P_x(X) = p^x(1-p)^{1-x}$ $\qquad$ con $x=0,1$

$$F_x(x)=\begin{cases}0 & 0<0\\ 1-p & 0\leq x<1\\ 1 & x\geq 1\end{cases}$$
$$E(X)=0*(1-p)+1*p=p$$
$$E(X^2)=0^2*(1-p)+1^2*p=p$$
$$var(X)=p-p^2=p(1-p)$$

Con p=1/2 la varianza assume valore massimo con p=1 o 1-p=1 diventa una varianza di Bernoulli



VARIABILE BINOMINIALE
Consideriamo n prove e supponiamo che siano indipendenti e che avvengano tutti nelle medesime condizioni. Supponiamo che ogni prova abbia 2 esiti possibili: insuccesso con probabilità 1-p, e successo con probabilità p, dove $0\leq p\leq 1$.
X = numero di successi nelle n prove

$$P(X=k)=\binom{n}{k}*p^k+(1-p)^{n-k} \qquad con\, k=0,1,\ldots,n$$

distribuzione binominale di parametri *n e p* $\qquad (\underbrace{S,S,\ldots,S}_{k},\underbrace{I,\ldots,I}_{n-k})=p^k*(1-p)^{n-k}$

$x\sim Binominale(n,p)$, con n= 1 ritroviamo la variabile di Bernoulli.

$$\sum_{k=0}^{n} P_x(k)=\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}=[p+(1-p)]^n=1 \qquad \textit{Questa formula è detto binomio di Newton}$$

Esercizi
- Una fabbrica produce viti che presentano un difetto, ognuno indipendentemente dalle altre con probabilità o,o1. Se le viti sono vendute in confezioni da 10, calcolare la probabilità che un pacchetto contenga almeno due viti difettose.
X= il numero di viti difettose nella confezione da 10

$$X\sim Binominale(n=10,\, p=0{,}01)$$

$$P(X\geq 2)=1-\{\,P(X=0)+P(X=1)\}=1-\{\binom{10}{0}(0{,}01)^0*(1-0{,}01)^{10-0}+\binom{10}{1}(0{,}01)^1(1-0{,}01)^{10-1}\}$$
$$\simeq 0{,}004$$

-In un gioco d'azzardo, un giocatore scommette su uno dei numeri compresi tra 1 e 6. Si lanciano 3 dadi. Se il numero su cui ha scommesso appare k volte(con k=1, 2, 3) il giocatore vince k euro. Se il numero non esce il giocatore perde 1 euro.
X= vincita del giocatore
S_x={-1, 1, 2, 3}

$$P(x=-1)=\binom{3}{0}\left(\frac{1}{6}\right)^0\left(1-\frac{1}{6}\right)^{3-0}=\frac{125}{216} \qquad P(x=1)=\binom{3}{1}\left(\frac{1}{6}\right)^1\left(1-\frac{1}{6}\right)^{3-1}=\frac{75}{216}$$
$$P(x=2)=\binom{3}{2}\left(\frac{1}{6}\right)^2\left(1-\frac{1}{6}\right)^{3-2}=\frac{15}{216} \qquad P(x=3)=\binom{3}{3}\left(\frac{1}{6}\right)^3\left(1-\frac{1}{6}\right)^{3-3}=\frac{1}{216}$$

$$E(X)=(-1)\frac{125}{216}+(1)\frac{75}{216}+(2)\frac{15}{216}+(3)\frac{1}{216}=\frac{-17}{216}$$ il gico non è equo

- Uno studente sta preparando un'esame orale. Se lo interroga l'assistente ha probabilità pari a 0, di rispondere ad ogni una delle domande. Se lo interroga il professore ha probabilità 0,4.
Le domande vengono porte in maniera indipendente da una dall'altra. Lo studente ritiene una probabilità doppia di fare l'esame con il professore piuttosto che con l'assistente, gli conviene che gli siano fatte 3 o 5 domande?

1° caso: 3 domande
X= numero risposte corrette nel caso in cui ho 3 domande

$$P(x\geq 2)=(P(x\geq 2|\text{lo studente è interrogato dal professore})P(\text{lo studente è interrogato dal professore}))$$
$$+(P(x\geq 2)|\text{lo studente è interrogato dall'assistente})P(lo\, studente\, è\, interrogato\, dall'\, assistente)=$$
$$=[\binom{3}{2}(0,4)^2(0,6)^{3-2}+\binom{3}{3}(0,4)^3(0,6)^0]*2*P(\frac{2}{3})+[\binom{3}{2}(0,8)^2(0,2)^{3-2}+\binom{3}{3}(0,8)^3(0,2)^0]*\frac{1}{3}=0,533$$

2° caso: 5 domande
Stesso procedimento del precedente. Risultato = 0,5256. Allo studente convengono 3 domande.

VALORE MEDIO E VARIANZA DELLA VARIABILE BINOMIALE
Teorema: $Se\ X\sim Binomiale(n,p)\, allora\, E(X)=n*p\, ,var(X)=n*p(1-p)$
Dim:

$$E(X)=\sum_{k=0}^{n}k*\binom{n}{k}p^k(1-p)^{n-k}=\sum_{k=1}^{n}k*\binom{n}{k}p^k(1-p)^{n-k}=\sum_{k=1}^{n}k*\frac{n!}{k!(n-k)}p^k(1-p)^{n-k}=$$
$$=\sum_{k=1}^{n}\cancel{k}*\frac{n(n-1)!}{\cancel{k}(k-1)!(n-k)!}p^k(1-p)^{n-k}=n\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}*p^k(1-p)^{n-k}=$$
$$=n\sum_{k=1}^{n}\binom{n-1}{k-1}*p^k(1-p)^{n-k}=\quad fissiamo\ j=k-1=$$
$$=n\sum_{j=0}^{n-1}\binom{n-1}{j}p^{j+1}(1-p)^{n-1-j}=n*p\sum_{j=0}^{n-1}\binom{n-1}{j}p^j(1-p)^{n-1-j}=$$
$$=\qquad abbiamo\, così\, riottenuto\, il\, binomio\, di\, newton=$$
$$=n*p[p+(1-p)]^{n-1}=n*p*1=n*p$$

Esercizi
- Un esperimento viene ripetuto n volte in condizioni di indipendenza. L'esperimento dà luogo a successo con probabilità p. Il numero medio di successi è 3, mentre la varianza è 2.
Calcolare quante sono le ripetizioni dell'esperimento che rendono la probabilità di avere successo maggiore di 1/27.

X=numero di successi nelle n prove

$$X\sim Binominale(n,p)\qquad n?\ \ p?$$
$$\left\{\begin{matrix}E(X)=n*p=3\\ Var(X)=n*p*(1-p)=2\end{matrix}\right\}\Rightarrow n=9\, e\, p=\frac{1}{3}$$
$$X\sim Binominale(n,\frac{1}{3})$$
$$P(X=n)=\frac{1}{3}^n\geq\frac{1}{27}=\frac{1}{3}^3\rightarrow n<3$$

- Un esperimento consiste nell'estrarre una biglia di ciascuna delle tre urne.Ogni urna contiene 3 biglie bianche ed 1 rossa. Calcolare la probabilità che:

1. Una delle biglie estratte sia bianca?
2. Almeno una delle biglie estratte siano bianche?

$X = numero\, di\, biglie\, estratte$

$X \sim Binominale(n=3, p=\frac{3}{4})$

1) $P(X=1)=\binom{3}{1}(\frac{3}{4})^{1}(1-\frac{3}{4})^{3-1}=3*\frac{3}{4}(\frac{1}{4})^{2}=\frac{9}{64}$

2) $P(X\geq 1)=1-p(X=0)=1-\binom{3}{0}(\frac{3}{4})^{0}(1-\frac{3}{4})^{3-0}=1-(\frac{1}{4})^{3}=\frac{63}{64}$

- Una moneta mostra testa con probabilità pari a p. La moneta viene lanciata 10 volte.
Se otteniamo 6 volte testa, calcolare la probabilità che i primi 3 esiti siano stati:

a) t c t
b) c t c

a) P((t c t) |nei dieci lanci abbiamo ottenuto 6 volte testa=

$$P\frac{((t\,c\,c)\cap(nei\,dieci\,lanci\,abbiamo\,ottenuto\,6\,volte\,T\,e\,4C))}{P(nei\,dieci\,lanci\,abbiamo\,ottenuto\,6\,volte\,testa)}=$$

$$=\frac{\binom{7}{5}p^{5}(1-p)^{2}*p(1-p)^{2}}{\binom{10}{6}p^{6}(1-p)^{10-6}}=\frac{\binom{7}{5}p^{6}(1-p)^{4}}{\binom{10}{6}p^{6}(1-p)^{4}}=\frac{\binom{7}{5}}{\binom{10}{6}}=\frac{1}{10}$$

b) è uguale ad *a*.

- Un esperimento consiste nell'effettuare lanci a caso indipendenti di una moneta truccata che ad ogni lancio mostra testa con probabilità 1-p.
X = numero totale di volte che esce croce in n lanci.

1. (X = k) con k= 1,…, n
2. Y= x+1, E(Y) e Var(Y)
3. E (1/ Y)

Soluzioni

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

2) $E(Y)=E(x+1)=E(x)+1=np+1$ dato da: $E(aX+b)=aE(X)+b$
$Var(Y)=Var(x+1)=Var(x)=np(1-p)$ dato da: $Var(aX+b)=a^{2}Var(X)$

3) $E(\frac{1}{Y})=E(\frac{1}{x+1})=\sum_{k=0}^{n}(\frac{1}{k+1}\binom{n}{k}p^{k}(1-p)^{n-k})$

- Sia X il numero di successi in *n* prove indipendenti quando la probabilità di successo è p, posto F_n= x/n. Determinare valore atteso e varianza di F_n. F_n= frequenza relativa.

$E(F_n)=E(\frac{X}{n})=\frac{1}{n}*E(X)=\frac{1}{n}*np=p$

$Var(F_n)=Var(\frac{X}{n})=\frac{1}{n^{2}}Var(X)=\frac{1}{n^{2}}np(1-p)=\frac{p(1-p)}{n}$ $\quad se\, n\quad tende\, a\quad infinito\, vale\, 0$

- Se $X \sim Binominale(n, p) e\, Y \sim Binominale(n, 1-p) \Rightarrow P(X=k)=P(Y=n-k) con\, k=0,\ldots,n$
Dim:

$$P(X=k)=\binom{n}{k} p^k (1-p)^{n-k}=\binom{n}{n-k}(1-p)^{n-k} p^{n-(n-k)}=P(Y=n-k)$$

DISTRIBUZIONE DI POISSON

$\binom{n}{k} p^k (1-p)^{n-k}$ per n $\rightarrow +\infty$, $p=0\, e\, np=\lambda$

$$\begin{aligned} P(Y=k) &= \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k} \\ &= \frac{(n)_k}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \cdot \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \cdot \frac{(1-\lambda/n)^n}{(1-\lambda/n)^k}. \end{aligned}$$

Per n grande risulta

$$\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \approx 1, \qquad \left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \qquad \left(1-\frac{\lambda}{n}\right)^k \approx 1.$$

Pertanto, quando n è grande e p è piccolo in modo che $n\,p=\lambda>0$ si ha

$$P(Y=k)=\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda} = \frac{(np)^k}{k!} e^{-np} \qquad (k=0,1,\ldots).$$

Esercizi:

- La probabilità di ricevere un full in una mano di poker è circa 0,0014. Calcolare la probabilità che in 1000 mani di poker ci siano stati almeno 2 full.
X = numero di full in n= 1000 .
con

$X \sim Binominale(n=1000, p=0{,}0014)$
$P(X\geq 2)=1-p(x<2)=1-[P(x=0)+P(x=1)]$
$1-[\binom{1000}{0}(0{,}0014)^0(1-0{,}0014)^{1000}+\binom{1000}{1}(0{,}0014)^1(1-0{,}0014)^{999}]\simeq 0{,}408265$

con Poisson

$np=1000*0{,}0014=1{,}4$
$P(x\geq 2)\simeq 1-[e^{-1,4}+(1{,}4)e^{-1,4}]=0{,}4082$

- Il numero di volte che una persona si prede l'influenza in un anno è una variabile aleatoria Y con distribuzione di Poisson di parametro λ=5. Viene commercializzata una nuova medicina che riduce il valore del parametro a λ=3 per il 75% della popolazione. Per l'altro 25% la medicina non è efficace.
Se un individuo prova la medicina e ha 2 influenze, quanto è probabile che la medicina sia stata efficace?

$$P(medicina\, è\, efficace \vee individuo\, ha\, 2\, influenze)=$$

$$\frac{P(individuo\, ha 2\, influenze \,|\, ma\, è\, efficace)*P(non\, è\, efficace)}{P(i.ha 2 inf.|ma\, è\, efficace)+P(nn\, è\, efficace)+P(i.ha 2 inf.|ma\, nn\, è\, efficace)*P(nn\, è\, efficace)}=$$

$$=\frac{P(X=2)*0{,}75}{P(X=2)0{,}75+P(Y=2)0{,}25}=\frac{[\frac{3^2}{2!}e^{-3}]0{,}75}{[\frac{3^2}{2!}e^{-3}]0{,}75+[\frac{5^2}{2!}e^{-5}0{,}25]}\simeq 0{,}886$$

Proprietà:
- E(X) = λ
- Var(X) = λ

Esercizi:
- Una copisteria assume 2 tipografi. Il numero medio di errori per articolo commessio dal primo tipografo è pari a 3, mentre è pari a 4,2 per il secondo. Se l'articolo può essere stato composto con uguale probabilità dai due tipografi, calcolare la probabilità che non ci siano errori.

X = numeri di errori del 1° tipografo $\sim Poisson(\lambda=3)$
Y= numeri di errori del 2° tipografo $\sim Poisson(\lambda=4{,}2)$

$$P(0\, errori)=P(X=0)*(l'articolo\, è\, stato\, fatto\, dal\, 1°\, tipografo)+$$
$$P(Y=0)*(l'articolo\, è\, stato\, fatto\, dal\, 2°\, tipografo)=$$
$$=\frac{(3)^0 e^{-3}}{0!}*\frac{1}{2}+\frac{(4{,}2)^0 e^{-4{,}2}}{0!}*\frac{1}{2}=\frac{1}{2}[e^{-3}+e^{-4{,}2}]\simeq 0{,}032$$

- Il numero di richieste di stampa che giunge ad una stampante aziendale è in media 3,2 al minuto. Approssimare la probabilità che non giungano più di 2 richieste.

X = numero di richieste che giungono alla stampante $\sim Poisson(\lambda=3{,}2)$

$$P(X\leq 2)=P(X=0)+P(X=1)+P(X=2)=$$
$$=[\frac{(3{,}2)^0}{0!}e^{-3{,}2}+\frac{(3{,}2)^1}{1!}e^{-3{,}2}+\frac{(3{,}2)^2}{2!}e^{-3{,}2}]\simeq 0{,}3799$$

VARIABILE GEOMETRICA

Supponiamo di ripetere (in condizioni di indipendenza) una prova, e la ripetiamo finché che non si verifica il primo successo. Supponiamo che in ogni prova il successo si realizzi con probabilità *p* e l'insuccesso con*1-p*.
X = numero di prove necessarie per ottenere il primo successo
$P(X = n) = (1 - p)^{n-1}p$

Ovvero, si ha X = n se le prime n−1 prove siano state un insuccesso e l'n-esima prova un successo. La formula si ricava anche per l'ipotesi di indipendenza tra le prove.

$$X\sim Geometrica(P)$$
$$\sum_{k=1}^{+\infty}(1-p)^{k-1}P\; si\, ritrova\, quindi\, la\, serie\, geometrica: \sum_{k=0}^{+\infty}c^k=\frac{1}{1-c}, con\ |c|<1$$

Esercizi:

- Un'urna contiene N biglie bianche e M biglie nere. Si estrae a caso una biglia alla volta, con reinserimento, fino a che non esce la prima biglia nera.
Calcolare la probabilità che si debbano estrarre:

a) esattamente n biglie
b) almeno k biglie

X = numero di biglie che si estraggono per ottenere la prima biglia nera.
Allora X ha distribuzione geometrica di parametro p = M/(M + N). Quindi:

$$\text{(a)} \qquad P(X=n) \;=\; (1-p)^{n-1}\,p;$$

$$\text{(b)} \qquad P(X>k) \;=\; p\sum_{n=k+1}^{\infty}(1-p)^{n-1} = p\,(1-p)^k\sum_{j=0}^{\infty}(1-p)^j$$

$$= p\,(1-p)^k\,\frac{1}{1-(1-p)} = (1-p)^k,$$

avendo posto $j=n-k-1$ e avendo ricordato che, per $-1<c<1$,

$$\sum_{j=0}^{\infty}c^j = \frac{1}{1-c}.$$

$P(X\geq k)=(1-p)^{k-1}$: questa probabilità è uguale ad avere insuccesso nelle prime k prove.
$P(X>k)=(1-p)^k$

Funzione di distribuzione

$$F_x(X)=P(X\leq x)=\begin{cases}0 & x<1\\ 1-(1-p)^k & k\leq x\leq k+1\end{cases}$$

$$p(x\leq k)=1-p-(x>k)=1-(1-p)^k$$

Proprietà di assenza di memoria

Se $X\sim Geometrica(P), \forall\, k, n\geq 0\; allora\; P(X>k+n|X>n)=p(X>k)$. Il fatto che abbiamo avuto n estrazioni non influenza la probabilità.

Dim:

La proprietà di assenza di memoria può esprimersi anche così: per ogni $k, n\geq 0$ risulta

$$P(X=n+k|X>k)=\frac{P(X=n+k)}{P(X>k)}=\frac{(1-p)^{n+k-1}p}{(1-p)^k}=(1-p)^{n-1}p=P(X=n).$$

Esercizi:

- Si lancia a caso una moneta. Calcolare la probabilit`a che esca testa per la prima volta dopo il 5° lancio sapendo che nei primi 3 lanci non esce testa.
X = numero di lanci necessari per ottenere testa per la prima volta

$$X\sim Geometrica\left(P=\frac{1}{2}\right)$$

$$P(X>5|X>3)=P(X>2)=\left(\frac{1}{2}\right)^2=\frac{1}{4}$$

Valore atteso e varianza

$$E(X)=\frac{1}{p} \qquad Var(X)=\frac{1-p}{p^2}$$

Esercizi:

- Un esperimento consiste nel lanciare una coppia di dadi non truccati.
X = numero di lanci necessari per cui si realizzi una coppia di numeri pari la prima volta.
Calcolare:

1. P(X = x) : densità discreta
2. la probabilità che siano necessari almeno 2 lanci per una coppia pari
3. il numero atteso di lanci per una coppia pari

$$1)\, X \sim Geometrica(P=\frac{1}{4}) \qquad P(X=x)=\frac{1}{4}(1-\frac{1}{4})^{x-1}$$

$$2)\, P(x\geq 2)=1-\frac{1}{4}=\frac{3}{4}$$

$$3)\, E(X)=4$$

VARIABILE ALEATORIA IPERGEOMETRICA

Nell’estrarre n biglie senza reinserimento da un’urna che contiene N biglie, di cui m sono bianche e N −m nere, sia X il numero di biglie bianche presenti tra le n estratte. Allora:

$$P(X=k)=\frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} \qquad k=0,1,\ldots,n. \qquad X\sim ipergeometrica(N,n,m)$$

$$=max(0,n+m-N)\leq k\leq min(n,m)$$

Confronto con la Binominale

$$\lim_{m\to+\infty} \inf_{n\to+\infty} \frac{\binom{m}{k}\binom{n-m}{n-k}}{\binom{N}{m}}=$$

=in questo caso è uguale ad una binominale se i valori di m e N sono molto grandi =

$$=\binom{n}{k}\left(\frac{m}{N}\right)^{k}\left(1-\frac{n}{N}\right)^{n-k}$$

valore atteso e varianza

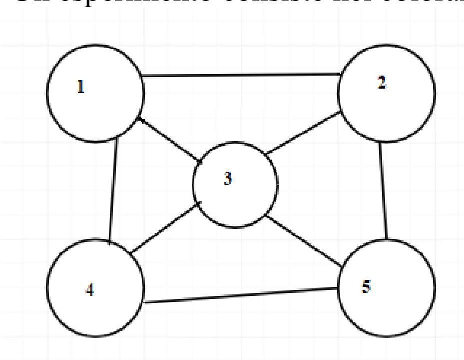
$$E(X)=n\,p \qquad \text{con } p=\frac{n}{N}$$

$$Var(X)=n\,p(1-p)(1-\frac{n-1}{N-1})$$

In caso che n = N la varianza vale 0 perchè assume come unico valore il valore medio.

Esercizi:

- Un esperimento consiste nel colorare di blu 3 vertici di questo grafo:



Successivamente si scelgono 2 vertici.
X = quanti tra i due nodi sono colorati di blu

1) $P(X=x)$
2) $F_x(X)$ $\qquad X \sim ipergeometrica(N=5, n=2, m=3)$
3) $E(X)$ e $Var(X)$

$$1)\, P(X=x) = \frac{\binom{3}{x}\binom{2}{2-x}}{\binom{5}{2}}$$

$$2)\, P_x(X) = \begin{cases} \frac{\binom{3}{0}\binom{2}{2}}{\binom{5}{2}} = \frac{1}{10} & x=0 \\ \frac{3\,2}{(10)} = \frac{6}{10} & x=1 \\ \frac{3\,1}{(10)} = \frac{3}{10} & x=2 \end{cases} \qquad F_x(X) = \begin{cases} 0 & x<0 \\ \frac{1}{10} & 0 \le x < 1 \\ \frac{7}{10} & 1 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

$$3)\, E(X) = n\,p = 2 * \frac{3}{5} = \frac{6}{5}$$

$$Var(X) = 2 * \frac{3}{5}\left(1 - \frac{1}{4}\right) = \frac{36}{100} = 0{,}36$$

- Una scatola contiene 10 sacchetti di cui 2 con un premio all'interno.
Alice seglie a caso 2 dei 10 sacchetti.
Successivamente Francesco sceglie a caso 1 degli 8 sacchetti rimanenti.
Calcolare la probabilità che:

1. Scelti dei sacchetti a caso da Alice ci siano k premi (k= 0, 1, 2)
2. Scelti dei sacchetti a cada tra i rimanenti 8 da Francesco ci sia un premio.

1) X = numeri di premi presenti nei 2 sacchetti scelti da Alice
A_k = {nei 2 sacchetti scelti da alice ci sono k premi} k = 0, 1, 2

$$P(A_k) = P(X=k) = \frac{\binom{2}{k}\binom{8}{2-k}}{\binom{10}{2}} = \begin{cases} \frac{28}{45} & k=0 \\ \frac{16}{45} & k=1 \\ \frac{1}{45} & k=2 \end{cases}$$

$X \sim Ipergeometrica(N=10, m=2, n=2)$

2) B = {nel sacchetto scelto da francesco c'è un premio}

$$P(B)=\sum_{k=0}^{2} P(B|A_k)P(A_k)=\sum_{k=0}^{2} \frac{2-k}{8} * \frac{\binom{2}{k}\binom{8}{2-k}}{\binom{10}{2}}=\frac{1}{5}=0{,}2$$

LA VARIABILE UNIFORME DISCRETA

S_x = {1, 2, 3, ..., N} $\quad P(X=x)=\frac{1}{N}, \forall x \in S_x$

un esempio è il lancio dei dadi.

$$E(X)=\frac{N+1}{2}$$

$$DIM: E(X)=\sum_{k=1}^{N} k*\frac{1}{N}=\frac{1}{N}\sum_{k=1}^{N} k=\frac{1}{N}*\frac{N(N+1)}{2}=\frac{N+1}{2}$$ somma dei primi N numeri naturali

$$Var(X)=E(x^2)-\{E(X)\}^2=\frac{(N+1)(2N+1)}{6}-\frac{N+1}{2}=\frac{n^2-1}{2}$$

$$DIM: E(X^2)=\sum_{k=1}^{N} k^2*\frac{1}{N}=\frac{1}{N}\sum_{k=1}^{N} k^2=\frac{1}{N}*\frac{(N+1)(2N+1)}{6}$$

somma dei quadrati dei primi N numeri naturali

VARIABILI ALEATORIE ASSOLUTAMENTE CONTINUE

X è assolutamente continua se esiste f:R ⟶ [0,+∞] t.c. $\forall B \subseteq R \quad$ $P(X \in B)=\int_b f_x(x)\,dx$

$f_x(x)$= densità di probabilità

Proprietà

- 1) $f(x) \geq 0 \quad \forall x \in R$
- 2) $\int_{-\infty}^{+\infty} f_x(x)dx=P(X \in (-\infty,+\infty))=1$ detta condizione di normalizzazione
- 3) P($a \leq X \leq b$)= $\int_a^b f_x(x)dx$ se b= a P(X=a)=0
- 4)P(X<x) =P($X \leq x$) = $F_x(x)$ = $\int_{-\infty}^{+\infty} f_x(t)\,dt$ t=densità di probabilità

  se X è assolutamente continua allora $F_x(x)$= $\int_{-\infty}^{x} f_x(t)\,dt$

  viceversa se x è un punto di continuità di $f_x(x)$⟶ $f_x(a)$= $\frac{d}{dx}*F_x(x)$

Esercizi:

-Fx(x)= $\begin{cases} c(4x-2x^2) & \text{a<x<2} \\ 0 & \text{altrimenti} \end{cases}$

1. determina c
2. calcolare P(X>1)

1) $\begin{cases} f_x(x) \geq 0 \quad \forall x \\ (\int_{-\infty}^{+\infty} f_x(x)dx) = 1 \end{cases} \rightarrow \quad c\int_0^2 [4x-2x^2]dx = c\int_0^2 4x\,dx - c\int_0^2 2x^2\,dx = 4c\left[\frac{x^2}{2}\right]_0^2 - 2c\left[\frac{x^3}{3}\right]_0^2 =$

=4c*2-2c* $\frac{8}{3} = 8c - \frac{16}{3}c = \frac{8}{3}c \rightarrow c = \frac{3}{8}$

2) $\int x^n\,dx = \frac{x^{(n+1)}}{n+1}$

P(X>1)=

$$\int_1^2 \frac{3}{8}(4x-2x^2)dx = \frac{3}{2}\int_1^2 x\,dx - \frac{3}{4}\int_1^2 x^2 dx = \frac{3}{2}\left[\frac{x^2}{2}\right]_1^2 - \frac{3}{4}\left[\frac{x^3}{3}\right]_1^2 = \frac{3}{2}\left\{2-\frac{1}{2}\right\} - \frac{3}{4}\left\{\frac{8}{3}-\frac{1}{3}\right\} = \frac{3}{2}*\frac{3}{2} - \frac{3}{8}*\frac{7}{3} = \frac{1}{2}$$

-X, $f_x(x)$, $F_x(x)$ trasformo la x in y $\rightarrow$ y= aX+b, a≠0

$F_y(y)$= P($Y \leq y$) = $P(aX+b \leq y) = P(aX \leq y-b) = \begin{cases} P(x) \leq \frac{y-b}{a} & \text{a>0} \\ P(x) \geq \frac{y-b}{a} & \text{a<0} \end{cases}$ = $F_x(X)$=P(X<=x)

= $\begin{cases} F_x(\frac{y-b}{a}) & \text{a>0} \\ 1-F_x(\frac{y-b}{a}) & \text{a<0} \end{cases}$ =$F_y(y)$= $\frac{d}{d_y} * F_y(y)$ = $\begin{cases} f_x(\frac{y-b}{a}) * \frac{1}{a} & \text{a>0} \\ -f_x(\frac{y-b}{a}) * \frac{1}{a} & \text{a<0} \end{cases}$ =

= $f_y(Y) = f_x(\frac{y-b}{a}) * \frac{1}{|a|}$ ← **FORMULA PER TROVARE LA DENSITÀ DI UNA Y APPLICATA AD UNA X**

-Trovare la densità di y=1-X

$f_x(x) = \begin{cases} 2x & 0 \leq X \leq 1 \\ 0 & \text{altrimenti} \end{cases}$

b a

y=1-1x

a=-1 b=1

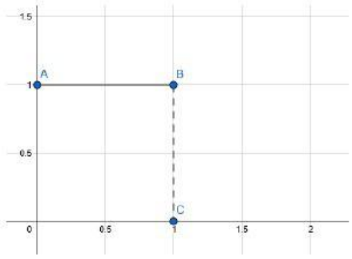
$F_y(Y) = f_x\left(\frac{y-1}{(-1)}\right) * \frac{1}{|-1|} = f_x(1-y) \quad \rightarrow \quad f_y(y) = \begin{cases} 2(1-y) & 0 \leq y \leq 1 \\ 0 & \text{altrimenti} \end{cases}$

ALTRE FORMULE

- E(X)= $\int_{-\infty}^{+\infty} x f_x(x)dx$

- $E(X^n)= \int_{-\infty}^{+\infty} x^n f_x(x)\,dx$
- $var(X)= E[(X-\mu)^2]= \int_{-\infty}^{+\infty} (x-\mu)^2 f_x(X)\,dx = E(x^2)-\{E(X)\}^2$
  $\mu=E(x)$
- $E(aX+b)= aE(x)+b$
- $var(aX+b)=a^2 * var(x)$
-

VARIABILI ALEATORIA UNIFORME
Assegno uguale probabilità ad intervalli che hanno la stessa lunghezza



$X \sim U(0,1)$ se ha $f_x(x)= \begin{Bmatrix} 1 & 0<x<1 \\ 0 & \text{altrimenti} \end{Bmatrix}$

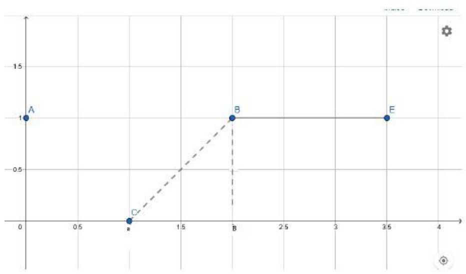
$P(\ a \le C \le b\ )= \left(\int_a^b f_x(X)\,dx - \int_a^b dx\right) = b-a$
$0<=a<b<=1$

Esempio
$P(\ 0 \le x \le \frac{1}{4}\ ) = P(\ \frac{1}{4} \le x \le \frac{1}{2}\ ) = \frac{1}{4}$

CASO GENERALE
$X \sim U(\alpha,\beta)$ se ha densità $f_x(x) = \begin{Bmatrix} \frac{1}{\beta-\alpha} & \alpha<x<\beta \\ 0 & \text{altrimenti} \end{Bmatrix}$

Esercizi:
$-\alpha<c<d<\beta$
$P(c<X<d)= \int_c^d \frac{1}{\beta-\alpha}\,dx = \frac{d-c}{\beta-\alpha}$

-funzione distribuzione → $F_x(x)=P(X<=x)= \int_{-\infty}^{+\infty} f_x(t)\,dt = \begin{Bmatrix} 0 & x<\alpha \\ \int_a^x \frac{1}{\beta-\alpha} & \alpha<x<\beta \\ 1 & x \le \beta \end{Bmatrix}$



-Momento → $E(x^n)= \int_\alpha^\beta x^n * \frac{1}{\beta-\alpha}\,dx = \frac{1}{\beta-\alpha}\left[\frac{x}{n+1}\right]_\alpha^\beta = \frac{\beta^{(n+1)}-\alpha^{(n+1)}}{(n+1)(\beta-\alpha)}$

-Valore atteso → $E(x)= \int_\alpha^\beta \frac{\beta^2-\alpha^2}{2(\beta-\alpha)} = \frac{\cancel{(\beta-\alpha)}(\beta+\alpha)}{2\cancel{(\beta-\alpha)}} = \frac{\beta-\alpha}{2}$

-Varianza → $var(x) = E(x^2) - \{E(x)\}^2 = \dfrac{\beta^3-\alpha^3}{3(\beta-\alpha)} - \left\{\left(\dfrac{\beta-\alpha}{2}\right)\right\}^2 =$

$\dfrac{\beta^2+\alpha\beta+\alpha^2}{3} - \dfrac{\alpha^2+2\alpha\beta+\alpha^2}{4} = \dfrac{(\beta-\alpha)^2}{12}$ la varianza e l'ampiezza dell'intervallo al quadrato fratto 12

Esercizi:

- I treni per una destinazione A passano alla stazione ogni 15 minuti a partire dalle 7. Quelli per una destinazione B passano ogni 15 minuti a partire dalle 7:05.

a) Un passeggero arriva alla stazione in un istante che è uniformemente distribuito dalle 7:00 alle 8:00, e sale sul primo treno che arriva. Calcolare la probabilità che salga sul treno diretto ad A.

b) Un passeggero arriva alla stazione in un istante che è uniformemente distribuito dalle 7 :10 alle 8:10, e sale sul primo treno che arriva. Calcolare la probabilità che salga sul treno diretto ad A.

Soluzioni:

a)

Tempo per A:	7:00	7:15	7:30	7:45	8:00
Tempo per B:	7:05	7:20	7:35	7:50	

$X =$ instante di arrivo del passeggero $\sim U(\underbrace{0}_{\alpha}, \underbrace{60}_{\beta})$

A = {il passeggero prende il treno per A}

$P(A) = P(5 < X < 15, 29 < X < 30, 35 < X < 45, 50 < X < 60) = \dfrac{10}{60} + \dfrac{10}{60} + \dfrac{10}{60} + \dfrac{10}{60} = \dfrac{2}{3}$

b)

Tempo per A:	7:00	7:15	7:30	7:45	8:00	8:15
Tempo per B:	7:05	7:20	7:35	7:50	8:05	

$X =$ instante di arrivo del passeggero $\sim U(\underbrace{0}_{\alpha}, \underbrace{60}_{\beta})$

A = {il passeggero prende il treno per A}

$P(A) = P(5 < X < 15, 29 < X < 30, 35 < X < 45, 50 < X < 60) = \dfrac{10}{60} + \dfrac{10}{60} + \dfrac{10}{60} + \dfrac{10}{60} = \dfrac{2}{3}$

- Un passeggero arriva alla fermata dell'autobus alle 10. Sapendo che il tempo di arrivo del bus è uniformemente distiribuito tra le 10:00 e le 10:30.

1. Calcolare la probabilità che deve aspetare più 10 minuti.
2. Se l'autobus non è ancora passato alle 10:15, calcolare la probabilità di aspettare 10 minuti.

Soluzioni:

1) $X =$ instante di arrivo dell'autobus $\sim U(\underbrace{0}_{\alpha}, \underbrace{20}_{\beta})$

$P(X > 10) = 1 - F_x(10) = 1 - \dfrac{10-0}{30-0} = 1 - \dfrac{1}{3} = \dfrac{2}{3}$

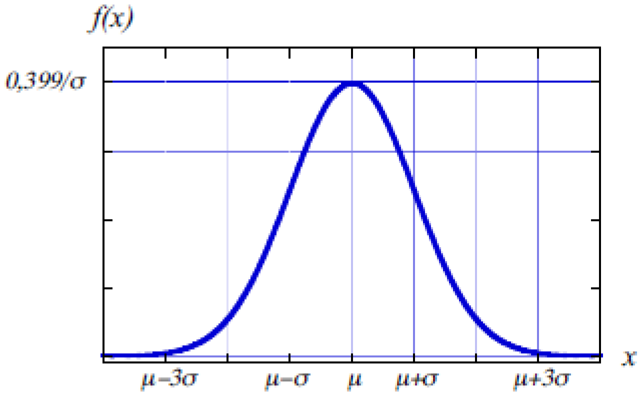
2) $P(X > 25 | X > 15) = \dfrac{P((X > 25) \cap (X > 15))}{P(X > 15)} = \dfrac{P(X > 25)}{P(X > 15)} = \dfrac{1 - F_x(25)}{1 - F_x(15)} = \dfrac{1 - \dfrac{25}{30}}{1 - \dfrac{15}{30}} = \dfrac{1}{3}$

VARIABILE ALEATORIE NORMALI

Una variabile aleatoria continua X è detta normale (o gaussiana) di parametri μ ∈ R e $\sigma^2 > 0$ se la densità di X è data da:

$$x \sim N(\mu,\sigma^2), \mu \in R, \sigma^2 > 0$$

$$f_x(X) = \frac{1}{\sqrt{2\Pi\sigma^2}} * \exp\frac{-(x-\mu)^2}{2\sigma^2}, X \in R$$



μ = valore atteso (E(X)) -> può essere negativo
σ^2 = varianza (Var(X)) -> deve essere sempre positivo

Variabile normale standard

$$Z \sim N(0,1) \qquad E(Z)=0 \quad \text{e} \quad Var(Z)=1$$

$$f(X) = \frac{1}{\sqrt{2\Pi}} * e^{\frac{-x^2}{2}}, x \in R$$

$$\Phi(X) = \frac{1}{\sqrt{2\Pi}} * \int_{-\infty}^{x} e^{\frac{-x^2}{2}} dt \leftarrow \textit{funzione di distribuzione}$$

Proprietà:

$$\Phi(-X) = 1 - \Phi(X)$$

dim:

$$\Phi(-X) = \frac{1}{\sqrt{2\Pi}} * \int_{-\infty}^{-x} e^{\frac{-t^2}{2}} dt \rightarrow \text{poniamo } y=-t \rightarrow \frac{1}{\sqrt{2\Pi}} * \int_{+\infty}^{x} e^{\frac{-y^2}{2}} (-dy) = \frac{1}{\sqrt{2\Pi}} * \int_{x}^{+\infty} e^{\frac{-y^2}{2}} =$$
$$= 1 - \Phi(X)$$

Funzione distribuita di una generica variabile normale

$$X \sim N(\mu,\sigma^2) \qquad \sigma = \sqrt{\sigma^2} = \textit{deviazione standard}$$

$$F_x(X) = P(X) \le X = \text{applico una trasformazione di variabile} = P\left(\frac{x-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P\left(Z \le \frac{x-\mu}{\sigma}\right) =$$
$$\Phi\left(\frac{x-\mu}{\sigma}\right)$$

proposizione:

se $X \sim N(\mu,\sigma^2) \Rightarrow \frac{x-\mu}{\sigma} = Z \sim N(0,1)$

dim che $\boldsymbol{E\left(\frac{x-\mu}{\sigma}\right)}$ **ha** $\boldsymbol{E(X)=0}$ **e** $\boldsymbol{Var(X)=1}$ **:**

$$E\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$$ per la regola: $E(aX+b) = aE(X)+b$

$$Var\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma^2}Var(X) = \frac{\sigma^2}{\sigma^2} = 1$$ per la regola: $= Var(aX+b) = a^2 Var(X)$

Esercizi:

- Sia X una variabile aleatoria normale di parametri $\mu = 3$ e $\sigma^2 = 9$

$X \sim N(\mu=3, \sigma^2=9)$

Determinare:

1. $P(2 < X < 5)$
2. $P(X > 0)$
3. $P(|X - 3| > 6)$

Soluzioni:

1.
$$\begin{aligned} P(2<X<5) &= P\left(\frac{2-3}{3} < \frac{X-3}{3} < \frac{5-3}{3}\right) = P\left(-\frac{1}{3} < Z < \frac{2}{3}\right) \\ &= \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{1}{3}\right) \approx \Phi(0{,}67) - [1 - \Phi(0{,}33)] \\ &= 0{,}7486 - (1 - 0{,}6293) = 0{,}3779 \end{aligned}$$

2.
$$P(X>0) = P\left(\frac{X-3}{3} > \frac{0-3}{3}\right) = P(Z > -1) = 1 - \Phi(-1) = \Phi(1) = 0{,}8413$$

3.
$$\begin{aligned} P(|X-3|>6) &= P(X>9) + P(X<-3) \\ &= P\left(\frac{X-3}{3} > \frac{9-3}{3}\right) + P\left(\frac{X-3}{3} < \frac{-3-3}{3}\right) \\ &= P(Z>2) + P(Z<-2) = 1 - \Phi(2) + \Phi(-2) \\ &= 2[1-\Phi(2)] = 2(1-0{,}9772) = 0{,}0456 \end{aligned}$$

- $X \sim N(\mu, \sigma^2=9)$ Determinare E(X) = μ se:
  $1)\, P(X \leq 2) = 0{,}791$
  $2)\, P(x > 0) = 0{,}281$

1. $0{,}791 = P(\frac{x-\mu}{3} \leq \frac{2-\mu}{3}) = P(Z \leq \frac{2-\mu}{3}) = \Phi(\frac{2-\mu}{3}) = \frac{2-\mu}{3} = 0{,}81 \rightarrow$
   $\rightarrow \mu = 2 - 0{,}81 * 3 = -0{,}43$

2. $0{,}281 = P(X>0) = P(Z > \frac{-\mu}{3}) = 1 - \Phi(\frac{-\mu}{3}) = 1 - [1 - (\frac{\mu}{3})] =$
   $\Phi(\frac{\mu}{3})$ non esiste il valore , possiamo ragionare su $= \Phi(\frac{-\mu}{3}) = 1 - 0{,}281 = 0{,}719 =$
   $= -\mu 3 = 0{,}58 \rightarrow \mu = -3 * 0{,}58 = -1{,}74$

- $X \sim N(\mu=5, \sigma^2)$ Se P(X > 9)= 0,2 determinare $\sigma^2 (Varianza) \rightarrow$
  $\rightarrow 0{,}2 = P(Z > \frac{9-5}{\sigma}) = 1 - [\Phi(\frac{4}{\sigma})] = \Phi(\frac{4}{\sigma}) = 0{,}8 = \frac{4}{\sigma} = 0{,}84 \rightarrow \sigma = \frac{4}{0{,}84} = 4{,}76 \rightarrow \sigma^2 = (4{,}76)^2$

- $X\, normale: E(\frac{X-3}{2}) = -2$ $\qquad Var(\frac{X-3}{2}) = 1$
  Determinare:
  $1)\, P(X \leq 0)$
  $2)\, P(X \leq 0 | -1 < x \leq 1)$

Ricaviamo prima μ e σ^2:

$$\mu=-2=E\left(\frac{x-3}{2}\right)=\frac{1}{2}E(X)-\frac{3}{2}=\frac{E(X)}{2}=-2+\frac{3}{2}=-\frac{1}{2}=E(X)=-1$$

$$1=Var\left(\frac{x-3}{2}\right)=\frac{1}{4}*Var(X)=4$$

μ = -1 e σ^2 = 4

1) $$P(x\leq 0)=P\left(\frac{x-(-1)}{2}\leq\frac{0-(-1)}{2}\right)=P\left(Z\leq\frac{1}{2}\right)=\Phi\left(\frac{1}{2}\right)=0{,}6915$$

2)P(X≤0 | -1≤X≤1)= $$\frac{P((X\leq 0)\cap(-1\leq X\leq 1))}{P(-1\leq X\leq 1)}=\frac{P(-1\leq X\leq 0)}{P(-1\leq X\leq 1)}=\frac{P\left(\frac{-1-(-1)}{2}\leq Z\leq\frac{0-(-1)}{2}\right)}{P\left(\frac{-1-(-1)}{2}\leq Z\leq\frac{1-(-1)}{2}\right)}=$$

$$\frac{\Phi\left(\frac{1}{2}\right)-\Phi(0)}{\Phi(1)-\Phi(0)}=\frac{0{,}6915-0{,}5}{0{,}8413-0{,}5}=0{,}5611$$

Distribuzione normale standard

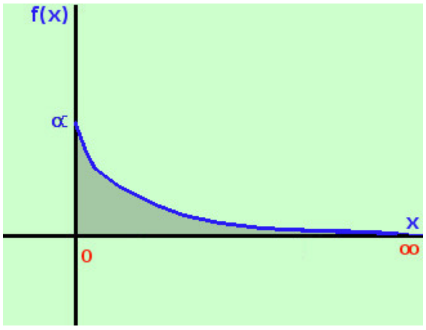
I valori di $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2}\, dz$ sono riportati per alcune scelte di x.

Ad esempio a $x = 2{,}54$ (ottenuto come $2{,}5 + 0{,}04$) corrisponde $\Phi(x) = 0{,}9945$.

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5474	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

VARIABILE ALEATORIA ESPONENZIALE

X~Exp(λ) , λ>0 se ha densità $f_x(x)= \lambda e^{-\lambda x}$, x>=0



$$F_x(x)=\int_{-\infty}^{x} f_x(t)dt=\begin{cases} 0 & x<0 \\ \int_0^x \lambda e^{-\lambda t}=[-e^{-\lambda t}]_0^x=1-e^{-\lambda x} & \text{se } x\geq 0 \end{cases}$$

E(x)= $\frac{1}{\lambda}$;   var(x)= $\frac{1}{\lambda^2}$;

PROPRIETÀ

- Assenza di memoria
  - X ~ Exp(λ) , ∀ s,t ≥ 0
    P(X>s+t | X>t) = P(X>S)
    dim.:
    P(X>s+t | X>t)= $\frac{P(X>s+t)\cap(X>t)}{P(X>t)}=\frac{P(X>s+t)}{P(X>t)}=\frac{1-F_x(t+s)}{1-F_x(t)}=1-\frac{[1-e^{-\lambda(t+s)}]}{1-[1-e^{-\lambda t}]}$ =
    = $\frac{e^{-\lambda(t+s)}}{e^{-\lambda t}}=-e^{-\lambda s}=1-[F_x(s)]=P(X>s)$

ESERCIZIO

Il tempo di vita di una macchina è una variabile aleatoria X esponenziale di valore atteso E(x)=6

1. Determinare $F_x(x)$;
2. Calcolare la probabilità che la macchina funzioni per più di 12 anni dall'acquisto
3. Calcolare la probabilità che la macchina funzioni per più di 12 anni dall'acquisto sapendo che è già vecchia di un anno

1. $\lambda=\frac{1}{6}\rightarrow F_x(x)=\begin{cases} 0 & \text{se } x<0 \\ 1-e^{-\frac{x}{6}} & \text{se } x\geq 0 \end{cases}$

2. P(X>12)= 1-F_x(12)= 1-1-e^{-2} = -e^{-2} ≃-0,1353

3.
s+t   t
P(X>13 | X>1) → per l'assenza di memoria = P(X>12)≃-0,1353

Esercizi:

- X~Exp(λ)

1. Determina $F_x(x)$ sapendo che var(x)=4
2. P(x ≤ 4 | X>2)

var(x)= $\frac{1}{\lambda^2}$ =4 $\rightarrow$ $\lambda^2=\frac{1}{2}$

1. $F_x(x)= \begin{cases} 0 & \text{se} < 0 \\ 1-e^{-\frac{x}{2}} & \text{se } x \geq 0 \end{cases}$

2. P(x ≤ 4 | X>2)= $\frac{P(2<X\leq 4)}{P(X>2)}=\frac{F_x(4)-F_x(2)}{1-F_x(2)} = \frac{[1-e^{-2}]-[1-e^{-1}]}{1-[1-e^{-1}]}=\frac{e^{-1}-e^{-2}}{e^{-1}}=1-e^{-1}$

VETTORI ALEATORI

Prendiamo due variabili (X, Y) aleatorie e discontinue, la loro densità discreta congiunta è :

$$p(x,y) = P(X=x, Y=y).$$

Notiamo che:

$$p(x,y) \geq 0 \quad \forall x, y, \qquad \sum_x \sum_y p(x,y) = 1.$$

Le densità discrete di X e di Y si ottengono da p(x, y) al seguente modo:

$$p_X(x) = P(X=x) = \sum_{y:\,p(x,y)>0} p(x,y);$$

$$p_Y(y) = P(Y=y) = \sum_{x:\,p(x,y)>0} p(x,y).$$

Esercizi:

- Vengono scelte a caso 3 biglie da un'urna contenente 3 biglie rosse, 4 bianche e 5 blu. Se denotiamo con X e Y , rispettivamente, il numero di biglie rosse e bianche scelte, la densità congiunta di X e Y , p(x, y) = P(X = x, Y = y), è data da:

$$p(0,0) = \binom{5}{3}\Big/\binom{12}{3} = \frac{10}{220}, \qquad p(0,1) = \binom{4}{1}\binom{5}{2}\Big/\binom{12}{3} = \frac{40}{220},$$

$$p(0,2) = \binom{4}{2}\binom{5}{1}\Big/\binom{12}{3} = \frac{30}{220}, \qquad p(0,3) = \binom{4}{3}\Big/\binom{12}{3} = \frac{4}{220},$$

$$p(1,0) = \binom{3}{1}\binom{5}{2}\Big/\binom{12}{3} = \frac{30}{220}, \qquad p(1,1) = \binom{3}{1}\binom{4}{1}\binom{5}{1}\Big/\binom{12}{3} = \frac{60}{220},$$

$$p(1,2) = \binom{3}{1}\binom{4}{2}\Big/\binom{12}{3} = \frac{18}{220}, \qquad p(2,0) = \binom{3}{2}\binom{5}{1}\Big/\binom{12}{3} = \frac{15}{220},$$

$$p(2,1) = \binom{3}{2}\binom{4}{1}\Big/\binom{12}{3} = \frac{12}{220}, \qquad p(3,0) = \binom{3}{3}\Big/\binom{12}{3} = \frac{1}{220},$$

Le probabilità $p(x, y)$ possono essere facilmente tabulate, come di seguito mostrato.

$$p(x,y) = P(X = x, Y = y) = \frac{\binom{3}{x}\binom{4}{y}\binom{5}{3-x-y}}{\binom{12}{3}} \qquad (x, y = 0, 1, 2, 3;\ x + y \leq 3)$$

$x \setminus y$	0	1	2	3	$P(X = x)$
0	10/220	40/220	30/220	4/220	84/220
1	30/220	60/220	18/220	0	108/220
2	15/220	12/220	0	0	27/220
3	1/220	0	0	0	1/220
$P(Y = y)$	56/220	112/220	48/220	4/220	1

Si ha:

$$P(X \geq 2) = \frac{28}{220} \approx 0{,}1273 \qquad P(Y \leq 1) = \frac{168}{220} \approx 0{,}7636$$

P(X = Y) = prendo la diagonale = 70 / 220

- Un esperimento consiste nello scegliere un nodo. X, Y sono il minimo e massimo tra il numero dei nodi che sono cammini da un singolo arco al nodo scelto.

Determinare $P_{xy}(X, Y)$

	X	Y
1	2	4
2	1	5
3	1	5
4	1	3
5	2	3

X \ Y	**3**	**4**	**5**	**P(X = x)**
1	1/5	0	2/5	3/5
2	1/5	1/5	0	2/5
P(Y = y)	2/5	1/5	2/5	1

Funzione di distribuzione congiunta

$P_{x\,y}(X, Y) = P(X = x, Y = y)$

$$F(x,y) = P(X \leq x, Y \leq y) = \sum_{u\,:\,u \leq x} \sum_{v\,:\,v \leq x} p(u,v).$$

La funzione di distribuzione di X può essere ottenuta da quella congiunta di (X, Y) come segue:

$$\lim_{y \to +\infty} F(x,y) = \lim_{y \to +\infty} P(X \leq x, Y \leq y) = P\left(\lim_{y \to +\infty} \{X \leq x, Y \leq y\}\right)$$
$$= P(X \leq x, Y < +\infty) = P(X \leq x) = F_X(x).$$

In maniera analoga la funzione di distribuzione di Y è data da:

$$\lim_{x \to +\infty} F(x,y) = F_Y(y).$$

Notaimo ache che: $\lim_{a \to +\infty} \sup_{b \to +\infty} F_{xy}(a,b) = 1$

Esercizi:

Si lanci 2 volte una moneta. X_1 = per n-volte esce testa X_2 = numero di variazioni del risultato

-
	X_1	X_2
t t	2	0
t c	1	1
c t	1	1
c c	0	0

$$P_{xy}(X,Y)=\begin{cases} \frac{1}{4} & x=0, y=0 \\ \frac{1}{4} & x=2, y=0 \\ \frac{2}{4} & x=1, y=1 \end{cases}$$

$X_1 \backslash X_2$	0	1	P(X = x)
0	1/4	0	1/4
1	0	2/4	2/4
2	1/4	0	1/4
P(Y = y)	1/2	1/2	1

$$F_{xy}(X,Y)=\begin{cases} 0 & x\in R, y<0 \\ 0 & x<0, y\in R \\ \frac{1}{4} & 0\le x<1, y\ge 0 \\ \frac{1}{4} & 1\le x<2, 0\le y<1 \\ \frac{3}{4} & 1\le x<2, y\ge 1 \\ \frac{1}{2} & x\ge 2, 0\le y<1 \\ 1 & x\ge 2, y\ge 1 \end{cases}$$

$$F_x(a)=\begin{cases} 0 & a<0 \\ \frac{1}{4} & 0\le a<1 \\ \frac{3}{4} & 1\le a<2 \\ 1 & a\ge 2 \end{cases} \qquad F_y(b)=\begin{cases} 0 & b<0 \\ \frac{1}{2} & a\le b<1 \\ 1 & b\ge 1 \end{cases}$$

Distribuzione multimodale

La distribuzione multinomiale si ottiene quando si ripete n volte un esperimento in condizioni di indipendenza, quando ogni esperimento può avere come risultato uno qualsiasi tra r possibili esiti, con probabilità $p_1, \ldots, p_r$, rispettivamente, tali che $p_1 + \ldots + p_r = 1$. Se denotiamo con X_i il numero degli n esperimenti che hanno dato come risultato l'esito i, allora si ha

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_r = n_r) = \frac{n!}{n_1!n_2!\cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}, \quad \text{con} \quad \sum_{i=1}^{r} n_i = n.$$

Tale formula è verificata notando che ogni successione degli esiti degli n esperimenti che portano al fatto che l'esito i si verifichi esattamente n_i volte, per $i = 1, 2, \ldots, r$, avrà probabilità pari a $p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$, grazie all'ipotesi di indipendenza. La formula finale segue ricordando che $n!/(n_1!n_2! \ldots n_r!)$ è il numero di permutazioni di n oggetti di cui n_1 sono uguali tra loro, n_2 sono uguali tra loro, ... , n_r sono uguali tra loro.

Osserviamo che per $r = 2$ la distribuzione multinomiale si riduce a quella binomiale.

Esempio. Nel lanciare 9 volte un dado equilibrato, indicando con X_k, $1 \le k \le 6$, il numero di volte che il risultato è k, si ha

$$P(X_1 = 3, X_2 = 2, X_3 = 2, X_4 = 1, X_5 = 1, X_6 = 0)$$
$$= \frac{9!}{3!2!2!1!1!0!}\left(\frac{1}{6}\right)^3\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^0 = \frac{9!}{3!2!2!}\left(\frac{1}{6}\right)^9 \approx 0{,}0015$$

e inoltre

$$P(X_1+X_2+X_3 = 7, X_4 = 1, X_5+X_6 = 1) = \frac{9!}{7!1!1!}\left(\frac{1}{2}\right)^7\left(\frac{1}{6}\right)^1\left(\frac{1}{3}\right)^1 = \frac{1}{2^5} = 0{,}3125.$$

Esempio. Siano p, q, $1-p-q$ le probabilità che una squadra di calcio vinca, perda, pareggi una partita. Supponendo di giocare n partite indipendenti, con probabilità costanti, posto $X =$"numero di vincite" e $Y =$"numero di sconfitte", risulta

$$P(X = x, Y = y) = \frac{n!}{x!y!(n-x-y)!}p^x q^y (1-p-q)^{n-x-y}.$$

Variabili aleatorie indipendenti

Proposizione:

- X,Y variabili discrete
- X,Y sono indipendenti <=> $P_{x,y}(x,y) = P_x(x) * P_y(y)$, $\forall x, y \in R$

DIM. → Ipotesi: $P(X \in A, y \in B) = P(X \in A) * P(Y \in B) \quad \forall A, B \subseteq R$

scelgo A={a} B={y}

P(X=x, Y=y)= $P(X \in \{x\}, Y \in \{y\}) = P(X \in \{x\}) * P(Y \in \{y\}) = P(X = x, Y = y) = P_X(X) * P_y(Y)$

$P_{x,y}(X,Y)$

DIM ← **P(** $X \in A, y \in B$ **) =**

$$\sum_{X \in A}\sum_{Y \in B} P_{(x,y)} = \sum_{X \in A}\sum_{Y \in B} P_x(x) * P_y(y) = \sum_{x \in A} P_x(x) * \sum_{x \in B} P_y(y) = P(X \in A) * P(X \in B)$$

Esempio. Nell'esperimento che consiste nel lancio di una moneta ripetuto 3 volte, sia X il numero di volte che esce testa e sia Y il numero di variazioni, ossia quanti lanci danno risultati diversi dal lancio precedente. Stabilire se X e Y sono indipendenti.

Soluzione. È evidente che X e Y hanno distribuzione binomiale con parametri $(n, p) = (3, \frac{1}{2})$ e $(2, \frac{1}{2})$, rispettivamente. Ricaviamo ora la densità congiunta di (X, Y):

ω	$X(\omega)$	$Y(\omega)$
ccc	0	0
cct	1	1
ctc	1	2
ctt	2	1
tcc	1	1
tct	2	2
ttc	2	1
ttt	3	0

$x \backslash y$	0	1	2	$p_X(x)$
0	1/8	0	0	1/8
1	0	1/4	1/8	3/8
2	0	1/4	1/8	3/8
3	1/8	0	0	1/8
$p_Y(y)$	1/4	1/2	1/4	1

Risulta ad esempio $p(0,0) = \frac{1}{8} \neq p_X(0)\,p_Y(0) = \frac{1}{8} \cdot \frac{1}{4}$;
ne segue che X e Y non sono indipendenti.

- X, Y variabili aleatorie indipendenti e identicamente distribuite

$$X \sim N(\mu=1,\sigma^2=4) \qquad Y \sim N(\mu=1,\sigma^2=4) \qquad calcolare\, P(|X|<1,|Y|<1)=$$

$$=P(|X|<1)*P(|Y|<1)=[P(|X|<1)]^2=[P(-1<X<1)]^2=[P(\frac{-1-1}{2}<\frac{X-1}{2}<\frac{1-1}{2})]^2=$$

$$=[P(-1<Z<0)]^2=[\Phi(0)-\Phi(-1)]^2=[\frac{1}{2}-[1-\Phi(1)]]^2=[\Phi(1)-\frac{1}{2}]^2=0{,}1165$$

- In un esperimento si generano 3 bit aleatorie descritte da variabili aleatorie di Bernoulli indipendenti: B_1, B_2, B_3. Ciascuna di parametro $p\in[0,1]$ **.**

X = B_1 or B_2 (or logico)

Y = B_2 and B_3 (and logico)

1. Determinare $P_{x\,y}(X, Y)$
2. Stabilire per quali valori di p X ed Y sono indipendenti.

B_1	B_2	B_3	X	Y	prob
0	0	0	0	0	$(1 - p)^3$
0	0	1	0	0	$p(1 - p)^2$
0	1	0	1	0	$p(1 - p)^2$
0	1	1	1	1	$p^2(1 - p)$
1	0	0	1	0	$p(1 - p)^2$
1	0	1	1	0	$p^2(1 - p)$
1	1	0	1	0	$p^2(1 - p)$
1	1	1	1	1	p^3

Le sequenze non sono equiprobabili

X\Y	0	1	P(X=x)
0	$p(1 - p)^2$	0	$p(1 - p)^2$
1	2p(1-p)	p^2	p^2+2p(1-p)
P(Y=y)	$(1-p)^2$+ 2p(1-p)	p^2	1

$$2)\, 0=p_{x\,y}(0,1)=P_x(0)P_y(1)=(1-p^2)\,p^2$$

$$quindi: \forall\, x,y\; P_{xy}(x,y)=P_x(x)P_x(y) \text{ se } p=0\; oppure\; p=1$$

Somma di variabili aleatorie indiependenti

Supponiamo che vengano eseguite n+m prove indipendenti, ognuna delle quali abbia probabilit`a pari a p di risultare in un successo. Se X è il numero di successi nelle prime n prove e Y il numero di successi nelle m prove successive, allora X e Y sono indipendenti, in quanto conoscere il numero dei successi nelle prime n prove non modifica la distribuzione del numero di successi nelle ulteriori m prove, in virtù dell'ipotesi di indipendenza delle prove.

Infatti, per $x = 0, 1, \ldots, n$ e $y = 0, 1, \ldots, m$ si ha

$$P(X = x, Y = y) = P(X = x)\, P(Y = y) = \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{y} p^y (1-p)^{m-y}.$$

Osserviamo che se $Z = X + Y$ è il numero totale di successi nelle $n + m$ prove, allora X e Z sono dipendenti, ossia non indipendenti.

DIM:
Se X e Y sono due variabili aleatorie binomiali indipendenti di parametri, rispettivamente, (n, p) e (m, p), mostrare che X + Y `e binomiale di parametri (n + m, p).

Dalle ipotesi fatte, per $k = 0, 1, \ldots, n+m$ segue

$$\begin{aligned} P(X+Y=k) &= \sum_{i=0}^{k} P(X=i, Y=k-i) = \sum_{i=0}^{k} P(X=i)\,P(Y=k-i) \\ &= \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} \binom{m}{k-i} p^{k-i}(1-p)^{m-k+i} \\ &= p^k (1-p)^{n+m-k} \sum_{i=0}^{k} \binom{n}{i}\binom{m}{k-i} = \binom{n+m}{k} p^k (1-p)^{n+m-k}, \end{aligned}$$

avendo utilizzato l'identità combinatoria (formula di Vandermonde)

$$\binom{n+m}{k} = \sum_{i=0}^{k} \binom{n}{i}\binom{m}{k-i}.$$

Proprietà:

1. $X \sim binomiale(n, p) \qquad Y \sim binomiale(m, p) \rightarrow X+Y \sim binomial(n+m, p)$
2. $X_{1,} \ldots, X_n$ indipendenti $\quad (X_i \sim N(\mu_i, \sigma_i^2)) \rightarrow \sum_{i=1}^{n} X_i \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$

Distribuzione Condizionate
Date X,Y discrete, si definisce la densità discreta condizionata di X dato Y=y

$P_{xy}(X \mid Y) = P(X=x \mid Y=y) = \dfrac{P(X=x, Y=y)}{P(Y=y)} = \dfrac{P_{x,y}(X,Y)}{P_y(Y)}$

Se X ed Y sono indipendenti
- $P_{xy}(X \mid Y) = P_x(x)$

dim.

$P_{xy}(X \mid Y) = \dfrac{Pxy(X \mid Y)}{P(Y=y)} =*$ per ipotesi: x,y indipendenti quindi

$*= \dfrac{Px(X) * \cancel{P_y(y)}}{\cancel{P_y(y)}} = P_x(x)$

Esercizi
Un esperimento consiste nell'effettuare 5 estrazioni da un urna che contiene 2 biglie nere e 3 bianche.
X=estrazione in cui esce la prima biglia nera
Y=numero di biglie nere uscite nelle prime 3 estrazioni

Determinare la densità discreta condiziona di X dato Y=y
y=0,1,2

					X	Y
n	n	b	b	b	1	2
n	b	n	b	b	1	2
n	b	b	n	b	1	1
n	b	b	b	n	1	1
b	n	n	b	b	2	2
b	n	b	n	b	2	1
b	n	b	b	n	2	1
b	b	n	n	b	3	1
b	b	n	b	n	3	1
b	b	b	n	n	4	0

x \ y	0	1	2	P(X=x)
1	0	2/10	2/10	4/10
2	0	2/10	1/10	3/10
3	0	2/10	0	2/10
4	1/10	0	0	1/10
P(Y=y)	1/10	6/10	3/10	1

$$P_{xy}(x|0)=\frac{P_{xy}(x,0)}{P_y(0)}=\begin{cases} 0 & \text{se x=1} \\ 0 & \text{se x=2} \\ 0 & \text{se x=3} \\ \dfrac{\frac{1}{10}}{\frac{1}{10}}=1 & \text{se x=4} \end{cases}$$

$$P_{xy}(x|1)=\frac{P_{xy}(x,1)}{P_y(1)}=\begin{cases} \frac{2}{6} & \text{se x=1} \\ \frac{2}{6} & \text{se x=2} \\ \frac{2}{6} & \text{se x=3} \\ 0 & \text{se x=4} \end{cases}$$

$$P_{xy}(x|2)=\frac{P_{xy}(x,2)}{P_y(2)}=\begin{cases} \frac{2}{3} & \text{se x=1} \\ \frac{1}{3} & \text{se x=2} \\ 0 & \text{se x=3} \\ 0 & \text{se x=4} \end{cases}$$

VALORE ATTESO DI SOMME DI VARIABILI ALEATORIE

X,Y variabili discrete indipendenti

$$\sum_x \sum_y g(x)h(y)P_{x,y}(X,Y)=$$

E[g(X) * h(Y)]= per l'ipotese di indipendenza $\rightarrow = \sum_x \sum_y g(x)h(y)Px,y(X,Y)=$

$$=\left(\sum_x g(x)P_x(X)\right)*\left(\sum_Y h(y)P_y(Y)\right)=$$

$$=E(g(x))*E(h(y))$$

Se X e Y sono indipendenti E[x*y]= E[x]*E[y]

COVARIANZA
Cov(x,y)=E{[X-E(X)] * [Y- E(Y)]}

Cov (x,y)>0 → X,Y si dicono positivamente correlate
Cov(x,y)< 0→ X,Y si dicono negativamente correlate
Cov(x,y) = 0 → X,Y si dicono non correlate

Cov(x,y) = E{[X-E(X)]*[Y-E(Y)]}= E[X*Y − X*E(Y) − Y*E(X) + E[X]E[Y]]
= E(XY) − E(X) *E(Y) − E(X)E(Y) + E(Y) E(X) = E(Y) − E(X)E(Y).

Proprietà:

1. Se X e Y sono indipendenti → Cov(X,Y) = 0 (non vale il contrario)
2. Cov(X,Y)= Cov(Y,X)
3. Cov(X,X)= var(X)
4. Cov (aX,Y)= a Cov(X,Y)
5. Cov $\left(\sum_{I=1}^{n} X_i, \sum_{j=1}^{m} X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$

Dim. 3

Cov(X,X)=E[(X-E(Y))*(X-E(X))]= E[$(X-E(X))^2$] var(X);

Dim. 4

Cov (aX,Y)= E[aXY]-E[aX] * E[Y] = E[X*Y]-aE(X) * E(Y) = a{E(X*Y)-E(X) * E(Y)}=
=a Cov(X,Y)

Dim 5

$E(X_i)= \mu$ $E(Y_j)= v_j$

$$Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = E\left[\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)\right] * \left[\sum_{j=1}^{m} Y_j - \left[\sum_{j=1}^{m} Y_j\right] - E\left(\sum_{j=1}^{m} Y_j\right)\right] =$$
$$= E\left\{\left[\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right] * \left[\sum_{j=1}^{m} Y_j - \sum_{j=1}^{m} v_j\right]\right\} =$$
$$= E\left\{\left[\sum_{i=1}^{n} (X_i - \mu_i)\right] * \left[\sum_{j=1}^{m} (Y_j - v_j)\right]\right\} =$$
$$= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m} (X_i - \mu_i)(Y_j - v_j)\right] =$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{m} \underbrace{E[(X_1 - \mu_i)(Y_j - v_j)] =}_{Cov(X_i, Y_j)}$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$$

Esercizi:

Esempio. Nel lancio di una moneta ripetuto 3 volte, sia X il numero di volte che esce testa e sia Y il numero di variazioni, ossia quanti lanci danno risultati diversi dal lancio precedente. Calcolare $Cov(X,Y)$.

Soluzione. Abbiamo già ricavato la densità congiunta di (X,Y), da cui si trae che X e Y non sono indipendenti; inoltre si ricava:

$x\backslash y$	0	1	2	$p_X(x)$
0	1/8	0	0	1/8
1	0	1/4	1/8	3/8
2	0	1/4	1/8	3/8
3	1/8	0	0	1/8
$p_Y(y)$	1/4	1/2	1/4	1

$$E(X) = 0\cdot\frac{1}{8} + 1\cdot\frac{3}{8} + 2\cdot\frac{3}{8} + 3\cdot\frac{1}{8} = \frac{3}{2}$$

$$E(Y) = 0\cdot\frac{1}{4} + 1\cdot\frac{1}{2} + 2\cdot\frac{1}{4} = 1$$

$$E(XY) = 1\cdot 1\cdot\frac{1}{4} + 1\cdot 2\frac{1}{8} + 2\cdot 1\cdot\frac{1}{4} + 2\cdot 2\cdot\frac{1}{8} = \frac{3}{2}$$

e pertanto: $Cov(X,Y) = E(XY) - E(X)\,E(Y) = \frac{3}{2} - \frac{3}{2}\cdot 1 = 0.$

- Esercizio precedente ma usando le proprietà:

X_i = esito i-esimo del lancio = $\left\{\begin{array}{lr} 1 & \text{se esce testa all'i-esimo lancio} \\ 0 & \text{altrimenti} \end{array}\right\} con\, i = 1,2,3$

$X = X_1 + X_2$ $\qquad\qquad$ $Y = X_2 + X_3$

$X_i \sim Bernoulli\left(\frac{1}{2}\right)$

$E(X_i) = \frac{1}{2}$ $\qquad\qquad$ $var(X_i) = \frac{1}{4}$

$cov(X,Y) = cov(X_1 + X_2, X_2 + X_3) =$ uso la 5° proprietà =

$$= \underbrace{cov(X_1, X_2)}_{essendo\ indipendenti\ valgono\ 0} + \underbrace{cov(X_1, X_3)}_{essendo\ indipendenti\ valgono\ 0} + cov(X_2, X_2) + \underbrace{cov(X_2, X_3)}_{essendo\ indipendenti\ valgono\ 0} = var(X_2) = \frac{1}{4}$$

- X: S_x ={-2, -1, 0, 1, 2} con probabilità 1/5 ciascuna. Y = X^2. Calcolare cov(X, Y)

X* \ *Y	**0**	**1**	**4**	**P(X = x)**
-2	0	0	1/5	1/5
-1	0	1/5	0	1/5
0	1/5	0	0	1/5
1	0	1/5	0	1/5
2	0	0	1/5	1/5
P(Y = y)	1/5	1/5	1/5	***1***

$$E(X,Y) = -2*4*\frac{1}{5} + (-1)*1*\frac{1}{5} + 1*1*\frac{1}{5} + 2*4*\frac{1}{5} = 0$$

$$E(X) = -2*\frac{1}{5} + (-1)*\frac{1}{5} + 1*\frac{1}{5} + 2*\frac{1}{5} = 0$$

cov(X, Y) = 0 – 0 * E(Y) = 0

VALORE ATTESO PER DUE VARIABILI

X,Y discrete con densità discreta congiunta $P_{x,y}(X,Y)$ allora E[g(x,y)]= $\sum_x \sum_y g(x,y) * P_{xy}(X,Y)$

Ponendo
g(x,y)=x+y
E[x+y]= $\sum_x \sum_y (x+y)*P_{xy}(X,Y) \;=\; \sum_x \sum_y x*P_{xy}(X,Y)+\sum_x \sum_y y*P_{xy}(X,Y) \;=$

$$\sum_x x\sum_y P_{xy}(X,Y)+\sum_y y\sum_x P_{xy}(X,Y)=\sum_x xP_x(X)+\sum_y y\,P(x,y)=E(x)+E(y)$$

Proprietà
- $E(x)=\sum_x x*P_x(X)$
- $E[g(x)]=\sum_x g(x)P_x(x)$
- $E(x+y)=E(x)+E(y)$
- $E(x_{1,\dots\dots},x_n)=E(x_1)+\dots\dots+E(x_n)$

Esercizi:
-X_1, …..,X_n indipendenti e identicamente distribuite (~F)

E(x) = μ i= 1….n

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n} X_i \qquad media\ \ campionaria$$

$$E(\overline{X})=E(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n}E(\sum_{i=1}^{n}X_i)=\frac{1}{n}\{\sum_{i=1}^{n}E(X_i)\}=\frac{1}{\cancel{n}}*\cancel{n}\,\mu=\mu$$

$$X\sim Binomiale(n,p)$$

\- $X_i=\begin{cases}1 & se\,l'\,i-esimo\ prova\ è\ un\,successo\\ 0 & altrimenti\ i=1,\dots,n\end{cases}$

X = numero di successi nella n prove.
Se sommo $X_1 + X_2 + ... + X_n = X$ --> E(X) = = $E(X_1) + E(X_2) + ... + E(X_n)$ = n * p

Legge della varianza

$$var(\sum_{i=1}^{n}X_i)=\sum_{i=1}^{n}var(X_i)+2\sum_{i<j}cov(X_i,Y_j).\mathrm{In}\ particolare:$$

1. Se X_i e X_j sono non correlate $\forall i,j \rightarrow var(\sum_{i=1}^{n}X_i)=\sum_{i=1}^{n}var(X_i)$

2. Se X_i, ..., X_n sono indipendenti $\rightarrow var(\sum_{i=1}^{n}X_i)=\sum_{i=1}^{n}var(X_i)$

Dim:

$$var(\sum_{i=1}^{n}X_i)=cov(\sum_{i=1}^{n}X_i,\sum_{j=1}^{n}X_j)=uso\,l'\,additività=\sum_{i=1}^{n}\sum_{j=1}^{n}cov(X_i,X_j)=$$

$$=\sum_{i=1}^{n}cov(X_i,X_i)+\sum_{i\neq j}^{n}cov(X_i,X_j)=\sum_{i=1}^{n}var(X_i)+2\sum_{i<j}cov(X_i,Y_j)\,per\ la\ proprietà\ di\ simmetria.$$

Esempio. Siano $X_1, \dots, X_n$ variabili indipendenti ed identicamente distribuite con valore atteso μ e varianza σ^2, e sia $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ la media campionaria. Le variabili

$$E(\overline{X})=\mu$$

$$var(\overline{X})=var(\frac{1}{n}\sum_{i=1}^{n} X_i)=\frac{1}{n^2}var(\frac{1}{n}\sum_{i=1}^{n} X_i)=\frac{1}{n^2}*\sum_{i=1}^{n} var(X_i)$$

perchè tutte le varianze sono indipendenti.=

$=\frac{1}{n^2}*n\sigma^2$ perchè tutte hanno la stasse varianza che è σ^2.=

$=\frac{\sigma^2}{n}$, *se facciamo tendere* $n \rightarrow +\infty$ *risulta uguale a* 0

$X \sim Binominale(n,p)$. *Definisco* $X_i=\begin{cases}1 & \text{se l'i-esima prova è un successo} \\ 0 & \text{altrimenti}\end{cases}$ *con* $i=1,\dots,n$

- $X=X_1+X_2+\dots+X_n * X_i$ *indipendenti*, $X_i \sim Bernoulli(p)$.

  $var(X)=var(\sum_{i=1}^{n} X_i)=\sum_{i=1}^{n} var(X_i)=\sum_{i=1}^{n} var\underbrace{(X_i)}_{p(1-p)}=n*p(1-p)$

Coefficiente di correlazione

$$\rho(X,Y)=\frac{cov(X,Y)}{\sqrt{var(X)Var(Y)}}$$

Proprietà:

1. ρ *è adimensionale* ($\rho = rho$)
2. $-1 \leq \rho \leq 1$
3. $|\rho|=1 \rightarrow Y=aX+b \begin{cases}\rho=1 & \rightarrow a>0 \\ \rho=-1 & \rightarrow a<0\end{cases}$

Esercizi:

Esercizio Nel lancio di un dado ripetuto n volte, sia X il numero di volte che esce 6 e Y il numero di volte che esce 5. Calcolare il coefficiente di correlazione di (X,Y).

Soluzione. Per $i,j = 1,2,\dots,n$ poniamo

$$X_i = \begin{cases}1 & \text{se esce 6 al lancio } i\text{-esimo} \\ 0 & \text{altrimenti,}\end{cases} \qquad Y_j = \begin{cases}1 & \text{se esce 5 al lancio } j\text{-esimo} \\ 0 & \text{altrimenti,}\end{cases}$$

da cui si segue che $X = \sum_{i=1}^{n} X_i$ e $Y = \sum_{j=1}^{n} Y_j$. Risulta

$$E(X_i) = E(Y_j) = \frac{1}{6}, \qquad E(X_iY_j) = \begin{cases}0 & \text{se } i=j, \\ \frac{1}{36} & \text{altrimenti,}\end{cases} \qquad \mathrm{Var}(X_i) = \mathrm{Var}(Y_j) = \frac{5}{36},$$

e pertanto

$$\mathrm{Cov}(X_i,Y_j) = E(X_iY_j) - E(X_i)E(Y_j) = \begin{cases}-\frac{1}{36} & \text{se } i=j, \\ 0 & \text{altrimenti,}\end{cases}$$

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(X_i,Y_j) = -\frac{n}{36},$$

e, poiché X e Y sono variabili binomiali,

$$\mathrm{Var}(X) = \mathrm{Var}(Y) = n\frac{5}{36}.$$

Infine segue:

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{-n/36}{n\,5/36} = -\frac{1}{5}.$$

Si noti che $\rho(X,Y)$ non dipende da n, a differenza di $\mathrm{Cov}(X,Y)$.

X \ Y	0	1	P(X=x)
0	p	$\frac{1}{2}-p$	$\frac{1}{2}$
1	$\frac{1}{2}-p$	p	$\frac{1}{2}$
P(Y=y)	$\frac{1}{2}$	$\frac{1}{2}$	1

1. Determinare i valori ammissibili di p
2. Stabilire per quali valori di p X ed Y siano indipendenti
3. Calcola il coefficiente di correlazione al variare di p

$$\sum_x \sum_y P_{x,y}(X,Y)=1 \qquad \forall p$$

1) $\begin{cases} 0 \le p \le 1 \\ 0 \le \frac{1}{2} \le 1 \end{cases} \rightarrow 0 \le p \le \frac{1}{2}$

2) $P_{xy}=P_x(X)*P_y(Y) \quad \forall x,y$

$P_{xy}(0,0)=p$

$P_x(0)*P(0)= \frac{1}{2}*\frac{1}{2}=\frac{1}{4}$ quindi poniamo p= $\frac{1}{4}$

X \ Y	0	1	P(X=x)
0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
P(Y=y)	$\frac{1}{2}$	$\frac{1}{2}$	1

$\forall x= 0,1 \quad \forall y=0,1$

$P_{xy}(X,Y)=P_x(X)*P_y(Y)$

Se p= ¼ → X ed Y sono indipendenti

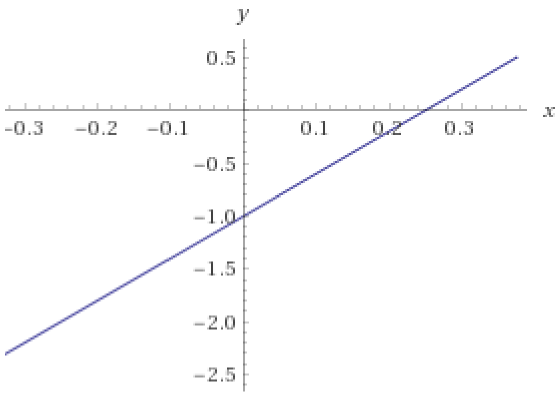
3) Cov(X,Y)=E(X*Y)-E(X)*E(Y)= p - $\frac{1}{4}$

E(X*Y)= 1*1*p=p; E(Y)=0* $\frac{1}{2}$ + $\frac{1}{2}$ = $\frac{1}{2}$; E(Y)=0* $\frac{1}{2}$ + $\frac{1}{2}$ = $\frac{1}{2}$

var(X)=E(X^2)-{E(X)}^{2} E(X^2)=0^2* $\frac{1}{2}$ +1^2* $\frac{1}{2}$ = $\frac{1}{2}$ quindi var(X)= $\frac{1}{2}$ * $\frac{1}{4}$ = $\frac{1}{4}$

var(Y)= $\frac{1}{4}$

$$\rho(X,Y)= \frac{p-\frac{1}{4}}{\sqrt{\frac{1}{4}*\frac{1}{4}}}=\frac{p-\frac{1}{4}}{\frac{1}{4}}=4p-1 \qquad 0\leq p=\frac{1}{2}$$



p= $\frac{1}{4} \rightarrow \rho(X,Y)=0$

p=0

Y \ X	0	1	P(X=x)
0	0	$\frac{1}{2}$	$\frac{1}{2}$
1	$\frac{1}{2}$	0	$\frac{1}{2}$
P(Y=y)	$\frac{1}{2}$	$\frac{1}{2}$	1

Se p=0

$\boxed{Y=1-X}$

p(Y=1-X)= $P_{xy}(0,1)$+ $P_{xy}(1,0)$= $\frac{1}{2}*\frac{1}{2}$ =1

p= $\frac{1}{2}$

Y \ X	0	1	P(X=x)
0	$\frac{1}{2}$	0	$\frac{1}{2}$
1	0	$\frac{1}{2}$	$\frac{1}{2}$
P(Y=y)	$\frac{1}{2}$	$\frac{1}{2}$	1

$P_{xy}(Y=X)=P_{xy}(0,0)=P_{xy}(1,1)=P_{xy}(0,0)+P_{xy}(1,1)=$

½ + ½ =1

$\boxed{Y=X}$

DISEGUAGLIANZA DI MARKOV

Sia X una variabile aleatoria non negativa. ∀a>0 p(X≥a)≤ $\frac{E(X)}{a}$

dim:

variabile indicatrice → I= $\begin{cases} 1 & \text{se } X \geq a \\ 0 & \text{altrimenti} \end{cases}$

I ≤ $\frac{X}{a}$ da questo deriva che:

E(I)≤ $E\left(\frac{X}{a}\right)$ = $\frac{1}{a}E(X)$

E(I)=1*P(I=1) + 0*P(I=1)=P(X≥a) quindi ottengo che p(X≥a)≤ $\frac{1}{a}E(X)$

DISEGUAGLIANZA DI CHEBYSHEV

X variabile aleatoria di valore atteso E(X)=μ finito e varianza var(x)= σ^2 finita

∀ k>0

1) P(|X-μ|≤k)≤ $\frac{\sigma^2}{k^2}$
2) P(|X-μ|) ≥ 1- $\frac{\sigma^2}{k^2}$

Dim 1

Usando la diseguaglianza di Markov P(Y≥a) ≤ $\frac{E(Y)}{k^2}$ Y= $(X-\mu)^2$ a=k^2

$$P((X-\mu)^2 \geq k^2) \leq \frac{E(X-\mu)^2}{k^2} = \frac{var(x)}{K^2} = \frac{\sigma^2}{k^2}$$

LEGGE DEBOLE DEI GRANDI NUMERI

$X_1,X_2,\ldots$.. variabili aleatorie indipendenti e identicamente distribuite

$E(X_i)= \mu$; $\sigma=var(X_i)$; μ,σ^2 finite $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\forall \epsilon > 0$

la legge dice che il $\lim_{n\to+\infty} P(|\overline{X}-\mu| \geq \epsilon)=0$

dim.

Assumiamo inoltre che $Var(X_i) = \sigma^2$ sia finita. Poiché E[$\overline{X_n}$] = μ e Var[$\overline{X_n}$] = σ^2/n, dalla disuguaglianza di Chebyshev si ha, per ogni ε > 0

$$P(|\overline{X_n} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\,\epsilon^2}$$

applicando il limite n → ∞ otteniamo che il $\lim_{n\to+\infty} \frac{\sigma^2}{nk^2}$ tende a 0 ma la probabilità

$P(|\overline{X_n} - \mu| \geq \epsilon)$ non può essere negativa ma al più può essere 0 perciò il lim =0 :

TEOREMA CENTRALE DEL LIMITE
Sia X1, X2, . . . una successione di variabili aleatorie indipendenti e identicamente distribuite , ognuna di media μ e varianza $\sigma^2 > 0$ finite

$E(X_i)= \mu$ e var(σ^2) μ ,σ^2 finite

$$Z_n = \frac{(X_1+X_2+....+X_n)-n\mu}{\sqrt{n\sigma^2}} = \frac{(X_1+X_2+....+X_n)-E(X_1+X_2+....+X_n)}{\sqrt{var(X_1+X_2+....+X_n)}}$$

$\forall a \in R \quad \lim_{n\to\infty} P(Z_n \le a) = \Phi(a)$ con n grande P(Zn≤a)≈ $\Phi(a)$

$\Phi(x)$ è la funzione di distribuzione di una variabile aleatoria normale standard.

ESERCIZIO
Un virus attacca una cartella contenente 200 files. Ciascun file è danneggiato con probabilità 0,2 indipendentemente dagli altri files . Calcolare la probabilità che siano danneggiati meno di 50 files

X=numero di files danneggiati
X~Binomiale(n=200,p=0,2)

P(X<50)= $P\left(\frac{\left(X-200\frac{2}{10}\right)}{\sqrt{200*\frac{2}{10}*\frac{8}{10}}}<50-\frac{40}{\sqrt{32}}\right)$ per il TCL e circa il p(Z< $\frac{10}{\sqrt{32}}$)= $\Phi(\frac{10}{\sqrt{32}})=0,95$ circa

ESERCIZIO
$X_1,X_2,….,X_{36}$ indipendente e identicamente distribuite

$E(X_i)$=10 sec. $var(X_i)$=16 sec^2

$Y=X_1,X_2,….,X_{36}$

1) Calcolare E(Y) e var(Y)
2) P(Y>6 minuti | Y<7 minuti) → calcolare in maniera approssimata

1) E(Y)= 36*10 sec= 360 sec
var(Y)= 36*16 sec^2 = 576 sec^2

2) ***P(Y>6 minuti | Y < 7 minuti)= P(Y>360 sec | Y<420 sec)=***

$$\frac{P(360<y<420)}{P(y<420)} = \frac{P(\frac{360-360}{24}<\frac{y-360}{24})<420-\frac{360}{24}}{P(\frac{y-360}{24}<\frac{420-360}{24})}$$ ***per il TCL***

$$\Phi\frac{(2,5)-\Phi(0)}{\Phi(2,5)}=circa\ 0,496$$

Statistica descrittiva
L'analisi statistica fornisce la chiave di lettura per interpretare dati a prima vista "rumorosi" e imperscrutabili, ricavandone informazioni reali, o almeno attendibili.
Alcune volte l'analisi statistica parte da un campione di dati definito
In altri casi occorre raccogliere i dati, ed il primo obiettivo consiste nell'ideare un procedimento ottimale per la loro raccolta
Questa parte della statistica è detta statistica descrittiva.

La statistica è solitamente finalizzata ad ottenere informazioni su un insieme completo di unità, detto popolazione.

Quando tale insieme è troppo grande (ad esempio, gli utenti di un social network) si cerca di ricavare informazioni sulla popolazione scegliendo ed esaminando un sottoinsieme di loro elementi, detto campione.
Il campione deve contenere informazioni sulla popolazione complessiva, e quindi deve essere in qualche senso rappresentativo della popolazione.

Tabelle e grafici delle frequenze
Le tecniche di organizzazione dei dati variano in relazione alle caratteristiche degli elementi da esaminare; vi sono
– caratteri qualitativi quando tali caratteristiche sono di natura non numerica, e
– caratteri quantitativi quando esse sono grandezze misurabili, e quindi distinguibili in caratteri quantitativi discreti e continui.

Supponiamo di avere un insieme $(x_1, x_2, \ldots, x_n)$ di n dati, che costituiscono un campione di ampiezza o numerosità n.
Se il carattere è di tipo discreto, ed i valori assumibili dal carattere sono in numero relativamente basso, conviene raggruppare i dati considerando l'insieme dei valori assumibili, detti solitamente modalità del carattere, ed associare ad ognuno di essi il numero di volte che esso compare nel campione $(x_1, x_2, \ldots, x_n)$.
Possiamo indicare con $v_1, v_2, \ldots, v_k$ le modalità di un carattere, ossia i valori distinti assunti dai dati.
Per semplicità supponiamo che le modalità siano ordinate in senso crescente: $v_1 < v_2 < \ldots < v_k$.

Ad esempio, consideriamo le temperature registrate alle ore 06:00 nelle principali località italiane in un giorno fissato:

Bari	16	Milano	17
Bologna	21	Napoli	20
Cagliari	16	Palermo	20
Catania	16	Roma	20
Firenze	18	Torino	14
Genova	21	Venezia	17

In tal caso il campione (di ampiezza n = 12) è dato da $(x_1, x_2, \ldots, x_n) = (16, 21, 16, 16, 18, 21, 17, 20, 20, 20, 14, 17)$

Vi sono k = 6 modalità: $v_1, v_2, \ldots, v_k = 14, 16, 17, 18, 20, 21$.

Possiamo definire vari tipi di frequenze per ogni modalità.
- f_i : frequenza assoluta della modalità vi , ossia il numero di dati del campione $(x_1, \ldots, x_n)$ che hanno valore vi , per $i = 1, 2, \ldots, k$ $0 \le f_i \le n$ per ogni i, $f_1 + f_2 + \cdots + f_k = n$.
Nel caso del campione

$$(x_1, x_2, \ldots, x_n) = (16, 21, 16, 16, 18, 21, 17, 20, 20, 20, 14, 17)$$

le frequenze assolute delle modalità

$v_1, v_2, \ldots, v_k = 14, 16, 17, 18, 20, 21$

sono:

$f_1 = 1, f_2 = 3, f_3 = 2, f_4 = 1, f_5 = 3, f_6 = 2.$

Possiamo definire anche le frequenze in forma cumulativa, ricordando che le modalità sono ordinate in senso crescente: $v_1 < v_2 < \ldots < v_k$.

- Fi : frequenza cumulativa assoluta delle modalità $v_1, \ldots, v_i$, ossia numero di dati del campione (x1, . . . , xn) che hanno valore minore o uguale a vi , per i = 1, 2, . . . , k

$$F_i = \sum_{j=1}^{i} f_j \qquad 0 \le F_i \le n \; per \; ogni \; i \, , F_n = n$$

Notiamo che vale la formula di ricorrenza: $F_i = F_i - 1 + f_i$ (i = 1, 2, . . . , k) con $F_0 = 0$

Esempio Il numero di errori riscontrati nel testing di n = 80 moduli software costituisce il campione qui riportato:

2, 7, 4, 2, 4, 5, 2, 4, 5, 3, 7, 6, 3, 2, 4, 6, 5, 6, 4, 5, 5, 5, 4, 8, 5, 6, 5, 3, 3, 4, 2, 3, 3, 5, 4, 7, 3, 2, 3, 6, 3, 4, 4, 2, 5, 4, 2, 4, 5, 4, 2, 5, 4, 3, 4, 7, 4, 6, 8, 4, 7, 4, 4, 8, 2, 4, 6, 3, 6, 4, 2, 4, 3, 6, 5, 4, 5, 4, 3, 5.

Si nota che vi sono k = 7 modalità: 2, 3, 4, 5, 6, 7, 8.

Nella tabella sono riportate le frequenze assolute e cumulative:

i	v_i	f_i	F_i
1	2	11	11
2	3	13	24
3	4	24	48
4	5	15	63
5	6	9	72
6	7	5	77
7	8	3	80

-p_i : frequenza relativa della modalità vi , ossia rapporto tra la frequenza assoluta f_i e l'ampiezza n del campione, per i = 1, 2, . . . , k

$$p_i = \frac{f_i}{n} \, , \quad 0 \le p_i \le 1 \quad per \; ogni \;\; i \, , \qquad p_1 + p_2 + \ldots + p_k = 1$$

- P_i : frequenza cumulativa relativa delle modalità v1, . . . , vi , ossia somma delle frequenze relative delle modalità v1, . . . , vi , per i = 1, 2, . . . , k

$$P_i = \sum_{j=1}^{i} P_j \, , \quad 0 \le P_i \le 1 \quad per \; ogni \;\; i \, , \qquad P_k = 1$$

Poichè

$$F_i = \sum_{j=1}^{i} f_j \, , \;\; \text{risulta} \quad P_i = \frac{1}{n} \sum_{j=1}^{i} f_j = \frac{F_I}{n} \qquad \text{per ogni i}$$

Riprendendo l'esempio precedente possiamo riportare anche le frequenze relative nella tabella.

i	v_i	f_i	p_i	F_i	P_i
1	2	11	0,1375	11	0,1375
2	3	13	0,1625	24	0,3
3	4	24	0,3	48	0,6
4	5	15	0,1875	63	0,7875
5	6	9	0,1125	72	0,9
6	7	5	0,0625	77	0,9625
7	8	3	0,0375	80	1,0

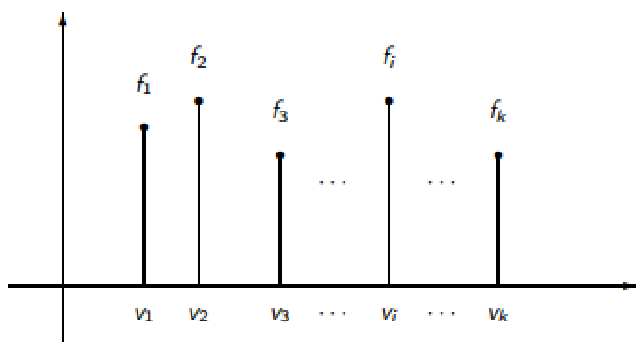
Algoritmo per determinare frequenze assolute e relative

```
 for j = 1 to k
          fj = 0;
          Fj = 0;
          pj = 0;
          Pj = 0
 for i = 1 to n
          j = MODAL{xi}
          fj = fj + 1
          F0 = 0
for j = 1 to k
          Fj = Fj−1 + fj
          pj = fj/n
          Pj = Fj/n
```

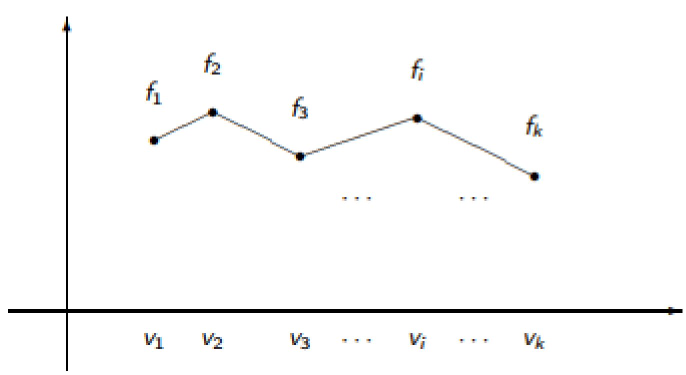
funzione MODAL{x_i}
restituisce l'indice j della modalità corrispondente al dato x_i .

Per rappresentare graficamente la distribuzione delle frequenze di un campione di dati sono molto usati tre tipi di grafici.

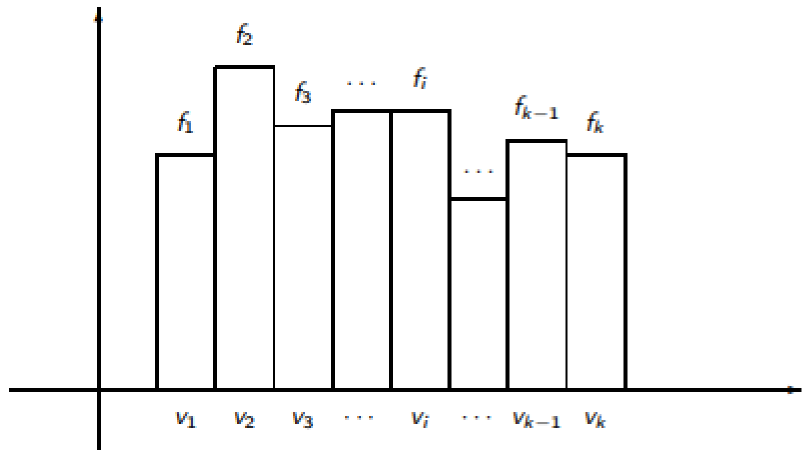
-Grafico ad aste (line graph) Sull'asse delle ascisse sono riportate le modalità; sull'asse delle ordinate è riportata la frequenza assoluta di ciascun valore, rappresentata da linee verticali.



-Grafico a linee (polygon graph) Sull'asse delle ascisse sono riportate le modalità; sull'asse delle ordinate vi è la frequenza assoluta di ciascun valore, ed i punti sono uniti da una spezzata.



-Grafico a barre (bar graph)
Sull'asse delle ascisse sono riportate le modalità; sull'asse delle ordinate è riportata la frequenza assoluta di ciascun valore, rappresentata da rettangoli.



Esempio Il numero di esami superati da n = 71 studenti costituisce il seguente campione:

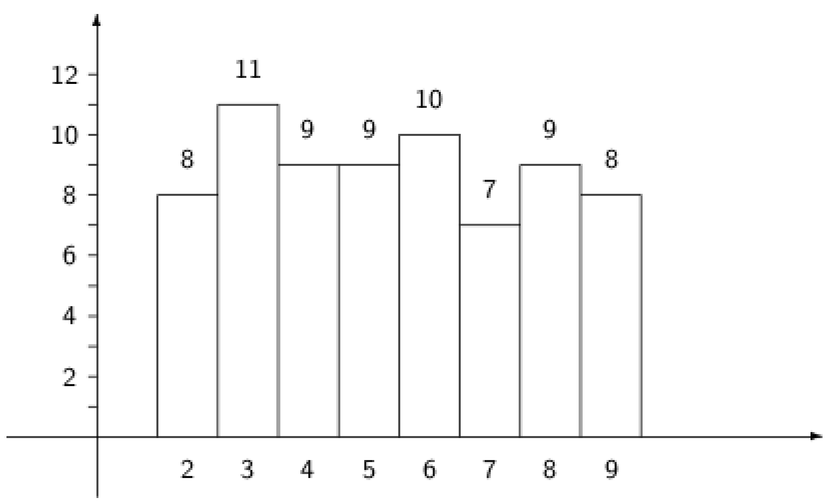
2, 3, 9, 2, 4, 4, 2, 7, 9, 7, 8, 6, 8, 6, 6, 5, 7, 4, 2, 9, 2, 4, 6, 4
5, 8, 5, 6, 5, 9, 3, 3, 2, 7, 6, 7, 9, 4, 5, 8, 6, 8, 6, 6, 5, 3, 2, 3
3, 4, 3, 6, 7, 8, 9, 7, 4, 3, 4, 3, 5, 5, 9, 9, 3, 3, 8, 5, 8, 2, 8

Si nota che vi sono k = 8 modalità: 2, 3, 4, 5, 6, 7, 8, 9.

Nella tabella sono riportate le frequenze assolute e relative:

i	v_i	f_i	p_i	F_i	P_i
1	2	8	0,1127	8	0,1127
2	3	11	0,1549	19	0,2676
3	4	9	0,1268	28	0,3944
4	5	9	0,1268	37	0,5211
5	6	10	0,1408	47	0,6620
6	7	7	0,0986	54	0,7606
7	8	9	0,1268	63	0,8873
8	9	8	0,1127	71	1,0

Grafico a barre corrispondente:



Un altro tipo di rappresentazione grafica tra le più comuni è il grafico a torta, utile per rappresentare caratteristiche qualitative.

Si costruisce tracciando un cerchio e suddividendolo in tanti settori circolari quante sono le categorie distinte di dati, con ogni settore avente angolo al centro proporzionale alla frequenza (relativa o assoluta, è lo stesso) della categoria corrispondente.

Esempio Dati sulle province della Campania (al 01/01/2015)

Provincia	Popolazione *residenti*	Superficie km^2	Densità $abitanti/km^2$	Numero Comuni
Napoli	3.118.149	1.178,93	2.645	92
Caserta	924.614	2.651,35	349	104
Salerno	1.108.509	4.954,16	224	158
Avellino	427.936	2.806,07	153	114
Benevento	282.231	2.080,44	136	78
TOTALE	5.861.529	13.670,95	429	550

Raggruppamento dei dati

Le metodologie statistiche esaminate finora sono utili soprattutto quando il numero di modalità non è troppo elevato. Quando questo requisito non è soddisfatto, i dati possono essere raccolti in gruppi di valori contigui, o classi.

La scelta di quante classi adottare è importante:

- se si prendono poche classi si perde troppa informazione sulla posizione dei dati all'interno delle classi;
- se si prendono troppe classi le frequenze di ciascuna classe assumono valori piccoli e non si riconosce la forma della distribuzione.

Istogramma

Valori tipici per il numero di classi sono tra 5 e 10, ma la scelta migliore va fatta in modo soggettivo ed empirico, anche provando più soluzioni.

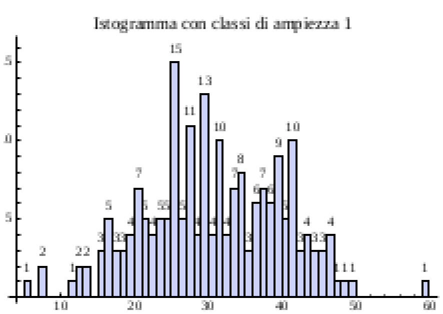
I bordi di una classe sono gli estremi del suo intervallo. Spesso i software statistici adottano la convenzione di includere i bordi di sinistra, nel senso che l'intervallo di una classe contiene il suo estremo sinistro e non il suo estremo destro, ad esempio [20, 30).

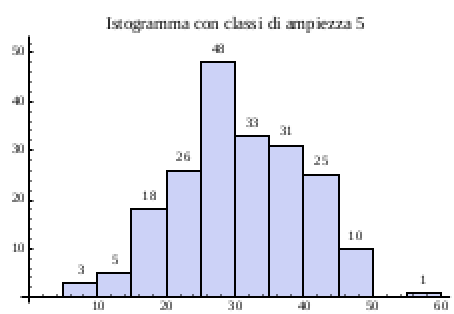
Il grafico a barre delle frequenze delle classi prende il nome di istogramma. Se invece sono riportate le frequenze relative, il grafico prende il nome di istogramma delle frequenze relative.

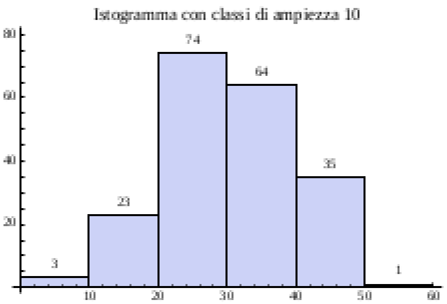
In tali definizioni ci si riferisce a classi tutte della stessa ampiezza. Se le classi non sono tutte della stessa ampiezza allora gli istogrammi si rappresentano rendendo le aree delle barre proporzionali alle frequenze delle classi

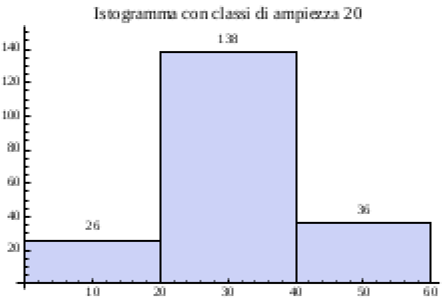
Esempio Un campione di numerosità $n = 200$: tempi di vita di componenti meccanici.

32	39	41	19	31	13	33	15	33	20	16	27	28	31	25	38	27
29	17	36	37	38	17	39	24	39	41	29	25	27	37	21	25	32
24	45	41	29	59	40	25	27	22	30	43	27	39	40	22	29	21
34	38	34	38	31	46	31	36	30	38	21	41	12	33	43	18	11
23	41	32	25	28	23	27	27	27	25	31	28	48	43	36	36	36
31	39	19	35	24	20	34	31	46	24	42	18	44	42	34	31	37
44	41	27	23	22	39	29	26	26	16	32	37	29	29	7	31	20
40	41	37	46	40	34	44	39	20	26	49	29	16	16	47	15	37
39	46	41	33	17	25	40	16	41	7	35	21	25	33	35	20	34
28	23	21	20	29	27	31	25	19	15	34	39	23	20	26	25	5
37	33	42	29	38	25	22	34	45	25	29	33	25	29	26	25	19
25	36	29	43	12	45	30	27	24	30	41	13	18				






Istogramma con classi di ampiezza 10
80
60
40
20
3
23
74
64
35
1
10
20
30
40
50
60


Istogramma con classi di ampiezza 20
140
120
100
80
60
40
20
26
138
36
10
20
30
40
50
60

Statistiche descrittive: indici di posizione
Le statistiche descrittive sono delle misure, o indici, che racchiudono informazioni essenziali per la descrizione di un campione di dati $x_1, x_2, \dots, x_n$ osservati. Le principali statistiche descrittive si dividono in indici di posizione e indici di variabilità.. Gli indici di posizione (o di tendenza centrale) sono misure che consentono di sintetizzare i dati osservati $x_1, x_2, \dots, x_n$ con un solo valore numerico che sia rappresentativo dei dati stessi. Gli indici di posizione più adoperati sono tre (la media, la mediana e la moda), da scegliere a seconda degli aspetti di maggiore interesse.

Media campionaria
Supponiamo di avere un campione di ampiezza o numerosità n, ossia un insieme $x_1, x_2, \dots, x_n$ di n dati. Definizione Si dice media campionaria e si denota con x, la quantità

$$\overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n} \qquad \text{(media aritmetica dei dati).}$$

Notiamo che, se si pone $y_i := ax_i + b$ con $i = 1, 2, \dots, n$
con a e b costanti, allora la media campionaria di y1, y2, . . . , yn è legata a quella dei dati iniziali dalla seguente relazione lineare: $\overline{y} = a\overline{x} + b$.
Infatti, se si pone $y_i := ax_i + b$ per $i = 1, 2, \dots, n$, con a e b costanti, allora la media campionaria di $y_1, y_2, \dots, y_n$ è data da:

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b)$$

$$= \frac{1}{n}\sum_{i=1}^{n} ax_i + \frac{1}{n}\sum_{i=1}^{n} b$$

$$= a\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}(nb)$$

$$= a\overline{x} + b.$$

Esercizio:
Riportiamo le reti segnate dai vincitori della classifica marcatori di Serie A di calcio negli anni dal 2004-05 al 2015-16: 24, 31, 26, 21, 25, 29, 28, 28, 29, 22, 22, 36
Per calcolare la media campionaria effettuiamo la trasformazione lineare $y_i = x_i - 25$, che conduce al seguente insieme di dati: −1, 6, 1, −4, 0, 4, 3, 3, 4, −3, −3, 11. La media campionaria dei dati trasformati si calcola facilmente:

$$\overline{y} = \frac{1}{12}(-1+6+1-4+0+4+3+3+4-3-3+11) = \frac{21}{12} = 1{,}75$$

Pertanto, essendo $\overline{y} = \overline{x} - 25$, ne segue che:

$$\overline{x} = \overline{y} + 25 = 1{,}75 + 25 = 26{,}75$$

Notiamo che se si conoscono le distribuzioni di frequenza dei dati $x_1, x_2, \ldots, x_n$, la media campionaria può essere calcolata anche come:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} v_i f_i = \sum_{i=1}^{k} v_i p_i \qquad \left(\text{invece di } \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

dove $v_1, v_2, \ldots, v_k$ sono le modalità dei dati, mentre f_i è la frequenza assoluta e $p_i = f_i/n$ la frequenza relativa della modalità i-esima.

La formula

$$\overline{x} = \frac{f_1}{n} v_1 + \ldots + \frac{f_k}{n} v_k$$

mostra che la media campionaria può essere vista come *media pesata* dei valori assunti dai dati, dove ogni valore ha come peso la sua frequenza relativa.

Mediana campionaria

Un'altra statistica che indica il centro di un insieme di dati è la mediana campionaria. È un numero che precede tanti dati quanti ne segue, ossia il valore centrale dei dati riordinati in senso crescente.

Definizione Dato un campione di ampiezza o numerosità n, ossia un insieme $x_1, x_2, \ldots, x_n$ di n dati, indichiamo con $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ tali valori riordinati in senso crescente.

Si dice mediana campionaria e si denota con $\tilde{x}$, la quantità:

$$\tilde{x} = \begin{cases} x_{(r+1)} & \text{se } n = 2r+1, \\ [x_{(r)} + x_{(r+1)}]/2 & \text{se } n = 2r; \end{cases}$$

Quindi, la mediana è data

- dall'elemento di posto centrale se n è dispari,
- dalla media aritmetica tra i due elementi centrali se n è pari.

Esercizio:

Consideriamo le frequenze delle età dei 76 studenti di un corso universitario.

Età (v_i)	19	20	21	22	23
Frequenza (f_i)	12	27	23	10	4

Poiché i dati sono n = 76, che è un numero pari, occorre calcolare la mediana campionaria come media aritmetica dei valori che occupano le posizioni 38 e 39 in ordine crescente:

$\tilde{x} = [x_{(38)} + x_{(39)}]/2 = 20$

Notiamo che la mediana campionaria è inferiore alla media campionaria $\overline{x} = 20{,}84$.

Media campionaria e mediana campionaria sono entrambe statistiche utili per descrivere i valori centrali dei dati.

La media fa uso di tutti i dati, ed in particolare è influenzata in maniera sensibile da valori eccezionalmente alti o bassi.

La mediana invece dipende direttamente solo da uno o due valori al centro della distribuzione e non risente dei dati estremi. La decisione di quale scegliere dipende dall'uso che se ne intende fare. Ad esempio, la media campionaria dei redditi delle persone di un comune è idonea a stimare il gettito fiscale complessivo, mentre la mediana campionaria è utile per la stima del potere d'acquisto del ceto residenziale medio.

Esercizio:

Profondità (in Km) dei terremoti registrati in Italia in 2 giorni (dati ordinati), con media e mediana campionaria:

$n = 27$ (2 settembre 2015): 5,0 6,5 7,1 7,5 8,1 8,2 9,2 9,5 9,7 9,9 10,0 10,0 10,0 $\boxed{10{,}0}$ 10,1 10,5 10,6 10,6 10,7 10,8 11,0 13,5 15,9 18,0 20,3 29,8 128,3

$$\overline{x} = \frac{420{,}8}{27} = 15{,}6 \qquad \tilde{x} = 10{,}0$$

$n = 33$ (3 settembre 2015): 1,0 1,9 2,3 5,7 7,3 7,7 7,7 8,1 8,1 8,4 8,5 8,6 9,4 9,9 10,0 10,0 $\boxed{10{,}0}$ 10,2 10,4 10,4 10,5 10,6 10,6 10,8 11,1 17,5 20,4 21,4 23,0 23,6 30,1 33,6 36,7

$$\overline{x} = \frac{415{,}5}{33} = 12{,}6 \qquad \tilde{x} = 10{,}0$$

I due campioni hanno uguale mediana campionaria, ma le medie sono molto differenti per un valore elevato nel primo campione.

Moda campionaria

Un'altra statistica che si usa per rappresentare una distribuzione di dati è la moda campionaria. Definizione La moda campionaria di un campione di dati, se esiste, è l'unico valore che ha frequenza massima, e si denota con $\hat{x}$. Se non vi è un solo valore con frequenza massima, ciascuno di essi è detto valore modale. La distribuzione dei dati si dice unimodale se la moda è unica; invece si dice bimodale se vi sono due valori modali.

Esercizio:

Frequenza in uscita delle 6 facce di un dado su 40 lanci

Valore $(v_i \equiv i)$	1	2	3	4	5	6
Frequenza (f_i)	9	8	5	5	6	7

$$\overline{x} = \frac{1}{40}\sum_{i=1}^{6} i\, f_i = 3{,}3 \qquad \tilde{x} = \frac{x_{(20)} + x_{(21)}}{2} = 3 \qquad \hat{x} = \operatorname{argmax} f_i = 1$$

Statistiche descrittive: indici di variabilità
Gli indici di variabilità (o di dispersione) sono delle misure che descrivono la variabilità dei dati osservati $x_1, x_2, \dots, x_n$, e che pertanto consentono di valutare l'informazione fornita dall'indice di posizione adoperato, dando così una descrizione più accurata della struttura dei dati. Ad esempio, per entrambe le sequenze di dati (0,5 0,8 2,0 2,7 4,0) e (1,4 1,7 2,0 2,1 2,8)
la media e la mediana sono pari a 2,0 ma tale informazione è più affidabile per la seconda sequenza, dove i dati sono meno dispersi.
Per valutare numericamente la variabilità di dati si adoperano spesso:

- varianza campionaria,
- deviazione standard campionaria,
- scarto medio assoluto,
- ampiezza del campo di variazione,
- coefficiente di variazione.

Varianza e deviazione standard campionarie
Supponiamo di avere un campione di ampiezza o numerosità n, ossia un insieme $x_1, x_2, \dots, x_n$ di n dati.
Una strategia per stabilire quanto i dati siano concentrati o dispersi attorno alla media campionaria consiste nel considerare le distanze dei dati dalla media campionaria, elevarle al quadrato e farne la media aritmetica. Questa è la definizione di varianza campionaria; però per ragioni tecniche si divide per $n - 1$ anziché per n. Definizione Si dice varianza campionaria e si denota con s^2, la quantità:

$$s^2 := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad (n > 1).$$

Esercizio:
Si desidera determinare la varianza campionaria dei due insiemi di dati A : (0,5 0,8 2,0 2,7 4,0) B : (1,4 1,7 2,0 2,1 2,8)

Dalla definizione di varianza campionaria, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$, essendo $\overline{x} = 2$ in entrambi i casi, segue:

$$s_A^2 = \frac{1}{4}\left[(-1{,}5)^2 + (-1{,}2)^2 + 0^2 + (0{,}7)^2 + (2{,}0)^2\right] = 2{,}045$$

$$s_B^2 = \frac{1}{4}\left[(-0{,}6)^2 + (-0{,}3)^2 + 0^2 + (0{,}1)^2 + (0{,}8)^2\right] = 0{,}275$$

Vi è una variabilità maggiore nei valori di A che non in quelli di B anche se gli insiemi di dati hanno la stessa media campionaria.

Il calcolo di s^2 è velocizzato se si tiene conto dell'indentità

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2.$$

Infatti

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n}(x_i^2 - 2\overline{x}x_i + \overline{x}^2)$$

$$= \sum_{i=1}^{n} x_i^2 - 2\overline{x} \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \overline{x}^2$$

$$= \sum_{i=1}^{n} x_i^2 - 2n\overline{x}^2 + n\overline{x}^2 \qquad (\text{poiché } \textstyle\sum_{i=1}^{n} x_i = n\overline{x})$$

$$= \sum_{i=1}^{n} x_i^2 - n\overline{x}^2.$$

Esercizio:

La seguente tabella mostra il PIL regionale del 2012 (in miliardi di euro) con dati ai prezzi di mercato

Lombardia	331,405	Sicilia	84,888	Sardegna	33,025
Lazio	169,483	Puglia	70,314	Abruzzo	30,048
Veneto	146,605	Liguria	44,064	Umbria	21,222
Emilia R.	140,914	Marche	40,192	Basilicata	10,516
Piemonte	124,926	Friuli V.G.	35,996	Molise	6,385
Toscana	105,895	Trentino A.A.	35,405	Valle d'Aosta	4,443
Campania	95,488	Calabria	33,282	Extra	2,514

Risulta

$$\overline{x} = 74{,}620 \qquad \sum_{i=1}^{n} x_i^2 = 237\,827{,}060$$

Pertanto la varianza campionaria è $s^2 = 6\,044{,}88$ essendo

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right) = \frac{1}{20}\left[237\,827{,}060 - 21\,(74{,}620)^2\right]$$

Il calcolo della varianza campionaria può essere anche semplificato se, con a e b costanti, si pone

$$y_i := ax_i + b, \qquad i = 1, 2, \ldots, n$$

e quindi

$$\overline{y} = a\overline{x} + b.$$

In tal caso risulta

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(ax_i + b - a\overline{x} - b)^2 = a^2 \sum_{i=1}^{n}(x_i - \overline{x})^2$$

e pertanto, dividendo per $n-1$, si ha

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 = a^2 \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = a^2 s_x^2.$$

Riassumendo, la trasformazione lineare yi := axi + b fornisce la seguente relazione tra le varianze campionarie: $s_y^2 = a^2 s_x^2$.
Sommare una costante b a ciascuno dei dati non fa cambiare la varianza campionaria.
Moltiplicare ciascuno dei dati per una costante a comporta che la varianza campionaria sia moltiplicata per il quadrato di tale fattore, con diverso effetto a seconda che sia.

$-1 < a < 1$ (la varianza campionaria diminuisce)

$a = \pm 1$ (la varianza campionaria resta invariata)

$|a| > 1$ (la varianza campionaria aumenta)

Esempio Il seguente elenco mostra i tempi di esecuzione x_i di un programma in $n = 10$ diverse prove:

18,2 16,1 16,0 12,5 12,4 9,6 8,9 8,5 8,1 7,8

Risulta $\sum_{i=1}^{10} x_i = 118{,}1$ e $\sum_{i=1}^{10} x_i^2 = 1\,526{,}53$.
Pertanto media campionaria e varianza campionaria sono:

$$\overline{x} = \frac{1}{10} 118{,}1 = 11{,}81 \qquad s^2 = \frac{1}{9}[1\,526{,}53 - 10(11{,}81)^2] = 14{,}64$$

Se si pone $y_i = x_i/10$ per ogni i, si ottiene un nuovo campione $(y_1, \ldots, y_{10})$ avente media e varianza campionaria date da:

$$\overline{y} = \frac{1}{10}\overline{x} = 1{,}181 \qquad s_y^2 = \frac{1}{100}s^2 = 0{,}1464$$

Se sono note le distribuzioni di frequenza dei dati, la varianza campionaria può essere calcolata anche nel seguente modo:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{k} f_i\,(v_i - \overline{x})^2.$$

Esempio Consideriamo le frequenze delle età dei 76 studenti:

Età (v_i)	19	20	21	22	23
Frequenza (f_i)	12	27	23	10	4

Poiché $\overline{x} = 20{,}84$ la varianza campionaria è

$$s^2 = \frac{1}{75}\big[12(19 - 20{,}84)^2 + 27(20 - 20{,}84)^2 + 23(21 - 20{,}84)^2$$

$$+10(22 - 20{,}84)^2 + 4(23 - 20{,}84)^2\big] = \frac{92{,}39}{75} = 1{,}23$$

Esaminiamo un altro indice di variabilità, definito come la radice quadrata della varianza campionaria.
Definizione La deviazione standard campionaria di un insieme di dati $x_1, x_2, \ldots, x_n$ è la quantità:

$$s := \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \qquad (n > 1).$$

Notiamo che la deviazione standard campionaria, così come la varianza campionaria:

- è tanto più grande quanto più i dati si discostano dalla media campionaria x, ovvero quanto più sono eterogenei,
- è uguale a 0 se e solo se i dati sono tutti uguali tra loro.

Scarto medio assoluto e ampiezza del campo di variazione
Esaminiamo due ulteriori indici di variabilità.
Definizione Lo scarto medio assoluto è dato dalla media aritmetica degli scarti assoluti dalla media campionaria:

$$s_a := \frac{1}{n}\sum_{i=1}^{n}|x_i - \overline{x}| = \frac{1}{n}\sum_{i=1}^{k} f_i\,|v_i - \overline{x}|.$$

Definizione Si dice ampiezza del campo di variazione l'indice dato dalla differenza tra i valori estremi delle osservazioni: $\boldsymbol{w := x_{(n)} - x_{(1)}}$, dove $x_{(n)}$ è il massimo e $x_{(1)}$ il minimo dei dati del campione. La deviazione standard campionaria, lo scarto medio assoluto e l'ampiezza del campo di variazione hanno la stessa dimensione dei dati xi e della media campionaria x, a differenza della varianzacampionaria (che ha dimensione pari al quadrato dei dati).

Esempio Prendiamo in esame il seguente campione di $n = 20$ dati:

2,9	2,9	2,9	3,0	3,0	3,0	3,0	3,1	3,1	3,1
3,2	3,2	3,2	3,3	3,3	3,3	3,3	3,4	3,4	3,5

La media campionaria è $\overline{x} = 3{,}155$. Dalla seguente tabella delle frequenze si ricava la varianza campionaria.

i	v_i	f_i	$\lvert v_i - \overline{x}\rvert$	$(v_i - \overline{x})^2$	$f_i\,(v_i - \overline{x})^2$	$f_i\,\lvert v_i - \overline{x}\rvert$
1	2,9	3	0,255	0,065	0,195	0,765
2	3,0	4	0,155	0,024	0,096	0,620
3	3,1	3	0,055	0,003	0,009	0,165
4	3,2	3	0,045	0,002	0,006	0,090
5	3,3	4	0,145	0,021	0,084	0,580
6	3,4	2	0,245	0,060	0,120	0,490
7	3,5	1	0,345	0,119	0,119	0,345

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i\,(v_i - \overline{x})^2 = \frac{1}{19}\sum_{i=1}^{7} f_i\,(v_i - 3{,}155)^2 = \frac{0{,}629}{19} = 0{,}0331.$$

La varianza campionaria si ottiene anche nel seguente modo:

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n(\overline{x})^2\right) = \frac{1}{19}\left[199{,}71 - 20(3{,}155)^2\right] = 0{,}0331.$$

Si ricava immediatamente la deviazione standard campionaria:

$$s = \sqrt{s^2} = \sqrt{0{,}0314} = 0{,}182.$$

Dalla tabella delle frequenze si trae anche lo scarto medio assoluto:

$$s_a = \frac{1}{n}\sum_{i=1}^{k} f_i\,|v_i - \overline{x}| = \frac{1}{20}\sum_{i=1}^{7} f_i\,|v_i - 3{,}155| = \frac{3{,}055}{20} = 0{,}153.$$

Infine, l'ampiezza del campo di variazione è

$$w = x_{(n)} - x_{(1)} = 3{,}5 - 2{,}9 = 0{,}6.$$

Trasformazioni lineari

Abbiamo già visto che talvolta può risultare conveniente effettuare delle trasformazioni dei dati x_1, x_2, . . . , x_n.

Le trasformazioni più semplici sono quelle di tipo lineare, che ricordiamo qui di seguito.

- *Cambiamento di origine*; se $y_i = x_i + b$ $(i = 1, 2, \ldots, n)$ si ha:

$$\overline{y} = \overline{x} + b, \qquad s_y^2 = s_x^2.$$

- *Cambiamento di scala*; se $y_i = ax_i$ $(i = 1, 2, \ldots, n)$ si ha:

$$\overline{y} = a\overline{x}, \qquad s_y^2 = a^2 s_x^2.$$

- *Cambiamento di origine e scala*; se $y_i = ax_i + b$ $(i = 1, 2, \ldots, n)$ si ha:

$$\overline{y} = a\overline{x} + b, \qquad s_y^2 = a^2 s_x^2.$$

Coefficiente di variazione

In certi casi si può usare un ulteriore indice di variabilità che consente di rapportare la variabilità dei dati alla loro media campionaria. Definizione Dato un campione (x_1, . . . , x_n) dove i dati possono essere solo non negativi, si dice coefficiente di variazione l'indice adimensionale di variabilità dato dal rapporto tra la deviazione standard campionaria e la media campionaria (se non nulla): $cv = \frac{s}{\overline{x}}$. Il coefficiente di variazione è un indice diretto della variabilità dei dati, indipendentemente dalla scala di misura che si adopera. Infatti, il cambiamento di scala $y_i = ax_i$ per ogni i , con a > 0, produce:

$$cv_y = \frac{s_y}{\overline{y}} = \frac{\sqrt{s_y^2}}{a\overline{x}} = \frac{\sqrt{a^2 s_x^2}}{a\overline{x}} = \frac{s_x}{\overline{x}} = cv_x.$$

Esercizio:
la deviazione standard di un campione di redditi in Lire è in genere diversa della deviazione standard degli stessi redditi in Euro, mentre il coefficiente di variazione è lo stesso nei due casi (essendo la conversione Lire-Euro effetto di cambiamento di scala). Invece, se effettua un cambiamento di origine e scala yi = axi + b per ogni i , con a > 0 e b 6 ≠ 0, risulta:

$$cv_y = \frac{s_y}{\overline{y}} = \frac{\sqrt{s_y^2}}{a\overline{x}+b} = \frac{\sqrt{a^2 s_x^2}}{a\overline{x}+b} = \frac{a s_x}{a\overline{x}+b} \neq cv_x .$$

Quindi il coefficiente di variazione è un indice non significativo per dati che provengono da scale definite per intervalli, ovvero per scale in cui il valore "zero" discende da convenzioni arbitrarie, come per la misurazione della temperatura Celsius e Fahrenheit.

Esempio Confrontiamo la variabilità in 2 coppie di sequenze di dati:

	dati	$\overline{x}$	s	cv
stessa	A_1 : 5 6 4 5 3 5	4,67	1,03	0,221
scala	B_1 : 5 1 3 5 1 2	2,83	1,83	0,648
scala	A_2 : 0,5 0,8 1,1 1,5 1,2 0,9	1,00	0,35	0,346
diversa	B_2 : 500 520 515 520 523 508	514,33	8,78	0,017

La deviazione standard della sequenza A_1 è minore della deviazione standard della sequenza B_1; inoltre anche il coefficiente di variazione di A_1 è minore di quello di B_1. Ciò mostra che la prima sequenza è caratterizzata da una minore variabilità.

La deviazione standard della sequenza A_2 è minore di quella di B_2, ma per il coefficiente di variazione accade il contrario. Ciò segue dal fatto che i dati hanno diverse scale di misura: anche se s è minore per i dati di A_2, il confronto dei coefficienti di variazione mostra che la variabilità rispetto alla media è maggiore per i dati di A_2.

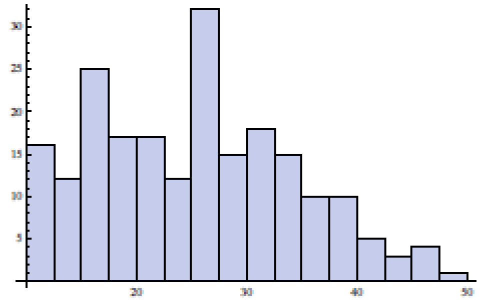
Indici di forma
Gli indici di forma misurano caratteristiche relative alla forma della distribuzione dei dati. I più usati sono l'indice di asimmetria e l'indice di curtosi, entrambi adimensionali.
Definizione Dato un campione ($x_1, \ldots, x_n$) avente media campionaria x e deviazione standard campionaria s, la seguente quantità è detta indice di asimmetria (o di skewness):

$$\gamma := \frac{1}{s^3}\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^3 .$$

Si usa per stabilire se i dati sono caratterizzati da simmetria centrale:

- $\gamma > 0$: la distribuzione di dati presenta una coda più lunga a destra (asimmetria positiva),
- $\gamma < 0$: la distribuzione presenta una coda più lunga a sinistra (asimmetria negativa).

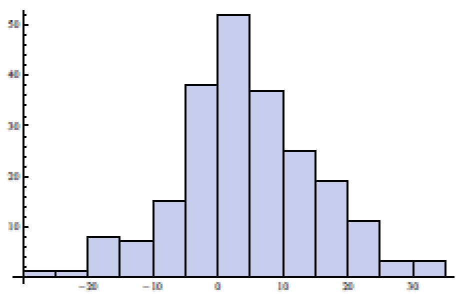
Esempio Il seguente istogramma suggerisce una asimmetria positiva nella distribuzione dei dati di un campione.



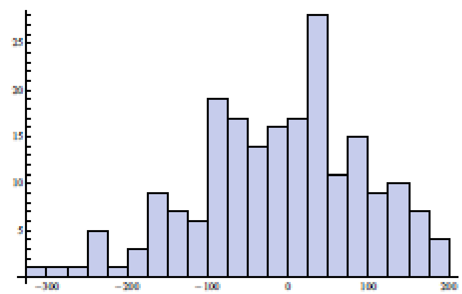
Risulta $\qquad \overline{x} = 25{,}13 \qquad s^2 = 82{,}14 \qquad s = 9{,}06$

L'indice di asimmetria è positivo in quanto la distribuzione di dati presenta una coda più lunga a destra:

$$\gamma = \frac{1}{s^3}\,\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3 = 0{,}319.$$



$\gamma = 0{,}006$



$\gamma = -0{,}388$

Definizione Dato un campione $(x_1, \ldots, x_n)$ avente media campionaria $\overline{x}$ e deviazione standard campionaria s, la seguente quantità è detta *indice di curtosi (o di appiattimento)*:

$$k := \left[\frac{1}{s^4}\,\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4\right] - 3$$

utile per stabilire se la distribuzione è poco o molto appiattita.

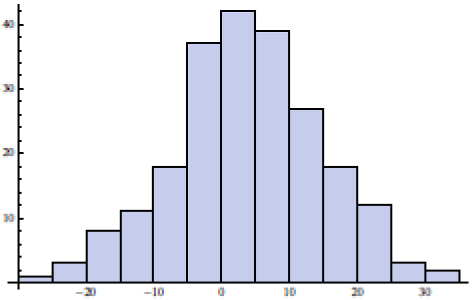
$k > 0$: è presente un eccesso di dati nelle classi centrali (i dati hanno distribuzione *leptocurtica*)

$k < 0$: è presente una carenza di dati nelle classi centrali (i dati hanno distribuzione *platicurtica*)

$k = 0$: è presente una distribuzione di dati come quella normale (i dati hanno distribuzione *normocurtica*)

In alcuni casi k si definisce come $\frac{1}{s^4}\,\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4$.

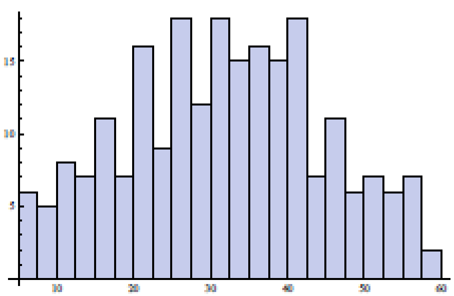
Esempio Consideriamo il seguente istogramma.

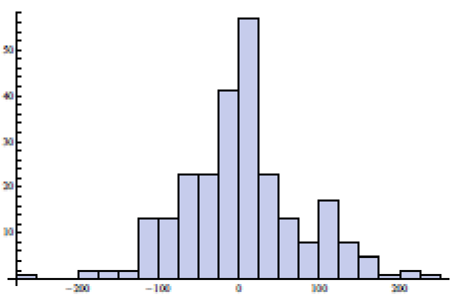


Risulta $\quad \overline{x} = 3{,}36 \qquad s^2 = 117{,}82 \qquad s = 10{,}85$

La distribuzione è approssimativamente normocurtica, essendo l'indice di curtosi prossimo a 0:

$$k = \left[\frac{1}{s^4}\,\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4\right] - 3 = 0{,}004.$$





$k = -0{,}738$

distribuzione
platicurtica

$k = 0{,}788$

distribuzione
leptocurtica

Percentili campionari e box plot
Il percentile k-esimo di un campione di dati è un valore che è maggiore di una percentuale k dei dati, e minore della restante percentuale 100 − k, dove k è un numero tra 0 e 100.
Definizione Sia k un numero intero con $0 \leq k \leq 100$. Dato un insieme di dati numerici, ne esiste sempre uno che è contemporaneamente maggiore o uguale di almeno il k per cento dei dati, e minore o uguale di almeno il 100 − k per cento dei dati. Se il dato con queste caratteristiche è unico, allora esso è per definizione il percentile k-esimo dell'insieme di dati considerato. Se invece non è unico, allora sono esattamente due, e in tal caso il percentile k-esimo è definito come la loro media aritmetica. Per determinare il percentile k-esimo di un campione di numerosità n occorre trovare un dato (o due dati) tali che, posto $p = k/100$, almeno np tra tutti i dati dell'insieme siano minori o uguali a loro; almeno $n(1 - p)$ tra tutti i dati dell'insieme siano maggiori o uguali a loro. Per prima cosa occorre disporre i dati in ordine crescente. Se np e $n(1 - p)$ non sono numeri interi occorre arrotondarli all'intero successivo.

Ad esempio, se $n = 22$, $k = 80$ (e quindi $p = 0{,}8$) allora risulta

$$np = 17{,}6 \qquad n(1-p) = 4{,}4$$

ed andranno arrotondati agli interi 18 e 5, rispettivamente.

Quindi, se abbiamo 22 dati ordinati allora il percentile 80-esimo è il numero tale che almeno 18 dati sono minori o uguali ad esso, ed almeno 5 dati sono maggiori o uguali ad esso. Ad esempio:

$$\underbrace{3,\ 5,\ 6,\ 6,\ 6,\ 7,\ 7,\ 7,\ 8,\ 9,\ 9,\ 9,\ 9,\ 9,\ 10,\ 15,\ 16,\ \boxed{16}}_{18\text{ dati}},\ 20,\ 22,\ 24,\ 24$$

In questo caso il percentile 80-esimo è 16, poiché occupa il posto 18-esimo nei dati disposti in ordine crescente.

Per esempio, se $n = 16$, $k = 25$ (e quindi $p = 0{,}25$) allora risulta

$$np = 4 \qquad n(1-p) = 12.$$

Quindi, se abbiamo 16 dati ordinati allora il percentile 25-esimo è la media aritmetica di due numeri, ossia il quarto e il quinto della sequenza ordinata in senso crescente. Ad esempio:

$$\underbrace{0,\ 2,\ 3,\ \boxed{3}}_{4\text{ dati}},\ \boxed{4},\ 5,\ 5,\ 5,\ 6,\ 6,\ 7,\ 8,\ 9,\ 10,\ 11,\ 11$$

In questo caso il percentile 25-esimo è

$$\frac{x_{(4)} + x_{(5)}}{2} = \frac{3+4}{2} = 3{,}5.$$

Il 50-esimo percentile coincide ovviamente con la mediana campionaria.
Definizione Il 25-esimo percentile si dice primo quartile. Il 50-esimo percentile si dice mediana campionaria o secondo quartile. Il 75-esimo percentile si dice terzo quartile. I quartili si denotano con Q_1, Q_2 e Q_3. Essi dividono il campione in quattro parti, ognuna contenente circa il 25% dei dati:

- i dati minori del primo quartile,
- i dati compresi tra il primo quartile e il secondo quartile,
- i dati compresi tra il secondo quartile e il terzo quartile,
- i dati maggiori del terzo quartile.

Esempio Determinare i quartili delle seguenti temperature, registrate dall'ARPAC a Vietri sul Mare (Marina di Vietri, secondo tratto)

Data	Ora	T. aria	T. acqua
2015-4-14	10:35	16	14,5
2015-5-12	09:50	24,5	17,6
2015-5-25	11:50	19	20,6
2015-6-09	11:45	27	24
2015-6-22	12:25	27,5	21,2
2015-7-07	13:05	29,5	27,3
2015-7-20	14:00	31,3	28,8
2015-7-28	11:30	28	27
2015-8-04	13:22	33	28,1
2015-8-24	11:55	29	27
2015-9-02	13:30	30	28,4
2015-9-14	11:42	29	25,1

Dati ordinati dei 2 campioni (di dimensione $n = 12$)
$x_{(i)}$ (T. aria):
16 19 **24,5** **27** 27,5 28 29 29 29,5 30 31,3 33
$y_{(i)}$ (T. acqua):
14,5 17,6 **20,6** **21,2** 24 25,1 27 27 27,3 28,1 28,4 28,8

1° quartile: $n = 12, \quad p = 0{,}25 \quad \Rightarrow \quad np = 3, \quad n(1-p) = 9$

occorre trovare un dato (o quei dati) tali che,
– almeno 3 tra tutti i dati dell'insieme siano minori o uguali a loro;
– almeno 9 tra tutti i dati dell'insieme siano maggiori o uguali a loro.

Quindi il quartile Q_1 dei 2 campioni è, rispettivamente:

$$\frac{x_{(3)} + x_{(4)}}{2} = \frac{24{,}5 + 27}{2} = 25{,}75 \qquad \frac{y_{(3)} + y_{(4)}}{2} = \frac{20{,}6 + 21{,}2}{2} = 20{,}9$$

Dati ordinati dei 2 campioni (di dimensione $n = 12$)

$x_{(i)}$ (T. aria):

16 19 24,5 27 27,5 **28** **29** 29 29,5 30 31,3 33

$y_{(i)}$ (T. acqua):

14,5 17,6 20,6 21,2 24 **25,1** **27** 27 27,3 28,1 28,4 28,8

2° quartile (mediana): $n = 12$, $p = 0{,}5 \Rightarrow np = 6$, $n(1-p) = 6$

occorre trovare un dato (o quei dati) tali che,
– almeno 6 tra tutti i dati dell'insieme siano minori o uguali a loro;
– almeno 6 tra tutti i dati dell'insieme siano maggiori o uguali a loro.

Quindi il quartile Q_2 dei 2 campioni è, rispettivamente:

$$\frac{x_{(6)} + x_{(7)}}{2} = \frac{28 + 29}{2} = 28{,}5 \qquad \frac{y_{(6)} + y_{(7)}}{2} = \frac{25{,}1 + 27}{2} = 26{,}05$$

Dati ordinati dei 2 campioni (di dimensione $n = 12$)

$x_{(i)}$ (T. aria):

16 19 24,5 27 27,5 28 29 29 **29,5** **30** 31,3 33

$y_{(i)}$ (T. acqua):

14,5 17,6 20,6 21,2 24 25,1 27 27 **27,3** **28,1** 28,4 28,8

3° quartile: $n = 12$, $p = 0{,}75 \Rightarrow np = 9$, $n(1-p) = 3$

occorre trovare un dato (o quei dati) tali che,
– almeno 9 tra tutti i dati dell'insieme siano minori o uguali a loro;
– almeno 3 tra tutti i dati dell'insieme siano maggiori o uguali a loro.

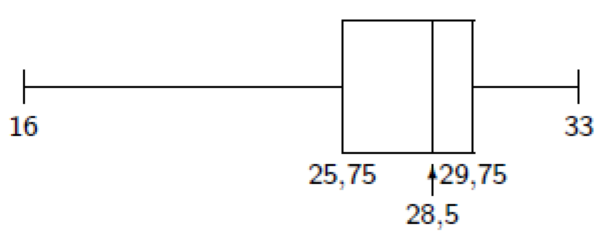
Quindi il quartile Q_3 dei 2 campioni è, rispettivamente:

$$\frac{x_{(9)} + x_{(10)}}{2} = \frac{29{,}5 + 30}{2} = 29{,}75 \qquad \frac{y_{(9)} + y_{(10)}}{2} = \frac{27{,}3 + 28{,}1}{2} = 27{,}7$$

Uno strumento grafico utile a visualizzare alcuni degli indici rappresentativi dei dati è il box plot. Si ottiene sovrapponendo ad una linea orizzontale che va dal minore al maggiore dei dati, un rettangolo che va dal primo al terzo quartile, con una linea verticale che lo divide al livello del secondo quartile.

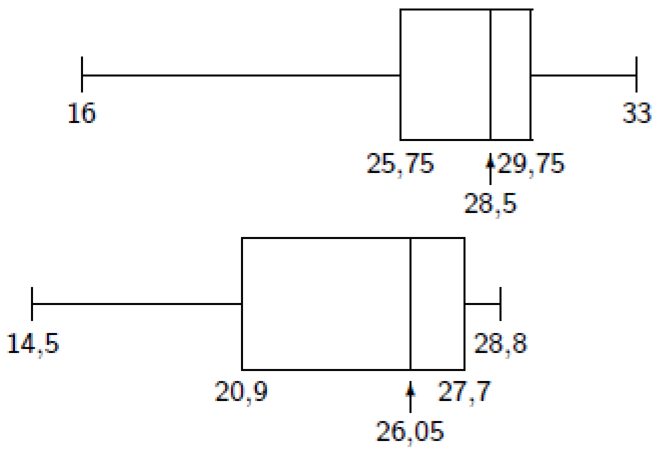
Esempio $x_{(i)}$ (T. aria):

16 19 24,5 27 27,5 28 29 29 29,5 30 31,3 33

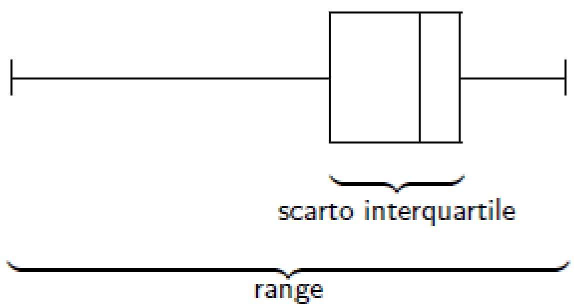


Nel caso di più campioni è utile affiancare i box plot rispettivi.

Esempio
$x_{(i)}$: 16 19 24,5 27 27,5 28 29 29 29,5 30 31,3 33
$y_{(i)}$: 14,5 17,6 20,6 21,2 24 25,1 27 27 27,3 28,1 28,4 28,8

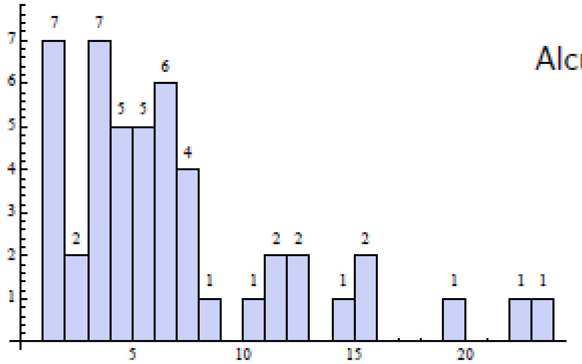


La lunghezza della linea orizzontale di un box plot è pari alla distanza tra il minimo e il massimo dei dati, $x_{(n)} - x_{(1)}$, e quindi corrisponde all'ampiezza del campo di variazione (o range). La lunghezza del solo rettangolo, pari alla distanza tra il primo e il terzo quartile, $Q_3 - Q_1$, è detta scarto interquartile.



Si dice *outlier* un dato anomalo, ossia molto distante dagli altri dati. Un outlier può essere dovuto alla variabilità nella misurazione o ad un errore sperimentale (e in tal caso potrebbe essere rimosso). Un outlier può essere rilevato in quanto tratto da una distribuzione con elevata curtosi (e in tal caso va usata cautela nell'ipotizzare che I dati provengano da una distribuzione regolare, ad esempio normale). Talvolta si definisce outlier ogni dato che, per qualche $k \geq 0$, si trova all'esterno dell'intervallo $[Q_1 - k (Q_3 - Q_1), Q_3 + k (Q_3 - Q_1)]$. In tal caso la linea orizzontale del box plot si restringe solo a tale intervallo, e gli outlier vengono rappresentati all'esterno.

Esempio Consideriamo il seguente campione di $n = 48$ dati, riferito agli intertempi (in anni) tra le eruzioni del Vesuvio dal 1631 al 1944: 6, 12, 11, 22, 3, 4, 5, 2, 1, 1, 3, 6, 7, 3, 6, 1, 6, 7, 14, 3, 6, 7, 3, 1, 8, 15, 10, 1, 1, 4, 2, 1, 4, 5, 12, 5, 11, 5, 6, 7, 4, 19, 4, 5, 3, 3, 23, 15.



Alcuni indici statistici:

$\overline{x} = 6{,}52$
$s^2 = 29{,}11$
$s = 5{,}39$
$\gamma = 1{,}44$
$k = 1{,}62$

Per ricavare i quartili esaminiamo le frequenze assolute:

v_i	1	2	3	4	5	6	7	8	10	11	12	14	15	19	22	23
f_i	7	2	7	5	5	6	4	1	1	2	2	1	2	1	1	1
F_i	7	9	16	21	26	32	36	37	38	40	42	43	45	46	47	48

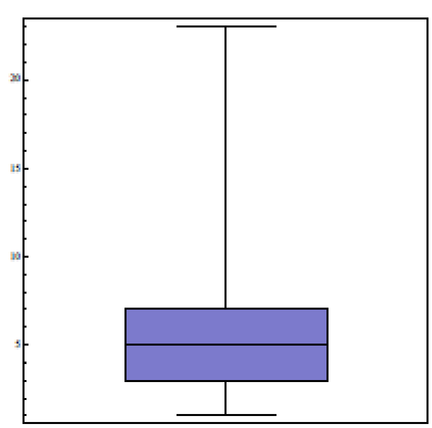
1) $n = 48, \;\; p = 0{,}25 \;\Rightarrow\;\; np = 12, \;\; n(1-p) = 36$
$Q_1 = \frac{x_{(12)}+x_{(13)}}{2} = \frac{3+3}{2} = 3$

2) $n = 48, \;\; p = 0{,}5 \;\Rightarrow\;\; np = 24, \;\; n(1-p) = 24$
$Q_2 = \frac{x_{(24)}+x_{(25)}}{2} = \frac{5+5}{2} = 5 \qquad (< \overline{x} = 6{,}52\ )$

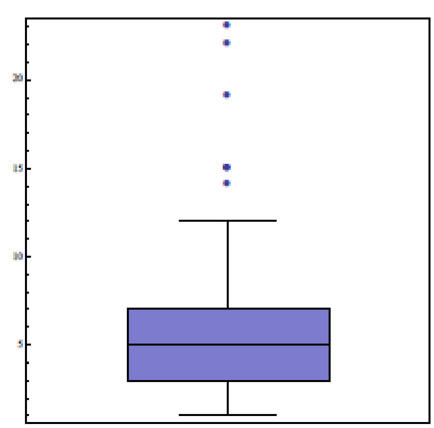
3) $n = 48, \;\; p = 0{,}75 \;\Rightarrow\;\; np = 36, \;\; n(1-p) = 12$
$Q_3 = \frac{x_{(36)}+x_{(37)}}{2} = \frac{7+8}{2} = 7{,}5$

Per $k = 1$, riteniamo outlier i 6 dati che sono all'esterno di

$$[Q_1 - k\,(Q_3 - Q_1)\,,\;\; Q_3 + k\,(Q_3 - Q_1)] = [-1{,}5\,, 12].$$



Box plot



Box plot con outliers

Disuguaglianza di Chebyshev
Spesso risulta utile stabilire quanti dati di un campione sono prossimi alla media campionaria. La disuguaglianza di Chebyshev fornisce una limitazione inferiore per tale numero quando l'ampiezza n del campione è nota. **Proposizione** (Disuguaglianza di Chebyshev) Dato un campione di dati $x_1, x_2, \dots, x_n$, con media campionaria $\overline{x}$ e deviazione standard campionaria $s > 0$, sia

$$S_k := \{i,\ 1 \le i \le n : |x_i - \overline{x}| < ks\}$$

l'insieme degli indici i corrispondenti ai dati x_i compresi tra $\overline{x} - ks$ e $\overline{x} + ks$. Allora, per ogni $k \ge 1$, risulta

$$\frac{|S_k|}{n} > 1 - \frac{1}{k^2},$$

dove $|S_k|$ denota il numero di elementi, o *cardinalità*, di S_k.

La disuguaglianza $\frac{|S_k|}{n} > 1 - \frac{1}{k^2}$ mostra che la percentuale di dati x_i compresi in $(\overline{x} - ks\,,\ \overline{x} + ks)$ è almeno pari a $1 - \frac{1}{k^2}$.

Si ha:

$$k = 1{,}5 \quad \Rightarrow \quad \frac{|S_{1,5}|}{n} > 1 - \frac{4}{9} = 0{,}5556$$

$$k = 2 \quad \Rightarrow \quad \frac{|S_2|}{n} > 1 - \frac{1}{4} = 0{,}75$$

$$k = 3 \quad \Rightarrow \quad \frac{|S_3|}{n} > 1 - \frac{1}{9} = 0{,}8889$$

Quindi, per un campione qualsiasi la percentuale di dati appartenenti
– all'intervallo $(\overline{x} - 1{,}5s\,,\ \overline{x} + 1{,}5s)$ è almeno pari al 55,56%
– all'intervallo $(\overline{x} - 2s\,,\ \overline{x} + 2s)$ è almeno pari al 75%
– all'intervallo $(\overline{x} - 3s\,,\ \overline{x} + 3s)$ è almeno pari all'88,89%

Esempio Marche di auto nuove più vendute in Italia (agosto 2015)

Fiat	11581	Lancia	2521	Hyundai	1770
Volkswagen	5025	Toyota/Lexus	2344	Kia	1746
Ford	4631	Citroen/Ds	2260	Mercedes	1609
Opel	3897	Audi	2177	Dacia	1342
Peugeot	3178	Nissan	2120	Alfa Romeo	1262
Renault	3124	Bmw	1834	Chrysler/Jeep/Dodge	1231

Dal campione si ricava: $\overline{x} = 2981 \qquad s = 2415$

Dalla disuguaglianza di Chebyshev segue che l'intervallo

$$(\overline{x} - 1{,}5s\,,\ \overline{x} + 1{,}5s) = (-642\,,\ 6604)$$

deve contenere almeno il 55,56% dei dati, mentre in realtà ne contiene 17 su 18, ossia il 94,44%.

Da quanto visto finora segue che possiamo usare vari tipi di intervalli come rappresentativi di un campione di dati.

facendo uso del primo e terzo quartile:

$$[Q_1 - k\,(Q_3 - Q_1)\,,\ Q_3 + k\,(Q_3 - Q_1)]$$

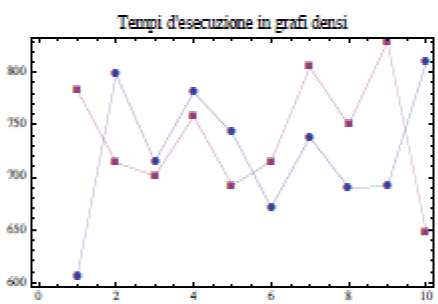
facendo uso della media campionaria e della deviazione standard campionaria:

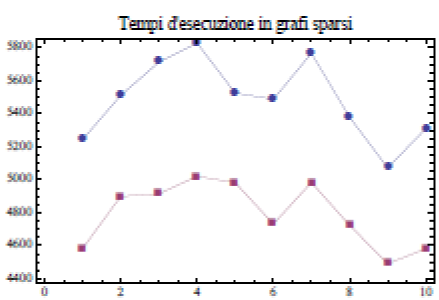
$$[\overline{x} - ks,\ \overline{x} + ks]$$

Esempio La tabella riporta i tempi di esecuzioni in 10 istanze di un algoritmo su grafi basato su 2 diverse discipline di gestione dei nodi: Queue e Deque, con implementazione su grafi densi e grafi sparsi.

ist.	Queue (g. densi)	Deque (g. densi)	Queue (g. sparsi)	Deque (g. sparsi)
1	606	783	5245	4582
2	798	714	5513	4899
3	715	701	5714	4915
4	781	758	5830	5021
5	742	692	5527	4983
6	671	715	5489	4731
7	738	805	5768	4984
8	690	750	5380	4723
9	691	829	5073	4490
10	809	648	5312	4580

Grafici a linee (blu: Queue ● / rosso: Deque ■)





statistiche su grafi densi		statistiche su grafi sparsi	
Queue	Deque	Queue	Deque
$Q_1 = 690$	$Q_1 = 701$	$Q_1 = 5312$	$Q_1 = 4582$
$Q_3 = 781$	$Q_3 = 783$	$Q_3 = 5714$	$Q_3 = 4983$
$\overline{x} = 724{,}1$	$\overline{x} = 739{,}5$	$\overline{x} = 5485{,}1$	$\overline{x} = 4790{,}8$
$s = 62{,}83$	$s = 55{,}78$	$s = 240{,}93$	$s = 194{,}67$

Individuiamo degli intervalli rappresentativi dei campioni considerati:

$$I_1 = [Q_1 - k(Q_3 - Q_1),\ Q_3 + k(Q_3 - Q_1)]$$

$$I_2 = [\overline{x} - ks,\ \overline{x} + ks]$$

	statistiche	su grafi densi	statistiche	su grafi sparsi
k	Queue	Deque	Queue	Deque
1	$I_1 = [599, 781]$	$I_1 = [619, 783]$	$I_1 = [4910, 5714]$	$I_1 = [4181, 4983]$
2	$I_1 = [508, 872]$	$I_1 = [537, 865]$	$I_1 = [4508, 6116]$	$I_1 = [3780, 5384]$
1	$I_2 = [661, 787]$	$I_2 = [684, 795]$	$I_2 = [5244, 5726]$	$I_2 = [4596, 4985]$
2	$I_2 = [598, 850]$	$I_2 = [628, 851]$	$I_2 = [5003, 5967]$	$I_2 = [4401, 5180]$

Si trae che gli intervalli rappresentativi della disciplina Queue sono sensibilmente maggiori di quelli di Deque nel caso di grafi sparsi, ma nel caso di grafi densi non si riscontra una differenza marcata.

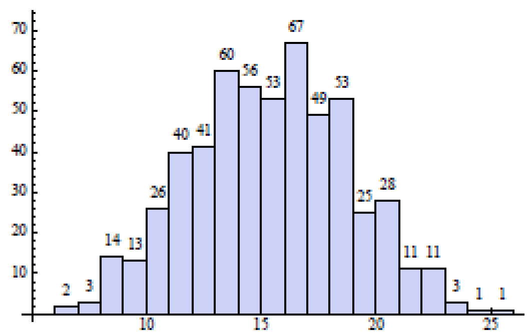
Campi normali

Gli istogrammi di campioni numerici forniti da esperimenti reali spesso hanno una forma caratteristica, con il grafico che presenta un solo massimo in corrispondenza della mediana, e che decresce da entrambi i lati simmetricamente, secondo una curva a campana. Un campione di dati che rispetta questi requisiti si dice normale. Un campione di dati con un istogramma simile si dice approssimativamente normale, e nel caso in cui sia fortemente sbilanciato si dice skewed (ossia asimmetrico) a sinistra o a destra a seconda del lato dove ha la coda più lunga.

Per un campione normale la percentuale di dati appartenenti

- all'intervallo ($\overline{x} - s$, $\overline{x} + s$) è circa il 68%
- all'intervallo ($\overline{x} - 2s$, $\overline{x} + 2s$) è circa il 95%
- all'intervallo ($\overline{x} - 3s$, $\overline{x} + 3s$) è circa il 99,7%

Votazioni di $n = 557$ studenti ad un test d'accesso universitario



Risulta: $\overline{x} = 14{,}9$ $\quad s = 3{,}4$

l'intervallo $(\overline{x} - s\ ,\ \overline{x} + s) = (11{,}5\ ,\ 18{,}3)$ contiene il 68% dei dati

l'intervallo $(\overline{x} - 2s\ ,\ \overline{x} + 2s) = (8{,}1\ ,\ 21{,}7)$ contiene il 93,7% dei dati

l'intervallo $(\overline{x} - 3s\ ,\ \overline{x} + 3s) = (4{,}7\ ,\ 25{,}1)$ contiene il 100% dei dati

quasi in accordo con le percentuali di un campione normale (68%, 95%, 99,7%) .

Dati bivariati e coefficiente di correlazione campionario

Talvolta non occorre trattare dati singoli, ma coppie di numeri (x_i, y_i), $i = 1, 2, \ldots, n$, tra i quali esiste qualche relazione.

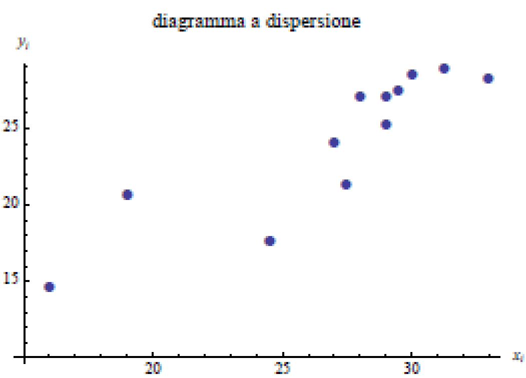
Dati di questa forma prendono il nome di campione *bivariato*.

Ad esempio: temperature registrate dall'ARPAC a Vietri sul Mare

Data	Ora	T. aria (x_i)	T. acqua (y_i)
2015-4-14	10:35	16	14,5
2015-5-12	09:50	24,5	17,6
2015-5-25	11:50	19	20,6
2015-6-09	11:45	27	24
2015-6-22	12:25	27,5	21,2
2015-7-07	13:05	29,5	27,3
2015-7-20	14:00	31,3	28,8
2015-7-28	11:30	28	27
2015-8-04	13:22	33	28,1
2015-8-24	11:55	29	27
2015-9-02	13:30	30	28,4
2015-9-14	11:42	29	25,1

Uno strumento utile a visualizzare campioni bivariati è il diagramma a dispersione, che consiste nella rappresentazioni sul piano cartesiano di tanti punti per quante sono le osservazioni (x_i , y_i), ognuno tracciato nelle corrispondenti coordinate.

Diagramma a dispersione dell'esempio considerato, con x_i = temperatura dell'aria e y_i = temperatura dell'acqua.



Una questione di grande interesse è stabilire se in un campione bivariato vi sia una correlazione (ossia una forma di dipendenza) tra i valori x e y. Si parla di correlazione positiva quando a valori elevati di x corrispondono valori elevati di y, e a valori bassi di x corrispondono valori bassi di y. Viceversa, si parla di correlazione negativa quando a valori elevati di x corrispondono valori bassi di y, e a valori bassi di x corrispondono valori elevati di y. Una risposta grossolana alla questione della correlazione di può ottenere osservando il diagramma a dispersione.

Per ottenere una misura quantitativa della correlazione costruiamo un nuovo indice statistico.
Dato un campione bivariato (x_i, y_i), $i = 1, 2, \ldots, n$, siano $\overline{x}$ e $\overline{y}$ le medie campionarie relative ai valori x e y, rispettivamente.
Se un valore x_i è grande rispetto ai valori tipici di x, allora la differenza $x_i - \overline{x}$ sarà positiva, mentre se x_i è piccolo, essa sarà negativa. Possiamo ragionare analogamente per i valori y_i.
Pertanto, il prodotto $(x_i - \overline{x})(y_i - \overline{y})$ sarà positivo per le osservazioni in cui x_i e y_i sono correlate positivamente, mentre sarà negativo per quelle in cui vi è correlazione negativa.
Quindi ci si attende che la somma $\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ sia positiva (o negativa) se l'intero campione di dati bivariati mostra una correlazione positiva (o negativa).

Definizione Sia dato un campione bivariato (x_i, y_i), $i = 1, 2, \ldots, n$, con medie campionarie $\overline{x}$ e $\overline{y}$, e con deviazioni standard campionarie s_x e s_y, per i soli dati x e y, rispettivamente. Allora si dice *coefficiente di correlazione campionario* la quantità

$$r := \frac{1}{(n-1)s_x s_y}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$
$$= \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

Quando $r > 0$ i dati si dicono *correlati positivamente*, mentre se $r < 0$ i dati si dicono *correlati negativamente*.
Notiamo che r è una quantità adimensionale.

Alcune proprietà del coefficiente di correlazione campionario:
Risulta sempre $-1 \leq r \leq 1$.
Se per opportune costanti a e b sussiste la relazione lineare

$$y_i = ax_i + b, \qquad \forall i = 1, 2, \ldots, n$$

con $a > 0$, allora $r = 1$.
Se per opportune costanti a e b sussiste la relazione lineare

$$y_i = ax_i + b, \qquad \forall i = 1, 2, \ldots, n$$

con $a < 0$, allora $r = -1$.
Se r è il coefficiente di correlazione campionario del campione (x_i, y_i), $i = 1, 2, \ldots, n$, allora lo è anche per il campione

$$(ax_i + b, cy_i + d), \qquad \forall i = 1, 2, \ldots, n$$

purché le costanti a e c abbiano lo stesso segno.

Il valore assoluto di r è una misura della forza della correlazione esistente tra i valori x_i e y_i.

Quando $|r| = 1$ vi è una relazione lineare perfetta, e i punti del diagramma di dispersione stanno tutti su una retta (con pendenza positiva o negativa a seconda che sia $r = 1$ o $r = -1$).

Quando $|r| = 0{,}8$ vi è una correlazione molto intensa, e i punti del diagramma di dispersione sono vicini ad una retta *interpolante* (avente pendenza positiva o negativa a seconda che sia $r = 0{,}8$ o $r = -0{,}8$).

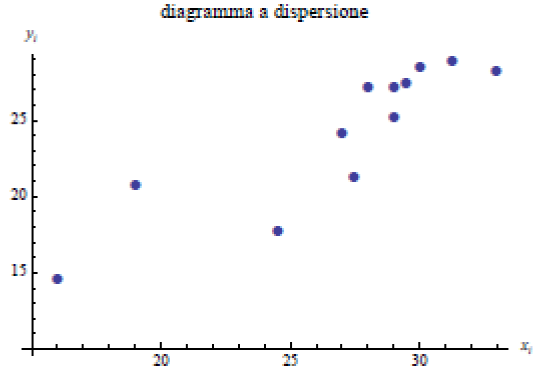
Valori di $|r|$ intorno a 0,3 indicano una correlazione molto debole.

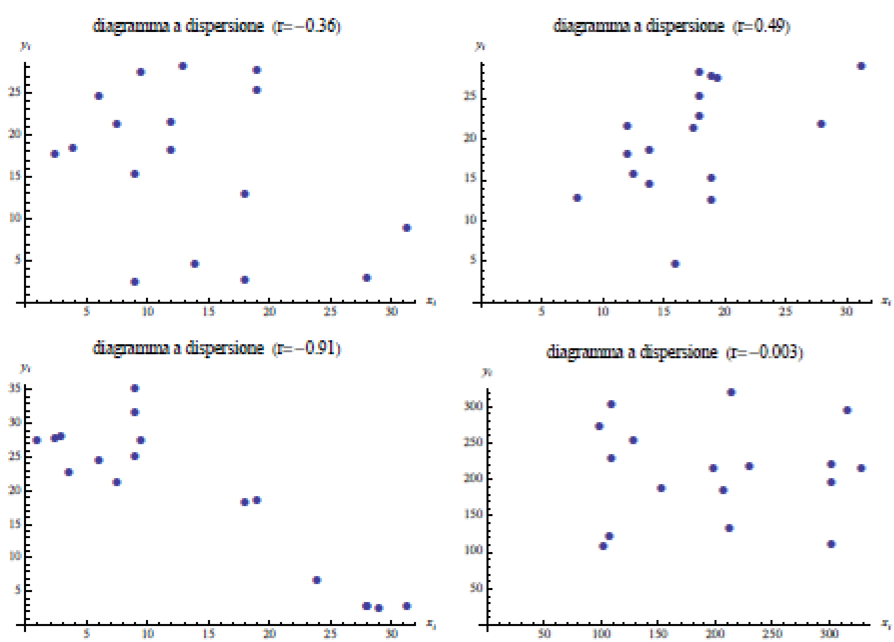
Ad esempio, per i dati esaminati in precedenza risulta

$$r = 0{,}88$$

x_i = temperatura dell'aria $\qquad$ y_i = temperatura dell'acqua

diagramma a dispersione





Talvolta lo studio della correlazione tra due caratteristiche evidenzia una forte correlazione. Ciò però non comporta necessariamente l'esistenza di un rapporto causa-effetto tra di esse. Se x e y sono grandezze correlate, l'analisi della correlazione non consente di stabilire se la causa di tale correlazione sia intrinseca nella natura delle variabili o se sia conseguenza dell'effetto di un'altra variabile che le influenza entrambe.
Supponiamo di aver riscontrato correlazione positiva tra il numero di visite mediche effettuate in una città e lo status socioeconomico dei residenti. Possono individuarsi tre spiegazioni plausibili:

1. la presenza di un numero elevato di visite mediche fa sì che gli individui siano maggiormente sani, e quindi più efficienti nel lavoro, risultandone quindi innalzati i guadagni e lo status socioeconomico;
2. gli individui caratterizzati da status elevato hanno possibilità di sottoporsi a frequenti visite mediche;
3. un'ulteriore variabile condiziona indipendentemente il numero di visite mediche e lo status socioeconomico: ad esempio città di grandi dimensioni presumibilmente offrono un numero elevato di centri medici e opportunità lavorative meglio remunerate.