UNIVERSITÀ DEGLI STUDI DI TRIESTE
Dipartimento di Ingegneria e Architettura

Corso di misure meccaniche, termiche e collaudi

Prof. Rodolfo Taccani
Prof. Lucia Parussini
Prof. Marco Bogar

a.a.2023-2024

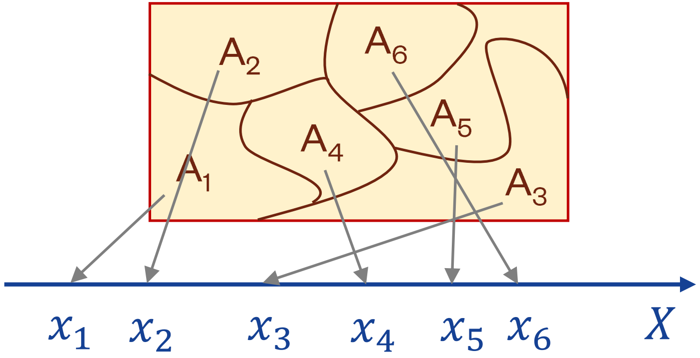
Outline

- Variabili casuali, funzione di densità di probabilità e distribuzione di probabilità
- Valore atteso e varianza
- Distribuzioni di probabilità

Variabili casuali, funzione di densità e distribuzione di probabilità

Variabili casuali

In generale, se consideriamo un esperimento il cui insieme di possibili risultati è E, diremo variabile casuale una applicazione X a valori reali definita sull'insieme E che associa ad ogni elemento di E un numero reale.



Variabili casuali, funzione di densità e distribuzione di probabilità

Variabili casuali

Sono modelli teorici utili a descrivere i fenomeni aleatori.

Sono sempre specificate da due entità:
- l'insieme dei valori assunti dalla variabile
- le probabilità associate a ciascun valore (o le densità di probabilità associata ad un intervallo di valori)

Gli eventi specificati dai valori assunti dalle variabili sono sempre incompatibili, e formano una classe completa (spazio degli eventi).

La somma delle probabilità (o l'integrale della funzione di densità di probabilità esteso a tutto il campo di esistenza della variabile) vale uno.

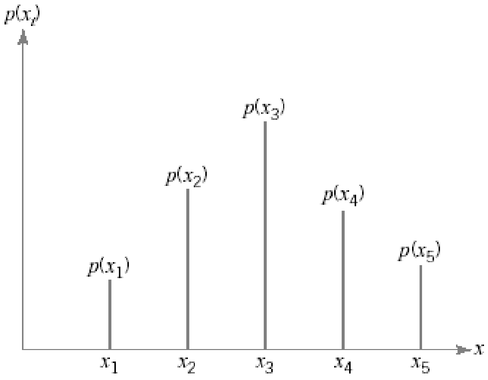
Una variabile casuale X può essere discreta (cioè assumere un insieme di valori discreti, x_i, con $i = 1,2, \dots, j, \dots$) oppure continua (cioè assumere un insieme continuo di valori).

Variabili casuali, funzione di densità e distribuzione di probabilità

Funzione di densità di probabilità di una variabile casuale discreta

Supponiamo che X sia una variabile casuale discreta, cioè che X possa assumere valori appartenenti all'insieme di numeri reali $\{x_1, x_2, \dots, x_j, ..\}$. Ad ognuno dei valori che X può assumere, possiamo assegnare una funzione di probabilità $p(x_i)$ tale che sia $p(x_i) \geq 0$ per ogni $i = 1,2, \dots, j, \dots$ e tale che sia

$$\sum_i p(x_i) = 1$$



Data una variabile casuale discreta $X: \Omega \to S \subseteq \mathbb{R}$ la *funzione di densità di probabilità* è la funzione $p: S \to [0,1]$

$$p(x) = \begin{cases} P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\}) & x \in S \\ 0 & x \in \mathbb{R} \backslash S \end{cases}$$

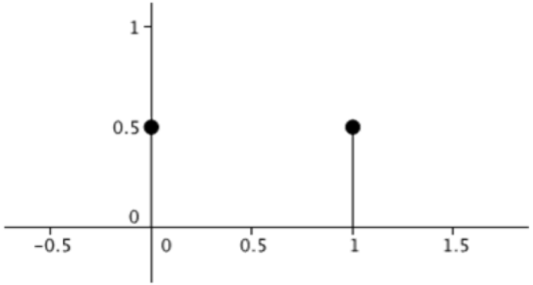
che associa ad ogni valore di x assunto dalla variabile casuale X la probabilità che la variabile X assuma esattamente quel valore.

Funzione di densità di probabilità di una variabile casuale discreta

Esempio:

Nel caso dell'esperimento casuale di lancio di una moneta bilanciata con spazio campionario Ω costituito dai punti campionari Croce e Testa: $\Omega = \{C, T\}$. Si assume la variabile casuale $X : \Omega \rightarrow S \subseteq \mathbb{R}$, che ad ogni punto dello spazio (croce e testa) associa uno ed un solo numero reale, ad esempio 0 a C ed 1 a T; simbolicamente abbiamo: $X(C) = 0$; $X(T) = 1$.
Abbiamo $S = \{0,1\}$ e la funzione di densità di probabilità:

$$p(x) = \begin{cases} p(0) = P(X = 0) = 1/2 & per\ 0 \in S \\ p(1) = P(X = 1) = 1/2 & per\ 1 \in S \\ 0 & per\ x \in \mathbb{R} \backslash S \end{cases}$$



Variabili casuali, funzione di densità e distribuzione di probabilità

Funzione di densità di probabilità di una variabile casuale discreta

Esempio:

lancio di 2 monete $\Omega^2 = \{(T,T),(C,T),(T,C),(C,C)\}$
Variabile casuale discreta: X numero totale degli esiti T; $X : \Omega^2 \to S \subseteq \mathbb{R}$; i suoi possibili valori sono:
0 se non si ottiene alcuna testa: $X(C,C) = 0$
1 se una delle due monete dà testa: $X[(C,T),(T,C)] = 1$
2 se entrambe le monete danno testa: $X(T,T) = 2$
Abbiamo le probabilità:

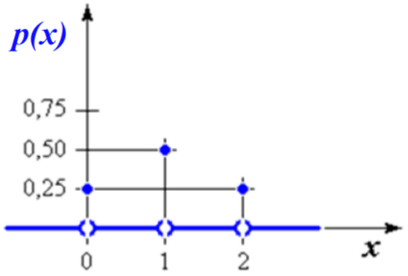
$P(X=0) = P(C,C) = 1/4$

$P(X=1) = P(C,T) = P(T,C) = \frac{2}{4} = 1/2$

$P(X=2) = P(T,T) = 1/4$

Abbiamo $S = \{0,1,2\}$ e la funzione di densità di probabilità:

$$p(x) = \begin{cases} p(0) = P(X=0) = \frac{1}{4} = 0.25 & per\ 0 \in S \\ p(1) = P(X=1) = \frac{1}{2} = 0.50 & per\ 1 \in S \\ p(2) = P(X=2) = \frac{1}{4} = 0.25 & per\ 2 \in S \\ 0 & per\ x \in \mathbb{R} \setminus S \end{cases}$$



Variabili casuali, funzione di densità e distribuzione di probabilità

Funzione di densità di probabilità di una variabile casuale continua

Per le variabili casuali continue non ha senso determinare la probabilità su ciascun valore assunto dalla variabile casuale semplice X (visto che è sempre nulla), ma è corretto parlare di probabilità che la variabile casuale assuma valori in un intervallo, anche piccolissimo, del tipo $(x, x + dx]$.

Infatti, una variabile casuale continua X è una funzione che può assumere tutti i valori compresi in un intervallo (a, b). I casi possibili sono di fatto infiniti e quindi:

$$P(X = x) = \frac{numero\ di\ casi\ favorevoli}{numero\ di\ casi\ possibili} = \frac{1}{\infty} = 0$$

Esempio: La probabilità che la temperatura nella stanza sia esattamente 20°C è nulla, invece sarà maggiore di zero la probabilità che la temperatura sia compresa fra 20°C e 20.1°C.

Variabili casuali, funzione di densità e distribuzione di probabilità

Funzione di densità di probabilità di una variabile casuale continua

Data la variabile casuale continua X che assume valori nell'intervallo $[a, b]$, $-\infty \leq a < b \leq \infty$, la *funzione di densità di probabilità (o PDF)* è la funzione che ad ogni numero reale associa il limite, per dx che tende a 0, del rapporto tra la probabilità che la variabile casuale assuma valori nell'intervallo $(x, x + dx]$ e l'ampiezza dx.

$$p: \mathbb{R} \to \mathbb{R}:\ x \to \lim_{dx \to 0} \left[\frac{P(x < X \leq x + dx)}{dx} \right]$$

La funzione di densità in x rappresenta quanto vale la probabilità *intorno ad* x in rapporto all'ampiezza di tale *intorno*. Il termine funzione di densità, serve proprio ad evocare quanto è *densa* la probabilità.

Variabili casuali, funzione di densità e distribuzione di probabilità

Funzione di densità di probabilità di una variabile casuale continua

Ogni evento deve essere ricondotto all'unione, negazione o intersezione di intervalli

$$P(X \leq x) = \int_{-\infty}^{x} p(x)dx = Z(x)$$

$Z(x)$ *funzione di ripartizione* (o *funzione cumulativa*)

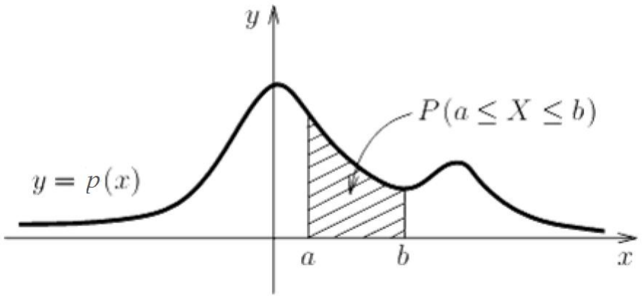
Quindi per ogni coppia di numeri reali (a, b), $-\infty \leq a < b \leq \infty$, si ha

$$P(a \leq X \leq b) = \int_{a}^{b} p(x)dx = Z(b) - Z(a)$$

La funzione di densità di probabilità è sempre $p(x) \geq 0$

L'area totale sottesa alla funzione è uguale a 1, ossia:

$$\int_{-\infty}^{+\infty} p(x)dx = 1$$

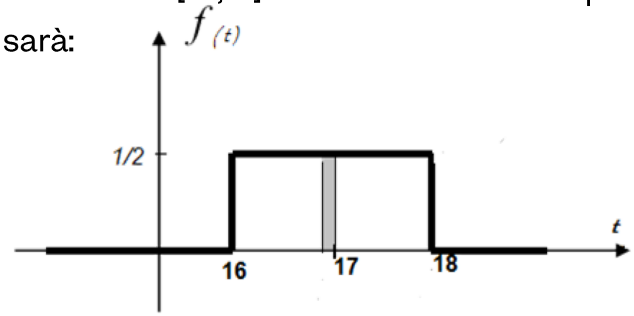


Variabili casuali, funzione di densità e distribuzione di probabilità

Funzione di densità di probabilità di una variabile casuale continua

Esempio:

A aspetta una telefonata da B il quale ha preannunciato che chiamerà, in un istante non meglio precisato, fra le 16:00 e le 18:00. A si deve però assentare dalle 16:45 alle ore 17:00. Qual'è la probabilità che la telefonata arrivi mentre A è assente?
L'istante della telefonata è una variabile casuale X continua. Per quanto ne sa A, tutti i momenti dalle 16:00 alle 18:00 sono equiprobabili, mentre fuori da questo intervallo la probabilità è zero. E' intuitivo che la densità abbia un valore costante c sull'intervallo [16, 18] ed ha il valore zero fuori di questo intervallo. Quanto vale la costante c? Deve essere tale da soddisfare la relazione $\int_{-\infty}^{+\infty} f(t)dt = 1$, ovvero l'area del rettangolo con base [16,18] e altezza c sia 1. Dunque $2c = 1$ e perciò abbiamo $c = 1/2$. La funzione di densità della variabile X, sarà:



$$f(t) = \begin{cases} 0 & per\ t < 16 \\ \frac{1}{2} & per\ 16 \leq t \leq 18 \\ 0 & per\ t > 18 \end{cases}$$

$P(16{:}45 \leq X \leq 17{:}00) = \frac{1}{4}\frac{1}{2}$ =0.125

Valore atteso

Consideriamo una variabile casuale X con una distribuzione discreta caratterizzata da un certo numero (finito o infinito) di valori x_i, dove $i = 1, \ldots, j, \ldots$. Indichiamo con la lettera $p(x_i)$ la funzione di probabilità della variabile casuale X e definiamo il valore atteso di X, che indicheremo con $E(X)$, mediante l'equazione:

$$E(X) = \sum_i x_i p(x_i)$$

Consideriamo una variabile casuale X con una distribuzione continua la cui densità di probabilità è data dalla funzione $p(x)$. Il valore atteso (o valor medio) è definito dall'equazione seguente:

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

Valore atteso

Proprietà del valore atteso

- Sia $Y = aX + b$, allora $E(Y) = E(aX + b) = aE(X) + b$

- Sia a tale che $P(X \geq a) = 1$, allora $E(X) \geq a$
- Sia b tale che $P(X \leq b) = 1$, allora $E(X) \leq b$

- Siano a e b tale che $P(a \leq X \leq b) = 1$, allora $a \leq E(X) \leq b$

- Siano X_A e X_B due variabili casuali, $E(X_A + X_B) = E(X_A) + E(X_B)$

- Siano $X_A, X_B, \ldots$ un insieme finito di variabili casuali, $E(X_A + X_B + \cdots) = E(X_A) + E(X_B) + \cdots$

- Allora, per una combinazione lineare di variabili casuali si ha:

$$E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)$$

Varianza

Consideriamo una variabile casuale X caratterizzata da una densità di probabilità $p(x)$ e da un valore atteso $\mu \equiv E(X)$. La varianza $Var(X)$ della distribuzione di probabilità è definita come:

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

Proprietà della varianza

- La varianza è uguale a zero se e solo se esiste un numero reale c tale che $P(X = c) = 1$.
- Sia $Y = aX + b$, allora $Var(Y) = Var(aX + b) = a^2 Var(X)$
- Per ogni variabile casuale X, si ha
$$Var(X) = E[(X-\mu)^2] = E[X^2] - [E(X)]^2 = E[X^2] - \mu^2$$

Varianza

Siano X_A e X_B due variabili casuali.

Definiamo covarianza delle due variabili casuali:

$$Cov(X_A, X_B) \equiv E\big[\big(X_A - E(X_A)\big)\big(X_B - E(X_B)\big)\big] = E[X_A X_B] - E(X_A)E(X_B)$$

Quindi:

$$Var(X_A + X_B) = E\left[\big(X_A + X_B - E(X_A + X_B)\big)^2\right] =$$

$$= E\left[\big(X_A - E(X_A)\big)^2\right] + E\left[\big(X_B - E(X_B)\big)^2\right] + 2E\big[\big(X_A - E(X_A)\big)\big(X_B - E(X_B)\big)\big]$$

$$Var(X_A + X_B) = Var(X_A) + Var(X_B) + 2\, Cov(X_A, X_B)$$

Varianza

Siano X_A e X_B due variabili casuali indipendenti.

Due variabili casuali sono dette indipendenti qualora la conoscenza dei valori assunti dall'una non influenzi in alcun modo la previsione sui valori che saranno assunti dall'altra. Poiché, per due variabili casuali indipendenti, abbiamo

$$E(X_A \cdot X_B) = E(X_A) \cdot E(X_B)$$

$$Cov(X_A, X_B) = 0$$

Possiamo dunque concludere che la varianza della somma di due variabili indipendenti è uguale alla somma delle varianze delle singole variabili:

$$Var(X_A + X_B) = Var(X_A) + Var(X_B)$$

Distribuzione binomiale

La distribuzione binomiale, detta anche distribuzione di Bernoulli, è la probabilità di ottenere k successi in n prove indipendenti.

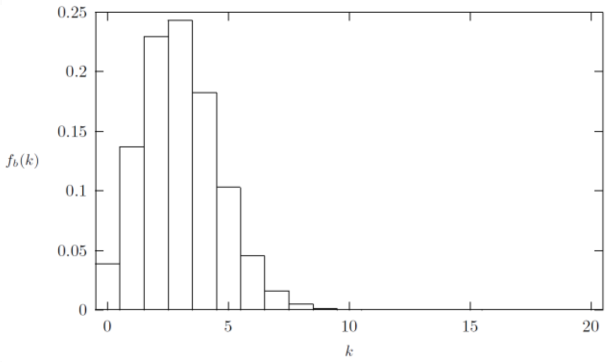
p: probabilità favorevole all'evento
$q = 1 - p$: probabilità sfavorevole all'evento.

$$f_b(k|n,p) = \begin{cases} \binom{n}{k} p^k q^{n-k} & per \;\; k = 0, \dots, n \\ 0 & altrimenti \end{cases}$$

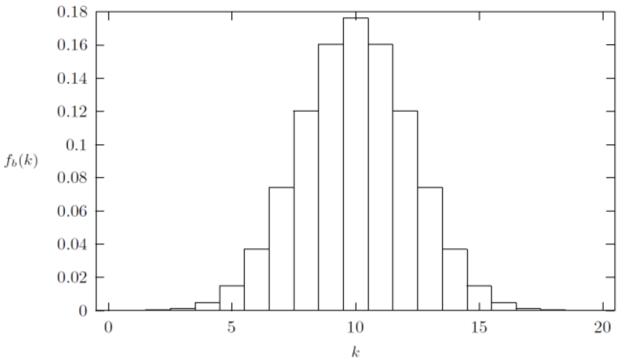
$$\mu = np = \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k}$$

$$\sigma^2 = npq$$

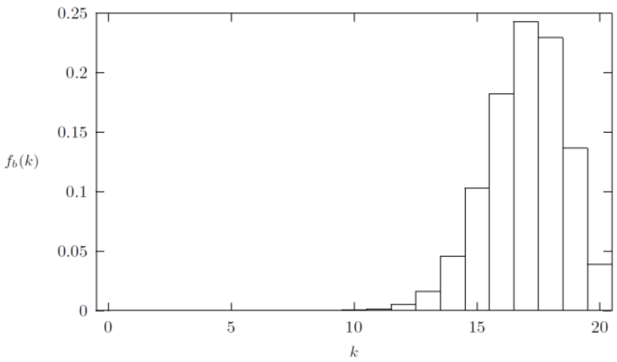
Distribuzione binomiale



Densità di probabilità binomiale con p=0.15 e n=20



Densità di probabilità binomiale con p=0.50 e n=20



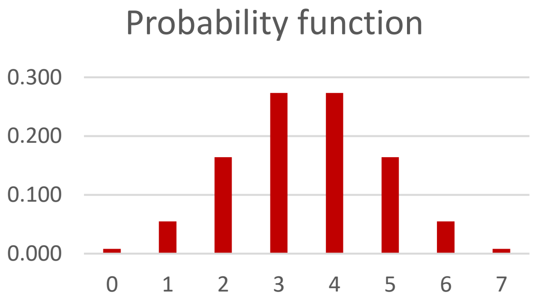
Densità di probabilità binomiale con p=0.85 e n=20

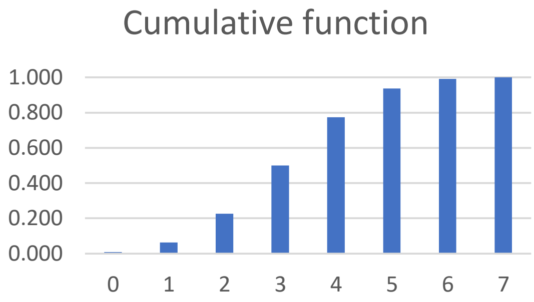
Distribuzione binomiale

Esempio: Una moneta è lanciata n =7 volte, studiare la distribuzione di probabilità della variabile binomiale X = numero di volte in cui compare testa. $X : \Omega \to \mathbb{R}, \Omega = \{C, T\}$.

$p = q =$ 1/2 con $x = 0, \dots, 7$

x	f(x)	Z(x)
0	0.008	0.008
1	0.055	0.063
2	0.164	0.227
3	0.273	0.500
4	0.273	0.773
5	0.164	0.938
6	0.055	0.992
7	0.008	1.000





Esempio: Un partecipante ad una gara di tiro con l'arco colpisce il bersaglio in media con la probabilità dell'80%. Calcolare il numero medio di centri che egli può aspettarsi con 20 tiri e calcolare anche la varianza:
probabilità di colpire p = 80/100 = 4/5
probabilità di non colpire q = 20/100 = 1/5
X = numero di volte in cui centra il bersaglio $X : \Omega \to \mathbb{R},\ \Omega = \{centro\ colpito, certro\ non\ colpito\}$

Media $\mu = np$ = 20 4/5 = 16
Varianza $\sigma^2 = npq$ = 20 4/5 1/5 = 3.2

Distribuzione ipergeometrica

La distribuzione binomiale descrive l'estrazione con reinserimento.

La distribuzione ipergeometrica è una distribuzione di probabilità discreta che descrive l'estrazione senza reinserimento.

La distribuzione descrive la variabile aleatoria che conta, per r elementi distinti estratti a caso (in modo equiprobabile) da un insieme A di cardinalità n, quanti sono nel sottoinsieme B di cardinalità h.

$\binom{n}{r}$ è il numero di possibili estrazioni di r elementi tra gli n di A

$\binom{h}{k}$ è il numero di possibili estrazioni di k elementi tra gli h di B

$\binom{n-h}{r-k}$ è il numero di possibili estrazioni dei restanti $r-k$ elementi tra gli $n-h$ non in B

Distribuzione ipergeometrica

$$f_i(k|n,h,r) = \begin{cases} \dfrac{\binom{h}{k}\binom{n-h}{r-k}}{\binom{n}{r}} & per\ \ k = 0, \dots, n \\ 0 & altrimenti \end{cases}$$

$$\mu = \frac{rh}{n}$$

$$\sigma^2 = \frac{r(n-r)h(n-h)}{n^2(n-1)}$$

Ad esempio: descrive, data un'urna contenente h palline bianche e $n-h$ palline nere, il numero di palline bianche k che vengono ottenute estraendo senza reinserimento r palline.

Distribuzione di Poisson

Se consideriamo le probabilità di un evento che capita molto raramente, utilizzando la distribuzione binomiale, dovremo fare un numero enorme di calcoli considerando probabilità dell'evento e probabilità contraria.

La distribuzione di Poisson, che è una approssimazione della distribuzione binomiale, consente di ridurre il costo computazionale.
Quando n tende a infinito, lasciando fisso np, la distribuzione binomiale tende alla distribuzione di Poisson.
In statistica quest'approssimazione viene solitamente accettata quando $n \geq 20$ e $p \leq 1/20$, oppure quando $n \geq 100$ e $np \leq 10$.

Esempio classico: Nel 1898 Ladislaus von Bortkiewicz pubblicò uno studio (The Law of Small Numbers) nel quale faceva vedere come il numero di soldati dell'esercito prussiano morti fra il 1875 e il 1894 a seguito di un calcio di cavallo, pur variabile da un anno all'altro e da un'armata all'altra, seguisse, in maniera molto fedele, una distribuzione di Poisson.

Distribuzione di Poisson

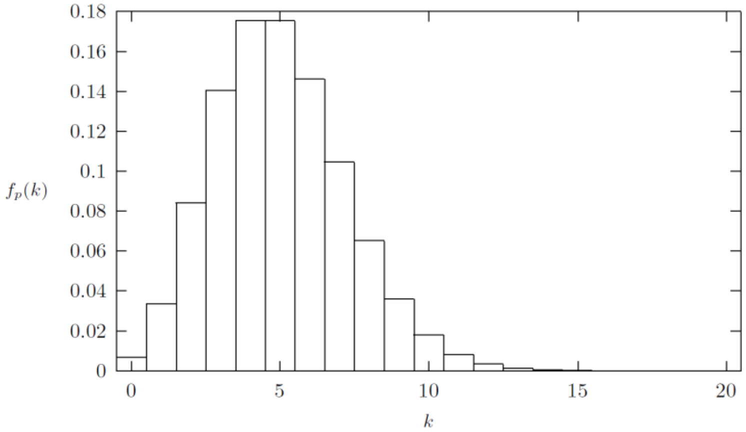
Densità di probabilità di Poisson:

$$f_p(k|\lambda) = \begin{cases} \dfrac{\lambda^k}{k!}\exp(-\lambda) & per \;\; k = 0, 1, \ldots \\ 0 & altrimenti \end{cases}$$

dove $\lambda = np > 0$.

$\mu = \lambda$

$\sigma^2 = \lambda$



Densità di probabilità di Poisson con λ=5.

Distribuzione di Poisson

Esempio: Nel suo studio, von Bortkiewicz analizzò i verbali di 10 reggimenti relativi ad un periodo di 20 anni, constatando che in totale 122 soldati erano morti a seguito di un calcio di cavallo. Quindi, in media, in ciascun reggimento ogni anno il numero di decessi per calcio di cavallo è stato pari a 0.61 (122/200 dove 200 sono il numero di reggimenti totali). La varianza è 0.6079.
La varianza di questa distribuzione sperimentale è quasi identica alla sua media, come atteso in una distribuzione poissoniana teorica. E' una buona indicazione che questi eventi seguono la legge di Poisson. Si calcoli la distribuzione di Poisson con λ=0.61.

Morti per un calcio di mulo	nessuno	un morto	due morti	tre morti	quattro morti	cinque morti
Numero reggimenti osservati	109	65	22	3	1	0
Numero reggimenti attesi	108.7	66.3	20.2	4.1	0.6	0.1
Probabilità	0.543	0.331	0.101	0.021	0.003	0.000

Distribuzione di Poisson

Esempio: La percentuale di pezzi difettosi prodotti da una macchina è, in media, dello 0.2%. Siccome la ditta esporta tali pezzi in confezioni di 1000, calcolare quanti pezzi in più dovranno essere messi in ogni confezione perché la probabilità di avere in una confezione meno di 1000 pezzi efficienti sia inferiore allo 0.01%.

La macchina in media produce 2 pezzi difettosi ogni mille, quindi la probabilità che un pezzo sia difettoso è **p= 0.002.**
Il numero di pezzi per confezione è **n= 1000.**
Essendo **n=1000** molto grande e **p=0.002** molto piccola possiamo approssimare bene utilizzando la distribuzione di Poisson. Costruiamo la variabile aleatoria di Poisson con λ=np=2 per 0, 1, 2, 3, 4, 5, ... pezzi e vediamo a quale numero corrisponde una probabilità inferiore a 0.01%. Otteniamo una probabilità del:
4.98% di avere 0 pezzi difettosi
13.53% di avere 1 pezzo difettoso
27.07% di avere 2 pezzi difettosi
27.07% di avere 3 pezzi difettosi
36.09% di avere 4 pezzi difettosi
9.02% di avere 5 pezzi difettosi
2.41% di avere 6 pezzi difettosi
0.21% di avere 7 pezzi difettosi
0.005% di avere 8 pezzi difettosi (probabilità che ogni 20 confezioni da 1000 pezzi ve ne sia una con 8 pezzi difettosi) $<$ 0.01%
0.0002% di avere 9 pezzi difettosi (probabilità che ogni 500 confezioni da 1000 pezzi ve ne sia una con 9 pezzi difettosi)
In **ogni confezione di 1000 pezzi vanno aggiunti 8 pezzi per essere ragionevolmente sicuri che la confezione contenga 1000 pezzi efficienti.**

Distribuzione gaussiana

- Molte distribuzioni che si incontrano nel mondo reale sono effettivamente di tipo normale, ovvero sono molto ben approssimate da una Normale: la Normale è un modello che descrive adeguatamente la distribuzione di numerosi fenomeni.
- Molte distribuzioni, di per sé anche lontane come forma dalla Normale, magari anche asimmetriche (purché unimodali), sono normalizzabili mediante una trasformazione di variabile (es. $w = \log(x)$).
- In generale si distribuiscono normalmente quei caratteri o fenomeni che sono il risultato di un gran numero di piccoli fattori, tra loro indipendenti (es. variabili biometriche, prodotto di serie, processo di misura).
- I valori prodotti da un processo di misura sono generalmente distribuiti normalmente: misurando ripetutamente lo stesso oggetto, lo strumento non produce sempre lo stesso valore, a causa del cosiddetto errore di misura, risultato della somma di un gran numero di piccoli fattori indipendenti che influenzano il processo.

Distribuzione gaussiana

La densità di probabilità di Gauss è definita dalla seguente equazione:

$$f_g(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

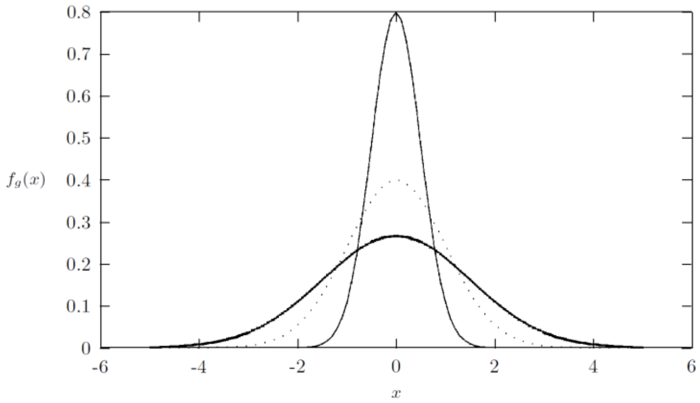
Si può sempre standardizzare la variabile normale: $z = \frac{x-\mu}{\sigma}$

La densità normale con media uguale a zero e con varianza uguale a 1, $f_g(z|0,1)$, è chiamata densità normale standard:

$$f_g(z|0,1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$

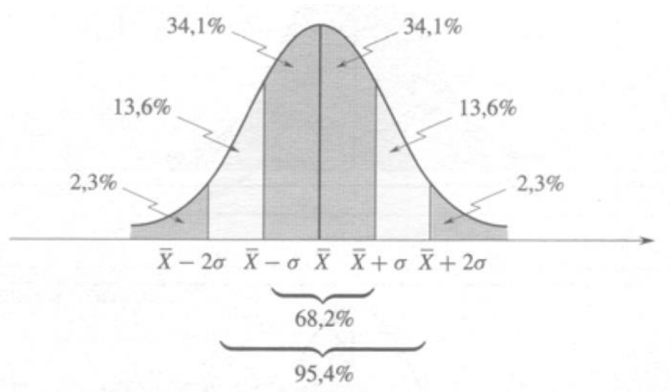
Indichiamo con $\phi_g(z)$ la funzione di ripartizione della densità normale standard:

$$\phi_g(z) = \int_{-\infty}^{z} f_g(t|0,1)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{1}{2}t^2\right) dt$$



Densità di probabilità normale, o gaussiana, con μ = 0. Le tre curve corrispondono a σ = 0.5 (la campana più alta e stretta), a σ = 1 (la densità normale standard) ed a σ = 1.5 (la campana più bassa e larga)

Distribuzione gaussiana



Il 68,26% dei valori è compreso fra $\bar{X}-\sigma$ e $\bar{X}+\sigma$.
Il 95,44% dei valori è compreso fra $\bar{X}-2\sigma$ e $\bar{X}+2\sigma$.
Il 99,73% dei valori è compreso fra $\bar{X}-3\sigma$ e $\bar{X}+3\sigma$.

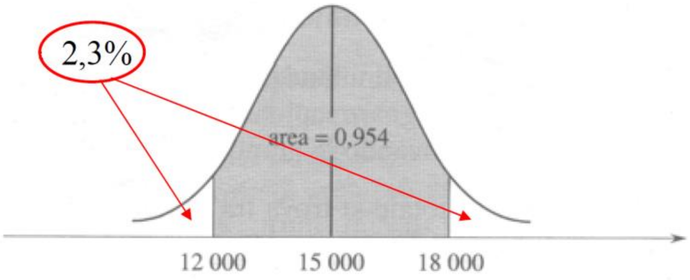
Distribuzione gaussiana

Esempio:

Supponiamo di sapere che il reddito delle famiglie degli studenti dell'Università di Trieste si distribuiscano normalmente con $\bar{X}$=15000 EUR e σ=1500 EUR.

Domande:

- Quante famiglie guadagnano più di 18000 EUR ?
- Se l'ARDIS vuole assegnare una borsa di studio al 2.3% degli studenti meno abbienti, che soglia di reddito massimo deve fissare ?



Osserviamo che:

$\bar{X} - 2\sigma$ = 15000-3000 = 12000

$\bar{X} + 2\sigma$ = 15000+3000 = 18000

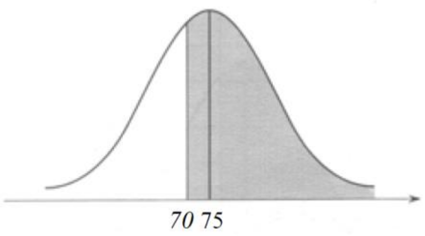
- Circa il 2,3% delle famiglie guadagnano più di 18000 EUR.
- La soglia di reddito massimo che l'ARDIS deve fissare è pari a 12000 EUR.

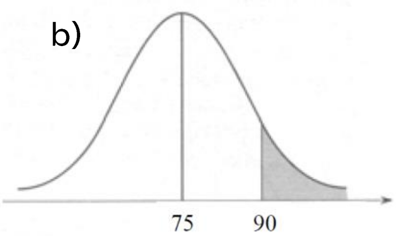
Distribuzione gaussiana

Esempio: La velocità delle auto rilevata dall'autovelox su una tangenziale si distribuisce normalmente con $\bar{X}$=75 km/h e σ =8 km/h.

a) Che percentuale di auto superano il limite di velocità di 70 km/h ?

b) A quanti automobilisti su 100 viene ritirata la patente (oltre 90 km/h) ?

a) 

b) 

a) $z = \frac{x-\bar{X}}{\sigma} = \frac{x-75}{8}$ per x=70 z= -0.625,

Dalle tabelle 1-$\phi_g(-0.625) = \phi_g$(0.625)=(0.7324+0.7357)1/2=0.7341

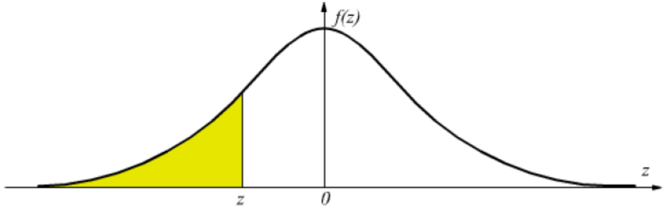
Il 73,41% degli autoveicoli in transito superano il limite di velocità di 70 km/h .

b) $z = \frac{x-\bar{X}}{\sigma} = \frac{x-75}{8}$ per x=90 z= 1.88,

Dalle tabelle 1-ϕ_g(1.88)=1-0.9699=0.0301

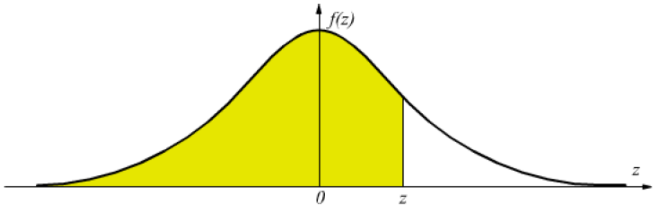
A 3 automobilisti su 100 viene ritirata la patente.

Distribuzione gaussiana



Probabilità cumulativa per valori negativi di z

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,1	0,001	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,001	0,001
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,002	0,0019
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,003	0,0029	0,0028	0,0027	0,0026
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,004	0,0039	0,0038	0,0037	0,0036
-2,5	0,0062	0,006	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,4	0,0082	0,008	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,011
-2,1	0,0179	0,0174	0,017	0,0166	0,0162	0,0158	0,0154	0,015	0,0146	0,0143
-2	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,025	0,0244	0,0239	0,0233
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,5	0,0668	0,0655	0,0643	0,063	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,102	0,1003	0,0985
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,123	0,121	0,119	0,117
-1	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,166	0,1635	0,1611
-0,8	0,2119	0,209	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,7	0,242	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,5	0,3085	0,305	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,281	0,2776
-0,4	0,3446	0,3409	0,3372	0,3336	0,33	0,3264	0,3228	0,3192	0,3156	0,3121
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,352	0,3483
-0,2	0,4207	0,4168	0,4129	0,409	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0	0,5	0,496	0,492	0,488	0,484	0,4801	0,4761	0,4721	0,4681	0,4641



Probabilità cumulativa per valori positivi di z

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5	0,504	0,508	0,512	0,516	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,591	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	0,648	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,67	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,695	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,719	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	0,758	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,791	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	0,834	0,8365	0,8389
1	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	0,877	0,879	0,881	0,883
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,898	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	0,937	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	0,975	0,9756	0,9761	0,9767
2	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	0,983	0,9834	0,9838	0,9842	0,9846	0,985	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,989
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	0,992	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,994	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	0,996	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,997	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	0,998	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	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,999	0,999
3,1	0,999	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

Distribuzione Chi-quadro χ^2 (di Pearson)

Si considerano ν variabili aleatorie indipendenti $Y_1, Y_2, \dots, Y_\nu$ distribuite secondo la legge normale con media $\mu = 0$ e varianza $\sigma^2 = 1$; allora la variabile χ^2 è:

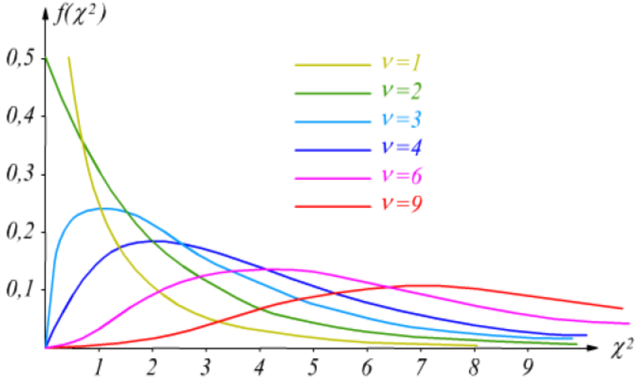
$$\chi^2 = Y_1^2 + Y_2^2 + \cdots + Y_\nu^2$$

La variabile χ^2 si distribuisce secondo la:

$$f_{\chi^2}(t|\nu) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} t^{\frac{\nu}{2}-1} \exp\left(-\frac{t}{2}\right)$$

$$\Gamma\left(\frac{\nu}{2}\right) = \left(\frac{\nu}{2} - 1\right)! \quad \text{se } \nu \text{ pari}$$

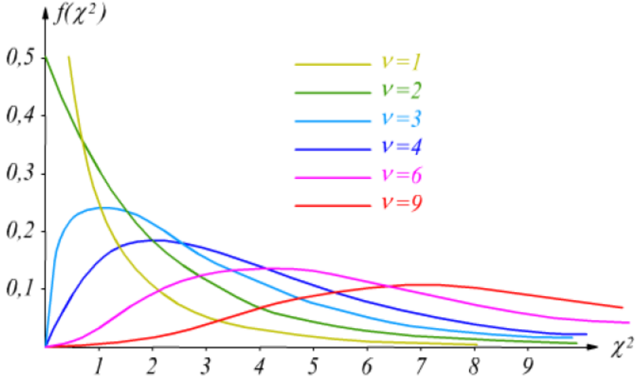
$$\Gamma\left(\frac{\nu}{2}\right) = \sqrt{\pi}\,\frac{(\nu-2)!!}{2^{\frac{\nu-1}{2}}} \quad \text{se } \nu \text{ dispari}$$



Distribuzione Chi-quadro χ^2 (di Pearson)

Le proprietà fondamentali della distribuzione chi-quadro sono:

- Asimmetria.
- Dipendenza dal parametro intero ν che indica i gradi di libertà: per ogni valore di ν si ha una curva diversa.
- La variabile χ^2 non può assumere valori negativi dato che è una somma di quadrati.
- E' completamente definita nel primo quadrante.



Distribuzione Chi-quadro χ^2 (di Pearson)

Sia X una variabile aleatoria con media $E(X) = \mu$ e varianza $\text{Var}(X) = \sigma^2$ e sia $X_1, X_2, \dots, X_n$ un campione casuale.
Si definisce **varianza campionaria** la quantità:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

La **deviazione standard campionaria** è la radice quadrata della varianza campionaria:

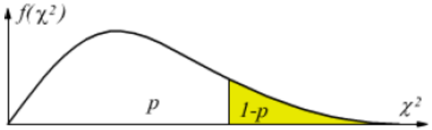
$$s = \sqrt{s^2}$$

La varianza campionaria ha una distribuzione Chi-quadro.

Esempi di applicazioni:
tra i test maggiormente conosciuti che utilizzano la distribuzione Chi-quadro ci sono

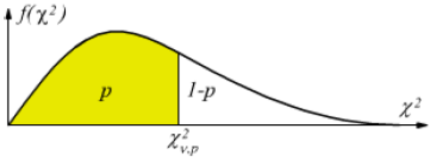
- ❑ Goodness of fit
- ❑ Test dell'indipendenza

Distribuzione Chi-quadro χ^2 (di Pearson)



valori critici della coda superiore (1-p) della distribuzione χ^2

v	0,5	0,25	0,1	0,05	0,025	0,01	0,005	0,001
1	0,45	1,32	2,71	3,84	5,02	6,63	7,88	10,83
2	1,39	2,77	4,61	5,99	7,38	9,21	10,6	13,82
3	2,37	4,11	6,25	7,81	9,35	11,34	12,84	16,27
4	3,36	5,39	7,78	9,49	11,14	13,28	14,86	18,47
5	4,35	6,63	9,24	11,07	12,83	15,09	16,75	20,52
6	5,35	7,84	10,64	12,59	14,45	16,81	18,55	22,46
7	6,35	9,04	12,02	14,07	16,01	18,48	20,28	24,32
8	7,34	10,22	13,36	15,51	17,53	20,09	21,96	26,12
9	8,34	11,39	14,68	16,92	19,02	21,67	23,59	27,88
10	9,34	12,55	15,99	18,31	20,48	23,21	25,19	29,59



valori critici della coda inferiore (p) della distribuzione χ^2

v	0.1	0.05	0.025	0.01	0.001
1	0.016	0.004	0.001	0.000	0.000
2	0,147	0,072	0.051	0.020	0.002
3	0,406	0,244	0,15	0,080	0.024
4	1.064	0,494	0,336	0,206	0.091
5	1.610	1.145	0,577	0,385	0,146
6	2.204	1.635	1.237	0,606	0,265
7	2.833	2.167	1.690	1.239	0,415
8	3.490	2.733	2.180	1.646	0,595
9	4.168	3.325	2.700	2.088	1.152
10	4.865	3.940	3.247	2.558	1.479

Distribuzione t-Student

Consideriamo due variabili casuali indipendenti:

Z: distribuita normalmente con media $\mu = 0$ e varianza $\sigma^2 = 1$
Y: distribuita secondo chi-quadro con ν gradi di libertà.

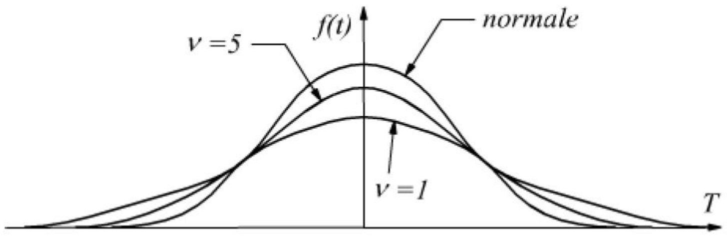
$$T = \frac{Z}{\sqrt{Y}}$$

ha una distribuzione t-Student con ν gradi di libertà

$$f_S(t|\nu) = \frac{1}{\sqrt{\nu}B\left(\frac{1}{2},\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

dove B è la funzione beta

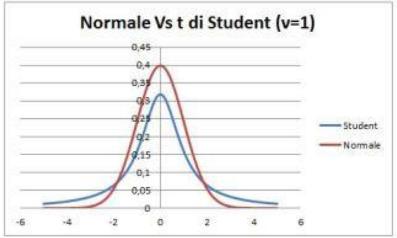
$$B\left(\frac{1}{2},\frac{\nu}{2}\right) = \frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{1}{2}+\frac{\nu}{2}\right)}$$

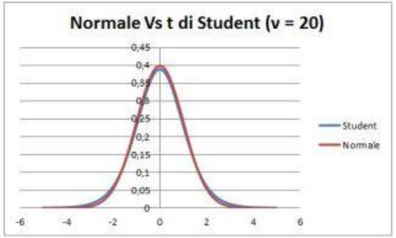


Distribuzione t-Student

Le caratteristiche di questa distribuzione sono:

- Simmetria rispetto al valor medio $\mu = 0$.
- Dipendenza dal parametro ν gradi di libertà.
- Tende alla distribuzione normale quando ν tende ad infinito ($\nu \to \infty$) (già con campioni di 30-50 dati praticamente coincidono).
- Per ogni valore di ν si ha una diversa distribuzione.
- Come nel caso della normale, l'area compresa tra la curva e l'asse delle ascisse vale 1.





Distribuzione t-Student

Esempi di applicazioni sono test di ipotesi su:

- **media di una popolazione** di cui non è conosciuta la deviazione standard della popolazione e calcolo dell'intervallo di confidenza

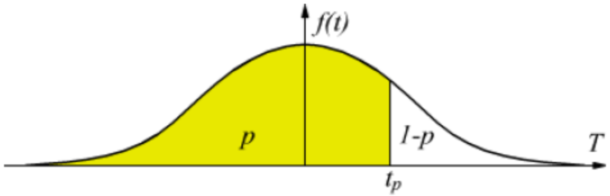
  Nel caso in cui la varianza della popolazione σ^2 non è nota:
  - se il numero n del campione è grande, si può sostituire σ^2 con la varianza del campione s^2
  - se, invece, l'ampiezza del campione è piccolo e tale campione proviene da una distribuzione normale, data una popolazione normale avente media μ da cui si estraggono campioni casuali con ampiezza n, indicando con $\bar{X}$ la media campionaria e con s lo scarto quadratico medio campionario (o deviazione standard campionaria), la variabile

  $$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

  é una variabile aleatoria avente distribuzione t-Student con grado di libertà $\nu = n - 1$.

- **differenza tra due medie**
- **differenza tra due medie di 2 campioni dipendenti**
- **coefficiente di correlazione**

Distribuzione t-Student



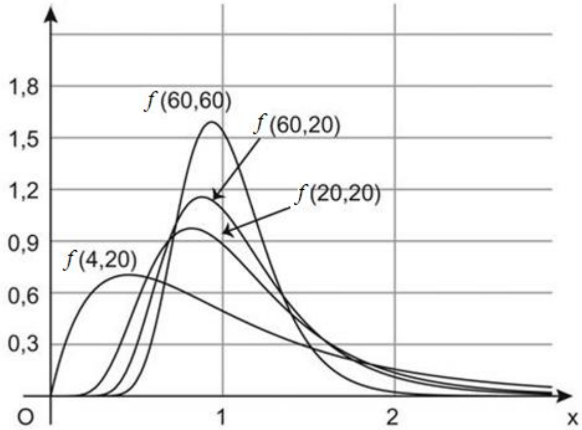
Distribuzione t di Student P(T<t)=p

v	0,75	0,8	0,85	0,9	0,95	0,975	0,99	0,995	0,999	0,9995
1	1	1,376	1,963	3,078	6,314	12,71	31,82	63,66	318,31	636,62
2	0,816	1,061	1,386	1,886	2,92	4,303	6,965	9,925	22,327	31,599
3	0,765	0,978	1,25	1,638	2,353	3,182	4,541	5,841	10,215	12,924
4	0,741	0,941	1,19	1,533	2,132	2,776	3,747	4,604	7,173	8,61
5	0,727	0,92	1,156	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	0,718	0,906	1,134	1,44	1,943	2,447	3,143	3,707	5,208	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,706	0,889	1,108	1,397	1,86	2,306	2,896	3,355	4,501	5,041
9	0,703	0,883	1,1	1,383	1,833	2,262	2,821	3,25	4,297	4,781
10	0,7	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	3,93	4,318
13	0,694	0,87	1,079	1,35	1,771	2,16	2,65	3,012	3,852	4,221

Distribuzione F di Fischer

La variabile aleatoria F si ottiene come rapporto di due variabili aleatorie χ^2 tra loro indipendenti, X_1 e X_2, divise per i rispettivi gradi di libertà ν_1 e ν_2 e spesso è indicata con $F(\nu_1, \nu_2)$:

$$F(\nu_1, \nu_2) = \frac{X_1/\nu_1}{X_2/\nu_2}$$



La sua funzione di densità è $f(x) = 0$ se $x \leq 0$ e, se $x > 0$, è

$$f_F(x) = \nu_1^{\left(\frac{\nu_1}{2}\right)} \nu_2^{\left(\frac{\nu_2}{2}\right)} \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \frac{x^{\left(\frac{\nu_1 - 2}{2}\right)}}{(\nu_1 x + \nu_2)^{\left(\frac{\nu_1 + \nu_2}{2}\right)}}$$

dove Γ è la funzione gamma di Eulero.

La forma del suo grafico, comunque asimmetrico, varia molto a seconda dei valori dei due parametri ν_1 e ν_2.

Distribuzione F di Fischer

La distribuzione di Fischer descrive l'andamento del rapporto tra le varianze delle due distribuzioni in funzione dei loro gradi di libertà ed è principalmente utilizzata per verificare la differenza statistica tra le due varianze.

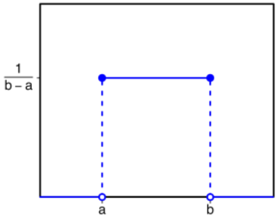
Il valore medio μ e la varianza σ^2 della distribuzione F di Fisher non sono sempre definiti.

$$\mu = \frac{\nu_2}{\nu_2 - 2} \text{ per } \nu_2 \geq 3$$

$$\sigma^2 = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)} \text{ per } \nu_2 \geq 5$$

Distribuzione rettangolare o distribuzione continua uniforme

E' una distribuzione di probabilità continua che è uniforme su un insieme, ovvero che attribuisce la stessa probabilità a tutti i punti appartenenti ad un dato intervallo $[a, b]$ contenuto nell'insieme.

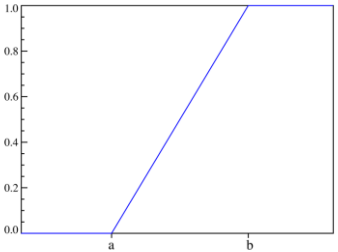


La densità di probabilità rettangolare è:

$f_r(x|a,b) = \frac{1}{b-a}$ su $[a,b]$, $f_r(x|a,b) = 0$ altrove

La funzione di ripartizione è:

$F_r(x|a,b) = \frac{x-a}{b-a}$ su $[a,b]$, $F_r(x|a,b) = 0$ per $x < a$, $F_r(x|a,b) = 1$ per $x > b$



$$\mu = (a+b)/2 \qquad \sigma^2 = (b-a)^2/12$$

Si può sempre normalizzare la variabile sull'intervallo unitario $[0,1]$: $u = a + (b-a)x$

$f_r(u|0,1) = 1$ su $[0,1]$, $f_r(u|0,1) = 0$ altrove

$F_r(u|0,1) = u$ su $[0,1]$, $F_r(u|0,1) = 0$ per $u < 0$, $F_r(u|0,1) = 1$ per $u > 1$

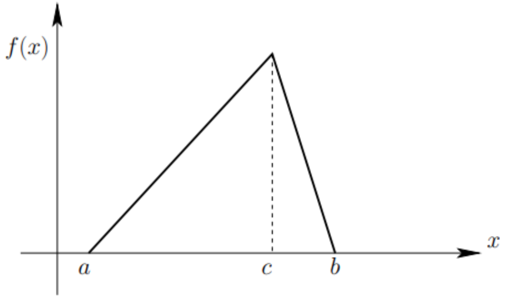
$\mu = 1/2 \qquad \sigma^2 = 1/12$

Distribuzione triangolare

E' una distribuzione di probabilità continua la cui funzione di densità di probabilità descrive un triangolo, ovvero è nulla sui due valori estremi ed è lineare tra questi ed un valore intermedio (la moda). Viene utilizzata come modello quando il campione a disposizione è molto ristretto, stimando il minimo, il massimo e la moda.

La densità di probabilità triangolare è:

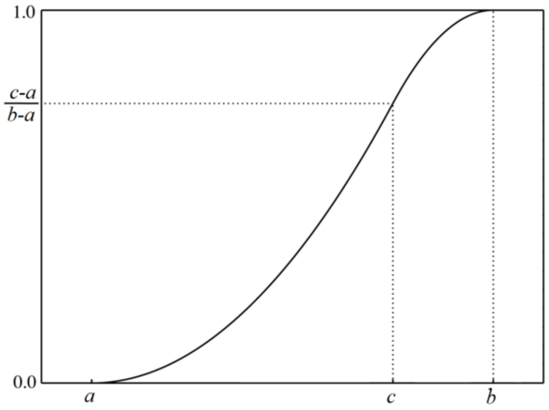
$$f_t(x|a,b,c) = \begin{cases} \dfrac{2}{b-a}\dfrac{x-a}{c-a} & se\ a \le x \le c \\ \dfrac{2}{b-a}\dfrac{b-x}{b-c} & se\ c \le x \le b \\ 0 & altrove \end{cases}$$



Distribuzione triangolare

La funzione di ripartizione è:

$$F_t(x|a,b,c) = \begin{cases} \dfrac{1}{b-a}\dfrac{(x-a)^2}{c-a} & se\ a \le x \le c \\ 1 - \dfrac{1}{b-a}\dfrac{(b-x)^2}{b-c} & se\ c \le x \le b \\ 0 & se\ x < a \\ 1 & se\ x > b \end{cases}$$



$$\mu = \frac{a+b+c}{3}$$

$$\sigma^2 = \frac{(a+b+c)^2 - (ab+ac+bc)}{18}$$

Distribuzione triangolare

Se la distribuzione di probabilità triangolare è simmetrica rispetto all'origine, detti $\pm\Delta$ i limiti del campo di variabilità:

$$f_t(x|\Delta) = \begin{cases} \frac{1}{\Delta} + \frac{1}{\Delta^2}x & se \ -\Delta \leq x \leq 0 \\ \frac{1}{\Delta} - \frac{1}{\Delta^2}x & se\ 0 \leq x \leq \Delta \\ 0 & altrove \end{cases}$$

$$F_t(x|\Delta) = \begin{cases} \frac{1}{2\Delta^2}(x+\Delta)^2 & se \ -\Delta \leq x < 0 \\ 1 - \frac{1}{2\Delta^2}(\Delta - x)^2 & se\ 0 < x \leq \Delta \\ 0 & se\ x < -\Delta \\ 1 & se\ x > \Delta \end{cases}$$

$$\mu = 0$$

$$\sigma^2 = \frac{\Delta^2}{6}$$

UNIVERSITÀ DEGLI STUDI DI TRIESTE

Dipartimento di
Ingegneria
e Architettura