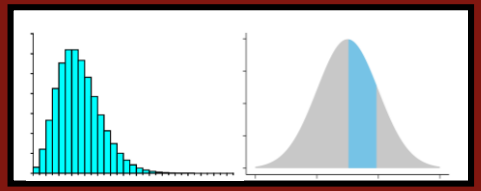
UNIVERSITÀ DEGLI STUDI DI TRIESTE

Dipartimento di Ingegneria e Architettura

Corso di misure meccaniche, termiche e collaudi

Prof. Lucia Parussini
Prof. Ronelly De Souza
Prof. Rodolfo Taccani



a.a.2026-2027

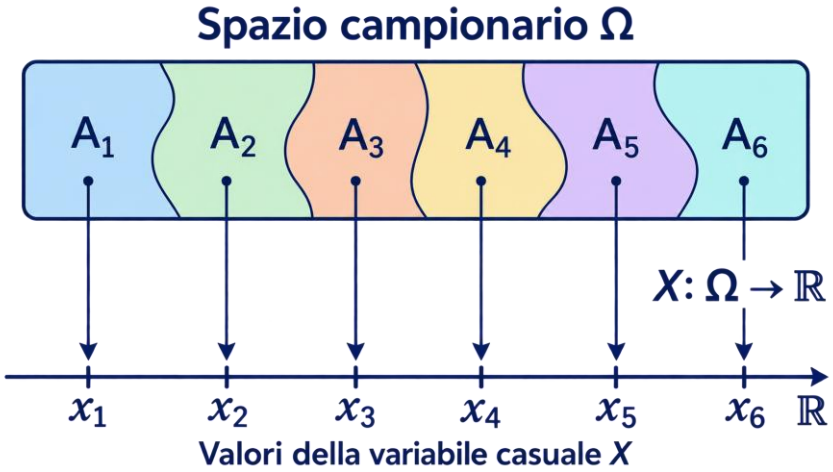
Outline

- Variabili casuali, funzione di densità di probabilità e distribuzione di probabilità
- Valore atteso, varianza e covarianza
- 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 Ω, diremo variabile casuale una applicazione X a valori reali definita sull'insieme Ω che associa ad ogni elemento di Ω 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

Distribuzione di probabilità di una variabile casuale discreta

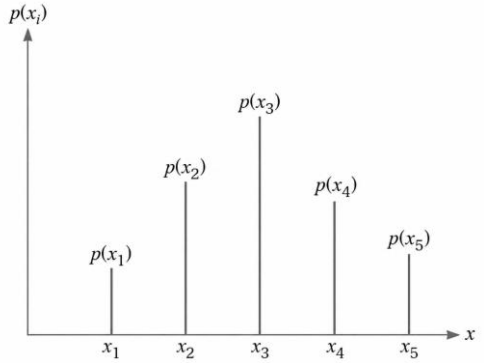
Supponiamo che X sia una variabile casuale discreta, cioè che X possa assumere valori appartenenti all'insieme di numeri reali $\{x_1, x_2, \dots, x_j, \dots\}$. Ad ognuno dei valori che X può assumere, possiamo assegnare una funzione di probabilità $p(x_i)$ tale che sia $p(x_i) \geq 0$ per ogni $i =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

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

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

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

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



Variabili casuali, funzione di densità e distribuzione di probabilità

Esempio Durante la messa a punto di un processo produttivo, si controllano due componenti meccanici. Per ciascun componente si assume che la probabilità di rispettare le tolleranze dimensionali sia 1/2 e che gli esiti dei due controlli siano indipendenti.

Indichiamo con C un componente **conforme** e con N un componente **non conforme**. Lo spazio dei possibili esiti dei due controlli è:

$$\Omega^2 = \{(C, C), (N, C), (C, N), (N, N)\}$$

Definiamo la **variabile casuale discreta** X come il numero totale di componenti conformi: C

$$X : \Omega^2 \to S \subseteq \mathbb{R}$$

I suoi possibili valori sono:

0 se nessuno dei due componenti è conforme: $X(N, N) = 0$

1 se un solo componente è conforme: $X(N, C) = X(C, N) = 1$

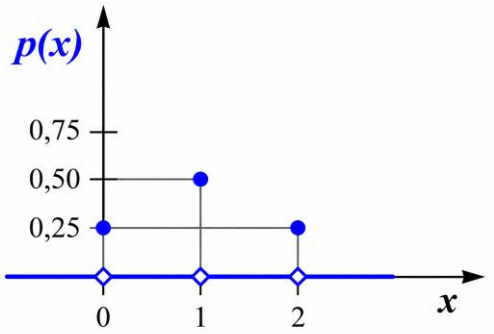
2 se entrambi sono conformi: $X(C, C) = 2$

Variabili casuali, funzione di densità e distribuzione di probabilità

Poiché i quattro esiti sono equiprobabili, ciascuno ha probabilità $1/4$.
Pertanto:
$P(X = 0) = P(N, N) = 1/4$
$P(X = 1) = P(N, C) = P(C, N) = 2/4 = 1/2$
$P(X = 2) = P(C, C) = 1/4$

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

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



Distribuzione 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.

$$p: \mathbb{R} \to [0,1] \qquad x \to \lim_{dx \to 0} \left[\frac{P(x < X \leq x + dx)}{dx} \right]$$

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

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

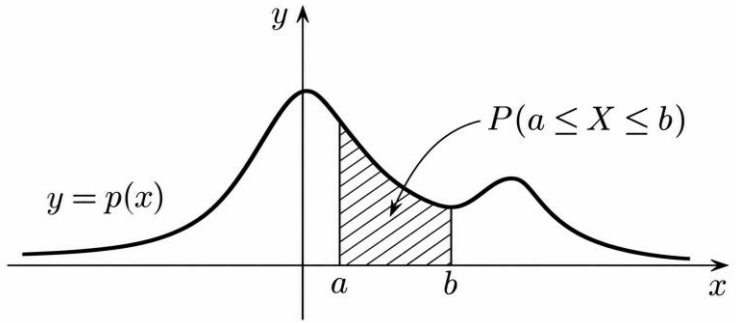
$F_X(x)$ **funzione di ripartizione** (o **funzione cumulativa**)

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

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

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

Esempio Registrazione di un evento durante una prova su un circuito idraulico

Durante una prova, una valvola viene azionata una sola volta, in un istante scelto casualmente con distribuzione uniforme nei primi 120 secondi. Il sistema di acquisizione della pressione è temporaneamente inattivo tra il secondo 45 e il secondo 60.

Qual è la probabilità che l'azionamento avvenga mentre il sistema di acquisizione è inattivo?

Indichiamo con X, variabile casuale continua, l'istante dell'azionamento, espresso in secondi dall'inizio della prova. La densità è costante nell'intervallo $[0,120]$ s e nulla al di fuori.

Variabili casuali, funzione di densità e distribuzione di probabilità

Quanto vale la costante c ?

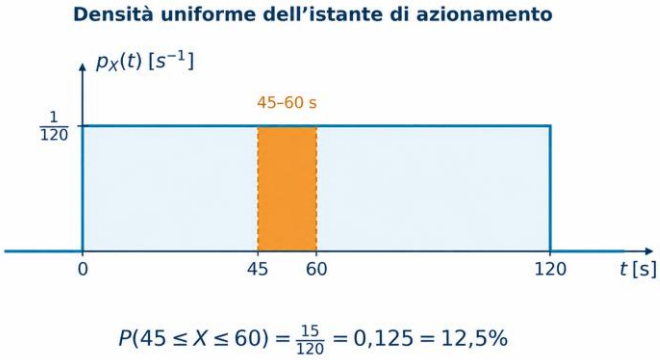
$$\int_{-\infty}^{+\infty} p_X(t)dt = 1$$

L'area totale sotto la densità deve essere uguale a 1, ovvero l'area del rettangolo con base $[0,120]$ e altezza c. Dunque $120\ c = 1$ e perciò abbiamo $c = \frac{1}{120}\, s^{-1}$.

La funzione di densità della variabile X, sarà:

$$p_X(t) = \begin{cases} 0 & per\ t < 0 \\ 1/120 & per\ 0 \le t \le 120 \\ 0 & per\ t > 120 \end{cases}$$

$$P(45 \le X \le 60) = \int_{45}^{60} p_X(t)dt = 15/120 = 0.125$$



Valore atteso

Consideriamo una variabile casuale X con una distribuzione discreta caratterizzata da un certo numero (finito o infinito) di valori x_i, dove $i = 1, \dots, j, \dots$. 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

- $Y = aX + b \Rightarrow E(Y) = E(aX + b) = aE(X) + b$

- a t.c. $P(X \geq a) = 1 \;\Rightarrow\; E(X) \geq a$
- b t.c. $P(X \leq b) = 1 \;\Rightarrow\; E(X) \leq b$
- a, b t.c. $P(a \leq X \leq b) = 1 \;\Rightarrow\; a \leq E(X) \leq b$

- 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, $E(X_A + X_B + \cdots) = E(X_A) + E(X_B) + \cdots$

- 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

- 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

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

Covarianza

Siano X_A e X_B due variabili casuali.

Definiamo covarianza delle due variabili casuali:

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

Quindi:

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

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

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

Covarianza

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.

Per due variabili casuali indipendenti, abbiamo

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

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

Ne segue

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

Distribuzione binomiale

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

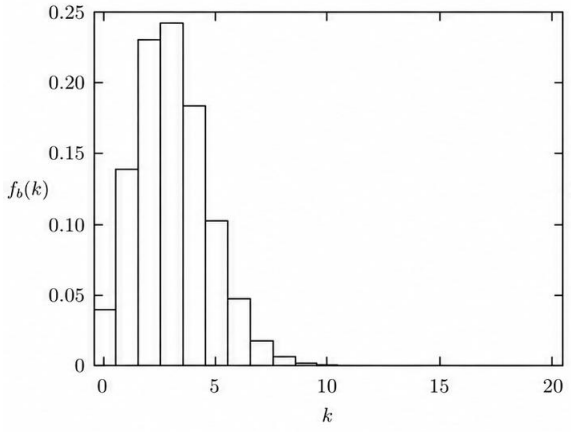
probabilità favorevole all'evento p
probabilità sfavorevole all'evento $q = 1 - p$

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

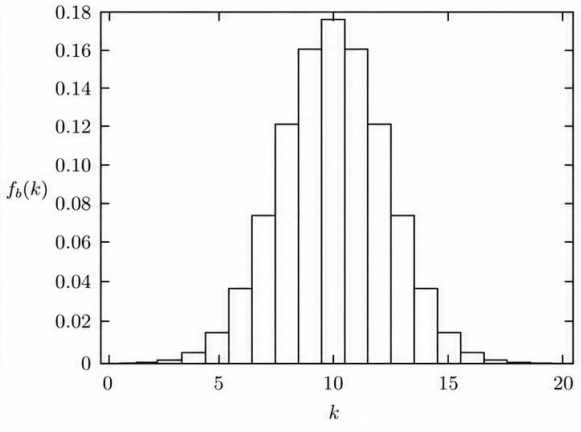
$$\mu = np = \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k}$$

$$\sigma^2 = npq$$

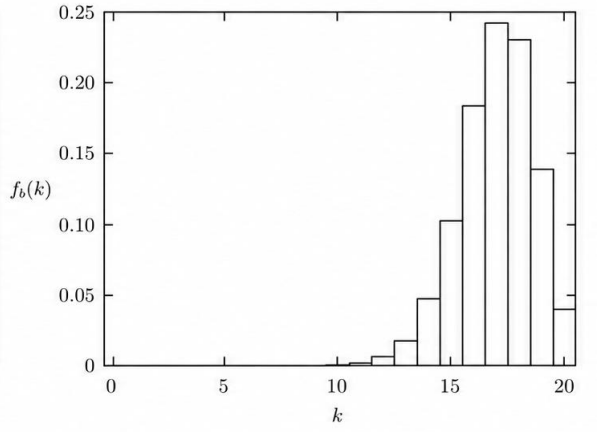
Distribuzione binomiale



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 Controllo di qualità

Un ingegnere controlla un lotto di 100 componenti prodotti da una fabbrica.
Il processo produttivo genera componenti difettosi con probabilità $p = 0.02$ (2%). Si assume che i difetti dei diversi componenti siano **indipendenti**.
Si selezionano 10 componenti a caso per il test.

Il piano di controllo stabilisce un **numero di accettazione** $c = 1$, cioè il massimo numero di componenti difettosi ammesso nel campione:

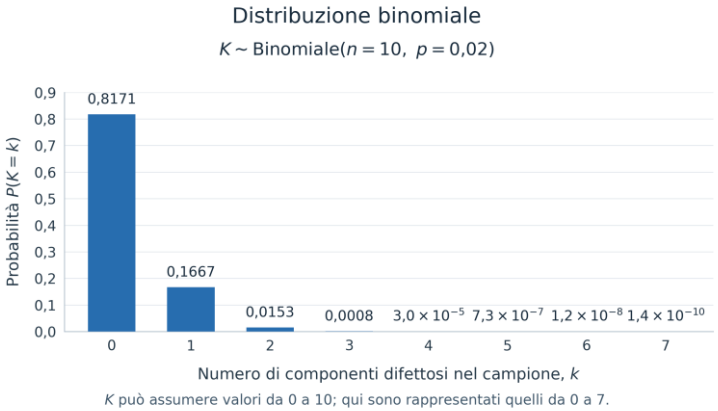
- il lotto viene **accettato** se si trovano zero oppure un componente difettoso;
- il lotto viene **rifiutato** se si trovano due o più componenti difettosi.

Indicando con K il numero di componenti difettosi nel campione, qual è la probabilità che il lotto venga accettato?

Distribuzione binomiale

Variabile casuale binomiale K numero di componenti difettosi nel campione di 10.

$$K \sim Binomiale(n = 10, p = 0.02)$$



$$P(K = 0) = 0.817 \ (81.7\%)$$
$$P(K = 1) = 0.167 \ (16.6\%)$$
$$P(K = 2) = 0.015 \ (1.5\%)$$
$$P(K \geq 3) = 0.001 \ (0.1\%)$$

$$P(accettazione) = P(K \leq 1) = P(K = 0) + P(K = 1) \approx 0.9838 = 98.4\%$$

La probabilità di **rifiuto** è quindi circa 1.6%

Distribuzione binomiale

Per **scegliere il piano di controllo da applicare quando la qualità del lotto è sconosciuta**, si deve valutare come si comporterebbe il controllo per diversi livelli di qualità.
Stabilisci quali livelli di difettosità consideri accettabili o inaccettabili. Poi scegli la dimensione del campione n e il numero di accettazione c, valutando i rischi α e β.

Ci sono due tipi principali di probabilità che interessano in questo contesto:
1. Probabilità di accettare un lotto buono Ti dice quanto è probabile che **un lotto conforme venga accettato**. Si vuole che questa probabilità sia **alta** (basso rischio del produttore, chiamato α **-risk**).
2. Probabilità di accettare un lotto cattivo Ti dice quanto è probabile che **un lotto difettoso venga comunque accettato**. Si vuole che questa probabilità sia **bassa** (basso rischio del cliente, chiamato β**-risk**).

$n = 10, p = 0.02 \quad P(accettazione) = P(K \leq 1) = 98.4\%$ rischio per il produttore molto basso
$n = 10, p = 0.10 \quad P(accettazione) = P(K \leq 1) = 73.6\%$ rischio per il cliente alto
$n = 20, p = 0.10 \quad P(accettazione) = P(K \leq 1) = 39.2\%$ rischio per il cliente un po' più basso

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.

Si considera un insieme A di N elementi, di cui D appartengono al sottoinsieme B. Si estraggono n elementi distinti, con tutti i possibili campioni di n elementi equiprobabili. La variabile aleatoria K conta quanti elementi del campione appartengono a B.

$\binom{N}{D}$ è il numero di possibili estrazioni di D elementi tra gli N di A

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

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

Distribuzione ipergeometrica

$$f_{ip}(k|N,n,D) = \begin{cases} \dfrac{\binom{n}{k}\binom{N-n}{D-k}}{\binom{N}{D}} & per\ \ k = 0,\dots,n \\ 0 & altrimenti \end{cases}$$

$$\mu = \frac{Dn}{N} \qquad \sigma^2 = \frac{D(N-D)n(N-n)}{N^2(N-1)}$$

Distribuzione ipergeometrica

Esempio Controllo di qualità

Un ingegnere controlla un lotto di 100 componenti. Per valutare il piano di controllo, si considera il caso in cui il lotto contenga esattamente 2 componenti difettosi.
Si selezionano 10 componenti a caso, senza reinserimento, per il test.

Il piano di controllo stabilisce un **numero di accettazione** $c = 1$:

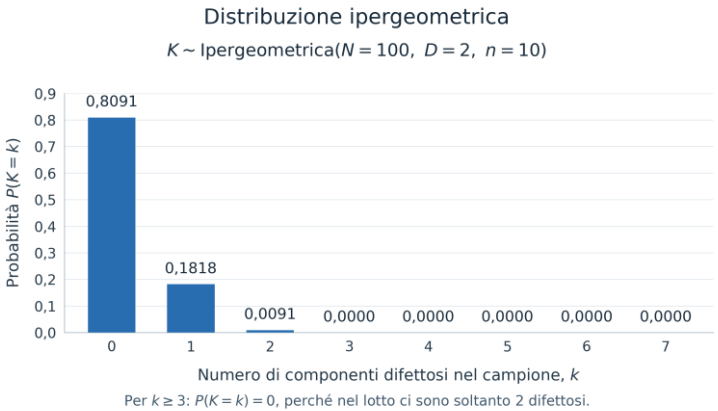
- il lotto viene **accettato** se nel campione si trovano **zero oppure un componente difettoso;**
- il lotto viene **rifiutato** se nel campione si trovano **entrambi i componenti difettosi.**

Indicando con K il numero di componenti difettosi nel campione, **qual è la probabilità che il lotto venga accettato?**

Distribuzione ipergeometrica

Variabile casuale ipergeometrica K numero di componenti difettosi nel campione di 10.

$$K \sim Ipergeometrica(N = 100, D = 2, n = 10)$$



$$P(K = 0) = 0.809 \ \ (0.817)$$
$$P(K = 1) = 0.182 \ \ (0.167)$$
$$P(K = 2) = 0.009 \ \ (0.015)$$
$$P(K \geq 3) = 0.000 \ \ (0.001)$$

Se il lotto è molto grande rispetto al campione ($N \gg n$), la differenza tra binomiale e ipergeometrica diventa trascurabile.

$$P(accettazione) = P(K \leq 1) = P(K = 0) + P(K = 1) \approx 0.9909 = 99.1\%$$

La probabilità di **rifiuto** è quindi circa 0.9%

Confronto distribuzione binomiale e distribuzione ipergeometrica

Binomiale o ipergeometrica: quando usarle?

Entrambe descrivono il **numero di componenti difettosi nel campione**, ma con ipotesi diverse.

Binomiale	Ipergeometrica
Si modellano **prove indipendenti**, con probabilità di difetto p costante.	Si modella un'estrazione **senza reinserimento** da un lotto finito con un numero fissato D di difettosi.
L'esito di una prova **non modifica** la probabilità delle successive.	Ogni estrazione modifica la composizione del lotto rimanente: le estrazioni sono **dipendenti**.
Nell'esempio, $p = 2\%$: il numero totale di difettosi nei 100 componenti può variare.	Nell'esempio, il lotto contiene **esattamente 2 difettosi su 100**.

Perché nell'ipergeometrica cambia la probabilità? La prima estrazione ha probabilità di difetto 2/100. Alla seconda: se è stato estratto un difettoso, la probabilità diventa **1/99**; se è stato estratto un componente efficiente, diventa **2/99**.

Collegamento tra i modelli: se il campione è molto piccolo rispetto al lotto, l'ipergeometrica può essere approssimata con una binomiale ponendo $p = D/N$.

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 è una approssimazione della distribuzione binomiale e 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

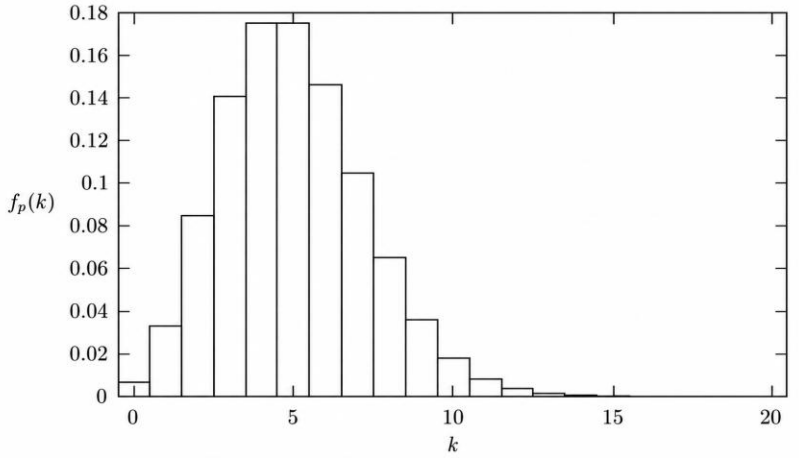
Densità di probabilità di Poisson:

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

dove $\lambda = np > 0$.

$\mu = \lambda$

$\sigma^2 = \lambda$



Densità di probabilità di Poisson con λ=5.

Distribuzione di Poisson

Esempio Una macchina produce pezzi con probabilità di difetto $p = 0.002$ (0.2%). Si assumono difetti indipendenti e probabilità di difetto costante. L'azienda vuole fornire **almeno 1000 pezzi efficienti per confezione.**

Quanti pezzi aggiuntivi occorre inserire affinché la probabilità di avere meno di 1000 pezzi efficienti sia inferiore allo 0.01%?

Indicando con s il numero di pezzi aggiuntivi, ogni confezione contiene: $n = 1000 + s$

Il numero K di pezzi difettosi segue una distribuzione binomiale.

Poiché n è grande e p è piccola, possiamo approssimarla con una distribuzione di Poisson di parametro: $\lambda = np = (1000 + s)0.002$

Distribuzione di Poisson

Per la soluzione, i pezzi efficienti sono $1000 + s - K$

La confezione è insufficiente quando $K > s$ ovvero il numero di pezzi difettosi è maggiore di quelli aggiunti.

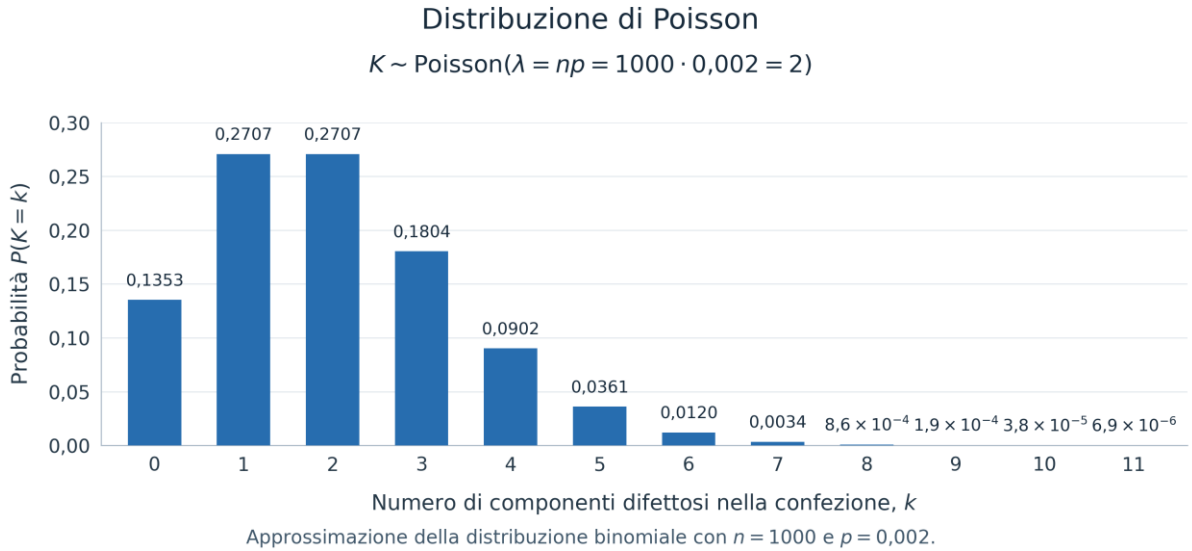
Dobbiamo quindi trovare il minimo s per cui

$$P(K > s) < 0.0001$$

Usiamo la distribuzione di Poisson con il parametro aggiornato al variare di s.

Distribuzione di Poisson

$s = 0, \lambda = np = 2$



$$P(K > 0) = 1 - 0.135 = 0.865 > 0.0001$$

Distribuzione di Poisson

k	$n = 1001$ λ=2.002	$n = 1002$ λ=2.004	$n = 1007$ λ=2.014	$n = 1008$ λ=2.016	$n = 1009$ λ=2.018
0	13.506488	13.479502	13.345379	13.318715	13.292104
1	27.039990	27.012923	26.877593	26.850529	26.823466
2	27.067030	27.066949	27.065737	27.065334	27.064877
3	18.062731	18.080722	18.170131	18.187904	18.205641
4	9.040397	9.058442	9.148661	9.166704	9.184746
5	3.619775	3.630623	3.685081	3.696015	3.706963
6	1.207798	1.212628	1.236959	1.241861	1.246775
7	0.345430	0.347158	0.355891	0.357656	0.359428
8	0.086444	0.086963	0.089595	0.090129	0.090666
9	0.019229	0.019364	0.020049	0.020189	0.020329
10	0.003850	0.003881	0.004038	0.004070	0.004102
11	0.000701	0.000707	0.000739	0.000746	0.000753

La tabella riporta $P(K = k)$ **in percentuale**.

Distribuzione di Poisson

Pezzi totali n	Pezzi aggiunti s	Evento da considerare	Probabilità di meno di 1000 pezzi efficienti
1001	1	$K > 1$	59.454%
1002	2	$K > 2$	32.441%
1007	7	$K > 7$	0.115%
1008	8	$K > 8$	0.025%
1009	**9**	$K > 9$	**0.005%**

Servono quindi almeno 9 pezzi aggiuntivi: usando l'approssimazione di Poisson, la probabilità scende sotto lo **0.01%** richiesto.
Il calcolo binomiale esatto conferma il risultato: con 9 pezzi aggiuntivi, la probabilità di carenza è circa **0.0049%**, inferiore allo **0.01%** richiesto.

Distribuzione gaussiana

La distribuzione normale: un modello fondamentale

- **Descrive molti fenomeni reali,** dalle grandezze naturali alle caratteristiche dei prodotti industriali.
- **La sua diffusione ha una spiegazione:** il Teorema del Limite Centrale giustifica, sotto opportune condizioni, l'approssimazione normale di somme e medie di molti contributi indipendenti.
- **È particolarmente utile nelle misure ripetute:** numerosi piccoli effetti casuali si sommano e possono produrre una distribuzione a campana, con valori concentrati attorno alla media.
- **È un riferimento per l'analisi statistica:** consente di calcolare probabilità, applicare numerosi metodi di analisi e simulare la variabilità dei processi.

Anche alcune distribuzioni asimmetriche possono diventare **approssimativamente normali dopo una trasformazione,** ad esempio $W = \log(X)$.

Distribuzione gaussiana

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

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

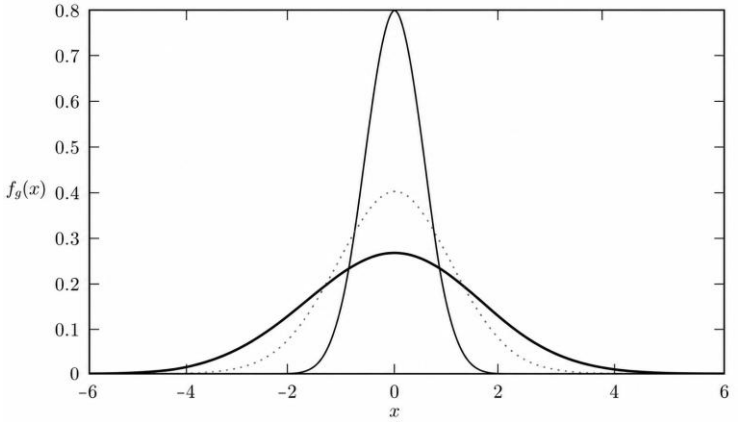
Si può sempre standardizzare la variabile normale: $z = \frac{x-\mu}{\sigma}$

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

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

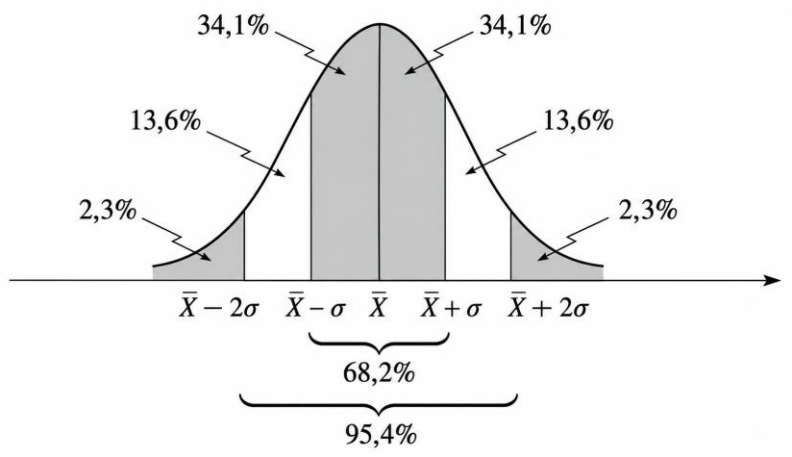
Indichiamo con $F_g(z)$ la funzione di ripartizione della densità normale standard

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



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



Distribuzione gaussiana

Esempio Tolleranza dimensionale di un componente meccanico

Un ingegnere controlla la **lunghezza di perni metallici** prodotti da una macchina CNC.

Il disegno tecnico richiede: $L = 50.00 \pm 0.10\ mm$ quindi il pezzo è accettabile se:

$$49.90\ mm \leq L \leq 50.10\ mm$$

Dalle misure sperimentali risulta che la lunghezza dei perni segue una **distribuzione normale**:

$$L \sim N(\mu = 50.02, \sigma = 0.04)\ mm$$

Distribuzione gaussiana

Calcoliamo i limiti in unità standard (z-score) $z = (x - \mu)/\sigma$ con $\mu = 50.02, \sigma = 0.04$

per il limite inferiore $(49.90\ mm)$ $z_1 = -3.0$
per il limite superiore $(50.10\ mm)$ $z_2 = 2.0$

La probabilità cumulativa è:

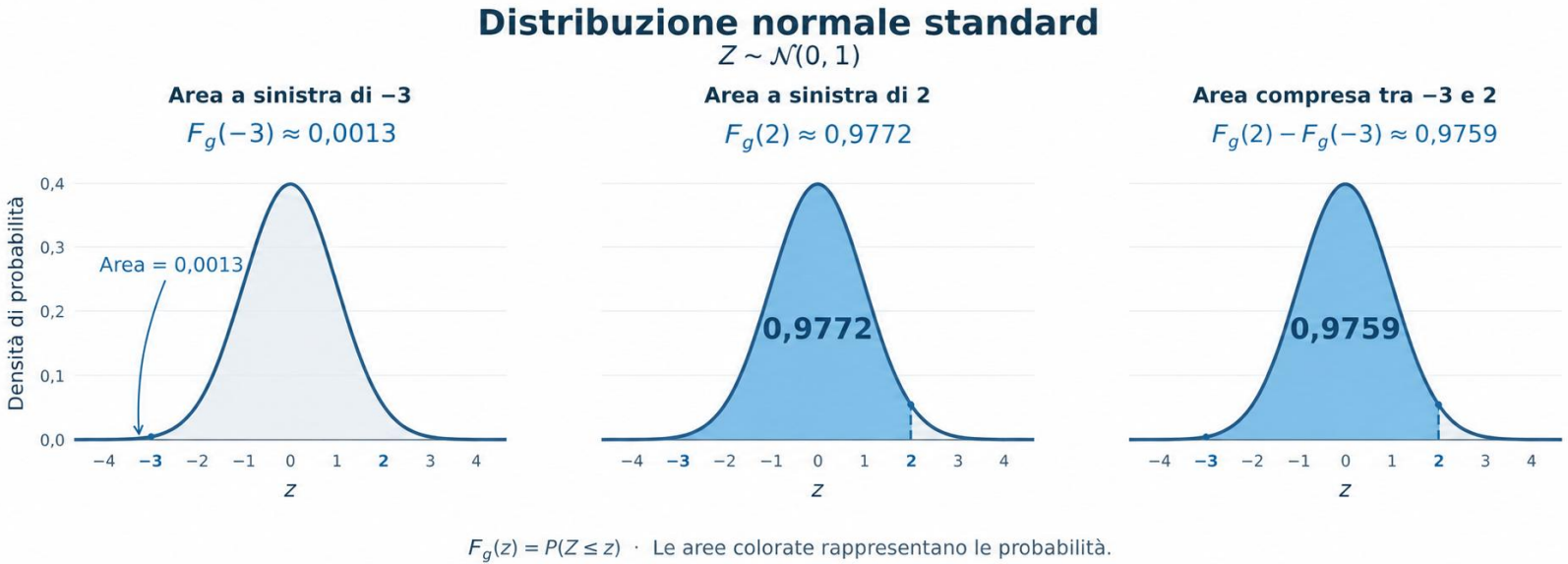
$$P(Z \leq z_1) = F_g(-3.0) = 0.0013 \qquad P(Z \leq z_2) = F_g(2.0) = 0.9772$$

quindi

$$P(z_1 \leq Z \leq z_2) = F_g(2.0) - F_g(-3.0) = 0.9759$$

La probabilità che $49.90\ mm \leq L \leq 50.10\ mm$ è 0.9759.

Distribuzione gaussiana



✓ 97,6% dei pezzi conforme

✗ 2,4% dei pezzi fuori tolleranza

Distribuzione Chi-quadro χ^2 (di Pearson)

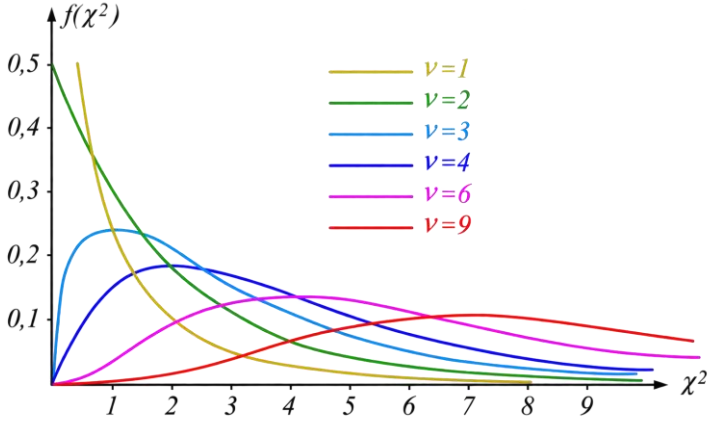
Si considerano ν variabili aleatorie indipendenti $Y_1, Y_2, \dots, Y_\nu$ distribuite secondo la legge normale con media $\mu = 0$ e varianza $\sigma^2 = 1$; allora la variabile χ^2 è:

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

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

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

$\mu = \nu \qquad \sigma^2 = 2\nu$



Distribuzione Chi-quadro χ^2 (di Pearson)

Proprietà

- asimmetria.
- dipendenza dal parametro intero ν che indica i gradi di libertà
- la variabile χ^2 non può assumere valori negativi dato che è una somma di quadrati.
- è completamente definita nel primo quadrante.

Esempio Sia X una variabile aleatoria con media $E(X) = \mu$ e varianza $\mathrm{Var}(X) = \sigma^2$ 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 varianza campionaria ha una distribuzione chi-quadro.

Esempi di applicazioni test di ipotesi su Goodness of fit, Independence

Distribuzione t-Student

Consideriamo due variabili casuali indipendenti: $Z \sim N(0,1)$ e $Y \sim \chi^2(\nu)$:

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

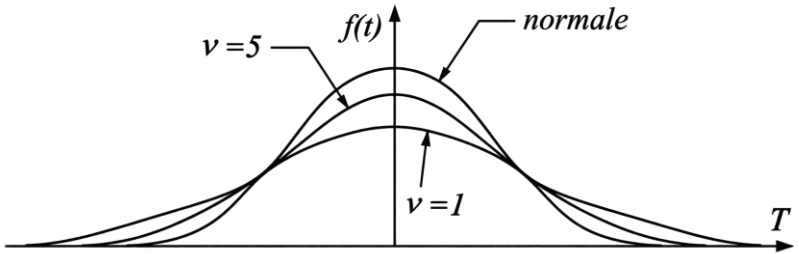
ha una distribuzione t-Student con ν gradi di libertà

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

dove B è la funzione beta

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

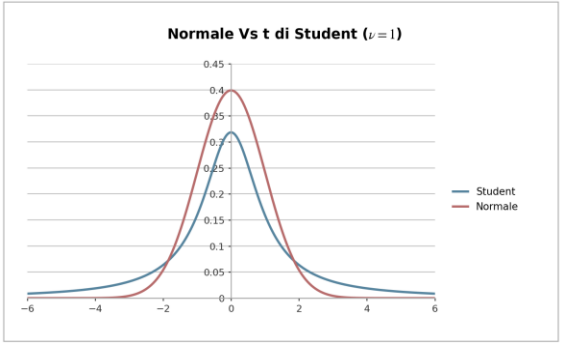
$$\mu = 0 \qquad \sigma^2 = \frac{\nu}{\nu - 2}, \nu > 2$$

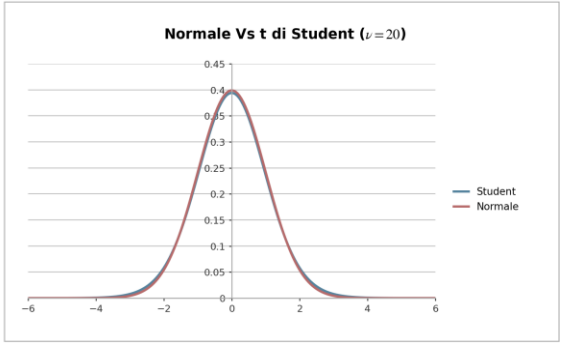


Distribuzione t-Student

Proprietà

- 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)





Esempi di applicazioni test di ipotesi su media di una popolazione di cui non è conosciuta la deviazione standard della popolazione, differenza tra due medie, differenza tra due medie di 2 campioni dipendenti, coefficiente di correlazione

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 $F(\nu_1, \nu_2)$:

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

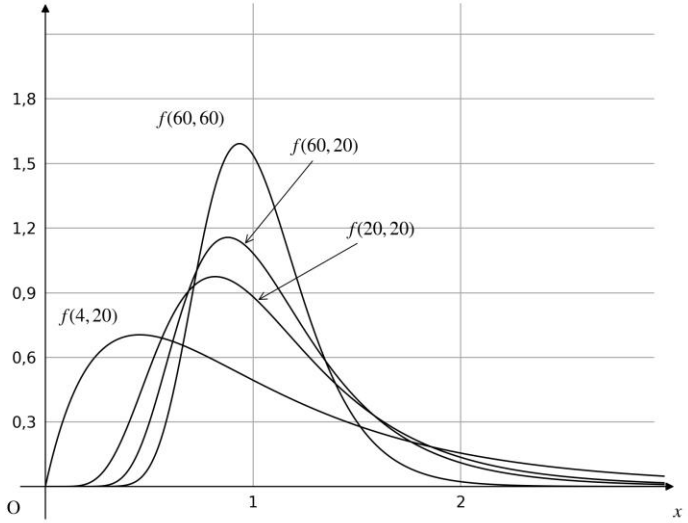
La sua funzione di densità è $f(x) = 0$ se $x \leq 0$ e, se $x > 0$, è

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

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



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

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

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

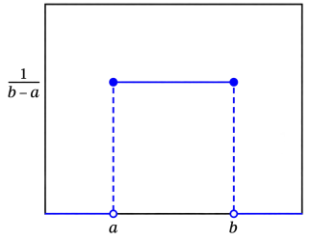
La distribuzione di Fischer descrive l'andamento del rapporto tra le varianze delle due distribuzioni in funzione dei loro gradi di libertà ed è principalmente utilizzata per verificare la differenza statistica tra le due varianze.

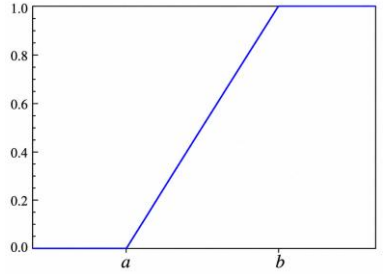
Distribuzione rettangolare o distribuzione continua uniforme

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 $x > b$

$$\mu = (a+b)/2 \qquad \sigma^2 = (b-a)^2/12 \qquad \sigma = (b-a)/(2\sqrt{3})$$

Si può standardizzare la variabile rettangolare: $z = (x-\mu)/\sigma$.

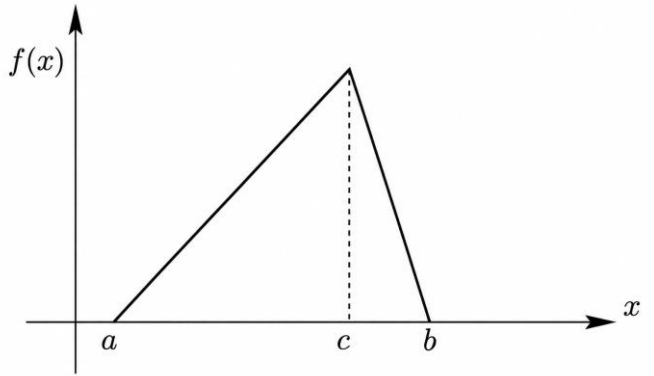
La distribuzione rettangolare standard ha $\mu = 0$ e $\sigma = 1$, è quindi definita nell'intervallo $[-\sqrt{3}, \sqrt{3}]$ con $f_r(z|-\sqrt{3},\sqrt{3}) = 1/(2\sqrt{3})$ su $[-\sqrt{3},\sqrt{3}]$, $f_r(z|-\sqrt{3},\sqrt{3}) = 0$ altrove.

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:

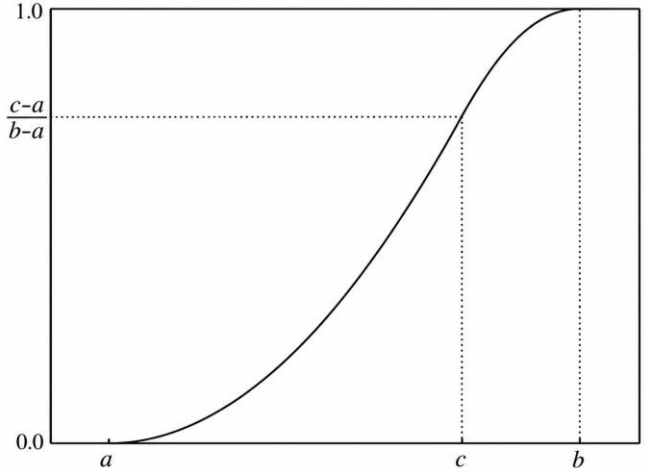
$$f_t(x|a,b,c) = \begin{cases} \frac{2}{b-a}\frac{x-a}{c-a} & se\ a \le x \le c \\ \frac{2}{b-a}\frac{b-x}{b-c} & se\ c \le x \le b \\ 0 & altrove \end{cases}$$



Distribuzione triangolare

La funzione di ripartizione è:

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



$$\mu = \frac{a+b+c}{3} \qquad \sigma^2 = \frac{a^2+b^2+c^2-(ab+ac+bc)}{18}$$

Distribuzione triangolare

Se la distribuzione di probabilità triangolare è **simmetrica** rispetto a c, si ha $c = (a+b)/2$ e quindi

$$\mu = \frac{a+b}{2}$$

$$\sigma = \frac{b-a}{2\sqrt{6}}$$

Si può standardizzare la variabile triangolare simmetrica: $z = (x-\mu)/\sigma$.

La distribuzione triangolare standard ha $\mu = 0$ e $\sigma = 1$, è quindi definita nell'intervallo $\left[-\sqrt{6}, \sqrt{6}\right]$ con

$$f_t\left(z\middle|\sqrt{6}\right) = \begin{cases} \frac{1}{\sqrt{6}} + \frac{1}{6}z & se\ -\sqrt{6} \le z \le 0 \\ \frac{1}{\sqrt{6}} - \frac{1}{6}z & se\ 0 \le z \le \sqrt{6} \\ 0 & altrove \end{cases}$$

Distribuzioni nella pratica

Le funzioni MATLAB richiedono **Statistics and Machine Learning Toolbox**; in Python la liberia **SciPy**.

Distribuzione	Ambiente	Probabilità / densità	Cumulata	Inversa della cumulata
Binomiale	MATLAB	binopdf(x,n,p)	binocdf(x,n,p)	binoinv(q,n,p)
	Python	st.binom.pmf(x,n,p)	st.binom.cdf(x,n,p)	st.binom.ppf(q,n,p)
Ipergeometrica	MATLAB	hygepdf(x,N,D,n)	hygecdf(x,N,D,n)	hygeinv(q,N,D,n)
	Python	st.hypergeom.pmf(x,N,D,n)	st.hypergeom.cdf(x,N,D,n)	st.hypergeom.ppf(q,N,D,n)
Poisson	MATLAB	poisspdf(x,lam)	poisscdf(x,lam)	poissinv(q,lam)
	Python	st.poisson.pmf(x,lam)	st.poisson.cdf(x,lam)	st.poisson.ppf(q,lam)

Distribuzioni nella pratica

Distribuzione	Ambiente	Probabilità / densità	Cumulata	Inversa della cumulata
Gaussiana	MATLAB	normpdf(x,mu,sigma)	normcdf(x,mu,sigma)	norminv(q,mu,sigma)
	Python	st.norm.pdf(x,mu,sigma)	st.norm.cdf(x,mu,sigma)	st.norm.ppf(q,mu,sigma)
Chi quadrato	MATLAB	chi2pdf(x,nu)	chi2cdf(x,nu)	chi2inv(q,nu)
	Python	st.chi2.pdf(x,nu)	st.chi2.cdf(x,nu)	st.chi2.ppf(q,nu)
t di Student	MATLAB	tpdf(x,nu)	tcdf(x,nu)	tinv(q,nu)
	Python	st.t.pdf(x,nu)	st.t.cdf(x,nu)	st.t.ppf(q,nu)
F di Fisher	MATLAB	fpdf(x,nu1,nu2)	fcdf(x,nu1,nu2)	finv(q,nu1,nu2)
	Python	st.f.pdf(x,nu1,nu2)	st.f.cdf(x,nu1,nu2)	st.f.ppf(q,nu1,nu2)
Triangolare	MATLAB	pdf(triM,x)	cdf(triM,x)	icdf(triM,q)
	Python	triP.pdf(x)	triP.cdf(x)	triP.ppf(q)
Rettangolare	MATLAB	unifpdf(x,a,b)	unifcdf(x,a,b)	unifinv(q,a,b)
	Python	st.uniform.pdf(x,a,b-a)	st.uniform.cdf(x,a,b-a)	st.uniform.ppf(q,a,b-a)

UNIVERSITÀ
DEGLI STUDI
DI TRIESTE

Dipartimento di
Ingegneria
e Architettura