Elementi di calcolo delle probabilità

definizione classica di probabilità (Laplace)

Nel caso di eventi *equiprobabili* la probabilità che un evento si realizzi è pari al rapporto fra il numero di casi favorevoli al verificarsi dell'evento e il numero totale di casi possibili:

$$P = \frac{\#\text{ casi favorevoli}}{\#\text{ casi possibili}}$$

è evidente come sia $0 \le P \le 1$ (P=1 corrisponde alla certezza)

ESEMPIO: consideriamo il lancio di 3 monete e chiediamoci quale sia la probabilità di ottenere 2 e solo 2 facce uguali.

Indichiamo con T testa e C croce. I possibili risultati sono in tutto 8:

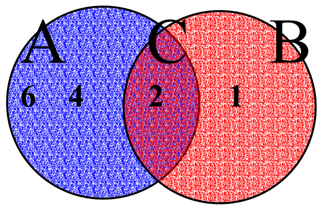
TTT TTC TCT TCC CTT CTC CCT CCC

Ci sono 6 casi favorevoli; la probabilità dell'evento è pari a 6/8 = ¾ = 75%

Se due eventi (A e B) hanno elementi in comune (C)
(p.es. pari e minore di 3 nel lancio di un dado)

P(A o B) = P(A) + P(B) –P(C)

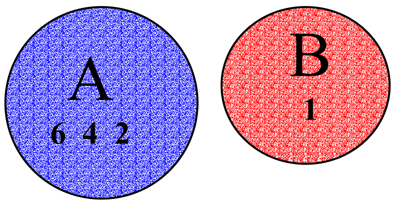
$$\frac{1}{2}+\frac{1}{3}-\frac{1}{6}=\frac{3+2-1}{6}=\frac{4}{6}$$



Se invece P(C)=0 (eventi incompatibili)
(p.es. pari e minore di 2)

P(A o B) = P(A) + P(B)

$$\frac{1}{2}+\frac{1}{6}=\frac{3+1}{6}=\frac{4}{6}$$

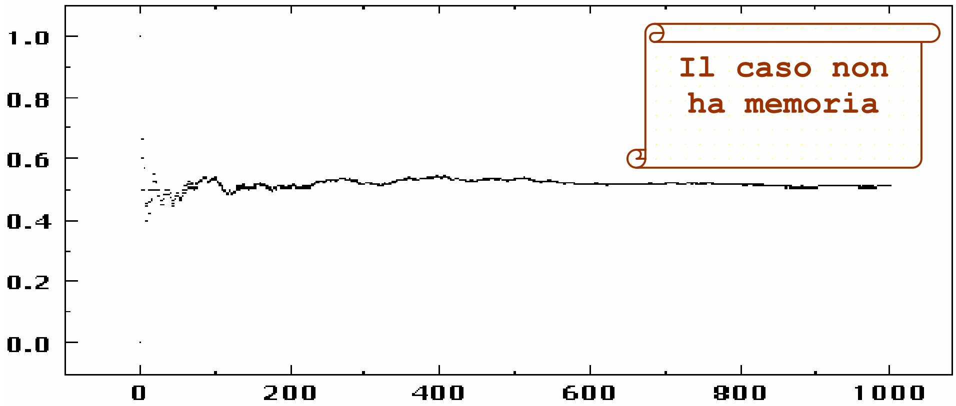


definizione frequentistica di probabilità (Venn)

lancio di una moneta

ordinate:
frequenza relativa: rapporto fra numero di "testa" (frequenza) e numero di lanci

ascisse:
prove: numero di lanci



Nei primissimi lanci la frequenza relativa varia notevolmente perché per alcuni lanci consecutivi può uscire sempre testa o sempre croce (fluttuazione statistica).
Dopo circa una cinquantina di lanci inizia tendenza verso il valore 50% e l'ampiezza delle fluttuazioni diminuisce (è rara una lunga sequenza di sole teste o sole croci).
Intuitivamente: per $N \rightarrow \infty$ la frequenza relativa tende alla probabilità di "testa" ma la natura casuale del fenomeno non impedisce che da un certo lancio in poi si ottenga una sequenza di molte teste o molte croci: indipendentemente dalla storia del fenomeno, ad ogni lancio la probabilità di testa è sempre la stessa.

Approssimativamente**: legge empirica del caso** (legge debole dei grandi numeri): quando il numero delle prove è molto grande la frequenza "praticamente" coincide con la probabilità:

$$\lim_{N \to \infty} \frac{k}{N} = p$$

(se N è grande $\frac{k}{N} \approx p$)

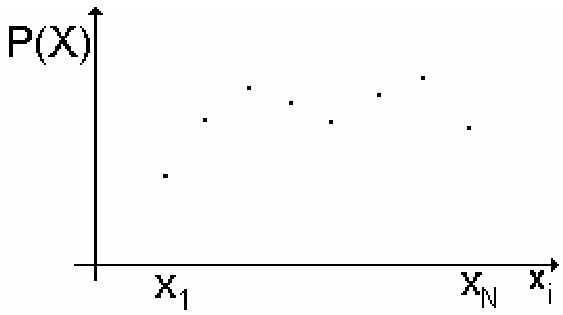
Più rigorosamente: **teorema di Bernoulli** (legge forte dei grandi numeri):
- se un evento E ha probabilità p di verificarsi
- se si eseguono N prove indipendenti e si ottengono k successi
- scelto un $\varepsilon > 0$ piccolo a piacere

$$\lim_{N \to \infty} P\left(\left| p - \frac{k}{N} \right| < \varepsilon \right) = 1$$

Il teorema afferma che al crescere di N la frequenza relativa k/N non tende alla probabilità p ma che ciò diventa probabilisticamente sempre più certo.

Variabili aleatorie

Si definisce variabile aleatoria X la variabile che assume i valori per cui è definita: $X = \{x_1, x_2, ..., x_N\}$ non in base a leggi deterministiche ma in base al caso secondo la probabilità $P(X=x_i)$.



Variabili aleatorie discrete

La funzione P(X) definita solo per i valori $x_1, x_2, ..., x_N$ viene detta distribuzione di probabilità della variabile aleatoria discreta X.

Essa è una probabilità e quindi:
- per ogni x_i risulta $1 \geq P(x_i) \geq 0$
- risulta $P(x_1)+P(x_2)+...+P(x_N) = \sum_{i=1,N} P(x_i) = 1$ (proprietà di chiusura)

Infatti, gli N valori x_i costituiscono un insieme completo (sono N in tutto) di valori e quindi è certo (P=1) che si verifichi almeno uno di essi.
Inoltre gli N valori sono incompatibili (se $X = x_i$ allora $X \neq x_j$) e quindi

$$P(x_1 \text{ o } x_2 \text{ o } ... \; x_N) = P(x_1) + P(x_2) + ... + P(x_N)$$

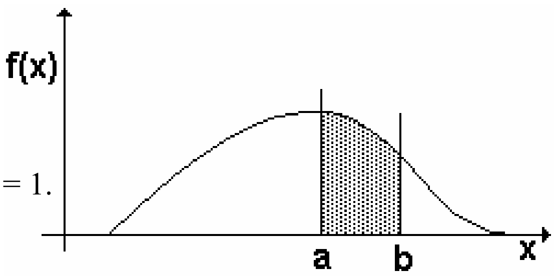
Variabili aleatorie continue

Se la v.a. X può assumere tutti i valori compresi in un intervallo allora è detta v.a. continua. In questo caso si può definire la probabilità infinitesima dP(x) che la v.a. X assuma un valore compreso fra x e x+dx: $P(x \leq X < x+dx) = dP(x) = f(x)\,dx$

Si definisce la $f(x) = \frac{dP(x)}{dx}$ densità di probabilità o funzione di distribuzione

La probabilità che X assuma un valore compreso fra **a** e **b** è quindi: $P(a \leq X \leq b) = \int_a^b f(x)\,dx$

Per la proprietà di chiusura $\int_{-\infty}^{+\infty} f(x)\,dx = P(\text{qualsiasi valore } x) = 1.$



Riassunti di una distribuzione

Spesso di una distribuzione è sufficiente conoscere alcuni valori caratteristici che ne riassumono l'andamento: il valore più probabile, un valore che indichi quanto la distribuzione sia simmetrica rispetto ad un valore centrale, ecc.).

Valore medio

Si definisce valore atteso o speranza matematica o valore medio o media (da non confondersi con la media aritmetica) della v.a. X la quantità:

$$\sum_{i=1,N} x_i P(x_i) \quad \text{(v.a. discreta)} \qquad\qquad \int_{-\infty}^{+\infty} x\, f(x)\, dx \quad \text{(v.a. continua)}$$

Il significato è quello di una media effettuata pesando ogni valore della v.a. X con la probabilità che essa assuma quel valore (analogia con il baricentro in meccanica dove le distanze vengono pesate con le masse).

Simbolicamente l'operazione di media si indica con E(X) (Expectation value).

Indicheremo con m il valore atteso di una distribuzione: E(X) = m

ESEMPIO: calcolo del valor medio della distribuzione di probabilità di X:

- nel caso del lancio di un dado:

$$E(X) = \sum_{k=1,6} k \times \frac{1}{6} = \frac{1+2+3+4+5+6}{6} = 21/6 = 3,5$$

- nel caso di una distribuzione uniforme continua fra a e b

$$E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{(b^2-a^2)}{2} \frac{1}{b-a} = \frac{a+b}{2}$$

Dato che E(X) è un operatore lineare (sommatoria o integrale), si può facilmente verificare che:

E(aX+b) $= \sum (a\, x_i + b)\, P(x_i) = \sum a\, x_i\, P(x_i) + \sum b\, P(x_i) =$ **a E(X) + b** [X discreta]
E(aX+b) $= \int (ax+b)\, f(x)\, dx = a \int x\, f(x)\, dx + b \int f(x)\, dx =$ **a E(X) + b** [X continua]

Scarto

Si definisce scarto (dalla media) la v.a. X-m cioè la differenza tra il valore della variabile aleatoria e il valore medio della distribuzione.
Il suo valore atteso è sempre E(X-m) = E(X)-E(m) = E(X)-m = m-m = 0 e quindi non può essere utilizzato come riassunto.

Varianza

Si definisce varianza della v.a. X il valore atteso del quadrato dello scarto dalla media:

$$\sigma^2(X) = E\{[X - E(X)]^2\} = E[(X-m)^2]$$

essa quantifica la larghezza di un distribuzione: più i valori maggiormente probabili sono vicini alla media e più $\sigma^2(X)$ è piccola; più ne sono distanti e più $\sigma^2(X)$ è grande.
Per semplicità di calcolo può essere utile ricordare che:
$\boldsymbol{\sigma^2(X)} = E[(X-m)^2] = E[X^2 - 2\, m\, X + m^2] = E(X^2) - 2\, m\, E(X) + m^2 =$ **E(X2) - m²**

Indicheremo con σ^2 la varianza di una distribuzione: $\sigma^2(X) = \sigma^2$

ESEMPIO: calcoliamo la varianza della distribuzione di probabilità di X:

- nel caso del lancio di un dado:

$$\sigma^2(X) = E(X^2) - m^2 = \sum_{k=1,6} k^2 \times \frac{1}{6} - \left(\frac{7}{2}\right)^2 = \frac{1+4+9+16+25+36}{6} - \frac{49}{4} = 35/12$$

- nel caso di una distribuzione uniforme continua fra a e b

$$\sigma^2(X) = E(X^2) - m^2 = \int_a^b x^2 \frac{1}{b-a} dx - \left(\frac{a+b}{2}\right)^2 = \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

È facile verificare che:

$$\boldsymbol{\sigma^2(aX+b)} = E\{[aX+b-E(aX+b)]^2\} = E\{[aX+ b - a\,m - b]^2\} =$$
$$= E\{[a(X-m)]^2\} = E\{a^2(X-m)^2\} = a^2\, E\{(X-m)^2\} = \boldsymbol{a^2\, \sigma^2(X)}$$

Deviazione standard

Si definisce deviazione standard o scarto quadratico medio la quantità $\sigma = \sqrt{\sigma^2(X)}$ che ha le stesse dimensioni fisiche della variabile aleatoria X.

Distribuzione uniforme continua fra a e b:

Si utilizza nei casi in cui i valori all'interno di un intervallo sono ugualmente probabili: **f(x) = K.**

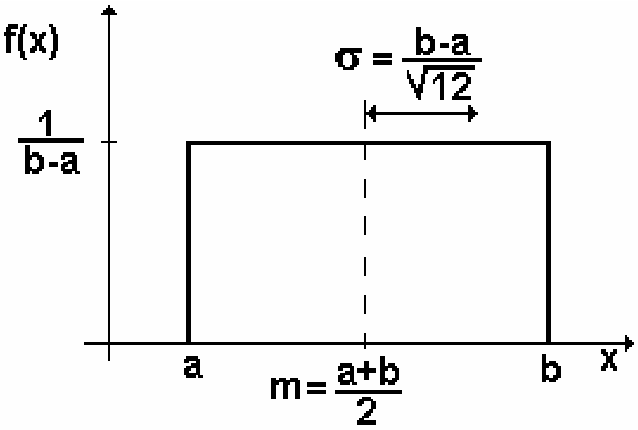
Dalla proprietà di chiusura: $\int_a^b K dx = 1$ segue $K = \frac{1}{b-a}$

$$E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}$$

$$\sigma^2(X) = \int_a^b x^2 \frac{1}{b-a} dx - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

$$\sigma(X) = \frac{b-a}{\sqrt{12}} = 0{,}29\,(b-a)$$

$$P(m-\sigma \leq X \leq m+\sigma) = \int_{\frac{a+b}{2}-\frac{b-a}{\sqrt{12}}}^{\frac{a+b}{2}+\frac{b-a}{\sqrt{12}}} \frac{1}{b-a} dx = \frac{2}{\sqrt{12}} = 57{,}7\%$$

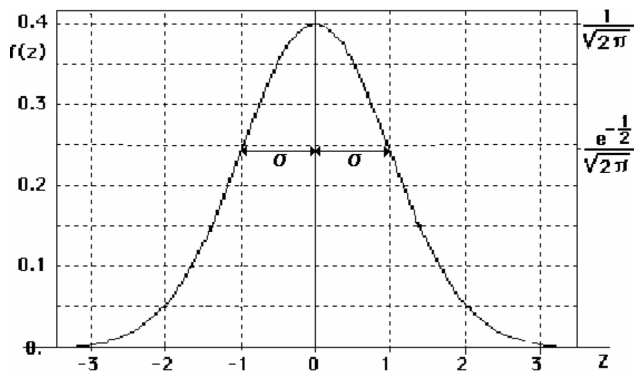


Utilizzeremo frequentemente i riassunti della distribuzione uniforme perché, in assenza di altre informazioni, è spesso ragionevole supporre a priori che non ci siano valori preferibili ad altri. Tutte le altre distribuzioni richiedono invece ipotesi particolari e conoscenze non sempre disponibili.

Distribuzione di Gauss (o normale)

è la distribuzione teorica degli errori casuali nonché della media aritmetica (norma) di un numero infinito di misure.

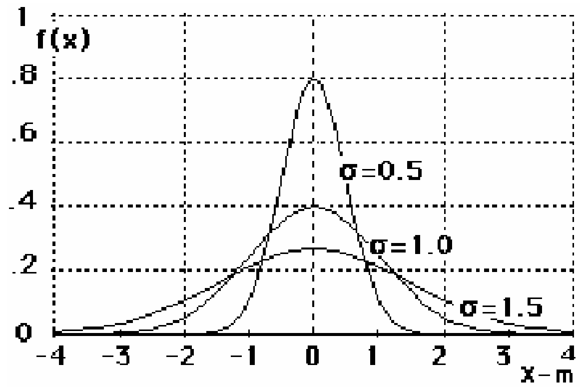
La densità di probabilità gaussiana è: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$



- La v.a. X può assumere tutti i valori compresi fra -∞ e ∞.
- La distribuzione è caratterizzata da m e σ^2

(spesso si utilizza la variabile ridotta $z = \frac{x-m}{\sigma}$)

Con qualche calcolo si può verificare che:



- è soddisfatta la proprietà di chiusura
- **Valore medio**: E(X) = m = **m**
- **Varianza** : $\sigma^2(X) = \sigma^2 = \boldsymbol{\sigma^2}$
- la curva è simmetrica rispetto a X = m;
- ha due flessi per X = m - σ e X = m + σ
- il massimo della funzione vale $\frac{1}{\sqrt{2\pi}\sigma}$

  (modulo di precisione): se raddoppia σ (minore precisione) si dimezza il valore del massimo .

- $P(a \le X \le b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$ non è calcolabile analiticamente.

Nella pratica per conoscere tale probabilità si utilizzano delle <u>tabelle</u> (ERF) espresse in funzione della variabile standardizzata $z = \frac{x-m}{\sigma}$: $\int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \int_{\frac{a-m}{\sigma}}^{\frac{b-m}{\sigma}} f(z)\, dz$ dove $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

Alcuni valori notevoli sono:

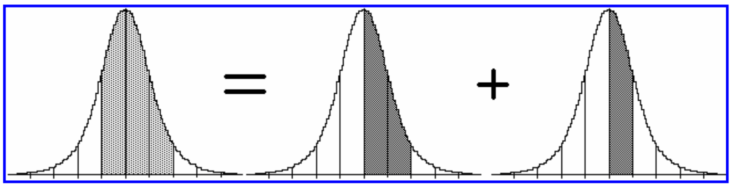
P(m - 1,0 σ ≤ X ≤ m + 1,0 σ) = P(-1,0 ≤ z ≤ 1,0) = **68,3 %**
P(m - 2,0 σ ≤ X ≤ m + 2,0 σ) = P(-2,0 ≤ z ≤ 2,0) = 95,4 %
P(m - 3,0 σ ≤ X ≤ m + 3,0 σ) = P(-3,0 ≤ z ≤ 3,0) = 99,7 %

<u>Funzione degli errori</u>

Poiché in presenza di soli errori casuali i risultati di misure sono distribuiti secondo la curva di Gauss, la funzione: $\int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{z'^2}{2}} dz'$ (che è tabulata) è detta anche funzione degli errori (ERF).

	0,00	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0949	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1369	0,1406	0,1443	0,1481	0,1518
0,4	0,1554	0,1591	0,1628	0,1664	0,1701	0,1737	0,1773	0,1808	0,1844	0,1880
0,5	0,1915	0,1950	0,1985	0,2020	0,2054	0,2089	0,2123	0,2157	0,2191	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2643	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2853
0,8	0,2881	0,2911	0,2939	0,2968	0,2996	0,3024	0,3051	0,3079	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3290	0,3315	0,3340	0,3365	0,3389
1,0	**0,3413**	0,3438	0,3462	0,3485	0,3509	0,3532	0,3555	0,3577	0,3600	0,3622
1,1	0,3644	0,3665	0,3687	0,3708	0,3729	0,3750	0,3770	0,3790	0,3810	0,3830
1,2	0,3850	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3998	0,4015
1,3	0,4032	0,4049	0,4066	0,4083	0,4099	0,4115	0,4131	0,4147	0,4162	0,4178
1,4	0,4193	0,4208	0,4222	0,4237	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4358	0,4370	0,4382	0,4395	0,4406	0,4418	0,4430	0,4441
1,6	0,4452	0,4463	0,4474	0,4485	0,4495	0,4506	0,4516	0,4526	0,4535	0,4545
1,7	0,4555	0,4564	0,4573	0,4582	0,4591	0,4600	0,4608	0,4617	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4679	0,4686	0,4693	0,4700	0,4706
1,9	0,4713	0,4720	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4762	0,4767
2,0	**0,4773**	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4813	0,4817
2,1	0,4822	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4858
2,2	0,4861	0,4865	0,4868	0,4872	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4899	0,4901	0,4904	0,4906	0,4909	0,4911	0,4914	0,4916
2,4	0,4918	0,4920	0,4923	0,4925	0,4927	0,4929	0,4931	0,4933	0,4935	0,4936
2,5	0,4938	0,4940	0,4942	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4954	0,4955	0,4956	0,4958	0,4959	0,4960	0,4961	0,4962	0,4963	0,4965
2,7	0,4966	0,4967	0,4968	0,4969	0,4970	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4975	0,4975	0,4976	0,4977	0,4978	0,4978	0,4979	0,4980	0,4980	0,4981
2,9	0,4982	0,4982	0,4983	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	**0,4987**	0,4987	0,4988	0,4988	0,4988	0,4989	0,4989	0,4990	0,4990	0,4990
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4999	0,4999

ESEMPIO: calcolare P(9,9 g ≤ M ≤ 10,2 g) con la massa M distribuita secondo una gaussiana con media m = 10,00 g e σ = 0,10 g.



$z_a = \frac{9{,}9 - 10}{0{,}1} = -1 \quad z_b = \frac{10{,}2 - 10}{0{,}1} = 2$ e quindi:

P(9,9 g ≤ M ≤ 10,2 g) = P(-1 ≤ z ≤ 2) =

= P(0 ≤ z ≤ 2) + P(-1 ≤ z ≤ 0) =

= P(0 ≤ z ≤ 2) + P(0 ≤ z ≤ 1) = ERF(2) + ERF(1) = 0,4773 + 0,3413 = 82,86 %

DISUGUAGLIANZA DI CHEBYCHEV

Molto spesso la distribuzione di una variabile aleatoria non è nota ma, dalla sola conoscenza di m e σ^2 è possibile trarre delle utilissime informazioni:

$P(m - K\sigma \le X \le m + K\sigma) \ge 1 - \frac{1}{K^2}$ (<u>disuguaglianza di Chebychev</u>)

dove K è una costante positiva

K	$P(m - K\sigma \le X \le m + K\sigma) \ge$
1	0%
1,5	55,6%
2	75,0%
2,5	84,0%
3	88,9%
4	93,7%
5	96,0%

$$\sum_{m-K\sigma \le x_i \le m+K\sigma} P(x_i) \ge 1 - \frac{1}{K^2}$$

$$\int_{m-K\sigma}^{m+K\sigma} f(x)\,dx \ge 1 - \frac{1}{K^2}$$

$\boxed{\text{nel 90\% circa dei casi } |x-m| < 3\sigma}$

TEOREMA DEL LIMITE CENTRALE

Questo importante teorema, in una delle sue formulazioni, può essere enunciato nel seguente modo:

<u>se</u> $X_1, X_2, ..., X_N$ sono N v.a. indipendenti con distribuzione qualsiasi e media e varianza dello stesso ordine di grandezza,

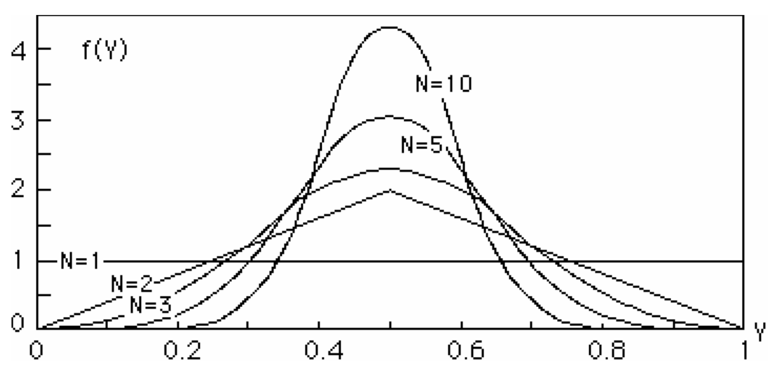
<u>allora</u> la v.a. $Y = X_1 + X_2 + ... + X_N$ ha una distribuzione di probabilità che per $N \to \infty$ tende ad essere gaussiana (se le variabili X_i seguono una distribuzione gaussiana allora anche Y segue una distribuzione gaussiana indipendentemente da N).

Quindi, avendo eseguito più misure di una stessa grandezza, la media aritmetica di tali misure tende ad avere una distribuzione gaussiana tanto più quanto più è elevato il numero delle misure effettuate.

ESEMPIO

consideriamo la funzione $Y = \frac{\sum X_i}{N}$ dove le v.a. X_i sono distribuite uniformemente fra 0 e 1: $E(Xi) = 0,5$ $\sigma(X_i) = 1/\sqrt{12}$

- Y ha media $E(Y) = E(X_i) = 0,5$ indipendentemente da N
- la varianza di Y è $\sigma^2(Y) = \frac{\sigma^2(X_i)}{N}$; quindi diminuisce al crescere di N



- secondo il teorema del limite centrale per N elevato Y ha distribuzione gaussiana.

Nell'esempio per N = 1 la distribuzione è uniforme e già per $N \approx 10$ diventa pressoché gaussiana.

Elementi di statistica

Scopo di ogni misurazione è la determinazione del valore vero di un misurando *M*.

Le imperfezioni degli strumenti, le variazioni delle condizioni ambientali e l'influenza dell'osservatore provocano inevitabili errori di misura a causa dei quali non è possibile trovare il valore vero *M*.

Si assume che i valori x_i ottenuti da diverse misurazioni siano i valori assunti da una variabile casuale X che obbedisce ad una distribuzione di probabilità caratterizzata in particolare dai parametri m e σ.

In assenza di errori sistematici il valore atteso m coincide col valore vero *M* del misurando.
La deviazione standard σ è una misura della variabilità di un singolo valore misurato dal valore atteso m.

STIMA DI PARAMETRI DI UNA DISTRIBUZIONE

Stima della media della distribuzione (*valor vero*)

La stima è data dalla media aritmetica del campione: $m \approx \overline{X} = \frac{\sum_{i=1,N} x_i}{N}$

Il valore atteso della media aritmetica è $E(\overline{X}) = E\left(\frac{\sum x_i}{N}\right) = \sum \frac{E(X_i)}{N} = \sum \frac{m}{N} = N\frac{m}{N} = m$

Se più valori x_i (i=1,N) coincidono conviene raggruppare gli N valori in M classi ognuna con frequenza n_j (j=1,M): $\sum_{i=1,N} x_i = \sum_{j=1,M} n_j x_j$.

$$\frac{\sum_{i=1,N} x_i}{N} = \frac{\sum_{j=1,M} n_j x_j}{N} \quad \text{ma} \quad \frac{\sum_{j=1,M} n_j x_j}{N} = \sum_{j=1,M} \frac{n_j x_j}{N} = \sum_{j=1,M} f_j\, x_j \approx \sum_{j=1,M} x_j\, P(x_j) = E(X) = m$$

essendo le frequenze relative $f_j = n_j/N$ una stima delle probabilità $P(x_j)$ si giustifica l'uso della media aritmetica nella stima del valor medio m.

ESEMPIO Supponiamo di aver ottenuto 10 misure di lunghezza:
80, 85, 85, 90, 90, 90, 95, 100, 100, 105 [cm]; calcoliamone la media aritmetica.
Raggruppando le N = 10 misure in M = 6 classi si ottiene:

j	x_j	n_j	$n_j\, x_j$
1	80	1	80
2	85	2	170
3	90	3	270
4	95	1	95
5	100	2	200
6	105	1	105
		10	920

$$m \approx \overline{X} = \frac{\sum_{i=1,N} x_i}{N} = \frac{\sum_{j=1,M} n_j x_j}{N}$$

pertanto il risultato è 92 cm

Una misura della **dispersione** dei dati intorno alla media aritmetica non può venire dalla somma degli scarti: $\sum_{i=1,N}(x_i-\overline{X})=\left(\sum_{i=1,N}x_i\right)-N\,\overline{X}=N\,\overline{X}-N\,\overline{X}=0$

Stima della varianza σ^2 di una distribuzione

- La stima è data dalla varianza campionaria o sperimentale $\sigma_s^2(X)=\dfrac{\sum_{i=1,N}(x_i-\overline{X})^2}{N-1}$
- Se venisse stimata con $\sigma'^2=\dfrac{\sum_{i=1,N}(x_i-\overline{X})^2}{N}$ si otterrebbe $E(\sigma_s'^2)=\dfrac{N-1}{N}\sigma^2$; per questo motivo a denominatore è presente N-1
- Per velocizzare i calcoli può essere utile ricordare che:

$$\sum_{i=1,N}(x_i-\overline{X})^2=\sum_{i=1,N}x_i^2-2\overline{X}\sum_{i=1,N}x_i+N\overline{X}^2=\sum_{i=1,N}x_i^2-2\overline{X}\,N\overline{X}+N\overline{X}^2=\sum_{i=1,N}x_i^2-N\overline{X}^2$$

- Raggruppando gli N valori x_i (i=1,N) in M classi di frequenza n_j (j=1,M):

$$\sigma_s^2(X)=\frac{\sum_{j=1,M}n_j(x_j-\overline{X})^2}{N-1}=\sum_{j=1,M}\frac{n_j}{N-1}(x_j-\overline{X})^2\approx\sum_{j=1,M}f_j(x_j-\overline{X})^2\approx\sum_{j=1,M}P(x_j)(x_j-m)^2=\sigma^2$$

ESEMPIO: Riprendiamo l'esempio precedente e costruiamo la tabella:

j	x_j[cm]	n_j	$n_j\,x_j$[cm]	x_j^2 [cm^2]	$n_jx_j^2$ [cm^2]
1	80	1	80	6400	6400
2	85	2	170	7225	14450
3	90	3	270	8100	24300
4	95	1	95	9025	9025
5	100	2	200	10000	20000
6	105	1	105	11025	11025
		10	920		85200

$$\sigma_s^2(X)=\frac{\sum_{i=1,N}(x_i-\overline{X})^2}{N-1}=\frac{\sum_{i=1,N}x_i^2-N\overline{X}^2}{N-1}=\frac{\sum_{j=1,M}n_jx_j^2-N\overline{X}^2}{N-1}=\frac{85200-10\times9{,}20^2}{9}=62{,}2\text{cm}^2$$

Stima della deviazione standard

Ovviamente: $\sigma(X)\approx\sigma_s(X)=\sqrt{\sigma_s^2(X)}=\sqrt{\dfrac{\sum_{i=1,N}(x_i-\overline{X})^2}{N-1}}$ e quindi: $\sigma_s(X)=\sqrt{\sigma_s^2(X)}=\sqrt{62{,}2\text{cm}^2}=7{,}89\text{cm}$

Stima della deviazione standard della media aritmetica

in analogia con $\sigma(\overline{X})=\dfrac{\sigma(X)}{\sqrt{N}}$ si ha come stima $\sigma_s(\overline{X})=\dfrac{\sigma_s(X)}{\sqrt{N}}=\sqrt{\dfrac{\sum_{i=1,N}(x_i-\overline{X})^2}{N(N-1)}}$

Il risultato di una serie di misure viene espresso dalla media aritmetica con associata un'incertezza espressa dalla deviazione standard stimata della media aritmetica: $\bar{X} \pm \sigma_s(\bar{X})$ (incertezza di tipo A, secondo la definizione delle norme ISO)

Si indica così che con alta probabilità il valore vero è all'interno dell'intervallo: $[\bar{X} - \sigma_s(\bar{X}); \bar{X} + \sigma_s(\bar{X})]$
La probabilità del risultato è detto **livello di confidenza**; l'intervallo nel quale si ritiene sia contenuto il valore vero è detto **intervallo di confidenza**.

Nell'**esempio** delle 10 misure di lunghezza il risultato è quindi (92,0 ± 2,5) cm dove viene attribuita alla media aritmetica un'incertezza $\sqrt{10}$ volte più piccola di quella associata alla singola misura (nel nostro caso $\sigma_s(X) =$ 7,89 cm).

Stima della distribuzione di probabilità della media aritmetica

1) se il risultato di una misura è determinato dalla concorrenza di numerosi fenomeni microscopici che, ognuno con la sua distribuzione di probabilità, alterano la grandezza in esame, il risultato della misura tenderà ad avere una distribuzione di tipo pressoché gaussiano;

2) a prescindere dall'effetto precedente, se consideriamo la media aritmetica di N misure questa segue una distribuzione gaussiana (per $N \to \infty$) indipendentemente da quale sia la distribuzione di partenza;

3) l'effetto combinato di **1)** (distribuzioni di partenza quasi gaussiane) e **2)** (media aritmetica di N misure) porta a distribuzioni praticamente gaussiane anche per medie aritmetiche di poche (5 - 10) misure

4) quindi il livello di confidenza dell'intervallo di confidenza $[\bar{X} - \sigma_s(\bar{X}); \bar{X} + \sigma_s(\bar{X})]$ è circa del **68%** (in realtà σ_s è solo una stima di σ; più correttamente occorrerebbe utilizzare la distribuzione t di Student che per un alto numero di misure coincide, ovviamente, con la curva di Gauss).