

PROBABILITA' E GAUSSIANA

Teoria delle probabilita': cenno

Dato un evento aleatorio, si definisce come frequenza relativa il num. delle volte che evento "capita" fratto il num. di prove totali:

$$Freq.rel. = N_{evento.capita}/N_{tot}. \quad (1)$$

Il limite della frequenza relativa per infinite prove, da' la **probabilita'**:

$$\lim_{N_{tot} \rightarrow \infty} Freq.rel. = Prob.. \quad (2)$$

Questa e' la definizione frequentistica, esistono anche altre definizioni, ad es. la teorica, etc. Esempio: la probabilita' che esca il valore "3" lanciando un dado e' 1/6.

Se la **variabile discreta** x puo' assumere N valori, la somma delle probabilita' e' $\sum_{i=1}^N p_i = 1$ per definizione stessa di probabilita'. La media o speranza matematica e' $\bar{x} = \sum_{i=1}^N x_i p_i$, la $Var(x) = \sum_{i=1}^N (x_i p_i - \bar{x})^2$, in pratica i p_i vengono usati come "pesi" (veda la "media pesata" in statistica descrittiva).

Sia x una **variabile continua** che puo' assumere tutti i valori nell'intervallo (a, b) . Sia $\Delta p = p(x, \Delta x)$ la probabilita' che x abbia un valore in un certo intervallo Δx contenuto in (a, b) ; per definizione $p(x, (a, b)) = 1$. Definiamo una funzione $y^*(x, \Delta x) = \Delta p / \Delta x$. La **densita' di probabilita'** e' allora:

$$y = \lim_{\Delta x \rightarrow 0} \Delta p / \Delta x = dp/dx. \quad (3)$$

Sara' allora $\bar{x} = \int_a^b xy(x)dx$ e per definizione di probabilita' $\int_a^b y(x)dx = 1$.

Caso equiprobabile: la y vs. x e' una retta orizzontale e $y = 1/(b - a)$. In altri casi la $y(x)$ puo' essere piu' o meno complicata. Sotto discutiamo la gaussiana ("Gaussian").

Errori accidentali e gaussiana

La legge di distribuzione degli errori accidentali e' ricavabile da un complesso di ipotesi postulate da Gauss:

- la deviazione macroscopica u del valore osservato da quello vero e' causato da n agenti, ciascuno dei quali porta una minuscola variazione, detta deviazione elementare;
- tutte le deviazioni elementari hanno egual valore assoluto e deviazioni di segno opposto sono equiprobabili;
- le cause di queste deviazioni elementari sono mutuamente indipendenti.

Si puo' dimostrare che il valore piu' probabile di u e' zero, cioe' sono piu' probabili errori macroscopici piccoli piuttosto che grandi (la cosiddetta **tendenza centrale**); u e' in modo equiprobabile positivo o negativo (**simmetria**). Si puo' anche dimostrare che la densita' di probabilita' di u e' data da:

$$y(u) = \frac{1}{\sqrt{2\pi}\sigma} e^{[-u^2/(2\sigma^2)]}, \quad (4)$$

dove σ e' una costante.

Se x e' una grandezza fisica il cui valore vero e' μ , i risultati delle misurazioni su di essa si distribuiranno con la stessa legge, ma attorno ad μ invece che a zero, dato che $u = x - \mu$. Si ha quindi la distribuzione **gaussiana**, detta anche distribuzione **normale**, che in grafico risulta essere una curva a campana, simmetrica, e va a zero agli estremi:

$$y(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}. \quad (5)$$

Si puo' dimostrare che il valore piu' probabile del valore vero μ e' \bar{x} e il valore piu' probabile di σ^2 e' proprio la varianza dei dati cioe' il quadrato dello scarto quadratico medio ($Var(x) = s^2$), quindi troveremo scritto:

$$y(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\bar{x})^2/(2\sigma^2)}. \quad (6)$$

Dato il valore vero μ e una misura x , la probabilita' che μ cada fra $x - \sigma$ e $x + \sigma$ e' 0.68=68%, ma vale anche che la probabilita' che x cada fra $\mu - \sigma$ e $\mu + \sigma$ e' 0.68=68%! A 2σ avremo probabilita' del 95%, a 3σ del 99% (vedi figure e tabella qui di seguito).

La **forma standardizzata** della distribuzione normale (cioe' con media=0 e varianza=1) si ottiene con il cambio di variabile $z = (x - \mu)/\sigma$, per cui:

$$y(z) = \frac{1}{\sqrt{2\pi}} e^{(-z^2/2)}. \quad (7)$$

Per $y(z)$ si ha la tabella per calcolare le varie probabilita' (qui di seguito).

Errore sulla media. Per capire quanto il valore medio si discosta dal valore vero (e viceversa) dovremo fare la varianza della media. Si dimostra che e' σ^2/N dove N e' il numero totale dei dati ($\eta = \sigma/\sqrt{N}$). In pratica, riesco a recuperare il valore vero μ tanto meglio quanto piu' numerose sono le misure. Il risultato sara' allora $R = \bar{x} \pm \sigma/\sqrt{N}$ e significa che la probabilita' che μ cada fra $\bar{x} - \sigma/\sqrt{N}$ e $\bar{x} + \sigma/\sqrt{N}$ e' del 68%.

Seguono altri appunti utili con grafici e tabella.

VARIABILI CONTINUE

Una variabile si dice continua se può assumere tutti i valori fra 2 valori dati.

$$x \in [a, b]$$

Probabilità di trovare x è $\frac{1}{\infty} = 0$


$$a \quad \text{-----} \quad b$$

$P(x)$ oggetto che può assumere valori equiprobabili in un intervallo a, b

$$\Delta x \rightarrow P(\Delta x) = \frac{\Delta x}{b-a}$$

ΔP probabilità che misura sia in un intervallo

$$\left[x + \frac{\Delta x}{2}, x - \frac{\Delta x}{2} \right]$$

Posso definire una funzione

$$y^*(x, \Delta x) = \frac{\Delta P}{\Delta x}$$

$$y(x) = \lim_{\Delta x \rightarrow 0} y^*(x, \Delta x) = \lim_{\Delta x \rightarrow 0} \frac{\Delta P}{\Delta x} \quad \text{converge}$$

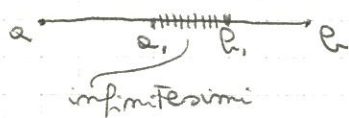
DENSITÀ DI PROBABILITÀ

Se eventi equiprobabili $\Delta P = \frac{\Delta x}{b-a} \quad \frac{\Delta P}{\Delta x} = \frac{1}{b-a}$

ΔP probabilità che variabile assuma un valore in Δx

per Δx piccoli $\rightarrow \Delta P$ piccoli

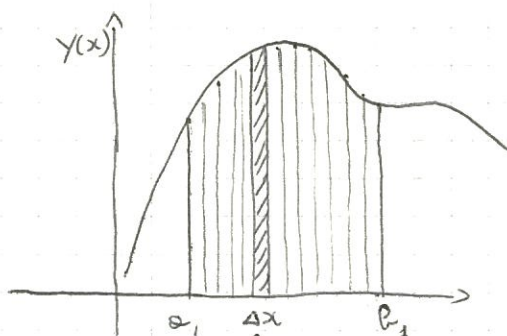
$$\Delta P = y(x) \cdot \Delta x$$



$$\Delta P = \Delta x \cdot y(x) \rightarrow \int_{a_1}^{b_1} y(x) \Delta x$$

Probabilità che risultato capiti

fra A e B $\bar{e} = 1$



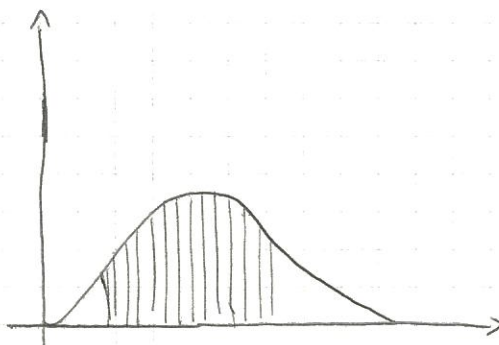
$$1 = \int_a^b y(x) dx$$

condizione di
normalizzazione

$$\sum P_i x_i \rightarrow \sum \underbrace{y(x) \Delta x}_{P_i} \cdot \underbrace{x}_{x_i}$$

Se $\Delta x \rightarrow 0$

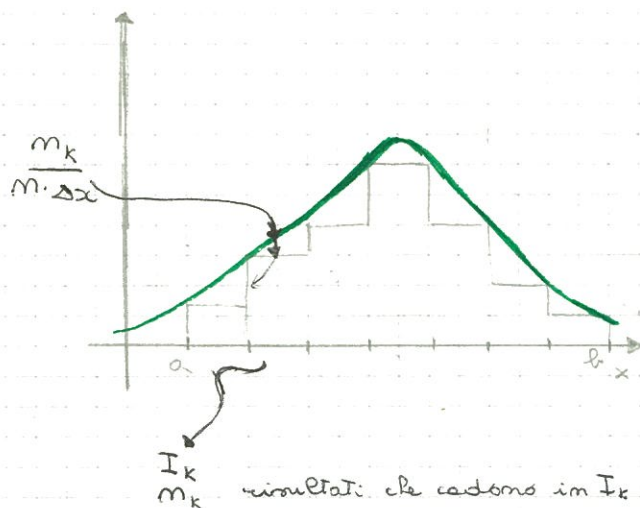
$$\rightarrow \int_a^b x y(x) dx$$



$$\text{Var}(x) = \int_a^b (x - [x])^2 y(x) dx = \int_a^b x^2 y(x) dx - \left(\int_a^b x y(x) dx \right)^2$$

m risultati: x_1, \dots, x_m

Per vedere come si distribuiscono faccio ISTOGRAMMA



$$\Delta x = \frac{b-a}{N}$$

$$k = 0, \dots, N$$

tendenza centrale

simmetria rispetto al centro

Passare al limite: stringere ampiezza del canale, aumentare misure

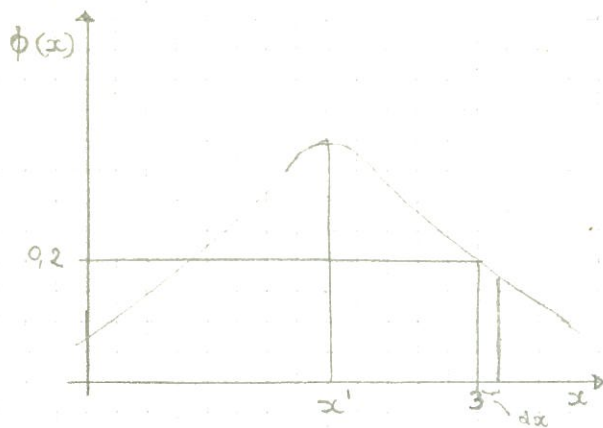
3 IPOTESI:

- 1) in ogni misura sono presenti degli errori
- 2) ogni errore macroscopico consiste di una somma di errori elementari δ
- 3) ogni δ ha la stessa probabilità di capitare tanto in meno quanto in più.

Da queste ipotesi consegue:

- 1) E' + probabile che somma dei δ sia $= 0$ poiché δ si compensano \Rightarrow sono più probabili errori macroscopici piccoli che grandi. (TENDENZA CENTRALE)
- 2) δ sia in meno che in più (SIMMETRIA)

Gaussiana



probabilità che cada in
intervallo $\bar{e} = 0.2 \cdot dx$

$$\phi(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-x')^2}{2\sigma^2}}$$

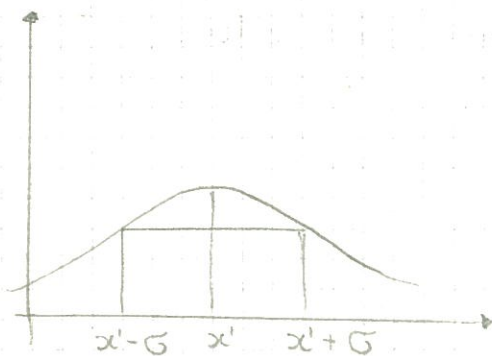
$\phi(x)$ = densità di probabilità

si ottiene come limite di una densità di frequenza



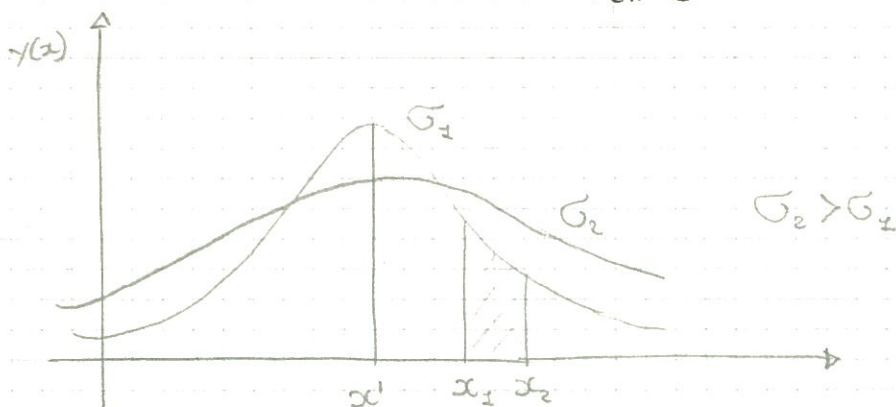
(integrale...)

la probabilità è rappresentata da
tutta l'area



probabilità che valore vero cada in intervallo fra $x'-\sigma$ e $x'+\sigma$
 $\bar{e} = 68\%$

$$y(x) = \frac{1}{\sqrt{2\pi} \sigma} \cdot e^{-\frac{(x-x')^2}{2\sigma^2}}$$



$$p \{ x \in [x_1, x_2] \} = \int_{x_1}^{x_2} y(x) dx \quad \text{che è poi la densità di probabilità}$$

Area sottesa dalle curve = 1

probabilità che x abbia un valore qualsiasi = 1

2 parametri x' e σ

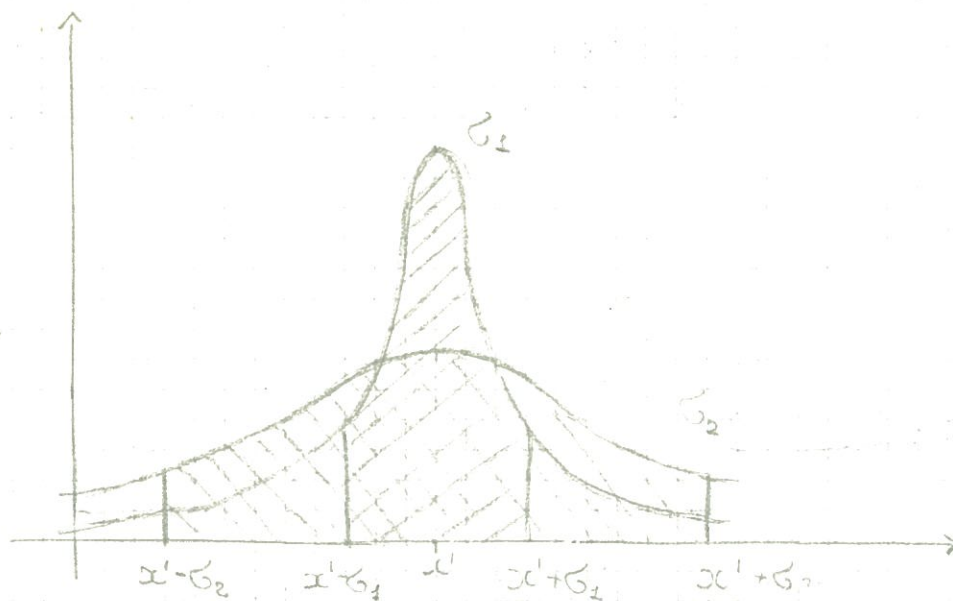
Posso cambiare forme delle curve, ma area sottesa sempre = 1

quando $y = \frac{1}{\sqrt{2\pi} \sigma}$ se $x = x'$ massima altezza

+ grande σ + bassa la curva

$$\sigma_2 > \sigma_1$$

! valori + diversi



più preciso quello con σ_1

$$p\{x \in [x' - \sigma_1, x' + \sigma_1]\} = 0,68 = 68\%$$

$$p\{x' \in [x - \sigma, x + \sigma]\} = 0,68 = 68\%$$



$[x - \sigma, x + \sigma]$ INTERVALLO DI CONFIDENZA al 68%

Ci fidiamo di trovarvi valore vero

m	p
1	0,68
2	0,95
3	0,99

$$[x - m\sigma, x + m\sigma]$$

informazioni + precise: 1) stessa probabilità con intervalli minore
2) probabilità maggiore nello stesso intervallo

4) aumento misure

$$n \quad x_1, \dots, x_n$$
$$\bar{x} = \frac{x_1 + \dots + x_n}{n}$$

$$\sigma_{\bar{x}}^2 \text{ var } \{\bar{x}\} = \text{var} \left\{ \frac{x_1 + x_2 + \dots + x_n}{n} \right\} = \text{var} \left\{ \frac{x_1}{n} \right\} + \dots + \text{var} \left\{ \frac{x_n}{n} \right\}$$

IPOTESI:

x_1, \dots, x_n sono statisticamente indipendenti:

$$= \frac{1}{n^2} \sum_{i=1}^n \text{var} \{x_i\} = \frac{n \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$\sigma_x = \frac{\sigma_x}{\sqrt{n}}$$

Ha senso quindi fare misure ripetute

$$\eta = \frac{\sigma}{\sqrt{n}}$$

media delle misure è $\frac{\sigma}{\sqrt{n}}$ + effluente della misura
singola per stabilire valore vero



z	0	1	2	3	4	5	6	7	8	9
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	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	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	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000