Università degli Studi di Bari
Dipartimento di Chimica

F.Mavelli

Laboratorio di Chimica Fisica

a.a. 2013-2014

Analisi Statistica dei Dati

Università degli Studi di Bari - Dipartimento di Chimica

Analisi Statistica

L'analisi statistica si applica al caso delle misure ripetute e permette di ottenere alcuni dei risultati che sono già stati esposti precedentemente:

stima della misura

valor medio

$$x_m = \frac{\sum_{i=1}^{N} x_i}{N} = \bar{x}$$

stima dell'errore

deviazione standard

$$s_x = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}}$$

deviazione standard della media

$$s_{\bar{x}} \; \frac{s_x}{\sqrt{N}}$$

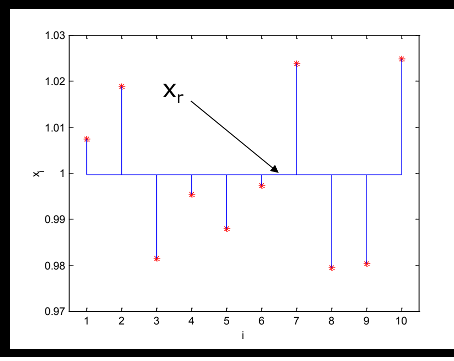
Università degli Studi di Bari - Dipartimento di Chimica

Analisi Statistica

La principale assunzione su cui si basa l'analisi statistica è che siano stati eliminati tutti gli errori sistematici

Assenza di errori sistematici

Gli unici errori presenti sono dunque quelli casuali che determinano scostamenti random intorno al valore vero x_r della grandezza in esame.



Università degli Studi di Bari - Dipartimento di Chimica

Analisi Statistica

L'analisi statistica dei risultati di una misura richiede che vengano effettuate numerose misure ripetute sullo stesso campione affinché i risultati dell'analisi siano corretti.

Distingueremo due casi:

valori discreti: il risultato della misura o dell'esperimento può assumere solo valori discreti, ad esempio valori interi positivi.

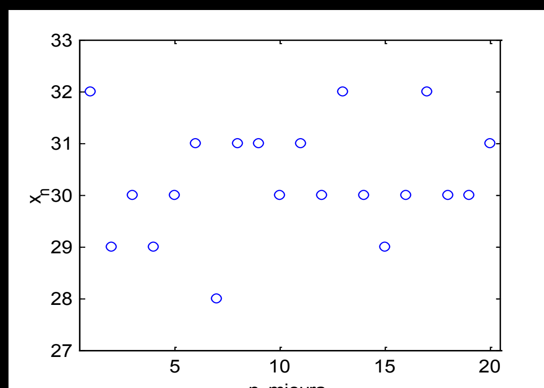
valori continui: il risultato della misura o dell'esperimento può essere un qualsiasi valore reale

F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

Università degli Studi di Bari - Dipartimento di Chimica

Valori discreti

Consideriamo un'ipotetica operazione di misura il cui risultato sia una varabile intera e immaginiamo di avere ottenuto la seguente serie di risultati da 20 operazioni di misura ripetute:



1) 32
2) 29
3) 30
4) 29
5) 30
6) 31
7) 28
8) 31
9) 31
10) 30
11) 31
12) 30
13) 32
14) 30
15) 29
16) 30
17) 32
18) 30
19) 30
20) 31

F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

Università degli Studi di Bari - Dipartimento di Chimica

Valori discreti

Possiamo ordinare i risultati in ordine crescente e raggrupparli per valore.

								n_x	F_x
28								1	0.05
29	29	29						3	0.15
30	30	30	30	30	30	30	30	8	0.40
31	31	31	31	31				5	0.25
32	32	32						3	0.15

quindi per ogni valore di x possiamo ricavare il numero di volte n_x che esso è apparso

$$\sum_{x=1}^{5} n_x = N = 20$$

e possiamo ricavare la frequenza F_x con cui esso è apparso

$$F_x = \frac{n_x}{N}$$

$$\sum F_x = \sum \frac{n_x}{N} = \frac{1}{N} \sum n_x = \frac{N}{N} = 1$$

Valori discreti

Le frequenze F_x rappresentano le probabilità stimate che il valore della grandezza x, determinato a seguito di una misura, risulti essere un certo valore, ad esempio:

x	F_x
28	0.05
29	0.15
30	0.40
31	0.25
32	0.15

$$\sum_{x=28}^{32} F_x = 1$$

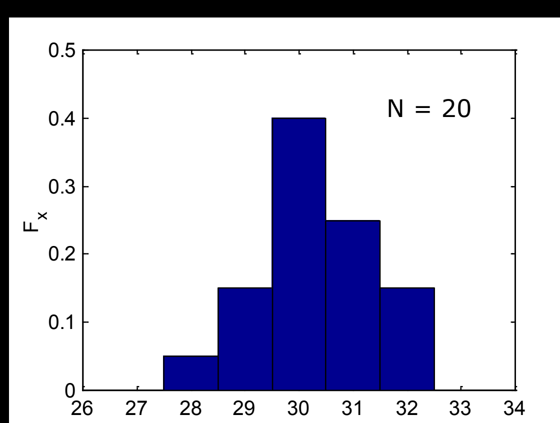
$F_{30} = 0.4$

esiste una probabilità del 40% che, a seguito di una misura, il valore di x risulti essere uguale a 30

si noti inoltre come le frequenze F_x risultino normalizzate, ossia la loro somma dia 1, e come questo dipenda dalla loro definizione.

Istogramma

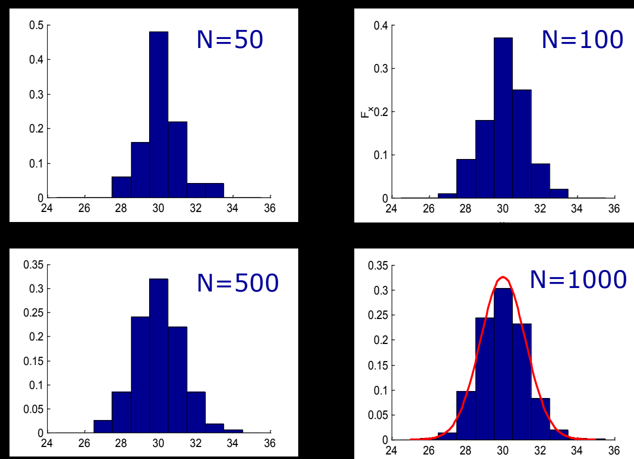
Possiamo riportare in grafico a barre (istogramma) i valori delle frequenze F_x contro i relativi valori x:



Università degli Studi di Bari - Dipartimento di Chimica

Istogramma

Aumentando il numero totale N di misure effettuate, le frequenze F_x approssimano meglio la funzione di probabilità



F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

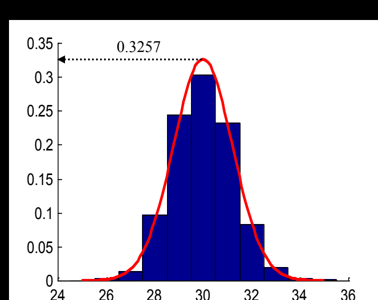
Università degli Studi di Bari - Dipartimento di Chimica

Funzione probabilità P(x)

Per cui abbiamo ottenuto che al tendere di N ad infinito le frequenze stimate tendono alla funzione probabilità P(x) così definita:

$$\lim_{N\to\infty} F_x = P(x)$$

P(x) ≡ *Probabilità*{che il risultato x_i di una misura sia proprio uguale ad x }



P(30) = 0.3257

↓

probabilità che il risultato della misura risulti uguale a 30: x_i = 30

Funzione probabilità P(x)

La funzione Probabilità deve soddisfare le seguenti proprietà:

$P(x) \geq 0$	è sempre positiva o nulla
$\sum_{x=0}^{\infty} P(x) = 1$	è normalizzata, cioè la sommatoria estesa a tutti i valori possibili deve essere uguale ad 1
$\sum_{x=A}^{B} P(x) =$	probabilità che la misura x_i abbia un valore compreso fra A e B (A < B)

P(30) = 0.3257
P(29) = 0.2334
P(29)+P(30) = 0.5591

probabilità che x_i risulti uguale a 29 o 30

Università degli Studi di Bari - Dipartimento di Chimica

Valore Medio

Si noti come il valore medio possa essere calcolato sia secondo la definizione, che utilizzando le frequenze F_x

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{\sum_{x}^{Valori} n_x x}{N} = \sum_{x}^{Valori} \frac{n_x}{N} x = \sum_{x}^{Valori} F_x x$$

Ad esempio nel caso precedente si ha:

$$\bar{x} = \frac{29+30+29+30+31+28+31+31+30+31+30+32+30+29+30+32+30+30+31}{20} =$$
$$= \frac{(1\cdot 28)+(3\cdot 29)+(8\cdot 30)+(5\cdot 31)+(3\cdot 32)}{20} =$$
$$= \left(\frac{1}{20}\cdot 28\right)+\left(\frac{3}{20}\cdot 29\right)+\left(\frac{8}{20}\cdot 30\right)+\left(\frac{5}{20}\cdot 31\right)+\left(\frac{3}{20}\cdot 32\right) =$$
$$= (0.05\cdot 28)+(0.15\cdot 29)+(0.40\cdot 30)+(0.25\cdot 31)+(0.15\cdot 32) \quad = 30.30$$

Università degli Studi di Bari - Dipartimento di Chimica

Varianza

Si noti come anche la varianza possa essere calcolata utilizzando le frequenze F_x

$$s_x^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1} = \frac{\sum_x n_x (x - \bar{x})^2}{N-1} = \left(\frac{N}{N-1}\right)\sum_x \frac{n_x}{N}(x - \bar{x})^2$$

$$= \left(\frac{N}{N-1}\right)\sum_x F_x (x - \bar{x})^2$$

e se N >> 1 allora risulta:

$$s_x^2 \approx \sum_x F_x (x - \bar{x})^2$$

Valori continui

Nel caso di un variabile continua x il primo problema che si pone nell'analisi statistica e come raggruppare i valori per determinare le frequenze delle occorrenze delle singole misure.

Bisogna discretizzare l'asse dei valori delle x in una serie di intervalli contigui: *classi*, e determinare il numero di misure x_i che cadono in ogni singola classe.

i	x_i
1)	28.0354
2)	30.3151
3)	28.7061
4)	31.7332
5)	29.0140
6)	30.6476
7)	30.2686
8)	28.8709
9)	27.3415
10)	29.9275

Valori continui

classi		valori misurati		
k	x_k	x_i	n_k	F_k
1)	27±0.5	27.3415	1	0.1
2)	28±0.5	28.0354	1	0.1
3)	29±0.5	28.7061, 28.8709, 29.0140	3	0.3
4)	30±0.5	29.9275, 30.3151, 30.2686	3	0.3
5)	31±0.5	30.6476	1	0.1
6)	32±0.5	31.7332	1	0.1

$N = 10$

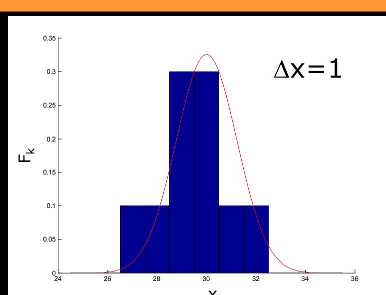
$F_k = n_k/N$

L'ampiezza Δx delle classi deve essere scelta in modo che in ogni classe cadano dei valori, ossia non ci siano classi vuote, ed il numero delle classi deve essere il più alto possibile (nella tabella $\Delta x = 1$).

F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

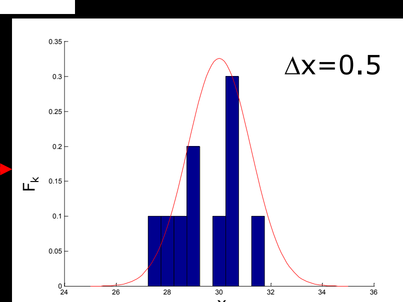
Università degli Studi di Bari - Dipartimento di Chimica

Scelta delle classi



Scelta delle classi corretta

Scelta delle classi non corretta



Valori Continui

Le frequenze F_k sono adesso le stime delle probabilità che una certa misura x_i cada nella classe k-esima $x_k \pm \Delta x/2$.

Tali frequenze saranno tanto più elevate quanto più è ampia la classe:

$$F_k = f_k \cdot \Delta x$$

a senso quindi introdurre la densità di frequenza f_k definita come la frequenza F_k diviso per l'intervallo Δx:

$$f_k = F_k / \Delta x$$

Densità di probabilità

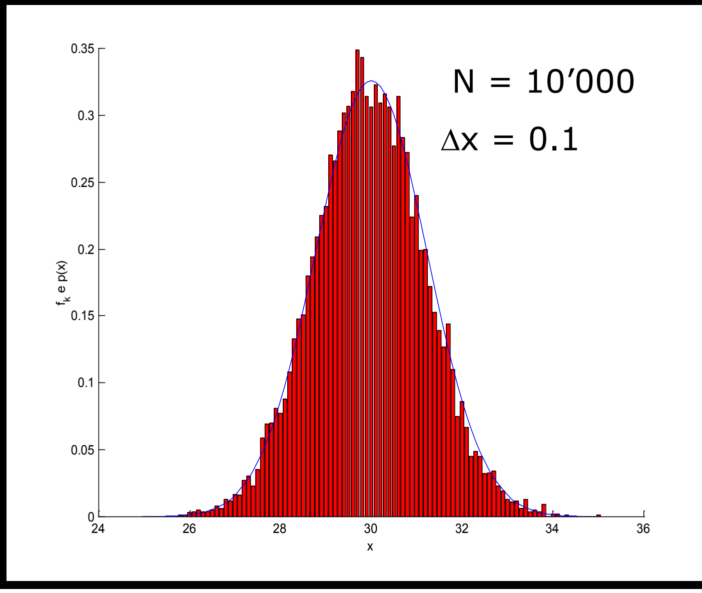
Al tendere di N ad infinito e Δx a zero la densità di frequenza f_k tende ad una funzione limite densità di probabilità p(x) così definita:

$$\lim_{\substack{N \to \infty \\ \Delta x \to 0}} \frac{F_k}{\Delta x} = \lim_{\substack{N \to \infty \\ \Delta x \to 0}} f_k = p(x)$$

p(x) dx ≡ *Probabilità* { che il risultato x_i di una misura sia compreso nell'intervallo: [x-dx/2, x+dx/2] }

Università degli Studi di Bari - Dipartimento di Chimica

Densità di probabilità



F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

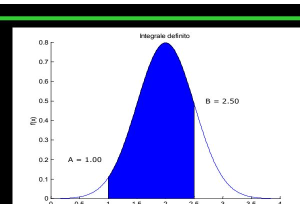
Università degli Studi di Bari - Dipartimento di Chimica

Densità di probabilità p(x)

La funzione densità di probabilità deve soddisfare le seguenti proprietà:

$p(x) \geq 0$	è sempre positiva o nulla
$\int_{-\infty}^{+\infty} p(x)dx = 1$	è normalizzata, cioè l’integrale esteso a tutto l’intervallo di definizione deve essere uguale ad 1
$\int_{A}^{B} p(x)dx =$	probabilità che la misura x_i abbia un valore compreso fra A e B: $x_i \in [A, B]$ (A<B)

va ricordato che l’integrale definito fra A e B della funzione $f(x)$ altro non è che l’area compresa fra la curva e l’asse delle X



Densità di probabilità p(x)

Data funzione densità di probabilità p(x) viene definito <x> valore medio della distribuzione l'integrale:

$$\langle x \rangle = \int x \cdot p(x) \cdot dx$$

mentre la varianza della distribuzione viene definita:

$$Var\{x\} = \int \left(x - \langle x \rangle\right)^2 \cdot p(x)\,dx$$

e da un'indicazione di quanto la distribuzione sia slargata intorno al valor medio.

F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

Università degli Studi di Bari - Dipartimento di Chimica

Densità di probabilità p(x)

Si dimostra inoltre facilmente che la varianza della distribuzione è data dalla differenza della media del quadrato meno il quadrato della media:

$$Var\{x\} = \int \left(x - \langle x \rangle\right)^2 \cdot p(x)\,dx = \langle x^2 \rangle - \langle x \rangle^2$$

Dimostrazione:

$$\begin{aligned} Var\{x\} &= \int \left(x - \langle x \rangle\right)^2 \cdot p(x)dx = \int \left(x^2 - 2x\langle x \rangle + \langle x \rangle^2\right) \cdot p(x)dx = \\ &= \int x^2 p(x)\,dx - 2\langle x \rangle \int x\,p(x)\,dx + \langle x \rangle^2 \int p(x)\,dx = \\ &= \langle x^2 \rangle - 2\langle x \rangle\langle x \rangle + \langle x \rangle^2 \cdot 1 = \langle x^2 \rangle - \langle x \rangle^2 \end{aligned}$$

Valor medio

Cerchiamo adesso di mettere in relazione la media aritmetiche di N misure ripetute con il valor medio della densità di probabilità. Esprimiamo il valor medio in funzione delle frequenze

$$\overline{x} = \sum_{k}^{M_{classi}} F_k x_k = \sum_{k}^{M_{classi}} \left(f_k \Delta x_k \right) x_k$$

mandando ad infinito il numero N di misure ed a zero la dimensione Δx delle classi:

$$\lim_{N\to\infty} \overline{x} = \lim_{\substack{N\to\infty \\ \Delta x\to 0}} \sum_{k}^{M_{classi}} (f_k \cdot \Delta x) x = \int x \cdot p(x) \cdot dx = \langle x \rangle$$

F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

Università degli Studi di Bari - Dipartimento di Chimica

Valor medio

Resta così dimostrato che il limite della media aritmetica per N→∞ tende proprio al valore vero x_r, nell'ipotesi che siano assenti gli errori sistematici e che quindi la funzione densità di probabilità p(x) sia simmetrica e centrata intorno al valore vero, ossia che x_r = <x>.

$$\lim_{N\to\infty} \overline{x} = \langle x \rangle = x_r$$

La media aritmetica è uno stimatore corretto del valor vero nel caso di un numero finito N di misure ripetute ed in assenza di errori sistematici

Varianza

Nel caso delle misure ripetute avevamo visto come l'errore poteva essere stimato con la deviazione standard:

Deviazione standard	Varianza
$s_x = \sqrt{\dfrac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}}$	$s_x^2 = \dfrac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}$

Vogliamo adesso trovare la relazione fra la varianza s_x^2 e la funzione densità di probabilità p(x).

Varianza

Per fare questo esprimiamo s_x^2 prima in termini delle frequenze osservate

$$s_x^2 \approx \frac{N \sum_{k}^{M_{classi}} F_k (x_k - \bar{x})^2}{N-1} = \frac{N \sum_{k}^{M_{classi}} (f_k \Delta x_k)(x_k - \bar{x})^2}{N-1}$$

e poi calcoliamo il limite per N che tende ad infinito e Δx che tende a zero

$$\lim_{N\to\infty} s_x^2 = \lim_{\substack{N\to\infty \\ \Delta x\to 0}} \frac{N \sum_{k}^{M_{classi}} (f_k \Delta x_k)(x_k - \bar{x})^2}{N-1} = \int (x - \bar{x})^2 \cdot p(x) dx$$

Varianza

per cui si ha:

$$\lim_{N\to\infty} s_x^2 = \int (\mathrm{x} - \overline{\mathrm{x}})^2 \cdot \mathrm{p}(\mathrm{x})\mathrm{dx} = Var\{\mathrm{x}\}$$

La varianza s_x^2 è uno stimatore corretto della varianza della funzione di distribuzione nel caso di un numero finito N di misure ripetute ed in assenza di errori sistematici

Conclusioni

L'analisi statistica ci ha mostrato che nel caso di errori casuali (in assenza quindi di errori sistematici) i valori delle misure ripetute si distribuiscono in maniera simmetrica intorno la valore vero x_r in accordo ad una funzione densità di probabilità p(x).

La funzione densità di probabilità p(x) può essere ricavata, effettuando infinite misure, come il limite delle frequenze ottenute per i singoli valori

$$\lim_{\substack{N\to\infty \\ \Delta x\to 0}} \frac{F_k}{\Delta x} = \lim_{\substack{N\to\infty \\ \Delta x\to 0}} f_k = p(x)$$

Conclusioni

Ottenuta la p(x) il valore vero x_r cercato può essere determinato effettuando un'operazione di media sulla funzione di distribuzione:

$$x_r = \int x \cdot p(x) \cdot dx$$

o, nel caso di un numero finito di misure ripetute, stimato con la formula della media aritmetica:

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Conclusioni

Mentre, la varianza $Var\{x\}$ può essere ottenuta come la differenza fra la media del quadrato ed il quadrato della media <x> = x_r:

$$Var\{x\} = \int (x - x_r)^2 \cdot p(x) \cdot dx$$

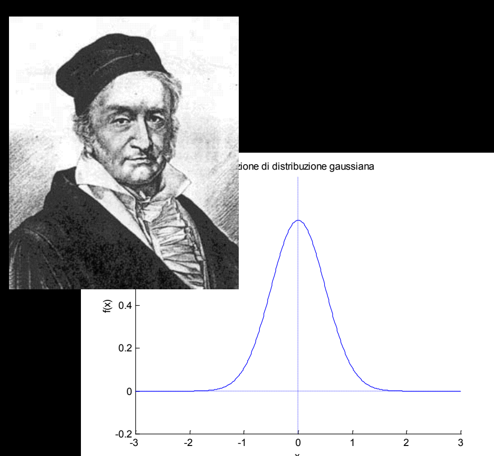
o, nel caso di un numero finito N di misure ripetute, stimata con la formula :

$$s_x^2 = \frac{\sum_{i=1}^{N} (x_i - \overline{x})^2}{N - 1}$$

Verrà adesso introdotta una particolare funzione densità di probabilità e, studiando le sue proprietà, si ricaverranno delle informazioni utili sul caso delle misure ripetute

Funzione di Gauss

1777 - 1855



$$G(x;\mu,\sigma)\frac{\exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right]}{\sqrt{2\pi}\sigma_x}$$

Università degli Studi di Bari - Dipartimento di Chimica

Funzione di Gauss

Introduciamo adesso la funzione di Gauss *f*(x), anche detta funzione guassiana:

Funzione gaussiana

$$f(x) = \exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right]$$

dove μ e σ_x sono costanti e inoltre $\sigma_x > 0$

e studiamone le proprietà.

F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

Università degli Studi di Bari - Dipartimento di Chimica

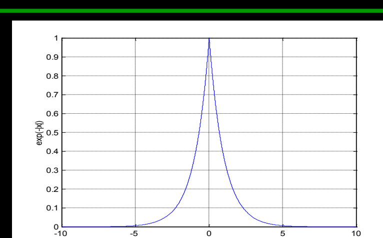
Funzione di Gauss

La funzione di Gauss ha valori sempre positivi compresi fra 0 e 1.

$$\forall x \in \Re : (x-\mu)^2 \geq 0 \rightarrow \exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right] \leq 1$$

$$\lim_{x\to\pm\infty} \exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right] = 0$$

$$\exp\left[-|x|\right]$$



Università degli Studi di Bari - Dipartimento di Chimica

Valore Massimo

La funzione di Gauss ha un massimo assoluto nel punto $x=\mu$.

$$f(\mu) = \exp\left[-\frac{(\mu-\mu)^2}{2\sigma_x^2}\right] = e^0 = 1$$

Dimostrazione:

Determiniamo la derivata prima

$$\frac{df(x)}{dx} = \frac{d}{dx}\exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right] = -\frac{(x-\mu)}{\sigma_x^2}\exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right]$$

$$\frac{df(x)}{dx} = -\frac{(x-\mu)}{\sigma_x^2}f(x)$$

F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

Università degli Studi di Bari - Dipartimento di Chimica

Valore Massimo

e cerchiamo gli zeri della derivata prima:

$$\frac{df(x)}{dx} = -\frac{(x-\mu)}{\sigma_x^2}f(x) = 0$$

$$f(x) = \exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right] \neq 0$$

$$-\frac{(x-\mu)}{\sigma_x^2}\exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right] = 0 \Leftrightarrow \frac{(x-\mu)}{\sigma_x^2} = 0 \Leftrightarrow x = \mu$$

$x = \mu$ zero della derivata prima

Valore Massimo

Determiamo la derivata seconda:

$$\frac{d^2 f(x)}{dx^2} = \frac{d}{dx}\left\{-\frac{(x-\mu)}{\sigma_x^2} f(x)\right\} =$$

$$= \frac{d}{dx}\left\{-\frac{(x-\mu)}{\sigma_x^2}\right\} f(x) + \left\{-\frac{(x-\mu)}{\sigma_x^2}\right\} \frac{df(x)}{dx} =$$

$$= \left\{-\frac{1}{\sigma_x^2}\right\} f(x) + \left\{-\frac{(x-\mu)}{\sigma_x^2}\right\}\left\{-\frac{(x-\mu)}{\sigma_x^2} f(x)\right\}$$

$$\frac{d^2 f(x)}{dx^2} = \left\{\frac{(x-\mu)^2}{\sigma_x^2} - 1\right\} \frac{f(x)}{\sigma_x^2}$$

F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

Università degli Studi di Bari - Dipartimento di Chimica

Valore Massimo

$$\frac{df(x)}{dx} = -\frac{(x-\mu)}{\sigma_x^2} f(x) \longrightarrow \frac{df(\mu)}{dx} = 0$$

$$\frac{d^2 f(x)}{dx^2} = \left\{\frac{(x-\mu)^2}{\sigma_x^2} - 1\right\} \frac{f(x)}{\sigma_x^2} \longrightarrow \frac{d^2 f(\mu)}{dx^2} = -\frac{1}{\sigma_x^2} < 0$$

punto di massimo della funzione

$$x = \mu \longrightarrow f(\mu) = \exp\left[-\frac{(\mu-\mu)^2}{2\sigma^2}\right] = e^0 = 1$$

Simmetria

La funzione gaussiana è simmetrica rispetto al valore massimo x=μ

$$f(\mu+\Delta x)=f(\mu-\Delta x)$$

Dimostrazione

$$f(\mu+\Delta x)=\exp\left[-\frac{[(\mu+\Delta x)-\mu]^2}{2\sigma_x^2}\right]=\exp\left[-\frac{\Delta x^2}{2\sigma_x^2}\right]$$

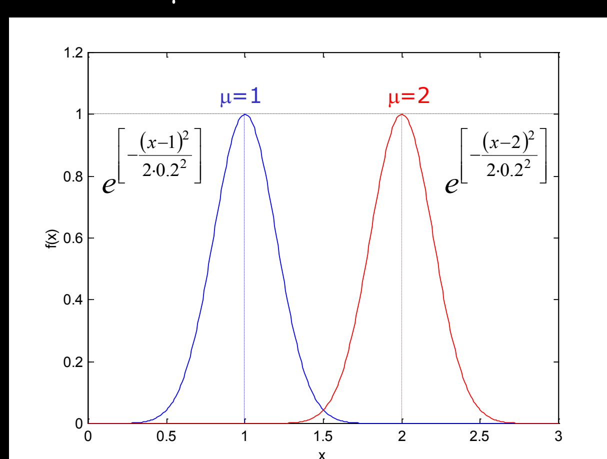
$$f(\mu-\Delta x)=\exp\left[-\frac{[(\mu-\Delta x)-\mu]^2}{2\sigma_x^2}\right]=\exp\left[-\frac{(-\Delta x)^2}{2\sigma_x^2}\right]$$

$$f(\mu+\Delta x)=e^{\left[-\frac{\Delta x^2}{2\sigma^2}\right]}=f(\mu-\Delta x)$$

Università degli Studi di Bari - Dipartimento di Chimica

Grafico

La gaussiana è quindi una funzione simmetrica, centrata sul suo valore massimo x=μ



Università degli Studi di Bari - Dipartimento di Chimica

Ampiezza a metà altezza

Calcoliamo il valore della funzione $f(x)$ nei punti $x = \mu \pm \sigma$:

$$f(\mu \pm \sigma) = \exp\left[-\frac{[(\mu \pm \sigma_x) - \mu]^2}{2\sigma_x^2}\right] =$$

$$= \exp\left[-\frac{(\pm \sigma_x)^2}{2\sigma_x^2}\right] = \exp\left[-\frac{1}{2}\right] = 0.6065$$

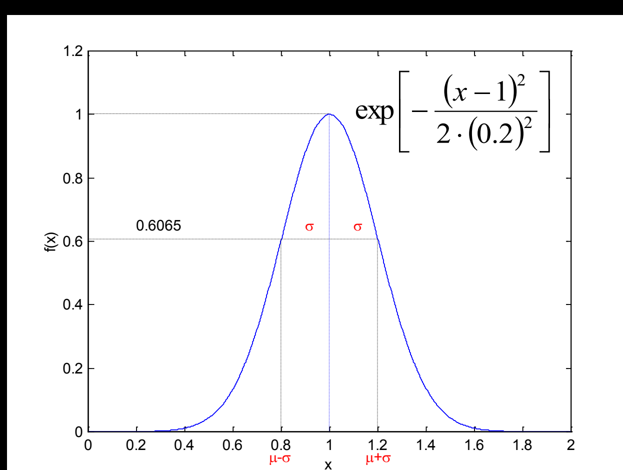
La semi ampiezza della curva ad una altezza di circa il 61% del suo valore massimo rappresenta proprio il valore del parametro σ.

F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

Ampiezza a metà altezza

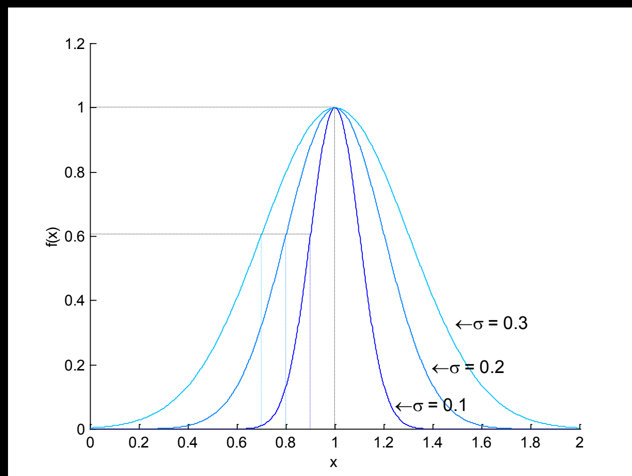
Università degli Studi di Bari - Dipartimento di Chimica

σ tiene quindi conto della larghezza della curva intorno al valore massimo



Parametro σ

Più σ e piccolo più la curva è stretta, più σ è elevato più la curva risulta slargata



F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

Università degli Studi di Bari - Dipartimento di Chimica

Integrale

L'integrale della gaussiana esteso a tutto l'intervallo di definizione della variabile indipendente ha un valore finito:

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{+\infty} \exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right]dx = \sqrt{2\pi}\sigma_x$$

Dimostrazione:

la dimostrazione viene fatta utilizzando il seguente integrale notevole

$$\int_{-\infty}^{+\infty} \exp\left[-\frac{z^2}{2}\right]dz = \sqrt{2\pi}$$

F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

Università degli Studi di Bari - Dipartimento di Chimica

Dimostrazione

Per cui effetuando un cambiamento di variabile:

$$z = \frac{x-\mu}{\sigma_x} \rightarrow x = z\sigma_x + \mu \rightarrow dx = \sigma_x dz$$

e ricalcolando gli estremi di integrazione:

$$x = \pm\infty \rightarrow z = \pm\infty$$

l'integrale della funzione gaussiana esteso a tutto l'intervallo di definizione può essere facilmente calcolato:

$$\int_{-\infty}^{+\infty} \exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right] dx \underset{dx=\sigma_x dz}{\overset{z=\left(\frac{x-\mu}{\sigma_x}\right)}{=}} \int_{-\infty}^{+\infty} \exp\left[-\frac{z^2}{2}\right] \sigma_x dz =$$

$$= \sigma_x \int_{-\infty}^{+\infty} \exp\left[-\frac{z^2}{2}\right] dz = \sigma_x \sqrt{2\pi}$$

F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

Università degli Studi di Bari - Dipartimento di Chimica

Gaussiana normalizzata

La funzione di Gauss $f(x)$ può essere quindi *normalizzata* dividendola per il valore dell'integrale esteso a tutto il campo di definizione:

$$N_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right]$$

La funzione $N_{\mu,\sigma}$ prende anche il nome di distribuzione normale.

I pedici μ e σ servono ad indicare i valori dei parametri della funzione per distinguerli dalla variabile indipendente *x*.

Funzione normalizzata

$$x \in \Re \rightarrow g(x) \in \Re$$

Normalizzare una generica funzione g(x) significa moltiplicare la funzione stessa per una costante C in modo che l'integrale della funzione, esteso a tutto l'intervallo di definizione della variabile x, risulti = 1.

$$\int_{-\infty}^{+\infty} Cg(x)dx = 1 \Rightarrow C\int_{-\infty}^{+\infty} g(x)dx = 1 \Rightarrow C = \frac{1}{\int_{-\infty}^{+\infty} g(x)dx}$$

$$g_N(x) = Cg(x) = \frac{g(x)}{\int_{-\infty}^{+\infty} g(y)dy}$$

Funzione Normalizzata

Gaussiana

Nel caso della gaussiana si ha infatti:

$$N_{\mu,\sigma}(x) = \frac{\exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right]}{\int_{-\infty}^{+\infty} \exp\left[-\frac{(y-\mu)^2}{2\sigma_x^2}\right]dy} = \frac{\exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right]}{\sqrt{2\pi}\sigma_x}$$

La Gaussiana Normalizzata gode quindi della proprietà che l'integrale esteso a tutto l'intervallo di definizione è unitario

$$\int_{-\infty}^{+\infty} N_{\mu,\sigma}(x)dx = \frac{\int_{-\infty}^{+\infty} \exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right]dx}{\sqrt{2\pi}\sigma_x} = \frac{\sqrt{2\pi}\sigma_x}{\sqrt{2\pi}\sigma_x} = 1$$

Gaussiana

La funzione gaussiana* soddisfa quindi le proprietà di una funzione densità di probabilità:

$\mathrm{N}_{\mu,\sigma_x}(x) \geq 0$ è sempre positiva o nulla

$\int_{-\infty}^{+\infty} \mathrm{N}_{\mu,\sigma_x}(x)dx = 1$ è normalizzata, cioè l'integrale esteso a tutto l'intervallo di definizione è uguale ad 1

Da qui in avanti con il termine gaussiana si intenderà la funzione di Gauss normalizzata.

Teorema di Laplace

La densità di probabilità di misure ripetute x dal valore vero μ in assenza di errori sistematici ed in presenza di errori random che siano:

- dovuti a cause di disturbo statisticamente indipendenti,
- equamente distribuiti in eccesso o in difetto,

si dimostra essere proprio la funzione gaussiana:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right]$$

Gaussiana

Quindi la funzione densità di probabilità gaussiana è proprio la funzione limite ottenuta facendo infinite misure ripetute:

$$\lim_{\substack{N\to\infty \\ \Delta x\to 0}} \frac{F_k}{\Delta x} = \lim_{\substack{N\to\infty \\ \Delta x\to 0}} f_k = \mathrm{N}_{\mu,\sigma_x}(x)$$

se siamo in assenza di errori sistematici e in presenza di cause di disturbo indipendenti.

Valor Medio

Abbiamo visto che data una funzione densità di probabilità $p(x)$ il valor medio della variabile x può essere calcolato come:

$$\langle x \rangle = \int x \cdot p(x) \cdot dx$$

nel caso della funzione gaussiana questo integrale diventa:

$$\langle x \rangle = \int_{-\infty}^{+\infty} x \cdot \mathrm{N}_{\mu,\sigma_x}(x) dx = \mu$$

il parametro μ è proprio il valor medio della distribuzione

Valor Medio

Dimostrazione

$$\langle x\rangle = \int_{-\infty}^{+\infty} x \cdot N_{\mu,\sigma_x}(x)dx = \int_{-\infty}^{+\infty} x \frac{\exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right]}{\sqrt{2\pi}\sigma_x}dx = \frac{I}{\sqrt{2\pi}\sigma_x}$$

dobbiamo quindi risolvere il seguente integrale I:

$$I = \int_{-\infty}^{+\infty} x \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma_x^2}\right]dx$$

Effettuiamo il cambio della variabile di integrazione:

$$z = \frac{x-\mu}{\sigma_x} \rightarrow x = z\sigma_x + \mu \rightarrow dx = \sigma_x dz$$

e ricalcolando gli estremi di integrazione:

$$x = \pm\infty \rightarrow z = \pm\infty$$

F.Mavelli- Laboratorio Chimica Fisica - a.a. 2013-2014

Università degli Studi di Bari - Dipartimento di Chimica

Valor Medio

si ottiene

$$I = \int_{-\infty}^{+\infty} (\sigma_x z + \mu) \cdot \exp\left[-\frac{z^2}{2}\right]\sigma_x dz =$$

$$= \sigma_x^2 \int_{-\infty}^{+\infty} z \exp\left[-\frac{z^2}{2}\right]dz + \sigma_x\mu \int_{-\infty}^{+\infty} \exp\left[-\frac{z^2}{2}\right]dz = \mu\sqrt{2\pi}\sigma_x$$

$$\int_{-\infty}^{+\infty} z \exp\left[-\frac{z^2}{2}\right]dz = 0$$

Integrale di una funzione dispari esteso ad un intervallo simmetrico

$$\int_{-\infty}^{+\infty} \exp\left[-\frac{z^2}{2}\right]dz = \sqrt{2\pi}$$

Integrale notevole

Valor Medio

da cui si ha il risultato cercato:

$$\langle x \rangle = \frac{I}{\sqrt{2\pi}\sigma_x}$$

$$I = \mu\sqrt{2\pi}\sigma_x$$

$$\langle x \rangle = \int_{-\infty}^{+\infty} x \cdot N_{\mu,\sigma_x}(x)dx = \mu$$

Il valor medio della distribuzione gaussiana corrisponde al parametro μ ossia al massimo della distribuzione

Varianza

Abbiamo visto che data una funzione densità di probabilità *p*(x) la varianza della variabile x può essere calcolata come:

$$Var\{x\} = \int \left(x-\langle x \rangle\right)^2 \cdot p(x)\ dx = \langle x^2 \rangle - \langle x \rangle^2$$

nel caso della funzione gaussiana questo integrale si dimostra essere uguale:

$$\langle x^2 \rangle - \langle x \rangle^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot N_{\mu,\sigma_x}(x)\,dx = \sigma_x^2$$

il parametro σ_x^2 è proprio la varianza della distribuzione σ_x la deviazione standard