

# Trattazione statistica dei dati

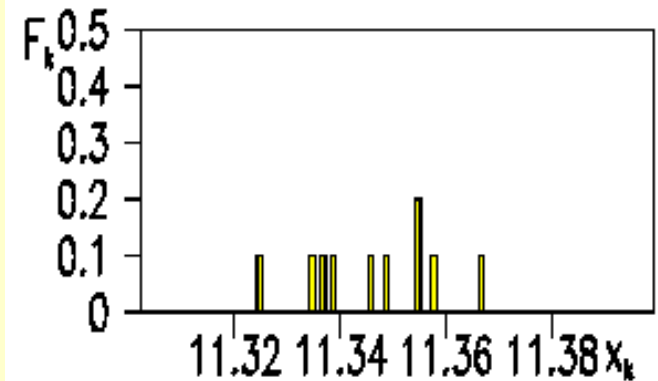
statistica matematica

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

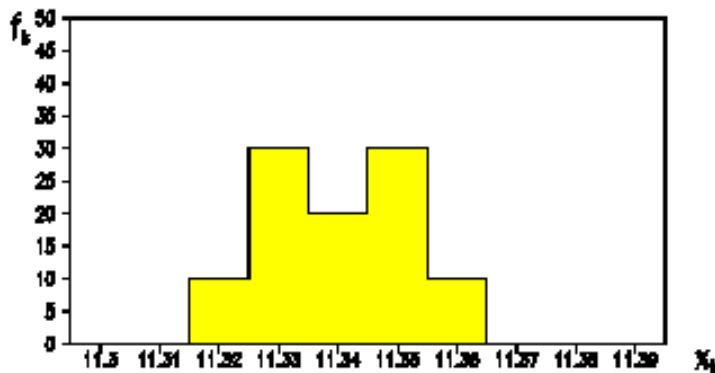
$N-1$  gradi di libertà

calibro con errore di sensibilità 0.001 cm

L'altezza delle barre non è più sufficiente a caratterizzare la distribuzione



istogramma ad intervalli



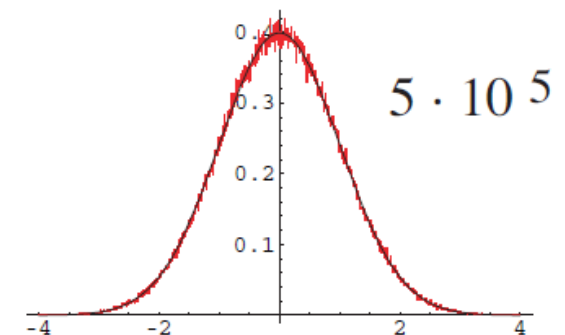
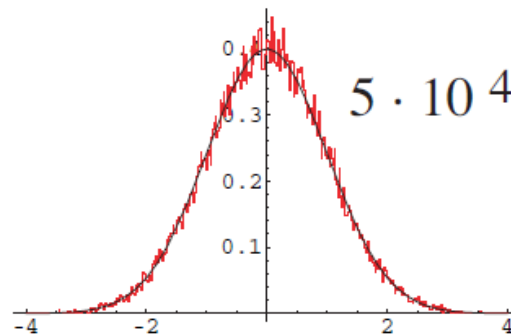
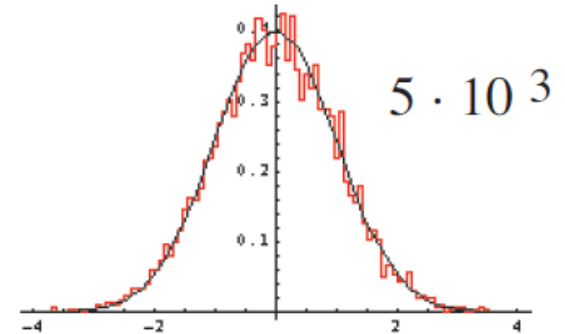
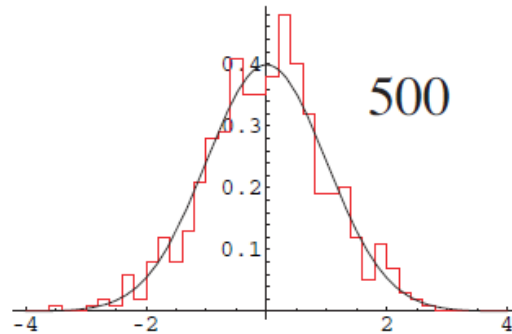
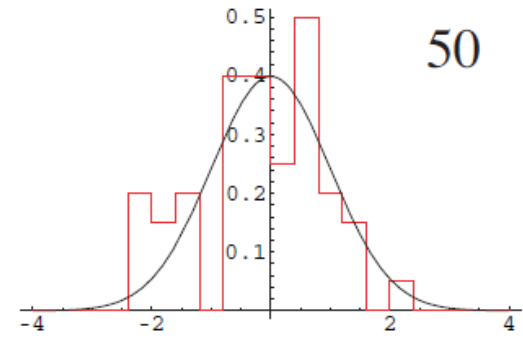
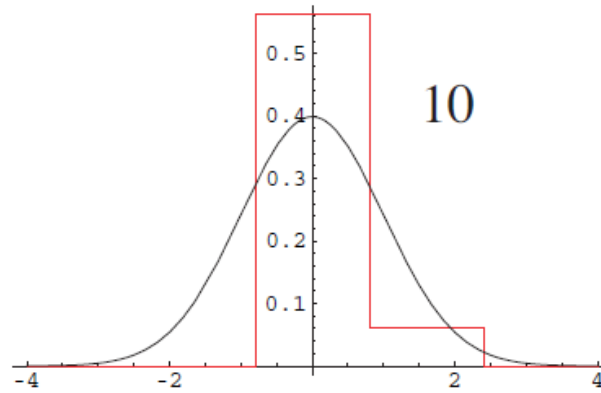
$$F_k = \frac{n_k}{N}$$

$f_k \Delta_k$  = frazione delle misure nell'intervallo  $k$ -esimo

# Trattazione statistica dei dati

All'aumentare del numero delle misure l'istogramma ad intervalli assumerà una forma più regolare e potremo anche ridurre la larghezza degli intervalli.

In tal modo l'istogramma si avvicinerà sempre più (senza però mai arrivarci) ad una curva continua



# Probabilità in esperimenti casuali

Esperimento casuale: dà un risultato diverso, non prevedibile, ad ogni misura

Variabile casuale: risultato dell'esperimento casuale

Probabilità  $P(\varepsilon)$  di un evento casuale  $\varepsilon$ :

$N$  misure,  $n_\varepsilon$  volte si verifica l'evento  $\varepsilon \rightarrow P(\varepsilon) = \lim_{N \rightarrow \infty} (n_\varepsilon / N)$

Alcune proprietà della probabilità:

1)  $P(S) = 1$  se  $S$  è l'insieme di tutti gli eventi possibili

2) Se  $A$  e  $B$  sono due sottinsiemi di  $S$  definiamo  $P(AB)$  la probabilità che si verifichino contemporaneamente  $A$  e  $B$ , ovvero che l'evento  $\varepsilon$  appartenga sia ad  $A$  che a  $B$ .  
Definiamo  $A$  e  $B$  mutuamente disgiunti se  $P(AB) = 0$  [ex: diversi risultati lancio dado]

3) Se  $A$  e  $B$  sono mutuamente disgiunti, la probabilità che il risultato dell'esperimento appartenga ad  $A$  o a  $B$   $P(A+B) = P(A)+P(B)$  [ex: Probabilità di avere 1 o 2 nel lancio del dado]

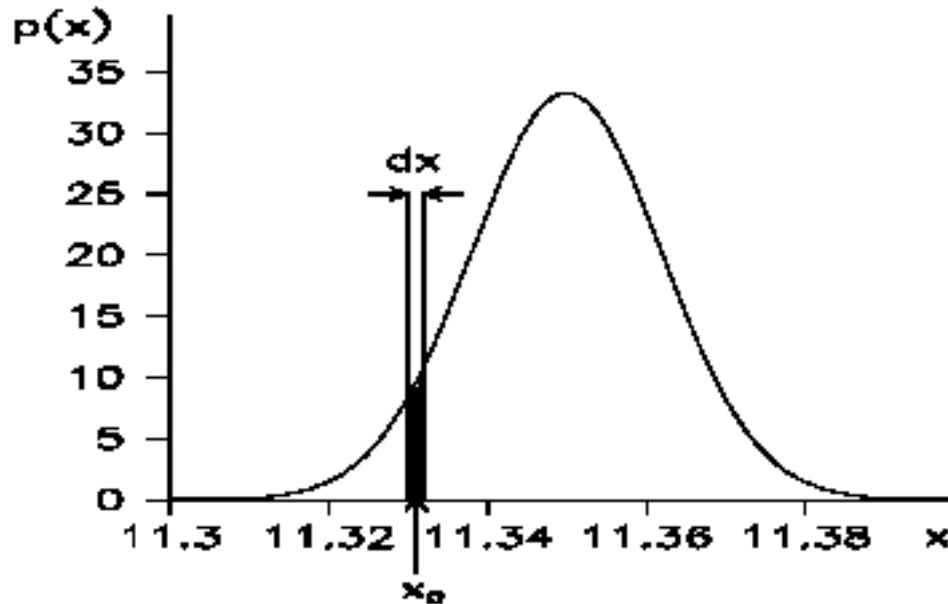
4) Probabilità condizionale  $P(A|B) = P(AB)/P(B)$  ovvero la probabilità che si verifichi  $A$  negli eventi per cui si conosce già che è verificato  $B$  [ex.: in un lancio del dado,  $B$  risultato pari,  $A$  risultato minore di 4  $\rightarrow P(A|B) = (1/6)/(1/2) = (1/3)$ ]

5) Definizione di indipendenza: gli eventi  $A$  e  $B$  sono mutuamente indipendenti se e solo se  $P(A|B)=P(A)$ , da cui segue che  $P(AB)=P(A)*P(B)$

# Trattazione statistica dei dati

Dalla statistica matematica → **“funzione limite”**  $p(x)$  = rappresentazione nel caso di un numero infinito di eventi

$p(x)$  = **“densità di probabilità”** →  $p(x_i) \cdot dx$  = probabilità che una singola misura dia un risultato tra  $x_i$  e  $x_i + dx$



Nel caso siano presenti molte sorgenti di errore accidentale (ognuna delle quali dia un contributo piccolo) la funzione limite prende il nome di **“distribuzione normale”** o **“funzione di Gauss”** e assume la forma:

$$p(x) = K e^{-\frac{(x-X)^2}{2\sigma^2}}$$

$K$  è il fattore di normalizzazione

$\sigma$  parametro di larghezza

# Trattazione statistica dei dati

## Distribuzione normale o funzione di Gauss

$$p(x) = K e^{-\frac{(x-X)^2}{2\sigma^2}}$$

$$\int_{-\infty}^{+\infty} p(x) dx = \int_{-\infty}^{+\infty} K e^{-\frac{(x-X)^2}{2\sigma^2}} dx = 1 \quad \rightarrow \quad K = \frac{1}{\sqrt{2\pi} \sigma}$$

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

$$p_{X,\sigma}(x)$$

$p_{X,\sigma}(x) \geq 0$  per ogni  $x$

$$\begin{array}{lll} p'_{X,\sigma}(x) = 0 & \Rightarrow & x = X \\ & \Rightarrow & x \rightarrow \pm\infty \\ & \Rightarrow & p_{X,\sigma}(x) \rightarrow 0 \\ & \Rightarrow & p_{X,\sigma}(X) = \frac{1}{\sqrt{2\pi} \sigma} \\ & \Rightarrow & p''_{X,\sigma}(X) < 0 \end{array}$$

$$p_{X,\sigma}(x_{1/2}) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x_{1/2}-X)^2}{2\sigma^2}} = \frac{1}{2} \frac{1}{\sqrt{2\pi} \sigma} \quad \rightarrow \quad (x_{1/2} - X) = \pm 1.18 \sigma$$

$$\bar{x} = \int_{-\infty}^{+\infty} x p_{X,\sigma}(x) dx = X$$

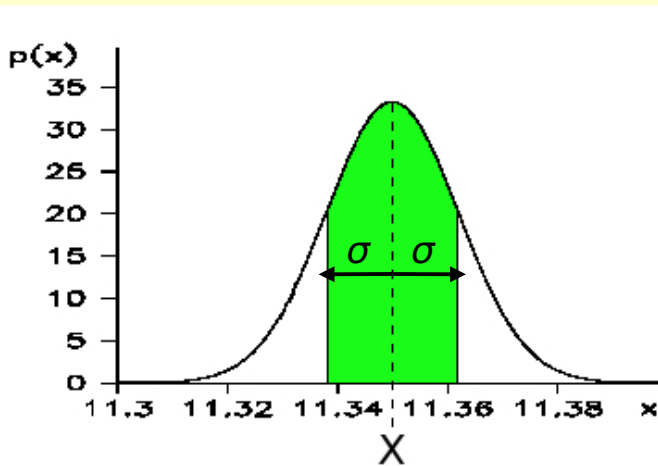
$$\sigma_x^2 = \int_{-\infty}^{+\infty} (x - X)^2 p_{X,\sigma}(x) dx = \sigma^2$$

Valori di aspettazione

# Trattazione statistica dei dati

## Distribuzione normale o funzione di Gauss

probabilità  $P(X, \sigma)$  che una singola misura dia un risultato tra  $X - \sigma$  e  $X + \sigma$



$$P(X, \sigma) = \int_{X-\sigma}^{X+\sigma} p_{X,\sigma}(x) dx = \frac{1}{\sqrt{2\pi} \sigma} \int_{X-\sigma}^{X+\sigma} e^{-\frac{(x-X)^2}{2\sigma^2}} dx$$

o, più in generale, la  $P(X, t\sigma)$ :

$$P(X, t\sigma) = \int_{X-t\sigma}^{X+t\sigma} p_{X,\sigma}(x) dx = \frac{1}{\sqrt{2\pi} \sigma} \int_{X-t\sigma}^{X+t\sigma} e^{-\frac{(x-X)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-\frac{z^2}{2}} dz$$

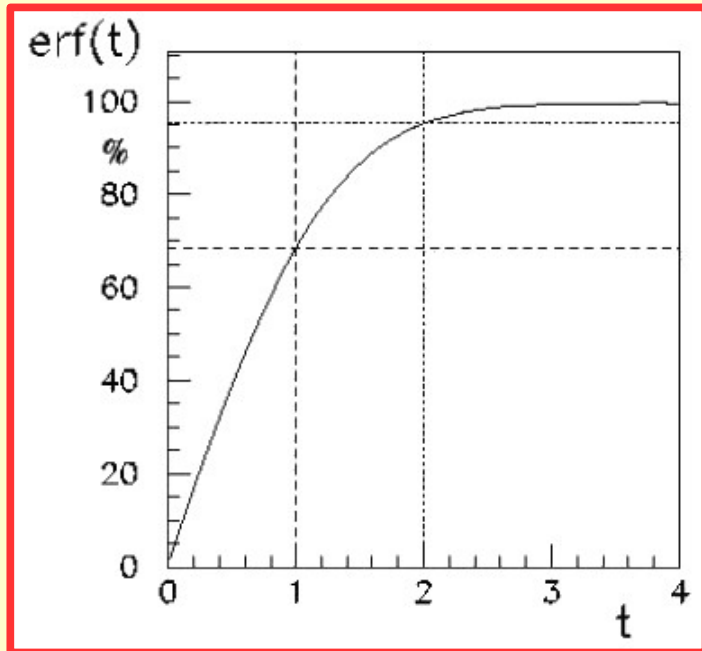
$$z = (x - X)/\sigma$$

funzione degli errori

$\text{erf}(t)$

# Trattazione statistica dei dati

## Funzione degli errori



$t$	0.00	0.02	0.04	0.06	0.08
0.0	0.00	1.60	3.19	4.78	6.38
0.1	7.97	9.55	11.13	12.71	14.28
0.2	15.85	17.41	18.97	20.51	22.05
0.3	23.58	25.10	26.61	28.12	29.61
0.4	31.08	32.55	34.01	35.45	36.88
0.5	38.29	39.69	41.08	42.25	43.81
0.6	45.15	46.47	47.78	49.07	50.35
0.7	51.61	52.85	54.07	55.27	56.46
0.8	57.63	58.78	59.91	61.02	62.11
0.9	63.19	64.24	65.28	66.29	67.29
1.0	68.27	69.23	70.17	71.09	71.99
1.1	72.87	73.73	74.57	75.40	76.20
1.2	76.99	77.75	78.50	79.23	79.95
1.3	80.64	81.32	81.98	82.62	83.24
1.4	83.85	84.44	85.01	85.57	86.11
1.5	86.64	87.15	87.64	88.12	88.59
1.6	89.04	89.48	89.90	90.31	90.70
1.7	91.09	91.46	91.81	92.16	92.49
1.8	92.81	93.12	93.42	93.71	93.99
1.9	94.26	94.51	94.76	95.00	95.23
2.0	95.45	95.66	95.86	96.06	96.25
2.1	96.43	96.60	96.76	96.92	97.07
2.2	97.22	97.36	97.49	97.62	97.74
2.3	97.86	97.97	98.07	98.17	98.27
2.4	98.36	98.45	98.53	98.61	98.69
2.5	98.76	98.83	98.89	98.95	99.01
2.6	99.07	99.12	99.17	99.22	99.26
2.7	99.31	99.35	99.39	99.42	99.46
2.8	99.49	99.52	99.55	99.58	99.60
2.9	99.63	99.65	99.67	99.69	99.71

# Trattazione statistica dei dati

Miglior stima del valore “vero” e del parametro di larghezza della funzione di Gauss

Lo sperimentatore non avrà mai la conoscenza esatta della densità di probabilità delle variabili su cui opera (il che richiederebbe un numero infinito di misure).

Lo sperimentatore ha a disposizione solo un campione di ampiezza  $N$  di dati misurati  $x_1, x_2, \dots, x_N$



Potrà quindi dare solo una stima approssimata dei parametri della distribuzione

Per stimare dal campione di dati un parametro della distribuzione si dovrà applicare al campione una funzione, che chiameremo **stimatore del parametro**. Tale stimatore, essendo una funzione di variabile casuale, sarà a sua volta una variabile casuale con una sua distribuzione.

Perché una funzione del campione  $s(x_1, x_2, \dots, x_N)$  possa essere considerata uno stimatore valido deve soddisfare alle seguenti proprietà:

- 1) **assenza di polarizzazione**: uno stimatore di un parametro  $z$  si dice non polarizzato se il suo valore di aspettazione coincide con  $z$  per qualsiasi valore di  $N$  (asintoticamente non polarizzato se tale coincidenza vale solo per  $N \rightarrow \infty$ )
- 2) **consistenza**: si chiede che lo stimatore individui, al crescere di  $N$ , il parametro cercato con sempre maggiore precisione (distribuzione dello stimatore sempre più stretta all'aumentare di  $N$ )



# Trattazione statistica dei dati

Miglior stima del valore “vero” e del parametro di larghezza della funzione di Gauss

Qual'è la miglior stima di  $X$  e  $\sigma$  ottenibile con  $N$  dati misurati  $x_1, x_2, \dots, x_N$ ????

Probabilità di ottenere una misura tra  $x_1$  e  $x_1+dx$

$$P(x_1, x_1 + dx) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x_1-X)^2}{2\sigma^2}} dx \propto \frac{1}{\sigma} e^{-\frac{(x_1-X)^2}{2\sigma^2}}$$

Probabilità di ottenere una misura tra  $x_2$  e  $x_2+dx$

$$P(x_2, x_2 + dx) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x_2-X)^2}{2\sigma^2}} dx \propto \frac{1}{\sigma} e^{-\frac{(x_2-X)^2}{2\sigma^2}}$$

Probabilità di ottenere la serie di valori  $x_1, x_2, \dots, x_N$

$$P(x_1, \dots, x_N) = P(x_1, x_1 + dx) P(x_2, x_2 + dx) \dots P(x_N, x_N + dx) \propto \frac{1}{\sigma^N} e^{-\frac{\sum_{i=1}^N (x_i - X)^2}{2\sigma^2}}$$

principio di massima verosimiglianza

Avendo realizzato una serie di  $N$  misure della grandezza ed avendo ottenuto i valori  $x_1, x_2, \dots, x_N$ , la probabilità di ottenere tale serie di valori deve essere massima (e quindi le migliori stime di  $X$  e  $\sigma$  sono quelle che la rendono massima)

$$\text{Miglior stima di } X \rightarrow (\partial P / \partial X) = 0 \rightarrow \frac{\sum_{i=1}^N (x_i - X)^2}{2\sigma^2} = \text{minimo} \Rightarrow \frac{\partial \sum_{i=1}^N (x_i - X)^2}{\partial X} = 0 \Rightarrow X = \frac{\sum_{i=1}^N x_i}{N} = \bar{x}$$

$$\text{Miglior stima di } \sigma \rightarrow (\partial P / \partial \sigma) = 0 \rightarrow \sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - X)^2}{N}} \rightarrow \dots \rightarrow \sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}} = \sigma_x$$

# Trattazione statistica dei dati

## Critero di Chauvenet

Criterio oggettivo per scartare misure sospette in una serie di  $N$  dati misurati  $x_1, x_2, \dots, x_N$

10.1, 10.0, 10.2, 10.3, 10.2, 10.1, 10.0, 10.2, 10.1, 10.2, 10.1, 10.3, 11.0, 11.5, 11.6

Primo passo (se possibile): ripetere le misure per vedere se le “anomalie” si ripresentano

Secondo passo (se il primo non è possibile o non ha dato risultati):

a) si determinano media e deviazione standard →

$$\bar{x} \simeq 10.4 \text{ mm}, \quad \sigma_x \simeq 0.5 \text{ mm}$$

b) si determina la deviazione dalla media della misura più sospetta →

$$d = 1.2 \text{ mm} \text{ ovvero } 2.4 \sigma_x$$

c) si determina la probabilità di ottenere misure che differiscano dal valore “vero” almeno di questa quantità →

$$\begin{aligned} P(\text{al di fuori di } 2.4 \sigma_x) &= 1 - P(\text{entro } 2.4 \sigma_x) = \\ &= 1 - 0.984 = 0.016 \simeq \frac{1}{63} \end{aligned}$$

d) si applica il Criterio di Chauvenet che dice:

“Se il numero atteso di misure con deviazione rispetto al valore medio maggiore o uguale a quello della misura sospetta è minore di 0.5, la misura sospetta può essere scartata”

e) si scarta la misura sospetta e si riparte dal punto a) → **NOOOO!**

Il Criterio di Chauvenet può essere applicato 1 sola volta e la sua applicazione deve portare allo scarto di un numero ridotto di misure!