Elementi di Teoria della probabilità.

A cura del Prof. Massimiliano Ferrara

Eventi e variabili casuali.

Oggetto della teoria delle probabilità è lo studio dei *fenomeni casuali* o *aleatori*, ossia quei fenomeni ripetibili che possono manifestarsi in più *modalità*, singolarmente imprevedibili.
L'insieme delle modalità costituisce lo *spazio dei risultati S*, e può essere costituito da un numero finito o infinito di elementi.
Si definisce *evento casuale* l'associazione di una o più di queste modalità.

Es: nel lancio di un dado vi sono sei modalità : 1, 2, 3, 4, 5, 6. Un evento casuale che si può definire è *l'uscita di un numero dispari*

L'insieme delle parti di S, $\mathcal{P}$(S*)* è l'insieme di tutti gli eventi casuali.
Se si riesce ad associare ad ogni evento un numero reale x, x prende il nome di *variabile casuale* su S.
Le variabili casuali possono essere finite o infinite discrete o continue.

Es: per la presenza degli errori, la misura di una grandezza può essere considerata un evento casuale. Il risultato della misura è una variabile casuale che può essere associato all'evento stesso.

La probabilità.

Definizione classica

Si definisce *probabilità* di un evento casuale il rapporto tra i casi favorevoli all'evento stesso e quelli possibili, purchè tutti equiprobabili. La probabilità è, quindi, un numero compreso tra zero ed uno, zero per l'evento impossibile ed uno per quello certo.
La definizione precedente, detta definizione *classica* di probabilità, ci permette di calcolare semplicemente le probabilità di semplici eventi casuali (nel caso finito), ma è intrinsecamente insoddisfacente. Infatti presuppone il concetto di equiprobabilità mentre definisce la probabilità (equiprobalità difficilmente individuabile nel caso di variabili continue).

Per questo motivo si è cercato di dare altre definizione di probabilità.

Definizione Empirica

Definiamo frequenza relativa *f(E)* di un evento casuale *E* , il rapporto tra il numero *n* di volte in cui E si è presentato e il numero *N* di casi reali osservati

$$f(E) = \frac{n}{N} \qquad (1)$$

la probabilità di *E*, *p(E),* si definisce come il

$$p(E) = \lim_{N\to\infty} f(E) = \lim_{N\to\infty}\left(\frac{n}{N}\right) \qquad (2)$$

dove il limite è, ovviamente, un limite in probabilità.
In verità anche questa definizione è insoddisfacente. Infatti non è detto, a priori, che la frequenza relativa debba convergere ed inoltre si basa sulla prova pratica.
Esistono altre definizioni di probabilità logicamente consistenti a cui accenneremo più avanti; per adesso ci baseremo sulla definizione empirica sia perché alcune dimostrazioni sono più semplici sia perché al limite (2) si giungerà anche attraverso una definizione *assiomatica* di probabilità.
Consideriamo due eventi *E* ed il suo complementare $\overline{E}$, tali che $E \cap \overline{E} = \phi$ ed $E \cup \overline{E} = S$.
Se la frequenza di *E* su *N* prove è *n* quella di $\overline{E}$ sarà

$$f(\overline{E}) = \frac{N-n}{N} = 1 - \frac{n}{N} = 1 - f(E)$$

e per la definizione empirica di probabilità si ha

$$p(\overline{E}) = 1 - p(E) \Rightarrow p(\overline{E}) + p(E) = 1$$

risultato che si estende facilmente al caso di più eventi concorrenti A, B,….Z

$$p(A) + p(B) + ... + p(Z) = 1 \qquad (4)$$

se gli eventi soddisfano le condizioni:

$$A \cap B \cap .... \cap Z = \phi \quad \text{e} \quad A \cup B \cup .... \cup Z = S$$

Consideriamo ora gli eventi *E* ed *F* ed i loro complementari $\overline{E}$ ed $\overline{F}$, se su *N* prove *l'evento composto EF* (prodotto logico dei due eventi semplici) si è manifestato nei quattro eventi fondamentali possibili *EF*, $E\overline{F}$, $\overline{E}F$, $\overline{E}\overline{F}$ con le frequenze assolute riportate nella tabella seguente

	F	$\overline{F}$
E	n_{11}	n_{12}
$\overline{E}$	n_{21}	n_{22}

Tab.1
allora si può scrivere:

$$f(EF) = \frac{n_{11}}{N}; f(E\overline{F}) = \frac{n_{12}}{N}$$
$$f(\overline{E}F) = \frac{n_{21}}{N}; f(\overline{E}\overline{F}) = \frac{n_{22}}{N}$$

ovvero

$$f(E) = \frac{n_{11} + n_{12}}{N} = f(EF) + f(E\overline{F})$$

$$f(F) = \frac{n_{11} + n_{21}}{N} = f(EF) + f(\overline{E}F)$$

e sempre in base alla definizione empirica di probabilità

$$p(E) = p(EF) + p(E\overline{F})$$
$$p(F) = p(EF) + p(\overline{E}F)$$

ed analogamente per $p(\overline{E})$ e $p(\overline{F})$.
Consideriamo l'evento *E+F* somma logica degli eventi semplici, definito come l'evento casuale consistente nel verificarsi di *E* o di *F* o di entrambi, dalla tabella 1 si ha

$$f(E+F) = \frac{n_{11} + n_{12} + n_{21}}{N} = \frac{(n_{11} + n_{12}) + (n_{11} + n_{21}) - n_{11}}{N} = f(E) + f(F) - f(EF)$$

ed in termini di probabilità

$$p(E+F) = p(E) + p(F) - p(EF) \quad (5)$$

nel caso E ed F si escludono mutuamente ($p(EF)=0; n_{11}=0$) vale la cosiddetta *legge della probabilità totale*

$$p(E+F) = p(E) + p(F)$$

che può generalizzarsi per induzione a più eventi esclusivi

$$p(A+B+....+Z) = p(A) + p(B) + ... + p(Z) \quad (6)$$

Possiamo, ora, introdurre il concetto di *probabilità condizionata* ossia per esempio la probabilità che si verifichi l'evento *E* nel caso in cui si sia già verificato l'evento *F*, probabilità che indicheremo col simbolo *p(E/F)*. Si ricava facilmente che

$$f(E/F) = \frac{n_{11}}{n_{11} + n_{21}} = \frac{n_{11}}{N} \frac{N}{n_{11} + n_{21}} = \frac{f(EF)}{f(F)}$$

analogamente $$f(F/E) = \frac{f(EF)}{f(E)}$$

da cui $$f(EF) = f(F) \cdot f(E/F) = f(E) \cdot f(F/E)$$

che passando al limite diventa

$$p(EF) = p(F) \cdot p(E/F) = p(E) \cdot p(F/E) \quad (7)$$

e se gli eventi E ed F sono indipendenti, cioè se è

$$p(E/F) = p(E) \quad \text{e} \quad p(F/E) = p(F)$$

allora vale la cosiddetta legge delle probabilità composta

$$p(EF) = p(E) \cdot p(F) \qquad (8)$$

che può generalizzarsi ad un numero qualsiasi di elementi indipendenti

Es: lancio di due dadi
Evento A: numero dispari sul primo dado
Evento B: numero dispari sul secondo dado
Evento C: punteggio complessivo dispari.
A e B sono ovviamente indipendenti.
Poiché risulta
$p(C/A) = p(C) = 1/2$ e
$p(C/B) = p(C) = 1/2$
anche A e C ed B e C sono indipendenti.
I tre eventi non lo sono nel loro complesso dato che il verificarsi di A e B esclude C

Definizione assiomatica di probabilità

Sia S l'insieme di tutti i possibili risultati di un fenomeno casuale, ed E un qualsiasi evento casuale $(E \subseteq S)$. Si definisce probabilità di E quel numero $p(E)$ associato ad E, tale che:

1) $p(E) \geq 0$
2) $p(S) = 1$
3)* $p(E_1 \cup E_2 \cup ....) = p(E_1) + p(E_2) + ...$
per ogni insieme finito o infinito di eventi $E_1, E_2, E_3,\dots$ e tali che
$E_i \cap E_j = \phi \quad \forall i \neq j$

*l'assioma 3 ha senso solo se l'insieme dei possibili risultati è finito o numerabile.

Dai tre assiomi è possibile ricavarsi tutto quanto è necessario comprese la legge della probabilità totale e composta che abbiamo ricavato a partire dalla definizione empirica. D'altra parte come abbiamo detto è possibile dimostrare, partendo dalla definizione assiomatica di probabilità la legge della convergenza della frequenza alla probabilità.
In seguito torneremo su questo argomento. Torniamo, invece, sul concetto di *variabile casuale* per adesso unidimensionale discreta.
In verità il caso più frequente in Fisica è quello di una variabile casuale continua, anche se a causa della sensibilità limitata degli strumenti l'intervallo continuo della variabile viene suddiviso in un numero finito di intervalli.

Variabili casuali discrete

Valore atteso o Speranza Matematica

Consideriamo la variabile casuale x discreta. Sia ν_j la frequenza assoluta con cui si è presentato il risultato x_j in N, risulterà

$$\sum_{j=1}^{M} \nu_j = N$$

e posto $f_j = \frac{\nu_j}{N}$ si ricava $\sum_{j=1}^{M} f_j = \sum_{j=1}^{M} \frac{\nu_j}{N} = \frac{1}{N} \sum_{j=1}^{M} \nu_j = 1$.

Come abbiamo visto la media aritmetica di N per un campione di N valori di una variabile casuale può essere espressa come

$$\bar{x} = \sum_{j=1}^{M} f_j \cdot x_j$$

Allo stesso modo può essere definita una nuova grandezza *E(x),* relativa all'intera popolazione, mediante la seguente relazione

$$E(x) = \sum_{j=1}^{M} p_j \cdot x_j \qquad (9)^*$$

*se l'insieme degli *xi* è infinito, o anche se la variabile casuale è continua non è detto che esista.

E(x) è nota come *speranza matematica* o *valore atteso* della variabile casuale *x* . Essa appare come una generalizzazione alla popolazione del concetto di media aritmetica di un campione. D'altra parte se si assumesse come definizione di probabilità quella empirica, *E(x)* rappresenterebbe il limite della media aritmetica del campione all'aumentare della sua dimensione, per tale motivo spesso viene anche detta *valor medio della popolazione*.
Analogamente si definisce la speranza matematica o valore atteso della variabile casuale $[x\text{-}E(x)]^2$

$$E\left\{[x - E(x)]^2\right\} = \sum_{i=1}^{N} p_i \cdot [x_i - E(x)]^2 \qquad (10)^*$$

*se l'insieme degli *xi* è infinito, o anche se la variabile casuale è continua non è detto che esista.

che può essere considerato la generalizzazione alla popolazione della varianza di un campione ed è appunto per questo che viene anche detto *varianza della popolazione* della variabile casuale *x* ed indicato con *Var(x)*.

Dimostriamo ora una serie di proprietà che ci saranno utili in seguito.

<u>*Combinazioni lineari di variabili casuali*</u>
Date due variabili casuali *x* e *y* ed una loro combinazione lineare $z=ax+by$ allora accade che

$$E(z)=aE(x)+bE(y)$$

Indichiamo con x_i i possibili valori di *x* e con y_j quelli di *y* e con p_i e q_j le rispettive probabilità. Indichiamo inoltre con *Pij* la probabilità che si abbia simultaneamente x_i ed y_j. Ora poiché un particolare valore della *x* può essere associato ad un qualsiasi valore della *y*, che sono tra loro mutamente esclusivi, ricordando la legge della probabilità totale (eq. 6 pag.3) risulterà

$$p_i = \sum_j P_{ij} \quad \text{e} \quad q_j = \sum_i P_{ij}$$

ora

$$\begin{aligned} E(z) = E(ax+by) &= \sum_{ij} P_{ij} \cdot (ax_i + by_j) = \sum_{ij} aP_{ij}x_i + \sum_{ij} bP_{ij}y_j = \\ &= a\sum_i (\sum_j P_{ij}) \cdot x_i + b\sum_j (\sum_i P_{ij}) \cdot y_j = a\sum_i p_i x_i + b\sum_j q_j y_j = \\ &= aE(x) + bE(y) \end{aligned}$$

Questo risultato può essere esteso, per induzione, alla combinazione lineare di un numero qualsiasi di variabili casuali.
Nel caso in cui le due variabili casuali *x* e *y* sono statisticamente indipendenti per cui è possibile affermare che la probabilità P_{ij} che risulti contemporaneamente $x=x_i$ e $y=y_j$ è data da (eq.8 Pag. 4)

$$P_{ij}=p_i\, q_j$$

È possibile dimostrare che la varianza della variabile casuale $z=ax+by$ è data da

$$Var(z)=a^2Var(x)+b^2Var(y)$$

Cominciamo col supporre che

$$E(x)=E(y)=0$$

e quindi

$$E(z)=aE(x)+bE(y)=0$$

ed indichiamo, per brevità, le varianze di x, y e z rispettivamente con σ_x^2, σ_y^2, e σ_z^2 . Possiamo scrivere :

$$\begin{aligned}
\sigma_z^2 &= E\left\{[z-E(z)]^2\right\} = E\left\{z^2\right\} = E\left\{[ax+by]^2\right\} = \\
&= \sum_{ij} P_{ij}(ax_i+by_j)^2 = \sum_{ij} p_i q_j (a^2x_i^2+b^2y_j^2+2abx_iy_j) = \\
&= a^2\sum_j q_j \sum_i p_i x_i^2 + b^2\sum_i p_i \sum_j q_j y_j^2 + 2ab\sum_i p_i x_i \cdot \sum_j q_j y_j = \\
&= a^2\sigma_x^2\sum_j q_j + b^2\sigma_y^2\sum_i p_i + 2ab\cdot E(x)\cdot E(y) = a^2\sigma_x^2+b^2\sigma_y^2
\end{aligned}$$

per estendere tale risultato a due variabili casuali con valore atteso diverso da zero notiamo per prima cosa che se ξ è una variabile casuale definita come

$$\xi=x+K$$

con K costante, allora risulta che

$E(x)=E(x)+K$ ed $\sigma_\xi^2 = E\left\{[\xi-E(\xi)]^2\right\} = E\left\{[x+K-E(x)-K]^2\right\} = E\left\{[x-E(x)]^2\right\} = \sigma_x^2$

Quindi, in generale, basta considerare due variabile ausiliarie

$$\xi=x-E(x) \quad ed \quad \eta=y-E(y)$$

la cui combinazione lineare

$$\theta=a\xi+b\eta$$

differisce da z per un fattore costante

$$aE(x)+bE(y)$$

per ottenere che

$$\sigma_\theta^2 = a^2\sigma_\xi^2 + b^2\sigma_\eta^2$$

e poiché

$$\sigma_\xi^2=\sigma_x^2 \qquad \sigma_\eta^2=\sigma_y^2 \qquad \sigma_\theta^2=\sigma_z^2$$

risulta anche

$$\sigma_z^2 = a^2\sigma_x^2+b^2\sigma_y^2$$

Il risultato ottenuto può essere esteso per induzione ad una qualsiasi combinazione lineare di un numero finito di variabili casuali statisticamente indipendenti.

Quindi possiamo affermare che una combinazione lineare, a coefficienti costanti, di un numero N finito di variabili casuali è una variabile casuale con valore atteso uguale alla combinazione lineare, con gli stessi coefficienti, dei valori attesi delle singole variabili

$$z = \sum_{i=1}^{N} a_i x_i \qquad\qquad E(z) = \sum_{i=1}^{N} a_i E(x_i) \qquad (11)$$

in più se le variabili sono statisticamente indipendenti, la varianza è la combinazione lineare delle rispettive varianze, con coefficienti pari ai quadrati dei rispettivi coefficienti

$$\sigma_z^2 = \sum_{i=1}^{N} a_i^2 \sigma_{xi}^2 \qquad (12)$$

Il caso della media

Importanti conseguenze possono subito essere ricavate per la media aritmetica dei valori ottenuti dalle misure ripetute di una grandezza fisica. Infatti la media aritmetica può essere riguardata come una combinazione lineare di variabili casuali

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

per cui il suo valore atteso sarà

$$E(\bar{x}) = \frac{1}{N} \sum_{i=1}^{N} E(x_i) = \frac{1}{N} \cdot N \cdot E(x) = E(x)$$

dove si è considerato che $E(x_1)=E(x_2)=\cdots\cdots=E(x)$

Quindi il valore atteso della popolazione delle medie aritmetiche dei campioni di dimensioni N estratti da una popolazione coincide con il valore atteso della popolazione stessa, in simboli

$$E(\bar{x}) = E(x) \qquad (13)$$

Inoltre poiché le N misure sono indipendenti, si ha anche

$$\sigma_{\bar{x}}^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sigma_{xi}^2 = \frac{1}{N^2} \cdot N \cdot \sigma_x^2$$

cioè

$$\sigma_{\bar{x}}^2 = \frac{1}{N^2} \sigma_x^2 \qquad (14)$$

In definitiva la variabile casuale media aritmetica (di campioni di N misure) ha varianza uguale alla varianza della popolazione da cui sono estratti i campioni, divisa per la dimensione N del campione. Per questo motivo la varianza della media tende a zero al crescere della dimensione del campione.

La legge dei grandi numeri

Abbiamo affermato che in base alla definizione assiomatica di probabilità si può dimostrare sotto determinate ipotesi che la frequenza relativa converge, all'aumentare del numero di prove, alla probabilità , cioè la definizione empirica di probabilità. Adesso ci accingiamo a farlo.
Consideriamo una variabile casuale x e siano $E(x)$ e σ^2 la sua speranza matematica e la sua varianza. Vogliamo calcolarci la probabilità che scelto a caso un valore di x esso differisca da $E(x)$ per più di un valore prefissato ε.
Per la legge della probabilità totale (eq.6 pag.3) essa è data da

$$P_r\left[|x-E(x)|\geq\varepsilon\right]=\sum_{|x_i-E(x)|\geq\varepsilon}p_i$$

d'altronde

$$\sigma^2=E\left\{[x-E(x)]^2\right\}=\sum_i p_i\cdot[x-E(x)]^2\geq\sum_{|x_i-E(x)|\geq\varepsilon}p_i\cdot[x-E(x)]^2\geq\sum_{|x_i-E(x)|\geq\varepsilon}p_i\cdot\varepsilon^2=$$
$$=\varepsilon^2\cdot\sum_{|x_i-E(x)|\geq\varepsilon}p_i=\varepsilon^2\cdot P_r(|x-E(x)|\geq\varepsilon)$$

che può riscriversi

$$P_r\left[|x-E(x)|\geq\varepsilon\right]\leq\frac{\sigma^2}{\varepsilon^2}$$

che, se prendiamo $\varepsilon=k\sigma$, diventa

$$P_r\left[|x-E(x)|\geq k\sigma\right]\leq\frac{1}{k^2}\qquad(15)$$

che è nota con il nome di disuguaglianza di $\check{C}eby\check{s}ev$, e che fissa un limite superiore per la probabilità esaminata, limite che è particolarmente significativo quando $k>1$.
Applichiamo la disuguaglianza di $\check{C}eby\check{s}ev$ alla variabile casuale media aritmetica $\bar{x}$.

$$P_r\left[|\bar{x}-E(\bar{x})|\geq\varepsilon\right]\leq\frac{\sigma_{\bar{x}}^2}{\varepsilon^2}\Rightarrow P_r\left[|\bar{x}-E(x)|\geq\varepsilon\right]\leq\frac{\sigma^2}{N\varepsilon^2}$$

questo vuol dire che comunque scelti due numeri positivi ε e δ, è possibile individuare un valore di N tale che risulti

$$\frac{\sigma^2}{N\varepsilon^2}\leq\delta\qquad\text{(basta prendere } N\geq\frac{\sigma^2}{\delta\varepsilon^2}\text{)}$$

e si potrà scrivere

$$P_r\left[|\bar{x}-E(x)|\geq\varepsilon\right]\leq\delta$$

Questo risultato è noto con il nome di *Teorema di* $\check{C}eby\check{s}ev$:
La media aritmetica di un campione finito di valori di una variabile casuale qualunque converge, all'aumentare della dimensione del campione, alla speranza matematica di quella variabile.

Questo quando ci si ricorda che

$$\bar{x} = \sum_{j=1}^{N} f_j \cdot x_j \qquad \text{e} \qquad E(x) = \sum_{j=1}^{N} p_j \cdot x_j$$

il risultato ottenuto conduce alla cosiddetta *legge dei grandi numeri* o *Teorema di Bernoulli: la frequenza relativa di qualunque evento casuale converge alla sua probabilità all'aumentare del numero delle prove.*

Il teorema di $\check{C}eby\check{s}ev$ afferma che

$$\lim_{N\to\infty} \bar{x} = E(x)$$

d'altronde $E(\bar{x})=E(x)$

Questo ci permette di affermare che non solo $\bar{x}$ converge, all'aumentare della dimensione del campione ad *E(x)*, ma che $\bar{x}$ mediamente coincide con *E(x)*. In altre parole dando come stima di *E(x)* la media di uno dei campioni si ha la stessa probabilità di sbagliare per eccesso o per difetto, probabilità che diminuisce all'aumentare della dimensione del campione (*stima imparziale-non distorta*).

In assenza di errori sistematici si può provare che per una variabile casuale distribuita gaussianamente (vedi pag.???) il valore atteso *E(x)* esiste e coincide col valore vero μ. In tal caso si può scrivere
$E(\bar{x})=E(x)=\mu$

Intuitivamente se indichiamo con μ il valore vero della grandezza x e con x_i *(i=1……N)* le N determinazioni sperimentali di x, allora l'errore commesso in ognuna delle xi è dato da

$$\varepsilon_i = x_i - m$$

Per cui l'errore della media è

$$\bar{\varepsilon} = \bar{x} - \mu = \frac{1}{N}\sum_i x_i - \mu = \frac{1}{N}\sum_i (x_i - \mu) = \frac{1}{N}\sum_i \varepsilon_i$$

e se gli errori sono casuali le ε_i tenderanno per $N \to \infty$ ad annullarsi a vicenda nella sommatoria che inoltre è divisa per N.

Approfondiamo, ora , il rapporto tra la varianza s^2 di un campione di N misure e la varianza σ^2 della popolazione da cui il campione è estratto.

$$\sigma^2 = \frac{1}{N}\sum_i (x_i - \mu)^2$$

con semplici passaggi si ottiene

$$\begin{aligned}
\frac{1}{N}\sum_i (x_i - \mu)^2 &= \frac{1}{N}\sum_i \left[(x_i - \bar{x}) + (\bar{x} - \mu)\right]^2 = \\
&= \frac{1}{N}\left[\sum (x_i - \bar{x})^2 + \sum_i (x_i - \mu)^2 + 2(\bar{x} - \mu)\sum_i (x_i - \bar{x})\right] = \\
&= \frac{1}{N}\left[\sum (x_i - \bar{x})^2 + N(\bar{x} - \mu)^2\right] = \\
&= s^2 + \frac{1}{N} N(\bar{x} - \mu)^2 \Rightarrow s^2 = \sigma^2 - (\bar{x} - \mu)^2
\end{aligned}$$

ora sugli infiniti campioni di dimensione N risulterà

$$E(s^2) = E(\sigma^2) - E\left[(\bar{x} - \mu)^2\right]$$

ossia

$$E(s^2) = \sigma^2 - E\left[(\bar{x} - E(\bar{x}))^2\right] \qquad \textit{(si ricordi che } E(\bar{x})=E(x)=\mu\textit{)}$$

$$E(s^2) = \sigma^2 - \sigma^2_{\bar{x}} < \sigma^2$$

Questo significa che il valore atteso della varianza s^2 di un campione è sistematicamente inferiore alla σ^2 dell'intera popolazione, di un fattore $\sigma^2_{\bar{x}}$.
Poiché

$$\sigma^2_{\bar{x}} = \frac{1}{N^2}\sigma^2$$

si può scrivere

$$E(s^2) = \frac{N-1}{N}\sigma^2 \qquad (16)$$

il che giustifica la scelta di utilizzare la relazione

$$s^2 = \frac{1}{N-1}\sum_i (x_i - \bar{x})^2 \qquad (17)$$

per avere una stima imparziale (non distorta) di σ^2.
E' possibile dimostrare, ma non lo faremo, che una volta assunto che la radice di s^2, ovvero la cosiddetta deviazione standard, sia il nostro metodo per valutare la dispersione di un campione di dati, la variabile media aritmetica è la stima del valor atteso *(o anche valor vero)* affetta dal minimo errore casuale, cioè avente la più piccola *deviazione standard* rispetto a qualsiasi altra misura di tendenza centrale.

Variabili casuali continue (unidimensionali)
Per una variabile casuale continua x non ha senso parlare della probabilità che in una prova esca il valore x_i. E' invece possibile associare ad ogni variabile continua x una funzione *f(x)* detta funzione densità di probabilità da cui si può dedurre la probabilità che in una prova x cada in un qualsiasi intervallo finito prefissato $[x_1, x_2]$, che sarà data:

$$\Pr\left(x \in [x_1, x_2]\right) = \int_{x_1}^{x_2} f(x)dx$$

Deve risultare comunque

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Condizione di normalizzazione.
E' bene sottolineare che la sola condizione di normalizzazione garantisce che qualsiasi funzione che rappresenti una densità si probabilità debba tendere a zero all'infinito, indipendentemente dalla natura del fenomeno casuale a cui è collegata. Cosa non sorprendente se ci si ricorda della disuguaglianza di $\check{C}eby\check{s}ev$.
IL significato della densità di probabilità diventa certamente più chiaro se lo colleghiamo a quello di densità di frequenza introdotta per riportare in istogramma i valori delle misure ripetute per una variabile continua.
Anche per una variabile continua è possibile definire il valore atteso

$$E(x) = \int_{-\infty}^{+\infty} xf(x)dx$$

e la varianza

$$Var(x) = \int_{-\infty}^{+\infty} (x - E(x)) f(x) dx$$

Dal teorema di $\check{C}eby\check{s}ev$ si ha:

$$\bar{x} = \sum_{i=1}^{N} f_i x_i = \sum_{i=1}^{N} F_i \Delta x_i \cdot x_i \xrightarrow{N\to\infty} \sum_i x_i f(x_i) dx_i \xrightarrow{N\to\infty} \int_{-\infty}^{+\infty} x f(x) dx = E(x)$$

Esempi di Funzioni di distribuzioni

<u>Binomiale</u>

Introduciamo la distribuzione *binomiale o di Bernoulli* con un esempio. Supponiamo di lanciare un tre dadi e di contare il numero di volte che esce l'asso. I possibili risultati sono 0, 1, 2, o 3. Se ripetiamo la prova un numero enorme di volte, allora troveremo la distribuzione limite che associa ad ogni singolo lancio (di tre dadi) la probabilità di ottenere ν assi, con ν=0, 1, 2 o 3.
Questo è un esperimento così semplice che ci permette facilmente di associare ad ogni evento la sua probabilità. Ovviamente se il dado è *buono* la probabilità di ottenere asso è 1/6.
Gettiamo i tre dadi e cerchiamo la probabilità di ottenere tre assi. Ora poiché i tre dadi rotolano indipendentemente si può scrivere:

$$p(3\ \textit{assi in tre lanci}) = \left(\frac{1}{6}\right)\cdot\left(\frac{1}{6}\right)\cdot\left(\frac{1}{6}\right) = \left(\frac{1}{6}\right)^3 \approx 0{,}5\%$$

$$p(3) = \left(\frac{1}{6}\right)^3$$

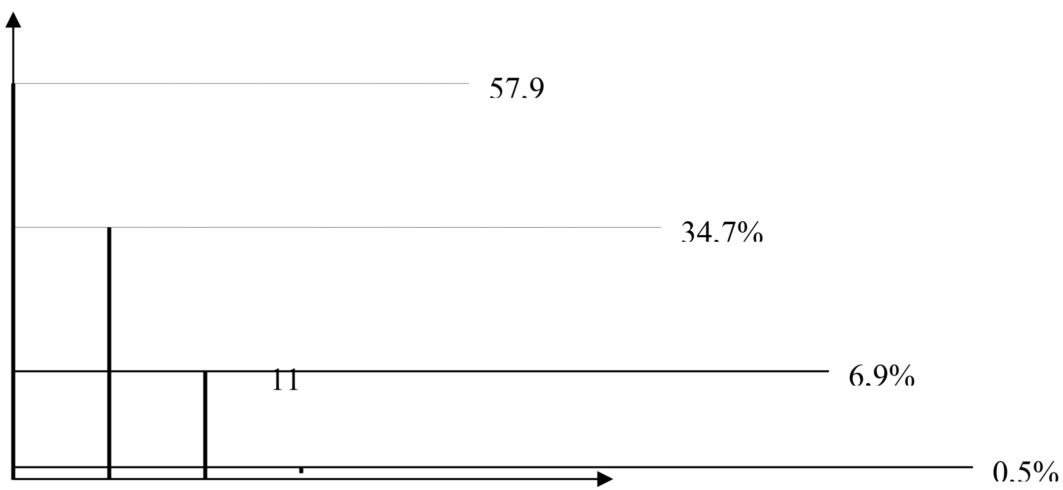
Calcolare $p(2)$ è un po' più difficile, infatti due assi possono ottenersi in parecchi modi. Cominciamo col considerare l'evento (*A, A, nonA*), esso ha probabilità

$$p(A,\ A,\ nonA) = \left(\frac{1}{6}\right)\cdot\left(\frac{1}{6}\right)\cdot\left(\frac{5}{6}\right) = \left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)$$

ma ugualmente probabili sono gli eventi (*A, nonA, A*) e (*nonA, A, A*) e quindi in totale possiamo scrivere che

$$p(2\ \textit{assi in tre lanci}) = 3\cdot\left(\frac{1}{6}\right)^2\cdot\left(\frac{5}{6}\right) \approx 6{,}9\%$$

Calcoli simili danno $p(1\ \textit{assi in tre lanci}) \approx 34{,}7\%$ $p(0\ \textit{assi in tre lanci}) \approx 57{,}9\%$



In generale immaginiamo di fare *n* prove indipendenti e denotiamo con *p* la probabilità che l'evento scelto (successo) accada e con *q=1-p* la probabilità che non accada (insuccesso) in una prova è possibile dimostrare seguendo la falsariga dell'esempio precedente che

$$p(\nu \text{ successi in n prove}) = b_{n,p}(\nu) = \binom{n}{\nu} p^{\nu} q^{n-\nu}$$

Dove il simbolo $b_{n,p}$ sta per binomiale dipendente da *n* e da *p* . La $b_{n,p}(\nu)$ soddisfa la condizione di normalizzazione infatti

$$\sum_{\nu=0}^{n} b_{n,p}(\nu) = \sum_{\nu=0}^{n} \binom{n}{\nu} p^{\nu} q^{n-\nu} = (p+q)^n = (p+1-p)^n = 1^n = 1$$

si può inoltre dimostrare che

$$E(\nu) = \sum_{\nu=0}^{n} \nu \cdot b_{n,p}(\nu) = np$$

Cioè, come ci si dovrebbe aspettare, se ripetiamo molte volte la prova, il numero medio di successi è proprio la probabilità di successo in una prova per il numero di prove effettuato.
Si può anche dimostrare che

$$\sigma_{\nu} = \sqrt{np(1-p)}$$

Quando p=1/2, il numero di successi attesi $E(\nu)$ è proprio n/2 (es. lancio di una moneta). Si può dimostrare che per p=1/2

$$b_{n,\frac{1}{2}}(\nu) = b_{n,\frac{1}{2}}(n-\nu)$$

ovvero la binomiale è simmetrica attorno al valore 1/2.

<u>*La distribuzione di Gauss*</u>
La funzione matematica

$$f(x) = e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

nota in fisica come *funzione di distribuzione di Gauss* o legge o *legge normale degli errori* è, come abbiamo accennato e come vedremo in seguito molto importante, non solo perché rappresenta la curva limite, ovvero, la funzione di distribuzione della densità di probabilità per le misure ripetute in fisica quando governate dall'errore casuale, ma anche perché molte funzioni di distribuzioni all'aumentare dei tentativi $(N \rightarrow \infty)$ tendono ad essa.

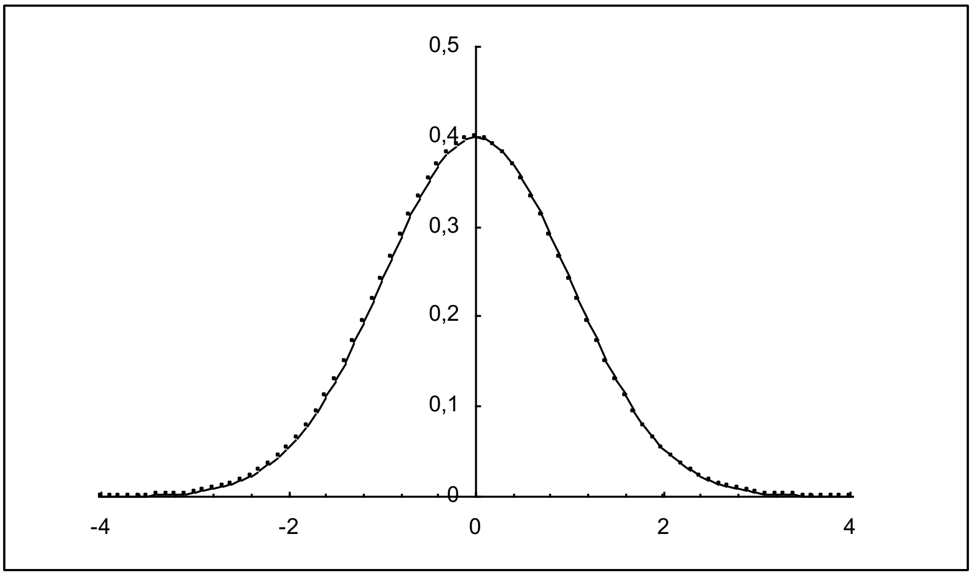
La funzione *di Gauss* ha un massimo per $x=\mu$ ed è simmetrica rispetto a questo valore, ha due flessi per $x=\mu-\sigma$ e per $x=\mu+\sigma$ e diminuisce simmetricamente da ambo i lati.
Perché possa rappresentare una funzione di distribuzione di densità di probabilità, deve per prima cosa soddisfare la condizione di normalizzazione.
Ora poiché si può far vedere che

$$\int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma\sqrt{2\pi}$$

basta moltiplicare la $f(x)$ per la costante $k = \dfrac{1}{\sigma\sqrt{2\pi}}$ per ottenere la funzione cercata

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (*)$$

correttamente normalizzata.
Nella figura è riportato il suo grafico nel caso particolare $\mu=0$ e $\sigma=1$.



La funzione di distribuzione (*) dipende dai due parametri μ e σ, centro e larghezza della distribuzione. In seguito ci soffermeremo sul significato di σ, per ora appare abbastanza chiaro che più piccolo è il valore di σ più stretta e piccata è la forma della funzione di distribuzione.
La speranza matematica o valore atteso per una funzione di Gauss è

$$E(x) = \int_{-\infty}^{+\infty} x \cdot f_{\mu,\sigma}(x) \cdot dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

da cui, integrando per sostituzione, si ha

$$E(x) = \mu$$

operando in maniera analoga si ha per $Var(x)$

$$Var(x) = E\left[(x - E(x))^2\right] = \left[(x-\mu)^2\right]\int_{-\infty}^{+\infty}(x-\mu)^2 \cdot f_{\mu,\sigma}(x)\cdot dx = \sigma^2$$

E' interessante calcolarsi la probabilità che il risultato di un tentativo (una misura) dia come risultato un valore della variabile x entro un certo intervallo predefinito, in particolare la probabilità che x cada nell'intervallo $\mu - t\sigma$ e $\mu + t\sigma$ con t>0, essa è data da

$$p(x \in [\mu - t\sigma; \mu + t\sigma]) = \int_{\mu - t\sigma}^{\mu + t\sigma} f_{\mu,\sigma}(x)\cdot dx = \frac{1}{\sigma\sqrt{2\pi}}\int_{\mu - t\sigma}^{\mu + t\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$$

che, operando la sostituzione $z=(x-\mu)/\sigma$, diventa

$$p(z \in [-t\sigma; +t\sigma]) = \frac{1}{\sqrt{2\pi}}\int_{-t\sigma}^{+t\sigma} e^{-\frac{z^2}{2}}\,dz$$

Questo integrale è fondamentale in fisica ed è spesso chiamato la *funzione degli errori* . Non può essere calcolato analiticamente, ma facilmente approssimato numericamente. Nella tabella seguente sono riportati i risultati per alcuni valori di t

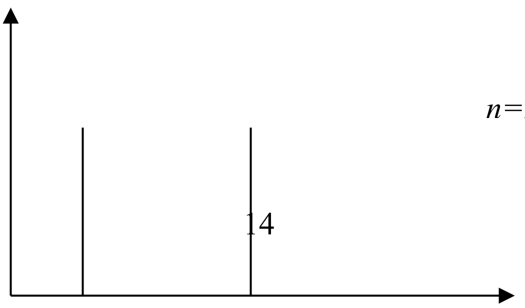
t	0	0,25	0,5	0,75	1	1,5	2	3	4
p(%)	0	20	38	55	68	87	95,4	99,7	99,9

Come si vede in corrispondenza di t=1 la probabilità è del 68% e poi si avvicina rapidamente al 100% .

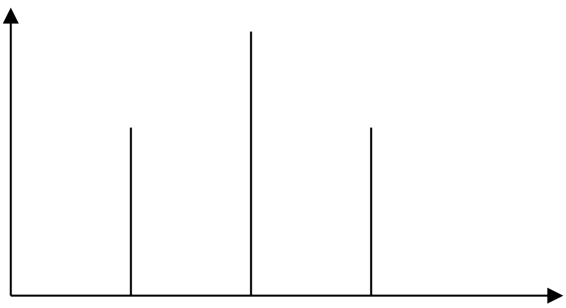
Errori casuali e legge di Gauss

Abbiamo più volte affermato che una misura soggetta a molti piccoli errori casuali è distribuita normalmente. Ora, con un semplice modello, vogliamo verificare questa affermazione.
Assumiamo che le nostre misure siano soggette ad errori sistematici trascurabili, ma che vi siano n sorgenti di errori casuali ed indipendenti e supponiamo che producano errori della stessa fissata dimensione e con la stessa probabilità p=1/2 di spingere il risultato verso l'alto o verso il basso. Per esempio, se il valore vero è μ e se c'è una sola sorgente di errore, allora i risultati possibili sono $x=\mu-\varepsilon$ e $x=\mu+\varepsilon$, entrambi ugualmente probabili; con due sorgenti di errore i risultati possibili sono $x=\mu-2\varepsilon$, $x=\varepsilon$, $x=\mu+2\varepsilon$.

n=1



n=2



In generale con n sorgenti di errore il nostro risultato è compreso tra $x=\mu-n\varepsilon$ e $x=\mu+n\varepsilon$. In una particolare misura se capita che ν sorgenti diano errori positivi e $(n-\nu)$ sorgenti errori negativi, si ha

$$x = \mu + \nu\varepsilon - (n-\nu)\varepsilon = \mu + (2\nu - n)\varepsilon = \mu - n\varepsilon + 2\varepsilon\nu \qquad (26.0)$$

La probabilità di questa occorrenza è data da

$$p(\nu \text{ errori positivi}) = b_{n,\frac{1}{2}}(\nu)$$

Che è simmetricamente distribuita attorno al valore vero μ, con probabilità date dalla funzione di distribuzione binomiale. Ora se il numero di sorgenti di errore, n, è grande e la dimensione dei singoli errori, ε, è piccola, allora le nostre misure sono distribuite normalmente.
Infatti, si può dimostrare che, per un valore fissato di p, quando n è grande, la distribuzione binomiale tende ad una distribuzione di Gauss con lo stesso valore atteso e stessa varianza.
Per n grande, quindi

$$b_{n,p}(\nu) \approx f_{\mu,\sigma}(\nu)$$

Con $E(\nu)=np$ e $\sigma_\nu^2=np(1-p)$

che nel nostro caso diventano $E(\nu)=1/2$ e $\sigma_\nu = \sqrt{n/4}$

Dalla (26.0) in base alla (11) possiamo scrivere che

$$E(x)=E(\mu-n\varepsilon+2\varepsilon\nu)=E(\mu)-E(\varepsilon n)+E(2\varepsilon\nu)=\mu-\varepsilon n+2\varepsilon E(\nu)=\mu-\varepsilon n+\varepsilon n=\mu$$

e poiché le misure sono indipendenti, in base alla (12) è anche

$$\sigma_x^2 = (2\varepsilon)^2\sigma_\nu^2 \Rightarrow \sigma_\nu = 2\varepsilon\sigma_\nu = \varepsilon\sqrt{n}$$

Per $n \to \infty$ ed $\varepsilon \to \infty$ il valore di $\sigma = \varepsilon\sqrt{n}$ rimane fissato e succedono due cose, in primo luogo la distribuzione binomiale si approssima ad una gaussiana con centro μ e larghezza σ_x, e poiché $\varepsilon \to \infty$ i possibili risultati della misura si fanno più vicini tra loro così che la distribuzione discreta si approssima ad una distribuzione continua, che è precisamente, la distribuzione di Gauss attesa.