Elementi di statistica

Le scienze esatte

- Nell'ambito delle **scienze esatte** (o deterministiche) non sembra esserci posto per la casualita': noti i **dati iniziali** e note le **leggi fisiche**, si possono prevedere esattamente i risultati :

condizioni iniziali ⇒ LEGGE FISICA ⇒ Risultato

- Se pero' le **condizioni iniziali** solo note solo parzialmente, la previsione sara' incerta. In tal caso …

⇒ **Massima conoscenza = Probabilita'**

- Nel mondo dell'infinitamente piccolo le condizioni iniziali non possono essere determinate in modo completo (**principio di indeterminazione**). Ne segue che nel mondo delle particelle elementari le leggi sono sempre **leggi di probabilita'.**

Il dado

• **Approccio classico** : la condizione iniziale e' troppo complessa da definire (non impossibile in linea di principio) …

• **Approccio probabilistico** : considerazione di simmetria inducono ad ipotizzare:

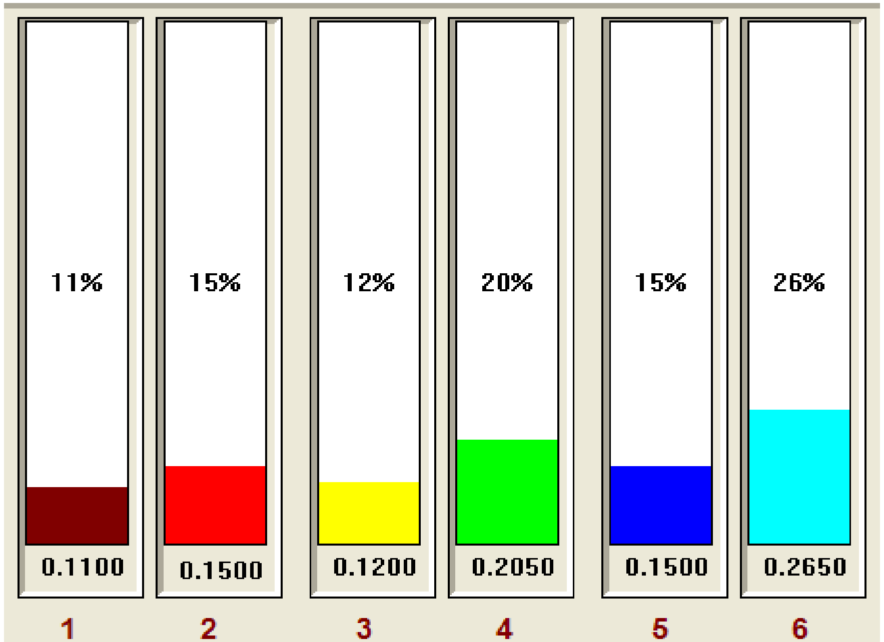
$$P(n) = 1/6 \qquad \text{con } \mathbf{n} = 1,2,3,4,5,6$$

→ Come **verificare** tale previsione teorica ? Galilei insegna: **con un esperimento !**

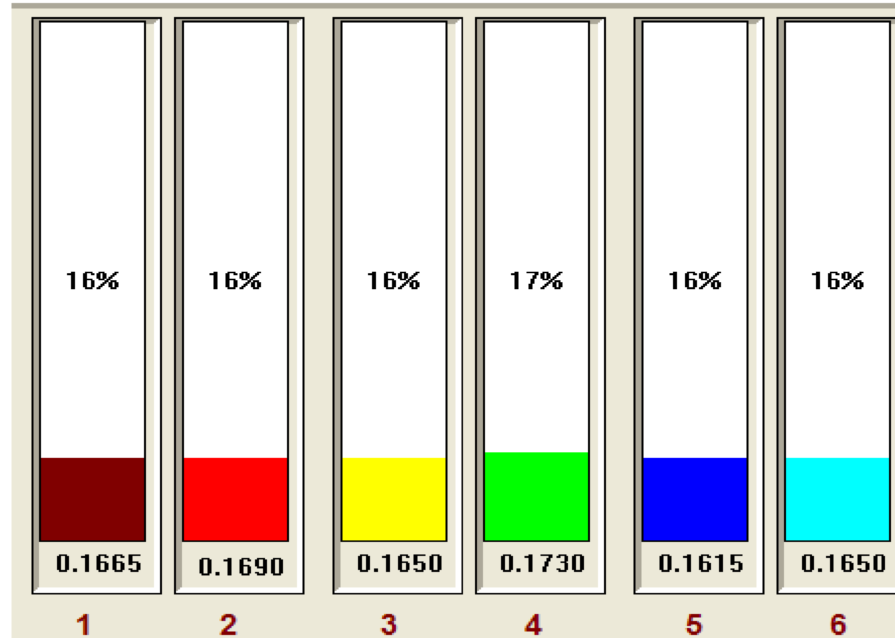
→ **Quanti lanci** effettuare ? **?**

lanciamo il dado !

N=200 lanci



N=2000 lanci



- Come si vede la **frequenza relativa** tende, al crescere delle prove, alla probabilita' prevista ($1/6 \cong 0{,}1666$). Il numero di volte che su N lanci esce un dato numero e' una **variabile casuale** di tipo **binomiale** (vedi piu' avanti).
- L'incertezza associata alle inevitabili fluttuazioni statistiche diminuisce col numero di lanci. Per N=2000 essa vale circa 0,008. Osservando i risultati ci accorgiamo che anche il peggiore risultato (**n=4**: $0{,}173 \pm 0{,}008$) concorda con la previsione (**diremo pertanto di aver verificato l'ipotesi P=1/6 con una precisione pari a 0,008/0,173=0,046 , cioe' del 4,6%).**

Non possiamo fare infinite prove o misure, ma sulla base di un campione possiamo estrarre informazioni utili

La statistica descrittiva

valor medio →

$$m = \frac{\sum_{i=1}^{n} x_i}{N}$$

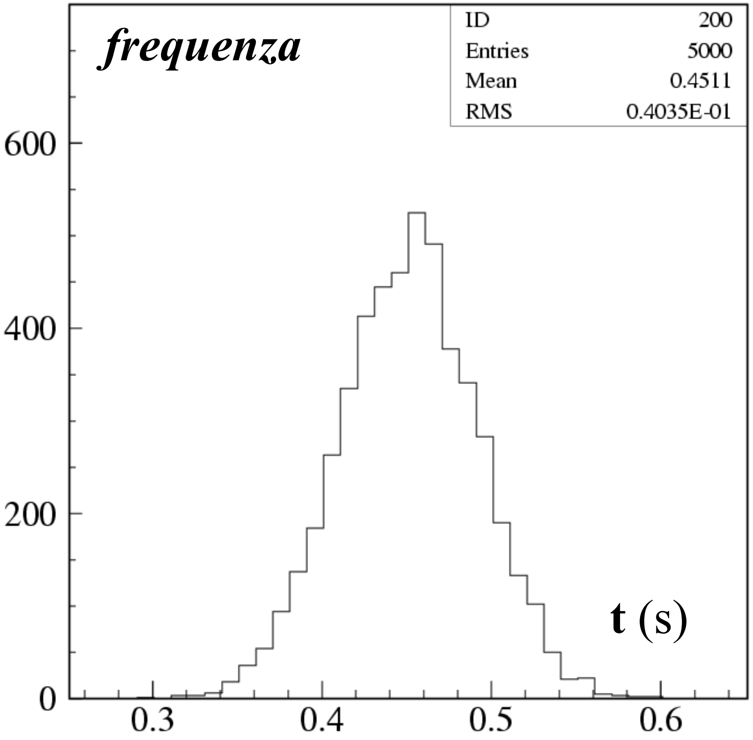
singola misura

esprime il valore piu' **attendibile**

deviazione standard o RMS →

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (m - x_i)^2}{N-1}}$$

esprime la variabilita' e quindi l'**incertezza**.



- L'istogramma accanto mostra la misura ripetuta (N=5000) del tempo di caduta di un oggetto sempre dalla stessa altezza. E' questo un esempio di **variabile casuale** di tipo **gaussiano.**

- Il valor medio che risulta da 5000 misure e' m=0,4511 s, mentre l'incertezza associata a tale valore medio e' $\sigma/\sqrt{5000}=0{,}04/\sqrt{5000}=0{,}0006$ s.

- Un campione piu' ampio (maggior numero di misure) diminuisce sempre l'incertezza: per tale motivo si parla di **errore statistico.**

La variabile binomiale

Si immagini una prova che abbia 2 possibili risultati: successo/insuccesso. Su N prove effettuate, il numero k di volte in cui si manifesta il successo (tra 0 e N) e' detta variabile (aleatoria) binomiale. Ad esempio ***:***

- lancio di un dado
- rivelazione di una data particella
- tipo di decadimento per una particella

Se **p** e' la probabilita' del successo il valor medio e la deviazione standard possono, nel caso di questa variabile, essere ricavati anche dalle espressioni:

$$m = Np$$

$$\sigma = \sqrt{Np(1-p)}$$

Es: *1000 particelle di tipo X attraversano un rivelatore. L'efficienza del rivelatore (cioe' la probabilita' di rilevare particelle X) e' del 65%. Il numero medio di particelle rivelate su 1000 che attraversano il rivelatore sara'* **m** *= 1000×0,65=650, mentre la deviazione standard sara':* $\sigma = \sqrt{1000 \times 0,65 \times (1-0,65)} = 15$

Un processo di decadimento

Il mesone ϕ puo' decadere in 4 'canali' differenti : K^+K^-, K_SK_L , $\pi^+\pi^-\pi^o$, $\eta\gamma$. Si vuole determinare la **probabilita' di ogni canale**. **Fissato** il numero totale di decadimenti, il **numero di decadimenti** in un certo canale e' una **variabile binomiale.**
Supponiamo di aver osservato **N=200** decadimenti e aver ottenuto i seguenti risultati :

canale	conteggio, *k*
K^+K^-	98
K_SK_L	68
$\pi^+\pi^-\pi^o$	30
$\eta\gamma$	4

errore statistico
7
7
5
2

$$m = k = 98$$

$$\sigma = \sqrt{Np(1-p)} = 7$$

200

p ≈ k/N=98/200

Come si esprime il risultato di un conteggio di questo tipo nel caso ad esempio del canale $\phi \rightarrow K^+K^-$?

$$\cancel{98 \pm 7}$$

Il numero assoluto non e' facilmente confrontabile con altri esperimenti !

$$\cancel{k} \rightarrow \frac{k}{N} = \frac{98}{200} = 0.49$$

La probabilita' e' invece facilmente confrontabile

$$\cancel{\sigma(k)} \rightarrow \sigma\left(\frac{k}{N}\right) = \frac{1}{N}\sqrt{Np(1-p)} = \sqrt{\frac{p(1-p)}{N}} = 0.03$$

$$BR = 0.49 \pm 0.03$$

errore statistico:
diminuisce al crescere di N

PS: La probabilita' nel caso dei decadimenti di particelle e' chiamata ***branching ratio*** (BR)

Gli errori sistematici

Ogni misura e' affetta da due tipi di errori : **statistico** e **sistematico**.
Nel caso in questione a cosa può essere dovuto quello sistematico ?

- difficile riconoscimento dell'evento

Un canale puo' essere confuso con un altro

$$BR = 0.49 \pm 0.03_{stat} \pm 0.01_{sist}$$

canale	BR		BR
K^+K^-	0.49	$\times\alpha$	0.47
$K_S K_L$	0.34	$\times\alpha$	0.32
$\pi^+\pi^-\pi^o$	0.15	$\times\alpha$	0.14
$\eta\gamma$	0.02	$\times\alpha$	0.02
...	0.05		0.05
totale	**1.05**		**1.00**

- **esistenza di un canale non direttamente osservabile:** $\Phi \rightarrow$ **...**

Se ad esempio e' noto che tale canale prende il 5%, allora l'errore sistematico di sopravvalutazione dei ***BR*** viene corretto $\rightarrow$ 0.95

Errore totale: $\Delta(BR) = \sqrt{0.03^2 + 0.01^2} = 0.032$ (somma in quadratura)

Combinare insieme le misure

- Perché combinare insieme più misure indipendenti ? **Per migliorare la precisione**
- Come si combinano piu' misure ? **Mediante una media pesata**

$$\overline{x} = \frac{\sum_i p_i \cdot x_i}{p_{tot}}$$

$$p_i = \frac{1}{\sigma_i^2}$$

$$\sigma_{\overline{x}} = \frac{1}{\sqrt{p_{tot}}}$$

Si abbiano due misure del ***BR***(ϕ→ K^+K^-) :

$$BR_1 = (0.49 \pm 0.03) \rightarrow \mathrm{p}_1 = 1111$$

$$BR_2 = (0.47 \pm 0.01) \rightarrow \mathrm{p}_2 = 10000$$

$$\overline{BR} = \frac{1111 \times 0.49 + 10000 \times 0.47}{11111} = 0.472$$

$$\sigma(\overline{BR}) = \frac{1}{\sqrt{11111}} = 0.009$$

$$BR = (0.472 \pm 0.009)$$