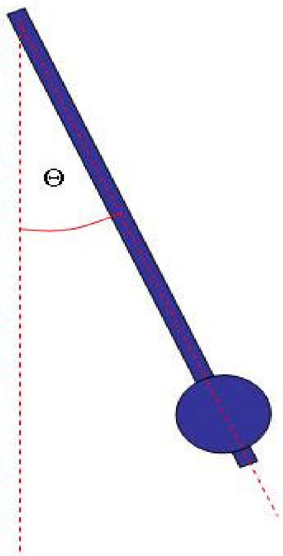
Università degli Studi di Udine
Laboratorio di Fisica Generale I
A.A: 2003/04
PRIMA ESPERIENZA DI LABORATORIO

L'esperienza consiste nella misura, mediante un cronometro, del periodo di oscillazione di un pendolo composto. **Essa è finalizzata all'apprendimento dei concetti fondamentali della statistica e al loro impiego nell'analisi dei dati sperimentali.** Si raccoglieranno tre campioni di 30, 100 e rispettivamente 300 misure ciascuno; per ogni misura si cronometra il singolo periodo.
Per ridurre la dipendenza del periodo dalla massima elongazione angolare Θ_{max}, ci si pone nella condizione delle piccole oscillazioni (Θ_{max}<15°).



ERRORI DI MISURA

Gli errori di misura sono di due tipi: **sistematici** e **casuali**.
Per spiegare cos'è un errore sistematico consideriamo un metro a nastro, in cui le divisioni siano state erroneamente tracciate più ravvicinate del dovuto, secondo un certo fattore percentuale *p* (per esempio siano più vicine del 5%). Come conseguenza, quando usiamo questo metro, le misure risulteranno maggiori di quanto dovrebbero, cioè lo strumento sovrastima le lunghezze. Un altro esempio di errore sistematico è un misuratore di resistenza elettrica che non sia azzerato correttamente (cioè che dia un valore di resistenza diverso da zero anche quando i terminali siano messi in corto circuito). E ancora: leggere la posizione dell'indice di una scala guardando in direzione non perpendicolare porta ad un errore sistematico di lettura (detto di parallasse). In conclusione un errore sistematico è dovuto all'azione di una sola causa che agisce costantemente in un verso.
Un errore casuale è invece il risultato dell'azione contemporanea di un numero molto grande, ma ciascuna di piccola entità, di cause diverse, che si sommano e si sottraggono differentemente ogni volta che eseguiamo una misura. Le misure casuali possono essere analizzate mediante la statistica e come regola generale il loro peso si può rendere tanto più piccolo (ovvero la misura si può rendere tanto più **precisa**), quanto maggiore è il numero di misure raccolte.

Questo non è il caso per gli errori sistematici. Riferiamoci di nuovo al nostro metro troppo corto: qualunque sia il numero di misure eseguite, tutte saranno affette da un errore percentuale per eccesso di $(1-p)/p$. L'unico modo per ridurre gli errori sistematici (cioè avere una misura più **accurata**) è usare uno strumento migliore.
In generale ogni misura è affetta sia da errori sistematici che casuali. Alcuni errori sono eliminabili, altri si possono ridurre, ma per quanto accurati e precisi si possa essere, ne rimangono altri che non possono essere eliminati. La cosa migliore che si può fare allora è stimare la loro grandezza. In quanto segue, concentreremo la nostra attenzione sugli errori casuali.

VARIABILI DISCRETE E CONTINUE

Per quanto detto sull'impossibilità di eliminare gli errori casuali, il valore della misura di una grandezza fisica può pensarsi, dal punto di vista statistico, come una variabile aleatoria.
Ci sono due tipi di variabili aleatorie:

1) Discrete, ad esempio l'uscita di una faccia particolare di un dado.
2) Continue, ad esempio il periodo del pendolo.

TABULAZIONE DEI DATI

I dati misurati possono essere raccolti in una tabella, come nell'esempio seguente, in cui sono riportate ottanta misure di periodo:

1.99	1.88	1.99	1.93	1.94	1.99	2.01	1.99	2.00	1.96
1.91	2.00	2.01	2.00	1.99	2.04	2.02	2.00	2.02	1.99
1.94	1.95	2.00	2.00	1.96	1.97	2.00	1.98	1.98	2.00
1.99	1.99	1.99	1.99	1.98	2.03	2.03	1.99	1.95	1.90
1.99	2.04	2.02	1.97	2.00	1.98	1.98	2.02	2.02	2.01
2.00	2.01	2.01	2.00	1.98	2.02	2.01	2.01	2.04	1.93
1.90	2.06	1.96	1.95	1.99	2.00	2.00	1.98	1.95	1.99
2.02	2.01	1.97	1.94	1.93	1.97	1.94	2.02	2.06	2.01

Un modo più utile di tabulare i dati si ottiene riportando il numero di occorrenze *n* per ciascun valore *v* misurato ed eventualmente la sua frequenza relativa *f*:

v	n	f (%)
1.85	0	0
1.86	0	0
1.87	0	0
1.88	1	1.25
1.89	0	0
1.90	2	2.5
1.91	1	1.25
1.92	0	0

v	n	f (%)
1.93	3	3.75
1.94	4	5
1.95	4	5
1.96	3	3.75
1.97	4	5
1.98	7	8.75
1.99	14	17.5
2.00	13	16.3

v	n	f (%)
2.01	9	11.3
2.02	8	10
2.03	2	2.5
2.04	3	3.75
2.05	0	0
2.06	2	2.5
2.07	0	0
2.08	0	0

PRESENTAZIONE DEI DATI. ISTOGRAMMAZIONE

L'istogramma è un diagramma bidimensionale molto usato per la visualizzazione dei dati. Sull'asse delle ascisse si riporta il valore della grandezza misurata (ad es. il tempo), sull'asse delle ordinate si riporta il numero di volte n_k che si è trovato un dato valore di tale grandezza, oppure la frequenza relativa di tale valore $f_k = n_k/N_{tot}$ ove N_{tot} è il numero totale di misure effettuate. L'asse delle ascisse viene suddiviso in un numero M di intervalli, detti **canali**, a cui si associano i valori che la variabile può assumere, secondo le modalità seguenti:

1) Per istogrammare una **variabile discreta**, si associa un canale ad ognuno dei valori che la variabile può assumere.
2) Per una **variabile continua** questo procedimento non si può applicare, in quanto è impossibile associare un canale ad ognuno dei valori che la variabile può assumere con continuità. Per ovviare a questo problema si può agire in due modi:
    a. suddividere l'intervallo dei valori della variabile in canali di ampiezza arbitraria e associare a un dato canale tutti i valori di misura maggiori o uguali all'estremo inferiore e minori dell'estremo superiore del canale;
    b. scegliere un intervallo di ampiezza minima uguale al valore della sensibilità dello strumento (o di apprezzamento dello sperimentatore).

    In entrambi i casi si riconduce l'istogrammazione di variabile continua al caso di variabile discreta.

Per una variabile continua l'intervallo di variazione di x si potrà pensare come unione di sottointervalli:

$$I = \bigcup_{k=1}^{M} I_k$$

ad ognuno dei quali sarà associato un valore (ad esempio il valore centrale x_k) che permetta di individuare il canale e un'ampiezza, Dx_k:

$$I_k = I_k(x_k, \Delta x_k)$$

Di solito si sceglie la stessa ampiezza per tutti gli intervalli, e la si pone uguale alla sensibilità dello strumento.

Nella figura seguente è riportato l'istogramma dei dati di periodo, l'ampiezza del canale è stata scelta uguale alla sensibilità del cronometro usato per eseguire la misura (*0.01 s*):

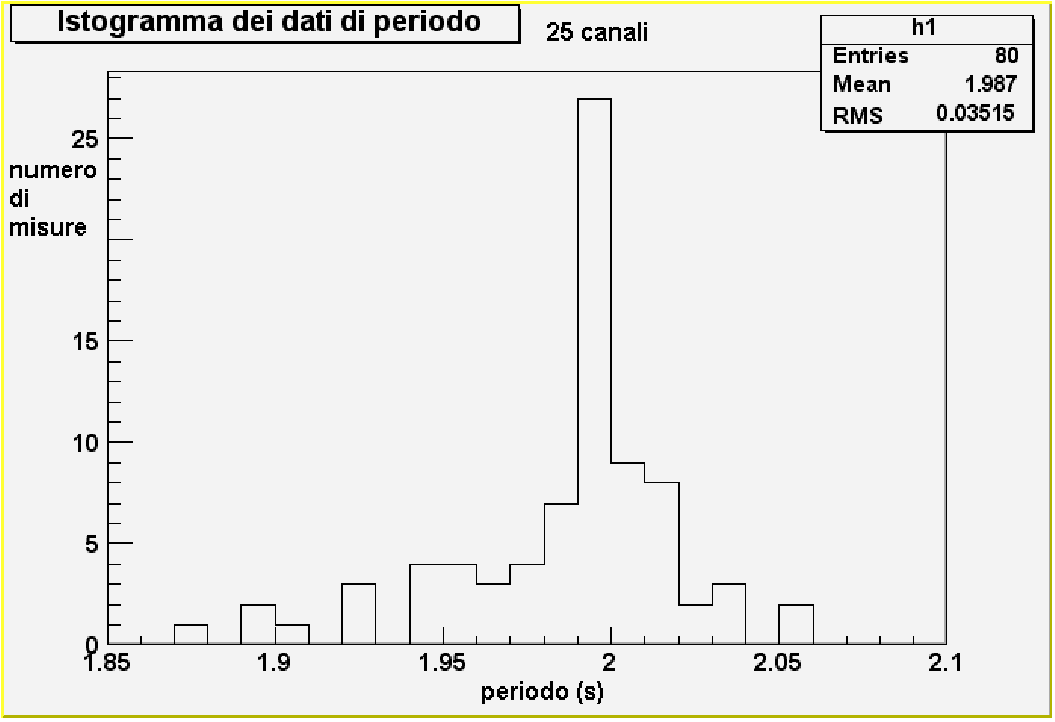


Fig. 1

Data l'arbitrarietà della scelta si può scegliere un'ampiezza più grande, ottenendo come risultato di raggruppare gli stessi dati in un numero minore di canali, come esemplificato nella figura seguente.

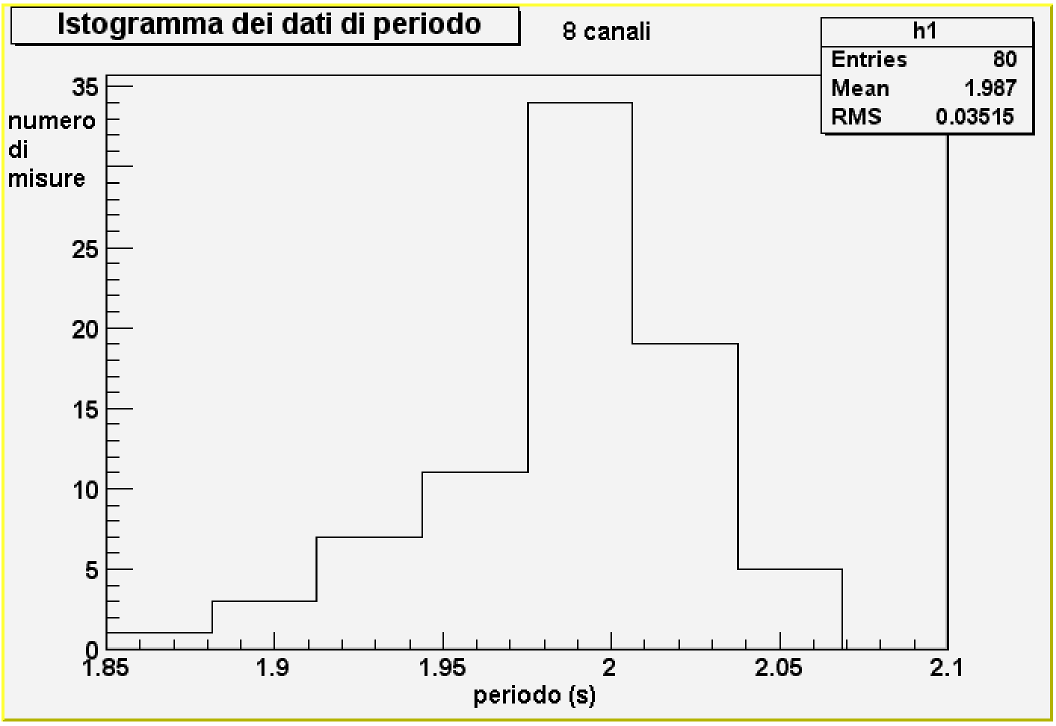


Fig. 2

Ovviamente non bisogna diminuire troppo il numero di canali, altrimenti si perde il dettaglio della distribuzione dei dati (basta pensare al caso limite in cui si abbia un solo canale)
Non ha poi molto senso usare canali con ampiezza inferiore alla sensibilità, in quanto in questo modo si introducono canali in soprannumero che rimangono vuoti, come è esemplificato nella figura seguente.

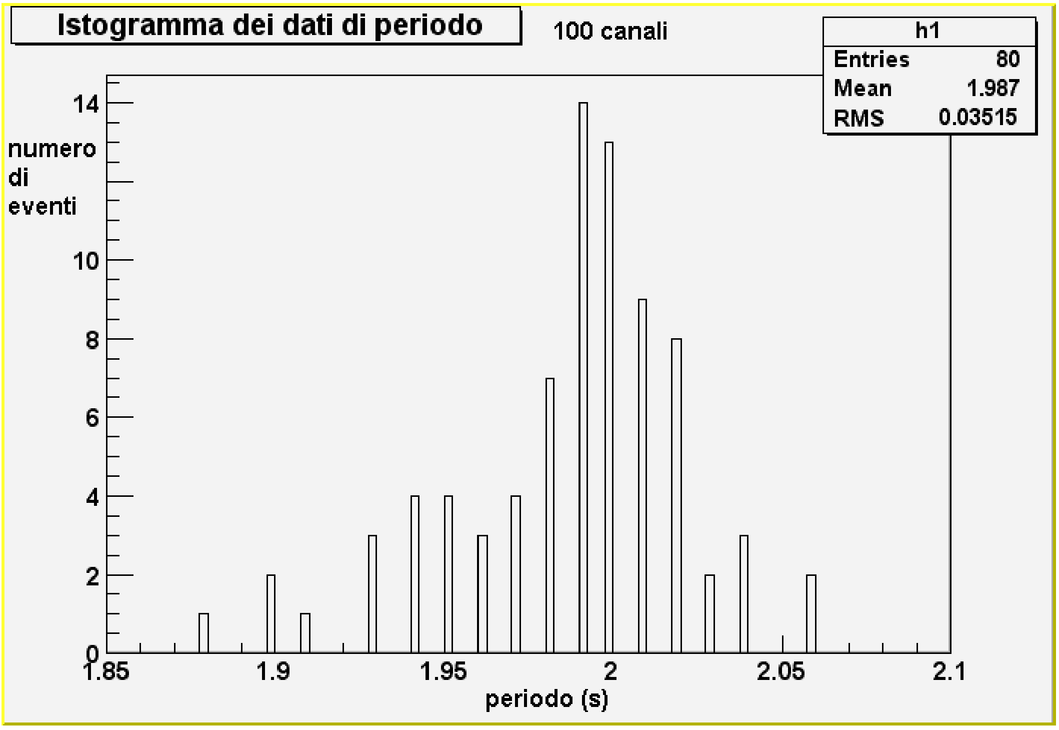


Fig. 3

Un'altra scelta possibile per l'ampiezza dei canali verrà indicata quando introdurremo il concetto di deviazione standard di un insieme di dati.

POPOLAZIONE E CAMPIONE.

Due concetti fondamentali della statistica sono quelli di **campione** e di **popolazione**. Riferiamoci per esempio al nostro pendolo: chiamiamo popolazione l'insieme di tutte le misure di periodo che possiamo eseguire sul pendolo. Questo insieme è infinito, in quanto non c'è limite al numero di misure che possiamo effettuare. Se poi eseguiamo un certo numero (finito) di misure di periodo, realizziamo un campione, ovvero 'estraiamo' un campione dalla popolazione. Scopo della statistica è di raccogliere informazioni sulla popolazione attraverso lo studio dei campioni. Ovvero: noi abbiamo accesso (empirico) solamente a uno o più campioni, ma siamo interessati alle proprietà della popolazione, cioè a proprietà che non dipendano dal particolare campione che abbiamo raccolto. Passare dal campione alla popolazione rappresenta dunque lo svincolamento dal particolare (per quanto ciò sia possibile) per attingere il generale.

Ricordando la definizione di frequenza relativa, calcoliamo la somma delle frequenze di tutti i possibili canali:

$$\sum_{j=1}^{M} f_j = \sum_{j=1}^{M} \frac{n_j}{N} = 1$$

Questa formula è molto generale, in quanto:

1) permette di includere anche i valori della variabile che non sono presenti nel campione, in tal caso basterà considerare pari a zero le frequenze associate a tali valori;
2) può essere estesa al caso in cui M sia infinito (sempre nel caso discreto).

Un'assunzione fondamentale della teoria statistica è che all'aumentare di N la frequenza f_j (concetto applicabile ai campioni) tenda alla probabilità p_j (che è invece un concetto proprio della popolazione). La formula precedente si traduce allora nella proprietà fondamentale di ogni funzione di probabilità di una popolazione, cioè la **normalizzazione** a 1:

$$\sum_{j=1}^{M} p_j = 1$$

che ha l'ovvio significato che la probabilità di ottenere un risultato, tra quelli possibili, eseguendo una misura è pari alla certezza (ovvero 1, matematicamente parlando).
Possiamo scrivere la condizione di normalizzazione anche per una variabile continua, ma dobbiamo prima fare una piccola parentesi e definire la frequenza cumulativa.

FREQUENZA CUMULATIVA

Per ogni valore x_j che una variabile discreta X può assumere, la frequenza cumulativa assoluta N_j è pari al numero di volte che il risultato della misura è stato minore o uguale a x_j:

$$N_j = \sum_{k=1}^{j} n_k$$

La frequenza cumulativa relativa F_j è il rapporto tra la frequenza assoluta e il numero totale N di misure:

$$F_j = \frac{N_j}{N} = \sum_{k=1}^{j} \frac{n_k}{N} = \sum_{k=1}^{j} f_k$$

Nella figura seguente è riportato l'istogramma cumulativo delle frequenze per i dati di periodo:

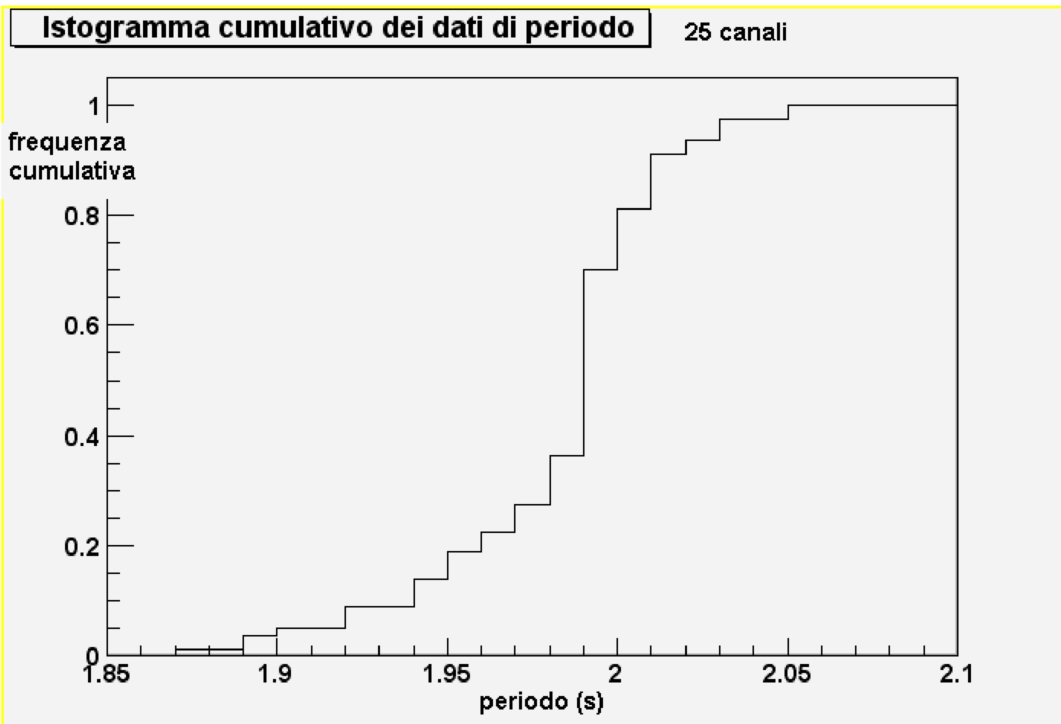


Fig. 4

Si assume ancora che all'aumentare di N la frequenza F_j tenda alla probabilità P_j:

$$\lim_{N \to \infty} F_j = P_j$$

Si consideri ora il rapporto

$$\frac{P_j - P_{j-1}}{x_j - x_{j-1}}$$

cioè la differenza tra le probabilità dei canali *j* e *j-1*, divisa per la differenza fra i valori centrali degli stessi intervalli. Se la variabile *x* è continua, possiamo pensare (almeno matematicamente) di diminuire indefinitamente l'ampiezza dei canali, cioè $(x_j - x_{j-1}) \to 0$, ottenendo il limite continuo dell'istogramma:

$$\lim_{\Delta x \to 0} \frac{P_j - P_{j-1}}{x_j - x_{j-1}} = \frac{dP}{dx}$$

che rappresenta non la probabilità, ma la **densità di probabilità** della variabile continua. Ovviamente questo procedimento è puramente ideale, in quanto qualunque sia lo strumento di misura impiegato, la sua risoluzione è comunque maggiore di zero; come si è notato in precedenza, non ha molto senso considerare un'intervallatura più fine di quella data dalla risoluzione. Se si vuole, è un procedimento iterativo, ove si considera una successione illimitata di strumenti, misure e istogrammi con risoluzione sempre più fine.

Un chiarimento sulle dimensioni fisiche della densità di probabilità: mentre la probabilità è un numero puro, senza dimensioni, la densità di probabilità ha le dimensioni dell'inverso della variabile *x*. Ad esempio se *x* ha la dimensione di un tempo $[T^1]$, la densità di probabilità associata a *x* ha le dimensioni dell'inverso di un tempo $[T^{-1}]$, e ciò perché nella definizione di densità, la variabile *x* sta a denominatore.

Siamo ora in grado di scrivere la condizione di normalizzazione per una variabile continua:

$$\int dP(x) = \int \frac{dP}{dx} dx = 1$$

Formalmente parlando, la sommatoria è stata sostituita dall'integrale, l'indice discreto *j* dall'indice continuo *x*, la probabilità discreta p_j da quella continua (e 'infinitesima') *dP(x).*
Per compattezza di scrittura useremo il simbolo *g(x)* per la funzione densità di probabilità:

$$g(x) = \frac{dP}{dx}$$

MEDIA E VARIANZA DI UN CAMPIONE

Consideriamo un campione di *N* dati:

$$\{x_i\}_{i=1,\dots N}$$

un primo importante concetto è quello di **media** del campione, definita come:

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

un altro modo utile di scrivere la media si ottiene raggruppando fra loro i dati uguali: ogni valore misurato x_j viene moltiplicato per il numero di volte n_j che è stato ottenuto:

$$\bar{x} = \frac{1}{N}\sum_{j=1}^{M} x_j n_j$$

ove *j = 1, … M*, indica gli *M* valori diversi tra gli *N* totali. Introducendo la frequenza relativa f_j (uguale a n_j/N) nella formula precedente, la media si può riscrivere:

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

Un secondo concetto importante è quello di **varianza** del campione, definita come:

$$s^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 = \sum_{j=1}^{M} f_j (x_j - \bar{x})^2$$

una proprietà della varianza è di essere sempre positiva o nulla.
Per il nostro campione di ottanta misure abbiamo rispettivamente:

$$\bar{x} = 1.987$$

$$s^2 = 0.00123$$

Si noti che mentre la media ha le stesse dimensioni fisiche della grandezza (nel nostro caso un tempo) la varianza ha le dimensioni del quadrato delle dimensioni della grandezza (nel nostro caso un tempo al quadrato). Questo ci porta a introdurre un altro concetto, quello di **scarto quadratico medio (o deviazione standard)**, definito come la radice quadrata della varianza:

$$sqm = \sqrt{s^2}$$

Lo *sqm* (indicato anche con la lettera s) ha le stesse dimensioni fisiche della media ed è un concetto molto utile per indicare in modo semplice quanto strettamente (o ampiamente) i dati si distribuiscano intorno a questa.
Per una variabile continua, si può scegliere l'intervallo di istogrammazione pari a *1/2* o *1/3* dello *sqm*, come alternativa a sceglierlo uguale alla sensibilità dello strumento di misura.

MEDIA E VARIANZA DELLA POPOLAZIONE

Analogamente a quanto fatto nel paragrafo precedente, definiremo ora la media e la varianza della popolazione.
Assegnata la funzione teorica p_j , che definisce la popolazione per una variabile discreta, la media della popolazione è così definita:

$$m = \sum_{j=1}^{M} x_j p_j$$

e analogamente assegnata la funzione teorica *dP/dx* per una variabile continua, la media è:

$$m = \int x dP = \int x \frac{dP}{dx} dx = \int x g(x) dx$$

Per entrambi i casi (continuo e discreto), si usa anche la notazione:

$$m = E(x)$$

ove *E* ha il significato di valore di aspettazione (dall'inglese 'expectation value') della variabile *X*.
Definiamo ora la varianza della popolazione per variabili discrete:

$$s^2 = \sum_{j=1}^{M} (x_j - m)^2 p_j$$

e continue:

$$s^2 = \int (x - m)^2 dP = \int (x - m)^2 \frac{dP}{dx} dx = \int (x - m)^2 g(x) dx$$

anche per la varianza si usa la notazione:

$$s^2 = E\left((x - m)^2\right)$$

con significato di valore di aspettazione della variabile $(X - m)^2$.

STIMA DI MEDIA E VARIANZA DELLA POPOLAZIONE

Possiamo finalmente gettare un ponte tra i campioni e la popolazione.
A questo proposito si possono dimostrare alcuni teoremi molto importanti.

1) La media di un campione è un **estimatore centrato** (o imparziale) della media della popolazione:

$$E(\bar{x}) = m$$

2) La varianza di un campione invece è un **estimatore scentrato** (o parziale) della varianza della popolazione:

$$E(s^2) \neq s^{\ 2}$$

infatti si può dimostrare che

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

Per questo motivo, per stimare la varianza della popolazione, si preferisce usare un estimatore leggermente diverso:

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

La sua media è infatti

$$E(s_{N-1}^2) = s^{\ 2}$$

Possiamo dunque fare l'importante affermazione che: **la media e $s^2{}_{\mathbf{N-1}}$ di un <u>campione</u> sono estimatori centrati della media e della varianza della <u>popolazione</u>.**

INTERPRETAZIONE PROBABILISTICA DELL' INTEGRALE DI g(x)

Ricordando la definizione di *g(x)*, possiamo interpretarne l'integrale su di un intervallo I_k:

$$\int_{x \in I_k} g(x)dx = \int_{x \in I_k} dP$$

come la probabilità che effettuando una misura, si ottenga un valore di *x* contenuto in tale intervallo.

ISTOGRAMMAZIONE DI UNA VARIABILE CONTINUA

Riferendoci all'equazione precedente, se l'intervallo considerato è abbastanza piccolo, si può valutare approssimativamente il valore di questa probabilità, nel modo seguente:

$$\int_{x \in I_k} g(x)dx \approx g(x_k)\Delta x_k$$

ove x_k è un valore interno all'intervallo e Δx_k è l'ampiezza dell'intervallo. Se abbiamo un campione di *N* misure, ci aspettiamo che nell'intervallo I_k cada un numero di misure circa uguale a:

$$n_k \approx N g(x_k)\Delta x_k$$

Quando si istogramma un campione di *N* misure, è molto utile confrontare il numero di dati che effettivamente cadono in ogni canale (valore misurato), con il numero calcolato con la formula precedente (ovviamente questo si può fare solo una volta nota la funzione *g(x)*). Tale formula ci dice che bisogna moltiplicare il valore assunto da *g* in un punto interno del canale (ad esempio il valore centrale), per il numero totale di dati e per l'ampiezza del canale.

FUNZIONE DENSITA' DI PROBABILITA' GAUSSIANA

La funzione gaussiana è definita come:

$$\frac{dP}{dx} = g(x) = Ce^{-\frac{(x-a)^2}{2b^2}}$$

cioè dipende da tre parametri: *C*, *a* e *b*; vedremo subito che solo due di questi parametri sono indipendenti. Usualmente si scelgono *a* e *b* come indipendenti.
Il significato del parametro *C* è quello di costante di normalizzazione, cioè va aggiustato in modo da soddisfare la richiesta di normalizzazione a 1 della distribuzione di probabilità:

$$\int \frac{dP}{dx} dx = \int C e^{-\frac{(x-a)^2}{2b^2}} dx = 1$$

estraendo la costante *C* dall'integrale e facendo il cambiamento di variabile:

$$u = \frac{x-a}{\sqrt{2}b}$$

si ottiene:

$$C\sqrt{2}b \int e^{-u^2} du = 1$$

ricordando che l'integrale vale $\sqrt{p}$ e risolvendo per *C*, si ottiene infine:

$$C = \frac{1}{\sqrt{2p}\, b}$$

Da questa espressione si vede come *C* dipenda da *b* (ma non da *a*).
Il significato del parametro *a* si ottiene considerando l'espressione della media:

$$m = \int x \frac{dP}{dx} dx = \int x \frac{1}{\sqrt{2p}\, b} e^{-\frac{(x-a)^2}{2b^2}} dx$$

operando lo stesso cambiamento di variabile visto in precedenza, si ottiene:

$$m = \frac{2}{\sqrt{2p}} b \int u e^{-u^2} dx + \frac{1}{\sqrt{p}} a \int e^{-u^2} dx$$

il primo integrale è nullo, per cui si ottiene:

$$m = a$$

cioè *a* rappresenta la media della distribuzione.
Il significato del parametro *b* si ottiene considerando l'espressione della varianza:

$$s^2 = \int (x-m)^2 \frac{dP}{dx} dx = \int (x-m)^2 \frac{1}{\sqrt{2p}\, b} e^{-\frac{(x-m)^2}{2b^2}} dx$$

con il solito cambiamento di variabile si ottiene:

$$s^2 = \frac{2}{\sqrt{p}} b^2 \int u^2 e^{-u^2} dx$$

siccome l'integrale vale $\sqrt{p}/2$, si ottiene infine:

$$s^2 = b^2$$

cioè *b* rappresenta lo *sqm* della distribuzione.
In conclusione, la funzione *g(x)* dipende dai due parametri (della popolazione) m e s:

$$g(x) = \frac{1}{\sqrt{2p}\, s} e^{-\frac{(x-m)^2}{2s^2}}$$

Essendo parametri della popolazione, essi non sono noti a priori. E' però possibile stimarli attraverso un campione di misure:

$$m -> \bar{x}$$
$$s^2 -> s_{N-1}^2$$

per cui in pratica si userà la seguente forma:

$$g(x) = \frac{1}{\sqrt{2ps^2}} e^{-\frac{(x-\bar{x})^2}{2s^2}}$$

ove **solo per semplicità di scrittura si è lasciato cadere l'indice *N-1* di s^2.**
Ritornando all'interpretazione probabilistica dell'integrale della densità di probabilità, possiamo affermare che l'espressione

$$\int_{m-s}^{m+s} g(x)dx = 68\%$$

rappresenta la probabilità che il risultato di una misura della grandezza x cada nell'intervallo compreso tra $(m-s)$ e $(m+s)$. Analogamente l'integrale

$$\int_{m-2s}^{m+2s} g(x)dx = 98\%$$

e l'integrale

$$\int_{m-3s}^{m+3s} g(x)dx = 99.7\%$$

rappresentano la probabilità che il risultato della misura cada entro l'intervallo $(m-2s, m+2s)$ e $(m-3s, m+3s)$ rispettivamente.

ESPERIMENTI SIMULATI. FIT DEI DATI

Grazie all'uso del computer e di opportuni programmi è possibile simulare l'esecuzione di un esperimento, o meglio è possibile generare dati che riproducano da vicino i risultati di un esperimento reale. Per rimanere nel nostro caso di misura di periodo, si parte dall'ipotesi che la variabile periodo segua una data distribuzione di probabilità, ad esempio gaussiana. Quindi si usa un generatore di numeri casuali distribuiti gaussianamente, specificando i parametri fondamentali: valor medio e deviazione standard della popolazione. Ovviamente questi valori vanno stimati in qualche modo, ad esempio analizzando un campione di dati reali. Si specifica inoltre la consistenza del campione e l'ampiezza dell'intervallo di istogrammazione. Si procede poi esattamente come fatto nel caso dei dati reali. Nella figura seguente sono istogrammati i dati relativi a tre esperimenti simulati, relativi a campioni di 30, 300 e 3000 dati rispettivamente (andando da sinistra a destra). Per tutti l'ampiezza di canale è di *0.01 s*, per aderenza al nostro esperimento reale, così come il valor medio (*1.987 s*) e la deviazione standard (*0.035 s*), entrambi stimati dal nostro campione reale.

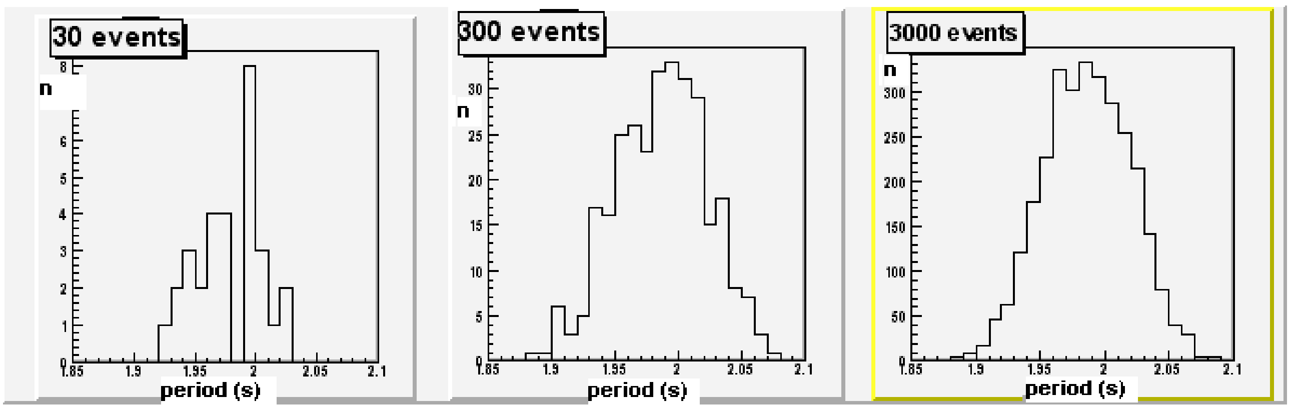


Fig. 5

Si può raffinare ulteriormente l'analisi, cercando una funzione che si adatti nel modo migliore ai dati. A tal fine bisogna innanzi tutto scegliere la forma della funzione teorica che descrive la popolazione. Nel nostro caso, visto come abbiamo generato i dati, sceglieremo ovviamente una gaussiana. Quindi il programma di 'adattamento' dei parametri ai dati (**fit**, in inglese), permette di trovare i valori migliori dei parametri.
Nella figura seguente si sono sovrapposte agli istogrammi di Fig. 5 le curve con i migliori valori dei parametri determinati dal programma.

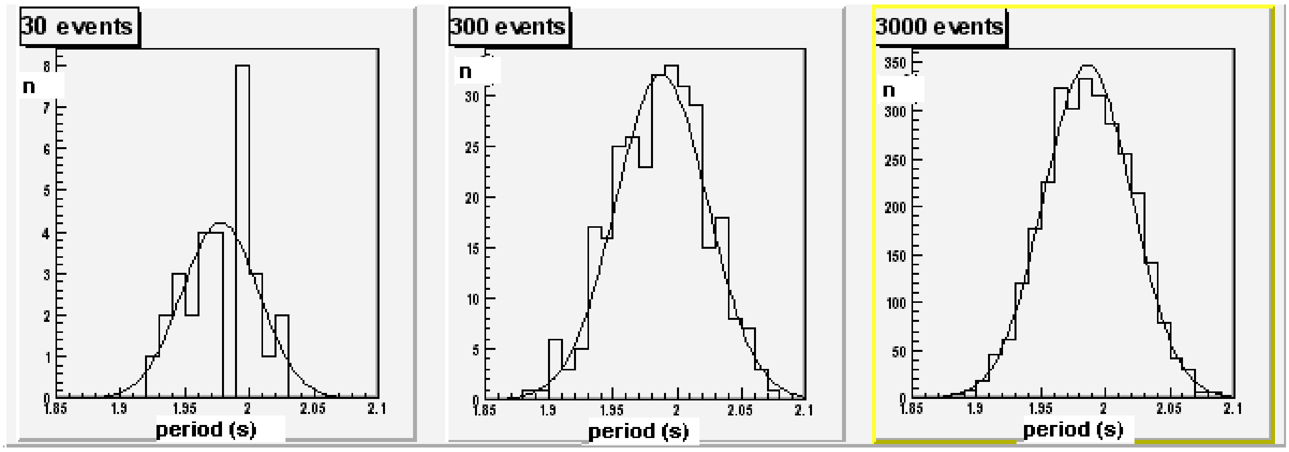


Fig. 6

I valori sono i seguenti:

N	**30**	**300**	**3000**
m	1.978	1.986	1.986
s	0.026	0.036	0.034

Essi sono in buon accordo con i valori che abbiamo usato per generare i dati, anche per il campione più piccolo con *N=30.*

TEST D'IPOTESI

Un capitolo importantissimo della statistica è il cosiddetto test d'ipotesi. Non abbiamo qui lo spazio per trattare questo argomento, ma possiamo almeno porre esplicitamente la domanda che sorge spontanea quando cerchiamo di adattare una funzione gaussiana ai dati di periodo che abbiamo raccolto nell'esperimento reale: **come facciamo a essere sicuri che la variabile periodo segue una legge di distribuzione gaussiana?** A priori non c'è nulla che ci assicuri che sia proprio così. La statistica ci permette però di affrontare la domanda **da un altro punto di vista**, e cioè:

1) ammettiamo che l'ipotesi sia vera
2) tramite opportuni algoritmi calcoliamo la probabilità di questa affermazione, cioè la verità dell'ipotesi.

Se questa probabilità risulta troppo bassa siamo ragionevolmente autorizzati a rigettare l'ipotesi, altrimenti siamo autorizzati a ritenerla consistente con i dati.
Non ci spingeremo così lontano con la nostra analisi, per ora ci basta ammettere la verità dell'ipotesi e ricavare i valori migliori dei parametri della distribuzione di probabilità.