

Sergio Ciampolillo

Maurizio Loreti

Dipartimento di Fisica
Università degli Studi di Padova

**Introduzione alle
Esperimentazioni di Fisica I**

Teoria degli errori e analisi dei dati



Marzo 1995
(Quinta Edizione)

Questo libro è stato completamente composto e stampato da uno degli autori (Maurizio Loreti) con un personal computer Commodore Amiga. Sono stati adoperati i programmi T_EX di Donald E. Knuth (nella versione L^AT_EX2_ε); i caratteri PostScript Adobe della famiglia Palatino ed il MathTime della T_EXplorators Corporation; ed una stampante Hewlett-Packard.

Indice

1	Introduzione	1
1.1	Il metodo scientifico	2
2	La misura	5
2.1	Misure dirette e misure indirette	5
2.2	Le unità di misura	6
2.3	Gli strumenti di misura	9
2.4	Errori di misura	10
2.5	Cifre significative ed arrotondamenti	15
2.6	Errore relativo	16
3	Elementi di teoria della probabilità	17
3.1	La probabilità: eventi e variabili casuali	17
3.2	La probabilità: definizioni	18
3.3	Proprietà della probabilità	19
3.4	Definizione assiomatica di probabilità	21
3.5	La legge dei grandi numeri	21
4	Elaborazione dei dati	23
4.1	Istogrammi	23
4.2	Stime di tendenza centrale	25
4.2.1	La moda	25
4.2.2	La mediana	28
4.2.3	La media aritmetica	28
4.2.4	Considerazioni complessive	29
4.2.5	Prima giustificazione della media	31
4.2.6	La media aritmetica espressa tramite le frequenze	31
4.3	Stime di dispersione	32

4.3.1	Semidispersione massima e quantili	32
4.3.2	Deviazione media assoluta	33
4.3.3	Varianza e deviazione standard	33
4.4	Giustificazione della media	34
5	Variabili casuali unidimensionali	37
5.1	Generalità	38
5.2	Valor medio e valor vero	39
5.3	Scarto ed errore quadratico medio	41
5.4	La varianza delle combinazioni lineari	42
5.5	L'errore della media	44
5.6	Stima dell'errore quadratico medio	44
5.7	Ancora sull'errore quadratico medio	45
6	La legge di Gauss	47
6.1	La densità di probabilità	47
6.2	La funzione di Gauss	50
6.3	Proprietà della legge normale	51
6.4	Lo scarto normalizzato	54
6.5	Il significato geometrico di σ	57
6.6	La curva di Gauss nella pratica	58
6.7	Esame dei dati	59
6.8	Sommario delle misure dirette	60
7	Le misure indirette	63
7.1	Risultato della misura	63
7.2	Combinazioni lineari di misure dirette	65
7.3	Errori nelle misure indirette	65
7.4	Errore dei prodotti di potenze	66
8	Altre tecniche di elaborazione dei dati	69
8.1	Media pesata	69
8.2	Interpolazione dei dati con una curva	72
8.2.1	Interpolazione lineare per due variabili	73
8.2.2	Stima a posteriori degli errori di misura	77
8.2.3	Interpolazione con una retta per l'origine	78
8.2.4	Interpolazione lineare nel caso generale	79
8.2.5	Interpolazione non lineare	81
A	L'errore della varianza	83

B	Covarianza e correlazione	89
B.1	La covarianza	89
B.2	Correlazione lineare	90
B.3	Applicazione all'interpolazione lineare	92
C	La distribuzione binomiale	95
D	La distribuzione di Poisson	99
E	Il modello di Laplace e la gaussiana	103
F	La verifica delle ipotesi	109
F.1	Compatibilità con un valore prefissato	109
F.2	La compatibilità di due valori misurati	112
F.3	I piccoli campioni e la distribuzione di Student	113
F.4	La distribuzione del χ^2	115
F.5	Compatibilità dei dati con una distribuzione	118
G	La funzione di verosimiglianza	121

Capitolo 1

Introduzione

Scopo della Fisica è lo studio dei fenomeni naturali, dei quali essa cerca per prima cosa di dare una descrizione, non solo qualitativa ma soprattutto quantitativa; tale studio si estende poi oltre la descrizione, e comprende l'indagine sulle relazioni reciproche tra più fenomeni, sulle cause che li producono e su quelle che ne determinano le modalità di presentazione.

Fine ultimo di tale studio è quello di formulare delle *leggi fisiche* capaci di dare del fenomeno stesso una descrizione razionale, quantitativa e (per quanto possibile) completa, che ci permetta di dedurre le caratteristiche con cui esso si verificherà a partire dalla conoscenza delle caratteristiche degli altri fenomeni che lo hanno causato, o che comunque con esso interagiscono.

Oggetto quindi della ricerca fisica devono essere delle *grandezze misurabili*, enti che cioè possiamo caratterizzare mediante la valutazione quantitativa di alcune caratteristiche, suscettibili di variare da caso a caso a seconda delle modalità con cui il fenomeno studiato si svolge¹.

¹Sono *grandezze misurabili* anche quelle connesse a oggetti non direttamente osservabili, ma su cui possiamo indagare attraverso lo studio delle influenze prodotte dalla loro presenza sull'ambiente che li circonda. Ad esempio i *quarks*, costituenti delle particelle elementari dotate di interazione forte, secondo le attuali teorie per loro stessa natura non potrebbero esistere isolati allo stato libero; le loro caratteristiche (carica, spin etc.) non sono quindi direttamente suscettibili di misura: ma sono ugualmente oggetto della ricerca fisica, in quanto sono osservabili e misurabili i loro effetti al di fuori della particella entro cui i quarks sono relegati.

1.1 Il metodo scientifico

Il linguaggio usato dal ricercatore per la formulazione delle leggi fisiche è il *linguaggio matematico*, che naturalmente si presta a descrivere i dati numerici che individuano i fenomeni, le loro variazioni ed i loro rapporti reciproci; il procedimento usato per giungere a tale formulazione è il *metodo scientifico*, la cui introduzione si fa storicamente risalire a Galileo Galilei. Esso può essere descritto distinguendone alcune fasi successive:

- una fase *preliminare* in cui, basandosi sul bagaglio delle conoscenze precedentemente acquisite, si determinano le grandezze rilevanti per la descrizione del fenomeno e quelle che presumibilmente influenzano le modalità con cui esso si presenterà;
- una fase *sperimentale* in cui si compiono osservazioni accurate del fenomeno, controllando e misurando sia le grandezze che lo possono influenzare sia quelle caratteristiche quantitative che lo individuano e lo descrivono, mentre esso viene causato in maniera (per quanto possibile) esattamente riproducibile; ed in questo consiste specificatamente il lavoro dei fisici sperimentali.
- una fase di *sintesi* o congettura, in cui, partendo dai dati numerici raccolti nella fase precedente, si inducono delle relazioni matematiche tra le grandezze misurate che siano in grado di render conto delle osservazioni stesse; si formulano cioè delle leggi fisiche ipotetiche controllando se esse sono in grado di spiegare il fenomeno.
- una fase *deduttiva* in cui dalle ipotesi formulate si traggono tutte le immaginabili conseguenze, particolarmente la previsione di fenomeni non ancora osservati (almeno non con la necessaria precisione); questo è specificatamente il compito dei fisici teorici.
- infine una fase di *verifica* delle ipotesi prima congettrate e poi sviluppate nei due passi precedenti, in cui si compiono ulteriori osservazioni sulle nuove speculazioni della teoria per accertarne l'esattezza.

Se nella fase di verifica si trova rispondenza con la realtà, l'ipotesi diviene una legge fisica accettata; se d'altra parte alcune conseguenze della teoria non risultano confermate, e non si trovano spiegazioni delle discrepanze tra quanto previsto e quanto osservato nell'ambito delle conoscenze acquisite, la legge dovrà essere modificata in parte, o rivoluzionata completamente per essere sostituita da una nuova congettura; si ritorna cioè alla fase di sintesi, e l'evoluzione della scienza comincia un nuovo ciclo.

Naturalmente, anche se non si trovano contraddizioni con la realtà ciò non vuol dire che la legge formulata sia esatta: è possibile che esperimenti effettuati in condizioni cui non si è pensato, o con strumenti di misura più accurati di quelli di cui si dispone ora, dimostrino in futuro che le nostre congetture erano in realtà sbagliate; come esempio, basti pensare alla legge galileiana del moto dei corpi ed alla teoria della relatività.

E' evidente quindi come le fasi di indagine e di verifica sperimentale costituiscano parte fondamentale dello studio dei fenomeni fisici; scopo di questo corso è quello di presentare la teoria delle misure e degli errori ad esse connessi.

Capitolo 2

La misura

Ad ogni grandezza fisica si deve, almeno in linea di principio, poter associare un valore numerico in modo univoco ed oggettivo, cioè riproducibile nelle stesse condizioni da qualsiasi osservatore; valore pari al rapporto fra la grandezza stessa e l'unità di misura per essa prescelta.

Per eseguire tale associazione dobbiamo disporre di strumenti e metodi che ci permettano di mettere in relazione da una parte la grandezza da misurare, e dall'altra l'unità di misura (oppure multipli o sottomultipli di essa); e ci dicano se esse sono uguali o, altrimenti, quale delle due è maggiore.

2.1 Misure dirette e misure indirette

La misura si dice *diretta* quando si confronta direttamente la grandezza misurata con l'unità di misura (*campione*) o suoi multipli o sottomultipli; per esempio la misura di una lunghezza mediante un regolo graduato è una misura diretta.

E' una misura diretta anche quella effettuata mediante l'uso di strumenti preparati (ad esempio la misura della temperatura mediante un termometro), che si fonda sulla proprietà dello strumento di reagire nella stessa maniera quando viene sottoposto alla medesima sollecitazione.

Misure *indirette* sono invece quelle in cui non si misura la grandezza che interessa, ma altre che risultino ad essa legate da una qualche relazione funzionale; così la velocità di un'automobile può essere misurata direttamente (tachimetro) o indirettamente, misurando spazi percorsi e tempi impiegati dai quali si risale poi alla velocità (media) con una operazione matematica.

2.2 Le unità di misura

Le grandezze fisiche si sogliono dividere in *fondamentali* e *derivate*. Con il primo di questi nomi si indicavano originariamente quelle grandezze misurate con strumenti e metodi sperimentali che richiedessero un confronto diretto con un campione, scelto arbitrariamente come unità di misura; mentre i valori delle seconde si intendevano determinati generalmente attraverso misure dirette di altre grandezze ad esse legate da relazioni algebriche, che ne fissavano anche le relative unità di misura.

In questo modo si sono definiti vari sistemi di misura coerenti come il Sistema Internazionale (**SI**) attualmente in uso, il quale assume come fondamentali lunghezza, massa, tempo, intensità di corrente elettrica, temperatura, intensità luminosa e quantità di materia, con le rispettive unità metro, kilogrammo, secondo, Ampère, grado Kelvin, candela e mole. Le unità delle altre grandezze sono univocamente determinate dalle relazioni algebriche che le legano alle grandezze fondamentali.

Se ciascuna unità fondamentale viene ridotta di un certo fattore, il valore della grandezza espresso nelle nuove unità risulta moltiplicato per un prodotto di potenze dei medesimi fattori. Così, per restare nell'ambito della meccanica, se riduciamo l'unità di lunghezza di un fattore L , l'unità di massa di un fattore M e quella di tempo di un fattore T , ed il valore di una grandezza fisica ne risulterà in conseguenza moltiplicato per

$$L^\lambda M^\mu T^\tau$$

si dice che la grandezza in questione ha le *dimensioni* di una lunghezza elevata alla potenza λ per una massa elevata alla potenza μ per un tempo elevato alla potenza τ .

Ad esempio la velocità (media) di un corpo in movimento è definita come il rapporto tra lo spazio da esso percorso in un certo intervallo di tempo e la durata di tale intervallo, ed è dunque una grandezza derivata. Una volta scelte le unità di misura delle lunghezze e dei tempi (per esempio metro e secondo), l'unità di misura delle velocità risulta fissata univocamente (1 metro al secondo). Se si alterano ad esempio l'unità di lunghezza moltiplicandola per il fattore $1/L = 1000$ (kilometro), quella di tempo moltiplicandola per il fattore $1/T = 3600$ (ora) e quella di massa moltiplicandola per un fattore $1/M = 1000$ (tonnellata), il valore di qualunque velocità nella nuova unità (kilometro all'ora) risulterà alterato rispetto al precedente di un fattore

$$L^1 T^{-1} M^0 = LT^{-1}$$

e si dice pertanto che le dimensioni fisiche di una velocità sono quelle di una lunghezza divisa per un tempo.

Come altro esempio si consideri l'energia cinetica di un corpo, definita come il lavoro compiuto dalla forza che si deve applicare per arrestarlo, e che è pari numericamente alla metà del prodotto della massa per il quadrato della velocità del corpo stesso:

$$K = \frac{1}{2}mv^2$$

Essa è pertanto una grandezza derivata, la cui unità di misura nel Sistema Internazionale è l'energia cinetica di un corpo di 2 Kg in moto traslatorio con velocità di 1 m/s (unità detta *Joule*). Passando al nuovo sistema di unità prima definito (assai inconsueto per un'energia), il valore di K risulta moltiplicato per il fattore $M^1L^2T^{-2}$; si dice dunque che un'energia ha le dimensioni di una massa per il quadrato di una lunghezza divisa per il quadrato di un tempo.

Queste proprietà di trasformazione sono legate alla cosiddetta *analisi dimensionale* ed alla *similitudine meccanica*, argomenti che esulano da questo corso.

Basti qui osservare che il numero di unità indipendenti non coincide necessariamente con quello delle grandezze assunte come "fondamentali"; così l'angolo piano e l'angolo solido sono entrambi privi di dimensioni in termini di grandezze fisiche fondamentali, e come tali dovrebbero avere come unità di misura derivata ($1m/1m$ e rispettivamente $1m^2/1m^2$) lo stesso "numero puro" 1, mentre esistono per essi due diverse unità: il radiante e lo steradiante, quasi essi avessero dimensioni proprie e distinte.

Né vi è alcunchè di necessario nella scelta delle grandezze fondamentali quale si è venuta configurando storicamente nel Sistema Internazionale, potendosi definire un sistema coerente anche con l'assegnazione di valori convenzionali alle costanti universali della fisica, come proposto agli inizi del secolo da Planck: così un sistema di unità "naturali" si potrebbe fondare, in linea di principio, ponendo eguali ad 1 la velocità della luce, il quanto d'azione (o costante di Planck), la costante di gravitazione universale, la costante di Boltzmann ed il quanto elementare di carica elettrica (ovverosia la carica dell'elettrone); ma, a parte considerazioni di opportunità e consuetudine, ciò che determina in ultima analisi fino a che punto si possa tradurre in pratica un simile programma, e quali grandezze siano da considerare fondamentali, è la riproducibilità dei campioni e la precisione con cui è possibile il confronto diretto tra grandezze omogenee.

E' emblematica a questo riguardo la storia dell'evoluzione delle unità di misura delle lunghezze, che anticamente erano le lunghezze degli arti dell'uomo quali il braccio, il cubito (risalente agli Egizi), il piede e la larghezza del pollice; ovvero delle medie di tali lunghezze su di un numero limitato di individui. L'ovvio vantaggio di una simile definizione è la disponibilità del campione in ogni tempo e luogo; l'altrettanto ovvio svantaggio è la grande variabilità del campione stesso, donde il ricorso a valori medi e infine a campioni artificiali costruiti con materiali

e accorgimenti che garantissero una minima variabilità della lunghezza col tempo, e con le condizioni esterne più o meno controllabili.

Così, dopo la parentesi illuministica che portò all'adozione della quarantamillesima parte del meridiano terrestre quale unità di lunghezza (metro), e fino al 1960, il metro campione fu la distanza tra due tacche tracciate su di un'opportuna sezione di una sbarra costituita da una lega metallica molto stabile; tuttavia le alterazioni spontanee della struttura microcristallina della sbarra fanno sì che diversi campioni, aventi la medesima lunghezza alla costruzione, presentino col tempo delle differenze apprezzabili col metodo di confronto. Inoltre l'uso di metodi ottici interferenziali finì per consentire un confronto più preciso delle lunghezze e condusse nel 1960 (come già suggerito da Babinet nel 1829!) a svincolare la definizione del metro dalla necessità di un supporto materiale macroscopico, col porlo eguale a $1'650'763,73$ volte l'effettivo campione: cioè la lunghezza d'onda nel vuoto della luce emessa in opportune condizioni da una sorgente atomica (riga arancio dell'isotopo del Krypton ^{86}Kr).

L'ulteriore rapido sviluppo della tecnologia, con l'avvento di laser molto stabili e di misure accuratissime delle distanze planetarie col metodo del radar, ha condotto recentemente (1984) ad una nuova definizione del metro, come distanza percorsa nel vuoto dalla luce in una determinata frazione ($1/299'792'458$) dell'unità di tempo (secondo); il che equivale ad assumere un valore convenzionale per il campione di velocità (la velocità della luce nel vuoto) ed a ridurre la misura della lunghezza fondamentale ad una misura di tempo.

Le misure di lunghezza hanno dunque percorso l'intero arco evolutivo, ed appare evidente come la complessa realtà metrologica odierna non sia più riflessa esattamente nella classificazione tradizionale di grandezze "fondamentali" e "derivate". Infatti la velocità assume ora in un certo senso il ruolo di grandezza fondamentale, e tuttavia una velocità non si misura praticamente mai per confronto diretto col campione (la velocità della luce nel vuoto); per converso le lunghezze sono spesso ancor oggi misurate per confronto con campioni, ma la lunghezza del campione primario (il metro) è determinata da una misura di tempo.

Per quanto riguarda l'unità di durata temporale, essa fu svincolata da un supporto macroscopico (il moto diurno della terra o i moti planetari) nel 1964 con l'adozione di un campione di frequenza atomica (in termini imprecisi il cosiddetto "orologio atomico al Cesio"), assegnando il valore convenzionale di $9'192'631'770$ cicli al secondo (Hertz) alla frequenza della radiazione elettromagnetica emessa in una particolare transizione tra due stati interni dell'atomo di ^{133}Cs .

Questa definizione del minuto secondo consente il confronto di intervalli di tempo con un errore relativo¹ inferiore ad una parte su 10^{13} . Se si considera che il quanto d'azione \hbar , che è la costante universale della meccanica (quantistica)

¹Vedi il paragrafo 2.6 alla fine del corrente capitolo.

determinata con maggior precisione dopo la velocità della luce e che sia da essa indipendente, è nota soltanto con una incertezza di alcune parti su 10^6 , si comprende quale iato si dovrebbe colmare per portare a compimento il programma di Planck anche con il tempo, così come lo si è realizzato per la lunghezza.

Ad uno stadio ancor meno avanzato è giunta l'evoluzione della misura di massa, il cui campione è tuttora costituito da un particolare oggetto macroscopico detto "kilogrammo-campione". Anche qui la precisione con cui si possono confrontare le masse supera di vari ordini di grandezza quella con cui è nota la costante di gravitazione universale, cui l'attribuzione di un valore convenzionale consentirebbe di ridurre le misure di massa a quelle di tempo e di lunghezza.

2.3 Gli strumenti di misura

Lo strumento di misura è un apparato che permette il confronto tra la grandezza misurata e l'unità di misura. Esso è costituito da un oggetto sensibile in qualche modo alla grandezza da misurare, che si può chiamare *rivelatore*; eventualmente da un dispositivo *trasduttore*, che traduce le variazioni della grandezza caratteristica del rivelatore in quelle di un'altra grandezza più facilmente accessibile allo sperimentatore; e da un dispositivo che presenta il risultato della misura ai sensi, generalmente alla vista, dello sperimentatore, direttamente o con una registrazione grafica o di altro genere.

Così in un *calibro*, strumento per la misura di spessori, il rivelatore è costituito dalla ganascia mobile col cursore ad essa solidale, e che può scorrere nella guida facente corpo unico con la ganascia fissa; mentre l'elemento indicatore è costituito dalla scala graduata in millimetri tracciata sulla guida e dal segno di fede inciso sul cursore, generalmente insieme ad una scala graduata ausiliaria (*nonio*) per la lettura delle frazioni di millimetro. La grandezza letta sulla scala è qui direttamente lo spessore oggetto della misura.

In un *termometro* a liquido l'elemento sensibile alla temperatura è il liquido contenuto nel bulbo; esso funge almeno in parte anche da trasduttore, perchè la proprietà termometrica che si osserva è il volume del rivelatore stesso. Il tubo capillare a sezione costante traduce le variazioni di volume del rivelatore in variazioni di lunghezza della colonna di liquido ivi contenuta; il menisco che separa il liquido dal suo vapore nel capillare funge da indicatore, assieme con la scala tracciata sulla superficie esterna del tubo stesso o sopra un regolo ad essa solidale. La grandezza letta sulla scala è la distanza del menisco da un segno di riferimento che può essere messa in corrispondenza con la temperatura per mezzo di una tabella di conversione; o, più spesso e comodamente, le temperature corrispondenti sono scritte sulla scala stessa accanto alle tacche della graduazione. Le caratteristiche più importanti di uno strumento sono le seguenti:

- La *prontezza*: è determinata dal tempo necessario perchè lo strumento risponda ad una variazione della sollecitazione; ad esempio, per avere una risposta corretta da un termometro si deve attendere che si raggiunga l'equilibrio termico tra il rivelatore e l'oggetto di cui si misura la temperatura.
- L'*intervallo d'uso*: è definito come l'insieme dei valori compresi tra la *soglia* e la *portata* dello strumento, cioè tra il minimo ed il massimo valore della grandezza che lo strumento può apprezzare in un singolo atto di misura.
- La *sensibilità*: si può definire come il reciproco della incertezza di lettura propria dello strumento, cioè della più piccola variazione della grandezza che può essere letta sulla scala, e che si assume generalmente corrispondente alla più piccola divisione della scala stessa—o ad una frazione apprezzabile di questa. La sensibilità può essere diversa in differenti punti della scala, o per diversi valori della grandezza; è un fattore che limita l'intervallo d'uso dello strumento, potendo divenire insufficiente al di sotto della soglia od al di sopra della portata.
- La *precisione* dello strumento: è legata alla riproducibilità del risultato della misura di una stessa grandezza. Esso può variare da una parte per difetti dello strumento dovuti alla costruzione, che non può mai essere perfetta, e al logoramento di alcune componenti in conseguenza dell'uso prolungato o improprio, o dell'invecchiamento; dall'altra parte, per la presenza di varie cause di disturbo ineliminabili anche in condizioni normali d'uso dello strumento stesso. La precisione si può definire come il reciproco dell'incertezza sul valore della grandezza che viene determinata dall'insieme di questi fattori.

Per sfruttare a pieno le possibilità di uno strumento di misura, è opportuno che la sensibilità non sia inferiore alla precisione; gli strumenti di uso corrente sono costruiti con una sensibilità circa eguale alla precisione in condizioni normali d'uso.

2.4 Errori di misura

Come accennato in relazione alla precisione di uno strumento, se si esegue una misura di una qualsiasi grandezza fisica si commettono inevitabilmente errori; conseguentemente il valore ottenuto per la grandezza misurata non è mai esattamente eguale al suo vero valore, che non sarà perciò mai noto con precisione arbitrariamente grande (diversamente da quanto accade con una costante matematica, come ad esempio π).

Quando si ripete la misura della stessa grandezza col medesimo strumento, nelle medesime condizioni e seguendo la medesima procedura, la presenza delle varie cause di errore che andremo ad esaminare produce delle differenze tra il valore misurato ed il valore vero; differenze variabili da una misura all'altra, ed in modo imprevedibile singolarmente. In conseguenza di ciò, i risultati di queste misure ripetute (se lo strumento è abbastanza sensibile) fluttueranno apprezzabilmente in maniera casuale in un certo intervallo, la cui ampiezza definisce la precisione delle misure stesse. Gli errori di questo tipo si dicono *errori casuali*, e la loro esistenza è facilmente accertabile con l'uso di un qualsiasi strumento sensibile.

Tuttavia certe cause di errore possono dar luogo a una discrepanza tra valore misurato e valore vero che si riproduce inalterata nelle misure ripetute di cui sopra, e la inosservabilità delle fluttuazioni non garantisce affatto che la discrepanza sia inferiore all'incertezza di lettura dello strumento; né si può esser certi che essa sia contenuta entro l'intervallo di variabilità degli errori casuali (quando esso sia maggiore dell'incertezza di lettura).

Gli errori di questo secondo tipo si dicono *errori sistematici* e sono i più insidiosi, perchè non immediatamente identificabili. Cause di errori sistematici possono essere quelle elencate nel seguito (ma la lista non è necessariamente completa):

1. *Difetti dello strumento, risalenti alla costruzione o conseguenti al suo deterioramento.* Ad esempio, in una bilancia con bracci di lunghezza diversa, l'eguaglianza dei momenti applicati ai due bracci ed assicurata dall'equilibrio del giogo non implica l'eguaglianza delle masse ad essi sospese: perchè una massa minore sospesa al braccio più lungo farà equilibrio ad una massa maggiore sospesa all'altro. (Questo errore si può anche classificare nel tipo 6, cioè come errore di interpretazione del risultato).

Altro esempio è quello di un goniometro *eccentrico*, cioè con la croce centrale o l'asse di rotazione in posizione diversa dal centro del cerchio recante la graduazione. Ciò può dar luogo, per esempio, a misure di angoli acuti sistematicamente in difetto o in eccesso a seconda della posizione del centro presunto rispetto agli assi $0^\circ - 180^\circ$ e $90^\circ - 270^\circ$ del goniometro.

Lo zero di una scala (ad esempio di un termometro) può essere spostato dalla posizione corretta di taratura, per cui tutte le letture saranno in difetto o in eccesso in dipendenza dalla direzione dello spostamento. Oppure la scala stessa dello strumento può essere difettosa: così, se il capillare di un termometro non ha sezione costante, anche se le posizioni corrispondenti a due punti fissi come 0°C e 100°C sono esatte, le temperature risultano in difetto in un tratto della scala ed in eccesso in un altro tratto.

2. *Uso dello strumento in condizioni errate*, cioè diverse da quelle previste per il suo uso corretto. Tale è l'uso di regoli, calibri e simili strumenti per misurare lunghezze, o di recipienti tarati per la misura di volumi, a temperature diverse da quella di taratura (generalmente fissata a 20 °C); infatti la dilatazione termica farà sí che lunghezza e volumi risultino alterati in difetto o in eccesso a seconda che si operi a temperatura superiore o inferiore. Si può naturalmente commettere un errore anche usando lo strumento a 20 °C, quando ciò che interessa in realtà è conoscere il valore di una grandezza dipendente dalla temperatura (la lunghezza di un oggetto, il volume di un corpo, la resistenza elettrica di un filo o qualsiasi altra) ad una temperatura diversa da 20 °C.
3. *Errori di stima da parte dello sperimentatore*: un esempio di questo tipo di errore si ha quando nello stimare una certa frazione di divisione di una scala graduata lo sperimentatore tende a valutarla sempre in difetto o sempre in eccesso; oppure, nel leggere la posizione di un indice mobile di fronte ad una scala, lo sperimentatore può tenere l'occhio sistematicamente alla sinistra o alla destra del piano passante per l'indice ed ortogonale alla scala (*errore di parallasse*).
4. *Perturbazioni esterne*; un esempio di errori di questo tipo è la presenza di corpi estranei, come la polvere, interposti tra le ganasce di un calibro e l'oggetto da misurare, il che porta a sovrastimarne lo spessore.
Un altro esempio è la misura della profondità del fondo marino o fluviale con uno scandaglio (filo a piombo) in presenza di corrente, la quale fa deviare il filo dalla verticale e porta sempre a sovrastimare la profondità se il fondo è orizzontale o quasi.
5. *Perturbazione del fenomeno osservato da parte dell'operazione di misura*. Tra gli errori di questo tipo si può citare la misura dello spessore di un oggetto con un calibro a cursore, o col più sensibile calibro a vite micrometrica (Palmer); l'operazione richiede l'accostamento delle ganasce dello strumento all'oggetto, ed in essa lo si comprime inevitabilmente con una forza sia pur piccola, e se ne provoca perciò una deformazione con leggera riduzione dello spessore.
6. *Uso di formule errate o approssimate nelle misure indirette*. Un esempio è offerto dalla misura indiretta dell'accelerazione di gravità g ottenuta dalla misura della lunghezza (cosiddetta ridotta) l di un apposito tipo di pendolo (di Kater) e dalla misura del suo periodo di oscillazione T , con la formula

$$g = \frac{4\pi^2 l}{T^2}$$

ottenuta dalla nota espressione del periodo

$$T_0 = 2\pi\sqrt{\frac{l}{g}}$$

Ma questa formula vale solo al limite per oscillazioni di ampiezza infinitesima, mentre una formula che meglio approssima la realtà è²

$$T(\theta) = 2\pi\sqrt{\frac{l}{g}} \left(1 + \frac{\theta^2}{16}\right) = T_0 \left(1 + \frac{\theta^2}{16}\right)$$

la quale mostra come il periodo sia una funzione leggermente crescente dell'ampiezza massima θ delle oscillazioni (qui espressa in radianti). L'uso della formula di prima approssimazione per determinare g comporta dunque una sottostima, che diviene tanto più sensibile quanto maggiore è θ , in quanto si usa in luogo di T_0 la durata T di una oscillazione reale avente ampiezza non nulla e perciò sempre superiore a T_0 .

La medesima misura è affetta anche da un altro errore sistematico, originato dal fatto che il pendolo non ruota oscillando attorno al filo orizzontale del coltello di sospensione, ma compie un moto in cui il profilo del taglio del coltello (che è approssimativamente un cilindro con raggio di curvatura minimo dell'ordine dei centesimi di millimetro) rotola sul piano di appoggio. A causa dell'impossibilità di una perfetta realizzazione meccanica dell'apparato, il fenomeno osservato è diverso da quello supposto che si intendeva produrre, e la sua errata interpretazione comporta invece una sovrastima di g . Infatti la formula del periodo, corretta per questo solo effetto, risulta essere

$$T = T_0\sqrt{1 - \frac{r}{a}}$$

(in cui r è il raggio di curvatura del filo del coltello ed a la distanza del centro di massa dal punto di appoggio) ed il T reale è sempre inferiore a T_0 .

Un modo per rivelare la presenza di errori sistematici insospettiti può essere quello di misurare, se possibile, la stessa grandezza con strumenti e metodi diversi; questi presumibilmente sono affetti da errori diversi e possono fornire perciò risultati differenti. Tuttavia neppure l'assenza di questo effetto dà la certezza che la misura sia esente da errori sistematici, ed essi sono generalmente individuati da

²Riguardo a questo punto ed al successivo, per una discussione approfondita del moto del pendolo si può consultare: G. Bruhat - Cours de Mécanique Physique - Ed. Masson, pagg. 311-321.

una attenta e minuziosa analisi critica sia dello strumento usato sia della procedura seguita nella misura.

Una volta scoperto, un errore sistematico può essere eliminato modificando lo strumento o la procedura, oppure apportando una opportuna correzione al risultato della misura (sebbene questa correzione comporti generalmente un aumento dell'errore casuale).

I primi cinque tipi elencati di possibili cause d'errore sistematico possono produrre anche errori casuali: così, per il primo tipo, gli inevitabili giochi meccanici e gli attriti tra parti dello strumento in moto relativo possono dar luogo a risultati fluttuanti; per quanto riguarda il secondo tipo, condizioni ambientali variabili e non del tutto controllabili (come temperatura e pressione) possono produrre variazioni imprevedibili del risultato.

Lo sperimentatore non ha un comportamento fisso e costante nelle valutazioni e nelle azioni compiute durante l'operazione di misura; come un esempio di questo terzo tipo di errori si consideri l'imprevedibile variabilità del tempo di reazione nell'avvio e nell'arresto di un cronometro a comando manuale.

Anche i disturbi esterni (quarto tipo), potendo essere di natura e intensità variabile, produrranno errori di un segno determinato (sistematici), ma di entità variabile ed imprevedibile, dunque in parte anche casuali.

Si aggiunga a ciò che disturbi casuali possono essere presenti nello strumento stesso per la costituzione corpuscolare della materia e per la natura fondamentalmente statistica di certe grandezze fisiche. Così l'equipaggio mobile, sospeso ad un filo lungo e sottile, di una bilancia a torsione di estrema sensibilità, avrà posizioni fluttuanti attorno a quella di equilibrio: non solo a causa del bombardamento incessante cui esso è sottoposto da parte delle molecole del gas circostante, ma anche nel vuoto assoluto per l'agitazione termica dei suoi stessi costituenti.

Infine, anche le cause del quinto tipo possono dar luogo ad errori casuali se il disturbo del fenomeno o dell'oggetto prodotto dall'operazione di misura è di entità variabile e non controllata.

Alle cause comuni con gli errori sistematici si deve qui aggiungere una tipica degli errori casuali, e consistente nella *imperfetta definizione della grandezza* che si intende misurare. Anche restando nell'ambito della fisica classica (e come accennato in relazione ai disturbi delle misure), certe grandezze, quali la pressione e la temperatura, sono in realtà legate a delle medie statistiche, come l'energia cinetica media molecolare; come tali esse hanno un'indeterminazione intrinseca, che tuttavia non si manifesta nelle misure relative ad oggetti e fenomeni macroscopici se non in casi eccezionali.

Ad un livello meno fondamentale, se si misura più volte con un calibro il diametro di un oggetto sferico può avvenire che i risultati siano leggermente diversi; questo perchè l'oggetto non può essere perfettamente sferico, ed ogni suo diametro ha una lunghezza generalmente diversa da quella di un altro.

Per concludere, gli errori casuali:

- Sono osservabili solo con uno strumento sufficientemente sensibile, cioè quando sono di entità maggiore dell'incertezza di lettura della scala.
- Possono essere ridotti; ad esempio affinando le caratteristiche dello strumento, o controllando più strettamente le condizioni del suo uso e dell'ambiente, e precisando meglio la procedura di esecuzione della misura: ma ciò con difficoltà crescente sempre più con la precisione; *non possono quindi mai essere eliminati*.
- Posseggono tuttavia certe regolarità statistiche, che studieremo nell'ambito di una teoria matematica che verrà affrontata nei prossimi capitoli; la loro entità può pertanto essere *stimata*.

Compito della *teoria dell'errore* è appunto quello di stimare l'errore presumibilmente commesso nell'atto della misura, a partire dai dati sperimentali stessi. Riassumendo:

Scopo della misura di una grandezza fisica è il valutare sia il rapporto della grandezza stessa con una certa unità di misura, sia l'errore da cui tale rapporto è presumibilmente affetto.

Il risultato delle misure dovrà quindi sempre essere espresso in una forma del tipo

$$l = 12.34 \pm 0.01 \text{ m}$$

in cui compaiano le tre parti *valore, errore ed unità di misura*.

2.5 Cifre significative ed arrotondamenti

Sempre per quanto riguarda il modo di esprimere il risultato di un insieme di misure, è un errore spingere la valutazione del risultato al di là della precisione sperimentale; in altre parole, se il calcolo dell'errore per la misura di una lunghezza indica incertezza sulla cifra ad esempio dei centimetri, è un errore dare nel risultato la cifra dei millimetri, o (peggio) dei decimi e centesimi di millimetro. Nei risultati intermedi possiamo tenere per i successivi calcoli tutte le cifre che vogliamo; ma, giunti *al risultato finale*, e solo una volta che l'errore sia stato calcolato, bisogna troncare il risultato stesso al livello dell'errore da noi stimato ed arrotondare. Così

$$\begin{array}{l} 12.34567 \pm 0.231 \quad \text{diventa} \quad 12.3 \pm 0.2 \quad \text{o} \quad 12.34 \pm 0.23 \\ 12.34567 \pm 0.00789 \quad \text{diventa} \quad 12.346 \pm 0.008 \quad \text{o} \quad 12.3457 \pm 0.0079 \end{array}$$

2.6 Errore relativo

Una volta valutato l'errore presumibile Δx (*errore assoluto*) della misura x_0 di una grandezza fisica x , il rapporto

$$\varepsilon = \frac{\Delta x}{x_0}$$

(indicato in *valore* od in *percentuale*) prende il nome di *errore relativo*.

L'errore relativo è importante perchè esprime la bontà della misura di una grandezza: è evidente come un errore assoluto di 1 cm assuma diverso significato se riferito alla misura di un tavolo o di una distanza astronomica, ed è la diversità degli errori relativi ad esprimere tale significato.

E' opportuno tuttavia osservare che l'errore relativo è privo di senso quando il valore vero della grandezza che si misura è nullo; pertanto si potrà parlare di errore relativo solo quando si possa escludere tale eventualità con pratica certezza, nel caso cioè che sia $x_0 \gg \Delta x$, ovvero che ε sia di almeno un ordine di grandezza minore dell'unità.

Capitolo 3

Elementi di teoria della probabilità

Abbiamo già notato come, per la ineliminabile presenza degli errori di misura, quello che otteniamo come risultato della stima del valore di una grandezza fisica non sia praticamente mai il valore vero della grandezza stessa; inoltre, se ripetiamo più volte la misura, non otteniamo mai, in generale, nemmeno lo stesso risultato.

Da questo si deduce che, sulla base di misure ripetute comunque effettuate, non si potrà mai affermare che un qualsiasi numero reale sia o non sia il valore vero della grandezza stessa. E' però evidente che tutti gli infiniti numeri reali non devono essere posti sullo stesso piano: alcuni di essi saranno più verosimili (intuitivamente i numeri vicini ai risultati delle nostre misure ripetute), altri (i più lontani) saranno meno verosimili.

Il problema della misura va dunque impostato in termini probabilistici, e potremo dire di averlo risolto quando, a partire dai dati sperimentali, saremo in grado di determinare un intervallo di valori avente una assegnata probabilità di contenere il valore vero.

Prima di proseguire, introduciamo dunque alcuni elementi della *teoria della probabilità*.

3.1 La probabilità: eventi e variabili casuali

Oggetto della teoria delle probabilità è lo studio di eventi *casuali* o *aleatori*, cioè eventi ripetibili (almeno in teoria) infinite volte, che possono manifestarsi in

più modi, imprevedibili singolarmente, e che si escludono a vicenda l'un l'altro; esempi tipici di eventi casuali sono il lancio di un dado o di una moneta, o l'estrazione di una carta da un mazzo. Come risultato del lancio della moneta o del dado, questi cadranno e si ridurranno in quiete con una determinata faccia rivolta verso l'alto; per la moneta le possibilità sono due, mentre per il dado sono sei.

Se si è in grado di fissare una legge di corrispondenza che permetta di associare ad ogni modalità E di presentarsi dell'evento casuale, scelta nell'insieme S di tutti i possibili risultati, uno ed un solo numero reale $x(E)$, questo numero prende il nome di *variabile casuale* definita nell'insieme S . Le variabili casuali possono assumere un numero finito od infinito di valori, possono essere discrete o continue; è da notare come, per la presenza degli errori, la misura di una grandezza possa essere considerata come un evento casuale, ed il numero che otteniamo una variabile casuale che associamo all'evento stesso.

3.2 La probabilità: definizioni

La definizione "classica" di probabilità è la seguente:

Si definisce come probabilità di un evento casuale il rapporto tra il numero di casi favorevoli al presentarsi dell'evento stesso ed il numero totale di casi possibili, purchè tutti i casi possibili siano ugualmente probabili.

Se ne ricava immediatamente il seguente

Corollario: *la probabilità di un evento casuale è un numero reale compreso tra zero e uno, che assume il valore zero per gli eventi impossibili ed uno per quelli certi.*

La definizione "classica" sembra sufficiente a permetterci di calcolare le probabilità di semplici eventi casuali che possano manifestarsi in un numero finito di modalità equiprobabili (ad esempio per i giochi d'azzardo), ma è insoddisfacente perchè racchiude in sé stessa una *tautologia*: si nota immediatamente come, per definire la probabilità, si presupponga che si sia già in grado di valutare l'equiprobabilità delle varie possibilità di manifestarsi dell'evento considerato. Nel caso di una variabile casuale continua, ciò si traduce nell'indeterminazione di quale tra le variabili topologicamente equivalenti (legate da trasformazioni continue) sia quella equiprobabile, cioè con probabilità per ogni intervallo proporzionale all'ampiezza dell'intervallo stesso.

Si possono dare della probabilità definizioni più soddisfacenti dal punto di vista logico, ad esempio la seguente (definizione *empirica*¹): definiamo la *frequenza*

¹Anche questa definizione (teorizzata da Von Mises) non è completamente soddisfacente dal punto di vista concettuale, ma è tra le più intuitive perchè tra le più vicine all'uso pratico.

relativa $f(E)$ con cui un evento casuale E si è presentato in un numero totale N di casi reali come il rapporto tra il numero n di volte in cui l'evento si è effettivamente prodotto (*frequenza assoluta*) ed il numero N delle prove effettuate; la probabilità di E si definisce euristicamente come l'estensione del concetto di frequenza relativa su un numero grandissimo di prove, cioè

$$p(E) \approx \lim_{N \rightarrow \infty} f(E) = \lim_{N \rightarrow \infty} \left(\frac{n}{N} \right)$$

3.3 Proprietà della probabilità

La mancata realizzazione dell'evento E costituisce l'*evento complementare* \bar{E} ; i due eventi E ed \bar{E} si escludono mutuamente, ed esauriscono l'insieme di tutti i possibili risultati di una prova od esperimento elementare del tipo considerato.

La frequenza relativa di \bar{E} su N prove è

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

da cui si ricava

$$p(\bar{E}) = 1 - p(E) \quad \text{o anche} \quad p(E) + p(\bar{E}) = 1$$

Generalizzando, se E, F, \dots, Z sono eventi casuali *mutuamente esclusivi* e che *esauriscono* l'insieme di tutti i possibili risultati, vale la

$$p(E) + p(F) + \dots + p(Z) = 1$$

Il risultato di una prova o esperimento più complesso può essere costituito dal verificarsi di due eventi simultanei in luogo di uno solo; come esempio, si consideri il lancio di una moneta e l'estrazione contemporanea di una carta da un mazzo. Se E indica l'apparizione della testa (\bar{E} allora sarà l'apparizione della croce) ed F l'estrazione di una carta nera (\bar{F} di una carta rossa), esistono quattro eventi fondamentali non ulteriormente decomponibili e che si escludono vicendevolmente: $EF, E\bar{F}, \bar{E}F$ e $\bar{E}\bar{F}$.

Il simbolo EF indica qui l'evento composto *prodotto logico* dei due eventi semplici E ed F , cioè il verificarsi sia dell'uno che dell'altro. Se ora, su N prove effettuate, la frequenza assoluta con cui i quattro eventi fondamentali si sono verificati è quella indicata nella seguente tabella:

	F	\bar{F}
E	n_{11}	n_{12}
\bar{E}	n_{21}	n_{22}

le rispettive frequenze relative saranno

$$\begin{aligned} f(EF) &= \frac{n_{11}}{N} & f(E\bar{F}) &= \frac{n_{12}}{N} \\ f(\bar{E}F) &= \frac{n_{21}}{N} & f(\bar{E}\bar{F}) &= \frac{n_{22}}{N} \end{aligned}$$

Facendo uso della definizione empirica di probabilità si trova, conformemente alle precedenti relazioni valide per eventi che si escludono:

$$\begin{aligned} E = EF + E\bar{F} &\quad \Rightarrow \quad f(E) = \frac{n_{11} + n_{12}}{N} \\ F = EF + \bar{E}F &\quad \Rightarrow \quad f(F) = \frac{n_{11} + n_{21}}{N} \end{aligned}$$

da cui si ricavano le

$$p(E) = p(EF) + p(E\bar{F}) \quad \text{e} \quad p(F) = p(EF) + p(\bar{E}F)$$

e simili per \bar{E} e \bar{F} . Se ora si applica la definizione empirica all'evento complesso $E + F$ *somma logica* degli eventi semplici E ed F (che consiste cioè nel verificarsi o di E o di F o di entrambi), si ottiene

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

da cui

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

Nel caso particolare di due eventi E ed F che si escludano mutuamente (cioè per cui sia $p(EF) = 0$ e $n_{11} = 0$) vale la cosiddetta *legge della probabilità totale*: $p(E + F) = p(E) + p(F)$. Questa si generalizza poi per induzione completa al caso di più eventi (sempre però *mutuamente esclusivi*), per la cui somma logica la probabilità è uguale alla somma delle probabilità degli eventi semplici:

$$p(E + F + \dots + Z) = p(E) + p(F) + \dots + p(Z)$$

La probabilità che si verifichi l'evento E nei casi in cui si sa già che si è verificato l'evento F si indica con il simbolo $p(E|F)$ e si chiama *probabilità condizionata*; si ricava per essa facilmente, usando la terminologia dell'esempio,

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

Vale quindi la

$$p(EF) = p(F) \cdot p(E|F)$$

Nel caso particolare in cui risulti $p(E|F) = p(E)$, il verificarsi o meno dell'evento F non modifica la probabilità che E ha di manifestarsi. In questo caso i due eventi si dicono *statisticamente indipendenti*, e per essi vale la seguente legge (della *probabilità composta*): $p(EF) = p(E) \cdot p(F)$.

Questa si generalizza facilmente per un evento costituito dal verificarsi contemporaneo di più di due eventi (sempre però *statisticamente indipendenti*), per il quale vale la

$$p(EF \cdots Z) = p(E) p(F) \cdots p(Z)$$

3.4 Definizione assiomatica di probabilità

Solo per completezza accenniamo qui alla cosiddetta *definizione assiomatica della probabilità*, che è matematicamente consistente:

Sia S l'insieme di tutti i possibili risultati di una prova, ed A un qualsiasi sottoinsieme di S . Si definisce "probabilità" di A un numero associato univocamente al sottoinsieme e che soddisfi le seguenti proprietà:

1. $p(A) \geq 0$ per ogni A ;
2. $p(S) = 1$;
3. $p(A + B) = p(A) + p(B)$ per ogni $A \subset S$ e $B \subset S$ tali che $A \cap B = \emptyset$.

Questa definizione, pur matematicamente consistente, non dice nulla su come assegnare dei valori alla probabilità; tuttavia su tali valori si possono fare delle ipotesi, verificabili analizzando gli eventi reali osservati.

3.5 La legge dei grandi numeri

Sempre per completezza, inseriamo un breve cenno alla cosiddetta legge dei grandi numeri. Difetto della definizione empirica di probabilità, oltre a quello di essere basata su di un esperimento, è quello di supporre una convergenza di $f(N)$, al crescere di N , verso un valore ben definito: valore che si assume poi come probabilità dell'evento.

Qualora si prenda come definizione di probabilità quella assiomatica, è effettivamente possibile dimostrare che, al crescere del numero di prove, la frequenza

relativa di un evento casuale converge verso la probabilità dell'evento stesso. E' tuttavia importantissimo sottolineare come questa legge (*dei grandi numeri*) non implichi una convergenza esatta nel senso dell'analisi: non implichi cioè che, scelto un qualunque numero positivo ε , sia possibile determinare in conseguenza un intero M tale che per ogni $N > M$ risulti sicuramente $|f(E) - p(E)| < \varepsilon$.

Si pensi in proposito alla chiara impossibilità di fissare un numero M tale che, quando si lanci un dado più di M volte, si sia *certi* di ottenere almeno un sei: al crescere di M crescerà la *probabilità* del verificarsi di questo evento, ma non si potrà mai raggiungere la certezza.

Nella legge dei grandi numeri la convergenza va intesa in senso *statistico*; ciò che si può dimostrare è che:

Dato un evento casuale E avente probabilità $\Pr(E)$ di manifestarsi, e che si presenti x volte su N prove eseguite, quindi con frequenza relativa $f(E) = x/N$; in corrispondenza a qualunque coppia di numeri positivi ε' ed ε'' scelti a piacere è possibile determinare in conseguenza un numero intero M tale che, per ogni $N > M$, risulti

$$\Pr\left(\left|\frac{x}{N} - \Pr(E)\right| > \varepsilon'\right) < \varepsilon''$$

In altre parole, aumentando il numero di prove si può rendere improbabile quanto si vuole che la frequenza relativa e la probabilità di un evento differiscano più di una quantità prefissata.

Capitolo 4

Elaborazione dei dati

In questo capitolo si discute dell'organizzazione da dare ai dati sperimentali, e su come si possano da essi ricavare quantità significative.

4.1 Istogrammi

Una volta che si disponga di più misure della stessa grandezza fisica, è opportuno cercare di organizzarle in modo che il loro significato risulti a colpo d'occhio evidente; la maniera più consueta di rappresentare graficamente le misure è quella di disporle in un *istogramma*.

Essendovi corrispondenza biunivoca tra numeri reali e punti di una retta orientata, ognuna delle nostre misure può essere rappresentata da un punto su di essa; l'istogramma è un particolare tipo di diagramma cartesiano in cui l'asse delle ascisse è dedicato a tale rappresentazione. Tuttavia è facile vedere come non tutti i valori della variabile siano permessi, perchè gli strumenti forniscono per loro natura un insieme discreto di valori, essendo limitati ad un numero finito di cifre significative.

Conviene allora indicare sull'asse delle ascisse tutti i possibili valori che possono risultare dalla misura, cioè punti separati da un intervallo che corrisponde alla cifra significativa più bassa dello strumento, o comunque alla più piccola differenza apprezzabile con esso se l'ultima cifra deve essere stimata dall'osservatore (ad esempio il decimo di grado stimato ad occhio su un goniometro avente la sensibilità del mezzo grado).

Nelle ordinate del diagramma si rappresenta la frequenza assoluta con la quale i diversi valori si sono presentati; questo si fa associando ad ognuna delle misu-

re un rettangolo avente area unitaria, che viene riportato con la base al di sopra dell'intervallo appropriato ogni volta che uno dei possibili valori è stato ottenuto. Nel caso consueto in cui l'asse delle ascisse venga diviso in intervalli aventi tutti la stessa ampiezza, tutti questi rettangoli avranno ovviamente la stessa altezza.

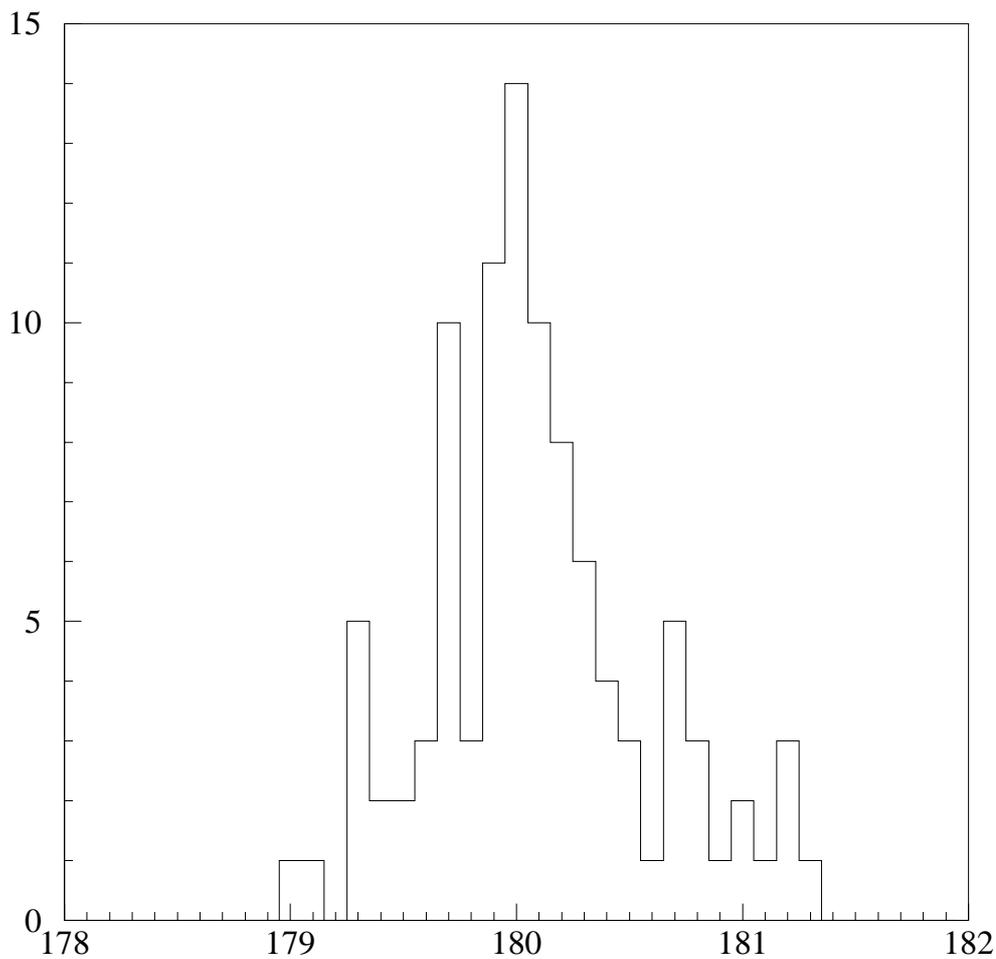


Figura 4.1: Esempio di istogramma (100 misure ripetute della somma degli angoli interni di un triangolo).

Se le frequenze assolute risultassero troppo piccole, può essere opportuno raggruppare le misure in *classi di frequenza*, ciascuna classe corrispondendo ad un intervallo multiplo opportuno del più piccolo rappresentabile discusso sopra.

Anzichè costruire l'istogramma riportandovi un risultato per volta, si posso-

no contare prima le frequenze in ciascuna classe e disegnare sopra ognuno degli intervalli un rettangolo avente altezza corrispondente alla frequenza ivi osservata. L'area dell'istogramma sopra ad ogni intervallo è proporzionale alla frequenza assoluta con cui si è osservato un valore che cade entro di esso: uguale, se si assume come unità di misura per le aree quella del rettangolo unitario; l'area totale sottesa dall'istogramma è rispetto a tale unità pari al numero di osservazioni N .

Un'altra rappresentazione, che è poco usata ma vantaggiosa perchè non richiede la previa (e in qualche misura arbitraria) definizione delle classi di frequenza, è quella della *frequenza cumulativa*, assoluta o relativa. Essa è definita per ogni valore dell'ascissa dal numero (assoluto o relativo) di volte per cui il risultato della misura è stato minore o uguale a x : si tratta dunque di una funzione monotona non decrescente con uno scalino pari rispettivamente ad 1 o a $1/N$ in corrispondenza ad ognuno degli N valori osservati. Risulta inoltre

$$0 = F(-\infty) \leq F(x) \leq F(+\infty) = \begin{cases} N & \text{(ass.)} \\ 1 & \text{(rel.)} \end{cases}$$

4.2 Stime di tendenza centrale

In presenza di N valori osservati di una grandezza fisica (che non siano tutti coincidenti), si pone il problema di definire un algoritmo che fornisca la stima migliore del valore vero della grandezza osservata; cioè di determinare quale, tra le infinite funzioni dei dati, ha la maggiore probabilità di darci il valore vero.

Ora, se supponiamo di avere eliminato tutti gli errori sistematici, è intuitivo come il valore di tale stima debba corrispondere ad una posizione centrale nella distribuzione dei valori osservati; sappiamo infatti che gli errori casuali hanno uguale probabilità di presentarsi in difetto ed in eccesso rispetto al valore vero e, *se il numero di misure è sufficientemente elevato*, ci aspettiamo (sulla base della legge dei grandi numeri) che la distribuzione effettiva delle frequenze non si discosti troppo da quella teorica delle probabilità. Dunque ci si attende che i valori osservati si distribuiscano simmetricamente rispetto al valore vero.

4.2.1 La moda

Nella statistica esistono varie stime della cosiddetta *tendenza centrale* di una distribuzione; una di queste stime è il valore corrispondente al massimo della frequenza, cioè il valore che si presenta il maggior numero di volte (ovvero la media dei valori *contigui* che presentassero tutti la medesima massima frequenza):

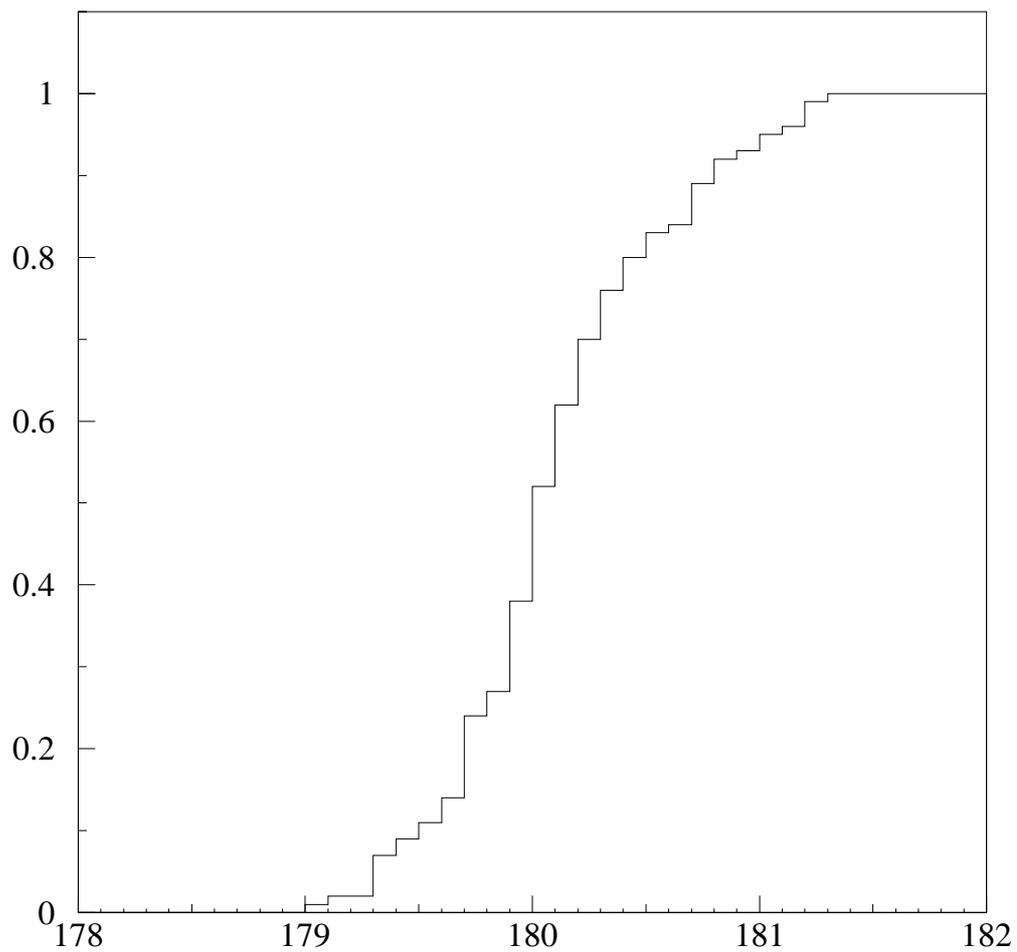


Figura 4.2: Frequenza cumulativa relativa per le stesse misure della figura precedente.

tale stima (se esiste) si chiama *moda* della distribuzione, e si indica con il simbolo \hat{x} .

In generale però la distribuzione potrebbe non avere massimo oppure averne più d'uno in intervalli non contigui (distribuzioni *multimodali*), anche se questo non dovrebbe essere il caso per le distribuzioni di misure ripetute.

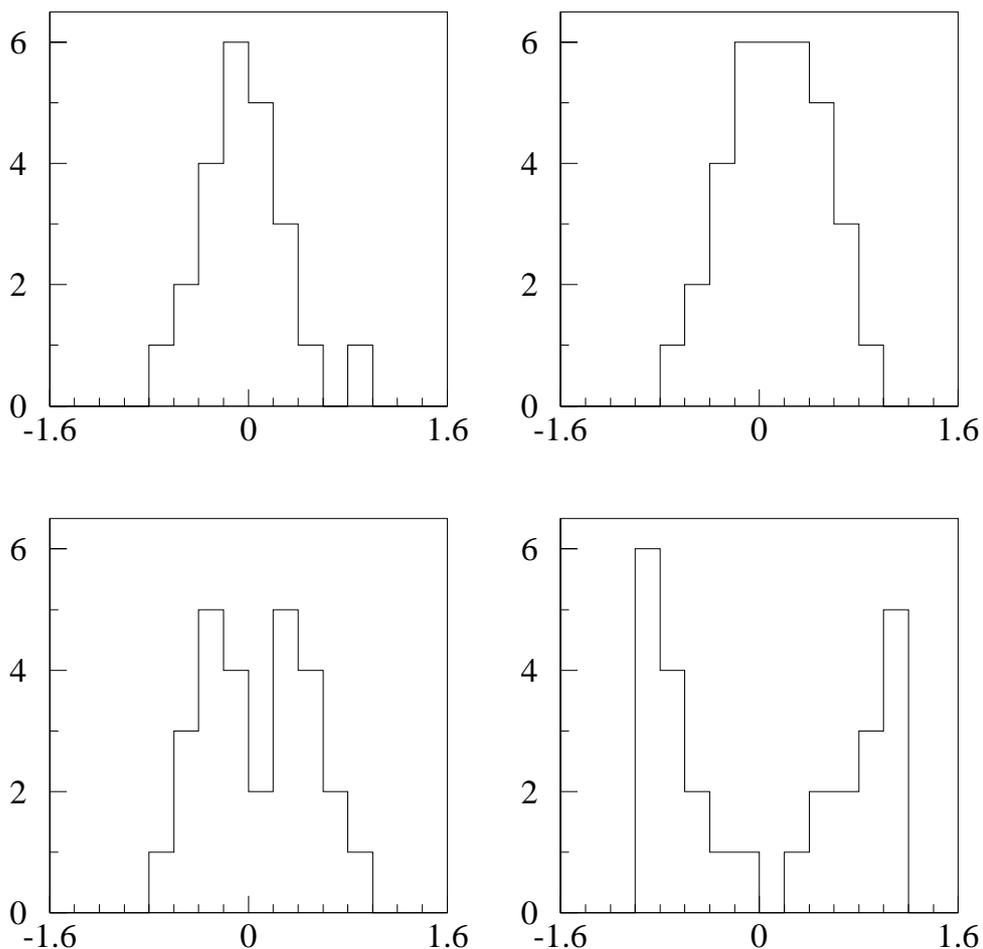


Figura 4.3: Due distribuzioni unimodali (in alto), una bimodale (in basso a sinistra), una senza moda (in basso a destra); quest'ultima distribuzione simula il campionamento a istanti casuali dell'elongazione di un pendolo.

Per questi motivi la moda non è di uso molto frequente, e non è opportuna in questo contesto anche per ragioni che saranno esaminate più avanti.

4.2.2 La mediana

Una stima di uso frequente nella statistica è la *mediana* di una distribuzione, \tilde{x} , definita come quel valore che divide l'istogramma dei dati in due parti di uguale area; in termini meno precisi, la mediana lascia un ugual numero di dati alla propria sinistra ed alla propria destra. Usando questa definizione, per trovare la mediana di un insieme di valori tutti distinti basta disporli in ordine crescente e prendere il valore centrale (per un numero dispari di misure; si prende la semisomma dei due valori centrali se le misure sono in numero pari).

Al contrario della moda, la mediana esiste sempre; nel diagramma della frequenza cumulativa è definita dall'ascissa corrispondente all'ordinata del 50%.

Si può dimostrare anche che la mediana \tilde{x} è quel valore di x che rende minima la somma dei valori assoluti degli scarti delle nostre misure x_i da x ; cioè tale che

$$\min \left(\sum_{i=1}^N |x_i - x| \right) = \sum_{i=1}^N |x_i - \tilde{x}|$$

4.2.3 La media aritmetica

La stima di gran lunga più usata del centro di una distribuzione è la *media aritmetica* dei valori osservati \bar{x} , definita attraverso la

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

Proprietà matematiche della media aritmetica sono le seguenti:

Proprietà 1: *La somma degli scarti di un insieme di valori dalla loro media aritmetica è identicamente nulla.*

Infatti dalla definizione risulta

$$\begin{aligned} \sum_{i=1}^N (x_i - \bar{x}) &= \sum_{i=1}^N x_i - \sum_{i=1}^N \bar{x} \\ &= \sum_{i=1}^N x_i - N\bar{x} \\ &= N\bar{x} - N\bar{x} \\ &\equiv 0 \end{aligned}$$

Proprietà 2: *La media aritmetica \bar{x} di un insieme di dati numerici x_1, x_2, \dots, x_N è quel valore di x rispetto al quale risulta minima la*

somma dei quadrati degli scarti dalle x_i ; cioè quel numero per il quale è verificata la

$$\min \left\{ \sum_{i=1}^N (x_i - x)^2 \right\} = \sum_{i=1}^N (x_i - \bar{x})^2$$

Infatti abbiamo

$$\begin{aligned} \sum_{i=1}^N (x_i - x)^2 &= \sum_{i=1}^N [(x_i - \bar{x}) + (\bar{x} - x)]^2 \\ &= \sum_{i=1}^N [(x_i - \bar{x})^2 + (\bar{x} - x)^2 + 2(x_i - \bar{x})(\bar{x} - x)] \\ &= \sum_{i=1}^N (x_i - \bar{x})^2 + \sum_{i=1}^N (\bar{x} - x)^2 + 2(\bar{x} - x) \sum_{i=1}^N (x_i - \bar{x}) \\ &= \sum_{i=1}^N (x_i - \bar{x})^2 + N(\bar{x} - x)^2 \\ &\geq \sum_{i=1}^N (x_i - \bar{x})^2 \end{aligned}$$

4.2.4 Considerazioni complessive

Oltre le tre stime citate di tendenza centrale ne esistono altre, di uso però limitato a casi particolari (*media geometrica*, *media armonica*) e che non hanno frequente uso né nella statistica né nella fisica, o derivate da altri valori medi (*radice quadratica media*).

Se la distribuzione dei dati non è troppo irregolare, le tre stime citate di tendenza centrale non sono molto lontane; esiste una relazione empirica che le lega e che è valida per distribuzioni non troppo asimmetriche:

$$\bar{x} - \hat{x} \approx 3(\bar{x} - \tilde{x})$$

cioè la differenza tra media aritmetica e moda è circa il triplo della differenza tra media aritmetica e mediana.

La scelta dell'uso dell'una o dell'altra stima statistica per determinare la tendenza centrale della distribuzione di un insieme di misure ripetute andrà decisa sulla base delle proprietà della stima stessa; più precisamente sulla base dello studio di come si distribuiscono varie stime che si riferiscano a *campioni* analoghi, cioè ad insiemi di misure della stessa grandezza fisica, presumibilmente affetti dallo stesso errore (eseguiti insomma in condizioni analoghe) e composti da uno stesso numero di dati.

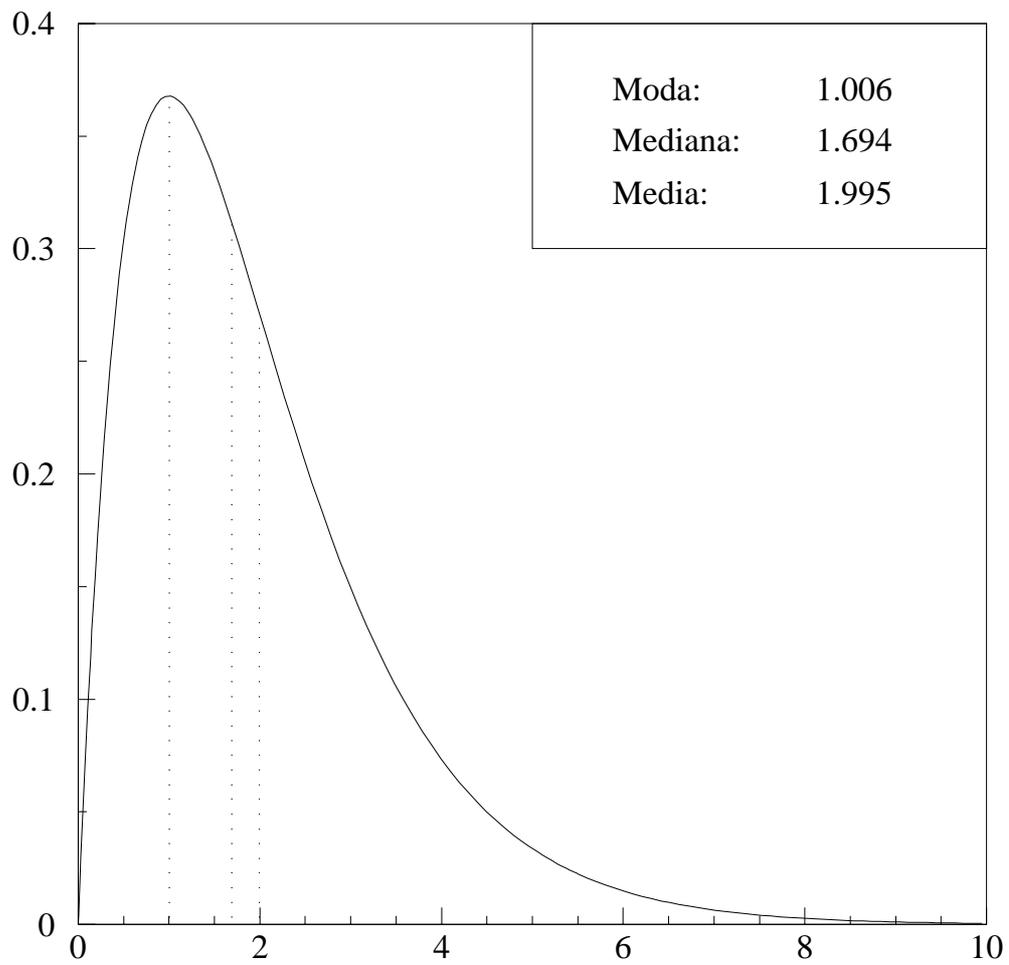


Figura 4.4: Le tre stime di tendenza centrale in una particolare distribuzione.

4.2.5 Prima giustificazione della media

La stima di tendenza centrale che è di uso generale per le misure ripetute è la *media aritmetica*: i motivi sono svariati e sostanzialmente legati alle proprietà statistiche della media stessa; di essi ci occuperemo ancora più avanti.

A questo punto possiamo già comunque verificare (anche se in maniera *non rigorosa*) che la media aritmetica di più misure ha un errore inferiore a quello delle misure stesse; indichiamo con x^* il valore vero della grandezza x , e con x_i ($i = 1, 2, \dots, N$) le N determinazioni sperimentali di x : l'errore assoluto commesso in ognuna delle misure x_i sarà dato da $\varepsilon_i = x_i - x^*$. L'errore assoluto della media aritmetica è allora dato da

$$\begin{aligned}\bar{\varepsilon} &= \bar{x} - x^* = \frac{1}{N} \sum_{i=1}^N x_i - x^* \\ &= \frac{1}{N} \sum_{i=1}^N (x_i - x^*) \\ &= \frac{1}{N} \sum_{i=1}^N \varepsilon_i\end{aligned}$$

Se gli errori sono solo casuali, saranno ugualmente probabili in difetto e in eccesso rispetto al valore vero; e se le misure sono numerose gli ε_i tenderanno ad eliminarsi a vicenda nella sommatoria, che inoltre è moltiplicata per un fattore $1/N$.

4.2.6 La media aritmetica espressa tramite le frequenze

Siano x_i , con $i = 1, \dots, N$, gli N valori del campione di cui vogliamo calcolare la media aritmetica; supponiamo che qualcuno dei valori ottenuti sia *ripetuto*, ad esempio che il valore x_1 si sia presentato n_1 volte, x_2 si sia presentato n_2 volte e così via: la media aritmetica si può calcolare come

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots}{N} \quad (N = n_1 + n_2 + \dots)$$

Indichiamo con x_j ($j = 1, 2, \dots, M$) gli M valori distinti di x presenti nel campione; n_j è la frequenza assoluta con cui abbiamo ottenuto il valore x_j nel corso delle nostre misure, ed il rapporto n_j/N è la frequenza relativa f_j dello stesso evento: allora possiamo scrivere

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

Formule in cui si sommano valori numerici (qui gli x_j) moltiplicati ciascuno per un fattore specifico (f_j) vanno sotto il nome generico di *formule di media pesata*: ogni valore distinto dagli altri contribuisce infatti al risultato finale con un *peso* relativo dato dal numero f_j .

E' bene osservare che si possono definire infinite medie pesate dei valori numerici x_j , corrispondenti alle differenti maniere di attribuire ad ognuno di essi un peso; ed anche che, generalmente, con il nome di "media pesata" ci si riferisce a quella formula che permette di calcolare la migliore stima del valor vero di una grandezza fisica sulla base di più misure aventi differente precisione (e che incontreremo nel paragrafo 8.1), e non alla formula precedente.

Fin qui la formula si presenta solo come un artificio per calcolare la media aritmetica di un insieme di valori risparmiando alcune operazioni; ma pensiamo di far tendere all'infinito il numero di misure effettuate: allora, come sappiamo, la frequenza relativa con cui ogni valore si è presentato tende stocasticamente alla probabilità rispettiva e, in definitiva, la media aritmetica delle misure tende ad un limite determinato:

$$\lim_{N \rightarrow \infty} \bar{x} = \sum_j p_j x_j$$

In definitiva, se siamo in grado di assegnare in qualche modo una probabilità al presentarsi di ognuno dei possibili valori di una misura, siamo anche in grado di calcolare il valore assunto dalla media aritmetica di quei valori nel limite di infinite misure effettuate. Di questa formula ci serviremo più avanti, una volta ricavata appunto (sotto opportune ipotesi) la probabilità di ottenere un certo risultato dalle misure di una grandezza fisica.

4.3 Stime di dispersione

La stima della tendenza centrale di un insieme di misure si intuisce legata al valore vero della grandezza misurata; così all'errore commesso nelle misure si intuisce legata un'altra caratteristica della distribuzione, la *dispersione*: la valutazione cioè della larghezza dell'intervallo in x in cui le misure stesse sono distribuite.

4.3.1 Semidispersione massima e quantili

La più grossolana delle stime statistiche di dispersione si effettua trovando il massimo ed il minimo valore osservato: la *semidispersione massima* è definita come la semidifferenza tra questi due valori,

$$\frac{x_{\max} - x_{\min}}{2}$$

Essa ha il difetto di ignorare la maggior parte dei dati e particolarmente quelli, generalmente preponderanti, prossimi al centro della distribuzione; inoltre normalmente aumenta all'aumentare del numero di misure, invece di tendere ad un valore determinato.

Grandezze frequentemente usate per caratterizzare una distribuzione nella statistica (non nella fisica) sono i quartili, i decili ed i percentili (collettivamente *quantili*), indicati con Q_i ($i = 1, 2, 3$); D_i ($i = 1, \dots, 9$); e P_i ($i = 1, \dots, 99$) rispettivamente. Essi sono definiti (analogamente alla mediana) come quei valori della x che dividono la distribuzione rispettivamente in 4, 10 e 100 parti di uguale area.

Ovviamente vale la

$$Q_2 \equiv D_5 \equiv P_{50} \equiv \bar{x}$$

Come stima della dispersione di una distribuzione è usato dagli statistici l'*intervallo semiinterquartilico* $Q = (Q_3 - Q_1)/2$, come pure la differenza $P_{90} - P_{10}$ tra il novantesimo ed il decimo percentile; tali intervalli esistono sempre, ma non sono padroneggiabili agevolmente negli sviluppi teorici.

4.3.2 Deviazione media assoluta

Altra stima di dispersione è la *deviazione media assoluta*, definita come

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

oppure, meno frequentemente, come

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

ma anch'essa non è facile a trattare a ragione della operazione non lineare costituita dal valore assoluto.

4.3.3 Varianza e deviazione standard

La più importante (e più usata), non solo in fisica, stima di dispersione è la *deviazione standard* (oppure *scarto* o *deviazione quadratica media*), definita come la radice quadrata della *varianza* s^2 :

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

Per distribuzioni non troppo asimmetriche la deviazione media assoluta è circa $\frac{4}{5}$ della deviazione standard, e l'intervallo semiinterquartilico è circa $\frac{2}{3}$ della stessa.

Per calcolare lo scarto quadratico medio di un campione senza l'aiuto di un calcolatore appositamente programmato, può risultare utile sfruttare la sua seguente proprietà:

$$\begin{aligned}
 N s^2 &= \sum_{i=1}^N (x_i - \bar{x})^2 \\
 &= \sum_{i=1}^N (x_i^2 + \bar{x}^2 - 2\bar{x}x_i) \\
 &= \sum_{i=1}^N x_i^2 + N\bar{x}^2 - 2\bar{x} \sum_{i=1}^N x_i \\
 &= \sum_{i=1}^N x_i^2 + N\bar{x}^2 - 2N\bar{x}^2 \\
 &= \sum_{i=1}^N x_i^2 - N\bar{x}^2
 \end{aligned}$$

da cui la formula

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

che permette un calcolo più agevole di s^2 accumulando successivamente i quadrati dei valori osservati anzichè quelli dei loro scarti dalla media.

4.4 Giustificazione della media

Stabiliamo quindi per convenzione che il nostro metodo per misurare la dispersione di un campione di dati è quello del calcolo della deviazione standard; accettiamo anche che *in qualche modo* questo numero sia legato all'errore presumibilmente commesso nel corso delle misure. Una definizione più precisa di ciò che si intende con le parole "errore commesso", ovvero sia l'interpretazione probabilistica dello scarto quadratico medio nei riguardi delle misure ripetute, verrà data più avanti.

Comunque, una volta assunto questo, possiamo approfondire il discorso già cominciato sulla media aritmetica come stima del centro della distribuzione e quindi del valor vero di una grandezza, nel caso di misure ripetute ed in assenza di errori sistematici. E' infatti possibile provare¹ che la media aritmetica è la

¹La dimostrazione risale a Gauss se ci si limita alle sole operazioni lineari sui dati, e solo ad anni recenti per un qualsiasi algoritmo operante su di essi; vedi in proposito l'appendice G.

stima del valore vero affetta *dal minimo errore* casuale, cioè avente la più piccola deviazione standard.

Riferendosi a quanto prima accennato, ciò significa che le medie aritmetiche di molti campioni analoghi di N misure avranno un istogramma più stretto delle mode, delle mediane e di qualsiasi altra misura di tendenza centrale desumibile dagli stessi campioni; la larghezza di tale istogramma (misurata, come abbiamo assunto, dal suo scarto quadratico medio) sarà messa in relazione con lo scarto quadratico medio delle misure da un teorema di cui ci occuperemo nel seguito. Da esso discenderà anche che l'errore statistico della media aritmetica converge a zero al crescere indefinito del numero di dati N .

Per concludere:

1. *Disponendo di più misure ripetute della stessa grandezza fisica, si assume come migliore stima del valore vero della stessa la loro media aritmetica.*
2. *Questa stima è più precisa di quanto non lo siano le singole misure, ed è tanto più attendibile quanto maggiore è il numero delle stesse.*
3. *Come valutazione dell'errore commesso nelle singole misure si assume il loro scarto quadratico medio; o meglio, per motivi che verranno chiariti in seguito, la quantità²*

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

²La differenza tra questa formula e quella prima citata non è praticamente avvertibile se N non è troppo piccolo.

Capitolo 5

Variabili casuali unidimensionali

Già sappiamo che, a causa degli errori, la misura di una grandezza fisica può essere considerata un evento casuale, e che il numero reale da noi ottenuto in conseguenza della misura stessa può essere considerato una variabile casuale definita sull'intero insieme dei possibili risultati.

Un insieme finito di operazioni di misura, i cui risultati costituiscono quello che in linguaggio statistico si dice *campione*, si può pensare come un particolare sottoinsieme formato da membri estratti a caso dall'insieme di tutte le infinite possibili operazioni di misura della stessa grandezza fisica, eseguite col medesimo strumento e le medesime procedure.

Quest'ultimo insieme nella terminologia statistica si dice *universo* o *popolazione*, ed è in effetti una finzione (si pensi all'universo di tutti i possibili lanci di un dado nella teoria dei giochi d'azzardo), nel senso che in realtà esso non è preesistente alle operazioni effettivamente eseguite, a differenza dell'insieme di tutti gli individui di una vera popolazione dalla quale si estrae realmente un campione in una ricerca demografica. Sebbene sia una finzione, questo concetto è tuttavia utile per applicare il calcolo delle probabilità alle caratteristiche di un campione.

In questo capitolo esamineremo il comportamento delle variabili casuali in generale, ed in particolare quello dei risultati delle misure; il tutto con lo scopo di mettere in evidenza i rapporti tra grandezze che si riferiscano ad un campione limitato e grandezze analoghe che siano invece riferite all'intera popolazione.

5.1 Generalità

Riprendiamo ora il concetto di variabile casuale già introdotto in precedenza nel paragrafo 3.1, e consideriamo alcuni esempi: se si associa ad ogni faccia di un dado un numero compreso tra 1 e 6 (il punteggio inciso sulla faccia stessa), si definisce una variabile casuale discreta; se l'evento casuale consiste nel lancio di due monete, indicando con E l'apparizione della testa nel lancio della prima e con F l'apparizione della testa nel lancio della seconda, il numero di teste osservate nell'evento è ancora una variabile casuale discreta, la cui definizione è data dalla tabella seguente:

Evento	x
EF	2
$E\bar{F}$	1
$\bar{E}F$	1
$\bar{E}\bar{F}$	0

Come si può notare, la corrispondenza tra variabile casuale ed insieme dei possibili eventi non è in questo caso biunivoca.

Se l'insieme di definizione S è continuo, la variabile casuale $x(E)$ può essere continua; è questo il caso più frequente nella fisica, ad esempio per le misure: ma anche in tal caso, a causa della sensibilità limitata degli strumenti, l'intervallo continuo di definizione della variabile x viene suddiviso in un numero finito M di intervalli, che vengono rappresentati dai valori centrali x_j della variabile casuale.

Detta v_j la frequenza assoluta con cui si è presentato il risultato x_j nelle N prove complessive, sarà

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

(potendo alcune frequenze v_j risultare nulle perchè i corrispondenti valori x_j non sono stati osservati nelle prove). Indicata con

$$f_j = \frac{v_j}{N}$$

la frequenza relativa del valore x_j nelle N prove, dalla prima relazione segue

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

esaurendo gli M valori x_j tutti i possibili risultati della misura.

Se il numero delle prove N è molto grande e viene fatto crescere a piacere, ciascuna f_j tende statisticamente al valore p_j (probabilità di osservare il valore x_j), e sarà ancora ovviamente

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

5.2 Valor medio e valor vero

Come sappiamo, il valore medio della variabile x sull'intera popolazione è dato dalla formula che abbiamo incontrato nel paragrafo 4.2.6; indicheremo nel testo questo valor medio (così come altri anch'essi riferiti all'intera popolazione) col simbolo $E(x)$.

Si può ora meglio precisare la distinzione fra errori casuali ed errori sistematici: i primi, visto che possono verificarsi con uguale probabilità in difetto ed in eccesso rispetto al valor vero, avranno valor medio nullo; mentre errori sistematici causeranno invece per definizione una differenza tra il valor medio della misura $E(x)$ ed il valor vero.

In assenza di errori sistematici assumiamo allora che valor medio e valor vero coincidano: ammettiamo insomma (lo proveremo più avanti per la distribuzione normale) che in tal caso $E(x)$ sia uguale a x^* ; potremo in definitiva scrivere

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{N} \sum_{j=1}^M v_j x_j = \sum_{j=1}^M f_j x_j$$

$$\lim_{N \rightarrow \infty} \bar{x} = \sum_{j=1}^M p_j x_j \equiv E(x)$$

$$E(x) \equiv x^*$$

Consideriamo due variabili casuali x e y , ed una loro qualsiasi combinazione lineare $z = ax + by$: vogliamo dimostrare ora che il valor medio della nuova variabile z è dato dalla combinazione lineare dei rispettivi valori medi di x e di y con gli stessi coefficienti a e b ; la dimostrazione si potrà poi facilmente estendere, per induzione completa, alla combinazione lineare di un numero qualsiasi di variabili.

Indichiamo con x_j i possibili valori per la prima variabile, e con y_k quelli per la seconda; indichiamo poi con p_j e q_k le probabilità di ottenere un determinato valore rispettivamente per la x e per la y .

Chiamiamo poi P_{jk} la probabilità che simultaneamente si abbia $x = x_j$ ed $y = y_k$; un particolare valore per la x potrà essere associato ad uno qualsiasi dei diversi

valori della y , che sono tra loro mutuamente esclusivi: in definitiva, per la legge della probabilità totale, risulterà

$$p_j = \sum_k P_{jk} \quad \text{e} \quad q_k = \sum_j P_{jk}$$

Per il valor medio di z avremo poi

$$\begin{aligned} E(ax + by) &= \sum_{jk} P_{jk} (ax_j + by_k) \\ &= \sum_{jk} aP_{jk}x_j + \sum_{jk} bP_{jk}y_k \\ &= a \sum_j \left(\sum_k P_{jk} \right) x_j + b \sum_k \left(\sum_j P_{jk} \right) y_k \\ &= a \sum_j p_j x_j + b \sum_k q_k y_k \\ &= aE(x) + bE(y) \end{aligned}$$

Una importante conseguenza può subito essere da qui ricavata per la media aritmetica \bar{x} di un campione di N misure: essa infatti si può considerare come una particolare combinazione lineare delle misure stesse, con coefficienti tutti uguali tra loro e pari ad $1/N$.

Prendendo dalla popolazione un differente campione di N misure, la loro media aritmetica \bar{x} sarà anch'essa in generale diversa: quale sarà il valor medio delle varie \bar{x} su un numero molto elevato di campioni di N misure estratti a caso dalla popolazione, al limite su tutti i campioni aventi la stessa dimensione fissa N che dalla popolazione è possibile ricavare?

$$\begin{aligned} E(\bar{x}) &= E\left(\frac{1}{N} \sum_{i=1}^N x_i\right) \\ &= \frac{1}{N} \sum_{i=1}^N E(x_i) \\ &= \frac{1}{N} \cdot NE(x) \\ &= x^* \end{aligned}$$

Cioè

Il valor medio delle medie aritmetiche dei campioni di dimensione finita N estratti da una popolazione, coincide con il valor medio dell'intera popolazione.

5.3 Scarto ed errore quadratico medio

La varianza s^2 di un campione di N misure si può esprimere come

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

Osserviamo che (per qualsiasi numero x^* e quindi anche per l'incognito valor vero) vale la seguente relazione matematica:

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N (x_i - x^*)^2 &= \frac{1}{N} \sum_{i=1}^N [(x_i - \bar{x}) + (\bar{x} - x^*)]^2 \\ &= \frac{1}{N} \left[\sum_{i=1}^N (x_i - \bar{x})^2 + \sum_{i=1}^N (\bar{x} - x^*)^2 + 2(\bar{x} - x^*) \sum_{i=1}^N (x_i - \bar{x}) \right] \\ &= \frac{1}{N} \left[\sum_{i=1}^N (x_i - \bar{x})^2 + N(\bar{x} - x^*)^2 \right] \\ &= s^2 + (\bar{x} - x^*)^2 \end{aligned}$$

(si è sfruttata qui la proprietà della somma algebrica degli scarti delle misure dalla loro media aritmetica di essere identicamente nulla; vedi anche la formula analoga già ricavata a pagina 29 nel paragrafo 4.2.3).

Cerchiamo ora di capire come le varianze s^2 dei campioni di dimensione N siano legate all'analogia grandezza σ^2 definita sull'intera popolazione, e calcolata rispetto al valor medio di essa, $E(x) = x^*$:

$$\sigma^2 = E \left\{ (x - x^*)^2 \right\} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (x_i - x^*)^2$$

Sfruttando la relazione trovata in precedenza si ha

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

e prendendo i valori medi di entrambi i membri (sugli infiniti campioni di dimensione N che si possono pensare estratti in modo casuale dalla popolazione originaria), otteniamo

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

Ricordando come il valor medio del quadrato degli scarti di una variabile (qui \bar{x}) dal suo valore medio (che è $E(\bar{x}) = x^*$) sia per definizione la varianza della

variabile stessa (indicata qui dal quadrato dell'errore quadratico medio $\sigma_{\bar{x}}$), si ricava infine:

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

Insomma:

- Il valor medio della varianza s^2 di un campione è sistematicamente inferiore all'analogha grandezza σ^2 che si riferisce all'intera popolazione.
- La differenza tra la varianza della popolazione σ^2 e la varianza di un campione di N misure da essa estratto è **in media** pari alla varianza della media del campione.

5.4 La varianza delle combinazioni lineari

Dimostriamo un teorema generale che riguarda la varianza di una combinazione lineare di più variabili casuali, che supporremo però *statisticamente indipendenti*. Dette x ed y due variabili che godano di tale proprietà, sappiamo che la probabilità P_{jk} che contemporaneamente sia $x = x_j$ e $y = y_k$ è data dal prodotto delle probabilità rispettive p_j e q_k .

Per semplificare i calcoli dimostriamo questo teorema dapprima nel caso particolare di due popolazioni x e y che abbiano medie, cioè valori veri, *nulli*; estenderemo poi il risultato a due variabili sempre statisticamente indipendenti ma aventi un qualunque valor medio. Ciò premesso, la combinazione lineare

$$z = ax + by$$

ha anch'essa media zero: infatti è

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

Risulta allora

$$\begin{aligned} \sigma^2(z) &= E \left\{ [z - E(z)]^2 \right\} \\ &= E(z^2) \\ &= E \left\{ (ax + by)^2 \right\} \\ &= \sum_{jk} P_{jk} (ax_j + by_k)^2 \\ &= \sum_{jk} p_j q_k (a^2 x_j^2 + b^2 y_k^2 + 2ab x_j y_k) \\ &= a^2 \sum_k q_k \sum_j p_j x_j^2 + b^2 \sum_j p_j \sum_k q_k y_k^2 + 2ab \sum_j p_j x_j \sum_k q_k y_k \\ &= a^2 \sigma_x^2 \sum_k q_k + b^2 \sigma_y^2 \sum_j p_j + 2ab E(x) E(y) \\ &= a^2 \sigma_x^2 + b^2 \sigma_y^2 \end{aligned}$$

Allo scopo di estendere la validità di quanto appena dimostrato a due variabili casuali x e y aventi medie anche differenti da zero, dimostriamo ora il seguente

Teorema: *Due variabili casuali che differiscano per un fattore costante hanno la stessa varianza.*

Infatti, se le due variabili casuali x e ξ soddisfano questa ipotesi, allora deve risultare:

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

Ora, date due variabili casuali x e y qualsiasi, ed una loro generica combinazione lineare $z = ax + by$, basta definire altre due variabili casuali ausiliarie

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

(che ovviamente soddisfano l'ipotesi di avere media zero): pertanto la loro combinazione lineare $\zeta = a\xi + b\eta$ (che differisce anch'essa da z per un fattore costante) sarà tale che

$$\sigma_{\zeta}^2 = a^2\sigma_{\xi}^2 + b^2\sigma_{\eta}^2$$

Ma per quanto detto, x e ξ hanno la stessa varianza; così y ed η , e z e ζ . Ne consegue come per *qualsiasi* coppia di variabili casuali (purchè statisticamente indipendenti) valga la seguente regola:

Una combinazione lineare, a coefficienti costanti, di variabili casuali statisticamente indipendenti ha varianza uguale alla combinazione lineare delle rispettive varianze, con coefficienti pari ai quadrati dei coefficienti rispettivi¹.

E' ovvio poi estendere (per induzione completa) questo risultato alla combinazione lineare di un numero finito qualsivoglia di variabili casuali tra loro indipendenti: se

$$F = ax + by + cz + \dots$$

allora

$$\sigma_F^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + c^2\sigma_z^2 + \dots$$

¹Una formula più generale che si può applicare a variabili casuali qualunque verrà dimostrata nell'appendice B.

5.5 L'errore della media

Torniamo ora ad occuparci dello studio delle proprietà statistiche della media aritmetica di un campione di N misure

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

e cerchiamo in particolare di determinarne la varianza. Applicando il teorema appena dimostrato, risulta

$$\begin{aligned} \sigma_{\bar{x}}^2 &= \frac{1}{N^2} \sum_{i=1}^N \sigma_{x_i}^2 \\ &= \frac{1}{N^2} \cdot N \sigma_x^2 \\ &= \frac{\sigma_x^2}{N} \end{aligned}$$

In definitiva abbiamo dimostrato che

- *Le medie aritmetiche di campioni di N misure hanno varianza pari alla varianza della popolazione da cui le misure provengono, divisa per la dimensione dei campioni.*
- *L'errore quadratico medio della media di un campione è minore dell'analogo errore delle singole misure, e tende a zero al crescere del numero di misure effettuato.*

5.6 Stima dell'errore quadratico medio

Vediamo ora come si può stimare la varianza dell'insieme delle infinite misure che possono essere fatte di una grandezza fisica a partire da un particolare campione di N misure. Abbiamo già dimostrato che

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

e sappiamo ora che la varianza della media del campione è data dalla

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

Risulta pertanto

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

e quindi

Mediamente la varianza di un campione di N misure è inferiore alla varianza della intera popolazione per un fattore $(N-1)/N$.

Questo è il motivo per cui, per avere una stima **mediamente** corretta di σ , si usa (come già anticipato) la quantità μ definita attraverso la

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

quantità la cui media su infiniti campioni risulta proprio σ^2 .

5.7 Ancora sull'errore quadratico medio

Diamo qui un'altra dimostrazione del teorema riguardante la stima corretta dell'errore quadratico medio di una popolazione a partire da un campione, seguendo una linea diversa e più vicina alle verifiche sperimentali che si possono fare avendo a disposizione numerosi dati.

Si supponga di avere M campioni contrassegnati dall'indice j (con j che assume i valori $1, \dots, M$); ciascuno di essi sia poi costituito da N misure ripetute della stessa grandezza x , contrassegnate a loro volta dall'indice i ($i = 1, \dots, N$): il valore osservato nella misura i -esima del campione j -esimo sia indicato insomma dal simbolo x_{ij} .

Indicando con x^* il valore vero di x , e con \bar{x}_j la media aritmetica del campione j -esimo, vale la

$$\begin{aligned} (x_{ij} - x^*)^2 &= [(x_{ij} - \bar{x}_j) + (\bar{x}_j - x^*)]^2 \\ &= (x_{ij} - \bar{x}_j)^2 + (\bar{x}_j - x^*)^2 + 2(\bar{x}_j - x^*)(x_{ij} - \bar{x}_j) \end{aligned}$$

Ora sommiamo su i tutte le N uguaglianze che si hanno per i valori dell'indice $i = 1, 2, \dots, N$ e dividiamo per N ; se indichiamo con s_j^2 la varianza del campione j -esimo, data da

$$s_j^2 = \frac{1}{N} \sum_{i=1}^N (x_{ij} - \bar{x}_j)^2$$

otteniamo alla fine

$$\frac{1}{N} \sum_{i=1}^N (x_{ij} - x^*)^2 = s_j^2 + (\bar{x}_j - x^*)^2 + \frac{2}{N} (\bar{x}_j - x^*) \sum_{i=1}^N (x_{ij} - \bar{x}_j)$$

L'ultima sommatoria a destra è la somma algebrica degli scarti delle misure del campione j -esimo dalla loro media aritmetica \bar{x}_j che sappiamo essere identicamente nulla. Dunque, per ogni j vale la

$$\frac{1}{N} \sum_{i=1}^N (x_{ij} - x^*)^2 = s_j^2 + (\bar{x}_j - x^*)^2$$

e se sommiamo membro a membro tutte le M uguaglianze che abbiamo per $j = 1, 2, \dots, M$ e dividiamo per M , risulta

$$\frac{1}{M} \sum_{j=1}^M \frac{1}{N} \sum_{i=1}^N (x_{ij} - x^*)^2 = \frac{1}{M} \sum_{j=1}^M s_j^2 + \frac{1}{M} \sum_{j=1}^M (\bar{x}_j - x^*)^2$$

Ora supponiamo di avere a disposizione moltissimi campioni e passiamo al limite per $M \rightarrow \infty$. Il primo membro (che rappresenta il valor medio, su tutti i dati e tutti gli infiniti campioni, del quadrato degli scarti dal valore *vero*) converge stocasticamente alla varianza della variabile casuale x ; il secondo termine a destra (valor medio, su tutti gli infiniti campioni, del quadrato degli scarti della media aritmetica del campione dal proprio valor vero) converge alla varianza delle medie dei campioni di N misure $\sigma_{\bar{x}}^2$.

Il primo termine a destra è il valor medio della varianza dei campioni di N misure e, sostituendo, infine si trova

$$\begin{aligned} \sigma^2 &= \lim_{M \rightarrow \infty} \frac{1}{NM} \sum_{ij} (x_{ij} - x^*)^2 \\ &= \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{j=1}^M s_j^2 + \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{j=1}^M (\bar{x}_j - x^*)^2 \\ &= E(s^2) + \sigma_{\bar{x}}^2 \end{aligned}$$

Ora, avendo già dimostrato che

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

si ricava facilmente

$$\sigma^2 = E(s^2) + \frac{\sigma^2}{N}$$

ovvero

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

che è il risultato già ottenuto.

Si noti che mentre molti teoremi della statistica sono validi solo *asintoticamente*, cioè per campioni numerosi o per un numero molto grande di variabili, questo teorema vale per ogni $N (\geq 2)$.

Capitolo 6

La legge di Gauss

Vogliamo ora investigare sulla distribuzione dei risultati di misure ripetute, nell'ipotesi che esse siano affette da errori esclusivamente casuali.

Le definizioni di probabilità che conosciamo sono adatte solo per una variabile che possa assumere un numero finito di valori discreti; vediamo innanzi tutto come il concetto di probabilità si possa generalizzare a variabili continue, variabili che possono cioè assumere tutti gli infiniti valori appartenenti ad un insieme continuo: tali si ammette generalmente siano i risultati delle misure delle grandezze fisiche, per poter applicare ad essi il calcolo differenziale ed integrale.

6.1 La densità di probabilità

Definiamo arbitrariamente delle classi di frequenza, suddividendo l'asse delle x in intervalli di ampiezze uguali qualunque, ed immaginiamo di fare un certo numero N di misure della grandezza fisica x . Come sappiamo, possiamo riportare le misure ottenute in istogramma tracciando, al di sopra dell'intervallo che rappresenta ogni classe, un rettangolo avente area uguale alla frequenza relativa con cui una misura cade in essa; l'altezza dell'istogramma in ogni intervallo è data quindi da tale frequenza divisa per l'ampiezza dell'intervallo di base, e l'area totale dell'istogramma stesso vale 1.

Se immaginiamo di far tendere all'infinito il numero di misure effettuate, in base alla legge dei grandi numeri ci aspettiamo un aggiustamento dell'istogramma in modo che l'area rappresentata sopra ogni intervallo tenda alla *probabilità* che il valore misurato cada entro di esso; le altezze tenderanno quindi al rapporto tra questa probabilità e l'ampiezza dell'intervallo di base dell'istogramma.

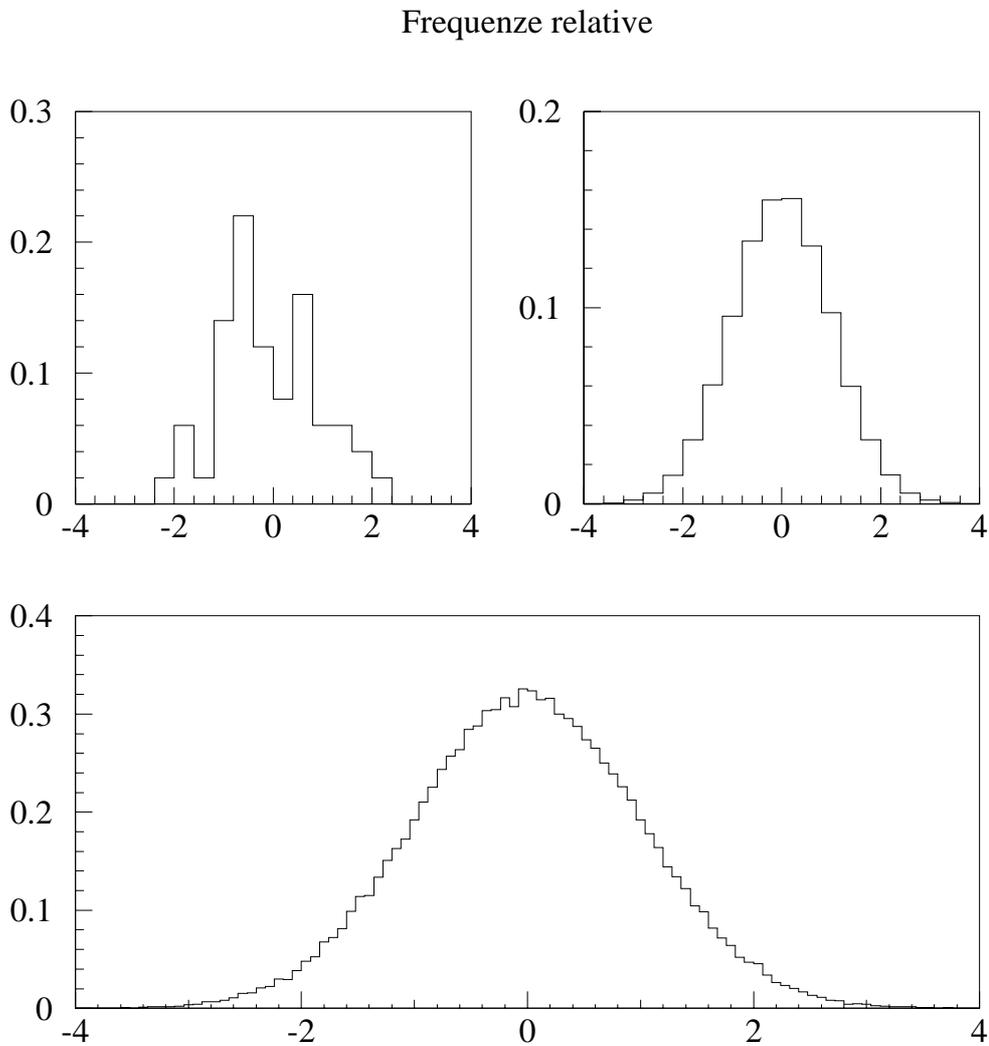


Figura 6.1: Nella prima figura, l'istogramma della grandezza x per un numero piccolo di misure; nella seconda, lo stesso istogramma per un numero molto grande di misure; nell'ultima, l'istogramma si approssima alla curva limite quando l'intervallo di base tende a zero.

Disponendo di un numero infinitamente grande di misure, ha senso diminuire l'ampiezza degli intervalli in cui l'asse delle x è stato diviso, e renderla piccola a piacere. Se l'intervallo corrispondente ad una data classe di frequenza tende a zero, la probabilità che una misura cada in esso tende ugualmente a zero; ma se esiste ed è finito il limite del rapporto tra probabilità ed ampiezza dell'intervallo, l'istogramma tenderà ad una curva continua la cui ordinata sarà in ogni punto data da tale limite.

L'ordinata di questa curva al di sopra di un intervallo infinitesimo dx vale quindi

$$y = f(x) = \frac{dp}{dx}$$

Le dimensioni della grandezza y sono quelle di una probabilità (un numero puro) divise per quelle della grandezza x ; la y prende il nome di *densità di probabilità*, o di *funzione di frequenza*, della x .

La variabile continua schematizza il caso in cui i valori osservabili (discreti per la sensibilità limitata degli strumenti) sono molto densi, separati cioè da intervalli molto piccoli, e assai numerosi. In questa situazione la probabilità di osservare uno solo di tali valori è anch'essa molto piccola, ed ha interesse soltanto la probabilità di osservare uno tra i numerosi valori della x che cadano in un dato intervallo $[x_1, x_2]$ di ampiezza grande rispetto alla risoluzione sperimentale.

Se dividiamo tale intervallo in un numero molto grande di sottointervalli infinitesimi dx , gli eventi casuali consistenti nell'appartenere il risultato della misura ad una delle classi di frequenza relative sono mutuamente esclusivi; di conseguenza, la probabilità che x appartenga all'intervallo finito $[x_1, x_2]$ è data dalla somma delle probabilità (infinitesime) rispettive $dp = f(x) dx$: e questa, per definizione, è l'integrale di $f(x) dx$ nell'intervallo $[x_1, x_2]$.

Insomma, qualunque sia l'intervallo $[x_1, x_2]$ vale la

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

In definitiva:

Per le variabili continue non si può parlare di probabilità attraverso le definizioni già esaminate. E' invece possibile associare ad ogni variabile continua x una funzione "densità di probabilità" $f(x)$, da cui si può dedurre la probabilità che la x cada in un qualsiasi intervallo prefissato: questa è data semplicemente dall'area sottesa dalla curva nell'intervallo in questione.

Analogamente al concetto sperimentale di frequenza cumulativa relativa, introdotto a pagina 25 nel paragrafo 4.1, si può definire la *funzione di distribuzione*

per una variabile continua x come

$$F(x) = \int_{-\infty}^x f(y) dy$$

Essa rappresenta la probabilità di osservare un valore non superiore ad x , e dovrà necessariamente soddisfare la $F(+\infty) \equiv 1$. Quindi:

L'integrale di una qualunque funzione che rappresenti una densità di probabilità nell'intervallo $[-\infty, +\infty]$ vale 1:

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

6.2 La funzione di Gauss

Dall'esame di molte distribuzioni sperimentali di valori ottenuti in misure ripetute in condizioni omogenee, si possono astrarre due proprietà generali degli errori casuali:

- La probabilità di ottenere un certo scarto dal valor vero deve essere funzione del modulo di tale scarto e non del suo segno, se valori in difetto ed in eccesso rispetto a quello vero si presentano con uguale probabilità; in definitiva la distribuzione degli scarti deve essere *simmetrica rispetto allo zero*.
- La probabilità di ottenere un certo scarto dal valor vero (in modulo) deve essere *decrescente* al crescere di tale scarto e *tendere a zero* quando esso tende all'infinito; questo perchè deve essere più probabile commettere errori piccoli che errori grandi, ed infinitamente improbabile commettere errori infinitamente grandi.

A queste due ipotesi sulla distribuzione delle misure affette da errori puramente casuali se ne può aggiungere una terza, valida per *tutte* le distribuzioni di probabilità; la cosiddetta *condizione di normalizzazione*, di cui abbiamo già parlato prima:

- L'area compresa tra la curva e l'asse delle ascisse da $-\infty$ a $+\infty$ deve valere 1.

Da queste tre ipotesi e dal principio della media aritmetica, Gauss dimostrò in modo euristico che la distribuzione degli scarti z delle misure affette da errori casuali è data dalla funzione

$$f(z) = \frac{h}{\sqrt{\pi}} e^{-h^2 z^2}$$

che da lui prese il nome (*funzione di Gauss* o *legge normale* di distribuzione degli errori).

Si deve originalmente a Laplace una prova più rigorosa ed indipendente dall'assunto della media aritmetica; una versione semplificata di questa dimostrazione è data nell'appendice E.

La funzione di Gauss ha una caratteristica forma a campana, simmetrica rispetto a $z = 0$ e decrescente tendendo a 0 per z che tende a $\pm\infty$, così come richiesto dalle ipotesi di partenza; essa dipende da un parametro h che prende il nome di *modulo di precisione* della misura: infatti quando h è piccolo la funzione è sensibilmente diversa da zero in una zona estesa dell'asse delle ascisse; mentre al crescere di h tale intervallo diminuisce e la curva si stringe sull'asse delle ordinate.

6.3 Proprietà della legge normale

Possiamo ora, come già anticipato nel paragrafo 4.2.6 dedicato alla media pesata, determinare il valor medio di una qualsiasi grandezza $w(z)$ legata alle misure, nel limite di un numero infinito di misure effettuate; questo valor medio, che avevamo scritto

$$\sum_i p_i w_i$$

per una variabile discreta, si deve ora scrivere per una variabile continua

$$\int_{-\infty}^{+\infty} f(z) w(z) dz$$

dove per $f(z)$ si intende la funzione di Gauss. Per ricavare questa formula basta pensare suddiviso l'asse delle z in un numero grandissimo di intervalli infinitesimi dz , ad ognuno dei quali è associata una probabilità che vale $f(z)dz$, e sostituire nella formula per variabili discrete.

Se vogliamo trovare il valor medio dello scarto z , questo è dato dalla

$$E(z) = \frac{h}{\sqrt{\pi}} \int_{-\infty}^{+\infty} z e^{-h^2 z^2} dz = 0$$

Il risultato è immediato considerando che $f(z)$ è una funzione simmetrica mentre z è antisimmetrica: in definitiva, ad ogni intervallino centrato su un dato valore $z > 0$ possiamo associarne uno uguale centrato sul punto $-z$, in cui il

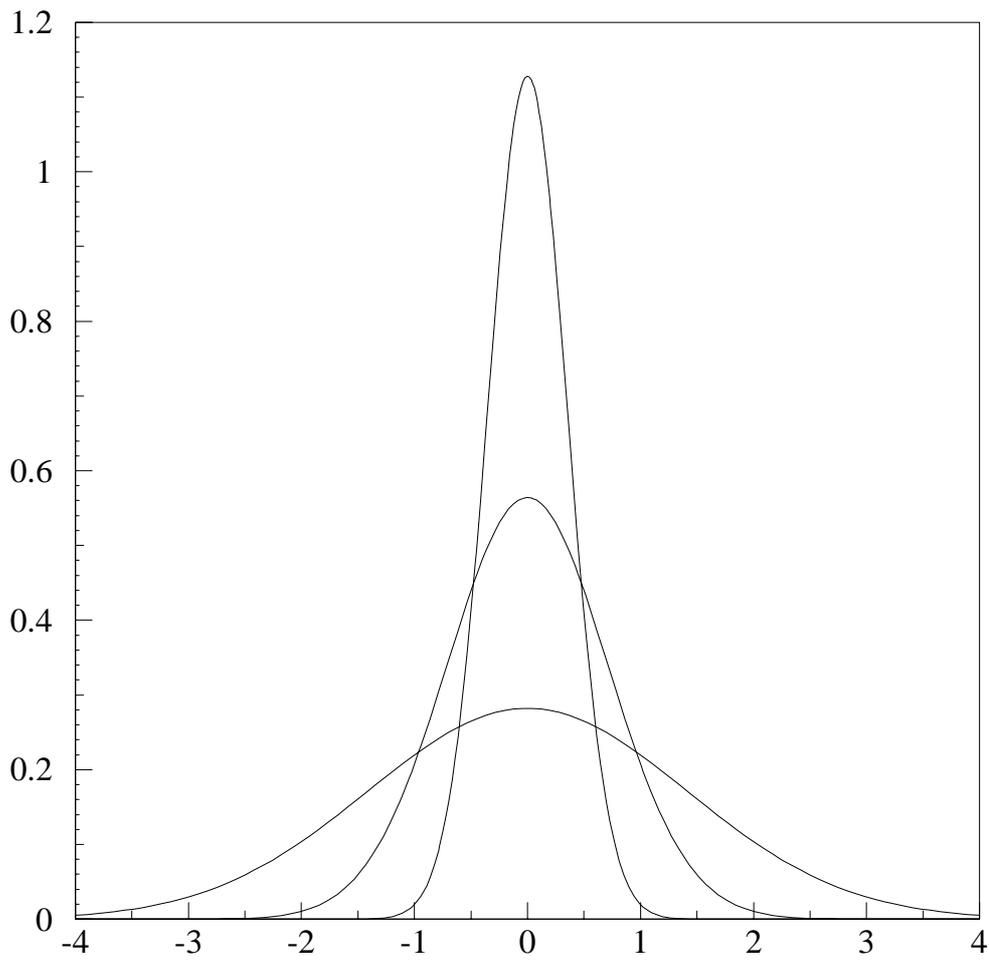


Figura 6.2: La funzione di Gauss per tre diversi valori di h .

prodotto $z f(z) dz$ assume valori uguali in modulo ma di segno opposto, così che la loro somma è zero. Essendo poi

$$E(z) = E(x - x^*) = E(x) - x^*$$

abbiamo così *dimostrato* che

Il valor medio della popolazione delle misure di una grandezza fisica affette solo da errori casuali coincide con il valor vero della grandezza misurata.

Cerchiamo ora il valor medio del modulo dello scarto $E(|z|)$:

$$\begin{aligned} E(|z|) &= \int_{-\infty}^{+\infty} |z| f(z) dz \\ &= 2 \int_0^{+\infty} z f(z) dz \\ &= \frac{2h}{\sqrt{\pi}} \int_0^{+\infty} z e^{-h^2 z^2} dz \\ &= \frac{h}{\sqrt{\pi}} \int_0^{+\infty} e^{-h^2 t} dt \\ &= \frac{1}{h\sqrt{\pi}} \end{aligned}$$

dove si è eseguito il cambio di variabile $t = z^2$.

Il valor medio del modulo degli scarti è quella grandezza che abbiamo chiamato “errore medio”: qui abbiamo ricavato la relazione tra l’errore medio di misure affette solo da errori casuali ed il modulo di precisione della misura h .

Il rapporto invece tra l’errore quadratico medio ed h si trova calcolando il valor medio del quadrato degli scarti:

$$\begin{aligned} E(z^2) &= \frac{h}{\sqrt{\pi}} \int_{-\infty}^{+\infty} z^2 e^{-h^2 z^2} dz \\ &= \frac{1}{h^2 \sqrt{\pi}} \int_{-\infty}^{+\infty} t^2 e^{-t^2} dt \\ &= -\frac{1}{2h^2 \sqrt{\pi}} \int_{-\infty}^{+\infty} t d(e^{-t^2}) \\ &= -\frac{1}{2h^2 \sqrt{\pi}} \left\{ [te^{-t^2}]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} e^{-t^2} dt \right\} \\ &= \frac{1}{2h^2} \end{aligned}$$

Per giungere al risultato, per prima cosa si è effettuata la sostituzione di variabile $t = hz$; poi si è integrato per parti; infine si è tenuto conto del fatto che

$$\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}$$

come si può ricavare dalla condizione di normalizzazione della funzione di Gauss per il particolare valore $h = 1$.

Concludendo:

- Per misure affette da errori distribuiti secondo la legge normale, il rapporto tra l'errore quadratico medio σ e l'errore medio a vale

$$\frac{\sigma}{a} = \sqrt{\frac{\pi}{2}} = 1.2533\dots$$

- Per misure affette da errori distribuiti secondo la legge normale, l'errore quadratico medio ed il modulo di precisione h sono legati dalla

$$\sigma = \frac{1}{h\sqrt{2}}$$

l'errore medio ed il modulo di precisione invece dalla

$$a = \frac{1}{h\sqrt{\pi}}$$

6.4 Lo scarto normalizzato

Introduciamo in luogo dello scarto z il risultato della misura x ; questo è legato a z dalla $z = x - x^*$ (relazione che implica anche $dz = dx$). In luogo del modulo di precisione h usiamo poi l'errore quadratico medio σ ; la funzione di Gauss si può allora scrivere nella forma

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x^*}{\sigma}\right)^2}$$

Definiamo ora una nuova variabile t , legata alla x dalla relazione

$$t = \frac{x - x^*}{\sigma} \quad \Rightarrow \quad dt = \frac{dx}{\sigma}$$

Essa prende il nome di *scarto normalizzato* della x ; vogliamo trovare la funzione di frequenza $f(t)$ della t nell'ipotesi che la x abbia distribuzione normale. Siano x_1 ed x_2 due qualunque valori della variabile x (con $x_1 < x_2$); sappiamo che

$$\Pr(x \in [x_1, x_2]) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-x^*}{\sigma}\right)^2} dx$$

Siano poi t_1 e t_2 i valori per la t che corrispondono a x_1 e x_2 ; sarà

$$\Pr(t \in [t_1, t_2]) = \int_{t_1}^{t_2} f(t) dt$$

Quando la x è compresa nell'intervallo $[x_1, x_2]$, allora (e soltanto allora) la t è compresa nell'intervallo $[t_1, t_2]$; pertanto la probabilità che x sia compresa in $[x_1, x_2]$ deve essere uguale alla probabilità che t sia compresa in $[t_1, t_2]$.

Eseguiamo sull'espressione della probabilità per x un cambiamento di variabile, sostituendovi la t :

$$\Pr(x \in [x_1, x_2]) = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{1}{2}t^2} dt$$

Confrontando le due espressioni (che, ricordiamo, devono valere per qualunque coppia di valori x_1 e x_2), si ricava immediatamente che deve essere

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$$

La cosa importante è che in questa espressione non compaiono né l'errore quadratico medio σ né alcuna altra grandezza dipendente dal modo in cui la misura è stata effettuata, ma *solo costanti*: in altre parole *lo scarto normalizzato ha una distribuzione di probabilità indipendente dalla precisione della misura*.

Di questa proprietà si fa uso, ad esempio, per comporre in un unico grafico campioni di misure aventi precisione diversa: se due osservatori misurano la stessa grandezza commettendo solo errori casuali, le distribuzioni delle loro misure saranno normali; ma se gli errori commessi sono diversi, raggruppando i due insiemi di osservazioni in un solo istogramma l'andamento di quest'ultimo non è gaussiano. Però gli scarti normalizzati hanno la stessa legge di distribuzione per entrambi i misuratori, indipendentemente dall'entità dei loro errori, e possono essere cumulati in un unico istogramma.

Altra conseguenza dell'indipendenza da σ della funzione di frequenza di t , è che la probabilità per una misura di avere scarto normalizzato compreso tra due valori costanti prefissati risulta indipendente dalla precisione della misura stessa;

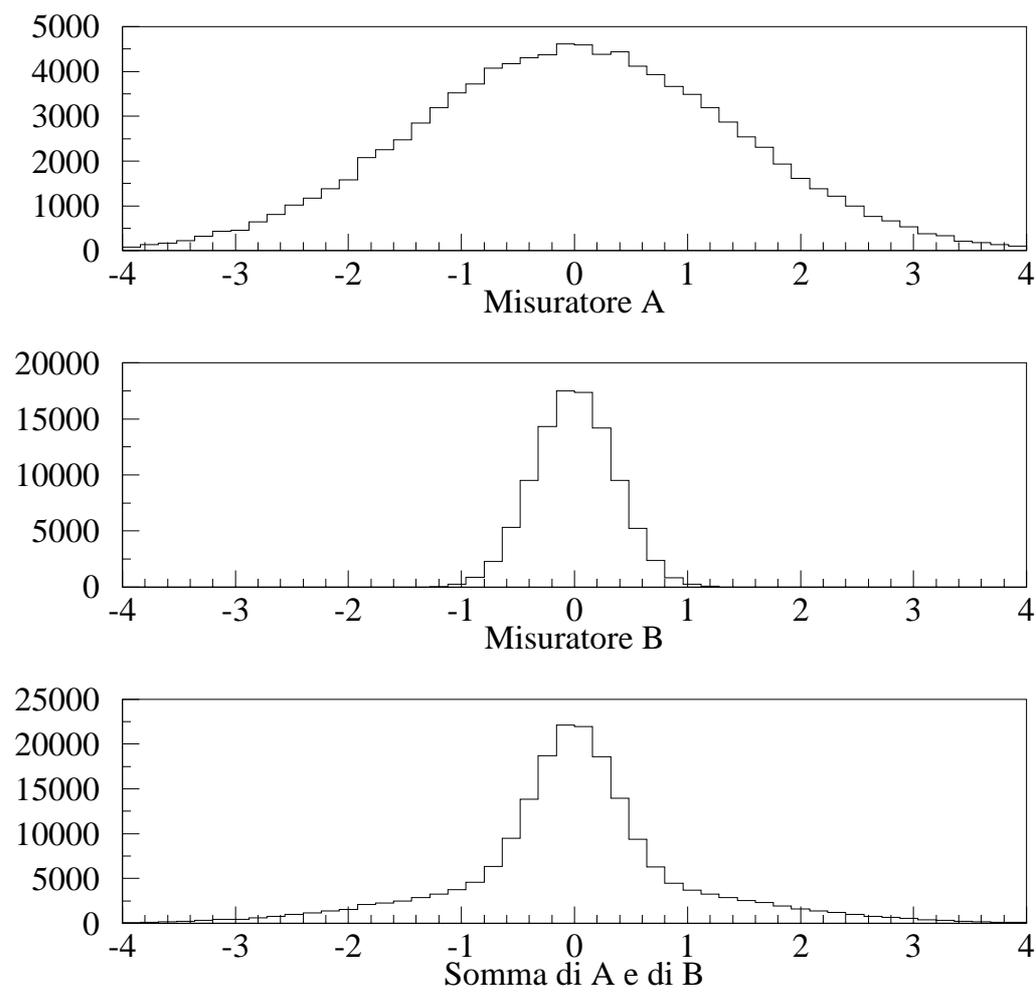


Figura 6.3: Istogrammi relativi a due campioni di misure aventi differente precisione; e l'istogramma relativo ai dati di entrambi i campioni.

ad esempio si ha

$$\Pr(t \in [-1, +1]) \equiv \Pr(x \in [-\sigma, +\sigma]) = \frac{1}{\sqrt{2\pi}} \int_{-1}^{+1} e^{-\frac{t^2}{2}} dt \approx 0.6827$$

$$\Pr(t \in [-2, +2]) \equiv \Pr(x \in [-2\sigma, +2\sigma]) = \frac{1}{\sqrt{2\pi}} \int_{-2}^{+2} e^{-\frac{t^2}{2}} dt \approx 0.9545$$

$$\Pr(t \in [-3, +3]) \equiv \Pr(x \in [-3\sigma, +3\sigma]) = \frac{1}{\sqrt{2\pi}} \int_{-3}^{+3} e^{-\frac{t^2}{2}} dt \approx 0.9973$$

Possiamo far uso di una di queste relazioni per dare un'interpretazione probabilistica dell'errore quadratico medio:

- *Le misure affette da errori casuali (e quindi normali) hanno una probabilità del 68% di cadere all'interno di un intervallo di semiampiezza σ centrato sul valore vero della grandezza misurata.*
- *L'intervallo di semiampiezza σ centrato su di una misura qualsiasi ha pertanto una probabilità del 68% di contenere il valore vero, semprechè gli errori siano casuali e normali.*

6.5 Il significato geometrico di σ

Calcoliamo ora la derivata prima della funzione di Gauss:

$$\frac{df}{dz} = -\frac{z}{\sqrt{2\pi}\sigma^3} e^{-\frac{1}{2}\frac{z^2}{\sigma^2}}$$

La funzione $f(z)$ è crescente ($f'(z) > 0$) quando z è negativa, e viceversa; ha quindi un massimo per $z = 0$, come richiesto dalle ipotesi fatte. La derivata seconda invece vale

$$\begin{aligned} \frac{d^2f}{dz^2} &= -\frac{1}{\sqrt{2\pi}\sigma^3} e^{-\frac{1}{2}\frac{z^2}{\sigma^2}} - \frac{z}{\sqrt{2\pi}\sigma^3} e^{-\frac{1}{2}\frac{z^2}{\sigma^2}} \left(-\frac{z}{\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma^3} e^{-\frac{1}{2}\frac{z^2}{\sigma^2}} \left(\frac{z^2}{\sigma^2} - 1\right) \end{aligned}$$

e si annulla quando $z = \pm\sigma$.

Da qui si può allora ricavare il *significato geometrico* dell'errore quadratico medio σ in relazione alla distribuzione normale:

L'errore quadratico medio σ può essere interpretato geometricamente come valore assoluto delle ascisse dei due punti di flesso della curva di Gauss.

6.6 La curva di Gauss nella pratica

Un campione di N misure di una grandezza fisica con valore vero x^* , affette da soli errori casuali normali con errore quadratico medio σ , avrà media \bar{x} prossima a x^* (sappiamo infatti che la varianza della media vale σ^2/N e tende a zero al crescere di N), e varianza s^2 prossima a σ^2 (anche la varianza di s^2 tende a zero al crescere di N : vedi in proposito l'appendice A).

Per N abbastanza grande, si può dunque assumere $s \approx \sigma$ ed interpretare lo stesso scarto quadratico medio del campione s , in luogo di σ (peraltro ignoto), come semiampiezza dell'intervallo di confidenza corrispondente ad una probabilità del 68%.

Purtroppo non è generalmente possibile capire, dall'andamento di un insieme di osservazioni, se fossero o meno presenti errori sistematici; un campione di misure ripetute, effettuate confrontando la lunghezza di un oggetto con un regolo graduato mal tarato, avrà distribuzione ancora normale: solo centrata attorno ad una media che non corrisponde al valor vero.

Al contrario, se la distribuzione delle misure non è normale sicuramente c'è qualcosa di sospetto nei dati che stiamo esaminando; sorge quindi il problema di stimare se un insieme di dati ha o non ha distribuzione gaussiana (o meglio, di stimare con quale probabilità provenga da una distribuzione gaussiana).

Per far questo si può ricorrere ad alcune proprietà matematiche della curva: ad esempio, si possono calcolare l'errore medio e l'errore quadratico medio per verificare se il loro rapporto ha un valore vicino a quello teorico; oppure si può calcolare la frazione di dati che cadono tra $\bar{x} - s$ e $\bar{x} + s$ e confrontare il numero ottenuto con il valore teorico di 0.68.

Il modo migliore di eseguire il confronto è però quello di disegnare assieme all'istogramma dei dati anche la curva teorica relativa; a questo livello il confronto può essere soltanto visuale, ma esistono metodi matematici (*metodo del chi quadro*) che permettono di stimare con esattezza la probabilità che i dati di un istogramma provengano da una data distribuzione, nel nostro caso quella normale.

Per sovrapporre la curva di Gauss ad un istogramma, occorre comunque moltiplicarne in ogni punto l'ordinata per un fattore costante. L'altezza dell'istogramma è infatti in ogni intervallo data da

$$y_i = \frac{n_i A}{\Delta x_i}$$

dove n_i è il numero di valori osservati nell'intervallo di centro x_i ed ampiezza Δx_i , mentre A è l'area del rettangolo corrispondente ad una osservazione.

Al tendere del numero N di misure effettuate all'infinito, risulta

$$\begin{aligned}\lim_{N \rightarrow \infty} \frac{n_i}{N} &= \Pr \left(x_i - \frac{\Delta x_i}{2} \leq x < x_i + \frac{\Delta x_i}{2} \right) \\ &= \frac{1}{\sigma \sqrt{2\pi}} \int_{x_i - \frac{\Delta x_i}{2}}^{x_i + \frac{\Delta x_i}{2}} e^{-\frac{1}{2} \left(\frac{x-x^*}{\sigma} \right)^2} dx\end{aligned}$$

e dunque

$$y_i \longrightarrow \frac{NA}{\Delta x_i} \frac{1}{\sigma \sqrt{2\pi}} \int_{x_i - \frac{\Delta x_i}{2}}^{x_i + \frac{\Delta x_i}{2}} e^{-\frac{1}{2} \left(\frac{x-x^*}{\sigma} \right)^2} dx$$

Cioè l'altezza dell'istogramma, in ognuna delle classi di frequenza, tende al valor medio sull'intervallo corrispondente della funzione di Gauss moltiplicato per un fattore costante NA . Allora la curva da sovrapporre all'istogramma sperimentale deve essere quella che corrisponde alla funzione

$$f(x) = \frac{NA}{s \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\bar{x}}{s} \right)^2}$$

(in luogo del valore vero x^* e dell'errore quadratico medio σ , generalmente ignoti, si pongono le loro stime, \bar{x} e s rispettivamente, ottenute dal campione stesso); osserviamo che $f(x)$ sottende la stessa area NA dell'istogramma.

Se gli intervalli hanno tutti la medesima ampiezza Δx , l'area del rettangolo elementare vale $A = \Delta x$, assumendo l'arbitraria unità di misura per le ordinate pari all'altezza costante del rettangolo elementare, e la funzione diviene

$$f(x) = \frac{N \Delta x}{s \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\bar{x}}{s} \right)^2}$$

6.7 Esame dei dati

Talvolta nella misura si compiono errori non classificabili né come casuali né come sistematici: ad esempio, dopo aver misurato un angolo di un triangolo con un goniometro, si può riportare come valore un numero diverso scambiando tra di loro due cifre contigue. La conseguenza sarà quella di ottenere per il risultato finale della misura (la somma degli angoli interni di un triangolo) un dato molto differente dagli altri, che si impone quindi alla nostra attenzione come *sospetto*.

Nasce quindi il desiderio di avere un criterio preciso in base al quale decidere se un dato possa o meno considerarsi *sospetto*, ed essere in conseguenza eliminato.

Normalmente la procedura consigliata è la seguente: dopo aver calcolato media e scarto quadratico medio, si eliminano dal campione i dati che differiscono

da \bar{x} per più di tre volte s . Sappiamo infatti che valori che si trovino nella regione oltre 3σ hanno probabilità molto bassa di presentarsi (del tre per mille circa); bisogna comunque osservare che questo modo di procedere è giustificato *solo* in presenza di un *numero piccolo* di dati.

Se le misure effettuate sono in numero per esempio di 60, ci si aspettano in media oltre 3σ solo 0.18 dati; se ne troviamo uno o più d'uno, possiamo ragionevolmente etichettarli come *sospetti* e scartarli. Le cose cambiano se ci troviamo di fronte invece ad un milione di misure, per le quali ci aspettiamo che ben 3000 cadano (per motivi perfettamente normali) al di fuori dell'intervallo di 3σ , e non possiamo quindi permetterci di scartare alcun dato particolare.

6.8 Sommario delle misure dirette

Per concludere, dovendo effettuare delle misure dirette:

- *Bisogna considerare criticamente le modalità della misura e le formule usate, e controllare le caratteristiche di costruzione e d'uso degli strumenti per mettere in evidenza la possibilità di errori sistematici; se questi sono presenti bisogna eliminarli, o modificando le operazioni da compiere o correggendo opportunamente i risultati.*
- *Potendo bisogna effettuare misure ripetute, perchè in questo caso sappiamo stimare ragionevolmente l'errore commesso (se non è possibile effettuare misure ripetute, si assumerà convenzionalmente come errore la sensibilità dello strumento); e bisogna effettuarne quante più possibile per aumentare in corrispondenza la validità statistica dei nostri risultati.*
- *Se il numero di misure effettuate è basso¹ si scartano quei dati che differiscano dal valor medio per più di 3 volte lo scarto quadratico medio s . Effettuata questa operazione si ricalcolano la media \bar{x} e lo scarto quadratico medio s , e si ricava da quest'ultimo la stima dell'errore della media $\sigma_{\bar{x}}$ costituita da $s_{\bar{x}} = s/\sqrt{N}$.*
- *Come valore più verosimile per la grandezza misurata si assume \bar{x} , e come errore di questo valore $s_{\bar{x}}$; se le misure sono in numero sufficiente e non si sono commessi errori sistematici, il significato dell'errore è quello di semiampiezza dell'intervallo di*

¹“Basso” si può ad esempio considerare un numero di misure tale che il numero atteso di eventi da scartare in base alla distribuzione normale sia inferiore all'unità.

*confidenza centrato sulla media avente probabilità di includere
il valore vero pari al 68%.*

Capitolo 7

Le misure indirette

Misure indirette, come sappiamo, sono quelle eseguite non sulla grandezza fisica che interessa determinare ma su altre grandezze che sono a quest'ultima legate da una qualche relazione funzionale, che ci permetterà poi di ricavarne il valore mediante il calcolo.

Assumiamo che queste grandezze vengano misurate direttamente: gli inevitabili errori commessi per ognuna di esse si ripercuoteranno poi attraverso i calcoli effettuati, e si *propagheranno* fino al risultato finale; l'entità dell'errore nelle misure indirette dipenderà dunque sia dal valore di quelli commessi nelle misure dirette, sia dalla forma analitica della funzione usata per il calcolo.

Consideriamo il caso del tutto generale di una qualsiasi funzione F di più variabili indipendenti x, y, z, \dots ; ammettendo che i valori di queste variabili si possano ottenere da misure di tipo diretto, vogliamo determinare un algoritmo per ricavare da tali misure una stima del valore vero di F ed una valutazione dell'errore commesso.

7.1 Risultato della misura

Innanzitutto, è chiaro che il valore vero F^* della grandezza F è quello che corrisponde ai valori veri delle variabili indipendenti da cui F dipende:

$$F^* = F(x^*, y^*, z^*, \dots)$$

Non avendo però a disposizione tali valori veri, tutto quello che possiamo fare è usare le migliori stime di cui disponiamo: cioè, supponiamo, i valori medi di

campioni di determinazioni ripetute di tutte queste grandezze; insomma, calcolare il valore \bar{F} assunto dalla funzione in corrispondenza dei valori $\bar{x}, \bar{y}, \bar{z}, \dots$ delle variabili.

Ricordiamo che il valore di una funzione di più variabili F in un qualsiasi punto si può ricavare dal valore assunto dalla F e dalle sue derivate successive in un punto diverso, attraverso la formula dello sviluppo in serie di Taylor:

$$F(x, y, \dots) = F(x_0, y_0, \dots) + \frac{\partial F}{\partial x}(x - x_0) + \frac{\partial F}{\partial y}(y - y_0) + \dots + \frac{\partial^2 F}{\partial x^2} \dots$$

(in cui per brevità si è ommesso di indicare che le derivate parziali vanno calcolate per i valori delle variabili $x = x_0, y = y_0$ e così via).

Se trascuriamo i termini di ordine superiore al primo, possiamo in particolare ricavare:

$$\bar{F} = F(\bar{x}, \bar{y}, \dots) \approx F(x^*, y^*, \dots) + \frac{\partial F}{\partial x}(\bar{x} - x^*) + \frac{\partial F}{\partial y}(\bar{y} - y^*) + \dots$$

Prendendo il valor medio di entrambi i membri, e tenendo presente nei passaggi che risulta $E(\bar{x} - x^*) = E(\bar{x}) - x^* \equiv 0$ in assenza di errori sistematici (e similmente risulta per le altre variabili), si ottiene

$$\begin{aligned} E\{F(\bar{x}, \bar{y}, \dots)\} &\approx F(x^*, y^*, \dots) + \frac{\partial F}{\partial x} E(\bar{x} - x^*) + \frac{\partial F}{\partial y} E(\bar{y} - y^*) + \dots \\ &= F(x^*, y^*, \dots) \equiv F^* \end{aligned}$$

In definitiva:

In media, il valore di una funzione F calcolato per le medie misurate delle variabili coincide col valor vero.

Ricordiamo che questa conclusione è valida solo approssimativamente, perchè nello sviluppo in serie di Taylor abbiamo trascurato tutti i termini di ordine superiore al primo; ma quali sono i limiti della validità della conclusione? In quali casi si possono cioè effettivamente considerare trascurabili i termini del second'ordine e degli ordini superiori?

Ognuno dei termini di ordine i nello sviluppo di F contiene una delle derivate i -esime della funzione, moltiplicata per un fattore del tipo $(\bar{x} - x^*)$ elevato alla i -esima potenza e divisa per il fattoriale di i ; sarà senz'altro lecito trascurare questi termini se le differenze tra i valori medi stimati per le variabili indipendenti ed i loro valori veri sono piccole, in altre parole *se gli errori commessi nelle misure dirette sono piccoli.*

7.2 Combinazioni lineari di misure dirette

Supponiamo che le misure dirette delle variabili indipendenti x, y, \dots siano esenti da errori sistematici, e che siano pertanto distribuite secondo la legge normale; ora, si può dimostrare il seguente

Teorema: *Se più variabili casuali hanno distribuzione normale, una qualsiasi loro combinazione lineare con coefficienti costanti è anch'essa distribuita secondo la legge normale.*

Consideriamo allora quelle particolari funzioni che sono le combinazioni lineari di più variabili:

$$F = k_1 x_1 + k_2 x_2 + k_3 x_3 + \dots = \sum_i k_i x_i$$

Abbiamo già visto nei paragrafi 5.2 e 5.4 che la media di una tale funzione è la combinazione lineare, con gli stessi coefficienti, delle medie delle variabili; e che la varianza di F è invece la combinazione lineare delle loro varianze con coefficienti pari ai quadrati dei rispettivi coefficienti:

$$E(F) = \sum_i k_i E(x_i) = \sum_i k_i x_i^* \equiv F^*$$

$$\sigma_F^2 = \sum_i k_i^2 \sigma_i^2$$

Ora, sapendo che la distribuzione della F è normale, siamo anche in grado di attribuire un significato più preciso al suo errore quadratico medio σ_F : quello cioè di semiampiezza dell'intervallo, avente centro sul valore stimato \bar{F} , che contiene il valor vero F^* con una probabilità del 68%.

In definitiva le formule precedenti risolvono il problema delle misure indirette per quelle particolari funzioni che sono le combinazioni lineari, permettendoci di calcolare per esse sia il valore stimato più verosimile che l'errore, e dandoci inoltre l'interpretazione probabilistica di questo errore.

7.3 Errori nelle misure indirette

Ora, qualsiasi funzione di più variabili si può considerare in prima approssimazione lineare se ci limitiamo a considerarla in un dominio di definizione abbastanza ristretto da poter trascurare i termini di ordine superiore al primo in uno sviluppo in serie di Taylor: in definitiva possiamo estendere le conclusioni del paragrafo precedente ad una qualsiasi funzione di più variabili

$$F(x, y, z, \dots) \approx F(\bar{x}, \bar{y}, \bar{z}, \dots) + \frac{\partial F}{\partial x} (x - \bar{x}) + \frac{\partial F}{\partial y} (y - \bar{y}) + \frac{\partial F}{\partial z} (z - \bar{z}) + \dots$$

per la cui varianza avremo

$$\sigma_F^2 \approx \left(\frac{\partial F}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial F}{\partial y}\right)^2 \sigma_y^2 + \left(\frac{\partial F}{\partial z}\right)^2 \sigma_z^2 + \dots$$

Questa formula è nota sotto il nome di *formula di propagazione degli errori*; ripetiamo che si tratta di una formula *approssimata*, e valida solo se non si commettono errori troppo grandi nelle misure dirette delle variabili indipendenti. La formula di propagazione è invece *esatta* nel caso particolare (esaminato nel paragrafo precedente) di una combinazione lineare di variabili casuali, caso questo nel quale tutte le derivate parziali di ordine superiore al primo sono identicamente nulle.

7.4 Errore dei prodotti di potenze

Applichiamo ora la formula di propagazione dell'errore a quella particolare classe di funzioni costituita dai prodotti di potenze delle variabili indipendenti: cioè alle funzioni del tipo

$$F(x, y, z, \dots) = x^\alpha \cdot y^\beta \cdot z^\gamma \dots$$

Calcoliamo innanzi tutto le derivate parziali di F ; risulta

$$\begin{cases} \frac{\partial F}{\partial x} = \alpha x^{\alpha-1} \cdot y^\beta \cdot z^\gamma \dots = \alpha \frac{F}{x} \\ \frac{\partial F}{\partial y} = x^\alpha \cdot \beta y^{\beta-1} \cdot z^\gamma \dots = \beta \frac{F}{y} \\ \dots \end{cases}$$

Introducendo questi valori delle derivate nella formula di propagazione dell'errore, avremo

$$\begin{aligned} \sigma_F^2 &= \left(\frac{\partial F}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial F}{\partial y}\right)^2 \sigma_y^2 + \dots \\ &= \alpha^2 \frac{F^2}{x^2} \sigma_x^2 + \beta^2 \frac{F^2}{y^2} \sigma_y^2 + \dots \end{aligned}$$

ed in definitiva

$$\left(\frac{\sigma_F}{F}\right)^2 = \alpha^2 \left(\frac{\sigma_x}{x}\right)^2 + \beta^2 \left(\frac{\sigma_y}{y}\right)^2 + \dots$$

relazione che permette di ricavare con semplici calcoli l'errore relativo di F dagli errori relativi commessi nella misura delle variabili indipendenti. Per quanto detto in precedenza questa relazione è solo una prima approssimazione, valida se le variabili indipendenti sono misurate con errori piccoli.

Capitolo 8

Altre tecniche di elaborazione dei dati

In questo capitolo prenderemo in considerazione alcuni speciali metodi di elaborazione dei dati: la media pesata di determinazioni sperimentali aventi diversa precisione, e la determinazione dei parametri da cui dipende l'equazione di una curva che deve descrivere una relazione tra più variabili interconnesse e misurate indipendentemente (curva interpolante i dati sperimentali).

8.1 Media pesata

Quando si abbiano a disposizione più determinazioni ripetute di una grandezza fisica, sappiamo che da esse si può ricavare un valore unico da usare come risultato finale attraverso il calcolo della media aritmetica; questa è la funzione dei dati con la distribuzione più stretta attorno al valor vero, e ci fornisce quindi la stima più verosimile di esso.

Però questo presuppone che i dati, essendo considerati tutti allo stesso modo nella formula, posseggano la stessa precisione sperimentale: ad esempio che siano stati valutati dallo stesso sperimentatore, con lo stesso strumento e nelle stesse condizioni; in altre parole, che le misure provengano da un'unica popolazione.

Può capitare invece di disporre di più determinazioni della stessa grandezza fisica fatte da sperimentatori diversi, od in condizioni sperimentali differenti: e di voler ugualmente estrarre da queste valutazioni, affette da differenti errori, un valore unico da usare come risultato complessivo.

Facendo le ipotesi che tutte le misure x_i siano tra loro statisticamente indi-

pendenti, ed inoltre affette da errori casuali distribuiti secondo la legge di Gauss, la densità di probabilità corrispondente all'evento casuale costituito dall'osservazione degli N valori x_1, x_2, \dots, x_N si può scrivere (applicando il teorema della probabilità composta)

$$\prod_{i=1}^N \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x^* - x_i}{\sigma_i} \right)^2}$$

dove x^* è il valore vero (ignoto) di x , e le σ_i sono gli errori quadratici medi (supposti noti) delle diverse determinazioni.

Si può dimostrare (vedi in proposito l'appendice G) che, anche sotto condizioni molto più generali di quelle ammesse, la migliore stima del valore vero x^* , ovvero quella di varianza minima, si ottiene prendendo quel valore di x che rende massima la funzione

$$\mathcal{L}(x_1, x_2, \dots, x_N | x) = \prod_{i=1}^N \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - x_i}{\sigma_i} \right)^2}$$

cioè la densità di probabilità di cui sopra, nella quale il valore vero x^* è sostituito dal parametro variabile x .

\mathcal{L} si chiama *funzione di verosimiglianza*: ed è facile rendersi conto che essa rappresenta la densità di probabilità associata all'evento casuale consistente nell'essere il numero x il valor vero della grandezza misurata, qualora di essa si siano ottenute le N stime indipendenti x_i , di errori rispettivi σ_i , supposte seguire la legge normale.

La stima più verosimile (quella di minima varianza) è dunque quella che, rendendo massima \mathcal{L} , individua quel numero che, sulla base delle osservazioni disponibili, possiede la *massima probabilità* di coincidere con il valor vero; ed il metodo consistente nel dare come miglior stima il valore che rende massima \mathcal{L} si dice *metodo della massima verosimiglianza*. Prendendo il logaritmo di \mathcal{L}

$$-2 \log \mathcal{L} = \sum_{i=1}^N \left(\frac{x - x_i}{\sigma_i} \right)^2 + 2 \sum_{i=1}^N \log \sigma_i + 2N \log \sqrt{2\pi}$$

ed osservando che il logaritmo stesso è una funzione crescente dell'argomento, si vede che il massimo di \mathcal{L} corrisponde al minimo di $-2 \log \mathcal{L}$; e la determinazione del valore più verosimile di x (nel caso di errori normali) si riduce allora al problema analitico di trovare il minimo della funzione

$$f(x) = \sum_{i=1}^N \left(\frac{x - x_i}{\sigma_i} \right)^2$$

(infatti nessuno degli altri termini dipende dall'incognita x). Risolviamo il problema facendo uso del calcolo infinitesimale:

$$\frac{df}{dx} = 2 \sum_{i=1}^N \left(\frac{x-x_i}{\sigma_i} \right) \frac{1}{\sigma_i} = 2 \left(x \sum_{i=1}^N \frac{1}{\sigma_i^2} - \sum_{i=1}^N \frac{x_i}{\sigma_i^2} \right)$$

$$\frac{d^2f}{dx^2} = 2 \sum_{i=1}^N \frac{1}{\sigma_i^2} > 0$$

Se per brevità poniamo

$$K = \sum_{i=1}^N \frac{1}{\sigma_i^2}$$

la condizione per l'estremante di $f(x)$ si scrive

$$\frac{df}{dx} = 2 \left(Kx - \sum_{i=1}^N \frac{x_i}{\sigma_i^2} \right) = 0$$

e la derivata prima di f si annulla quando la variabile x assume il valore

$$\bar{x} = \frac{1}{K} \sum_{i=1}^N \frac{x_i}{\sigma_i^2}$$

Il fatto che la derivata seconda sia positiva assicura poi che si tratta effettivamente di un punto di minimo; si vede come \bar{x} sia una *media pesata* dei valori misurati x_i , ottenuta assegnando ad ognuno di essi peso relativo inversamente proporzionale al *quadrato* dell'errore rispettivo.

Per determinare poi l'errore del risultato \bar{x} , è in questo caso possibile usare in tutta generalità la formula della propagazione degli errori: infatti \bar{x} è una particolare funzione delle variabili x_i , di ognuna delle quali conosciamo per ipotesi l'errore quadratico medio σ_i ; ed inoltre dipende *linearmente* da ognuna di queste N variabili, e questo fa sì che la formula di propagazione sia in questo caso *esatta* e non approssimata (dando insomma risultati sempre validi, indipendentemente dall'entità degli errori commessi). Ora, risultando

$$\frac{\partial \bar{x}}{\partial x_j} = \frac{1}{K} \frac{\partial}{\partial x_j} \left(\sum_{i=1}^N \frac{x_i}{\sigma_i^2} \right) = \frac{1}{K \sigma_j^2}$$

avremo

$$\sigma_{\bar{x}}^2 = \sum_{j=1}^N \left(\frac{\partial \bar{x}}{\partial x_j} \right)^2 \sigma_j^2 = \frac{1}{K^2} \sum_{j=1}^N \frac{1}{\sigma_j^2} = \frac{1}{K}$$

cioè

$$\sigma_{\bar{x}}^2 = \frac{1}{\sum_{i=1}^N \frac{1}{\sigma_i^2}}$$

Per la osservata linearità della formula, la media pesata \bar{x} (nelle ipotesi ammesse) è una variabile casuale normale come le singole x_i ; ed il suo errore quadratico medio $\sigma_{\bar{x}}$ ha dunque l'analogia interpretazione di semiampiezza dell'intervallo con centro in \bar{x} avente probabilità pari al 68% di contenere il valore vero x^* .

E' da notare come, prima di comporre tra loro determinazioni indipendenti della stessa grandezza, sia opportuno controllare che queste siano (entro i rispettivi errori) tra loro *compatibili*; analogamente a quanto si fa per le misure ripetute, è preferibile scartare dati che non vadano d'accordo con gli altri entro i limiti della pura casualità.

Il caso di N misure ripetute effettuate nelle medesime condizioni sperimentali non è altro che il caso particolare in cui tutti gli errori quadratici medi σ_i sono uguali tra di loro: la media pesata si riduce allora alla media aritmetica. Questo prova l'asserto del paragrafo 4.4 (giustificazione della media); abbiamo finalmente *dimostrato* che la media aritmetica è il valore *più verosimile* della grandezza misurata: cioè quello che ha la massima probabilità di coincidere con il valor vero sulla base del nostro campione di misure, e che rappresenta per ciò stesso la stima di minima varianza.

8.2 Interpolazione dei dati con una curva

Può in alcuni casi capitare di conoscere la forma analitica della legge fisica che mette in relazione tra loro due variabili, e di dover stimare da dati misurati il valore di uno o più parametri da cui tale funzione dipende.

Ad esempio, nel moto dei corpi soggetti all'azione di una forza costante, le velocità assunte in istanti successivi dal corpo crescono linearmente rispetto ai tempi trascorsi, secondo la nota formula $v = v_0 + at$; misurando in istanti successivi del moto tempi e velocità, i punti aventi per coordinate cartesiane i valori determinati per queste due grandezze devono disporsi approssimativamente lungo una linea retta: e sarebbero tutti quanti esattamente allineati se fosse possibile misurare senza commettere errori.

In questo ultimo caso sarebbe possibile ricavare immediatamente dal grafico il valore dell'accelerazione del moto, che corrisponderebbe al coefficiente angolare (alla pendenza) della retta tracciata; vedremo ora come, pur commettendo errori, sia comunque possibile ricavare una stima sia dei valori dei parametri da cui l'equazione del moto dipende, sia degli errori inerenti a tale valutazione.

C'è una qualche analogia fra questo problema e quello delle misure indirette, nel senso che in entrambi i casi esiste una relazione funzionale tra più grandezze fisiche; tuttavia, mentre in quel caso la funzione era completamente nota e veniva usata per trovare il valore di una di quelle grandezze una volta misurati quelli di tutte le altre, qui si suppone di conoscere soltanto la forma della funzione: ma sono ignoti uno o più parametri da cui pure essa dipende, e si usano i valori osservati di tutte le grandezze per stimare quelli dei parametri stessi.

8.2.1 Interpolazione lineare per due variabili

Cominciamo col supporre che le variabili oggetto della misura siano due sole, e che la legge che le mette in relazione reciproca sia di tipo lineare:

$$y = a + bx$$

Supponiamo poi che siano state effettuate misure del valore della x e di quello corrispondente assunto dalla y in diverse condizioni, così che si disponga in definitiva di N coppie di valori tra loro corrispondenti (x_i, y_i) ; abbiamo già detto che, una volta riportati sul piano cartesiano punti con queste coordinate, essi si dovranno disporre approssimativamente lungo una linea retta.

Si può dimostrare che vale, sul piano, qualcosa di analogo a quanto abbiamo già asserito riguardo alla media aritmetica di misure ripetute di una stessa grandezza fisica (cioè, geometricamente, su di una retta, visto che quelle determinazioni potevano essere univocamente rappresentate da dei punti su di una retta orientata); infatti

- *Sulla base delle misure effettuate, non si può escludere con certezza che alcuna delle infinite rette del piano corrisponda a quella vera su cui le nostre osservazioni si disporrebbero in assenza di errori; tuttavia esse non appaiono tutte quante ugualmente verosimili, e la verosimiglianza sarà in qualche modo in relazione alla “distanza complessiva” tra i nostri punti e la retta stessa.*
- *Nel caso particolare che siano verificate le seguenti ipotesi:*
 1. *una sola delle variabili coinvolte (ad esempio la y) è affetta da errori;*
 2. *gli errori quadratici medi delle misure dei diversi valori di y sono tutti uguali (o comunque non molto differenti);*
 3. *questi errori seguono la legge normale di distribuzione;*
 4. *le N determinazioni effettuate sono tra loro statisticamente indipendenti;*

dimostriamo ora che per “distanza” si deve intendere la lunghezza del segmento parallelo all’asse y compreso tra il punto misurato e la retta esaminata.

Infatti, detto σ_y l’errore quadratico medio delle y_i , la funzione di verosimiglianza è

$$\mathcal{L}(x_1, y_1, x_2, y_2, \dots, x_N, y_N | a, b) = \prod_{i=1}^N \frac{1}{\sigma_y \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{a + bx_i - y_i}{\sigma_y} \right)^2}$$

Per scrivere questa espressione si è fatto uso di tutte le ipotesi postulate; in particolare, il fatto che le x_i siano misurate senza errore ci permette di affermare che il valor vero assunto in corrispondenza dalla y è $a + bx_i$ qualora la legge fisica che lega le due variabili sia data dall’equazione $y = a + bx$.

Questa funzione di verosimiglianza rappresenta la densità di probabilità collegata all’evento casuale consistente nell’essere la legge fisica che lega x ad y rappresentata dall’equazione $y = a + bx$, qualora si siano ottenuti gli N valori misurati (x_i, y_i) , e sotto le quattro ipotesi su elencate.

Il metodo della massima verosimiglianza ci permette poi di affermare che la stima, appunto, più verosimile (la retta che corrisponde al massimo della probabilità) sarà anche la stima di minima varianza. Prendendo il logaritmo di entrambi i membri risulta

$$-2 \log \mathcal{L} = \frac{1}{\sigma_y^2} \sum_{i=1}^N (a + bx_i - y_i)^2 + 2N \log \sigma_y + 2N \log \sqrt{2\pi}$$

I valori più verosimili dei parametri a e b sono quelli per cui è massima \mathcal{L} , ovvero è minima $-2 \log \mathcal{L}$: il problema dell’interpolazione lineare dunque si riduce (se sono soddisfatte le ipotesi citate) a quello di trovare tra le infinite rette del piano quella che rende minima la funzione

$$f(a, b) = \sum_{i=1}^N [(a + bx_i) - y_i]^2$$

(essendo tutti gli altri termini indipendenti dalle due incognite a e b).

L’interpretazione geometrica è evidente: la retta soluzione del nostro problema è (come già preannunciato) quella che rende minima la somma dei quadrati delle distanze, misurate parallelamente all’asse y , dall’insieme dei punti misurati. Per trovare il valore dei coefficienti dell’equazione di tale retta, calcoliamo ora le derivate prime della funzione f :

$$\frac{\partial f}{\partial a} = 2 \sum_{i=1}^N (a + bx_i - y_i) = 2 \left(Na + b \sum_{i=1}^N x_i - \sum_{i=1}^N y_i \right)$$

$$\frac{\partial f}{\partial b} = 2 \sum_{i=1}^N (a + bx_i - y_i)x_i = 2 \left(a \sum_{i=1}^N x_i + b \sum_{i=1}^N x_i^2 - \sum_{i=1}^N x_i y_i \right)$$

Imponendo che le due derivate prime siano contemporaneamente nulle, dovranno essere verificate le

$$\begin{cases} a \cdot N + b \cdot \sum_i x_i = \sum_i y_i \\ a \cdot \sum_i x_i + b \cdot \sum_i x_i^2 = \sum_i x_i y_i \end{cases}$$

Questo sistema di due equazioni in due incognite ammette, come si può verificare, sempre una ed una sola soluzione, purchè vi siano almeno due punti sperimentali non coincidenti; esaminando poi le derivate seconde, si troverebbe che essa corrisponde effettivamente ad un minimo. La soluzione è

$$\begin{cases} a = \frac{1}{\Delta} \left[(\sum_i x_i^2) \cdot (\sum_i y_i) - (\sum_i x_i) \cdot (\sum_i x_i y_i) \right] \\ b = \frac{1}{\Delta} \left[N \cdot (\sum_i x_i y_i) - (\sum_i x_i) \cdot (\sum_i y_i) \right] \end{cases}$$

in cui si è posto per brevità

$$\Delta = N \sum_i x_i^2 - (\sum_i x_i)^2$$

(queste formule sono note sotto il nome di *formule dei minimi quadrati*).

Per quanto attiene al calcolo degli errori commessi nella valutazione di a e b in base ai dati, osserviamo che entrambi si ricavano da relazioni lineari in ognuna delle variabili affette da errore che, nelle nostre ipotesi, sono le sole y_i ; possiamo dunque adoperare la formula della propagazione degli errori, che è in questo caso esatta. Se indichiamo con σ_y^2 la varianza comune a tutte le y_i , risulta per a

$$\begin{aligned} \sigma_a^2 &= \sum_{j=1}^N \left(\frac{\partial a}{\partial y_j} \right)^2 \sigma_y^2 \\ &= \sigma_y^2 \sum_{j=1}^N \frac{1}{\Delta^2} \left(\sum_{i=1}^N x_i^2 - x_j \sum_{i=1}^N x_i \right)^2 \\ &= \frac{\sigma_y^2}{\Delta^2} \sum_{j=1}^N \left[\left(\sum_{i=1}^N x_i^2 \right)^2 + x_j^2 \left(\sum_{i=1}^N x_i \right)^2 - 2x_j \left(\sum_{i=1}^N x_i \right) \left(\sum_{i=1}^N x_i^2 \right) \right] \\ &= \frac{\sigma_y^2}{\Delta^2} \left[N \left(\sum_{i=1}^N x_i^2 \right)^2 + \left(\sum_{i=1}^N x_i^2 \right) \left(\sum_{i=1}^N x_i \right)^2 - 2 \left(\sum_{i=1}^N x_i \right)^2 \left(\sum_{i=1}^N x_i^2 \right) \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{\sigma_y^2}{\Delta^2} \left[N \left(\sum_{i=1}^N x_i^2 \right)^2 - \left(\sum_{i=1}^N x_i \right)^2 \left(\sum_{i=1}^N x_i^2 \right) \right] \\
&= \frac{\sigma_y^2}{\Delta^2} \left(\sum_{i=1}^N x_i^2 \right) \left[N \left(\sum_{i=1}^N x_i^2 \right) - \left(\sum_{i=1}^N x_i \right)^2 \right] \\
&= \sigma_y^2 \frac{\sum_i x_i^2}{\Delta}
\end{aligned}$$

e, similmente, per b :

$$\begin{aligned}
\sigma_b^2 &= \sum_{j=1}^N \left(\frac{\partial b}{\partial y_j} \right)^2 \sigma_y^2 \\
&= \sigma_y^2 \sum_{j=1}^N \frac{1}{\Delta^2} \left(Nx_j - \sum_{i=1}^N x_i \right)^2 \\
&= \frac{\sigma_y^2}{\Delta^2} \sum_{j=1}^N \left[N^2 x_j^2 + \left(\sum_{i=1}^N x_i \right)^2 - 2Nx_j \sum_{i=1}^N x_i \right] \\
&= \frac{\sigma_y^2}{\Delta^2} \left[N^2 \left(\sum_{i=1}^N x_i^2 \right) + N \left(\sum_{i=1}^N x_i \right)^2 - 2N \left(\sum_{i=1}^N x_i \right)^2 \right] \\
&= \frac{N\sigma_y^2}{\Delta^2} \left[N \left(\sum_{i=1}^N x_i^2 \right) - \left(\sum_{i=1}^N x_i \right)^2 \right] \\
&= \sigma_y^2 \frac{N}{\Delta}
\end{aligned}$$

In definitiva, a e b hanno errori quadratici medi dati dalle

$$\begin{cases} \sigma_a = \sigma_y \sqrt{\frac{\sum_i x_i^2}{\Delta}} \\ \sigma_b = \sigma_y \sqrt{\frac{N}{\Delta}} \end{cases}$$

ed il fatto poi che a e b siano funzioni lineari di variabili che seguono la legge di Gauss ci permette ancora di affermare che anch'esse sono distribuite secondo la legge normale—e di attribuire così ai loro errori il consueto significato.

8.2.2 Stima a posteriori degli errori di misura

E' da osservare come nelle formule dei minimi quadrati non compaia il valore di σ : la soluzione del problema dell'interpolazione lineare è indipendente dall'entità degli errori di misura; ed i coefficienti della retta interpolante possono essere calcolati anche se gli errori sulle y non sono noti (purchè naturalmente si assuma che siano tutti uguali tra loro).

Se non è a priori nota la varianza delle y , essa può però essere stimata a partire dai dati stessi una volta eseguita l'interpolazione lineare; infatti gli stessi ragionamenti fatti per le variabili casuali unidimensionali potrebbero essere ripetuti con le opportune modifiche sul piano, per giungere a risultati analoghi.

In una dimensione abbiamo a suo tempo potuto collegare l'errore commesso alla dispersione dei dati rispetto al *valore* stimato della grandezza misurata; sul piano è in effetti ancora possibile calcolare l'errore commesso, partendo dalla dispersione dei dati misurata rispetto alla *retta* stimata che passa attraverso di essi: dati disposti mediamente lontano da questa retta indicheranno errori maggiori rispetto a dati ben allineati (e quindi vicini alla retta interpolante).

In una dimensione abbiamo visto che la dispersione dei dati, misurata dal valor medio del quadrato degli scarti rispetto alla loro media aritmetica (nostra migliore stima per la grandezza misurata), era sistematicamente in difetto rispetto alla corrispondente grandezza riferita all'intera popolazione delle misure. Sul piano si può analogamente dimostrare che il valor medio del quadrato delle distanze dei punti misurati dalla retta nostra migliore stima è ancora sistematicamente in difetto rispetto alla varianza riferita alla popolazione delle misure stesse.

Così come abbiamo dimostrato che, al fine di correggere questa sottostima (in media) per le misure ripetute, occorre dividere la somma dei quadrati degli scarti per $N - 1$ invece che per N , si potrebbe analogamente dimostrare che una corretta stima dell'errore dei punti misurati si ha, in media, dividendo l'analogha somma per $N - 2$; in definitiva, che la corretta stima di σ_y è data dalla formula

$$\sigma_y^2 = \frac{\sum_i [(a + bx_i) - y_i]^2}{N - 2}$$

In essa a numeratore compare la somma dei quadrati delle distanze dei punti misurati (x_i, y_i) dalla retta di equazione $a + bx$; distanze calcolate ancora secondo la direzione parallela all'asse delle ordinate.

Questa formula¹ permette una corretta stima dell'errore dei dati interpolati, qualora sia impossibile (o scomodo) determinarli per altra via; l'errore è stimato dai residui dei dati sperimentali, ed è quindi scientificamente affidabile.

¹Una formula equivalente (ma più semplice) per il calcolo di σ_y si può trovare nella pagina 93, alla fine dell'appendice B.

Il fatto che la corretta stima dell'errore si ottenga dividendo per $N - 2$ invece che per N deve essere messo in relazione con il fatto che gli scarti, invece che rispetto al valor vero, sono calcolati rispetto ad un valore stimato che dipende da *due* parametri, che sono a loro volta stati preventivamente determinati sulla base dei dati sperimentali: cioè i due coefficienti a e b dell'equazione della retta.

Nell'analogo caso della stima dell'errore quadratico medio di una variabile casuale unidimensionale, gli scarti erano calcolati rispetto ad un valore che, unica grandezza necessaria, veniva preventivamente determinato sulla base delle misure: appunto la media aritmetica.

In generale, disponendo di N dati sperimentali dai quali possiamo determinare un valore dell'errore quadratico medio che dipende da M parametri che debbano essere preventivamente derivati dai dati stessi, la modifica da apportare alla formula per ottenere una corretta valutazione dell'errore della popolazione consiste nel dividere la somma dei quadrati degli scarti per un fattore $N - M$.

8.2.3 Interpolazione con una retta per l'origine

Se conosciamo altri vincoli cui debba soddisfare la legge che mette in relazione i valori delle variabili misurate x e y , possiamo imporre che la retta corrispondente appartenga ad un particolare sottoinsieme delle rette del piano; ad esempio, un caso che si può presentare è che la retta sia vincolata a passare per una posizione particolare, che supporremo qui essere l'origine degli assi coordinati.

Una generica retta per l'origine ha equazione $y = mx$; ammesso ancora che gli errori commessi riguardino soltanto la misura della y e non quella della x , che tutti i vari y_i abbiano errori distribuiti secondo la legge normale e tra loro eguali, e che le misure siano tra loro indipendenti, il problema dell'interpolazione lineare si riduce a trovare tra le infinite rette passanti per l'origine quella che rende minima la somma dei quadrati degli scarti (secondo y) dai punti misurati: cioè il minimo della funzione

$$f(m) = \sum_{i=1}^N (mx_i - y_i)^2$$

La derivata prima di f vale

$$\frac{df}{dm} = 2 \sum_{i=1}^N (mx_i - y_i)x_i = 2 \left(m \sum_{i=1}^N x_i^2 - \sum_{i=1}^N x_i y_i \right)$$

Imponendo che essa sia nulla, l'estremante si ha per

$$m = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

e corrisponde in effetti ad un minimo. La legge di propagazione degli errori (esatta anche in questo caso) ci dà

$$\frac{dm}{dy_j} = \frac{x_j}{\sum_i x_i^2}$$

$$\sigma_m^2 = \sum_{j=1}^N \left(\frac{dm}{dy_j} \right)^2 \sigma_y^2 = \frac{\sigma_y^2}{(\sum_i x_i^2)^2} \sum_{j=1}^N x_j^2 = \frac{\sigma_y^2}{\sum_i x_i^2}$$

e la formula per il calcolo degli errori a posteriori diventa

$$\sigma_y^2 = \frac{\sum_i (mx_i - y_i)^2}{N - 1}$$

visto che il parametro da cui l'errore quadratico medio dipende e che deve essere stimato sulla base dei dati è uno soltanto: m .

8.2.4 Interpolazione lineare nel caso generale

Le condizioni 1) e 2) sugli errori delle grandezze misurate x e y date nel paragrafo 8.2.1 non potranno ovviamente mai essere verificate esattamente; come ci si deve comportare quando nemmeno in prima approssimazione le possiamo considerare vere?

Se gli errori quadratici medi delle y_i sono tra loro diversi, non è più possibile raccogliere a fattor comune $1/\sigma^2$ nell'espressione del logaritmo della verosimiglianza, e ciascun addendo sarà diviso per il corrispondente errore σ_i . In definitiva la retta più verosimile si trova cercando il minimo della funzione

$$f(a, b) = \sum_{i=1}^N \left[\frac{(a + bx_i) - y_i}{\sigma_i} \right]^2$$

Questo avviene quando

$$\begin{cases} a = \frac{1}{\Delta} \left[\left(\sum_i \frac{x_i^2}{\sigma_i^2} \right) \cdot \left(\sum_i \frac{y_i}{\sigma_i^2} \right) - \left(\sum_i \frac{x_i}{\sigma_i^2} \right) \cdot \left(\sum_i \frac{x_i y_i}{\sigma_i^2} \right) \right] \\ b = \frac{1}{\Delta} \left[\left(\sum_i \frac{1}{\sigma_i^2} \right) \cdot \left(\sum_i \frac{x_i y_i}{\sigma_i^2} \right) - \left(\sum_i \frac{x_i}{\sigma_i^2} \right) \cdot \left(\sum_i \frac{y_i}{\sigma_i^2} \right) \right] \end{cases}$$

in cui si è posto

$$\Delta = \left(\sum_i \frac{1}{\sigma_i^2} \right) \cdot \left(\sum_i \frac{x_i^2}{\sigma_i^2} \right) - \left(\sum_i \frac{x_i}{\sigma_i^2} \right)^2$$

Le varianze di a e di b saranno poi date dalle

$$\begin{cases} \sigma_a^2 = \frac{1}{\Delta} \sum_i \frac{x_i^2}{\sigma_i^2} \\ \sigma_b^2 = \frac{1}{\Delta} \sum_i \frac{1}{\sigma_i^2} \end{cases}$$

Si deve tuttavia osservare che per applicare questo metodo è necessario conoscere per altra via tutte le N varianze σ_i^2 . Ciò può essere molto laborioso o addirittura impossibile, e non risulta conveniente rinunciare ad una stima unica e ragionevole σ_y^2 di queste varianze per tener conto di una variazione, generalmente debole, delle σ_i in un intervallo limitato di valori della x .

Volendo tener conto dell'errore su entrambe le variabili x ed y , non è generalmente possibile usare un metodo, descritto in alcuni testi, consistente nel cercare la retta che rende minima la somma dei quadrati delle distanze dai punti, misurate però *ortogonalmente* alla retta stessa: a prescindere dalla complicazione della soluzione di un sistema di equazioni non lineari, resta il fatto che se x ed y sono due grandezze fisiche diverse, o anche soltanto misurate con strumenti e metodi diversi, i loro errori quadratici medi sono generalmente diversi; mentre la distanza sul piano attribuisce lo stesso peso agli scarti in x ed a quelli in y .

Per applicare questo metodo si dovrebbe conoscere, per via indipendente, almeno il rapporto tra σ_x e σ_y ; e rappresentare i punti (x_i, y_i) non sul piano (x, y) , bensì su quello delle variabili ridotte $(x/\sigma_x, y/\sigma_y)$.

Per solito nella pratica si preferisce considerare affetta da errore una soltanto delle variabili, ad esempio la y , la scelta cadendo generalmente su quella determinata in maniera *più indiretta*, e che risente perciò degli errori di tutte le altre grandezze misurate direttamente; così, in un diagramma velocità-tempo trascorso o velocità-spazio percorso, si assumerà affetta da errore la sola velocità.

Un eventuale errore sulla x si *propagherà* attraverso la relazione funzionale anche alla y , e, se l'errore quadratico medio σ_y è stimato dai dati sperimentali, esso congloberà anche l'indeterminazione dovuta alla x .

Per meglio chiarire il concetto, consideriamo la legge $v = v_0 + gt$ di caduta di un grave. Se t si ammette esente da errore, una differenza tra il valore vero di v e quello misurato può solo essere imputata ad imperfetta misura di v ; supponiamo invece che ognuno dei valori misurati di t possa presentarsi con un certo scarto non nullo rispetto al valor vero corrispondente: anche se le misure di v fossero esenti da errore, questo vorrebbe dire che i possibili valori che potremmo di v determinare cadrebbero ugualmente in certi intervalli di ampiezza non nulla.

Poichè anche le velocità saranno misurate con errori finiti, l'indeterminazione di misura si cumulerà con quella proveniente dall'incertezza sul tempo in cui

la misura è stata effettuata; ed in definitiva tutto avviene come se l'istante della misura fosse perfettamente noto e la velocità affetta da un errore maggiore.

8.2.5 Interpolazione non lineare

Formule analoghe a quelle trovate si possono ricavare per risolvere il problema dell'interpolazione di curve di ordine superiore al primo (parabole, cubiche, polinomiali in genere) ad un insieme di dati sperimentali, sempre usando il metodo della massima verosimiglianza.

Nel caso poi ci si trovasse di fronte ad una curva di equazione diversa da un polinomio, in parecchi casi è possibile *linearizzare* la relazione cambiando variabile: così, ad esempio, se due grandezze hanno tra loro una relazione di tipo esponenziale, il logaritmo ne avrà una di tipo lineare:

$$y = Ke^{-bx} \quad \Longleftrightarrow \quad \ln y = \ln K - bx = a - bx$$

Appendice A

L'errore della varianza

Può a volte essere utile valutare l'errore della stima della varianza ricavata da un campione di dati sperimentali. Facendo un esempio concreto, supponiamo di disporre di un ampio insieme di valutazioni della stessa grandezza fisica: $N \cdot M$ misure ripetute $x_1, x_2, \dots, x_{N \cdot M}$. Dividiamo questi valori in M sottoinsiemi costituiti da N dati ciascuno, e per ognuno di questi M sottocampioni calcoliamo la media aritmetica dei dati; otterremo così M medie parziali, che indicheremo con i simboli $\bar{x}_1, \dots, \bar{x}_M$.

Lo scopo di queste operazioni può essere quello di verificare che le medie di questi sottocampioni sono distribuite su un intervallo di valori più ristretto di quello su cui si distribuisce l'insieme dei dati originali: in sostanza, per verificare che le medie di N dati hanno errore quadratico medio inferiore a quello dei dati di partenza.

L'errore delle medie dei sottocampioni può essere stimato calcolandone la varianza:

$$\sigma_{\bar{x}}^2 = \frac{1}{M-1} \sum_{i=1}^M (\bar{x}_i - \langle \bar{x} \rangle)^2 \quad (\text{sperimentale})$$

intendendo con $\langle \bar{x} \rangle$ la media delle M medie parziali, che coinciderà necessariamente con la media complessiva dell'intero campione di $N \cdot M$ dati.

Questo valore può essere poi confrontato con quello previsto dalla teoria per la varianza della media di un gruppo di dati, allo scopo di verificare in pratica l'adeguatezza della teoria stessa; tale previsione teorica è come sappiamo data dal rapporto tra la varianza di ognuno dei dati che contribuiscono alla media ed il

numero dei dati stessi:

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{N} \quad (\text{teorico})$$

Come stima di σ si può usare l'errore quadratico medio dell'insieme di tutti gli $N \cdot M$ dati; ma, naturalmente, perchè il confronto tra questi due numeri abbia un significato, *occorre conoscere gli errori* da cui sia la valutazione sperimentale che la previsione teorica di $\sigma_{\bar{x}}$ sono affette.

Consideriamo (come già fatto precedentemente) una popolazione a media zero per semplificare i calcoli:

$$E(x) \equiv x^* = 0$$

I risultati si potranno in seguito facilmente estendere ad una popolazione qualsiasi, tenendo presente il teorema ed i ragionamenti di pagina 43. La varianza di una variabile casuale x , indicata di seguito come $\text{Var}(x)$, si può scrivere come

$$\text{Var}(x) = E(x^2) - [E(x)]^2$$

usando questa formula per calcolare la varianza di s^2 , avremo

$$\text{Var}(s^2) = E(s^4) - [E(s^2)]^2$$

Ora

$$\begin{aligned} s^4 &= \left[\frac{\sum_i x_i^2}{N} - \left(\frac{\sum_i x_i}{N} \right)^2 \right]^2 \\ &= \frac{1}{N^2} (\sum_i x_i^2)^2 - \frac{2}{N^3} (\sum_i x_i^2) (\sum_i x_i)^2 + \frac{1}{N^4} (\sum_i x_i)^4 \end{aligned}$$

Sviluppiamo uno per volta i tre termini a secondo membro; per il primo risulta

$$\begin{aligned} (\sum_i x_i^2)^2 &= (\sum_i x_i^2) (\sum_j x_j^2) \\ &= \sum_i \left(x_i^2 \sum_{j=i} x_j^2 \right) + \sum_i \left(x_i^2 \sum_{j \neq i} x_j^2 \right) \\ &= \sum_i x_i^4 + \sum_{j \neq i} x_i^2 x_j^2 \\ &= \sum_i x_i^4 + 2 \sum_{j < i} x_i^2 x_j^2 \end{aligned}$$

La prima sommatoria comprende N addendi distinti; la seconda è estesa a tutte le possibili *combinazioni* dei valori distinti di i e j presi a due a due: è costituita quindi da

$$C_2^N = \frac{N!}{2!(N-2)!} = \frac{N(N-1)}{2}$$

addendi distinti.

Il fattore 2 che compare davanti ad essa è dovuto al fatto che una coppia di valori degli indici si presentava nella sommatoria su $i \neq j$ una volta come $x_i^2 x_j^2$ e un'altra come $x_j^2 x_i^2$, termini diversi per l'ordine ma con lo stesso valore. In definitiva, passando ai valori medi e tenendo conto dell'indipendenza statistica di x_i e x_j quando è $i \neq j$, risulta

$$E \left\{ \left(\sum_i x_i^2 \right)^2 \right\} = N E(x^4) + N(N-1) \left[E(x^2) \right]^2$$

Con simili passaggi, si ricava per il secondo termine

$$\begin{aligned} \left(\sum_i x_i^2 \right) \left(\sum_j x_j \right)^2 &= \left(\sum_i x_i^2 \right) \left(\sum_j x_j^2 + \sum_{j \neq k} x_j x_k \right) \\ &= \sum_i x_i^4 + \sum_{i \neq j} x_i^2 x_j^2 + \sum_{i \neq j} x_i^3 x_j + \sum_{i \neq j \neq k} x_i^2 x_j x_k \end{aligned}$$

dove gli indici aventi simboli diversi si intendono avere anche valori sempre diversi tra loro nelle sommatorie.

Il valor medio del terzo e del quarto termine si annulla essendo $E(x) = 0$; inoltre gli addendi nella prima sommatoria sono in numero di N e quelli nella seconda in numero di $N(N-1)/2$ e vanno moltiplicati per un fattore 2. Pertanto anche

$$E \left\{ \left(\sum_i x_i^2 \right) \left(\sum_i x_i \right)^2 \right\} = N E(x^4) + N(N-1) \left[E(x^2) \right]^2$$

Infine avremo, con la medesima convenzione sugli indici,

$$\begin{aligned} \left(\sum_i x_i \right)^4 &= \left(\sum_i x_i \right) \left(\sum_j x_j \right) \left(\sum_k x_k \right) \left(\sum_l x_l \right) \\ &= \sum_i x_i^4 + \sum_{i \neq j} x_i^3 x_j + \sum_{i \neq j} x_i^2 x_j^2 + \sum_{i \neq j \neq k} x_i^2 x_j x_k + \sum_{i \neq j \neq k \neq l} x_i x_j x_k x_l \end{aligned}$$

I valori medi del secondo, quarto e quinto termine (che contengono potenze dispari delle x) sono nulli. Gli addendi nella prima sommatoria sono in numero di N ; nella terza vi sono $N(N-1)/2$ termini distinti: ma ciascuno appare in 6 modi diversi solo per l'ordine, corrispondenti al numero C_2^4 di combinazioni dei quattro indici originari i, j, k ed l presi a due a due. Allora

$$E \left(\sum_i x_i \right)^4 = N E(x^4) + 3N(N-1) \left[E(x^2) \right]^2$$

Riprendendo la formule di partenza,

$$E(s^4) = \frac{(N-1)^2}{N^3} E(x^4) + \frac{(N-1)(N^2-2N+3)}{N^3} [E(x^2)]^2$$

Per il valor medio di s^2 , già sappiamo come risulti per la varianza del campione

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

inoltre

$$\sigma^2 = E\{(x-x^*)^2\} = E(x^2)$$

(essendo $x^* = 0$) e

$$\sigma_{\bar{x}}^2 = E\{(\bar{x}-x^*)^2\} = \frac{\sigma^2}{N}$$

da cui abbiamo ottenuto a suo tempo la

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

Per la varianza di s^2 , che vogliamo determinare:

$$\begin{aligned} \text{Var}(s^2) &= E(s^4) - [E(s^2)]^2 \\ &= \frac{(N-1)^2}{N^3} E(x^4) + \\ &\quad + \left[\frac{(N-1)(N^2-2N+3)}{N^3} - \frac{(N-1)^2}{N^2} \right] [E(x^2)]^2 \\ &= \frac{(N-1)^2}{N^3} E(x^4) - \frac{(N-1)(N-3)}{N^3} [E(x^2)]^2 \\ &= \frac{N-1}{N^3} \left\{ (N-1)E(x^4) - (N-3)[E(x^2)]^2 \right\} \end{aligned}$$

Questa relazione ha validità generale. *Nel caso poi che la popolazione ubbidisca alla legge normale* potremo calcolare il valor medio di x^4 usando la forma analitica della funzione di Gauss: per distribuzioni normali risulta in generale, per qualsiasi intero k ,

$$E(x^{2k}) = \frac{(2k)!}{2^k k!} [E(x^2)]^k$$

con

$$E(x^2) = \sigma^2$$

Per la varianza di s^2 se ne ricava

$$E(x^4) = 3\sigma^4$$

e, sostituendo,

$$\text{Var}(s^2) = \frac{2(N-1)}{N^2} [E(x^2)]^2 = \frac{2(N-1)}{N^2} \sigma^4$$

Insomma l'errore quadratico medio della varianza s^2 del campione vale

$$\sigma_{s^2} = \frac{\sqrt{2(N-1)}}{N} \sigma^2$$

La varianza, invece, della stima della varianza della popolazione

$$\sigma^2 = \frac{N}{N-1} s^2$$

vale

$$\text{Var}(\sigma^2) = \left(\frac{N}{N-1}\right)^2 \text{Var}(s^2) = \frac{2}{N-1} \sigma^4$$

L'errore stimato sulla stima della varianza della popolazione è

$$\sigma_{\sigma^2} = \sqrt{\frac{2}{N-1}} \sigma^2$$

Sottolineiamo ancora come queste formule che permettono di calcolare, per una popolazione avente distribuzione normale, gli errori quadratici medi sia della varianza di un campione di N misure che della stima della varianza della popolazione ricavata da un campione di N misure, siano *esatte*.

Se si vuole invece calcolare l'errore da attribuire agli *errori quadratici medi*, cioè alle quantità s e σ radici quadrate delle varianze di cui sopra, non è possibile dare delle formule esatte: la ragione ultima è che il valor medio di s non può essere espresso in forma semplice in termini di grandezze caratteristiche della popolazione.

Per questo motivo è *sempre meglio riferirsi ad errori di varianze* piuttosto che ad errori di scarti quadratici medi; comunque, in prima approssimazione, l'errore di σ si può ricavare da quello su σ^2 usando la formula di propagazione:

$$\text{Var}(\sigma) \approx \left(\frac{1}{\frac{d\sigma^2}{d\sigma}}\right)^2 \text{Var}(\sigma^2) = \frac{1}{4\sigma^2} \text{Var}(\sigma^2) = \frac{\sigma^2}{2(N-1)}$$

cioè

$$\sigma_{\sigma} \approx \frac{\sigma}{\sqrt{2(N-1)}}$$

Il fatto che questa formula sia approssimata risulta chiaramente se si considera che la relazione tra σ^2 e σ è non lineare.

Appendice B

Covarianza e correlazione

B.1 La covarianza

Per due variabili casuali x ed y si definisce la *covarianza*, che si indica con uno dei due simboli $\text{Cov}(x,y)$ o K_{xy} , nel seguente modo:

$$\text{Cov}(x,y) = E\{[x - E(x)][y - E(y)]\} = E(xy) - E(x)E(y)$$

Per provare l'equivalenza delle due forme, basta osservare che¹

$$\begin{aligned}\text{Cov}(x,y) &= E\{[x - E(x)][y - E(y)]\} \\ &= \sum_{ij} P_{ij} [x_i - E(x)] [y_j - E(y)] \\ &= \sum_{ij} P_{ij} x_i y_j - E(x) \sum_{ij} P_{ij} y_j - E(y) \sum_{ij} P_{ij} x_i + \\ &\quad + E(x)E(y) \sum_{ij} P_{ij} \\ &= E(xy) - E(x) \sum_j q_j y_j - E(y) \sum_i p_i x_i + E(x)E(y) \\ &= E(xy) - E(x)E(y)\end{aligned}$$

ricordando alcune relazioni già ricavate nel capitolo 5, e valide per variabili casuali *qualunque*: in particolare, anche *non* statisticamente indipendenti.

E' chiaro come per variabili *statisticamente indipendenti* la covarianza sia nulla: infatti per esse vale la

$$E(xy) = \sum_{ij} P_{ij} x_i y_j = \sum_{ij} p_i q_j x_i y_j = E(x)E(y)$$

¹Nel seguito useremo per la varianza, per le probabilità di ottenere i vari valori x_i o y_j e così via, le stesse notazioni già introdotte nel capitolo 5.

Possiamo ora calcolare la varianza delle combinazioni lineari di due variabili casuali qualunque, estendendo la formula già trovata nel capitolo 5 nel caso particolare di variabili statisticamente indipendenti; partendo ancora da due variabili x e y con media zero per semplificare i calcoli, per la loro combinazione lineare $z = ax + by$ valgono le:

$$\begin{aligned}
 E(z) &= aE(x) + bE(y) = 0 \\
 \text{Cov}(x,y) &= E(xy) - E(x)E(y) = E(xy) \\
 \text{Var}(z) &= E\{[z - E(z)]^2\} \\
 &= E[(ax + by)^2] \\
 &= \sum_{ij} P_{ij} (ax_i + by_j)^2 \\
 &= a^2 \sum_{ij} P_{ij} x_i^2 + b^2 \sum_{ij} P_{ij} y_j^2 + 2ab \sum_{ij} P_{ij} x_i y_j \\
 &= a^2 \sum_i p_i x_i^2 + b^2 \sum_j q_j y_j^2 + 2ab E(xy) \\
 &= a^2 \text{Var}(x) + b^2 \text{Var}(y) + 2ab \text{Cov}(x,y)
 \end{aligned}$$

Questa si estende immediatamente a variabili casuali con media qualsiasi: introducendo ancora le variabili ausiliarie

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

per le quali già sappiamo che valgono le

$$\begin{aligned}
 E(\xi) &= E(\eta) = 0 \\
 \text{Var}(\xi) &= \text{Var}(x) \quad \text{e} \quad \text{Var}(\eta) = \text{Var}(y)
 \end{aligned}$$

Basta osservare infatti che vale anche la

$$\text{Cov}(x,y) = E\{[x - E(x)][y - E(y)]\} = \text{Cov}(\xi, \eta)$$

B.2 Correlazione lineare

Per due variabili casuali qualunque si definisce poi il *coefficiente di correlazione lineare* $\text{Corr}(x,y)$ (o ρ_{xy} , o semplicemente ρ) come

$$\rho_{xy} \equiv \text{Corr}(x,y) = \frac{\text{Cov}(x,y)}{\sqrt{\text{Var}(x)\text{Var}(y)}} = \frac{\text{Cov}(x,y)}{\sigma_x \sigma_y}$$

Il coefficiente di correlazione di due variabili è nullo quando le variabili stesse sono statisticamente indipendenti, ed è comunque compreso tra i due limiti -1 e

+1; quest'ultima proprietà si può dimostrare calcolando la varianza della variabile casuale $z = \sigma_y x - \sigma_x y$ ed osservando che essa deve essere una quantità non negativa:

$$\begin{aligned}\text{Var}(z) &= \sigma_y^2 \text{Var}(x) + \sigma_x^2 \text{Var}(y) - 2\sigma_x \sigma_y \text{Cov}(x, y) \\ &= 2 \text{Var}(x) \text{Var}(y) - 2\sigma_x \sigma_y \text{Cov}(x, y) \\ &\geq 0\end{aligned}$$

da cui

$$\text{Corr}(x, y) \leq 1$$

Compiendo analoghi passaggi sulla $z = \sigma_y x + \sigma_x y$ si trova poi che deve essere anche $\text{Corr}(x, y) \geq -1$.

Se il coefficiente di correlazione lineare raggiunge uno dei valori estremi ± 1 , risulta $\text{Var}(z) = 0$ e dunque deve essere

$$z = \sigma_y x \mp \sigma_x y \equiv E(z) = \text{costante}$$

cioè x ed y devono essere legati da una relazione funzionale di tipo *lineare*.

Vale anche l'inverso: partendo infatti dall'ipotesi che valga la $y = a + bx$, con b finito e non nullo, ne consegue che

$$\begin{aligned}E(y) &= a + bE(x) \\ \text{Var}(y) &= b^2 \text{Var}(x) \\ E(xy) &= E(ax + bx^2) = aE(x) + bE(x^2) \\ \text{Cov}(x, y) &= E(xy) - E(x)E(y) \\ &= aE(x) + bE(x^2) - E(x)[a + bE(x)] \\ &= b \{ E(x^2) - [E(x)]^2 \} \\ &= b \text{Var}(x) \\ \text{Corr}(x, y) &= \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x) \text{Var}(y)}} \\ &= \frac{b \text{Var}(x)}{\sqrt{b^2 [\text{Var}(x)]^2}} \\ &= \frac{b}{|b|} = \pm 1\end{aligned}$$

Il segno del coefficiente di correlazione è quello del coefficiente angolare della retta. Sono da notare due cose: innanzi tutto il rapporto $b/|b|$ perde significato quando $b = 0$ o quando $b = \infty$, cioè quando la retta è parallela ad uno degli assi

coordinati: in questi casi ($x = \text{cost.}$ o $y = \text{cost.}$) una delle due grandezze non è in realtà una variabile casuale, e l'altra è dunque indipendente da essa; è facile vedere che tanto il coefficiente di correlazione tra x e y quanto la covarianza valgono zero, essendo $E(xy) \equiv E(x)E(y)$ in questo caso.

Anche quando esiste una relazione funzionale esatta tra x e y , se questa non è rappresentata da una funzione lineare il coefficiente di correlazione non raggiunge i valori estremi ± 1 ; per questa ragione appunto esso si chiama "coefficiente di correlazione *lineare*".

B.3 Applicazione all'interpolazione lineare

Riprendiamo adesso il problema dell'interpolazione lineare, già discusso nel capitolo 8: si sia cioè compiuto un numero N di misure indipendenti di coppie di valori di due grandezze fisiche x e y , tra le quali si ipotizza che esista una relazione funzionale di tipo lineare data da $y = a + bx$.

Supponiamo inoltre che siano valide le ipotesi esposte nel paragrafo 8.2.1; in particolare che le x_i siano prive di errore, e che le y_i siano affette da errori normali e tutti uguali tra loro.

Sebbene i valori della x siano scelti dallo sperimentatore e privi di errore, e non siano pertanto variabili casuali; e sebbene la variabilità delle y sia dovuta non solo agli errori casuali di misura ma anche alla variazione della x , introduciamo ugualmente in maniera puramente formale le medie e le varianze degli N valori x_i e y_i , date dalle espressioni

$$\bar{x} = \frac{\sum x_i}{N} \quad \text{e} \quad \text{Var}(x) = \frac{\sum (x_i - \bar{x})^2}{N} = \frac{\sum x_i^2}{N} - \bar{x}^2$$

(e simili per la y); e la covarianza di x e y , data dalla

$$\text{Cov}(x,y) = \frac{\sum x_i y_i}{N} - \bar{x}\bar{y}$$

Queste grandezze permettono di riscrivere le equazioni risolutive del problema dell'interpolazione lineare per un insieme di dati, che abbiamo già incontrato nel paragrafo 8.2.1 a pagina 75, nella forma

$$\begin{cases} a + b\bar{x} & = \bar{y} \\ a\bar{x} + b[\text{Var}(x) + \bar{x}^2] & = \text{Cov}(x,y) + \bar{x}\bar{y} \end{cases}$$

Ricavando dalla prima equazione $a = \bar{y} - b\bar{x}$ e sostituendo nella seconda, dopo aver semplificato alcuni termini si ottiene la soluzione

$$b = \frac{\text{Cov}(x,y)}{\text{Var}(x)} \equiv \text{Corr}(x,y) \sqrt{\frac{\text{Var}(y)}{\text{Var}(x)}}$$

La retta interpolante ha quindi equazione

$$y = a + bx = (\bar{y} - b\bar{x}) + bx \quad \text{o anche} \quad (y - \bar{y}) = b(x - \bar{x})$$

in cui b ha il valore su scritto.

Introduciamo ora le due variabili casuali ausiliarie $\xi = x - \bar{x}$ e $\eta = y - \bar{y}$, per cui valgono le

$$\bar{\xi} = 0 \quad \text{e} \quad \text{Var}(\xi) = \text{Var}(x)$$

(ed analoghe per η ed y), e la

$$\text{Cov}(\xi, \eta) = \text{Cov}(x, y)$$

Indichiamo poi con \hat{y}_i il valore della y sulla retta interpolante in corrispondenza dell'ascissa x_i :

$$\hat{y}_i = a + bx_i = \bar{y} + b(x_i - \bar{x})$$

e con δ_i la differenza $\hat{y}_i - y_i$. Risulta che

$$\begin{aligned} \sum \delta_i^2 &= \sum \left\{ [\bar{y} + b(x_i - \bar{x})] - y_i \right\}^2 \\ &= \sum (b\xi_i - \eta_i)^2 \\ &= b^2 \sum \xi_i^2 + \sum \eta_i^2 - 2b \sum \xi_i \eta_i \\ &= Nb^2 \text{Var}(\xi) + N \text{Var}(\eta) - 2Nb \text{Cov}(\xi, \eta) \\ &= Nb^2 \text{Var}(x) + N \text{Var}(y) - 2Nb \text{Cov}(x, y) \\ &= N \text{Var}(x) \left[\frac{\text{Cov}(x, y)}{\text{Var}(x)} \right]^2 + N \text{Var}(y) - 2N \frac{\text{Cov}(x, y)}{\text{Var}(x)} \text{Cov}(x, y) \\ &= N \left\{ \text{Var}(y) - \frac{[\text{Cov}(x, y)]^2}{\text{Var}(x)} \right\} \\ &= N \text{Var}(y) (1 - r^2) \end{aligned}$$

in cui r è il coefficiente di correlazione lineare calcolato usando il campione dei valori misurati di x e y .

Visto che quest'ultimo, nel calcolo dell'interpolazione lineare fatto con le calcolatrici da tasca, viene in genere dato come sottoprodotto dell'algoritmo, la sua conoscenza permette (usando questa formula) di ottenere facilmente l'errore a posteriori sulle ordinate interpolate (studiato nel paragrafo 8.2.2):

$$\mu_y = \sqrt{\frac{\sum \delta_i^2}{N-2}} = \sqrt{\frac{N}{N-2} \text{Var}(y) (1 - r^2)}$$

oltre a fornire una grossolana stima dell'allineamento dei punti; quanto più infatti esso è rigoroso, tanto più r si avvicina a ± 1 .

Il valore μ_y dell'errore rappresenta sempre anche una stima dell'allineamento dei punti, a differenza del coefficiente di correlazione lineare, per cui $r = \pm 1$ implica allineamento perfetto, ma non inversamente: potendo essere ad esempio (punti su di una retta parallela all'asse x) $r = 0$ e $\text{Var}(y) = 0$ ancora con allineamento perfetto.

E' opportuno qui osservare che r non è il coefficiente di correlazione fra gli errori delle grandezze x ed y ; infatti, per ipotesi, la x non è affetta da errore e, se pur lo fosse, la correlazione fra gli errori sarebbe nulla qualora ciascuna x_i fosse misurata indipendentemente dalla corrispondente y_i o, altrimenti, sarebbe tanto più prossima ai valori estremi ± 1 quanto maggiore fosse il numero di cause d'errore comuni alle misure di x e di y .

Invece r è il coefficiente di correlazione per l'insieme dei punti aventi coordinate date dalle coppie di valori misurati, e nell'ipotesi di effettiva dipendenza lineare delle grandezze x ed y sarebbe sempre rigorosamente uguale a ± 1 se non intervenissero gli errori sperimentali.

Appendice C

La distribuzione binomiale

Consideriamo un evento casuale ripetibile E , avente probabilità costante p di verificarsi; indichiamo con $q = 1 - p$ la probabilità del non verificarsi di E (cioè la probabilità dell'evento complementare \bar{E}). Vogliamo ora determinare la probabilità $P(x; N)$ che in N prove ripetute E si verifichi esattamente x volte ($0 \leq x \leq N$).

L'evento casuale costituito dal presentarsi di E per x volte (ed ovviamente dal presentarsi di \bar{E} per le restanti $N - x$) è un evento complesso che può verificarsi in diverse maniere, corrispondenti a tutte le diverse possibili sequenze di successi e fallimenti; queste sono ovviamente mutuamente esclusive, ed in numero pari a quello delle possibili combinazioni di N oggetti a x a x , che vale

$$C_x^N = \frac{N!}{x!(N-x)!}$$

Essendo poi ognuna delle prove statisticamente indipendente dalle altre (infatti la probabilità di E non cambia di prova in prova), ognuna delle possibili sequenze di x successi ed $N - x$ fallimenti ha una probabilità di presentarsi che vale $p^x q^{N-x}$; in definitiva

$$P(x; N) = \frac{N!}{x!(N-x)!} p^x q^{N-x}$$

Questa distribuzione di probabilità $P(x; N)$ per una variabile casuale discreta x si chiama *distribuzione binomiale* o *di Bernoulli*; vogliamo ora determinarne alcune costanti caratteristiche.

Verifichiamo per prima cosa che vale la condizione di normalizzazione: sfrut-

tando la formula per lo sviluppo delle potenze del binomio, risulta

$$\sum_{x=0}^N P(x;N) = \sum_{x=0}^N \frac{N!}{x!(N-x)!} p^x q^{N-x} = (p+q)^N \equiv 1$$

Vogliamo ora calcolare il valor medio della variabile x (ossia il numero medio di successi in N prove): a questo scopo introduciamo innanzi tutto la variabile casuale y_i , che rappresenta il numero di successi nella i -esima delle N prove eseguite. E' chiaro che y_i può assumere i due soli valori 1 (con probabilità p) e 0 (con probabilità q); il suo valor medio è perciò

$$E(y_i) = 1 \cdot p + 0 \cdot q = p$$

e la sua varianza

$$\text{Var}(y_i) = E(y_i^2) - [E(y_i)]^2 = 1 \cdot p + 0 \cdot q - p^2 = p(1-p) = pq$$

osservando che anche y_i^2 può assumere i due soli valori 1 e 0, sempre con le probabilità rispettive p e q . Il numero totale x di successi nelle N prove è legato alle variabili casuali y_i dalla

$$x = \sum_{i=1}^N y_i$$

Risulta quindi

$$E(x) = E\left(\sum_{i=1}^N y_i\right) = \sum_{i=1}^N E(y_i) = Np$$

La varianza della distribuzione binomiale vale poi

$$\text{Var}(x) = \text{Var}\left(\sum_{i=1}^N y_i\right) = \sum_{i=1}^N \text{Var}(y_i) = Npq$$

Come è evidente anche dalla figura, se il numero di prove N è sufficientemente elevato la distribuzione binomiale è bene approssimata da una distribuzione normale avente la stessa media Np e la stessa varianza Npq .

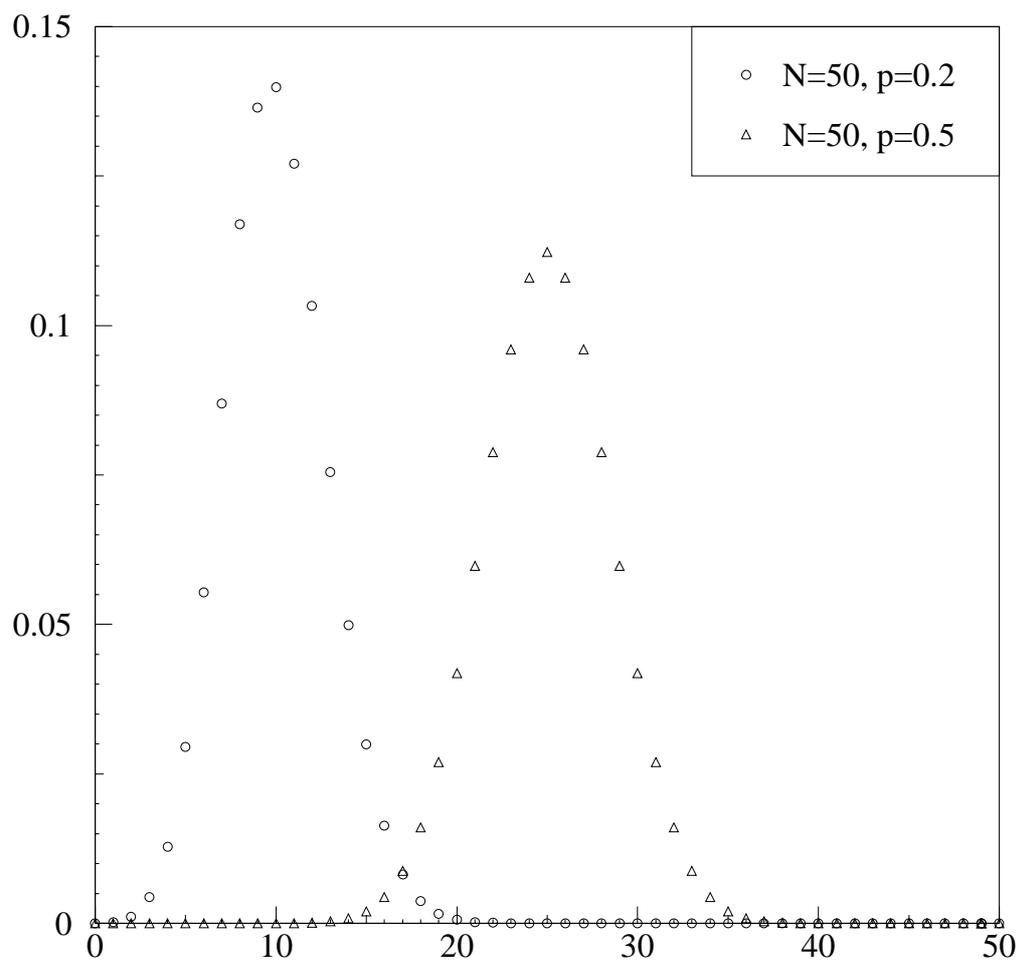


Figura C.1: La distribuzione binomiale, per un numero di prove $N = 50$ e due differenti valori della probabilità p .

Appendice D

La distribuzione di Poisson

Consideriamo il seguente problema: sia E un evento casuale che avvenga rispettando le seguenti ipotesi:

1. La probabilità del verificarsi dell'evento E in un intervallo di tempo¹ molto piccolo (al limite infinitesimo) dt è proporzionale alla durata di tale intervallo.
2. Il verificarsi o meno dell'evento in un certo intervallo temporale è indipendente dal verificarsi o meno dell'evento prima o dopo di esso.
3. La probabilità che più di un evento si verifichi in un tempo infinitesimo dt è infinitesima di ordine superiore rispetto a dt .

Vogliamo ora, sotto queste ipotesi, ricavare la probabilità $P(x, t)$ che in un intervallo di tempo finito, di durata t , si verifichi esattamente un numero prefissato x di eventi E . Usando questa simbologia, la prima ipotesi fatta sul processo casuale in esame si scrive

$$P(1, dt) = \lambda dt \quad \Rightarrow \quad P(0, dt) = 1 - \lambda dt$$

¹Sebbene ci si riferisca, per esemplificare le nostre considerazioni, ad un processo *temporale* (e si faccia poi l'esempio del numero di decadimenti in un intervallo costante di tempo per una sostanza radioattiva come quello di una variabile casuale che segue la distribuzione di Poisson), gli stessi ragionamenti naturalmente si applicano anche a fenomeni fisici riferiti ad intervalli di differente natura, per esempio di spazio.

Così anche il numero di urti per unità di lunghezza delle molecole dei gas segue la distribuzione di Poisson (se si ammette che la probabilità di un urto nel percorrere un intervallo infinitesimo di spazio sia proporzionale alla sua lunghezza, ed analogamente per le altre ipotesi).

e, viste le altre ipotesi ed applicando in conseguenza i teoremi delle probabilità totali e composte, la probabilità di avere x eventi in un intervallo di tempo lungo $t + dt$ è data, a meno di infinitesimi di ordine superiore, da

$$\begin{aligned} P(x, t + dt) &= P(x-1, t)P(1, dt) + P(x, t)P(0, dt) \\ &= P(x-1, t)\lambda dt + P(x, t)(1 - \lambda dt) \end{aligned}$$

cioè

$$\frac{P(x, t + dt) - P(x, t)}{dt} \equiv \frac{d}{dt}P(x, t) = -\lambda P(x, t) + \lambda P(x-1, t)$$

Ora, quando $x = 0$, essendo chiaramente nulla la probabilità di avere un numero negativo di eventi E in un tempo qualsiasi, risulta in particolare

$$\frac{d}{dt}P(0, t) = -\lambda P(0, t)$$

da cui

$$P(0, t) = e^{-\lambda t}$$

(la costante di integrazione si determina imponendo che $P(0, 0) = 1$). Da questa relazione si può ricavare $P(1, t)$ e, con una serie di integrazioni successive, $P(x, t)$: risulta

$$P(x, t) = \frac{(\lambda t)^x}{x!} e^{-\lambda t}$$

In questa espressione x è l'unica variabile casuale, e t funge da parametro: se introduciamo la nuova costante $\alpha = \lambda t$, possiamo scrivere

$$P(x) = \frac{\alpha^x}{x!} e^{-\alpha}$$

Questa distribuzione di probabilità per una variabile casuale (discreta) x prende il nome di *distribuzione di Poisson*; segue questa distribuzione, ad esempio, la variabile casuale che indica il numero di decadimenti in una certa massa di sostanza radioattiva in un dato tempo, processo questo per il quale risultano soddisfatte le tre ipotesi di partenza. Verifichiamo ora la condizione di normalizzazione:

$$\sum_{x=0}^{+\infty} P(x) = e^{-\alpha} \sum_{x=0}^{+\infty} \frac{\alpha^x}{x!} = e^{-\alpha} e^{\alpha} \equiv 1$$

(riconoscendo nella sommatoria l'espressione di uno sviluppo in serie di Mac Laurin della funzione esponenziale). Calcoliamo poi il valor medio di x :

$$\begin{aligned}
 E(x) &= \sum_{x=0}^{+\infty} x \frac{\alpha^x}{x!} e^{-\alpha} \\
 &= \sum_{x=1}^{+\infty} x \frac{\alpha^x}{x!} e^{-\alpha} \\
 &= \alpha e^{-\alpha} \sum_{x=1}^{+\infty} \frac{\alpha^{x-1}}{(x-1)!} \\
 &= \alpha e^{-\alpha} \sum_{y=0}^{+\infty} \frac{\alpha^y}{y!} \\
 &= \alpha e^{-\alpha} e^{\alpha} \\
 &= \alpha
 \end{aligned}$$

Nei passaggi si è prima osservato che il primo termine della sommatoria ($x = 0$) è nullo, e si è poi introdotta la nuova variabile $y = x - 1$. Troviamo ora il valor medio di x^2 : con passaggi analoghi, si ottiene

$$\begin{aligned}
 E(x^2) &= \sum_{x=0}^{+\infty} x^2 \frac{\alpha^x}{x!} e^{-\alpha} \\
 &= \alpha \sum_{x=1}^{+\infty} x \frac{\alpha^{x-1}}{(x-1)!} e^{-\alpha} \\
 &= \alpha \sum_{x=1}^{+\infty} [(x-1) + 1] \frac{\alpha^{x-1}}{(x-1)!} e^{-\alpha} \\
 &= \alpha \left[\sum_{y=0}^{+\infty} y \frac{\alpha^y}{y!} e^{-\alpha} + \sum_{y=0}^{+\infty} \frac{\alpha^y}{y!} e^{-\alpha} \right] \\
 &= \alpha \left[\sum_{y=0}^{+\infty} y P(y) + \sum_{y=0}^{+\infty} P(y) \right] \\
 &= \alpha(\alpha + 1)
 \end{aligned}$$

La varianza di x risulta allora anch'essa data da

$$\text{Var}(x) = E(x^2) - [E(x)]^2 = \alpha(\alpha + 1) - \alpha^2 = \alpha$$

Anche la distribuzione di Poisson è bene approssimata da una distribuzione normale quando α è abbastanza elevato.

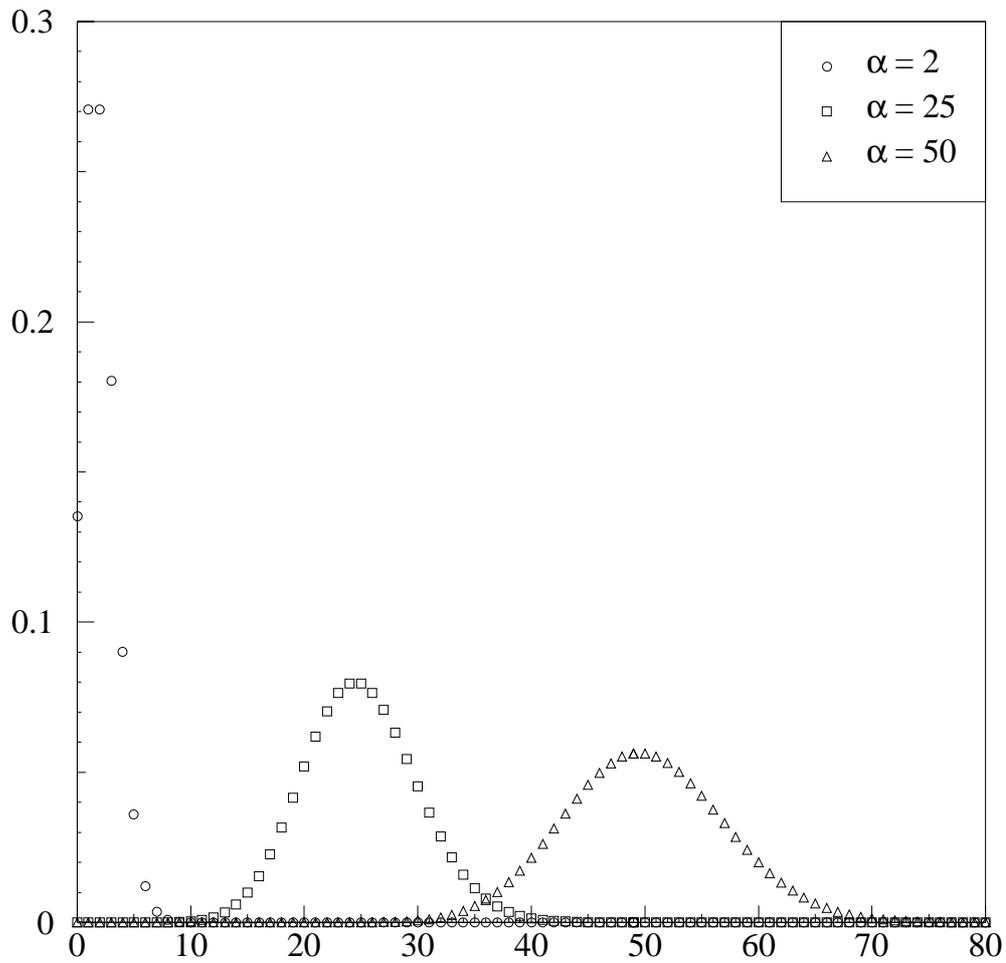


Figura D.1: La distribuzione di Poisson, per tre diversi valori del valor medio α .

Appendice E

Il modello di Laplace e la gaussiana

Pensiamo di eseguire una misura di una grandezza fisica (il cui valor vero indicheremo con il simbolo x^*), e sia x il risultato ottenuto; in generale x è diverso da x^* per la presenza degli errori di misura, che supporremo siano di natura puramente casuale.

Questi errori casuali di misura possono essere schematizzati come un insieme estremamente grande, al limite infinito, di disturbi contemporanei molto piccoli, al limite infinitesimi, ognuno dei quali tende ad alterare di pochissimo il risultato della misura; si considerino in particolare le seguenti ipotesi (*modello semplificato di Laplace per gli errori di misura*):

1. *Ognuna delle cause di disturbo presenti introdurrà nella misura una variazione rispetto al valor vero di modulo fisso ε , con uguale probabilità in difetto o in eccesso.*
2. *Ognuna delle variazioni nella misura dovute a queste cause di disturbo è statisticamente indipendente dalle altre.*

Ognuna delle N cause indipendenti di disturbo produce quindi la variazione $+\varepsilon$ con probabilità $p = \frac{1}{2}$ oppure $-\varepsilon$ con probabilità $q = 1 - p = \frac{1}{2}$; se M tra le N perturbazioni sono positive (e le altre $N - M$ negative), il valore osservato sarà

$$x = x^* + M\varepsilon - (N - M)\varepsilon = x^* + (2M - N)\varepsilon$$

La probabilità di un dato valore di M sulle N prove è data dalla distribuzione

binomiale (vedi appendice C), e vale

$$P(M;N) = \frac{N!}{M!(N-M)!} p^M q^{N-M}$$

Il valor medio di M è dato da Np , e la sua varianza da Npq . Indichiamo poi con il simbolo λ lo scarto di M dal suo valor medio

$$M = Np + \lambda$$

In corrispondenza al variare di M tra 0 ed N , λ varia tra i limiti $-Np$ e $+Nq$; risulta poi anche

$$N - M = N - Np - \lambda = Nq - \lambda$$

e la probabilità di ottenere un certo valore di λ su N prove vale

$$P(\lambda;N) = \frac{N!}{(Np + \lambda)!(Nq - \lambda)!} p^{Np + \lambda} q^{Nq - \lambda}$$

Valor medio e varianza di λ valgono

$$E(\lambda) = E(M) - Np \equiv 0$$

e

$$\text{Var}(\lambda) = \text{Var}(M) = Npq$$

L'andamento generale della probabilità in funzione di M si può trovare considerando il rapporto tra i valori di P che corrispondono a due valori successivi di M :

$$\begin{aligned} \frac{P(M+1;N)}{P(M;N)} &= \frac{N! p^{M+1} q^{N-M-1}}{(M+1)!(N-M-1)!} \frac{M!(N-M)!}{N! p^M q^{N-M}} \\ &= \frac{N-M}{M+1} \frac{p}{q} \end{aligned}$$

$P(M;N)$ risulterà minore, uguale o maggiore di $P(M+1;N)$ a seconda che $(M+1)q$ risulti minore, uguale o maggiore di $(N-M)p$; ossia, essendo $p+q=1$, a seconda che M sia minore, uguale o maggiore di $Np - q$.

Insomma, chiamato μ il più piccolo intero non minore di $Np - q$, la sequenza di valori $P(0;N), P(1;N), \dots, P(\mu;N)$ è crescente, mentre quella dei valori $P(\mu+1;N), P(\mu+2;N), \dots, P(N;N)$ è decrescente. Il massimo valore della probabilità si ha in corrispondenza ad un intero μ che soddisfi la

$$Np - q \leq \mu \leq Np - q + 1 = Np + p$$

e che è unico, salvo il caso che i due estremi dell'intervallo siano entrambi numeri interi: in questo caso si hanno due valori massimi, uguali, in corrispondenza di entrambi. Concludendo: il caso più probabile è che l'evento E si presenti in una sequenza di N prove Np volte, ed il valore di λ con la massima probabilità di presentarsi è 0.

Cerchiamo ora di determinare se esiste e quanto vale il limite della probabilità di ottenere un certo risultato al crescere indefinito del numero delle prove. Per ottenere questo, introduciamo la *formula approssimata di De Moivre e Stirling*¹ per il fattoriale:

$$N! = N^N e^{-N} \sqrt{2\pi N} (1 + \varepsilon_N) \approx \sqrt{2\pi N} N^{N+\frac{1}{2}} e^{-N}$$

con

$$0 \leq \varepsilon_N < \frac{1}{11N}$$

E' lecito trascurare il resto ε_N quando l'argomento del fattoriale è elevato: per $N = 10$ l'errore commesso è già inferiore all'1%. Per usare la formula di De Moivre e Stirling nel nostro caso, sviluppiamo

$$\begin{aligned} (Np + \lambda)! &\approx \sqrt{2\pi} (Np + \lambda)^{Np + \lambda + \frac{1}{2}} e^{-Np - \lambda} \\ &= \sqrt{2\pi} \left(1 + \frac{\lambda}{Np}\right)^{Np + \lambda + \frac{1}{2}} e^{-Np - \lambda} (Np)^{Np + \lambda + \frac{1}{2}} \end{aligned}$$

e, similmente,

$$(Nq - \lambda)! \approx \sqrt{2\pi} \left(1 - \frac{\lambda}{Nq}\right)^{Nq - \lambda + \frac{1}{2}} e^{-Nq + \lambda} (Nq)^{Nq - \lambda + \frac{1}{2}}$$

Queste approssimazioni sono valide quando gli argomenti dei fattoriali, $Np + \lambda$ e $Nq - \lambda$, sono abbastanza grandi: cioè quando λ non è vicino ai valori limite $-Np$ e Nq ; accettata la loro validità (e ritorneremo su questo punto tra poco), sostituendo si ha

$$P(\lambda; N) = \frac{1}{\sqrt{2\pi Npq}} \left(1 + \frac{\lambda}{Np}\right)^{-(Np + \lambda + \frac{1}{2})} \left(1 - \frac{\lambda}{Nq}\right)^{-(Nq - \lambda + \frac{1}{2})}$$

Questa espressione è certamente valida quando $|\lambda|$ non è troppo grande, e per $\lambda = 0$ fornisce la probabilità del valore medio di M ($M = Np$), che risulta

$$P(0; N) = \frac{1}{\sqrt{2\pi Npq}}$$

¹Per la dimostrazione, vedi ad esempio: G. Castelnuovo - Calcolo delle probabilità (Zanichelli), in appendice.

Questa probabilità tende a zero come $1/\sqrt{N}$ al crescere di N ; dato che la somma delle probabilità relative a tutti i casi possibili deve essere 1, si deve concludere che il numero di valori di λ per cui la probabilità non è trascurabile rispetto al suo massimo deve divergere come \sqrt{N} al crescere di N , sebbene il numero di tutti i possibili valori (che è $N+1$) diverga invece come N .

L'espressione approssimata di $P(\lambda; N)$ non è valida per valori di λ prossimi agli estremi $\lambda = -Np$ e $\lambda = Nq$ (è infatti divergente); tuttavia tali valori hanno probabilità infinitesime di presentarsi al crescere di N . Infatti $P(-Np; N) = q^N$ e $P(Nq; N) = p^N$, ed entrambi tendono a zero quando N tende all'infinito essendo sia p che q inferiori all'unità.

Concludendo: la formula approssimata da noi ricavata è valida già per valori relativamente piccoli di N , e per N molto grande si può ritenere esatta per tutti i valori dello scarto λ con probabilità non trascurabile di presentarsi, valori che sono mediamente dell'ordine dell'errore quadratico medio \sqrt{Npq} e che quindi divergono solo come \sqrt{N} . Consideriamo ora il fattore

$$\kappa = \left(1 + \frac{\lambda}{Np}\right)^{-(Np + \lambda + \frac{1}{2})} \left(1 - \frac{\lambda}{Nq}\right)^{-(Nq - \lambda + \frac{1}{2})}$$

che nell'espressione approssimata di $P(\lambda; N)$ moltiplica il valor massimo $P(0; N)$, e se ne prenda il logaritmo:

$$\log \kappa = - \left(Np + \lambda + \frac{1}{2}\right) \log \left(1 + \frac{\lambda}{Np}\right) - \left(Nq - \lambda + \frac{1}{2}\right) \log \left(1 - \frac{\lambda}{Nq}\right)$$

Ora, poichè sia λ/Np che λ/Nq sono in modulo minori dell'unità (salvi i due casi estremi, di probabilità come sappiamo infinitesima), si possono sviluppare i due logaritmi in serie di Mac Laurin:

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

Il primo termine di $\log \kappa$ diventa

$$\begin{aligned} & - \left(Np + \lambda + \frac{1}{2}\right) \left(\frac{\lambda}{Np} - \frac{\lambda^2}{2N^2p^2} + \frac{\lambda^3}{3N^3p^3} - \dots\right) = \\ & = -\lambda + \left(\frac{\lambda^2}{2Np} - \frac{\lambda^2}{Np}\right) - \left(\frac{\lambda^3}{3N^2p^2} - \frac{\lambda^3}{2N^2p^2} + \frac{\lambda}{2Np}\right) + \dots \\ & = -\lambda - \frac{\lambda^2}{2Np} - \frac{\lambda}{2Np} + \frac{\lambda^3}{6N^2p^2} + \dots \end{aligned}$$

ed il secondo

$$\begin{aligned} -\left(Nq - \lambda + \frac{1}{2}\right) \left(-\frac{\lambda}{Nq} - \frac{\lambda^2}{2N^2q^2} - \frac{\lambda^3}{3N^3q^3} - \dots\right) = \\ = \lambda - \frac{\lambda^2}{2Nq} + \frac{\lambda}{2Nq} - \frac{\lambda^3}{6N^2q^2} - \dots \end{aligned}$$

e sommando si ottiene

$$\log \kappa = -\frac{\lambda^2}{2Npq} - \frac{\lambda}{2N} \left(\frac{1}{p} - \frac{1}{q}\right) + \frac{\lambda^3}{6N^2} \left(\frac{1}{p^2} - \frac{1}{q^2}\right) + \dots$$

Da questo sviluppo risulta che il solo termine che si mantiene finito al divergere di N , e per valori di λ dell'ordine di \sqrt{Npq} , è il primo; gli altri due scritti convergono a zero come $1/\sqrt{N}$, e tutti gli altri omessi almeno come $1/N$.

In conclusione, per valori dello scarto per cui la probabilità non è trascurabile (grosso modo $|\lambda| < 3\sqrt{Npq}$), al divergere di N il logaritmo di κ è bene approssimato da

$$\log \kappa \approx -\frac{\lambda^2}{2Npq}$$

e la probabilità dello scarto dalla media λ da

$$P(\lambda) \approx \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{1}{2} \frac{\lambda^2}{Npq}}$$

Per la variabile M sarà invece

$$P(M) \approx \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{1}{2} \frac{(M-Np)^2}{Npq}}$$

Nel caso particolare del modello semplificato di Laplace per gli errori di misura, $p = q = \frac{1}{2}$ e pertanto i termini di ordine $1/\sqrt{N}$ sono identicamente nulli: l'approssimazione è già buona per $N \geq 25$; nel caso generale $p \neq q$, essa è invece accettabile per $Npq \geq 9$. Introducendo lo scarto quadratico medio di M e di λ

$$\sigma = \sqrt{Npq}$$

l'espressione si può scrivere

$$P(\lambda) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{\lambda^2}{\sigma^2}}$$

che è la celebre *legge normale* o di Gauss.

Tornando ancora al modello semplificato di Laplace degli errori di misura, il risultato x ha uno scarto dal valor vero che vale

$$x - x^* = \varepsilon(2M - N) = \varepsilon(2Np + 2\lambda - N) = 2\varepsilon\lambda$$

e possiede varianza

$$\sigma_x^2 \equiv \text{Var}(x - x^*) = 4\varepsilon^2 \text{Var}(\lambda) = 4\varepsilon^2\sigma^2$$

La probabilità di un certo risultato $x = x^* + 2\varepsilon\lambda$ vale infine

$$P(x) = P(\lambda) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\lambda^2}{\sigma^2}} = \frac{2\varepsilon}{\sigma_x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x^*}{\sigma_x}\right)^2}$$

La x è una grandezza discreta che varia per multipli di ε ; nel limite su accennato diventa una variabile continua, e $P(x)$ è infinitesima con ε perdendo così significato; si mantiene invece finita la densità di probabilità, che si ottiene dividendo $P(x)$ per l'ampiezza 2ε dell'intervallo che separa due valori contigui di x :

$$f(x) = \frac{P(x)}{2\varepsilon} = \frac{1}{\sigma_x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x^*}{\sigma_x}\right)^2}$$

ed ha infatti le dimensioni fisiche di $1/\sigma_x$, ovvero di $1/x$.

Al medesimo risultato per $f(x)$ si perverrebbe anche nell'ipotesi più generale che gli errori elementari siano distribuiti comunque, ed anche diversamente l'uno dall'altro, purchè ciascuno abbia una varianza dello stesso ordine di grandezza degli altri ed infinitesima al divergere del numero delle cause di errore.

Appendice F

La verifica delle ipotesi

Una volta eseguita una misura si può voler controllare se i nostri risultati possono confermare o rigettare una determinata ipotesi riguardante il fenomeno fisico che li ha prodotti; naturalmente, visto che risultati di una misura comunque lontani dal valor vero sono sempre possibili (eventualmente con probabilità molto piccole), una qualunque ipotesi sulla grandezza fisica misurata potrà essere confermata o rigettata dai dati solo ad un certo livello di probabilità.

F.1 Compatibilità con un valore prefissato

Per cominciare, si potrebbe voler controllare se un determinato valore numerico a priori attribuibile alla grandezza fisica in esame è o non è confermato dai risultati della misura; cioè se quel valore è o non è *compatibile* con i nostri risultati—più precisamente, a che livello di probabilità (o, per usare la terminologia statistica, a che *livello di confidenza*) è con essi compatibile.

Ammettiamo che gli errori di misura seguano la legge normale; sappiamo che la probabilità per il risultato di cadere in un qualunque intervallo prefissato dell'asse reale si può calcolare integrando la funzione di Gauss fra gli estremi dell'intervallo stesso. Riferiamoci per comodità alla variabile *scarto normalizzato*

$$t = \frac{x - E(x)}{\sigma}$$

che sappiamo già dal paragrafo 6.4 essere distribuita secondo una legge che è indipendente dall'entità degli errori di misura.

Se fissiamo arbitrariamente un numero positivo τ , possiamo calcolare la probabilità che si verifichi l'evento casuale consistente nell'ottenere, in una particolare misura, un valore di t che in modulo superi τ ; come esempio particolare, le condizioni $|t| > 1$ o $|t| > 2$ già sappiamo che si verificano con probabilità rispettivamente del 31.73% e del 4.55%, visto che l'intervallo $-1 < t < 1$ corrisponde al 68.27% dell'area della curva normale, e quello $-2 < t < 2$ al 95.45% .

Se consideriamo poi un campione di N misure, avente valor medio \bar{x} e proveniente da questa stessa popolazione di varianza σ^2 , è immediato capire come la variabile

$$t = \frac{\bar{x} - E(x)}{\frac{\sigma}{\sqrt{N}}}$$

soddisferà a queste stesse condizioni: accadrà cioè nel 31.73% dei casi che $|t|$ sia maggiore di $\tau = 1$, e nel 4.55% dei casi che $|t|$ sia superiore a $\tau = 2$.

Per converso, se fissiamo arbitrariamente un qualunque valore ammissibile P per la probabilità, possiamo calcolare in conseguenza un numero τ , tale che la probabilità di ottenere effettivamente da un particolare campione un valore dello scarto normalizzato t superiore ad esso (in modulo) sia data dal numero P . Ad esempio, fissato un valore del 5% per P , il limite per t che se ne ricava è $\tau = 1.96$: insomma

$$\int_{-1.96}^{+1.96} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} = 0.95$$

e solo nel cinque per cento dei casi si ottiene un valore di t che supera (in modulo) 1.96.

Se si fissa per convenzione un valore della probabilità che indichi il confine tra un avvenimento accettabile ed uno inaccettabile nei limiti della pura casualità, possiamo dire che l'ipotesi consistente nell'essere un certo numero il valor vero della grandezza misurata sarà compatibile o incompatibile con i nostri dati a seconda che lo scarto normalizzato relativo a tale numero sia inferiore o superiore al valore di τ che a quella probabilità corrisponde; e diremo che la compatibilità (o incompatibilità) è riferita a quel certo livello di confidenza prescelto.

La difficoltà è che tutti questi ragionamenti coinvolgono una quantità numerica (lo scarto quadratico medio) relativa *alla popolazione* e per ciò stesso in generale ignota; in tal caso, per calcolare lo scarto normalizzato relativo ad un certo valore numerico ξ non possiamo che servirci, in luogo di σ , della corrispondente *stima* ricavata dal campione, s :

$$t = \frac{\xi - \bar{x}}{\frac{s}{\sqrt{N}}}$$

e quindi si deve presupporre di avere un campione di dimensioni tali che questa stima si possa ritenere ragionevole, ossia sufficientemente vicina ai corrispondenti

valori relativi alla popolazione a meno di fluttuazioni casuali abbastanza poco probabili.

In generale si ammette che almeno 30 dati siano necessari perchè questo avvenga; in corrispondenza a tale dimensione del campione l'errore della media è circa 5.5 volte inferiore a quello dei dati, e l'errore relativo di s è approssimativamente del 13%.

Bisogna anche porre attenzione alla esatta natura dell'ipotesi che si intende verificare. Per un valore limite di $\tau = 1.96$ abbiamo visto che il 95% dell'area della curva normale è compreso tra $-\tau$ e $+\tau$: superiormente a $+\tau$ si trova il 2.5% di tale area; ed anche inferiormente a $-\tau$ se ne trova un'altra porzione pari al 2.5%.

Se l'ipotesi da verificare riguarda l'essere differenti tra loro due entità (il presupposto valor vero della grandezza misurata e la media aritmetica dei nostri dati, nell'esempio precedente) quel valore di τ corrisponde in effetti ad una verifica relativa ad un livello di confidenza del 5% (*two-tailed test*); ma se l'ipotesi riguarda l'essere un valore numerico superiore (od inferiore) alla nostra media aritmetica (ad esempio, i dati misurati potrebbero essere relativi al rendimento di una macchina, e si vuole verificare l'ipotesi che tale rendimento misurato sia superiore ad un valore prefissato), allora un limite $\tau = 1.96$ corrisponde in effetti ad un livello di confidenza del 2.5% (*one-tailed test*): nell'esempio fatto, soltanto l'intervallo $[-\infty, -\tau]$ deve essere preso in considerazione per il calcolo della probabilità.

Alcuni limiti relativi a diversi livelli di confidenza si possono trovare nelle tabelle F.1 e F.2.

P (%)	τ_2	τ_1
10.0	1.64485	1.28155
5.0	1.95996	1.64485
2.0	2.32635	2.05375
1.0	2.57583	2.32635
0.5	2.81297	2.57583
0.2	3.09023	2.87816
0.1	3.29053	3.09023

Tabella F.1: Alcuni valori della probabilità P e dei corrispondenti limiti τ sullo scarto normalizzato, per verifiche two-tailed (τ_2) o one-tailed (τ_1).

τ	P_2 (%)	P_1 (%)
0.5	61.708	30.854
1.0	31.731	15.866
1.5	13.361	6.681
2.0	4.550	2.275
2.5	1.242	0.621
3.0	0.270	0.135

Tabella F.2: I valori della probabilità per verifiche two-tailed (P_2) ed one-tailed (P_1) che corrispondono a valori prefissati dello scarto normalizzato τ .

F.2 La compatibilità di due valori misurati

Un altro caso frequente è quello in cui si hanno a disposizione due campioni di misure, e si vuole verificare l'ipotesi statistica che essi provengano da popolazioni aventi lo stesso valor medio; un caso particolare è quello dell'ipotesi consistente nell'essere i due campioni composti da misure della stessa grandezza fisica, che hanno prodotto differenti stime come effetto della presenza in entrambi degli errori, che assumiamo ancora seguire la legge normale.

Siano ad esempio un primo campione di N misure x_i , ed un secondo campione di M misure y_j ; indichiamo con \bar{x} e \bar{y} le medie dei due campioni, con σ_x e σ_y le varianze delle popolazioni da cui tali campioni provengono, e con $\delta = \bar{x} - \bar{y}$ la differenza tra le due medie.

Sappiamo già che i valori medi e le varianze delle medie dei campioni sono legati ai corrispondenti valori relativi alle popolazioni dalle

$$E(\bar{x}) = E(x) \quad , \quad E(\bar{y}) = E(y)$$

e

$$\text{Var}(\bar{x}) = \frac{\sigma_x^2}{N} \quad , \quad \text{Var}(\bar{y}) = \frac{\sigma_y^2}{M}$$

per cui risulterà, se i campioni sono tra loro statisticamente indipendenti e se si ammette valida l'ipotesi (da verificare) che abbiano la stessa media,

$$E(\delta) = E(\bar{x} - \bar{y}) = E(x) - E(y) = 0$$

e

$$\text{Var}(\delta) = \text{Var}(\bar{x} - \bar{y}) = \frac{\sigma_x^2}{N} + \frac{\sigma_y^2}{M}$$

Inoltre, essendo \bar{x} , \bar{y} (e quindi δ) combinazioni lineari di variabili normali, seguiranno anch'esse la legge normale; e la verifica dell'ipotesi che i campioni pro-

vengano da popolazioni aventi la stessa media si traduce nella verifica dell'ipotesi che δ abbia valor vero nullo.

Tale verifica, essendo δ distribuita secondo la legge normale, si esegue come abbiamo visto nel paragrafo precedente: si fissa arbitrariamente un valore del livello di confidenza, si determina il corrispondente valore limite degli scarti normalizzati, e lo si confronta con il valore di

$$\frac{\delta - E(\delta)}{\sigma_\delta} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{N} + \frac{\sigma_y^2}{M}}}$$

Ovviamente vale anche qui l'osservazione fatta nel paragrafo precedente: non conoscendo le varianze delle popolazioni σ_x e σ_y , siamo costretti ad usare in loro vece le stime ottenute dai campioni, s_x ed s_y ; e questo si ammette generalmente lecito quando la dimensione di entrambi i campioni è almeno pari a 30.

F.3 I piccoli campioni e la distribuzione di Student

Solo per completezza, accenniamo alla verifica di ipotesi statistiche mediante piccoli campioni; ovvero campioni costituiti da un numero di dati così esiguo da farci ritenere che non si possa ottenere da essi con ragionevole probabilità una buona stima delle varianze delle rispettive popolazioni (sempre però supposte normali).

Indicando con \bar{x} la media aritmetica di un campione di dimensione N estratto a caso da una popolazione normale, e con s la stima della varianza della popolazione ottenuta dal campione stesso, cioè

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

e definita la variabile

$$t = \frac{\bar{x} - E(x)}{\frac{s}{\sqrt{N}}}$$

si può dimostrare che la funzione densità di probabilità relativa alla variabile casuale t è data dalla

$$f(t; N) = \frac{T_N}{\left(1 + \frac{t^2}{N-1}\right)^{\frac{N}{2}}}$$

(*distribuzione di Student*).

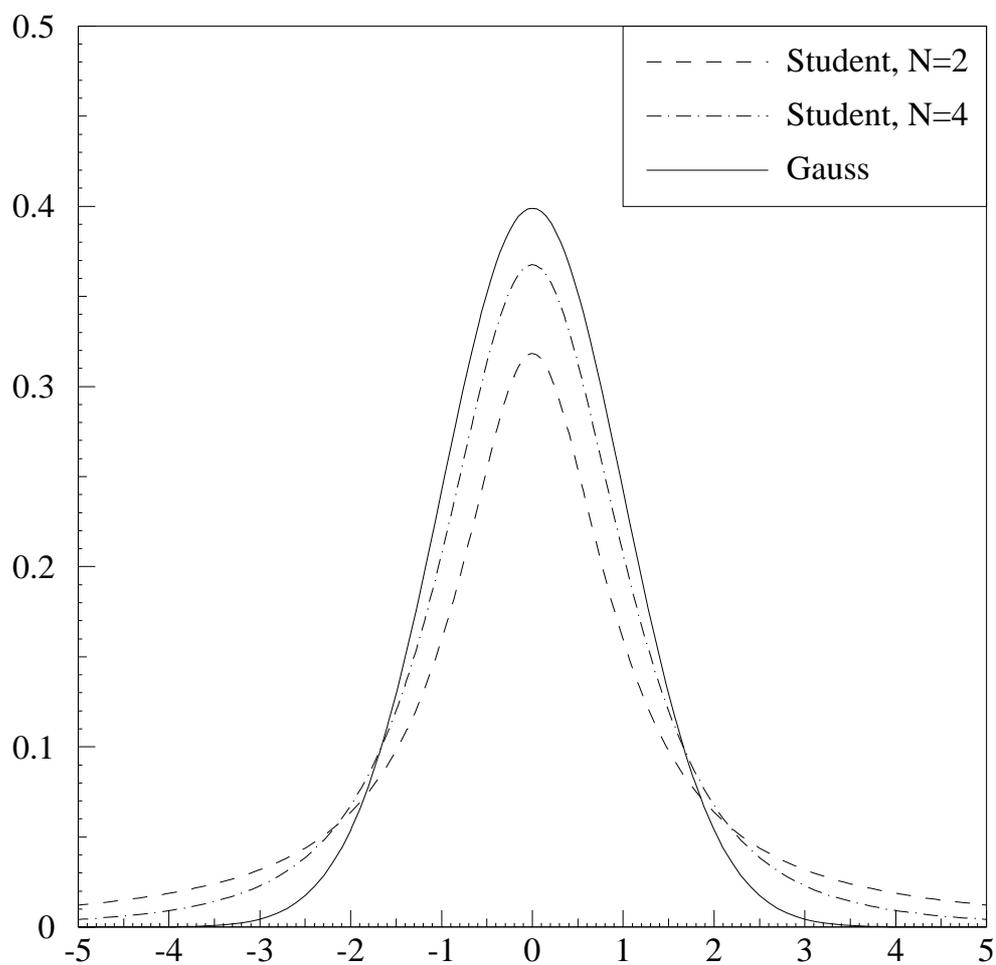


Figura F.1: La distribuzione di Student per $N = 2$ ed $N = 4$, confrontata con la funzione normale.

Il coefficiente T_N è una costante che viene fissata dalla condizione di normalizzazione; se N viene poi fatto tendere all'infinito il denominatore della funzione tende a $e^{t^2/2}$, e dunque la distribuzione di Student tende alla distribuzione normale (con media 0 e varianza 1). Anche la forma della funzione ricorda molto quella della Gaussiana, come appare evidente dalla figura F.1.

Se i campioni a disposizione non hanno dimensioni accettabili, una volta calcolato lo scarto normalizzato relativo alla differenza tra la media di un campione ed un valore prefissato, o tra le medie di due campioni (così come indicato nei due paragrafi precedenti), occorrerà confrontare il suo valore con i limiti degli intervalli di confidenza relativi alla distribuzione di Student e non alla distribuzione normale: limiti che si trovano tabulati in vari testi di statistica¹.

F.4 La distribuzione del χ^2

Se le N variabili casuali x_i , tra loro statisticamente indipendenti, sono distribuite secondo la legge normale con media 0 e varianza 1, si può dimostrare che la nuova variabile casuale

$$X = \sum_{i=1}^N x_i^2$$

(ovviamente non negativa) è distribuita con una densità di probabilità data dalla

$$\frac{dp}{dX} = f(X;N) = K X^{\left(\frac{N}{2}-1\right)} e^{-\frac{X}{2}}$$

(*distribuzione del chi-quadro*); la costante K viene fissata dalla condizione di normalizzazione, ed il parametro N prende il nome di *numero di gradi di libertà* della distribuzione.

Il valor medio e la varianza di una variabile casuale distribuita come il χ^2 a N gradi di libertà sono, come si potrebbe trovare integrando le opportune espressioni,

$$E(X) = N \qquad \text{e} \qquad \text{Var}(X) = 2N$$

Inoltre si può dimostrare che, quando N assume valori sufficientemente grandi, la distribuzione del χ^2 è ben approssimata da una distribuzione normale avente la stessa media N e la stessa varianza $2N$; tale approssimazione si può ritenere in pratica già buona quando N è superiore a 30.

Ovviamente, se le x_i sono variabili casuali indipendenti e provenienti da una stessa distribuzione normale con media μ e varianza σ^2 , la nuova variabile casuale

$$X = \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^2$$

¹Ad esempio in una delle appendici di: Spiegel - Probabilità e statistica - Collana Schaum, Etas-Kompass editrice.

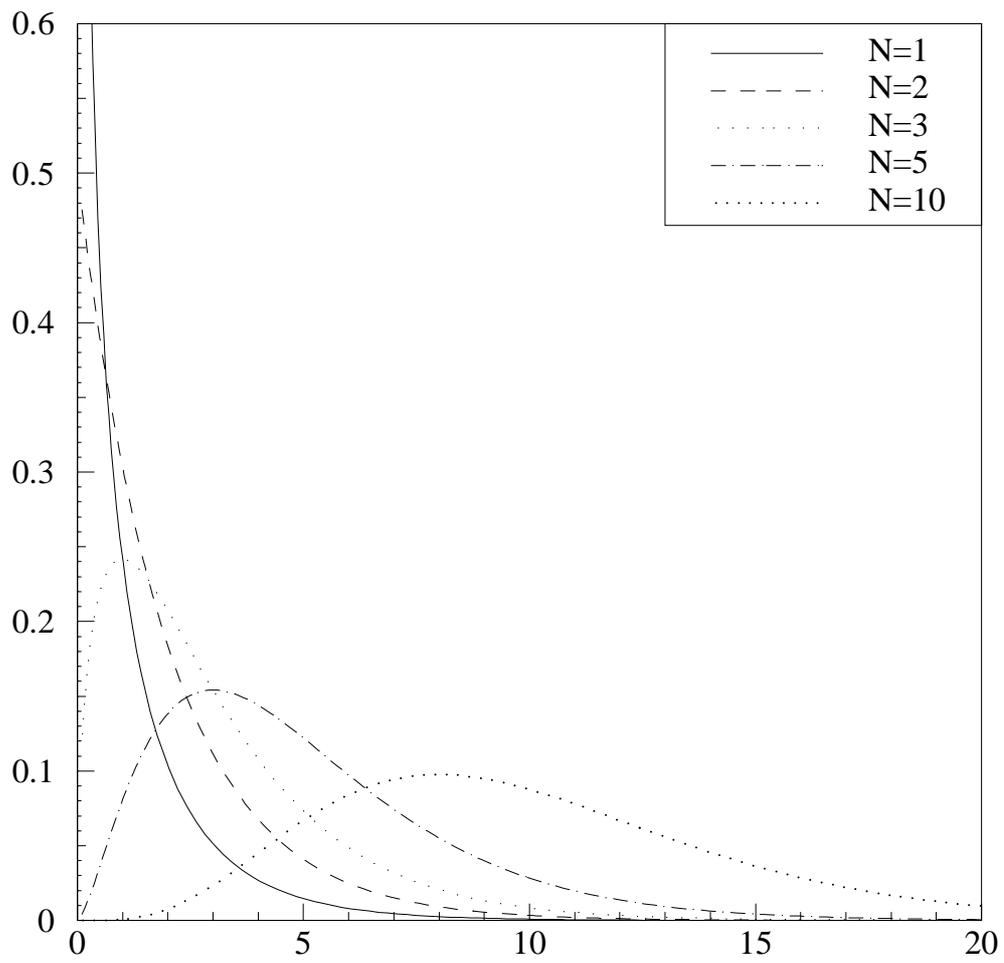


Figura F.2: La distribuzione del χ^2 per alcuni valori del parametro N .

è distribuita come il χ^2 a N gradi di libertà.

E' immediato poi, sulla base della definizione, capire che, se X ed Y sono due variabili casuali indipendenti distribuite come il χ^2 con N ed M gradi di libertà rispettivamente, la loro somma $Z = X + Y$ è una variabile casuale ancora distribuita come il χ^2 , però con $N + M$ gradi di libertà (*regola di somma del χ^2*). Si potrebbe analogamente dimostrare (sempre a riguardo della somma $Z = X + Y$ di due variabili casuali indipendenti), che se Z ed X seguono la distribuzione del χ^2 , ad $N + M$ ed N gradi di libertà rispettivamente, la terza variabile Y deve essere distribuita come il χ^2 ad M gradi di libertà.

Ora, supponiamo che le x_i (con $i = 1, \dots, N$) siano variabili casuali indipendenti distribuite secondo la legge normale con media 0 e varianza 1, ed indichiamo al solito con \bar{x} la loro media aritmetica; vogliamo dimostrare ora che la variabile casuale

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

è distribuita come il χ^2 con $N - 1$ gradi di libertà. A questo scopo introduciamo la variabile ausiliaria

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

che ha, come si può facilmente calcolare applicando le formule già ricavate nel capitolo 5, anch'essa distribuzione normale con media 0 e varianza 1: risulta

$$\begin{aligned} \sum_{i=1}^N (x_i - \bar{x})^2 &= \sum_{i=1}^N x_i^2 - N\bar{x}^2 \\ &= \sum_{i=1}^N x_i^2 - Y^2 \end{aligned}$$

per cui

$$\sum_{i=1}^N x_i^2 = \sum_{i=1}^N (x_i - \bar{x})^2 + Y^2$$

Ora, nell'espressione precedente, il primo termine è distribuito come il χ^2 a N gradi di libertà, mentre Y^2 è distribuito come il χ^2 ad un grado di libertà; si potrebbe dimostrare che X ed Y sono tra loro statisticamente indipendenti e da questo consegue, tenendo conto di quanto prima affermato, la validità della nostra tesi.

In generale, se le x_i provengono tutte da un'unica popolazione normale avente varianza σ^2 , la variabile casuale

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

è ancora distribuita come il χ^2 a $N - 1$ gradi di libertà.

E' interessante confrontare questa formula con quella precedentemente ricavata, e riguardante la stessa espressione in cui però gli scarti erano calcolati rispetto alla media della popolazione μ ; nel primo caso la distribuzione era ancora quella del χ^2 , ma con N gradi di libertà: riferendoci alla media del campione i gradi di libertà diminuiscono di una unità.

Questo è conseguenza di una legge generale, secondo la quale il numero di gradi di libertà da associare a variabili che seguono la distribuzione del χ^2 è dato dal numero di termini sommati (qui N , uno per ogni determinazione indipendente x_i) diminuito del numero di parametri che compaiono nella formula e che sono stimati dai dati stessi (qui uno: appunto la media della popolazione, stimata usando la media aritmetica delle misure).

F.5 Compatibilità dei dati con una distribuzione

Supponiamo di avere dei dati raccolti in un istogramma, e di voler verificare l'ipotesi che i dati provengano da una certa distribuzione; ad esempio, la distribuzione normale.

Ora per una misura la probabilità p_i di cadere nell'intervallo i -esimo di ampiezza prefissata Δx e corrispondente alla generica classe di frequenza usata per la realizzazione dell'istogramma, è data dal valor medio della funzione densità di probabilità nell'intervallo stesso, moltiplicato per Δx .

Il numero di misure effettivamente ottenute in una classe di frequenza su N prove obbedisce alla distribuzione binomiale: il loro valor medio è quindi Np_i , e la loro varianza $Np_i(1 - p_i)$; quest'ultimo termine si può approssimare ancora con Np_i se si ammette che le classi di frequenza siano sufficientemente ristrette da poter trascurare i termini in p_i^2 rispetto a quelli in p_i (cioè se $p_i \ll 1$).

In questo caso il numero di misure in ciascuna classe segue approssimativamente la distribuzione di Poisson; questa è infatti la funzione che mostra come si presentano, su un grande numero di osservazioni, eventi aventi piccola probabilità di verificarsi singolarmente in ognuna: distribuzione nella quale l'errore quadratico medio è effettivamente dato dalla radice quadrata del valor medio: $\sigma = \sqrt{Np_i(1 - p_i)} \simeq \sqrt{Np_i}$.

Nei limiti in cui il numero di misure attese in una classe è sufficientemente elevato da poter confondere la distribuzione binomiale con una distribuzione normale, la quantità

$$X \simeq \sum_{i=1}^N \frac{(n_i - Np_i)^2}{Np_i} = \sum_{i=1}^N \frac{(O_i - A_i)^2}{A_i}$$

cioè la somma, su tutte le classi di frequenza, del quadrato della differenza tra il numero di misure ivi *attese* ($A_i = Np_i$) ed ivi *effettivamente osservate* ($O_i = n_i$), diviso per la varianza del numero di misure attese (approssimata da $Np_i = A_i$), ha la distribuzione del χ^2 con $N - 1$ gradi di libertà; il motivo per quest'ultima affermazione è che esiste un vincolo sulle O_i , quello di avere per somma il numero totale di misure effettuate N (che viene usato nella formula che definisce X per calcolare il numero di misure attese in ogni intervallo, A_i).

Questo si può in pratica supporre verificato se A_i in ogni intervallo è almeno pari a 5; o meglio, se il numero di classi di frequenza in cui ci si aspetta un numero di misure minore di 5 è trascurabile rispetto al totale (meno del 20%).

Quindi, per la verifica dell'ipotesi statistica che i dati vengano dalla distribuzione usata per il calcolo delle A_i , basta valutare X ; e poi verificare che non superi il valore di taglio corrispondente alla coda superiore della distribuzione del χ^2 ad $N - 1$ gradi di libertà, coda avente area pari al livello di confidenza desiderato: valore rilevabile dalle apposite tabelle².

²Nel già citato Spiegel - Probabilità e statistica - Collana Schaum, Etas-Kompass editrice, sono presenti anche queste tabelle. E' bene anche ricordare che quando N è superiore a 30 si può far riferimento alla distribuzione normale con media N ed errore quadratico medio $\sqrt{2N}$.

Appendice G

La funzione di verosimiglianza

Si supponga di aver compiuto N osservazioni indipendenti relative ad una grandezza fisica x , e di aver trovato i valori x_i , con $i = 1, 2, \dots, N$. Ciascuna delle variabili casuali x_i abbia poi densità di probabilità $f_i(x_i|\theta)$, funzione nota che dipende dal parametro θ di valore vero θ^* ignoto, definita in un intervallo dell'asse reale delle x_i con estremi indipendenti da θ (che potremo assumere essere $\pm\infty$ ponendo eventualmente $f_i(x_i|\theta) \equiv 0$ esternamente all'intervallo di definizione).

Una stima di una generica funzione nota del parametro, $\tau(\theta)$ (che supporremo con derivata non nulla), è una funzione dei soli valori osservati $t(x_1, x_2, \dots, x_N)$; dunque a sua volta una variabile casuale con associata una funzione densità di probabilità che indicheremo con $g(t|\theta)$. La stima si dice *imparziale* (o *indistorta*) quando il suo valor medio

$$\begin{aligned} E(t) &= \int_{-\infty}^{+\infty} t g(t|\theta) dt \\ &= \int_{-\infty}^{+\infty} dx_1 f_1(x_1|\theta) \cdots \int_{-\infty}^{+\infty} dx_N f_N(x_N|\theta) t(x_1, x_2, \dots, x_N) \end{aligned}$$

è uguale al rispettivo valor vero:

$$E(t) = \tau(\theta)$$

Una importante proprietà della stima t è la sua varianza

$$\begin{aligned} \sigma_t^2 &= \int_{-\infty}^{+\infty} [t - \tau(\theta)]^2 g(t|\theta) dt \\ &= \int_{-\infty}^{+\infty} dx_1 f_1(x_1|\theta) \cdots \int_{-\infty}^{+\infty} dx_N f_N(x_N|\theta) [t(x_1, x_2, \dots, x_N) - \tau(\theta)]^2 \end{aligned}$$

perchè *la minima varianza* sarà il criterio di scelta fra stime diverse di $\tau(\theta)$.

Il teorema che segue (**teorema di Cramér-Rao**) mostra che esiste un limite inferiore per la varianza di una stima. Osserviamo per prima cosa che la densità di probabilità per la N -pla (x_1, x_2, \dots, x_N) risulta

$$\prod_{i=1}^N f_i(x_i|\theta^*)$$

per il teorema della probabilità composta; se in luogo del valor vero θ^* si pone il parametro variabile θ , si ottiene la *funzione di verosimiglianza*

$$\mathcal{L}(x_1, x_2, \dots, x_N|\theta) = \prod_{i=1}^N f_i(x_i|\theta)$$

La condizione di normalizzazione di ciascuna f_i comporta che l'integrale della verosimiglianza su tutti i domini delle variabili x_i valga 1:

$$\begin{aligned} \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \cdots \int_{-\infty}^{+\infty} dx_N \mathcal{L}(x_1, x_2, \dots, x_N|\theta) &= \\ &= \int_{-\infty}^{+\infty} dx_1 f_1(x_1|\theta) \int_{-\infty}^{+\infty} dx_2 f_2(x_2|\theta) \cdots \int_{-\infty}^{+\infty} dx_N f_N(x_N|\theta) \\ &= \prod_{i=1}^N \int_{-\infty}^{+\infty} dx_i f_i(x_i|\theta) \\ &\equiv 1 \end{aligned}$$

indipendentemente dal valore di θ .

Derivando sotto il segno di integrale rispetto a θ , dato che i domini delle $f_i(x_i|\theta)$ non dipendono da θ , si ottiene

$$\int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \cdots \int_{-\infty}^{+\infty} dx_N \frac{\partial \mathcal{L}}{\partial \theta} = 0$$

da cui, dividendo e moltiplicando l'integrando per \mathcal{L} , risulta

$$\begin{aligned} \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \cdots \int_{-\infty}^{+\infty} dx_N \mathcal{L} \left(\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta} \right) &= \\ &= \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \cdots \int_{-\infty}^{+\infty} dx_N \mathcal{L} \frac{\partial \log \mathcal{L}}{\partial \theta} \\ &= \int_{-\infty}^{+\infty} dx_1 f_1(x_1|\theta) \cdots \int_{-\infty}^{+\infty} dx_N f_N(x_N|\theta) \frac{\partial \log \mathcal{L}}{\partial \theta} \\ &= 0 \end{aligned}$$

ossia

$$E\left(\frac{\partial \log \mathcal{L}}{\partial \theta}\right) = 0$$

Se t è imparziale,

$$E(t) = \int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_N t(x_1, x_2, \dots, x_N) \mathcal{L}(x_1, x_2, \dots, x_N | \theta) = \tau(\theta)$$

da cui, derivando ambo i membri rispetto a θ ,

$$\int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \cdots \int_{-\infty}^{+\infty} dx_N t \frac{\partial \mathcal{L}}{\partial \theta} = \tau'(\theta)$$

Dividendo e moltiplicando poi l'integrando per la verosimiglianza \mathcal{L} , risulta

$$\begin{aligned} \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \cdots \int_{-\infty}^{+\infty} dx_N t \frac{\partial \mathcal{L}}{\partial \theta} &= \\ &= \int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_N t \mathcal{L} \left(\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta} \right) \\ &= \int_{-\infty}^{+\infty} dx_1 f_1(x_1 | \theta) \cdots \int_{-\infty}^{+\infty} dx_N f_N(x_N | \theta) t \frac{\partial \log \mathcal{L}}{\partial \theta} \\ &= E\left(t \frac{\partial \log \mathcal{L}}{\partial \theta}\right) \end{aligned}$$

e, in definitiva,

$$E\left(t \frac{\partial \log \mathcal{L}}{\partial \theta}\right) = \tau'(\theta)$$

Infine, sommando membro a membro questa equazione e la precedente moltiplicata per $\tau(\theta)$, si ottiene

$$E\left(t \frac{\partial \log \mathcal{L}}{\partial \theta}\right) - \tau(\theta) E\left(\frac{\partial \log \mathcal{L}}{\partial \theta}\right) = \tau'(\theta)$$

ovvero

$$E\left\{[t - \tau(\theta)] \frac{\partial \log \mathcal{L}}{\partial \theta}\right\} = \tau'(\theta)$$

Se ora si definiscono il rapporto

$$R(\theta) = \frac{E\left\{[t - \tau(\theta)] \frac{\partial \log \mathcal{L}}{\partial \theta}\right\}}{E\left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta}\right)^2\right]} = \frac{\tau'(\theta)}{E\left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta}\right)^2\right]}$$

(che è una costante dipendente da θ ; osserviamo anche che deve risultare $R(\theta) \neq 0$) e la variabile casuale

$$z = [t - \tau(\theta)] - R(\theta) \frac{\partial \log \mathcal{L}}{\partial \theta}$$

il cui quadrato risulta essere

$$z^2 = [t - \tau(\theta)]^2 - 2R(\theta)[t - \tau(\theta)] \frac{\partial \log \mathcal{L}}{\partial \theta} + R^2(\theta) \left(\frac{\partial \log \mathcal{L}}{\partial \theta} \right)^2$$

prendendo il valor medio di z^2 si ottiene

$$\begin{aligned} E(z^2) &= E \left\{ [t - \tau(\theta)]^2 \right\} - 2R(\theta) E \left\{ [t - \tau(\theta)] \frac{\partial \log \mathcal{L}}{\partial \theta} \right\} + \\ &\quad + R^2(\theta) E \left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta} \right)^2 \right] \\ &= \sigma_t^2 - 2 \frac{\tau'(\theta)}{E \left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta} \right)^2 \right]} \tau'(\theta) + \\ &\quad + \left\{ \frac{\tau'(\theta)}{E \left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta} \right)^2 \right]} \right\}^2 E \left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta} \right)^2 \right] \\ &= \sigma_t^2 - 2 \frac{[\tau'(\theta)]^2}{E \left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta} \right)^2 \right]} + \frac{[\tau'(\theta)]^2}{E \left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta} \right)^2 \right]} \\ &= \sigma_t^2 - \frac{[\tau'(\theta)]^2}{E \left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta} \right)^2 \right]} \end{aligned}$$

Ma il valor medio del quadrato di una qualsiasi variabile casuale non può essere negativo, e dunque

$$0 \leq E(z^2) = \sigma_t^2 - \frac{[\tau'(\theta)]^2}{E \left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta} \right)^2 \right]}$$

ed infine

$$\sigma_t^2 \geq \frac{[\tau'(\theta)]^2}{E \left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta} \right)^2 \right]} = [\tau'(\theta)]^2 \frac{R(\theta)}{\tau'(\theta)} = \tau'(\theta) R(\theta)$$

cioè:

Nessuna funzione dei valori osservati $t(x_1, x_2, \dots, x_N)$, che sia stima imparziale di una funzione del parametro $\tau(\theta)$, può avere varianza inferiore ad un limite determinato.

La varianza minima si raggiunge se e soltanto se $E(z^2)$ è nullo, il che è possibile solo se z è nulla ovunque, cioè se

$$z = t - \tau(\theta) - R(\theta) \frac{\partial \log \mathcal{L}}{\partial \theta} \equiv 0$$

o, altrimenti detto, se la derivata logaritmica della verosimiglianza è proporzionale alla variabile casuale $t - \tau(\theta)$:

$$\boxed{\frac{\partial \log \mathcal{L}}{\partial \theta} = \frac{t - \tau(\theta)}{R(\theta)}}$$

Col metodo della massima verosimiglianza si assume, come stima del valor vero θ^* del parametro θ , quel valore $\hat{\theta}$ che rende massima la verosimiglianza \mathcal{L} per i valori osservati delle variabili, x_1, x_2, \dots, x_N .

Ora, nel caso esista una stima di minima varianza t per la funzione $\tau(\theta)$, per quanto visto sopra l'equazione di massima verosimiglianza diviene

$$\frac{\partial \log \mathcal{L}}{\partial \theta} = \frac{t - \tau(\theta)}{R(\theta)} = 0$$

e le soluzioni $\hat{\theta}$ sono tutte e sole quelle dell'equazione

$$\tau(\theta) = t(x_1, x_2, \dots, x_N)$$

La derivata seconda di $\log \mathcal{L}$ è in tal caso

$$\begin{aligned} \frac{\partial^2 \log \mathcal{L}}{\partial \theta^2} &= - \frac{\tau'(\theta) R(\theta) + R'(\theta) [t - \tau(\theta)]}{R^2(\theta)} \\ &= - \frac{\sigma_t^2 + R'(\theta) [t - \tau(\theta)]}{R^2(\theta)} \end{aligned}$$

ma se $\theta = \hat{\theta}$ è anche $t - \tau(\hat{\theta}) = 0$ e risulta

$$\left(\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2} \right)_{\theta=\hat{\theta}} = - \frac{\sigma_t^2}{R^2(\hat{\theta})} < 0$$

cioè per tutte le soluzioni $\theta = \hat{\theta}$ la verosimiglianza è massima.

Ora, se la funzione $\log \mathcal{L}$ è regolare, tra due massimi deve esistere un minimo; dato che non esistono minimi, ne consegue che il massimo è unico ed in corrispondenza al valore della funzione τ^{-1} inversa di $\tau(\theta)$ e calcolata in $t(x_1, x_2, \dots, x_N)$:

$$\hat{\theta} = \tau^{-1} [t(x_1, x_2, \dots, x_N)]$$

Si ponga attenzione al fatto che per N finito questa non è una stima di θ con varianza minima, che esiste solo per la funzione $\tau(\theta)$ ed è appunto la $t(x_1, x_2, \dots, x_N)$.

La statistica $t(x_1, x_2, \dots, x_N)$ (come viene anche indicata una funzione dei dati) di minima varianza è un caso particolare di statistica *sufficiente* per il parametro θ , come è chiamata una funzione dei valori osservati, se esiste, che riassume in sé tutta l'informazione che i dati possono fornire sul valore del parametro.

Se x_1, x_2, \dots, x_N sono i valori osservati di N variabili casuali *normali* con lo stesso valor medio λ e varianze rispettive σ_i supposte note, la verosimiglianza è

$$\mathcal{L} = \prod_{i=1}^N \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x_i - \lambda}{\sigma_i} \right)^2}$$

il suo logaritmo

$$\log \mathcal{L} = -\frac{1}{2} \sum_{i=1}^N \frac{(x_i - \lambda)^2}{\sigma_i^2} - \sum_{i=1}^N \log (\sigma_i \sqrt{2\pi})$$

e la sua derivata rispetto al parametro λ

$$\frac{\partial \log \mathcal{L}}{\partial \lambda} = \sum_{i=1}^N \frac{x_i - \lambda}{\sigma_i^2} = \left(\sum_{i=1}^N \frac{1}{\sigma_i^2} \right) \left(\frac{\sum_{i=1}^N \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} - \lambda \right)$$

Pertanto la media dei dati, pesati con coefficienti inversamente proporzionali alle varianze, è una stima di minima varianza per λ . Se le N varianze sono poi tutte uguali tra loro e di valore σ^2 , risulta

$$\frac{\partial \log \mathcal{L}}{\partial \lambda} = \frac{1}{\sigma^2} \left[\left(\sum_{i=1}^N x_i \right) - N\lambda \right] = \frac{N}{\sigma^2} (\bar{x} - \lambda) = \frac{\bar{x} - \lambda}{R}$$

ed in tal caso la *media aritmetica* del campione è una stima di minima varianza per λ . Sempre in tal caso è poi

$$R(\lambda) \equiv R = \frac{\sigma^2}{N}$$

con

$$\tau(\lambda) \equiv \lambda \quad \text{e} \quad \tau'(\lambda) = 1$$

dunque

$$\text{Var}(\bar{x}) = \tau' R = \frac{\sigma^2}{N}$$

come da altra parte già si sapeva.

Qui la media del campione è un esempio di statistica sufficiente per λ ; infatti non ha alcuna importanza quali siano i singoli valori x_i : ma se le medie di due diversi campioni sono uguali, le conclusioni che si possono trarre sul valore di λ sono le medesime.

Supponendo di conoscere il valor medio λ , la stima della varianza σ^2 si ottiene cercando lo zero della derivata logaritmica

$$\frac{\partial \log \mathcal{L}}{\partial \sigma} = \frac{1}{\sigma^3} \left[\sum_{i=1}^N (x_i - \lambda)^2 \right] - \frac{N}{\sigma} = \frac{N}{\sigma^3} \left\{ \left[\frac{1}{N} \sum_{i=1}^N (x_i - \lambda)^2 \right] - \sigma^2 \right\}$$

la quale ha la forma richiesta perchè la soluzione

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \lambda)^2$$

sia una stima di σ^2 con minima varianza, data da

$$\text{Var} \left\{ \frac{1}{N} \sum_{i=1}^N (x_i - \lambda)^2 \right\} = \tau' R = 2\sigma \frac{\sigma^3}{N} = \frac{2\sigma^4}{N}$$

essendo $R(\sigma) = \sigma^3/N$, $\tau(\sigma) = \sigma^2$ e $\tau'(\sigma) = 2\sigma$: questo risultato è lo stesso trovato nell'appendice A.

Il valore di λ tuttavia non è generalmente noto, e l'uso della media aritmetica del campione \bar{x} comporta una distorsione che si corregge, come si è visto, ponendo $N - 1$ in luogo di N .

