

Maurizio Loreti

Dipartimento di Fisica
Università degli Studi di Padova

Teoria degli Errori e Fondamenti di Statistica

Introduzione alla Fisica Sperimentale



Dicembre 2006
(Edizione privata fuori commercio)

Questo libro è stato completamente composto e stampato dall'autore. Sono stati adoperati i programmi T_EX di Donald E. Knuth e E_T_EX di Leslie Lamport (nella versione E_T_EX2_ε); i caratteri tipografici sono quelli della famiglia Lucida, disegnati da Bigelow & Holmes e distribuiti dalla Y&Y Inc. in versione PostScript®. Il libro viene distribuito sotto la licenza GNU GPL, contenuta nell'appendice F.

IMPORTANTE:

questo testo non è ancora definitivo.

Viene mantenuto sotto CVS, e questo mi permette di risalire a tutti i cambiamenti effettuati a partire dalla data e dall'ora in cui T_EX è stato eseguito:

il 18 dicembre 2006 alle 12:18.

Maurizio Loreti

© Copyright 1987–2005 Maurizio Loreti

© Copyleft 2005–∞ (a free book distributed under the GNU GPL)

available at <http://wwwcdf.pd.infn.it/labo/INDEX.html>

Indice

Elenco delle figure	vii
Prefazione	ix
Prefazione alla sesta edizione	xi
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	12
2.5 Cifre significative ed arrotondamenti	17
2.6 Errore relativo	18
3 Elementi di teoria della probabilità	19
3.1 La probabilità: eventi e variabili casuali	19
3.2 La probabilità: definizioni	20
3.3 Proprietà della probabilità	21
3.3.1 L'evento complementare	22
3.3.2 Probabilità totale	22
3.3.3 Probabilità condizionata e probabilità composta . . .	24
3.3.4 Il teorema di Bayes	25
3.4 Definizione assiomatica della probabilità	27
3.4.1 Le leggi della probabilità e la definizione assiomatica	27
3.5 La convergenza statistica	28

4	Elaborazione dei dati	31
4.1	Istogrammi	31
4.2	Stime di tendenza centrale	33
4.2.1	La moda	35
4.2.2	La mediana	35
4.2.3	La media aritmetica	37
4.2.4	Considerazioni complessive	38
4.2.5	Prima giustificazione della media	40
4.2.6	La media aritmetica espressa tramite le frequenze . .	41
4.3	Stime di dispersione	42
4.3.1	Semidispersione massima e quantili	42
4.3.2	Deviazione media assoluta (errore medio)	43
4.3.3	Varianza e deviazione standard	43
4.4	Giustificazione della media	44
5	Variabili casuali unidimensionali discrete	47
5.1	Generalità	48
5.2	Speranza matematica	49
5.3	Il valore medio delle combinazioni lineari	50
5.4	La varianza delle combinazioni lineari	51
5.5	L'errore della media dei campioni	54
5.6	La legge dei grandi numeri	55
5.6.1	La disuguaglianza di Bienaymé-Čebyšef	55
5.6.2	Il teorema di Čebyšef	57
5.6.3	Il teorema di Bernoulli	57
5.7	Valore medio e valore vero	58
5.8	Scarto ed errore quadratico medio	59
5.9	Stima della varianza della popolazione	61
5.10	Ancora sull'errore quadratico medio	61
6	Variabili casuali unidimensionali continue	65
6.1	La densità di probabilità	65
6.2	La speranza matematica per le variabili continue	69
6.3	I momenti	69
6.4	Funzione generatrice e funzione caratteristica	71
6.4.1	Funzioni caratteristiche di variabili discrete	75
6.5	Cambiamento di variabile casuale	77
6.6	I valori estremi di un campione	78

7	Variabili casuali pluridimensionali	81
7.1	Variabili casuali bidimensionali	81
7.1.1	Momenti, funzione caratteristica e funzione generatrice	83
7.1.2	Cambiamento di variabile casuale	84
7.1.3	Applicazione: il rapporto di due variabili casuali indipendenti	85
7.1.4	Applicazione: il decadimento debole della Λ^0	86
7.1.5	Applicazione: il decadimento debole K_{e3}^0	87
7.1.6	Ancora sui valori estremi di un campione	87
7.2	Cenni sulle variabili casuali in più di due dimensioni	88
8	Esempi di distribuzioni teoriche	93
8.1	La distribuzione uniforme	93
8.1.1	Applicazione: decadimento del π^0	94
8.1.2	Applicazione: generazione di numeri casuali con distribuzione data	96
8.1.3	Esempio: valori estremi di un campione di dati a distribuzione uniforme	99
8.2	La distribuzione normale	101
8.3	La distribuzione di Cauchy	104
8.3.1	Il rapporto di due variabili normali	107
8.4	La distribuzione di Bernoulli	108
8.4.1	Applicazione: decadimenti radioattivi	113
8.4.2	Applicazione: il rapporto di asimmetria	114
8.4.3	La distribuzione binomiale negativa	115
8.5	La distribuzione di Poisson	116
8.5.1	Applicazione: esperimenti “negativi”	122
8.5.2	Applicazione: ancora il rapporto di asimmetria	122
8.5.3	La distribuzione esponenziale	124
8.5.4	La distribuzione di Erlang	126
8.5.5	La distribuzione composta di Poisson	128
8.5.6	Esempio: l’osservazione di un quark isolato	129
8.5.7	Applicazione: segnale e fondo	130
8.6	La distribuzione log-normale	132
8.7	La distribuzione normale in più dimensioni	135
9	La legge di Gauss	141
9.1	La funzione di Gauss	141
9.2	Proprietà della legge normale	142
9.3	Lo scarto normalizzato	146

9.4	Il significato geometrico di σ	149
9.5	La curva di Gauss nella pratica	149
9.6	Esame dei dati	152
9.7	Sommario delle misure dirette	153
9.8	Il teorema del limite centrale	154
9.8.1	Applicazione: numeri casuali normali	157
10	Le misure indirette	161
10.1	Risultato della misura	161
10.2	Combinazioni lineari di misure dirette	163
10.3	La formula di propagazione degli errori	164
10.4	Errore dei prodotti di potenze	165
10.5	Errori massimi	166
11	Stime di parametri	167
11.1	Stime e loro caratteristiche	167
11.2	La stima di massima verosimiglianza	170
11.2.1	Un esempio di stima sufficiente	173
11.3	Media pesata	175
11.4	Interpolazione dei dati con una curva	178
11.4.1	Interpolazione lineare per due variabili	179
11.4.2	Stima a posteriori degli errori di misura	183
11.4.3	Interpolazione con una retta per l'origine	184
11.4.4	Interpolazione lineare nel caso generale	186
11.4.5	Interpolazione non lineare	188
11.5	Altre applicazioni della stima di massima verosimiglianza	188
11.5.1	Stima di probabilità	188
11.5.2	Media e varianza di una popolazione normale	190
11.5.3	Range di una popolazione uniforme	191
11.5.4	Stima della vita media di una particella	192
12	La verifica delle ipotesi (I)	195
12.1	La distribuzione del χ^2	195
12.2	Verifiche basate sulla distribuzione del χ^2	203
12.2.1	Compatibilità dei dati con una distribuzione	203
12.2.2	Il metodo del minimo χ^2	205
12.2.3	Test di omogeneità per dati raggruppati	210
12.2.4	Un esempio: diffusione elastica protone-protone	212
12.3	Compatibilità con un valore prefissato	214
12.4	I piccoli campioni e la distribuzione di Student	217
12.5	La compatibilità di due valori misurati	220

12.6	La distribuzione di Fisher	222
12.6.1	Confronto tra varianze	224
12.7	Il metodo di Kolmogorov e Smirnov	225
13	La verifica delle ipotesi (II)	227
13.1	Un primo esempio	229
13.2	Il lemma di Neyman-Pearson	232
13.3	Tests di massima potenza uniforme	234
13.4	Il rapporto delle massime verosimiglianze	235
13.5	Applicazione: ipotesi sulle probabilità	238
13.6	Applicazione: valore medio di una popolazione normale . . .	240
A	Cenni di calcolo combinatorio	243
A.1	Il lemma fondamentale del calcolo combinatorio	243
A.2	Fattoriale di un numero intero	244
A.3	Disposizioni	244
A.4	Permutazioni	245
A.5	Permutazioni con ripetizione	245
A.6	Combinazioni	245
A.7	Partizioni ordinate	246
B	L'errore della varianza	249
C	Covarianza e correlazione	255
C.1	La covarianza	255
C.2	La correlazione lineare	259
C.3	Propagazione degli errori per variabili correlate	261
C.4	Applicazioni all'interpolazione lineare	262
C.4.1	Riscrittura delle equazioni dei minimi quadrati	262
C.4.2	Verifica di ipotesi sulla correlazione lineare	264
C.4.3	La correlazione tra i coefficienti della retta	265
C.4.4	Stima puntuale mediante l'interpolazione lineare . .	267
C.4.5	Verifica di ipotesi nell'interpolazione lineare	268
C.4.6	Adeguatezza dell'interpolazione lineare o polinomiale in genere	269
C.4.7	Il <i>run test</i> per i residui	270
C.5	Applicazioni alla stima di parametri	274
D	Il modello di Laplace e la funzione di Gauss	277
E	La funzione di verosimiglianza	283

F	La licenza GNU GPL (General Public License)	293
F.1	The GNU General Public License	293
F.2	Licenza pubblica generica del progetto GNU	301
G	Tabelle	311
H	Bibliografia	323
	Indice analitico	327

Elenco delle figure

4a	Istogramma di un campione di misure	32
4b	Frequenza cumulativa relativa di un campione di misure . .	34
4c	Distribuzioni unimodali, bimodali e amodali	36
4d	La distribuzione di Maxwell-Boltzmann	39
6a	Il comportamento limite delle frequenze relative	66
8a	Le aree elementari sulla superficie di una sfera di raggio R .	95
8b	Il metodo dei rigetti — esempio	97
8c	La distribuzione normale standardizzata	101
8d	La distribuzione di Cauchy	105
8e	La distribuzione binomiale	110
8f	La distribuzione di Poisson	120
8g	Limiti superiori sul segnale in presenza di fondo noto	132
8h	La distribuzione log-normale	134
8i	Funzione normale bidimensionale	135
8j	Funzione normale bidimensionale (curve di livello)	136
8k	Funzione normale bidimensionale (probabilità condizionate)	137
9a	Dipendenza da h della distribuzione di Gauss	143
9b	Istogrammi di dati con differente precisione	148
11a	Stime consistenti ed inconsistenti, imparziali e deviate	169
12a	La distribuzione del χ^2	197
12b	La funzione di distribuzione del χ^2	206
12c	La funzione di distribuzione del χ^2 ridotto	207
12d	Urto elastico protone-protone.	212
12e	La distribuzione di Student	218

13a	Un esempio: errori di prima e seconda specie	231
C1	Esempio di interpolazione lineare per un insieme di 12 punti.	271

Prefazione

Quando ho cominciato a tenere uno dei due corsi di Esperimentazioni di Fisica I (per il primo anno del Corso di Laurea in Fisica), ormai molti anni fa, non sono riuscito a trovare un libro di testo in cui fosse contenuto, della materia, tutto quello che io ritenevo fosse necessario per la formazione di uno studente che si supponeva destinato alla carriera del fisico; e, dopo aver usato per qualche tempo varie dispense manoscritte, mi sono deciso a riunirle in un testo completo in cui la materia fosse organicamente esposta.

Per giungere a questo è stato fondamentale l'aiuto datomi dal docente dell'altro corso parallelo di Esperimentazioni, il Prof. Sergio Ciampolillo, senza il quale questo libro non sarebbe probabilmente mai venuto alla luce; e tanto più caro mi è stato questo suo aiuto in quanto lui stesso mi aveva insegnato nel passato la statistica, completa di tutti i crismi del rigore matematico ma esposta con la mentalità di un fisico e mirata ai problemi dei fisici, nei lontani anni in cui frequentavo la Scuola di Perfezionamento (a quel tempo il Dottorato di Ricerca non era ancora nato).

Assieme abbiamo deciso l'impostazione da dare al testo e, nel 1987, preparata la prima edizione; che era ciclostilata, e che veniva stampata e distribuita a prezzo di costo agli studenti a cura dello stesso Dipartimento di Fisica. Il contenuto è stato poi più volte ampliato e rimaneggiato da me (e, all'inizio, ancora dal Prof. Ciampolillo: che ha partecipato a tutte le successive edizioni fino alla quarta compresa); dalla seconda alla quarta edizione, poi, il testo è stato edito a cura della Libreria Progetto.

L'esposizione della materia è vincolata dalla struttura del corso: un testo organico dovrebbe ad esempio presentare dapprima la probabilità e poi la statistica; ma gli studenti entrano in laboratorio sin dal primo giorno, e fanno delle misure che devono pur sapere come organizzare e come trattare per estrarne delle informazioni significative. Così si è preferito parlare subito degli errori di misura e dell'organizzazione dei dati, per poi dimostrare soltanto

alla fine (quando tutta la matematica necessaria è stata infine esposta) alcune delle assunzioni fatte; si veda a tale proposito l'esempio della media aritmetica, che gli studenti adoperano fin dal primo giorno ma il cui uso viene del tutto giustificato soltanto nel paragrafo 11.3 di questo libro.

Questo testo non contiene soltanto materia oggetto di studio nel primo anno del Corso di Laurea: su richiesta di docenti degli anni successivi, nel passato erano state aggiunte alcune parti (collocate tra le appendici) che potessero loro servire come riferimento. Ho poi largamente approfittato, sia dell'occasione offertami dall'uscita di questa quinta edizione che del fatto di dover tenere anche un corso di Statistica per la Scuola di Dottorato in Fisica, per includere nel testo degli argomenti di teoria assiomatica della probabilità e di statistica teorica che vanno assai più avanti delle esigenze di uno studente del primo anno: questo perché se, negli anni successivi, le necessità dell'analisi spingeranno dei fisici ad approfondire dei particolari argomenti di statistica, queste aggiunte sono sicuramente le basi da cui partire.

Ho cercato insomma di trasformare un testo ad uso specifico del corso di Esperimentazioni di Fisica I in una specie di riferimento base per la statistica e l'analisi dei dati: per questo anche il titolo è cambiato, e "Introduzione alle Esperimentazioni di Fisica I" è diventato un più ambizioso "Introduzione alla Fisica Sperimentale"; e "Teoria degli errori e analisi dei dati" un più veritiero "Teoria degli Errori e Fondamenti di Statistica". Ma, anche se le nuove aggiunte (addirittura per un raddoppio complessivo del contenuto originale) sono mescolate alle parti che utilizzo nel corso, ho cercato di far sì che queste ultime possano essere svolte indipendentemente dalla conoscenza delle prime.

Può stupire invece che manchi una parte di descrizione e discussione organica delle esperienze svolte e degli strumenti usati: ma gli studenti, come parte integrante del corso, devono stendere delle relazioni scritte sul loro operato che contengano appunto questi argomenti; e si è preferito evitare che trovassero già pronto un riferimento a cui potessero, per così dire, ispirarsi.

Maurizio Loreti

Gennaio 1998
(Quinta edizione)

Prefazione alla sesta edizione

Gli anni sono passati, e non ho mai smesso di modificare questo testo: per prima cosa ho corretto gli errori segnalati dai colleghi e dagli studenti, ed ho sfruttato parecchi dei loro consigli. Poi ho aggiunto nuovi paragrafi: sia per descrivere distribuzioni teoriche che, sia pur poco usate nella Fisica, ogni tanto vi compaiono; sia per introdurre piccoli esempi ed applicazioni giocattolo usati nel tenere per la seconda volta un corso di Statistica alla Scuola di Dottorato in Fisica.

Come conseguenza, questo testo è a parer mio parecchio migliorato rispetto a quello pubblicato nel 1998 dalla Decibel-Zanichelli; purtroppo le scarse vendite hanno indotto la casa editrice a rinunciare ad una seconda edizione che seguisse la prima. Conseguentemente ho deciso di mettere nella sua forma attuale (la “sesta edizione”) questo libro a disposizione della comunità su Internet: sperando che possa ancora servire a qualcuno. La licenza è quella GPL, inclusa nella appendice F; in sostanza è permesso modificare a piacimento e ridistribuire in qualunque forma questo libro, purché la ridistribuzione comprenda il sorgente.

Un'ultima considerazione personale: adesso il Corso di Laurea in Fisica è articolato in “tre più due” anni di studio; ed è con un certo piacere che ho visto come tutti gli studenti che hanno preparato con noi, sui dati dell'esperimento CDF, le loro tesi di laurea di primo livello abbiano potuto trovare aiuto in queste pagine per completarle, usando ad esempio sia metodi di verifica delle ipotesi basati sul rapporto delle funzioni di verosimiglianza che il test di Kolmogorov e Smirnov.

Maurizio Loreti

Dicembre 2006
(Sesta edizione)

*“Where shall I begin, please, your Majesty?” he asked.
“Begin at the beginning,” the King said, gravely,
“and go on till you come to the end: then stop.”*

*Charles L. Dodgson (Lewis Carroll)
Alice in Wonderland (1865)*

Capitolo 1

Introduzione

Scopo della Fisica è lo studio dei fenomeni naturali, dei quali essa cerca per prima cosa di dare una descrizione; che deve essere non solo qualitativa, ma soprattutto *quantitativa*. Questo richiede di individuare, all'interno del fenomeno, quelle grandezze fisiche in grado di caratterizzarlo univocamente; e di ottenere, per ognuna di esse, i valori che si sono presentati in un insieme significativo di casi reali. Lo studio si estende poi *oltre* la semplice descrizione, e deve comprendere 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 ricerca è quello di formulare delle *leggi fisiche* che siano in grado di dare, del fenomeno in esame, una descrizione razionale, quantitativa e (per quanto possibile) completa; e che permettano di dedurre univocamente le caratteristiche con cui esso si verificherà 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 possano essere caratterizzati dalla valutazione quantitativa di alcune loro caratteristiche, suscettibili di variare da caso a caso a seconda delle particolari 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 la quale 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 in modo naturale si presta a descrivere le relazioni tra 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 sia le grandezze rilevanti per la descrizione del fenomeno che 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); e questo è specificatamente il compito dei fisici teorici.
- Infine una fase di *verifica* delle ipotesi prima congetturate 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 moderna teoria della relatività.

È 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 suoi multipli o sottomultipli); 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; come esempio, la misura di una lunghezza mediante un regolo graduato è una misura diretta.

È una misura diretta anche quella effettuata mediante l'uso di *strumenti pretarati* (ad esempio la misura della temperatura mediante un termometro), che si basa sulla proprietà dello strumento di reagire sempre 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 valutata sia direttamente (con il tachimetro) sia 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 le seconde venivano generalmente determinate in modo indiretto, ovverosia (come appena detto) attraverso misure dirette di altre grandezze ad esse legate da relazioni algebriche: che permettevano non solo di calcolarne i valori, ma ne fissavano nel contempo anche le unità di misura.

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

Se ciascuna unità fondamentale viene ridotta di un certo fattore, il valore della grandezza espresso nelle nuove unità dovrà essere 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 risultasse in conseguenza moltiplicato per

$$L^{\lambda} M^{\mu} T^{\tau} ,$$

si dirà 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 .

Pensiamo alla velocità (media) di un corpo in movimento, che 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 il metro ed il secondo), l'unità di misura delle velocità risulta fissata univocamente (metro al secondo).

Se si alterano ad esempio l'unità di lunghezza moltiplicandola per un fattore $1/L = 1000$ (chilometro), quella di tempo moltiplicandola per un 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 (chilometro all'ora) risulterà alterato rispetto al precedente di un fattore

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

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:

$$K = \frac{1}{2} m v^2 .$$

Essa è pertanto una grandezza derivata, la cui unità di misura nel Sistema Internazionale è l'energia cinetica di un corpo avente massa di 2 Kg ed 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 $M^1 L^2 T^{-2}$; si dice dunque che un'energia ha le dimensioni di una massa, moltiplicata per il quadrato di una lunghezza e 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 ($1 \text{ m}/1 \text{ m}$ e rispettivamente $1 \text{ m}^2/1 \text{ m}^2$) lo stesso “numero puro” 1, mentre esistono per essi due diverse unità: il radiante e lo steradiano, quasi essi avessero dimensioni proprie e distinte.

Né vi è alcunché 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 delle leggi fisiche (come proposto agli inizi del secolo da Max Planck): così un sistema di unità “naturali” si potrebbe fondare, in linea di principio, ponendo uguali ad 1 la velocità della luce nel vuoto, 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 quindi da considerare fondamentali, è la riproducibilità dei campioni e la precisione con cui è possibile il confronto diretto tra grandezze omogenee.

È emblematica a questo riguardo la storia dell'evoluzione delle unità di misura delle lunghezze: queste anticamente erano riferite a parti del corpo umano quali il braccio, il cubito (già usato dagli 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 dapprima a valori medi ed infine a campioni artificiali costruiti con materiali e accorgimenti che garantissero una minima variabilità della loro lunghezza, col tempo e con le condizioni esterne più o meno controllabili.

Così, dopo la parentesi illuministica che portò all'adozione della quarantamilionesima 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 costruita usando 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 con l'andar del tempo differenze apprezzabili dai moderni metodi di misura. Inoltre l'uso di metodi ottici interferenziali finì per consentire un confronto più preciso delle lunghezze, e condusse nel 1960 (come suggerito da Babinet già nel 1829!) a svincolare la definizione del metro dalla necessità di un supporto materiale macroscopico, col porlo uguale 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 arancione 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. È implicita nella definizione anche la fiducia nell'indipendenza della velocità della luce nel vuoto sia dal sistema di riferimento dell'osservatore che dal "tipo" di luce (frequenza, stato di polarizzazione e così via); ipotesi queste che sono necessarie conseguenze delle moderne teorie della fisica.

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 "fon-

damentali” 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 ancora oggi misurate per confronto con campioni, ma la lunghezza del campione primario (il metro) è a sua volta 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 atomico (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 quantici 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) determinata con maggior precisione dopo la velocità della luce nel vuoto e che sia da essa indipendente, è noto soltanto con una incertezza dell'ordine di 0.2 parti per milione², 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 ancora meno avanzato è giunta l'evoluzione delle misure di massa, il cui campione è tuttora costituito da un particolare oggetto macroscopico detto “chilogrammo-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 prescelta. 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 *indicatore*

¹Vedi il paragrafo 2.6 alla fine del corrente capitolo.

²Attualmente (2004), l'errore relativo sul valore comunemente usato di \hbar (e che vale $1.054'571'68 \times 10^{-34} \text{ J s}$) è di 1.7×10^{-7} .

che presenta il risultato della misura ai sensi (generalmente alla vista) dello sperimentatore: o direttamente o mediante 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, a sua volta generalmente collegato 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, perché la proprietà termometrica che viene usata 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; oppure, 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 perché lo strumento risponda in modo completo 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 per il logoramento di alcune componenti in conseguenza dell'uso prolungato o improprio, o dell'invecchiamento; e, inoltre, per la presenza di varie cause di disturbo ineliminabili anche in condizioni normali d'uso dello strumento stesso.

Tutto questo fa sì che misure ripetute di una stessa grandezza fisica si distribuiscano in un intervallo più o meno ampio; la precisione si può definire come il reciproco dell'incertezza sul valore della grandezza che viene determinata dall'insieme di questi fattori: ed è sostanzialmente legata all'entità degli *errori casuali*, di cui parleremo tra poco nel paragrafo 2.4.

- L'*accuratezza* dello strumento; ossia la sua capacità di fornire valori corrispondenti a quello realmente posseduto dalla grandezza in esame. In altre parole, se lo strumento è accurato ci si aspetta che i risultati di misure ripetute della stessa grandezza fisica siano equamente distribuiti in un intorno del valore vero; questa caratteristica degli strumenti sarà, come vedremo, legata alla presenza di *errori sistematici* da essi introdotti (di questi, e delle loro possibili cause parleremo sempre nel paragrafo 2.4).

Ci si attende da uno sperimentatore serio che sappia individuare le cause di scarsa accuratezza nei suoi strumenti (ad esempio un'errata taratura dello zero della scala) ed in qualche modo neutralizzarle; così da ricondursi, in ogni caso, a *risultati accurati*.

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

Anche se, per questo motivo, generalmente la sensibilità e la precisione in uno strumento hanno valori simili, fate attenzione a non confondere i due concetti: la sensibilità è una caratteristica intrinseca degli strumenti, e rimane perciò costante in ogni situazione; mentre la precisione delle nostre misure dipende, è vero, dal tipo di strumento usato (e quindi dalla sua sensibilità) — ma anche dalle modalità contestuali di impiego e dal tipo di grandezza misurata.

Così su un orologio nella cui scala non siano riportate che poche divisioni (l'inverso della sensibilità sia ad esempio di 60 o 15 minuti) non è difficile stimare l'ora con una approssimazione che invece è dell'ordine di pochi minuti; mentre un cronometro in grado di apprezzare il decimillesimo di secondo, se azionato a mano, difficilmente può raggiungere una precisione inferiore al decimo.

Similmente, un regolo lungo un metro e graduato al millimetro può essere usato per valutare le dimensioni di un quaderno (con un singolo atto di misura); oppure (riportandolo varie volte di seguito a se stesso) le dimensioni di un edificio. È evidente come, pur essendo lo strumento lo stesso (quindi la sensibilità non varia) la precisione delle misure debba essere completamente diversa nei due casi.

2.4 Errori di misura

Come già 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 uguale al suo vero valore, che non ci potrà perciò mai essere 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 tra poco ad esaminare) produce delle differenze casuali 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 definirà 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 in una serie di misure ripetute: e la inosservabilità delle fluttuazioni non garantisce affatto che tale 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, perché non risultano 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'uguaglianza dei momenti applicati ai due bracci ed assicurata dall'equilibrio del giogo non implica l'uguaglianza delle masse ad essi sospese: perché una massa minore sospesa al braccio più lungo produrrà una azione atta ad equilibrare quella esercitata da una massa maggiore sospesa all'altro (questo errore si potrebbe anche classificare nel tipo 6, cioè come errore di interpretazione del risultato).

Un altro esempio è quello di un goniometro *eccentrico*, cioè avente la croce centrale o l'asse di rotazione in posizione diversa da quella del centro del cerchio recante la graduazione: ciò determina come conseguenza misure di angoli acuti sistematicamente errate per difetto o per eccesso a seconda della posizione del centro presunto rispetto agli assi 0° - 180° e 90° - 270° 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 a seconda del verso di tale 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 fossero esatte, le temperature lette risulterebbero 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 le lunghezze, o di recipienti tarati per la misura dei volumi, a temperature diverse da quella di taratura (generalmente fissata a 20°C); infatti la dilatazione termica farà sì che lunghezze 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 divisio-

ne di una scala graduata, lo sperimentatore tende a valutarla sempre in difetto o sempre in eccesso; oppure quando, nel leggere la posizione di un indice mobile posto di fronte ad una scala graduata (non sullo stesso piano), lo sperimentatore tenga il proprio occhio sistematicamente alla sinistra o alla destra del piano passante per l'indice ed ortogonale alla scala stessa (*errore di parallasse*). Proprio per evitare questi errori di parallasse, dietro gli indici mobili degli strumenti più precisi si pone uno specchio che aiuta l'osservatore a posizionarsi esattamente davanti ad esso.

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: questo 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; questa fa deviare il filo dalla verticale e porta sempre a sovrastimare la profondità se il fondo è approssimativamente orizzontale.

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 effettuandola si comprime inevitabilmente quest'ultimo 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_0 , utilizzando la formula

$$g = 4\pi^2 \frac{l}{T_0^2} \quad (2.1)$$

ottenuta dalla nota espressione del periodo

$$T_0 = 2\pi \sqrt{\frac{l}{g}} \quad (2.2)$$

Ma questa formula vale solo, al limite, per oscillazioni di ampiezza

infinitesima; mentre una formula che meglio approssima la realtà è³

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

ed essa mostra come il periodo T sia una funzione leggermente crescente dell'ampiezza massima θ delle oscillazioni (qui espressa in radianti). L'uso della formula (2.1) di prima approssimazione per determinare g comporta dunque una sua sottostima, che diviene tanto più sensibile quanto maggiore è θ : questo 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'altra causa di errore sistematico, originata 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 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 al T_0 definito nell'equazione (2.2).

Un modo per rivelare la presenza di errori sistematici insospettati può essere quello di misurare, se possibile, la stessa grandezza con strumenti e metodi diversi; questi presumibilmente sono affetti da errori aventi cause diverse 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 solo da una attenta e minuziosa analisi critica: sia dello strumento usato, sia della procedura seguita nella misura.

³Riguardo 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 volta scoperto, un errore sistematico può essere eliminato: modificando o lo strumento o la procedura, oppure ancora apportando una opportuna correzione al risultato della misura (sebbene questo comporti generalmente un aumento dell'errore casuale: il fattore di correzione deve essere ricavato sperimentalmente, e quindi sarà affetto da un suo errore intrinseco).

Le prime cinque categorie sopra citate come possibili cause di errori sistematici, 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 sia nelle valutazioni che 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 ulteriore e 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; in quanto 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 di misura in misura; questo perché 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 migliorando 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 proprietà 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 delle nostre misure, è *un errore* spingere la valutazione del risultato stesso 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 o

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. Così⁴

$$\begin{array}{llll} 12.34567 \pm 0.231 & \text{diventa} & 12.3 \pm 0.2 & \text{o} & 12.34 \pm 0.23 ; \\ 12.34567 \pm 0.00789 & \text{diventa} & 12.346 \pm 0.008 & \text{o} & 12.3457 \pm 0.0079 . \end{array}$$

2.6 Errore relativo

Una volta valutato l'errore presumibile Δx (errore *assoluto*) da cui è affetta la misura x_0 di una grandezza fisica x , il rapporto

$$\epsilon = \frac{\Delta x}{|x_0|} \quad (2.3)$$

(indicato in *valore* od in *percentuale*) prende il nome di *errore relativo*; essendo definito attraverso il modulo del valore stimato della grandezza in esame, l'errore relativo è una quantità sicuramente positiva.

L'errore relativo è importante perché, in un certo senso, esprime la *qualità* della misura di una grandezza: è evidente come un errore assoluto stimato in 1 cm assuma ben diverso significato se riferito alla misura di un tavolo o di una distanza astronomica — ed è appunto la differenza fra gli errori relativi a suggerirci tale interpretazione.

È opportuno tuttavia osservare che l'errore relativo definito nella (2.3) 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 inferiore all'unità.

⁴Come vedremo nelle ultime righe dell'appendice B, normalmente per l'errore si dà una sola cifra significativa; o al massimo due, se le misure sono state veramente molte — o anche per diminuire il disagio psicologico legato al “buttare via qualcosa” del frutto delle proprie fatiche...

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. È però evidente come tutti gli infiniti numeri reali non debbano essere posti sullo stesso piano: alcuni di essi saranno più verosimili (intuitivamente i numeri vicini ai risultati delle nostre misure ripetute), altri (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 dei fenomeni *casuali* o *aleatori*: cioè fenomeni ripetibili (almeno in teoria) infinite volte e che possono manifestarsi in più modalità, imprevedibili singolarmente, che si escludono a vicenda l'una con l'altra; esempi tipici di fenomeni 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, essi 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.

Il complesso delle possibili modalità con cui un fenomeno casuale si può verificare costituisce l'insieme (o *spazio*) dei *risultati*, S ; esso può essere costituito da un numero finito o infinito di elementi.

Definiremo poi come *evento casuale* l'associazione di una o più di queste possibili modalità: ad esempio, lo spazio dei risultati per il fenomeno "lancio di un dado" è un insieme composto da sei elementi; ed uno degli eventi casuali che è possibile definire (e che corrisponde al realizzarsi dell'uno o dell'altro di tre dei sei possibili risultati) consiste nell'uscita di un numero dispari. L'insieme di tutti i possibili eventi (o *spazio degli eventi*) \mathcal{E} è dunque l'insieme di tutti i sottoinsiemi di S (*insieme potenza* o *insieme delle parti* di S); compresi l'insieme vuoto \emptyset ed S stesso, che si chiamano anche rispettivamente *evento impossibile* ed *evento certo*.

Se si è in grado di fissare una legge di corrispondenza che permetta di associare ad ogni modalità di un fenomeno casuale scelta nell'insieme S uno ed un solo numero reale x , questo numero prende il nome di *variabile casuale* definita su S . Le variabili casuali possono assumere un numero finito od infinito di valori, e possono essere discrete o continue; è da notare che, per la presenza degli errori, la misura di una grandezza fisica può essere considerata come un evento casuale — ed il risultato numerico che da tale misura otteniamo è una variabile casuale che possiamo associare 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, purché tutti questi casi possibili siano ugualmente probabili.

e se ne ricava immediatamente il seguente

COROLLARIO: *la probabilità di un evento casuale è un numero 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 intrinsecamente insoddisfacente perché racchiude in sé stessa una *tautologia*: si nota immediatamente come, per definire la probabilità, essa presupponga che si sia già in grado di valutare l'equiprobabilità delle varie modalità con cui può manifestarsi l'evento considerato. Nel caso di una variabile casuale continua, ciò si traduce nell'indeterminazione di quale tra le variabili topologicamente equivalenti (ossia 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, teorizzata da von Mises²): definiamo la *frequenza 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

Proseguendo in questa nostra esposizione, useremo ora la definizione empirica per ricavare alcune proprietà delle probabilità di eventi casuali: queste stesse proprietà, come vedremo nel paragrafo 3.4.1, possono essere ricavate a partire dalla *definizione assiomatica* (matematicamente soddisfacente, e che verrà presentata nel paragrafo 3.4). Il motivo per cui ci basiamo sulla definizione empirica è sia la maggiore semplicità delle dimostrazioni che la concretezza e l'intuitività dei ragionamenti, che si possono facilmente esemplificare con semplici procedure pratiche come il lancio di monete e dadi.

¹Anche questa definizione non è completamente soddisfacente dal punto di vista concettuale (come vedremo più in dettaglio nel paragrafo 3.5); ma è tra le più intuitive, perché tra le più vicine all'uso pratico.

²Richard von Mises fu un matematico che visse dal 1883 al 1953; compì ricerche nei campi della probabilità e della statistica, ma soprattutto in quello della matematica applicata alla meccanica dei fluidi (nel 1913 istituì all'Università di Vienna il primo corso al mondo sul volo, e nel 1915 progettò un aereo che pilotò personalmente nel corso della I guerra mondiale).

3.3.1 L'evento complementare

La mancata realizzazione dell'evento E costituisce l'*evento complementare* ad E , che indicheremo con \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

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

Analogamente si può dimostrare che, se A, B, \dots, Z sono eventi casuali *mutuamente esclusivi* e che *esauriscono l'insieme di tutti i possibili risultati*, vale la

$$p(A) + p(B) + \dots + p(Z) = 1 . \quad (3.1)$$

3.3.2 Probabilità totale

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 quello consistente nel 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, partendo dalle seguenti identità:

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

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

che devono valere le

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

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

ed altre due 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 , definito come l'evento casuale consistente nel verificarsi *o dell'uno o dell'altro di essi o di entrambi*, otteniamo

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

da cui, passando al limite,

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

Nel caso particolare di due eventi E ed F che si escludano mutuamente (cioè per cui sia $p(EF) = 0$ e $n_{11} \equiv 0$) 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:

$$p(A + B + \cdots + Z) = p(A) + p(B) + \cdots + p(Z) \quad . \quad (3.2)$$

3.3.3 Probabilità condizionata e probabilità composta

La probabilità che si verifichi l'evento E nel caso 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 precedente, l'identità

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

con l'analoga

$$f(F|E) = \frac{f(EF)}{f(E)} ;$$

e vale quindi, passando al limite, la

$$p(EF) = p(F) \cdot p(E|F) = p(E) \cdot p(F|E) . \quad (3.3)$$

Nel caso particolare di due eventi casuali tali che il verificarsi o meno dell'uno non alteri la probabilità di presentarsi dell'altro, ovverosia per cui risulti $p(E|F) = p(E)$ e $p(F|E) = p(F)$, questi si dicono tra loro *statisticamente indipendenti*³; e per essi vale la seguente legge (della *probabilità composta*):

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

Questa si generalizza facilmente (sempre per induzione completa) ad un evento complesso costituito dal verificarsi contemporaneo di un numero qualsiasi di eventi semplici (sempre però tutti statisticamente indipendenti tra loro); per il quale vale la

$$p(A \cdot B \cdots Z) = p(A) \cdot p(B) \cdots p(Z) . \quad (3.4)$$

Più in particolare, gli eventi casuali appartenenti ad un insieme di dimensione N (con $N > 2$) si dicono *tutti* statisticamente indipendenti tra loro quando la probabilità del verificarsi di uno qualsiasi di essi non è alterata dal fatto che uno o più d'uno degli altri si sia già presentato.

Come esempio si consideri il lancio indipendente di due dadi, ed i seguenti tre eventi casuali: A , consistente nell'uscita di un numero dispari sul primo dado; B , consistente nell'uscita di un numero dispari sul secondo dado; e C , consistente nell'uscita di un punteggio complessivo dispari. È facile

³Il concetto di indipendenza statistica tra eventi casuali fu definito per la prima volta nel 1718 da Abraham de Moivre (purtroppo noto al grosso pubblico solo per aver correttamente predetto il giorno della propria morte servendosi di una formula matematica), nel suo libro "The Doctrine of Chance".

vedere che questi eventi casuali sono, se considerati a due a due, statisticamente indipendenti: A e B per ipotesi, A e C perché $p(C|A) = \frac{1}{2} = p(C)$, ed infine B e C perché anche $p(C|B) = \frac{1}{2} = p(C)$; ma gli stessi tre eventi, se vengono considerati nel loro complesso, *non* sono *tutti* statisticamente indipendenti — perché il verificarsi di A assieme a B rende poi impossibile il verificarsi di C .

3.3.4 Il teorema di Bayes

Supponiamo che un dato fenomeno casuale A possa dare luogo a N eventualità mutuamente esclusive A_j , che esauriscano inoltre la totalità delle possibilità; e sia poi un differente fenomeno casuale che possa condurre o al verificarsi o al non verificarsi di un evento E . Osservando la realizzazione di entrambi questi fenomeni, se E si verifica, assieme ad esso si dovrà verificare anche una ed una sola delle eventualità A_j ; applicando prima la legge della probabilità totale (3.2) e poi l'equazione (3.3), si ottiene

$$p(E) = \sum_{j=1}^N p(E \cdot A_j) = \sum_{j=1}^N p(A_j) \cdot p(E|A_j) . \quad (3.5)$$

Ora, riprendendo la legge fondamentale delle probabilità condizionate (3.3), ne ricaviamo

$$p(A_i|E) = \frac{p(A_i) \cdot p(E|A_i)}{p(E)}$$

e, sostituendovi la (3.5), si giunge alla

$$p(A_i|E) = \frac{p(A_i) \cdot p(E|A_i)}{\sum_j [p(A_j) \cdot p(E|A_j)]} \quad (3.6)$$

L'equazione (3.6) è nota con il nome di *teorema di Bayes*, e viene spesso usata nel calcolo delle probabilità; talvolta anche, come adesso vedremo, quando le A_j non siano tanto eventi casuali in senso stretto, quanto piuttosto *ipotesi* da discutere per capire se esse siano o meno rispondenti alla realtà.

Facendo un esempio concreto, si abbiano due monete: una “buona”, che presenti come risultato la testa e la croce con uguale probabilità (dunque pari a 0.5); ed una “cattiva”, con due teste sulle due facce. Inizialmente si sceglie una delle due monete; quindi avremo due eventualità mutuamente esclusive: A_1 (è stata scelta la moneta “buona”) e A_2 (è stata scelta la moneta “cattiva”) con probabilità rispettive $p(A_1) = p(A_2) = 0.5$. Se l'evento casuale E consiste nell'uscita di una testa, ovviamente $p(E|A_1) = 0.5$ e $p(E|A_2) = 1$.

Se ora facciamo un esperimento, lanciando la moneta una volta e ottenendo una testa, quale è la probabilità che nell'effettuare la scelta iniziale si sia presa quella "buona"? La risposta è data dal teorema di Bayes, da cui si ottiene:

$$\begin{aligned}
 p(A_1|E) &= \frac{p(A_1) \cdot p(E|A_1)}{p(A_1) \cdot p(E|A_1) + p(A_2) \cdot p(E|A_2)} \\
 &= \frac{0.5 \cdot 0.5}{0.5 \cdot 0.5 + 0.5 \cdot 1} \\
 &= \frac{0.25}{0.75} \\
 &= \frac{1}{3} .
 \end{aligned}$$

Ovviamente, se si volesse progettare un esperimento reale, sarebbe meglio associarlo al lanciare la moneta N volte (con $N > 1$): o si ottiene almeno una croce, ed allora è sicuramente vera A_1 ; o, invece, si presenta l'evento E consistente nell'ottenere N teste in N lanci. In quest'ultimo caso, $p(E|A_2) = 1$ e $p(E|A_1) = 1/2^N$ se i lanci sono indipendenti tra loro; utilizzando ancora l'equazione (3.6), si ricava che la probabilità di aver scelto la moneta "buona", $p(A_1)$, è data da $1/(1 + 2^N)$ — e di conseguenza $p(A_2) = 2^N/(1 + 2^N)$ è la probabilità che si sia scelta la moneta "cattiva".

Qui il teorema di Bayes viene utilizzato per *verificare una ipotesi statistica*: ovvero per calcolare la probabilità che l'una o l'altra di un insieme di condizioni A_j che si escludono a vicenda sia vera, sulla base di osservazioni sperimentali riassunte dal verificarsi di E ; ma questo ci risulta possibile solo perché si conoscono *a priori* le probabilità di *tutte* le condizioni stesse $p(A_j)$.

Se, viceversa, queste non sono note, la (3.6) ci dà ancora la probabilità che sia vera l'una o l'altra delle ipotesi A_j se sappiamo che si è verificata la condizione sperimentale E ; ma essa non si può ovviamente calcolare, a meno di fare opportune ipotesi sui valori delle $p(A_j)$: ad esempio assumendole tutte uguali, il che è chiaramente arbitrario. Per essere più specifici, non potremmo servirci di un esperimento analogo a quelli delineati e del teorema di Bayes per calcolare la probabilità che una particolare moneta da 1 euro ricevuta in resto sia o non sia "buona": a meno di non conoscere a priori $p(A_1)$ e $p(A_2)$, le probabilità che una moneta da 1 euro scelta a caso tra tutte quelle circolanti nella nostra zona sia "buona" o "cattiva".

3.4 Definizione assiomatica della probabilità

Per completezza, accenniamo infine alla cosiddetta *definizione assiomatica della probabilità*⁴, che è matematicamente consistente:

Sia S l'insieme di tutti i possibili risultati di un fenomeno casuale, ed E un qualsiasi evento casuale definito su S (ossia un qualsiasi sottoinsieme $E \subseteq S$). Si definisce come “probabilità” di E un numero, $p(E)$, associato univocamente all'evento stesso, che soddisfi alle seguenti tre proprietà:

1. $p(E) \geq 0$ per ogni E ;
2. $p(S) = 1$;
3. $p(E_1 \cup E_2 \cup \dots) = p(E_1) + p(E_2) + \dots$ per qualsiasi insieme di eventi E_1, E_2, \dots , in numero finito od infinito e a due a due senza alcun elemento in comune (ossia tali che $E_i \cap E_j = \emptyset$ per ogni $i \neq j$).

Questa definizione, pur matematicamente consistente⁵, non dice nulla su come assegnare dei valori alla probabilità; tuttavia su tali valori si possono fare delle ipotesi, verificabili poi analizzando gli eventi reali osservati.

3.4.1 Le leggi della probabilità e la definizione assiomatica

Dalla definizione assiomatica è possibile ricavare, come abbiamo già prima accennato, le stesse leggi cui siamo giunti a partire dalla definizione empirica. Infatti:

- Essendo $S \cup \emptyset = S$, la proprietà 3 (applicabile perché $S \cap \emptyset = \emptyset$) implica $p(S) + p(\emptyset) = p(S)$; da cui ricaviamo, vista la proprietà 2,

$$p(\emptyset) = 0 \quad .$$

- Se $A \supset B$, essendo in questo caso $A = B \cup (A \cap \bar{B})$, applicando la proprietà 3 (il che è lecito dato che $B \cap (A \cap \bar{B}) = \emptyset$) si ottiene $p(A) = p(B) + p(A \cap \bar{B})$; e, vista la proprietà 1,

$$A \supset B \quad \Rightarrow \quad p(A) \geq p(B) \quad .$$

⁴Questa definizione è dovuta all'eminente matematico russo Andrei Nikolaevich Kolmogorov; vissuto dal 1903 al 1987, si occupò principalmente di statistica e di topologia. Fu enunciata nel suo libro del 1933 *Grundbegriffe der Wahrscheinlichkeitsrechnung*.

⁵Volendo essere del tutto rigorosi, questa definizione risulta valida solo se l'insieme dei possibili risultati è composto da un numero finito o da un'infinità numerabile di elementi; la reale definizione assiomatica della probabilità è leggermente differente (ed ancora più astratta).

- Dati due insiemi A e B , visto che qualunque essi siano valgono le seguenti identità:

$$\begin{aligned} A &= (A \cap B) \cup (A \cap \bar{B}) \\ B &= (A \cap B) \cup (\bar{A} \cap B) \\ (A \cup B) &= (A \cap B) \cup (A \cap \bar{B}) \cup (\bar{A} \cap B) \end{aligned}$$

e applicando a queste tre relazioni (dopo aver verificato che gli insiemi a secondo membro sono tutti disgiunti) la proprietà 3 e sommando e sottraendo opportunamente i risultati, si ottiene la *legge della probabilità totale* nella sua forma più generale:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) .$$

Definendo poi $p(E|A)$ (con $p(A) \neq 0$) come

$$p(E|A) = \frac{p(E \cap A)}{p(A)} , \quad (3.7)$$

è facile riconoscere che anche essa rappresenta una probabilità: essendo $p(E \cap A) \geq 0$ e $p(A) > 0$, $p(E|A)$ soddisfa alla proprietà 1; essendo $S \cap A = A$, $p(S|A) = p(A)/p(A) = 1$, e $p(E|A)$ soddisfa alla proprietà 2; infine, se E_1, E_2, \dots sono insiemi a due a due disgiunti,

$$\begin{aligned} p(E_1 \cup E_2 \cup \dots | A) &= \frac{p[(E_1 \cup E_2 \cup \dots) \cap A]}{p(A)} \\ &= \frac{p[(E_1 \cap A) \cup (E_2 \cap A) \cup \dots]}{p(A)} \\ &= \frac{p(E_1 \cap A)}{p(A)} + \frac{p(E_2 \cap A)}{p(A)} + \dots \\ &= p(E_1|A) + p(E_2|A) + \dots \end{aligned}$$

e $p(E|A)$ soddisfa anche alla proprietà 3. Dalla (3.7) si ottiene infine la *legge della probabilità composta* nella sua forma più generale,

$$p(A \cap B) = p(A|B) \cdot p(B) = p(B|A) \cdot p(A) .$$

3.5 La convergenza statistica

Difetto della definizione empirica di probabilità, oltre a quello di essere basata su di un esperimento, è quello di presupporre a priori una convergenza della frequenza relativa f , al crescere di N , verso un valore ben definito: valore che si assume poi come probabilità dell'evento.

Qualora si assuma come definizione di probabilità quella assiomatica, è effettivamente possibile dimostrare (come vedremo più avanti nel paragrafo 5.6, ed in particolare nel sottoparagrafo 5.6.3) come, al crescere del numero di prove, la frequenza relativa di un *qualunque* evento casuale converga verso la probabilità dell'evento stesso.

È tuttavia assai importante sottolineare come questa legge (*legge dei grandi numeri*, o *teorema di Bernoulli*) 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, se si effettuano N prove, per ogni $N > M$ risulti *sicuramente* $|f(E) - p(E)| < \epsilon$. 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 il concetto di convergenza va inteso invece in senso *statistico* (o *debole*, o *stocastico*); si dice che all'aumentare del numero di prove N una grandezza x tende statisticamente al limite X quando, scelta una qualsiasi coppia di numeri positivi ϵ e δ , si può in conseguenza determinare un numero intero M tale che, se si effettua un numero di prove N maggiore di M , la probabilità che x differisca da X per più di ϵ risulti minore di δ . Indicando col simbolo $\Pr(E)$ la probabilità di un evento E , la definizione di convergenza statistica è

$$\forall \epsilon, \delta > 0 \quad \rightarrow \quad \exists M : \quad N > M \Rightarrow \Pr(|x - X| \geq \epsilon) \leq \delta . \quad (3.8)$$

Nel paragrafo 5.6 vedremo che, dato un qualunque evento casuale E avente probabilità $\Pr(E)$ di manifestarsi, si può dimostrare che la sua frequenza relativa $f(E)$ su N prove converge statisticamente a $\Pr(E)$ all'aumentare di N ; o, in altre parole, come aumentando il numero di prove si possa rendere tanto improbabile quanto si vuole che la frequenza relativa e la probabilità di un qualunque evento casuale E 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 un insieme di più misure della stessa grandezza fisica (nella statistica si parla in genere di un *campione* di misure), 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 una corrispondenza biunivoca tra i numeri reali ed i punti di una retta orientata, ognuna delle nostre misure può essere rappresentata su di essa da un punto; l'istogramma è un particolare tipo di diagramma cartesiano in cui l'asse delle ascisse è dedicato a tale rappresentazione. Tuttavia è facile rendersi conto del fatto che non tutti i valori della variabile sono in realtà permessi, perché gli strumenti forniscono per loro natura un insieme discreto di valori essendo limitati ad un numero finito di cifre significative.

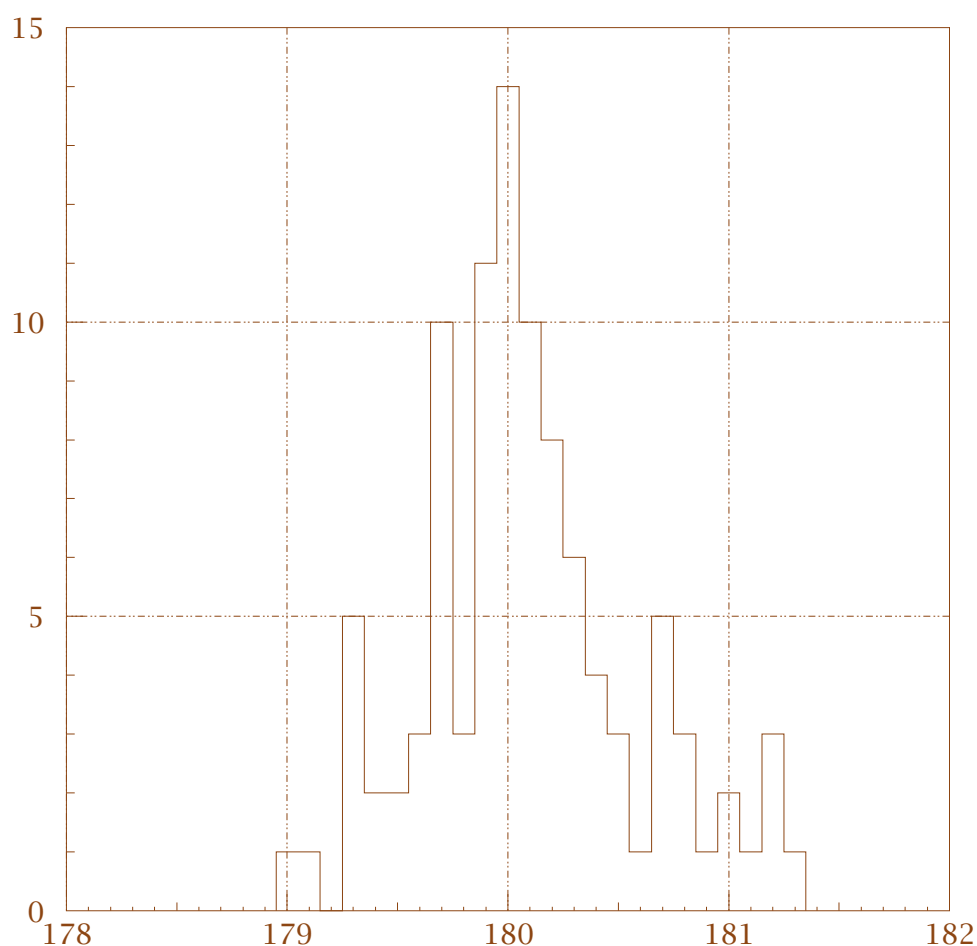
Convienne allora mettere in evidenza sull'asse delle ascisse tutti i possibili valori che possono essere ottenuti da una misura reale; cioè punti separati da un intervallo che corrisponde alla cifra significativa più bassa dello strumento, o comunque alla più piccola differenza apprezzabile con esso se l'ultima cifra deve essere stimata dall'osservatore (ad esempio il decimo di grado stimato ad occhio su un goniometro avente scala al mezzo grado).

Nelle ordinate del diagramma si rappresenta poi la frequenza assoluta

con la quale i diversi valori si sono presentati; questo si fa associando ad ognuna delle misure 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: di modo che è possibile, dall'altezza di una colonna di rettangoli unitari sovrapposti, risalire al numero di dati del campione aventi un determinato valore.

FIGURA 4a - 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 corrispon-

dendo ad un intervallo multiplo opportuno del più piccolo rappresentabile discusso sopra.

Anziché costruire l'istogramma riportandovi un risultato per volta, si possono contare prima le frequenze in ciascuna classe e disegnare sopra ognuna di esse un rettangolo avente area corrispondente alla frequenza ivi osservata. L'area dell'istogramma sopra ad un qualsiasi 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 di altezza unitaria. L'area totale sottesa dall'istogramma è, sempre rispetto a tale unità, pari al numero di osservazioni N .

Un'altra rappresentazione, che è poco usata ma vantaggiosa perché 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 x , 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 di ognuno degli N valori osservati. Risulta inoltre

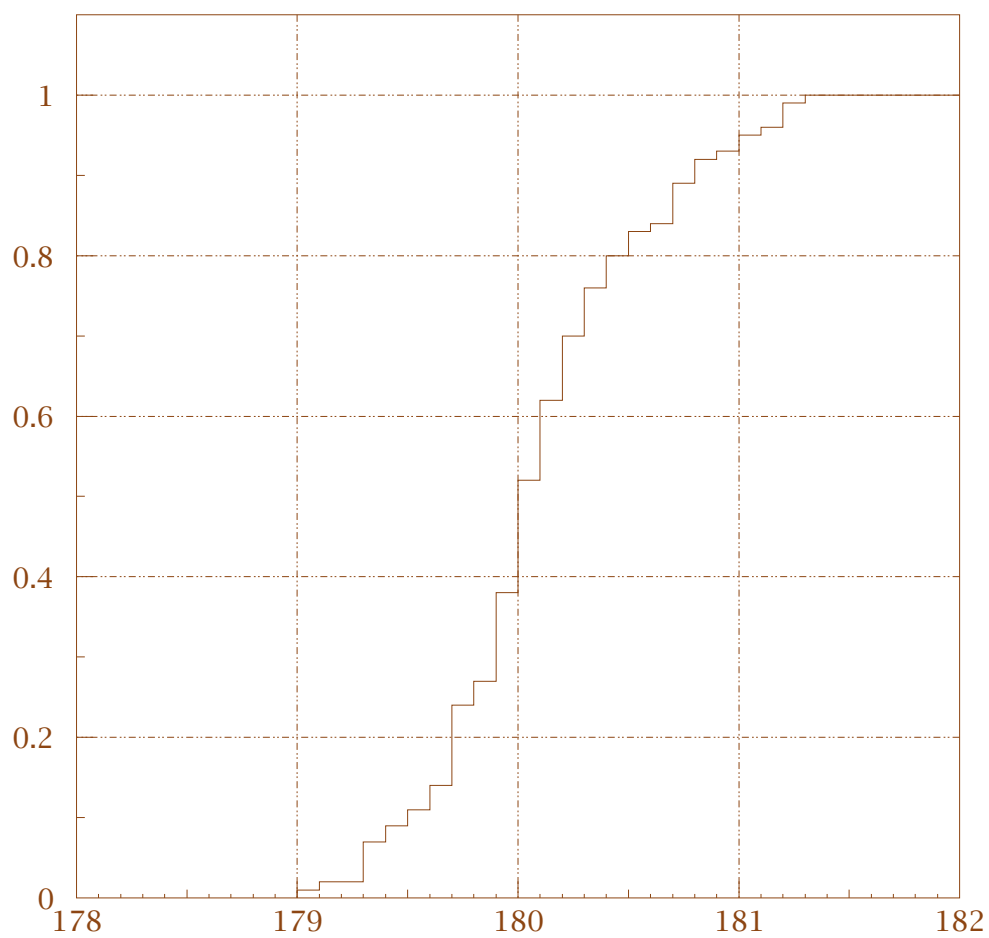
$$0 = F(-\infty) \leq F(x) \leq F(+\infty) = \begin{cases} N & \text{(ass.)} \\ 1 & \text{(rel.)} \end{cases}$$

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 ascissa in posizione centrale rispetto alla 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.

FIGURA 4b - Frequenza cumulativa relativa per le stesse misure della figura 4a.



4.2.1 La moda

Nella statistica esistono varie stime della cosiddetta *tendenza centrale* di un campione; una di queste stime è il valore corrispondente al massimo della frequenza, cioè il valore che si è presentato il maggior numero di volte (ovvero la media dei valori *contigui* che presentassero tutti la medesima massima frequenza): tale stima (se esiste) si chiama *moda* del campione, e si indica con il simbolo \hat{x} .

In generale però la distribuzione potrebbe non avere massimo (distribuzioni *amodali*), 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. Talvolta si dice che la distribuzione non ha moda anche se il massimo esiste, ma si presenta ad uno degli estremi dell'intervallo che contiene le misure; non essendo in tal caso la moda, ovviamente, una stima di tendenza centrale.

Per tutti 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

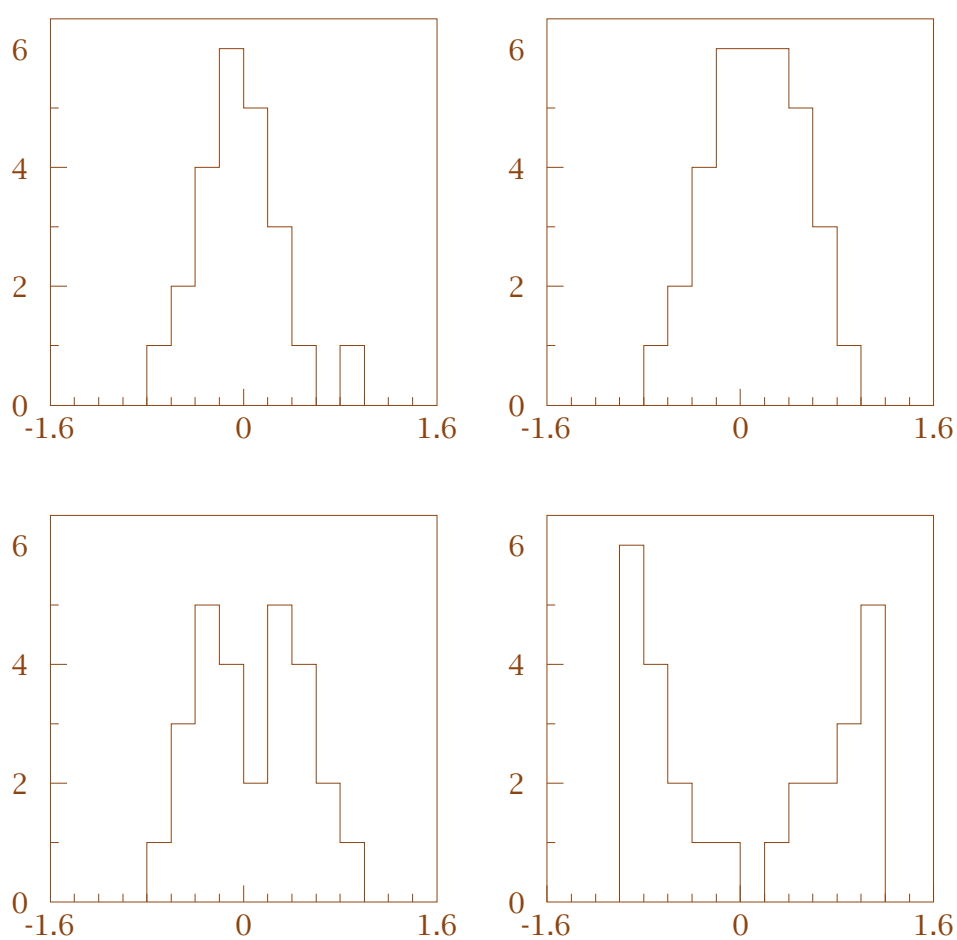
Un'altra stima di tendenza centrale di uso frequente nella statistica (anche se non nella fisica) è la *mediana* di un campione: indicata col simbolo \tilde{x} , è definita come quel valore che divide l'istogramma dei dati in due parti di uguale area¹; in termini meno precisi, la mediana lascia un uguale numero di dati alla propria sinistra ed alla propria destra². Usando questa forma della 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 relativa 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

¹Il valore della mediana di un insieme di dati, così definito, dipende dalla scelta delle classi di frequenza; per questo motivo la mediana in genere non si adopera tanto per i campioni sperimentali di dati, quanto per le distribuzioni teoriche.

²Basta applicare le due definizioni ad un insieme di dati composto dai tre valori $\{0, 1, 1\}$ per rendersi conto della differenza.

FIGURA 4c - 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.



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 della tendenza centrale di un campione è la *media aritmetica* \tilde{x} dei valori osservati, definita attraverso la

$$\tilde{x} = \frac{1}{N} \sum_{i=1}^N x_i \ . \quad (4.1)$$

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

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

ed infine

$$\sum_{i=1}^N (x_i - \tilde{x}) \equiv 0 \ . \quad (4.2)$$

PROPRIETÀ 2: *la media aritmetica \tilde{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*

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

Infatti abbiamo

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

da qui, sfruttando l'equazione (4.2), si ottiene

$$\sum_{i=1}^N (x_i - x)^2 = \sum_{i=1}^N (x_i - \bar{x})^2 + N(\bar{x} - x)^2 \quad (4.3)$$

e finalmente

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

4.2.4 Considerazioni complessive

Oltre le tre stime citate di tendenza centrale ne esistono altre, di uso però limitato a casi particolari e che non hanno frequente uso né nella statistica né nella fisica; per soli motivi di completezza citiamo qui:

- la *media geometrica*, g , definita come la radice N -esima del prodotto degli N valori rappresentati nel campione:

$$g^N = \prod_{i=1}^N x_i ;$$

- la *media armonica*, h , definita come il reciproco del valore medio dei reciproci dei dati:

$$\frac{1}{h} = \frac{1}{N} \sum_{i=1}^N \frac{1}{x_i} ;$$

- la *media quadratica*, q , definita come la radice quadrata del valore medio dei quadrati dei dati:

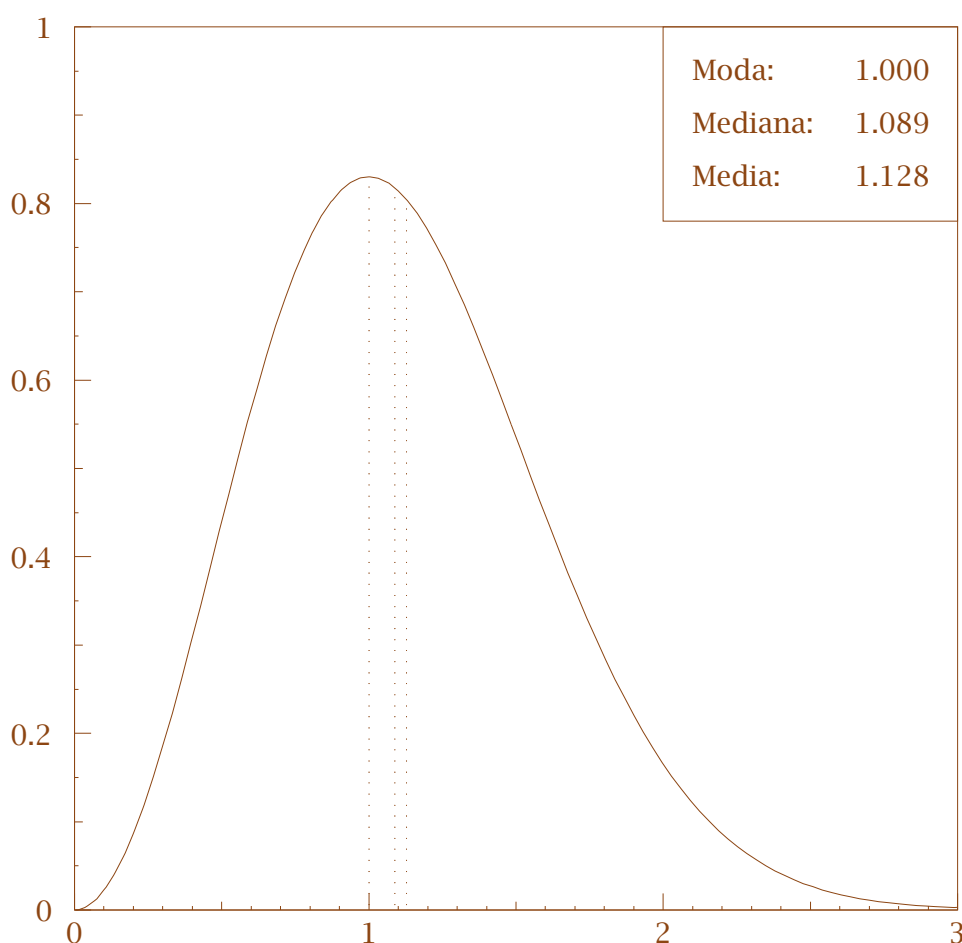
$$q = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} .$$

Se la distribuzione dei dati non è troppo irregolare, le prime tre stime citate per la tendenza centrale (moda, mediana e media aritmetica) 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.

FIGURA 4d - I valori delle tre principali stime di tendenza centrale per la distribuzione di Maxwell-Boltzmann; l'unità per le ascisse è il parametro α che compare nell'equazione (4.4).



Come esempio, nella figura 4d è mostrato l'andamento di una distribuzione di probabilità per una variabile (continua) che è di interesse per la

fisica; e nel grafico sono messi in evidenza i valori per essa assunti dalle tre stime di tendenza centrale considerate. Si tratta della funzione di frequenza detta di *Maxwell-Boltzmann*, e secondo essa sono ad esempio distribuiti, in un gas perfetto, i moduli delle velocità delle molecole: l'equazione della curva è

$$y = f(v) = \frac{4}{\sqrt{\pi}} \alpha^{\frac{3}{2}} v^2 e^{-\alpha v^2} \quad (4.4)$$

(in cui α è una costante dipendente dalla massa delle molecole e dalla temperatura del gas).

La scelta dell'uso dell'una o dell'altra stima statistica per determinare la tendenza centrale di un campione di misure ripetute andrà decisa sulla base delle proprietà delle stime stesse; 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 affette dallo stesso errore (eseguite insomma in condizioni simili) e composti da uno stesso numero di dati. La stima che sceglieremo dovrebbe essere la migliore, nel senso già usato all'inizio di questo paragrafo 4.2: quella che ha la maggiore probabilità di darci il valore vero della grandezza misurata.

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. In particolare vedremo nel paragrafo 11.3 che la media aritmetica è effettivamente la stima *migliore*, nel senso ora chiarito di questa frase.

A questo punto possiamo già comunque renderci conto (anche se in maniera *non rigorosa*) che la media aritmetica di più misure dovrebbe avere 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 $\epsilon_i = x_i - x^*$. L'errore assoluto della media aritmetica è allora dato da

$$\bar{\epsilon} = \bar{x} - x^*$$

e, sfruttando la (4.1),

$$\bar{\epsilon} = \frac{1}{N} \sum_{i=1}^N x_i - x^* = \frac{1}{N} \sum_{i=1}^N (x_i - x^*) = \frac{1}{N} \sum_{i=1}^N \epsilon_i .$$

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 quindi 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 casuale: 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 .

È bene osservare come si possano definire infinite medie pesate dei valori numerici x_j , corrispondenti alle infinite differenti maniere di attribuire ad ognuno di essi un peso; ed anche che, in genere, con il nome di “media pesata” ci si riferisce a quella particolare formula che permette di calcolare la migliore stima del valore vero di una grandezza fisica sulla base di più misure aventi differente precisione (l’equazione (11.7), che incontreremo più avanti nel paragrafo 11.3), e non alla formula precedente.

Fin qui tale 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. In tal caso, se assumiamo che la frequenza relativa con cui ogni valore si è presentato tenda stocasticamente alla probabilità rispettiva, in definitiva otteniamo che la media aritmetica delle misure deve anch’essa tendere 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 un campione 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

Abbiamo sviluppato il paragrafo 4.2 partendo dall'intuizione (giustificata con l'aiuto delle caratteristiche degli errori casuali e della legge dei grandi numeri) che la tendenza centrale di un insieme di misure è legata al valore vero della grandezza misurata. Così, similmente, si intuisce che agli errori introdotti nell'eseguire le nostre misure è legata un'altra grandezza caratteristica del campione, cioè la sua *dispersione*: ovvero la valutazione della larghezza dell'intervallo in x in cui le misure stesse sono distribuite attorno al valore centrale.

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. Il doppio della semidispersione massima

$$R = x_{\max} - x_{\min}$$

è anch'esso usato come stima della dispersione di un campione, e viene chiamato *range*.

Grandezze frequentemente usate per caratterizzare una distribuzione nella statistica (non nella fisica) sono i *quantili*, i *decili* ed i *percentili* (collettivamente *quantili*), indicati con Q_i ($i = 1, 2, 3$); con D_i ($i = 1, \dots, 9$); e con 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

$$Q_2 \equiv D_5 \equiv P_{50} \equiv \tilde{x} .$$

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 (errore medio)

Altra stima di dispersione è la *deviazione media assoluta* (o *errore medio*), 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 da 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*) s ; che si definisce 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 i $\frac{4}{5}$ della deviazione standard, e l'intervallo semiinterquartilico è circa i $\frac{2}{3}$ 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:

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

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 anziché 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 nel paragrafo 9.3.

Comunque, una volta assunto questo, possiamo approfondire il discorso già cominciato sulla media aritmetica come stima del centro della distribuzione e quindi del valore vero di una grandezza, nel caso di misure ripetute ed in assenza di errori sistematici. È infatti possibile provare³ che la media

³La 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 E.

aritmetica è la 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 di quella grandezza 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 quantità⁴*

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

⁴La differenza tra questa formula e quella prima citata non è praticamente avvertibile se N non è troppo piccolo.

Capitolo 5

Variabili casuali unidimensionali discrete

Già sappiamo (come osservato nel paragrafo 3.1) che, a causa degli inevitabili 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'insieme di tutti i 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 elementi estratti a caso dall'insieme di tutte le infinite possibili operazioni di misura che potrebbero essere effettuate sulla stessa grandezza fisica, eseguite col medesimo strumento e sfruttando le medesime procedure.

Quest'ultimo insieme nella terminologia della 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 un'entità preesistente alle operazioni effettivamente eseguite; a differenza dell'insieme di tutti gli individui di una vera popolazione, dalla quale si estrae realmente un campione per eseguire una ricerca demografica. Sebbene sia una finzione, questo concetto è tuttavia utile per poter applicare la teoria della 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): tra le altre cose, metteremo in evidenza i rapporti tra grandezze statistiche che si riferiscano ad un campione limitato e grandezze analoghe che siano invece

riferite all'intera popolazione (*teoria del campionamento*); e dimostreremo la validità della legge dei grandi numeri.

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 invece 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 x di teste osservate nell'evento è ancora una variabile casuale discreta, la cui definizione è data dalla tabella seguente:

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

e, come si può notare, la corrispondenza tra la variabile casuale e l'insieme dei possibili risultati non è in questo caso biunivoca.

Se l'insieme di definizione 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 in pratica 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 perché 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 deve tendere statisticamente al valore p_j (probabilità di osservare il valore x_j), e sarà ancora

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

come dovevamo ovviamente attenderci ricordando l'equazione (3.1).

5.2 Speranza matematica

Come sappiamo dal paragrafo 4.2.6, il valore medio della variabile x su di un campione finito è dato dall'equazione

$$\bar{x} = \sum_i f_i x_i$$

dove la sommatoria si intende estesa a tutti i valori che la x può assumere, essendo nulle le frequenze di quelli che non si sono effettivamente presentati; definiamo in maniera analoga una nuova grandezza $E(x)$, relativa all'intera popolazione, mediante la

$$E(x) = \sum_i p_i x_i . \quad (5.1)$$

$E(x)$ (che si chiama *speranza matematica* della variabile casuale x) ci appare quindi come una generalizzazione alla popolazione del concetto di media aritmetica e, se si assumesse come definizione di probabilità quella empirica, sarebbe in base ad essa il limite (statistico) del valore medio del campione all'aumentare della sua dimensione; per cui lo chiameremo anche, meno appropriatamente, *valore medio di x sull'intera popolazione*.

È da notare come non ci sia alcuna garanzia dell'esistenza di $E(x)$ se l'insieme dei possibili valori x_i non è finito (in particolare se x è una variabile continua); in effetti esistono delle distribuzioni di probabilità usate anche in fisica (ad esempio la *distribuzione di Cauchy*, che studieremo più avanti nel paragrafo 8.3) per le quali la sommatoria della (5.1) non converge, e che non ammettono quindi speranza matematica.

La speranza matematica per la variabile casuale $[x - E(x)]^2$ (ossia la generalizzazione alla popolazione della varianza di un campione) si indica poi col simbolo $\text{Var}(x)$:

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

e ad essa ci riferiremo come *varianza della popolazione della variabile casuale* x ; come $E(x)$, e per gli stessi motivi, anch'essa potrebbe non esistere per quelle variabili che assumono un numero infinito di possibili valori.

Le considerazioni dei paragrafi seguenti si applicano ovviamente solo a popolazioni di variabili casuali per le quali esista finita la speranza matematica e, qualora la si consideri, la varianza. Inoltre non useremo mai la definizione empirica di probabilità, ma quella assiomatica; e vedremo come, partendo da essa, si possa dimostrare la legge detta “dei grandi numeri” già enunciata nel paragrafo 3.5: ossia la convergenza, all'aumentare del numero di prove effettuate, della frequenza di un qualsiasi evento casuale alla sua probabilità.

5.3 Il valore medio delle combinazioni lineari

Consideriamo due variabili casuali x e y , aventi speranza matematica $E(x)$ ed $E(y)$ rispettivamente; ed una loro qualsiasi combinazione lineare a coefficienti costanti $z = ax + by$. Vogliamo dimostrare ora che la speranza matematica della nuova variabile z esiste, ed è data dalla combinazione lineare delle speranze matematiche di x e di y con gli stessi coefficienti a e b .

Indichiamo con x_j i possibili valori della prima variabile, e con y_k quelli della 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, applicando la legge della probabilità totale (equazione (3.2)) risulterà

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

Per la speranza matematica $E(z)$ di z avremo poi

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

È immediato poi estendere, per induzione completa, questa dimostrazione alla combinazione lineare di un numero qualsiasi di variabili casuali: se abbiamo

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

allora

$$E(F) = aE(x) + bE(y) + cE(z) + \dots \quad (5.2)$$

Una importante conseguenza può subito essere ricavata applicando l'equazione (5.2) alla 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 la speranza matematica di \bar{x} , ovverosia il valore medio delle varie \bar{x} su un numero molto elevato di campioni di N misure estratti a caso dalla popolazione — e, al limite, su tutti i campioni (aventi la stessa dimensione fissa N) che dalla popolazione è possibile ricavare?

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

ed infine

$$E(\bar{x}) = E(x) \quad (5.3)$$

cioè:

Il valore medio della popolazione delle medie aritmetiche dei campioni di dimensione finita N estratti da una popolazione coincide con il valore medio della popolazione stessa.

5.4 La varianza delle combinazioni lineari

Dimostriamo ora un altro teorema generale che riguarda la varianza di una combinazione lineare di più variabili casuali, che supporremo però *stati-*

sticamente indipendenti. Usando gli stessi simboli già introdotti nel paragrafo 5.3, e dette x ed y due variabili casuali che godano di tale proprietà, sappiamo dall'equazione (3.4) che la probabilità P_{jk} che contemporaneamente risulti sia $x = x_j$ che $y = y_k$ 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 *speranza matematica nulla*; estenderemo poi il risultato a due variabili (sempre statisticamente indipendenti) aventi speranza matematica qualunque. Ciò premesso, la combinazione lineare

$$z = ax + by$$

ha anch'essa speranza matematica zero: infatti applicando l'equazione (5.2) risulta

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

e si può allora ricavare (indicando con i simboli σ_x^2 , σ_y^2 e σ_z^2 le varianze di x , y e z rispettivamente):

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

ed infine

$$\sigma_z^2 = a^2 \sigma_x^2 + b^2 \sigma_y^2 . \quad (5.4)$$

Allo scopo di estendere la validità dell'equazione (5.4) appena dimostrata a due variabili casuali x e y aventi speranza matematica anche differente 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:

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

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 speranza matematica zero): pertanto la loro combinazione lineare $\zeta = a\xi + b\eta$, che differisce anch'essa da z per un fattore costante e pari ad $aE(x) + bE(y)$, avrà varianza che, in conseguenza della (5.4), sarà data dalla

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

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 (purché però statisticamente indipendenti) vale la relazione (5.4), che possiamo enunciare nel modo seguente:

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

È ovvio poi estendere (per induzione completa) questo risultato alla combinazione lineare di un numero finito qualsivoglia di variabili casuali, che siano però sempre tra loro tutte statisticamente indipendenti: se

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

¹O, come si usa dire in sintesi, *gli errori si combinano quadraticamente*. Una formula più generale, che si può applicare a coppie di variabili casuali qualunque, verrà dimostrata nell'appendice C.

allora

$$\sigma_F^2 = a^2 \sigma_x^2 + b^2 \sigma_y^2 + c^2 \sigma_z^2 + \dots \quad (5.5)$$

5.5 L'errore della media dei campioni

Torniamo ora ad occuparci dello studio delle proprietà statistiche della media aritmetica di un campione di N misure indipendenti estratto da una popolazione,

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

e cerchiamo in particolare di determinarne la varianza. Applicando l'equazione (5.5) appena dimostrata, risulta

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

ed infine

$$\boxed{\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{N}} \quad (5.6)$$

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.*

e conseguentemente

- *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 La legge dei grandi numeri

Le relazioni (5.3) e (5.6) sono state dimostrate sulla base della definizione di speranza matematica, e senza presupporre la convergenza verso di essa della media dei campioni (né quella delle frequenze verso la probabilità); vediamo ora come la legge dei grandi numeri (cui abbiamo già accennato nel paragrafo 3.5) si possa da esse dedurre.

5.6.1 La disuguaglianza di Bienaymé-Čebyšef

Sia una variabile casuale x , e siano $E(x)$ e σ^2 la speranza matematica e la varianza della sua popolazione; vogliamo ora determinare la probabilità che un valore di x scelto a caso differisca (in valore assoluto) da $E(x)$ per più di una assegnata quantità (positiva) ϵ . Questa è ovviamente data, in base alla legge della probabilità totale (3.2), dalla

$$\Pr(|x - E(x)| \geq \epsilon) = \sum_{|x_i - E(x)| \geq \epsilon} p_i$$

dove la sommatoria è estesa solo a quei valori x_i che soddisfano a tale condizione. Ora, sappiamo che

$$\sigma^2 = E\{[x - E(x)]^2\} = \sum_i p_i [x_i - E(x)]^2 ;$$

se si restringe la sommatoria ai soli termini x_i che differiscono (in modulo) da $E(x)$ per più di ϵ , il suo valore diminuirà o, al massimo, rimarrà invariato: deve risultare insomma

$$\sigma^2 \geq \sum_{|x_i - E(x)| \geq \epsilon} p_i [x_i - E(x)]^2 \geq \sum_{|x_i - E(x)| \geq \epsilon} p_i \epsilon^2 = \epsilon^2 \sum_{|x_i - E(x)| \geq \epsilon} p_i$$

e da questa relazione si ottiene la *disuguaglianza di Bienaymé-Čebyšef*²

$$\Pr(|x - E(x)| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \quad (5.7)$$

e, se si pone $\epsilon = k \sigma$,

$$\Pr(|x - E(x)| \geq k \sigma) \leq \frac{1}{k^2} \quad (5.8)$$

²Irénée-Jules Bienaymé, francese, fu un matematico e statistico vissuto dal 1796 al 1878; Pafnuty Lvovič Čebyšef, matematico russo vissuto dal 1821 al 1894, si occupò di analisi, teoria dei numeri, probabilità, meccanica razionale e topologia.

(se nella dimostrazione si sostituissero le frequenze relative alle probabilità e la media aritmetica ad $E(x)$, si troverebbe che una analoga relazione vale anche per ogni campione di valori sperimentali x_i rispetto alla media aritmetica \bar{x} ed alla varianza del campione s^2).

La (5.8) fissa un *limite superiore* per la probabilità esaminata, limite che deve valere per *qualsiasi* variabile casuale; con $k \leq 1$ non si ottiene alcuna informazione significativa da essa, ma con $k > 1$ si vede che il maggiorante della probabilità tende a zero all'aumentare di k . In particolare, per *qualsiasi* variabile casuale la probabilità di uno scarto dal valore medio non inferiore in valore assoluto a 2σ non può superare $\frac{1}{4} = 25\%$; e quella di uno scarto non inferiore in valore assoluto a 3σ non può superare $\frac{1}{9} \approx 11.1\%$.

Si deve notare che non si è fatta alcuna ipotesi sulla distribuzione, a parte l'esistenza della sua varianza σ^2 e della sua speranza matematica $E(x)$; in termini così generali il limite superiore (5.8) non può essere ridotto, ma non è escluso che (per una particolare distribuzione) la probabilità per la variabile da essa descritta di differire dal suo valore medio sia più piccola ancora di quella fissata dalla disuguaglianza di Bienaymé-Čebyšef. Ad esempio, se esiste finita la quantità

$$\mu_4 = E \{ [x - E(x)]^4 \}$$

(momento del quarto ordine rispetto alla media), con passaggi analoghi si troverebbe che

$$\Pr \{ [x - E(x)]^4 \geq \epsilon \} \leq \frac{\mu_4}{\epsilon^4}$$

e, quindi, che

$$\Pr \{ [x - E(x)]^4 \geq k \sigma^4 \} \leq \frac{\mu_4}{k^4 \sigma^4} .$$

Imponendo altre condizioni (anche non molto restrittive) alla distribuzione di probabilità, si potrebbe ridurre ulteriormente (in quantità anche notevole) il limite superiore stabilito in generale dalla (5.8); e stimare così anche la probabilità di uno scarto della variabile casuale dal suo valore medio inferiore a σ . Risale ad esempio a Gauss (1821) la dimostrazione che per una variabile continua avente distribuzione unimodale (con massimo in x_0), e per la quale esista finita la quantità $\sigma_0^2 = E[(x - x_0)^2]$, la probabilità di uno scarto dalla moda x_0 non inferiore in valore assoluto ad una quantità prefissata non può superare la frazione $\frac{4}{9}$ del limite di Bienaymé-Čebyšef:

$$\Pr \{ |x - x_0| \geq k \sigma \} \leq \frac{4}{9 k^2} .$$

Se la distribuzione è anche *simmetrica*, moda e media coincidono entrambe col centro di simmetria; e σ_0 è uguale alla deviazione standard σ . Per distribuzioni di questo genere, quindi, il limite superiore per la probabilità di uno scarto che non sia inferiore a k volte l'errore quadratico medio scende a $\frac{4}{9} \approx 44.4\%$ per $k = 1$; a $\frac{1}{9} \approx 11.1\%$ per $k = 2$; ed a $\frac{4}{81} \approx 4.9\%$ per $k = 3$ (e vedremo poi nel paragrafo 9.3 che per le misure affette da errori puramente casuali i limiti superiori sono ancora più stringenti di questi).

5.6.2 Il teorema di Čebyšef

Adesso applichiamo la (5.7) alla variabile casuale \bar{x} , media aritmetica di un campione di dimensione N di valori che supponiamo essere statisticamente indipendenti:

$$\Pr(|\bar{x} - E(\bar{x})| \geq \epsilon) \leq \frac{\sigma_{\bar{x}}^2}{\epsilon^2}; \quad (5.9)$$

ma valendo, per questa variabile casuale, le

$$E(\bar{x}) = E(x) \quad \text{e} \quad \text{Var}(\bar{x}) = \frac{\sigma^2}{N},$$

sostituendo nella (5.9) otteniamo

$$\Pr(|\bar{x} - E(x)| \geq \epsilon) \leq \frac{\sigma^2}{N\epsilon^2}. \quad (5.10)$$

Ora, scelti comunque due numeri positivi ϵ e δ , si può trovare in conseguenza un valore di N per cui il secondo membro della (5.10) risulti sicuramente minore di δ : basta prendere $N > M = \lceil \sigma^2 / (\delta \epsilon^2) \rceil$. In base alla definizione (3.8), questo significa che vale il

TEOREMA (DI ČEBYŠEF): *il valore medio di un campione finito di valori di una variabile casuale qualunque converge statisticamente, all'aumentare della dimensione del campione, alla speranza matematica per quella variabile.*

5.6.3 Il teorema di Bernoulli

Sia un qualsiasi evento casuale E avente probabilità p di verificarsi; indichiamo con $q = 1 - p$ la probabilità del non verificarsi di E (cioè la probabilità dell'evento complementare \bar{E}).

Consideriamo poi un insieme di N prove nelle quali si osserva se E si è o no verificato; ed introduciamo una variabile casuale y , definita come il

numero di volte in cui E si è verificato in una di tali prove. Ovviamente y può assumere i due soli valori 1 (con probabilità p) e 0 (con probabilità q); la sua speranza matematica è perciò data da

$$E(y) = 1 \cdot p + 0 \cdot q = p . \quad (5.11)$$

La frequenza relativa f dell'evento E nelle N prove si può chiaramente esprimere (indicando con y_i il valore assunto dalla variabile casuale y nella i -esima di esse) come

$$f = \frac{1}{N} \sum_{i=1}^N y_i ,$$

ossia è data *dal valore medio della y sul campione di prove*, \bar{y} ; ma quest'ultimo (per il teorema di Čebyšef³) deve convergere statisticamente, all'aumentare di N , alla speranza matematica per y : che vale proprio p . Riassumendo, abbiamo così dimostrato il

TEOREMA (DI BERNOULLI, O LEGGE “DEI GRANDI NUMERI”): *la frequenza relativa di qualunque evento casuale converge (statisticamente) alla sua probabilità all'aumentare del numero delle prove.*

5.7 Valore medio e valore vero

Anche se non ci basiamo sulla definizione empirica di probabilità, ma su quella assiomatica, possiamo ora presupporre la convergenza della media aritmetica dei campioni di misure alla speranza matematica della grandezza misurata, che ora a buon diritto possiamo chiamare “valore medio del risultato della misura sull'intera popolazione”.

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 valore vero, avranno valore medio nullo; mentre errori sistematici causeranno invece per definizione una differenza tra il valore medio delle misure $E(x)$ ed il valore vero. In assenza di errori sistematici assumiamo allora che valore medio e valore vero coincidano: ammettiamo insomma (lo proveremo più avanti per la distribuzione normale) che in tal

³Il teorema di Čebyšef vale per tutte le variabili casuali per le quali esistano sia la speranza matematica che la varianza: la prima è espressa dall'equazione (5.11), la seconda sarà ricavata più tardi nell'equazione (8.8) a pagina 109.

caso $E(x)$ esista e sia uguale a x^* . Sappiamo insomma che risulta

$$\begin{aligned}\bar{x} &= \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{N} \sum_{j=1}^M v_j x_j = \sum_{j=1}^M f_j x_j \\ \lim_{N \rightarrow \infty} \bar{x} &\equiv E(x) = \sum_j p_j x_j\end{aligned}$$

e postuliamo che

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

inoltre sappiamo che anche

$$E(\bar{x}) = E(x) \equiv x^* .$$

Ossia, non solo \bar{x} converge ad $E(x)$ all'aumentare della dimensione del campione; ma, qualunque sia il valore di quest'ultima grandezza, **mediamente** \bar{x} coincide con $E(x)$. Ripetendo varie volte la misura ed ottenendo così più campioni con differenti medie aritmetiche, dando come stima di $E(x)$ la media di uno dei nostri campioni avremo insomma la stessa probabilità di sbagliare per difetto o per eccesso⁴.

5.8 Scarto ed errore quadratico medio

L'ultimo punto da approfondire riguarda la relazione tra la varianza s^2 di un campione di N misure e quella σ^2 della popolazione da cui il campione proviene. Ora, s^2 si può esprimere come

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

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

⁴Questo nella terminologia statistica si esprime dicendo che la media dei campioni è una stima *imparziale* della media della popolazione; al contrario della varianza del campione che, come vedremo nel prossimo paragrafo, è una stima *parziale* (o *distorta*) della varianza della popolazione (il concetto verrà poi approfondito nel paragrafo 11.1).

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^N (x_i - x^*)^2 &= \frac{1}{N} \sum_{i=1}^N [(x_i - \bar{x}) + (\bar{x} - x^*)]^2 \\
&= \frac{1}{N} \left[\sum_{i=1}^N (x_i - \bar{x})^2 + \sum_{i=1}^N (\bar{x} - x^*)^2 + 2(\bar{x} - x^*) \sum_{i=1}^N (x_i - \bar{x}) \right] \\
&= \frac{1}{N} \left[\sum_{i=1}^N (x_i - \bar{x})^2 + N(\bar{x} - x^*)^2 \right] \\
&= s^2 + (\bar{x} - x^*)^2
\end{aligned} \tag{5.12}$$

(si è sfruttata qui l'equazione (4.2), secondo la quale la somma algebrica degli scarti delle misure dalla loro media aritmetica è identicamente nulla; vedi anche l'analoga formula (4.3) nel paragrafo 4.2.3).

Cerchiamo ora di capire come le varianze s^2 dei campioni di dimensione N siano legate all'analoga grandezza, σ^2 o $\text{Var}(x)$, definita sull'intera popolazione, e però calcolata rispetto al valore medio di essa, $E(x) = x^*$:

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

Sfruttando la relazione (5.12) in precedenza trovata, 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

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

Ricordando come il valore medio del quadrato degli scarti di una variabile (qui \bar{x}) dal suo valore medio (che è $E(\bar{x}) = E(x) = x^*$) sia per definizione la varianza della variabile stessa (che indicheremo qui come quadrato dell'errore quadratico medio $\sigma_{\bar{x}}$), si ricava infine:

$$E(s^2) = \sigma^2 - \sigma_{\bar{x}}^2 < \sigma^2 . \tag{5.13}$$

Insomma:

- Il valore medio della varianza s^2 di un campione è sistematicamente inferiore all'analoga 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.

5.9 Stima della varianza della popolazione

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. Riprendiamo l'equazione (5.13); in essa abbiamo già dimostrato che

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

e sappiamo dalla (5.6) che la varianza della media del campione vale

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

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 *imparziale* (ossia *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 il cui valore medio su infiniti campioni risulta proprio σ^2 .

5.10 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

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

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 valore 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 (valore medio, su tutti gli infiniti campioni, del quadrato degli scarti della media aritmetica del campione dal proprio valore vero) converge alla varianza delle medie dei campioni di N misure $\sigma_{\bar{x}}^2$.

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

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

Ora, avendo già dimostrato che

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

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 (≥ 2).

Capitolo 6

Variabili casuali unidimensionali continue

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

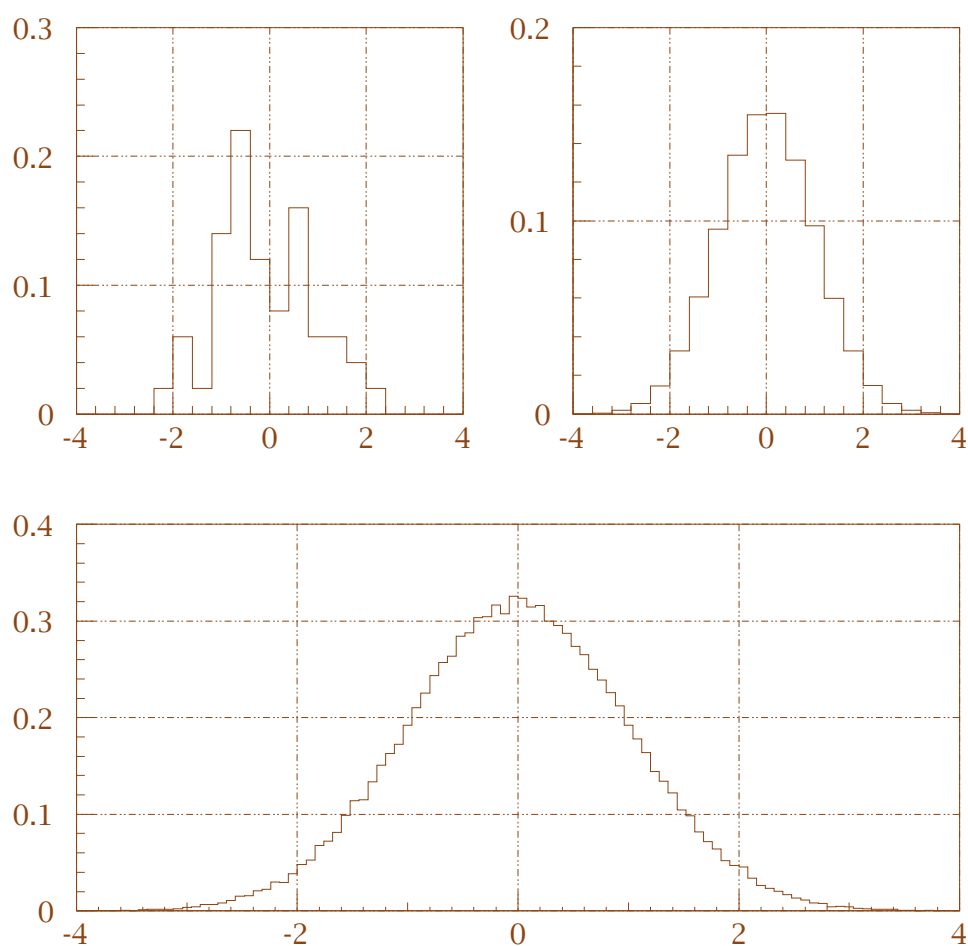
6.1 La densità di probabilità

Definiamo arbitrariamente delle classi di frequenza, suddividendo l'asse delle x in intervalli di ampiezze che, per semplicità, supponiamo siano tutte uguali; 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 relativa¹ con cui una misura è caduta 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 uno.

¹Non vi è alcuna differenza nell'usare frequenze relative o assolute: essendo esse proporzionali l'una all'altra, l'aspetto dell'istogramma è il medesimo — cambia solo la scala dell'asse delle ordinate.

FIGURA 6a - 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.

Frequenze relative



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.

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 dp ed ampiezza dx 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

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

e 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 (sempre 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 estremamente piccola — ed ha interesse soltanto la probabilità che venga osservato uno tra i molti possibili valori della x che cadono 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 di ampiezza dx , gli eventi casuali consistenti nell'appartenere il risultato della misura ad una delle classi di frequenza relative sono mutualmente esclusivi; di conseguenza, vista l'equazione (3.2), 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 $f(x)$ rispetto ad x nell'intervallo $[x_1, x_2]$.

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

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

e, in definitiva:

Per le variabili continue non si può parlare di probabilità attraverso le definizioni già esaminate. È 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 finito prefissato: questa è data semplicemente dall’area sottesa dalla curva nell’intervallo in questione.

Analogamente al concetto sperimentale di frequenza cumulativa, introdotto a pagina 33 nel paragrafo 4.1, si può definire la *funzione di distribuzione* per una variabile continua x come

$$F(x) = \int_{-\infty}^x f(t) dt .$$

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

CONDIZIONE DI NORMALIZZAZIONE: *l’integrale di una qualunque funzione che rappresenti una densità di probabilità, nell’intervallo $[-\infty, +\infty]$ vale 1.*

$$\int_{-\infty}^{+\infty} f(x) dx = 1 . \quad (6.1)$$

È da enfatizzare come il solo fatto che valga la condizione di normalizzazione, ossia che converga l’integrale (6.1), è sufficiente a garantire che una qualsiasi funzione che rappresenti una densità di probabilità *debba tendere a zero* quando la variabile indipendente tende a più o meno infinito; e questo senza alcun riferimento alla particolare natura del fenomeno casuale cui essa è collegata. Questo non è sorprendente, visto che la disuguaglianza (5.7) di Bienaymé-Čebyšef implica che a distanze via via crescenti dal valore medio di una qualsiasi variabile casuale corrispondano probabilità via via decrescenti, e che si annullano asintoticamente.

Al lettore attento non sarà sfuggito il fatto che, per introdurre il concetto di densità di probabilità, ci si è ancora una volta basati sul risultato di un esperimento reale (l’istogramma delle frequenze relative in un campione); e si è ipotizzato poi che la rappresentazione di tale esperimento si comporti in un determinato modo quando alcuni parametri (il numero di misure e la sensibilità sperimentale) vengono fatti tendere a limiti che, nella pratica, sono irraggiungibili.

Questo è in un certo senso analogo all’enfasi che abbiamo prima posto sulla definizione empirica della probabilità, in quanto più vicina all’esperienza reale di una definizione totalmente astratta come quella assiomatica; per un matematico la densità di probabilità di una variabile casuale continua è

invece definita semplicemente come una funzione non negativa, integrabile su tutto l'asse reale e che obbedisca alla condizione di normalizzazione. Il passo successivo consiste nell'associare ad ogni intervallo infinitesimo dx la quantità $dp = f(x)dx$, e ad ogni intervallo finito $[x_1, x_2]$ il corrispondente integrale: integrale che, come si può facilmente controllare, soddisfa la definizione assiomatica di probabilità.

6.2 La speranza matematica per le variabili continue

Possiamo ora determinare l'espressione della speranza matematica di una generica variabile casuale continua x ; questa grandezza, che avevamo già definito nell'equazione (5.1) come

$$E(x) = \sum_i p_i x_i$$

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

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

dove per $f(x)$ si intende la funzione densità di probabilità della variabile casuale x .

Per ricavare questa formula, basta pensare di aver suddiviso l'asse delle x in un numero grandissimo di intervalli estremamente piccoli di ampiezza dx , ad ognuno dei quali è associata una probabilità anch'essa estremamente piccola che vale $dp = f(x) dx$; e sostituire poi nella formula per variabili discrete. In base al teorema di pagina 57 (il teorema di Čebyšef), le medie aritmetiche dei campioni finiti di valori della grandezza x tendono proprio a questo $E(x)$ all'aumentare indefinito di N .

La speranza matematica di una qualsiasi grandezza $W(x)$ funzione della variabile casuale x sarà poi

$$E[W(x)] = \int_{-\infty}^{+\infty} W(x) \cdot f(x) dx . \quad (6.2)$$

6.3 I momenti

Per qualunque variabile casuale x si possono definire, sempre sulla popolazione, i cosiddetti *momenti*: il momento di ordine k rispetto all'origine,

λ_k , è la speranza matematica di x^k ; ed il momento di ordine k rispetto alla media, μ_k , è la speranza matematica di $[x - E(x)]^k$. In formula (con ovvio significato dei simboli):

$$\lambda_k = E(x^k) = \sum_i p_i x_i^k$$

e

$$\mu_k = E\{[x - E(x)]^k\} = \sum_i p_i [x_i - E(x)]^k$$

per una variabile discreta (analogamente, usando le frequenze, si possono definire i momenti rispetto all'origine ed alla media aritmetica di un campione); oppure

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

e

$$\mu_k = E\{[x - E(x)]^k\} = \int_{-\infty}^{+\infty} [x - E(x)]^k f(x) dx$$

per una variabile continua.

Chiaramente, se la popolazione è costituita da un numero infinito di elementi (quindi, in particolare, per le variabili continue), non è detto che i momenti esistano; inoltre $E(x) \equiv \lambda_1$ e $\text{Var}(x) \equiv \mu_2 \equiv \lambda_2 - \lambda_1^2$. Dalla definizione consegue immediatamente che, per qualsiasi popolazione per cui esista $E(x)$,

$$\begin{aligned} \mu_1 &= \int_{-\infty}^{+\infty} [x - E(x)] f(x) dx \\ &= \int_{-\infty}^{+\infty} x f(x) dx - E(x) \int_{-\infty}^{+\infty} f(x) dx \\ &= E(x) - E(x) \\ &\equiv 0 . \end{aligned}$$

È poi facile dimostrare che, per popolazioni simmetriche rispetto alla media, tutti i momenti di ordine dispari rispetto ad essa, se esistono, valgono zero: basta considerare come, negli integrali, i contributi infinitesimi di ognuno degli intervallini si possano associare a due a due in modo che si annullino vicendevolmente. Il valore del momento del terzo ordine rispetto alla media aritmetica può quindi essere considerato una sorta di misura dell'asimmetria di una distribuzione.

In pratica però si preferisce usare, in luogo di μ_3 , un parametro adimensionale; definendo il cosiddetto *coefficiente di asimmetria* (o *skewness*, in inglese) come

$$\gamma_1 = \frac{\mu_3}{(\sqrt{\mu_2})^3} = \frac{\mu_3}{\sigma^3}$$

(dove $\sigma = \sqrt{\mu_2}$ è la radice quadrata della varianza); γ_1 è nullo per densità di probabilità simmetriche rispetto alla media, oppure ha segno positivo (o negativo) a seconda che i valori della funzione di frequenza per la variabile casuale in questione si trovino “sbilanciati” verso la destra (o verso la sinistra) rispetto al valore medio.

Dal momento del quarto ordine rispetto alla media si può ricavare un altro parametro adimensionale talvolta usato per caratterizzare una distribuzione: il *coefficiente di curtosi* γ'_2 , definito come

$$\gamma'_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4} \quad (6.3)$$

e che è ovviamente sempre positivo. Esso misura in un certo senso la “rapidità” con cui una distribuzione di probabilità converge a zero quando ci si allontana dalla zona centrale in cui essa assume i valori più alti (individuata dal valore di $E(x) \equiv \lambda_1$): o, se si preferisce, l'importanza delle sue “code” laterali; infatti, quanto più rapidamente la funzione converge a zero in queste code, tanto più piccolo sarà il valore di γ'_2 . Come si potrebbe ricavare integrandone la funzione di frequenza (che troveremo più avanti nel paragrafo 8.2), il coefficiente di curtosi della distribuzione normale calcolato usando la (6.3) vale 3; per questo motivo si preferisce generalmente definirlo in modo differente, usando la

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 .$$

Questo fa sì che esso valga zero per la funzione di Gauss, e che assuma poi valori di segno negativo o positivo per funzioni che convergano a zero nelle code in maniera rispettivamente più “rapida” o più “lenta” della distribuzione normale.

6.4 Funzione generatrice e funzione caratteristica

La speranza matematica della funzione e^{tx} per una variabile casuale continua x prende il nome di *funzione generatrice dei momenti* della variabile stessa; la indicheremo nel seguito col simbolo $M_x(t)$. Il motivo del nome è

che risulta, indicando con $f(x)$ la densità di probabilità di x :

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx \quad (6.4)$$

(per una variabile continua, oppure

$$M_x(t) = \sum_i p_i e^{tx_i}$$

per una variabile discreta); e, ricordando sia lo sviluppo in serie di McLaurin della funzione esponenziale

$$e^{tx} = \sum_{k=0}^{\infty} \frac{(tx)^k}{k!}$$

che la definizione dei momenti rispetto all'origine, *se questi esistono tutti fino a qualsiasi ordine* risulta anche

$$M_x(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \lambda_k$$

da cui

$$\left. \frac{d^k M_x(t)}{dt^k} \right|_{t=0} = \lambda_k$$

e, in definitiva, derivando successivamente la funzione generatrice si possono ricavare tutti i momenti della funzione di frequenza da cui essa discende. Se interessa invece uno dei momenti non rispetto all'origine, ma rispetto al valore medio λ , basta considerare l'altra funzione

$$\bar{M}_x(t) = E[e^{t(x-\lambda)}] = e^{-t\lambda} M_x(t) \quad (6.5)$$

e si trova facilmente che risulta

$$\left. \frac{d^k \bar{M}_x(t)}{dt^k} \right|_{t=0} = \mu_k .$$

La speranza matematica della funzione e^{itx} si chiama invece *funzione caratteristica* della variabile casuale x , e si indica con $\phi_x(t)$:

$$\phi_x(t) = E(e^{itx}) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx \quad (6.6)$$

(per una variabile continua, e

$$\phi_x(t) = \sum_k p_k e^{itx_k} \quad (6.7)$$

per una variabile discreta); e, se esistono i momenti di qualsiasi ordine rispetto all'origine, risulta anche

$$\phi_x(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \lambda_k \quad (6.8)$$

dalla quale si ricava

$$\left. \frac{d^k \phi_x(t)}{dt^k} \right|_{t=0} = i^k \lambda_k . \quad (6.9)$$

Queste funzioni sono importanti in virtù di una serie di teoremi, che citeremo qui senza dimostrarli:

- I momenti (se esistono fino a qualunque ordine) caratterizzano univocamente una variabile casuale; se due variabili casuali hanno gli stessi momenti fino a qualsiasi ordine, la loro densità di probabilità è identica.
- La funzione generatrice esiste solo se esistono i momenti fino a qualsiasi ordine; e anch'essa caratterizza univocamente una variabile casuale, nel senso che se due variabili hanno la stessa funzione generatrice la loro densità di probabilità è identica.
- La $\phi_x(t)$ prima definita si chiama anche *trasformata di Fourier* della funzione $f(x)$; anch'essa caratterizza univocamente una variabile casuale nel senso su detto. Le proprietà che contraddistinguono una funzione che rappresenti una densità di probabilità implicano poi che la funzione caratteristica, a differenza della funzione generatrice dei momenti, *esista sempre* per qualsiasi variabile casuale; la (6.9) è però valida solo se i momenti esistono fino a qualsiasi ordine. Inoltre, se è nota la $\phi_x(t)$, *la si può sempre invertire* (riottenendo da essa la f) attraverso la

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ixt} \phi_x(t) dt \quad (6.10)$$

(*trasformata inversa di Fourier*).

Vogliamo infine ricavare una relazione che ci sarà utile più avanti: siano le N variabili casuali continue x_k (che supponiamo tutte statisticamente indipendenti tra loro), ognuna delle quali sia associata ad una particolare funzione caratteristica $\phi_k(t)$; il problema che vogliamo affrontare consiste nel determinare la funzione caratteristica della nuova variabile casuale S , definita come loro somma:

$$S = \sum_{k=1}^N x_k .$$

Il valore di ogni x_k sarà univocamente definito dai possibili risultati di un qualche evento casuale E_k ; per cui la S si può pensare univocamente definita dalle possibili *associazioni* di tutti i risultati di questi N eventi — associazioni che, in sostanza, corrispondono alle possibili posizioni di un punto in uno spazio cartesiano N -dimensionale, in cui ognuna delle variabili x_k sia rappresentata su uno degli assi.

Visto che i valori x_k sono (per ipotesi) tra loro tutti statisticamente indipendenti, la probabilità di ottenere una particolare N -pla è data dal prodotto delle probabilità relative ad ogni singolo valore: e, se indichiamo con $f_k(x_k)$ la funzione densità di probabilità della generica x_k , la probabilità di ottenere un determinato valore per la S è data da

$$dP \equiv g(S) dS = \prod_{k=1}^N f_k(x_k) dx_k$$

(dS rappresenta un intorno (ipercubico) infinitesimo del punto S , di coordinate cartesiane $\{x_k\}$ nello spazio N -dimensionale prima descritto, corrispondente agli N intorni unidimensionali dx_k dei valori assunti dalle N variabili casuali x_k); e la densità di probabilità per la S vale quindi

$$g(S) = \prod_{k=1}^N f_k(x_k) .$$

La funzione caratteristica di S è, dall'equazione di definizione (6.6),

$$\begin{aligned} \phi_S(t) &= \int_{-\infty}^{+\infty} e^{itS} g(S) dS \\ &= \int_{-\infty}^{+\infty} \prod_{k=1}^N e^{itx_k} f_k(x_k) dx_k \end{aligned}$$

ed infine

$$\phi_S(t) = \prod_{k=1}^N \phi_k(t)$$

(6.11)

Quindi la funzione caratteristica della somma di N variabili casuali statisticamente indipendenti è pari al prodotto delle loro funzioni caratteristiche.

6.4.1 Funzioni caratteristiche di variabili discrete

Invece della funzione caratteristica definita attraverso la (6.7), e che è una funzione complessa di variabile reale, talvolta, per variabili casuali discrete, viene usata una rappresentazione equivalente ricorrendo alla variabile complessa

$$z = e^{it} .$$

Sostituendo questa definizione di z nella (6.7) si ottiene la *funzione caratteristica di variabile complessa*

$$\phi_x(z) = \sum_k p_k z^{x_k} = E(z^x) ,$$

che ha proprietà analoghe a quelle della funzione caratteristica di variabile reale $\phi_x(t)$. In particolare, definendo una variabile casuale w come somma di due altre variabili x e y discrete e tra loro indipendenti, la funzione caratteristica di variabile complessa $\phi_w(z)$ è ancora il prodotto delle due funzioni caratteristiche $\phi_x(z)$ e $\phi_y(z)$: infatti

$$\begin{aligned} \phi_w(z) &= \sum_{jk} \Pr(x_j) \Pr(y_k) z^{(x_j+y_k)} \\ &= \sum_j \Pr(x_j) z^{x_j} \cdot \sum_k \Pr(y_k) z^{y_k} \\ &= \phi_x(z) \cdot \phi_y(z) ; \end{aligned}$$

e, generalizzando per induzione completa, la somma S di un numero prefissato N di variabili casuali discrete e tutte tra loro indipendenti

$$S = \sum_{k=1}^N x_k$$

è anch'essa associata alla funzione caratteristica di variabile complessa

$$\phi_S(z) = \prod_{k=1}^N \phi_{x_k}(z) .$$

Nel caso particolare, poi, in cui le N variabili provengano dalla stessa popolazione,

$$\phi_S(z) = [\phi_x(z)]^N . \quad (6.12)$$

Cosa accade se il numero N di variabili casuali da sommare non è costante, *ma è anch'esso una variabile casuale* (ovviamente discreta)? In altre parole, vogliamo qui di seguito trovare la rappresentazione analitica della funzione caratteristica della *somma di un numero casuale di variabili casuali discrete, indipendenti ed aventi tutte la stessa distribuzione*. Supponiamo che la N sia associata ad una funzione caratteristica

$$\phi_N(z) = E(z^N) = \sum_N \Pr(N) z^N ; \quad (6.13)$$

la probabilità di ottenere un determinato valore per la S vale

$$\Pr(S) = \sum_N \Pr(N) \Pr(S|N)$$

e di conseguenza la funzione caratteristica di variabile complessa associata alla S che, per definizione, è data dalla

$$\phi_S(z) = E(z^S) = \sum_S \Pr(S) z^S$$

si potrà scrivere anche

$$\begin{aligned} \phi_S(z) &= \sum_S z^S \cdot \sum_N \Pr(N) \Pr(S|N) \\ &= \sum_N \Pr(N) \cdot \sum_S \Pr(S|N) z^S \\ &= \sum_N \Pr(N) \cdot [\phi_x(z)]^N . \end{aligned}$$

Nell'ultimo passaggio si è sfruttato il fatto che la sommatoria su S rappresenta la speranza matematica di z^S *condizionata dall'aver assunto N un determinato valore*; rappresenta quindi la funzione caratteristica della S quando N ha un valore costante prefissato, che appunto è data dalla (6.12).

Ricordando poi la (6.13), la funzione caratteristica cercata è infine data dalla *funzione di funzione*

$$\boxed{\phi_S(z) = \phi_N[\phi_x(z)]} \quad (6.14)$$

È immediato riconoscere che, se N non è propriamente una variabile casuale e può assumere un unico valore N_0 , essendo tutte le $\Pr(N)$ nulle meno $\Pr(N_0) = 1$,

$$\phi_N(z) = z^{N_0}$$

e

$$\phi_S(z) = \phi_N[\phi_x(z)] = [\phi_x(z)]^{N_0}$$

e la (6.14) ridiventa la meno generale (6.12).

6.5 Cambiamento di variabile casuale

Supponiamo sia nota la funzione $f(x)$ densità di probabilità della variabile casuale x ; e sia y una nuova variabile casuale definita in funzione della x attraverso una qualche relazione matematica $y = y(x)$. Ci proponiamo di vedere come, da queste ipotesi, si possa ricavare la densità di probabilità $g(y)$ della nuova variabile y .

Supponiamo dapprima che la corrispondenza tra le due variabili continue sia *biunivoca*: ossia che la $y = y(x)$ sia una funzione monotona in senso stretto, crescente o decrescente, e di cui quindi esista la funzione inversa che indicheremo con $x = x(y)$; ed inoltre supponiamo che la $y(x)$ sia derivabile. Questo, dovendo risultare $y'(x) \neq 0$ in conseguenza dell'ipotesi fatta, implica che sia derivabile anche la $x(y)$ e che risulti

$$x'(y) = \frac{1}{y'[x(y)]} .$$

L'asserita biunivocità della corrispondenza tra le due variabili assicura che, se la prima è compresa in un intervallo infinitesimo di ampiezza dx centrato sul generico valore x , allora e solo allora la seconda è compresa in un intervallo di ampiezza $dy = |y'(x)| dx$ (il valore assoluto tiene conto del fatto che la $y(x)$ può essere sia crescente che decrescente) centrato attorno al valore $y = y(x)$. Questo a sua volta implica che le probabilità (infinitesime) degli eventi casuali consistenti nell'essere la x o la y appartenenti a tali intervalli debbano essere uguali: ossia che risulti

$$f(x) dx = g(y) dy = g(y) |y'(x)| dx$$

identicamente per ogni x , il che è possibile soltanto se

$$g(y) = \frac{f(x)}{|y'(x)|} = \frac{f[x(y)]}{|y'[x(y)]|} = f[x(y)] \cdot |x'(y)| . \quad (6.15)$$

Se la relazione che lega y ad x non è invece biunivoca, i ragionamenti sono più complicati e devono essere fatti tenendo conto della natura della particolare funzione in esame; ad esempio, se

$$y = x^2 \quad \text{e quindi} \quad x = \pm\sqrt{y}$$

un particolare valore per la y corrisponde a *due* eventualità (mutuamente esclusive) per la x ; perciò

$$g(y) dy = [f(-\sqrt{y}) + f(\sqrt{y})] dx$$

e quindi

$$g(y) = [f(-\sqrt{y}) + f(\sqrt{y})] \cdot |x'(y)| = \frac{f(-\sqrt{y}) + f(\sqrt{y})}{2\sqrt{y}}.$$

Per quello che riguarda la funzione generatrice dei momenti e la funzione caratteristica associate a variabili casuali definite l'una in funzione dell'altra, se ci limitiamo a considerare una *trasformazione lineare* del tipo $y = ax + b$, vale la

$$\begin{aligned} M_y(t) &= E(e^{ty}) \\ &= E[e^{t(ax+b)}] \\ &= e^{tb} E(e^{tax}) \end{aligned}$$

da cui infine ricaviamo la

$$M_y(t) = e^{tb} M_x(at) \quad (6.16)$$

per la funzione generatrice dei momenti; e potremmo ricavare l'analogia

$$\phi_y(t) = e^{itb} \phi_x(at) \quad (6.17)$$

per la funzione caratteristica (si confronti anche la funzione (6.5), prima usata per ricavare i momenti rispetto alla media, e che si può pensare ottenuta dalla (6.4) applicando alla variabile casuale una traslazione che ne porti il valore medio nell'origine).

6.6 I valori estremi di un campione

Sia x una variabile casuale continua, di cui siano note sia la funzione di frequenza $f(x)$ che la funzione di distribuzione $F(x)$; e sia disponibile un campione di dimensione N di valori *indipendenti* di questa variabile casuale. Supponiamo inoltre, una volta ottenuti tali valori, di averli disposti *in ordine crescente*: ovvero in modo che risulti $x_1 \leq x_2 \leq \dots \leq x_N$. Vogliamo qui, come esercizio, determinare la funzione di frequenza *del generico di questi*

valori ordinati, x_i : funzione che verrà nel seguito identificata dal simbolo $f_i(x)$.

Supponiamo che x_i sia compreso nell'intervallo infinitesimo $[x, x + dx]$; la scelta di un certo i divide naturalmente il campione (ordinato) in tre sottoinsiemi, ovvero:

1. x_i stesso, che può essere ottenuto (dall'insieme non ordinato dei valori originariamente a disposizione) in N maniere differenti; si sa inoltre che è compreso nell'intervallo $[x, x + dx]$ — evento, questo, che avviene con probabilità $f(x) dx$.
2. I primi $(i - 1)$ valori: questi possono essere ottenuti, dagli $N - 1$ elementi restanti dall'insieme non ordinato dei valori originari, in C_{i-1}^{N-1} modi distinti²; ognuno di essi è inoltre minore di x , e questo avviene con probabilità data da $F(x)$.
3. I residui $(N - i)$ valori: questi sono univocamente determinati dalle due scelte precedenti; inoltre ognuno di essi è maggiore di x , e questo avviene con probabilità $[1 - F(x)]$.

In definitiva, applicando i teoremi della probabilità totale e della probabilità composta, possiamo affermare che risulta

$$f_i(x) dx = N \binom{N-1}{i-1} [F(x)]^{i-1} [1 - F(x)]^{N-i} f(x) dx ; \quad (6.18)$$

in particolare, i valori estremi x_1 e x_N hanno densità di probabilità date da

$$f_1(x) = N [1 - F(x)]^{N-1} f(x)$$

e da

$$f_N(x) = N [F(x)]^{N-1} f(x) .$$

² C_K^N è il numero delle combinazioni di classe K di N oggetti; si veda in proposito il paragrafo A.6.

Capitolo 7

Variabili casuali pluridimensionali

Può avvenire che un evento casuale complesso E sia decomponibile in N eventi semplici E_i , ognuno dei quali a sua volta sia descrivibile mediante una variabile casuale x_i (che supporremo continua); le differenti modalità dell'evento E si possono allora associare univocamente alla N -pla dei valori delle x_i , ossia alla posizione di un punto in uno spazio cartesiano N -dimensionale.

7.1 Variabili casuali bidimensionali

Nel caso multidimensionale più semplice, $N = 2$, se supponiamo che la probabilità dP per la coppia di variabili casuali x ed y di trovarsi nell'intorno (infinitesimo) di un certo punto dello spazio bidimensionale sia proporzionale all'ampiezza dell'intorno stesso e dipenda dalla sua posizione, possiamo definire la *densità di probabilità* (o *funzione di frequenza congiunta*, $f(x, y)$), attraverso la

$$dP = f(x, y) dx dy ;$$

e, analogamente a quanto fatto nel caso unidimensionale, definire poi attraverso di essa altre funzioni. Ad esempio la *funzione di distribuzione congiunta*,

$$F(x, y) = \int_{-\infty}^x du \int_{-\infty}^y dv f(u, v)$$

che dà la probabilità di ottenere valori delle due variabili non superiori a quantità prefissate; le *funzioni di frequenza marginali*

$$g(x) = \int_{-\infty}^{+\infty} f(x, y) dy \quad \text{e} \quad h(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

che rappresentano la densità di probabilità di ottenere un dato valore per una delle due variabili *qualunque sia il valore assunto dall'altra*; ed infine le *funzioni di distribuzione marginali*

$$G(x) = \int_{-\infty}^x g(t) dt = F(x, +\infty) \quad \text{e} \quad H(y) = \int_{-\infty}^y h(t) dt = F(+\infty, y) .$$

La *condizione di normalizzazione* si potrà poi scrivere

$$F(+\infty, +\infty) = 1 .$$

Per un insieme di due variabili si possono poi definire le funzioni di frequenza *condizionate*, $\pi(x|y)$ e $\pi(y|x)$; esse rappresentano la densità di probabilità dei valori di una variabile quando già si conosce il valore dell'altra. Per definizione deve valere la

$$f(x, y) dx dy = g(x) dx \cdot \pi(y|x) dy = h(y) dy \cdot \pi(x|y) dx$$

per cui tra probabilità condizionate, marginali e congiunte valgono la

$$\pi(y|x) = \frac{f(x, y)}{g(x)} \quad \text{e la} \quad \pi(x|y) = \frac{f(x, y)}{h(y)} .$$

Due variabili casuali sono, come sappiamo, statisticamente indipendenti tra loro quando il fatto che una di esse abbia un determinato valore non altera le probabilità relative ai valori dell'altra: ovvero quando

$$\pi(x|y) = g(x) \quad \text{e} \quad \pi(y|x) = h(y) ; \quad (7.1)$$

e questo a sua volta implica che

$$f(x, y) = g(x) \cdot h(y) \quad (7.2)$$

Non è difficile poi, assunta vera la (7.2), giungere alla (7.1); in definitiva:

Due variabili casuali continue sono statisticamente indipendenti tra loro se e solo se la densità di probabilità congiunta è fattorizzabile nel prodotto delle funzioni marginali.

7.1.1 Momenti, funzione caratteristica e funzione generatrice

Analogamente a quanto fatto per le variabili casuali unidimensionali, in uno spazio degli eventi bidimensionale in cui rappresentiamo le due variabili $\{x, y\}$ aventi densità di probabilità congiunta $f(x, y)$, si può definire la speranza matematica (o valore medio) di una qualunque funzione $\psi(x, y)$ come

$$E[\psi(x, y)] = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \psi(x, y) f(x, y) ;$$

i momenti rispetto all'origine come

$$\lambda_{mn} = E(x^m y^n)$$

e quelli rispetto alla media come

$$\mu_{mn} = E[(x - \lambda_{10})^m (y - \lambda_{01})^n] .$$

Risulta ovviamente:

$$\lambda_{00} \equiv 1$$

$$\lambda_{10} = E(x)$$

$$\lambda_{01} = E(y)$$

$$\mu_{20} = E\{[x - E(x)]^2\} = \text{Var}(x)$$

$$\mu_{02} = E\{[y - E(y)]^2\} = \text{Var}(y)$$

$$\mu_{11} = E\{[x - E(x)][y - E(y)]\}$$

La quantità μ_{11} si chiama anche *covarianza* di x ed y ; si indica generalmente col simbolo $\text{Cov}(x, y)$, e di essa ci occuperemo più in dettaglio nell'appendice C (almeno per quel che riguarda le variabili discrete). Un'altra grandezza collegata alla covarianza è il cosiddetto *coefficiente di correlazione lineare*, che si indica col simbolo r_{xy} (o, semplicemente, con r): è definito come

$$r_{xy} = \frac{\mu_{11}}{\sqrt{\mu_{20} \mu_{02}}} = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y} ,$$

e si tratta di una grandezza adimensionale compresa, come vedremo, nell'intervallo $[-1, +1]$. Anche del coefficiente di correlazione lineare ci occuperemo estesamente più avanti, e sempre nell'appendice C.

La funzione caratteristica per due variabili, che esiste sempre, è la

$$\phi_{xy}(u, v) = E \left[e^{i(ux+vy)} \right] ;$$

se poi esistono tutti i momenti, vale anche la

$$\left. \frac{\partial^{m+n} \phi_{xy}}{\partial u^m \partial v^n} \right|_{u=0, v=0} = (i)^{m+n} \lambda_{mn} .$$

La funzione generatrice, che esiste solo se tutti i momenti esistono, è poi definita come

$$M_{xy}(u, v) = E \left[e^{(ux+vy)} \right]$$

e per essa vale la

$$\left. \frac{\partial^{m+n} M_{xy}}{\partial u^m \partial v^n} \right|_{u=0, v=0} = \lambda_{mn} .$$

7.1.2 Cambiamento di variabile casuale

Supponiamo di definire due nuove variabili casuali u e v per descrivere un evento casuale collegato a due variabili continue x ed y ; e questo attraverso due funzioni

$$u = u(x, y) \quad \text{e} \quad v = v(x, y) .$$

Se la corrispondenza tra le due coppie di variabili è biunivoca, esistono le funzioni inverse

$$x = x(u, v) \quad \text{e} \quad y = y(u, v) ;$$

se inoltre esistono anche le derivate parziali prime della x e della y rispetto alla u ed alla v , esiste anche non nullo il *determinante Jacobiano*

$$\frac{\partial(x, y)}{\partial(u, v)} = \det \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

dotato della proprietà che

$$\frac{\partial(x, y)}{\partial(u, v)} = \left[\frac{\partial(u, v)}{\partial(x, y)} \right]^{-1}$$

In tal caso, dalla richiesta di invarianza della probabilità sotto il cambiamento di variabili,

$$f(x, y) dx dy = g(u, v) du dv$$

si ottiene la funzione densità di probabilità congiunta per u e v , che è legata alla $f(x, y)$ dalla

$$g(u, v) = f[x(u, v), y(u, v)] \cdot \left| \frac{\partial(x, y)}{\partial(u, v)} \right| \quad (7.3)$$

7.1.3 Applicazione: il rapporto di due variabili casuali indipendenti

Come esempio, consideriamo due variabili casuali x ed y indipendenti tra loro e di cui si conoscano le funzioni di frequenza, rispettivamente $f(x)$ e $g(y)$; e si sappia inoltre che la y non possa essere nulla. Fatte queste ipotesi, useremo la formula precedente per calcolare la funzione di frequenza $\varphi(u)$ della variabile casuale u rapporto tra x ed y . Definite

$$u = \frac{x}{y} \quad \text{e} \quad v = y,$$

la corrispondenza tra le coppie di variabili è biunivoca; e le funzioni inverse sono la

$$x = uv \quad \text{e la} \quad y = v.$$

Le funzioni di frequenza congiunte delle due coppie di variabili sono, ricordando la (7.2) e la (7.3)

$$f(x, y) = f(x) g(y) \quad \text{e} \quad \varphi(u, v) = f(x) g(y) \left| \frac{\partial(x, y)}{\partial(u, v)} \right|$$

rispettivamente; e, calcolando le derivate parziali,

$$\frac{\partial(x, y)}{\partial(u, v)} = \det \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix}$$

per cui

$$\varphi(u, v) = f(uv) g(v) |v| .$$

In conclusione, la funzione di distribuzione della sola u (la funzione marginale) è la

$$\varphi(u) = \int_{-\infty}^{+\infty} \varphi(u, v) dv = \int_{-\infty}^{+\infty} f(uv) g(v) |v| dv \quad (7.4)$$

7.1.4 Applicazione: il decadimento debole della Λ^0

La particella elementare Λ^0 decade, attraverso processi governati dalle interazioni deboli, nei due canali

$$\Lambda^0 \rightarrow p + \pi^- \quad \text{e} \quad \Lambda^0 \rightarrow n + \pi^0 ;$$

il suo decadimento è quindi un evento casuale che può essere descritto dalle due variabili c (carica del nucleone nello stato finale, 1 o 0 rispettivamente) e t (tempo di vita della Λ^0).

La teoria (confermata dagli esperimenti) richiede che la legge di decadimento sia la stessa per entrambi gli stati finali, ovvero esponenziale con la stessa vita media τ ; e che il cosiddetto *branching ratio*, cioè il rapporto delle probabilità di decadimento nei due canali citati, sia indipendente dal tempo di vita e valga

$$\frac{\Pr(\Lambda^0 \rightarrow p + \pi^-)}{\Pr(\Lambda^0 \rightarrow n + \pi^0)} = 2 .$$

In altre parole, le probabilità marginali e condizionate per le due variabili (una discreta, l'altra continua) devono essere: per la c

$$g(1) = g(1|t) = \frac{2}{3} \quad \text{e} \quad g(0) = g(0|t) = \frac{1}{3}$$

o, in maniera compatta,

$$g(c) = g(c|t) = \frac{c+1}{3} ;$$

per il tempo di vita t ,

$$h(t) = h(t|0) = h(t|1) = \frac{1}{\tau} e^{-\frac{t}{\tau}} .$$

La probabilità congiunta delle due variabili casuali è, infine,

$$f(c, t) = g(c) \cdot h(t) = \frac{c+1}{3\tau} e^{-\frac{t}{\tau}} .$$

7.1.5 Applicazione: il decadimento debole K_{e3}^0

I decadimenti K_{e3}^0 consistono nei due processi deboli di decadimento del mesone K^0

$$K^0 \rightarrow e^- + \pi^+ + \bar{\nu}_e \quad \text{e} \quad K^0 \rightarrow e^+ + \pi^- + \nu_e ;$$

essi possono essere descritti dalle due variabili casuali c (carica dell'elettrone nello stato finale, $c = \mp 1$) e t (tempo di vita del K^0). La teoria, sulla base della cosiddetta "ipotesi $\Delta Q = \Delta S$ ", prevede che la funzione di frequenza congiunta sia

$$f(t, c) = \frac{N(t, c)}{\sum_c \int_0^{+\infty} N(t, c) dt}$$

ove si è indicato con $N(t, c)$ la funzione

$$N(t, c) = e^{-\lambda_1 t} + e^{-\lambda_2 t} + 2c \cos(\omega t) e^{-\frac{\lambda_1 + \lambda_2}{2} t} \quad ; \quad (7.5)$$

nella (7.5), le costanti λ_1 e λ_2 rappresentano gli inversi delle vite medie dei mesoni K_1^0 e K_2^0 , mentre ω corrisponde alla differenza tra le loro masse.

Si vede immediatamente che la (7.5) non è fattorizzabile: quindi le due variabili *non sono* tra loro indipendenti. In particolare, le probabilità marginali sono date dalla

$$h(t) = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} (e^{-\lambda_1 t} + e^{-\lambda_2 t})$$

e dalla

$$g(c) = \frac{1}{2} + \frac{4c \lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)^2 + 4\omega^2} \quad ;$$

mentre le probabilità condizionate sono, da definizione, la

$$h(t|c) = \frac{f(t, c)}{g(c)} \quad \text{e la} \quad g(c|t) = \frac{f(t, c)}{h(t)} .$$

7.1.6 Ancora sui valori estremi di un campione

Come esempio, e ricordando il paragrafo 6.6, calcoliamo la densità di probabilità congiunta dei due valori estremi x_1 e x_N di un campione *ordinato* e di dimensione N ; questo sotto l'ipotesi che i dati appartengano a una popolazione avente funzione di frequenza $f(x)$ e funzione di distribuzione $F(x)$ entrambe note.

x_1 può provenire dal campione a disposizione in N maniere distinte; una volta noto x_1 , poi, x_N può essere scelto in $(N - 1)$ modi diversi; e, infine, ognuno dei dati restanti è compreso tra x_1 e x_N : e questo avviene con probabilità $[F(x_N) - F(x_1)]$. Ripetendo i ragionamenti del paragrafo 6.6, si ricava

$$f(x_1, x_N) = N(N - 1) [F(x_N) - F(x_1)]^{N-2} f(x_1) f(x_N) \quad (7.6)$$

che *non è fattorizzabile*: quindi i valori minimo e massimo di un campione *non* sono indipendenti tra loro. Introducendo le variabili ausiliarie

$$\begin{cases} \xi = N \cdot F(x_1) \\ \eta = N \cdot [1 - F(x_N)] \end{cases} \quad \text{con} \quad \begin{cases} d\xi = N f(x_1) dx_1 \\ d\eta = -N f(x_N) dx_N \end{cases}$$

ed essendo $F(x_N) - F(x_1)$ identicamente uguale a $1 - [1 - F(x_N)] - F(x_1)$, dalla (7.6) si ricava

$$f(\xi, \eta) = \frac{N-1}{N} \left(1 - \frac{\xi + \eta}{N}\right)^{N-2}$$

che, ricordando il limite notevole

$$\lim_{x \rightarrow +\infty} \left(1 + \frac{k}{x}\right)^x = e^k ,$$

asintoticamente diventa

$$f(\xi, \eta) \xrightarrow{N \rightarrow \infty} e^{-(\xi+\eta)} \equiv e^{-\xi} e^{-\eta} .$$

Quindi ξ ed η (come anche di conseguenza x_1 e x_N) sono statisticamente indipendenti *solo asintoticamente*, all'aumentare indefinito della dimensione del campione.

7.2 Cenni sulle variabili casuali in più di due dimensioni

Estendendo a spazi cartesiani a più di due dimensioni il concetto di densità di probabilità, possiamo pensare di associare ad un evento casuale E descritto da N variabili continue x_1, x_2, \dots, x_N una funzione f di tutte queste variabili; la probabilità che, simultaneamente, ognuna di esse cada in un intervallo infinitesimo attorno ad un determinato valore sarà poi data da

$$dP = f(x_1, x_2, \dots, x_N) dx_1 dx_2 \cdots dx_N .$$

Usando la legge della probabilità totale e la definizione dell'operazione di integrazione, è poi immediato riconoscere che la probabilità dell'evento casuale consistente nell'essere ognuna delle x_i compresa in un determinato intervallo finito $[a_i, b_i]$ è data da

$$P = \int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} dx_2 \cdots \int_{a_N}^{b_N} dx_N \cdot f(x_1, x_2, \dots, x_N) \quad .$$

Similmente, poi, se consideriamo il sottoinsieme delle prime M variabili x_i (con $M < N$), la probabilità che ognuna di esse cada all'interno di intervallini infinitesimi attorno ad una M -pla di valori prefissati, *indipendentemente* dal valore assunto dalle altre $N - M$ variabili, è data da

$$\begin{aligned} dP &\equiv f^M(x_1, \dots, x_M) dx_1 \cdots dx_M \\ &= dx_1 \cdots dx_M \int_{-\infty}^{+\infty} dx_{M+1} \int_{-\infty}^{+\infty} dx_{M+2} \cdots \int_{-\infty}^{+\infty} dx_N \cdot f(x_1, x_2, \dots, x_N) \end{aligned}$$

dove gli integrali definiti sulle $N - M$ variabili che non interessano si intendono estesi a tutto l'asse reale; potendosi senza perdere in generalità assumere che tale sia il loro dominio di esistenza, definendo eventualmente la f come identicamente nulla al di fuori del reale intervallo di variabilità se esse fossero limitate.

La f^M definita dalla equazione precedente prende il nome di *densità di probabilità marginale* delle M variabili casuali x_1, \dots, x_M ; infine la condizione di normalizzazione si scriverà

$$\int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \cdots \int_{-\infty}^{+\infty} dx_N \cdot f(x_1, x_2, \dots, x_N) = 1 \quad .$$

Definendo, analogamente a quanto fatto nel paragrafo 7.1, la densità di probabilità delle M variabili casuali x_j (con $j = 1, 2, \dots, M$ e $M < N$) *condizionata* dai valori assunti dalle restanti $N - M$ variabili attraverso la

$$f(x_1, x_2, \dots, x_M | x_{M+1}, x_{M+2}, \dots, x_N) = \frac{f(x_1, x_2, \dots, x_N)}{f^M(x_{M+1}, x_{M+2}, \dots, x_N)} \quad (7.7)$$

il concetto di indipendenza statistica può facilmente essere generalizzato a *sottogruppi* di variabili: diremo che le M variabili x_j sono statisticamente indipendenti dalle restanti $N - M$ quando la probabilità che le x_1, x_2, \dots, x_M assumano determinati valori non dipende dai valori assunti dalle $x_{M+1}, x_{M+2}, \dots, x_N$ — e dunque quando la densità condizionata (7.7) è identicamente uguale alla densità marginale $f^M(x_1, x_2, \dots, x_M)$.

Esaminando la (7.7) si può facilmente capire come, perché questo avvenga, occorra e basti che la densità di probabilità complessiva sia *fattorizzabile* nel prodotto di due termini: il primo dei quali sia funzione solo delle prime M variabili ed il secondo dei quali dipenda soltanto dalle altre $N - M$; ovviamente ognuno dei fattori coincide con le probabilità marginali, per cui la condizione è espressa matematicamente dalla formula

$$f(x_1, \dots, x_N) = f^M(x_1, \dots, x_M) \cdot f^M(x_{M+1}, \dots, x_N)$$

e, in particolare, le variabili sono *tutte* indipendenti tra loro se e solo se risulta

$$f(x_1, x_2, \dots, x_N) = f^M(x_1) \cdot f^M(x_2) \cdots f^M(x_N) . \quad (7.8)$$

Nel caso che esista un differente insieme di N variabili y_i in grado di descrivere lo stesso fenomeno casuale E , il requisito che la probabilità di realizzarsi di un qualunque sottoinsieme dei possibili risultati (l'integrale definito, su una qualunque regione Ω dello spazio ad N dimensioni, della funzione densità di probabilità) sia invariante per il cambiamento delle variabili di integrazione, porta infine a ricavare la formula di trasformazione delle densità di probabilità per il cambiamento di variabili casuali nel caso multidimensionale:

$$f(y_1, y_2, \dots, y_N) = f(x_1, x_2, \dots, x_N) \cdot \left| \frac{\partial(x_1, x_2, \dots, x_N)}{\partial(y_1, y_2, \dots, y_N)} \right| \quad (7.9)$$

dove con il simbolo

$$\left| \frac{\partial(x_1, x_2, \dots, x_N)}{\partial(y_1, y_2, \dots, y_N)} \right|$$

si è indicato il valore assoluto del determinante Jacobiano delle x rispetto alle y :

$$\frac{\partial(x_1, x_2, \dots, x_N)}{\partial(y_1, y_2, \dots, y_N)} = \det \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_N} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_N} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial x_N}{\partial y_1} & \frac{\partial x_N}{\partial y_2} & \cdots & \frac{\partial x_N}{\partial y_N} \end{vmatrix}$$

che esiste sempre non nullo se la trasformazione tra l'insieme delle funzioni x_i e quello delle funzioni y_i è biunivoca; e che gode, sempre in questa ipotesi, della proprietà che

$$\frac{\partial(y_1, y_2, \dots, y_N)}{\partial(x_1, x_2, \dots, x_N)} = \left[\frac{\partial(x_1, x_2, \dots, x_N)}{\partial(y_1, y_2, \dots, y_N)} \right]^{-1}.$$

Capitolo 8

Esempi di distribuzioni teoriche

In questo capitolo presentiamo alcune funzioni teoriche che rappresentano densità di probabilità di variabili casuali unidimensionali (continue e discrete) che hanno importanza per la fisica.

8.1 La distribuzione uniforme

Il caso più semplice, dal punto di vista teorico, è quello di una variabile casuale x che possa assumere solo valori compresi in un intervallo finito avente estremi costanti prefissati, $[a, b]$; e ivi con probabilità uguale per ogni punto¹.

Questo implica che la densità di probabilità $f(x)$ di questa variabile debba essere definita come

$$\begin{cases} f(x) = 0 & \text{per } x < a \text{ e per } x > b; \\ f(x) = \frac{1}{b-a} = \text{cost.} & \text{per } a \leq x \leq b. \end{cases}$$

(il valore costante di $f(x)$ quando $x \in [a, b]$ è fissato dalla condizione di

¹La frase è intuitiva, ma impropria; si intende qui che la probabilità, per la variabile casuale, di cadere in un intervallino di ampiezza (infinitesima) prefissata dx e centrato su un qualsivoglia punto del dominio di definizione, ha sempre lo stesso valore.

normalizzazione). La funzione di distribuzione $F(x)$ della x è data da

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & \text{per } x < a; \\ \frac{x-a}{b-a} & \text{per } a \leq x \leq b; \\ 1 & \text{per } x > b. \end{cases}$$

I valori della media e della varianza della variabile casuale x , come si può facilmente calcolare, valgono

$$\begin{cases} E(x) = \frac{a+b}{2} \\ \text{Var}(x) = \frac{(b-a)^2}{12} \end{cases} \quad (8.1)$$

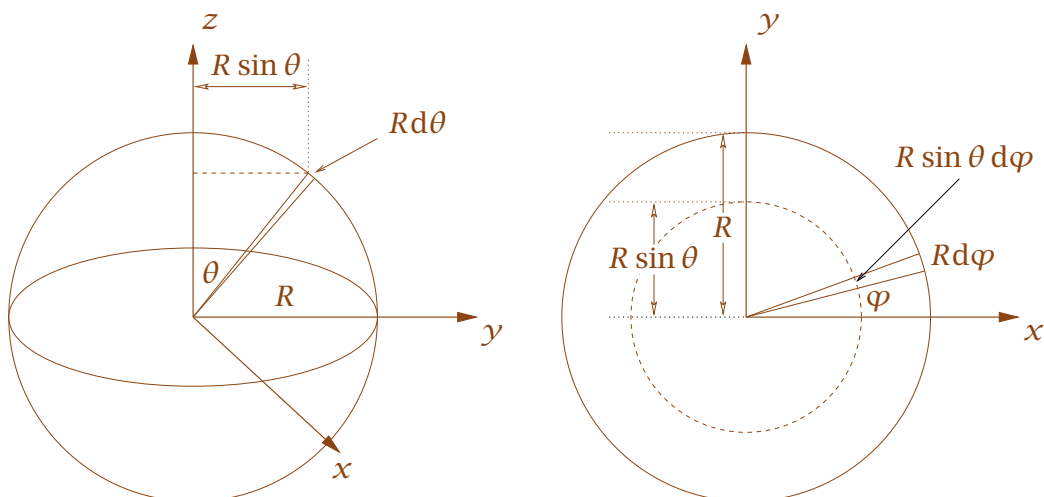
Per vedere una prima applicazione pratica della distribuzione uniforme, supponiamo di misurare una grandezza fisica usando uno strumento *digitale*: ad esempio una bilancia con sensibilità inversa di 1 grammo. Se, per semplicità, escludiamo la presenza di errori sistematici, il fatto che il display digitale indichi (ad esempio) 10 grammi significa solo che la massa dell'oggetto pesato è maggiore o uguale a questo valore e minore di 11 grammi²; e tutti i valori interni a questo intervallo ci appaiono inoltre come ugualmente plausibili. Per questo motivo, viste le (8.1), in casi di questo genere si attribuisce all'oggetto pesato una massa di 10.5 g con un errore di $1/\sqrt{12} \approx 0.3$ g.

8.1.1 Applicazione: decadimento del π^0

Esistono, nella fisica, variabili casuali che seguono la distribuzione uniforme: ad esempio, se una particella instabile non dotata di momento angolare intrinseco (come il mesone π^0), originariamente in quiete in un punto (che supporremo sia l'origine degli assi coordinati), decade, i prodotti di decadimento si distribuiscono uniformemente tra le varie direzioni possibili; sostanzialmente per motivi di simmetria, perché non esiste nessuna direzione privilegiata nel sistema di riferimento considerato (ovverosia nessuna caratteristica intrinseca del fenomeno che possa servire per definire uno, o più d'uno, degli assi coordinati).

²La maggior parte degli strumenti digitali *tronca* il valore mostrato e si comporta appunto in questo modo; altri invece *arrotondano* il risultato e, se questo fosse il caso, vorrebbe dire che la massa dell'oggetto pesato è maggiore o uguale a 9.5 g e minore di 10.5 g.

FIGURA 8a - Le aree elementari sulla superficie di una sfera di raggio R (in coordinate polari).



Con riferimento alla figura 8a, pensiamo introdotto un sistema di coordinate polari $\{R, \theta, \varphi\}$: l'elemento infinitesimo di area, dS , sulla sfera di raggio R , che corrisponde a valori della *colatitudine* compresi tra θ e $\theta + d\theta$, e dell'*azimuth* compresi tra φ e $\varphi + d\varphi$, è uno pseudorettangolo di lati $R d\theta$ ed $R \sin \theta d\varphi$; quindi, a meno del segno,

$$|dS| = R^2 \sin \theta d\theta d\varphi = -R^2 d(\cos \theta) d\varphi$$

mentre l'angolo solido corrispondente vale

$$d\Omega = \frac{|dS|}{R^2} = \sin \theta d\theta d\varphi = -d(\cos \theta) d\varphi .$$

L'asserita uniformità nell'emissione dei prodotti di decadimento si traduce nella condizione che la probabilità, per essi, di essere contenuti in un qualsiasi angolo solido, sia proporzionale all'ampiezza di quest'ultimo:

$$dP = K d\Omega = K' d(\cos \theta) d\varphi$$

(ove K e K' sono due opportune costanti); ovvero si richiede che le due variabili casuali

$$u = \cos \theta$$

e

$$v = \varphi$$

abbiano distribuzione uniforme, e siano inoltre statisticamente indipendenti tra loro (questo in conseguenza dell'equazione (7.8)).

8.1.2 Applicazione: generazione di numeri casuali con distribuzione data

Supponiamo che la variabile casuale x abbia densità di probabilità $f(x)$ e funzione di distribuzione $F(x)$: vogliamo ora dimostrare che la variabile casuale $y = F(x)$ è distribuita uniformemente nell'intervallo $[0, 1]$ qualunque siano $f(x)$ e $F(x)$. Chiaramente y può appartenere solo a tale intervallo; ed inoltre, essendo funzione integrale di $f(x)$, è dotata della proprietà di essere continua e derivabile in tutto l'insieme di definizione e con derivata prima data da

$$y' = F'(x) = f(x)$$

così che, ricordando l'equazione (6.15), la densità di probabilità della nuova variabile y è data (ove $f(x)$ non sia nulla) dalla

$$g(y) = \frac{f(x)}{y'(x)} = \frac{f(x)}{f(x)} \equiv 1$$

come volevamo dimostrare.

Supponiamo sia nota la densità di probabilità $f(x)$ di una qualche variabile casuale x ; e che si vogliano ottenere dei numeri che si presentino secondo una legge di probabilità data appunto da questa $f(x)$. I moderni calcolatori numerici sono in grado di generare sequenze di numeri casuali³ che hanno distribuzione uniforme in un intervallo dipendente dall'implementazione dell'algoritmo, e che possono a loro volta essere usati per produrre numeri casuali con distribuzione uniforme nell'intervallo $[0, 1]$; se y è uno di tali numeri, e se si è in grado di invertire, numericamente od analiticamente, la funzione di distribuzione $F(x)$ della variabile casuale x , i numeri

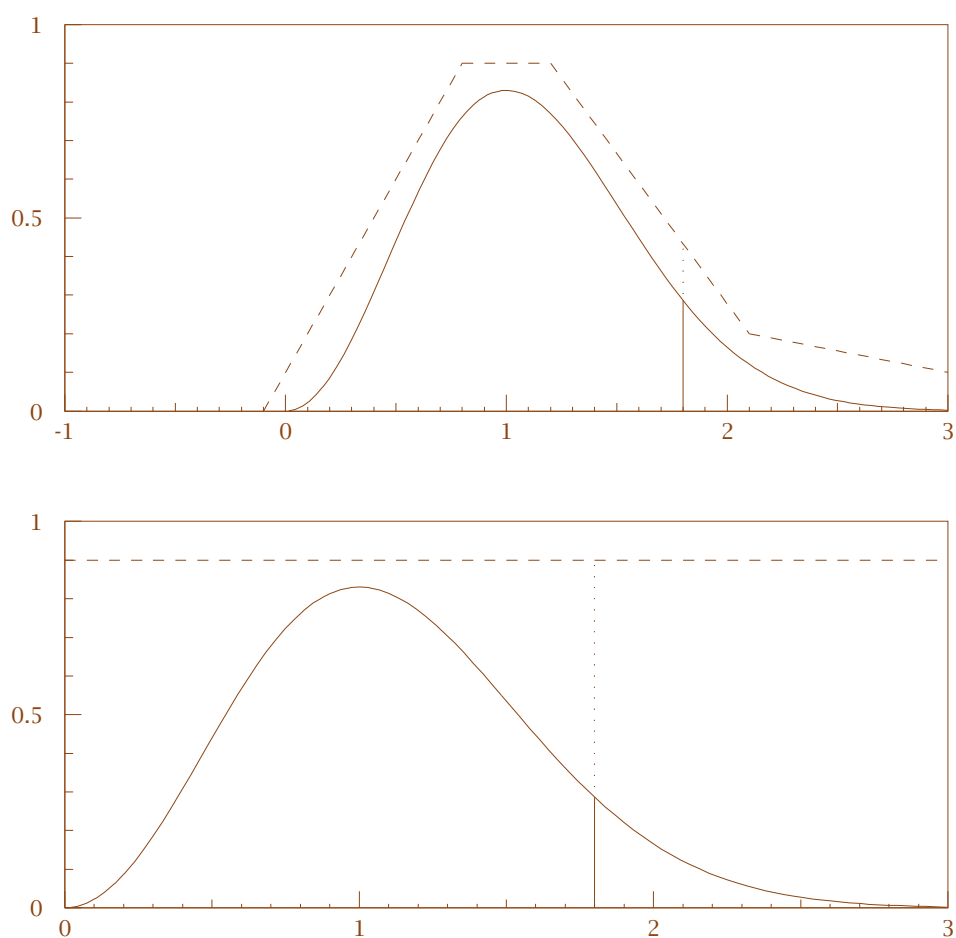
$$x = F^{-1}(y)$$

hanno densità di probabilità data da $f(x)$, come appunto richiesto.

Generalmente le funzioni di distribuzione $F(x)$ non si sanno invertire per via analitica; un metodo numerico spesso impiegato, e che richiede la sola preventiva conoscenza della $f(x)$ (quindi non bisogna nemmeno saper calcolare la $F(x)$, per non parlare della sua inversa) è illustrato qui di seguito (*metodo dei rigetti*). Si faccia riferimento alla figura 8b: sia x limitata in un intervallo chiuso $[x_{\min}, x_{\max}]$ (nella figura, $x_{\min} = 0$ e $x_{\max} = 3$); e si conosca

³O meglio *pseudo-casuali*: ovverosia prodotti da un algoritmo ripetibile, quindi non propriamente "imprevedibili"; ma in modo tale che le loro proprietà statistiche siano indistinguibili da quelle di una sequenza casuale propriamente detta.

FIGURA 8b - La scelta di un numero a caso con distribuzione prefissata mediante tecniche numeriche (la densità di probabilità è la stessa della figura 4d); la funzione maggiorante è una spezzata (superiormente) o la retta $y = 0.9$ (inferiormente).



una funzione $y = \varphi(x)$ *maggiorante* della $f(x)$, ossia una funzione che risulti comunque non inferiore alla f per qualunque $x \in [x_{\min}, x_{\max}]$.

Nel caso si sappia scegliere, sul piano $\{x, y\}$, un punto con distribuzione uniforme nella parte di piano limitata inferiormente dall'asse delle ascisse, superiormente dalla funzione $y = \varphi(x)$, e, lateralmente, dalle due rette di equazione $x = x_{\min}$ ed $x = x_{\max}$, basta accettare tale punto se la sua ordinata risulta non superiore alla corrispondente $f(x)$; e rigettarlo in caso contrario, iterando il procedimento fino a che la condizione precedente non è soddisfatta: le ascisse x dei punti accettati seguono la funzione di distribuzione $f(x)$.

Infatti, i punti accettati saranno distribuiti uniformemente nella parte di piano limitata dalla $y = f(x)$; quindi, in un intervallino infinitesimo centrato su una particolare x , vi sarà un numero di punti accettati proporzionale all'altezza della curva sopra di esso — ovvero sia ogni ascissa x viene accettata con densità di probabilità che è proprio $f(x)$.

La scelta, infine, di un punto che sia distribuito uniformemente nella parte di piano limitata dalla funzione $y = \varphi(x)$ si sa sicuramente effettuare se $\varphi(x)$ è stata scelta in modo che si sappia invertire la sua funzione integrale

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt$$

così che si possa associare, a qualsiasi valore A compreso tra 0 e $\Phi(+\infty)$, quella $x = \Phi^{-1}(A)$ che lascia alla propria sinistra un'area A al di sotto della funzione $y = \varphi(x)$ (una scelta banale è quella di prendere come maggiorante una retta, o meglio una spezzata — come illustrato nella figura 8b).

In tal caso basta scegliere un numero A con distribuzione uniforme tra i limiti $\Phi(x_{\min})$ e $\Phi(x_{\max})$; trovare la $x = \Phi^{-1}(A)$ che soddisfa la condizione precedente; ed infine scegliere una y con distribuzione uniforme tra 0 e $\varphi(x)$. Non è difficile rendersi conto che il punto (x, y) soddisfa alla condizione richiesta di essere distribuito uniformemente nella parte del semipiano $y > 0$ limitata superiormente dalla funzione $y = \varphi(x)$: a questo punto non rimane che calcolare la $f(x)$ ed accettare x se $y \leq f(x)$.

Se proprio non si è in grado di effettuare una scelta migliore, anche una retta del tipo $y = \text{cost.}$ può andar bene; basta tener presente che l'algoritmo viene sfruttato tanto più efficacemente quanto più $y = \varphi(x)$ è vicina alla $f(x)$ (in tal caso il numero di rigetti è minore).

Per questo motivo, una scelta del tipo $\varphi(x) = \text{cost.}$ è assolutamente da evitare se la $f(x)$ è sensibilmente diversa da zero solo in una parte ristretta dell'intervallo di definizione (perché in tal caso la scelta uniforme di x all'interno dell'area su detta ci farebbe trascorrere gran parte del tempo ad esa-

minare valori poco probabili rispetto alla $\varphi(x)$, che vengono in conseguenza quasi sempre rifiutati).

8.1.3 Esempio: valori estremi di un campione di dati a distribuzione uniforme

Come ulteriore esempio, applichiamo le conclusioni dei paragrafi 6.6 e 7.1.6 ad un campione di valori proveniente da una distribuzione uniforme. Usando le espressioni per $f(x)$ e $F(x)$ che conosciamo, ed essendo⁴

$$1 - F(x) = \frac{b - x}{b - a} ,$$

la (6.18) diventa

$$f_i(x) = N \binom{N-1}{i-1} \frac{(x-a)^{i-1} (b-x)^{N-i}}{(b-a)^N}$$

e, in particolare, per i due valori minimo e massimo presenti nel campione le densità di probabilità si scrivono

$$f_1(x) = N \frac{(b-x)^{N-1}}{(b-a)^N}$$

e

$$f_N(x) = N \frac{(x-a)^{N-1}}{(b-a)^N} .$$

Come conseguenza, la speranza matematica di x_N vale

$$\begin{aligned} E(x_N) &= \int_a^b x \cdot f_N(x) dx \\ &= \frac{N}{(b-a)^N} \int_a^b [a + (x-a)] (x-a)^{N-1} dx \\ &= \frac{N}{(b-a)^N} \left[a \frac{(x-a)^N}{N} + \frac{(x-a)^{N+1}}{N+1} \right]_a^b \\ &= a + \frac{N}{N+1} (b-a) \\ &= b - \frac{1}{N+1} (b-a) . \end{aligned}$$

⁴All'interno dell'intervallo $[a, b]$; per brevità ometteremo, qui e nel seguito, di specificare che, al di fuori di questo intervallo, le densità di probabilità sono identicamente nulle e le funzioni di distribuzione valgono o zero od uno.

Allo stesso modo si troverebbe

$$E(x_1) = a + \frac{1}{N+1} (b-a) ;$$

e, per il generico x_i ,

$$E(x_i) = a + \frac{i}{N+1} (b-a) .$$

Dopo gli opportuni calcoli, si potrebbero ricavare anche le varianze rispettive: che valgono

$$\text{Var}(x_1) = \text{Var}(x_N) = \frac{N}{(N+1)^2 (N+2)} (b-a)^2$$

e

$$\text{Var}(x_i) = \frac{i \cdot (N-i+1)}{(N+1)^2 (N+2)} (b-a)^2 .$$

È immediato calcolare la speranza matematica della semisomma del più piccolo e del più grande valore presenti nel campione

$$d = \frac{x_1 + x_N}{2}$$

che vale

$$E(d) = \frac{E(x_1) + E(x_N)}{2} = \frac{a+b}{2} ;$$

come pure quella del cosiddetto *range*,

$$R = x_N - x_1$$

per il quale

$$E(R) = E(x_N) - E(x_1) = (b-a) \left(1 - \frac{2}{N+1}\right) .$$

Per il calcolo delle varianze, invece, si deve ricorrere alla distribuzione congiunta (7.6), dalla quale si può ricavare

$$\text{Var}(d) = \frac{(b-a)^2}{2(N+1)(N+2)}$$

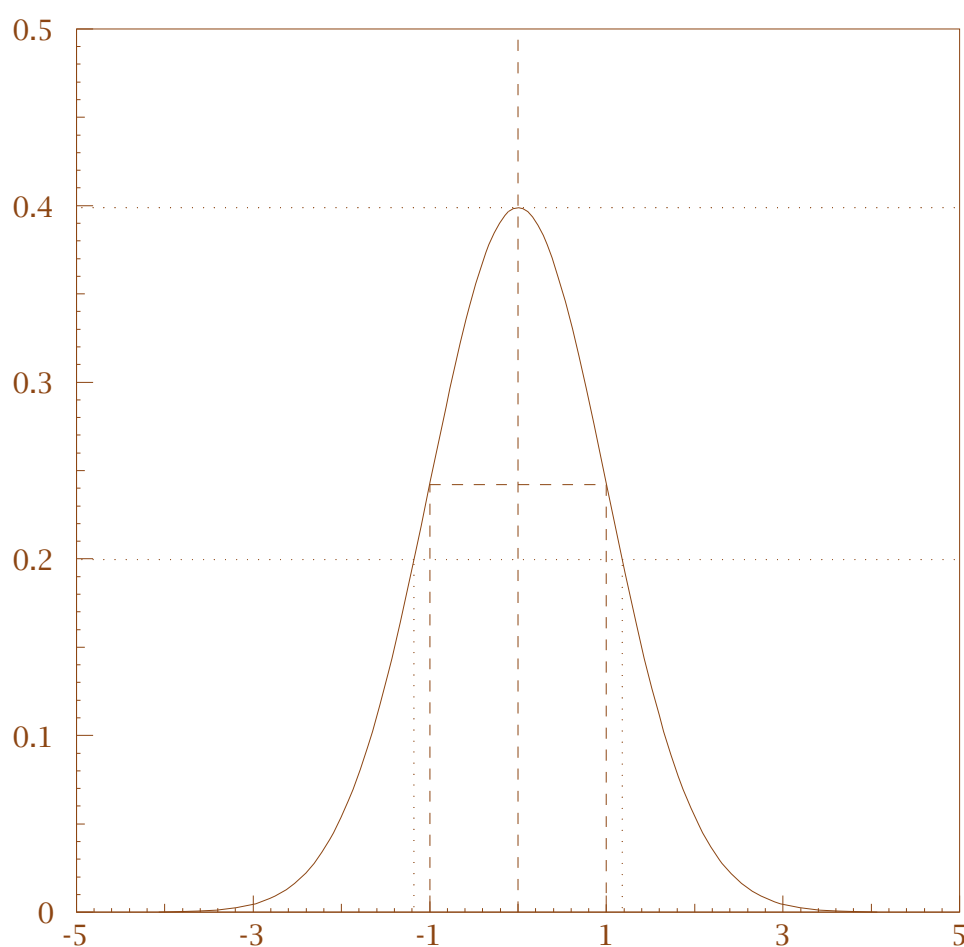
e

$$\text{Var}(R) = \frac{2(N-1)}{(N+1)^2 (N+2)} (b-a)^2 .$$

8.2 La distribuzione normale

La *funzione normale* (o *funzione di Gauss*), che esamineremo poi in dettaglio nel prossimo capitolo mettendo l'accento sui suoi legami con le misure ripetute delle grandezze fisiche, è una funzione di frequenza per la x che dipende da due parametri μ e σ (con la condizione $\sigma > 0$) definita come

FIGURA 8c - L'andamento della funzione $N(x;0,1)$ per la *variabile normale standardizzata* (ossia con media 0 e varianza 1).



$$f(x) \equiv N(x; \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2}.$$

L'andamento della funzione normale è quello delineato nella figura 8c: quando $x = \mu$ si ha un punto di massimo, nel quale la funzione ha il valore

$\hat{y} = (\sigma\sqrt{2\pi})^{-1} \approx 0.4/\sigma$. La *larghezza a metà altezza*⁵ è pari all'ampiezza dell'intervallo che separa i due punti x_1 ed x_2 di ascissa $\mu \pm \sigma\sqrt{2\ln 2}$ e di ordinata $y_1 = y_2 = \hat{y}/2$: e vale quindi $2\sigma\sqrt{2\ln 2} \approx 2.35\sigma$.

La funzione generatrice dei momenti è definita attraverso l'equazione (6.4) e, nel caso della distribuzione normale, abbiamo

$$\begin{aligned} M_x(t) &= \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \frac{e^{t\mu}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{t(x-\mu)} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \frac{e^{t\mu}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\left[\frac{\sigma^2 t^2}{2} - \frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}\right]} dx \\ &= e^{\left(t\mu + \frac{\sigma^2 t^2}{2}\right)} \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-(\mu+\sigma^2 t)}{\sigma}\right]^2} dx . \end{aligned}$$

Riconoscendo nell'argomento dell'integrale la funzione $N(x; \mu + \sigma^2 t, \sigma)$, ovvero la funzione normale relativa ai parametri $\mu + \sigma^2 t$ e σ , è immediato capire che esso vale 1 in conseguenza della condizione di normalizzazione; quindi la funzione generatrice dei momenti, per la distribuzione normale, è data da

$$M_x(t) = e^{\left(t\mu + \frac{\sigma^2 t^2}{2}\right)} \quad (8.2)$$

e, con passaggi simili, si potrebbe trovare la funzione caratteristica della distribuzione normale: che vale

$$\phi_x(t) = e^{\left(it\mu - \frac{\sigma^2 t^2}{2}\right)} . \quad (8.3)$$

Sfruttando la (8.2) è facile calcolare la speranza matematica della distribuzione normale:

$$E(x) = \left. \frac{dM_x(t)}{dt} \right|_{t=0} = \mu ;$$

la funzione generatrice dei momenti rispetto alla media μ vale allora

$$\bar{M}_x(t) = e^{-t\mu} M_x(t) = e^{\frac{\sigma^2 t^2}{2}} \quad (8.4)$$

e dalla (8.4) si ricava poi la varianza della x ,

$$\text{Var}(x) = \left. \frac{d^2 \bar{M}_x(t)}{dt^2} \right|_{t=0} = \sigma^2 .$$

⁵In genere indicata con la sigla FWHM, acronimo di *full width at half maximum*; è un parametro talvolta usato nella pratica per caratterizzare una curva, perché facile da misurare su un oscilloscopio.

Vista la simmetria della funzione, tutti i suoi momenti di ordine dispari rispetto alla media sono nulli; mentre quelli di ordine pari soddisfano alla formula generale (valida per qualsiasi intero k)

$$\mu_{2k} = E \{ [x - E(x)]^{2k} \} = \frac{(2k)!}{2^k k!} \mu_2^k \quad (8.5)$$

con

$$\mu_2 = E \{ [x - E(x)]^2 \} = \sigma^2 .$$

Nel caso particolare di una variabile normale con valore medio $\mu = 0$ e varianza $\sigma^2 = 1$ (*variabile normale standardizzata*), la funzione generatrice dei momenti diventa

$$M_x(t) \equiv \bar{M}_x(t) = e^{\frac{t^2}{2}}$$

e la funzione caratteristica

$$\phi_x(t) = e^{-\frac{t^2}{2}} .$$

Dimostriamo ora il seguente importante

TEOREMA: *combinazioni lineari di variabili casuali normali e tutte statisticamente indipendenti tra loro sono ancora distribuite secondo la legge normale.*

Siano N variabili normali x_k (con $k = 1, \dots, N$), e siano μ_k e σ_k^2 i loro valori medi e le loro varianze rispettivamente; consideriamo poi la nuova variabile casuale y definita dalla

$$y = \sum_{k=1}^N a_k x_k$$

(ove le a_k sono coefficienti costanti). La funzione caratteristica di ognuna delle x_k è, dalla (8.3),

$$\phi_{x_k}(t) = e^{\left(it\mu_k - \frac{\sigma_k^2 t^2}{2} \right)}$$

e quella della variabile ausiliaria $\xi_k = a_k x_k$, dall'equazione (6.17),

$$\phi_{\xi_k}(t) = \phi_{x_k}(a_k t) = e^{\left(ia_k t \mu_k - \frac{\sigma_k^2 a_k^2 t^2}{2} \right)} .$$

Infine, la funzione caratteristica della \mathcal{Y} vale, essendo

$$\mathcal{Y} = \sum_{k=1}^N \xi_k$$

e ricordando l'equazione (6.11), applicabile perché anche le ξ_k sono indipendenti tra loro, otteniamo

$$\begin{aligned} \phi_{\mathcal{Y}}(t) &= \prod_{k=1}^N \phi_{\xi_k}(t) \\ &= \prod_{k=1}^N e^{(it a_k \mu_k - \frac{1}{2} t^2 a_k^2 \sigma_k^2)} \\ &= e^{[it(\sum_k a_k \mu_k) - \frac{1}{2} t^2 (\sum_k a_k^2 \sigma_k^2)]} \\ &= e^{(it\mu - \frac{t^2 \sigma^2}{2})} \end{aligned}$$

ove si è posto

$$\mu = \sum_{k=1}^N a_k \mu_k \quad \text{e} \quad \sigma^2 = \sum_{k=1}^N a_k^2 \sigma_k^2 .$$

Questa è appunto la funzione caratteristica di una nuova distribuzione normale; e, in virtù di uno dei teoremi enunciati nel paragrafo 6.4, quanto dimostrato prova la tesi.

8.3 La distribuzione di Cauchy

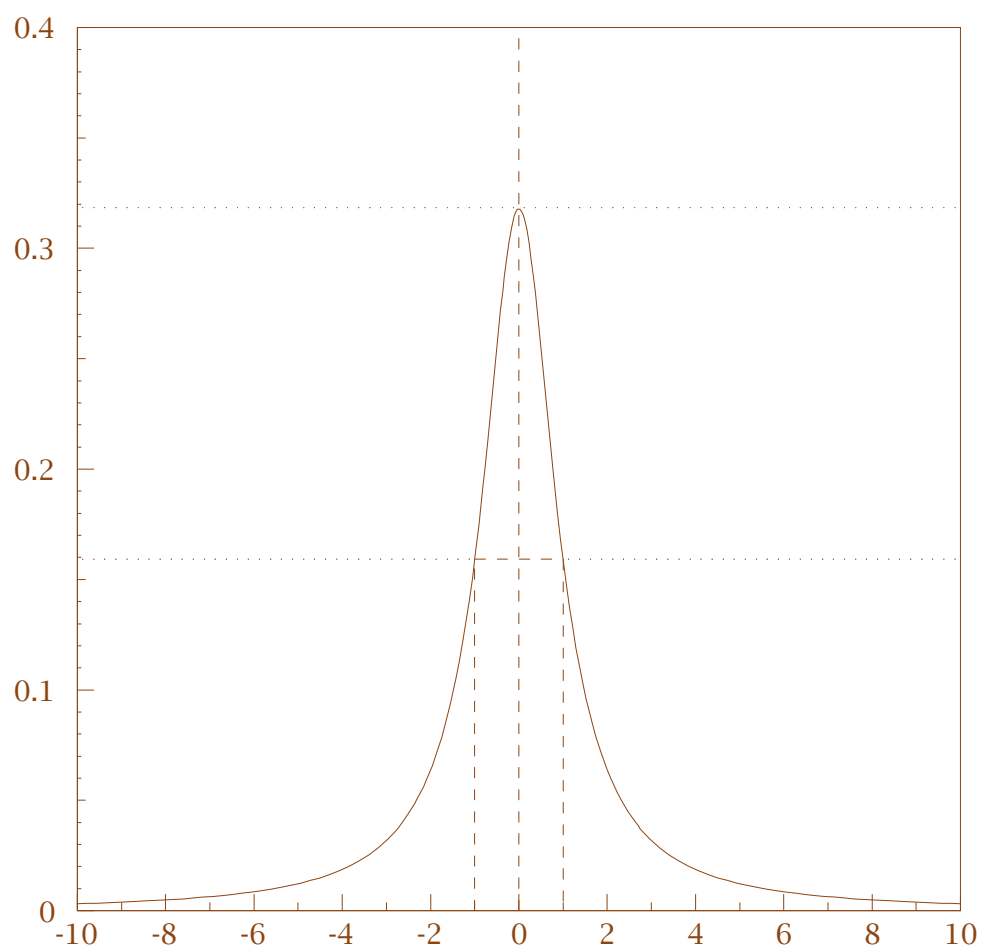
La distribuzione di Cauchy (o *distribuzione di Breit-Wigner*, nome con il quale è più nota nel mondo della fisica) è definita da una densità di probabilità che corrisponde alla funzione, dipendente da due parametri θ e d (con la condizione $d > 0$),

$$f(x; \theta, d) = \frac{1}{\pi d} \frac{1}{1 + \left(\frac{x-\theta}{d}\right)^2} . \quad (8.6)$$

Anche se la (8.6) è integrabile, e la sua funzione integrale, ovverosia la funzione di distribuzione della x , vale

$$F(x; \theta, d) = \int_{-\infty}^x f(t) dt = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-\theta}{d}\right)$$

FIGURA 8d - L'andamento della distribuzione di Cauchy, per $\theta = 0$ e $d = 1$.



nessuno dei momenti esiste, nemmeno la media.

θ è la mediana della distribuzione e d ne misura la larghezza a metà altezza, come è rilevabile ad esempio dalla figura 8d. La funzione caratteristica della distribuzione di Cauchy è la

$$\phi_x(t; \theta, d) = e^{i\theta t - |t| d} ;$$

per la cosiddetta *variabile standardizzata*,

$$u = \frac{x - \theta}{d}$$

funzione di frequenza, funzione di distribuzione e funzione caratteristica valgono rispettivamente

$$\begin{cases} f(u) = \frac{1}{\pi (1 + u^2)} \\ F(u) = \frac{1}{2} + \frac{1}{\pi} \arctan u \\ \phi_u(t) = e^{-|t|} \end{cases}$$

Secondo la funzione (8.6) sono, ad esempio, distribuite le intensità nelle righe spettrali di emissione e di assorbimento degli atomi (che hanno una ampiezza non nulla); e la massa invariante delle risonanze nella fisica delle particelle elementari. È evidente però come nella fisica la distribuzione di Cauchy possa descrivere questi fenomeni solo in prima approssimazione: infatti essa si annulla solo per $x \rightarrow \pm\infty$, ed è chiaramente priva di significato fisico una probabilità non nulla di emissione spettrale per frequenze negative, o di masse invarianti anch'esse negative nel caso delle risonanze.

Per la distribuzione di Cauchy *troncata*, ossia quella descritta dalla funzione di frequenza (per la variabile standardizzata)

$$f(u | -K \leq u \leq K) = \begin{cases} 0 & |u| > K \\ \frac{1}{2 \arctan K} \frac{1}{(1 + u^2)} & |u| \leq K \end{cases}$$

(*discontinua* in $u = \pm K$), esistono invece i momenti: i primi due valgono

$$E(u | -K \leq u \leq K) = 0$$

e

$$\text{Var}(u | -K \leq u \leq K) = \frac{K}{\arctan K} - 1$$

Se le x_k sono N variabili casuali indipendenti che seguono la distribuzione di Cauchy con parametri θ_k e d_k , una generica loro combinazione lineare

$$y = \sum_{k=1}^N a_k x_k$$

segue la stessa distribuzione: infatti la funzione generatrice per le x_k è

$$\phi_{x_k}(t) = e^{i\theta_k t - |t|d_k}$$

e, definendo $\xi_k = a_k x_k$ e ricordando la (6.17),

$$\phi_{\xi_k}(t) = \phi_{x_k}(a_k t) = e^{ia_k \theta_k t - |t| \cdot |a_k| d_k} ;$$

infine, applicando la (6.11),

$$\phi_y(t) = \prod_{k=1}^N \phi_{\xi_k}(t) = e^{i\theta_y t - |t|d_y}$$

ove si è posto

$$\theta_y = \sum_{k=1}^N a_k \theta_k \quad \text{e} \quad d_y = \sum_{k=1}^N |a_k| d_k .$$

Una conseguenza importante è che il valore medio di un campione di misure proveniente da una popolazione che segua la distribuzione di Cauchy con certi parametri θ e d (in questo caso tutte le a_k sono uguali e valgono $1/N$) è distribuito anch'esso secondo Cauchy *e con gli stessi parametri*; in altre parole *non si guadagna nessuna informazione* accumulando più di una misura (e calcolando la media aritmetica del campione)⁶.

8.3.1 Il rapporto di due variabili normali

Siano due variabili casuali x ed y che seguano la distribuzione normale standardizzata $N(0, 1)$; e sia inoltre la y definita su tutto l'asse reale ad eccezione dell'origine ($y \neq 0$). La densità di probabilità congiunta di x e y è la

$$f(x, y) = N(x; 0, 1) \cdot N(y; 0, 1) = \frac{1}{2\pi} e^{-\frac{1}{2}x^2} e^{-\frac{1}{2}y^2} ;$$

⁶Esistono altre tecniche, basate però sull'uso della mediana, che permettono di migliorare la conoscenza del valore di θ disponendo di più di una misura.

definendo

$$u = \frac{x}{y} \quad \text{e} \quad v = y$$

e ricordando la (7.4), la densità di probabilità $\varphi(u)$ della u è la

$$\begin{aligned} \varphi(u) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2v^2} e^{-\frac{1}{2}v^2} |v| \, dv \\ &= \frac{1}{\pi} \int_0^{+\infty} e^{-\frac{1}{2}v^2(1+u^2)} v \, dv \\ &= \frac{1}{\pi(1+u^2)} \int_0^{+\infty} e^{-t} \, dt \\ &= \frac{1}{\pi(1+u^2)} \left[-e^{-t} \right]_0^{+\infty} \\ &= \frac{1}{\pi(1+u^2)} \end{aligned}$$

Per eseguire l'integrazione si è effettuata la sostituzione

$$t = \frac{1}{2} v^2 (1 + u^2) \quad \Rightarrow \quad dt = (1 + u^2) v \, dv$$

e si riconosce immediatamente nella $\varphi(u)$ la densità di probabilità di una variabile (standardizzata) di Cauchy: *il rapporto tra due variabili normali segue la distribuzione di Cauchy.*

8.4 La distribuzione di Bernoulli

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 (deve necessariamente risultare $0 \leq x \leq N$).

L'evento casuale costituito dal presentarsi di E per x volte (e quindi 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

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

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

$$P(x; N) = \frac{N!}{x! (N - x)!} p^x q^{N-x} . \quad (8.7)$$

Questa distribuzione di probabilità $P(x; N)$ per una variabile casuale discreta x si chiama *distribuzione binomiale* o *di Bernoulli*⁷; vogliamo ora determinarne alcune costanti caratteristiche.

Verifichiamo per prima cosa che vale la condizione di normalizzazione: sfruttando 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 la speranza matematica della variabile x (ossia il numero di successi attesi, in media, in N prove): per questo useremo la stessa variabile casuale ausiliaria già considerata nel paragrafo 5.6.3, y , che rappresenta il numero di successi nella generica delle N prove eseguite.

Avevamo a suo tempo già calcolato, sempre nel paragrafo 5.6.3, la speranza matematica della y

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

e, osservando che anche y^2 può assumere i due soli valori 1 e 0, sempre con le probabilità rispettive p e q ,

$$E(y^2) = 1 \cdot p + 0 \cdot q = p$$

e quindi la varianza della y esiste e vale

$$\text{Var}(y) = E(y^2) - [E(y)]^2 = p - p^2 = p(1 - p) = pq . \quad (8.8)$$

Il numero totale x di successi nelle N prove è legato ai valori y_i della y in ognuna di esse dalla

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

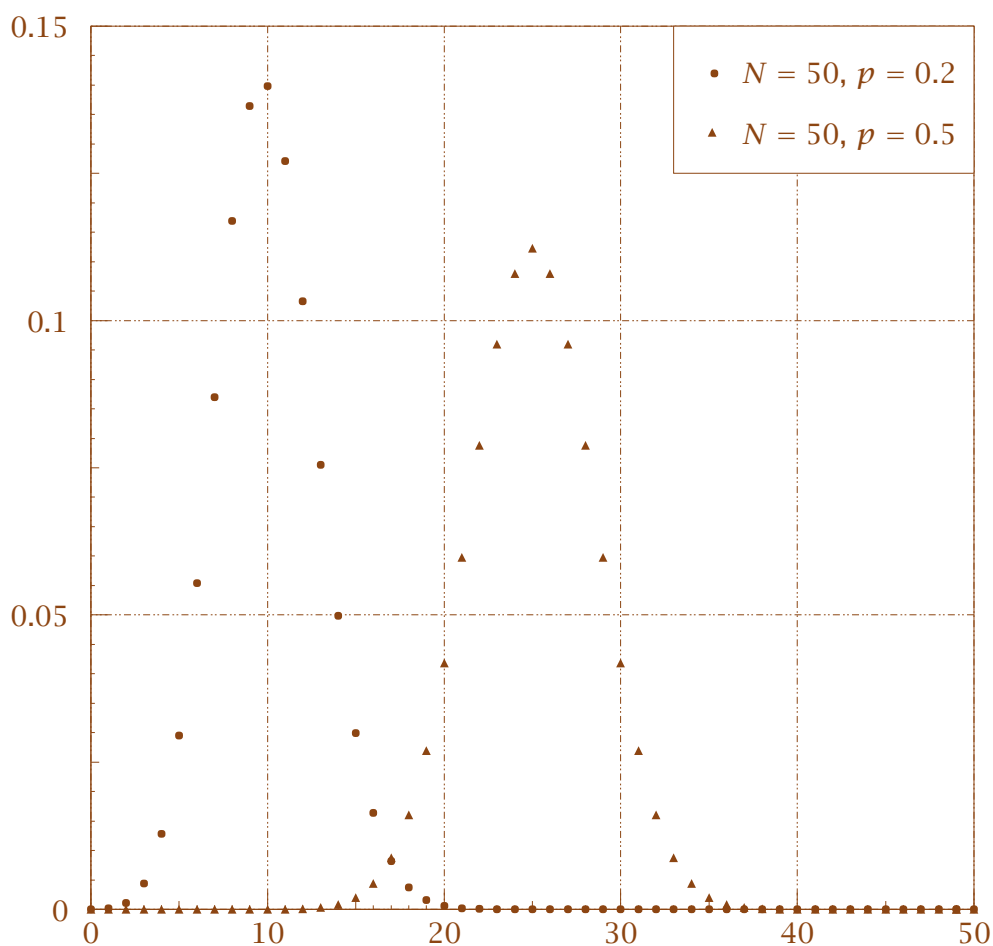
⁷I Bernoulli furono una famiglia originaria di Anversa poi trasferitasi a Basilea, numerosi membri della quale ebbero importanza per le scienze del diciassettesimo e del diciottesimo secolo; quello cui vanno attribuiti gli studi di statistica ebbe nome Jacob (o Jacques), visse dal 1654 al 1705, e fu zio del più noto Daniel (cui si deve il teorema di Bernoulli della dinamica dei fluidi).

e risulta quindi, per speranza matematica e varianza della distribuzione binomiale,

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

$$\text{Var}(x) = \sigma_x^2 = \text{Var}\left(\sum_{i=1}^N y_i\right) = \sum_{i=1}^N \text{Var}(y_i) = Npq .$$

FIGURA 8e - La distribuzione binomiale, per un numero di prove $N = 50$ e due differenti valori della probabilità p .



Come è evidente dalla figura 8e, la forma della distribuzione binomiale è molto simile a quella di una curva di Gauss; si può in effetti dimostrare che

quando N tende all'infinito la distribuzione di probabilità dei possibili valori tende ad una distribuzione normale avente la stessa media Np e la stessa varianza Npq . Infatti la funzione generatrice dei momenti della distribuzione di Bernoulli è

$$\begin{aligned} M_x(t) &= E(e^{tx}) \\ &= \sum_{x=0}^N e^{tx} \binom{N}{x} p^x q^{N-x} \\ &= \sum_{x=0}^N \binom{N}{x} (pe^t)^x q^{N-x} \\ &= (pe^t + q)^N \end{aligned}$$

o anche, ricordando che $q = 1 - p$,

$$M_x(t) = \left[1 + p(e^t - 1) \right]^N$$

e se, per semplificare i calcoli, ci riferiamo alla *variabile standardizzata*

$$z = \frac{x - E(x)}{\sigma_x} = \frac{x - Np}{\sqrt{Npq}} = ax + b$$

ove si è posto

$$a = \frac{1}{\sqrt{Npq}} \quad \text{e} \quad b = -\frac{Np}{\sqrt{Npq}}$$

applicando la (6.16) si trova

$$\begin{aligned} M_z(t) &= e^{tb} M_x(at) \\ &= e^{-\frac{Np}{\sqrt{Npq}}t} \left[1 + p \left(e^{\frac{t}{\sqrt{Npq}}} - 1 \right) \right]^N \end{aligned}$$

da cui, passando ai logaritmi naturali,

$$\begin{aligned}
\ln M_z(t) &= -\frac{Np}{\sqrt{Npq}} t + N \ln \left[1 + p \left(e^{\frac{t}{\sqrt{Npq}}} - 1 \right) \right] \\
&= -\frac{Np}{\sqrt{Npq}} t + N \left[p \left(e^{\frac{t}{\sqrt{Npq}}} - 1 \right) - \frac{p^2}{2} \left(e^{\frac{t}{\sqrt{Npq}}} - 1 \right)^2 + \right. \\
&\quad \left. + \frac{p^3}{3} \left(e^{\frac{t}{\sqrt{Npq}}} - 1 \right)^3 + \dots \right] \\
&= -\frac{Np}{\sqrt{Npq}} t + N \left\{ p \left[\frac{t}{\sqrt{Npq}} + \frac{1}{2} \frac{t^2}{Npq} + \frac{1}{6} \frac{t^3}{(Npq)^{\frac{3}{2}}} + \dots \right] - \right. \\
&\quad \left. - \frac{p^2}{2} \left[\frac{t}{\sqrt{Npq}} + \frac{1}{2} \frac{t^2}{Npq} + \frac{1}{6} \frac{t^3}{(Npq)^{\frac{3}{2}}} + \dots \right]^2 + \right. \\
&\quad \left. + \frac{p^3}{3} \left[\frac{t}{\sqrt{Npq}} + \frac{1}{2} \frac{t^2}{Npq} + \frac{1}{6} \frac{t^3}{(Npq)^{\frac{3}{2}}} + \dots \right]^3 + \dots \right\} \\
&= N \left\{ p \left[\frac{1}{2} \frac{t^2}{Npq} + \frac{1}{6} \frac{t^3}{(Npq)^{\frac{3}{2}}} + \dots \right] - \right. \\
&\quad \left. - \frac{p^2}{2} \left[\frac{t^2}{Npq} + \frac{t^3}{(Npq)^{\frac{3}{2}}} + \dots \right] + \frac{p^3}{3} \left[\frac{t^3}{(Npq)^{\frac{3}{2}}} + \dots \right] \right\} \\
&= N \left[\frac{1}{2} \frac{t^2}{Npq} p(1-p) + \frac{t^3}{(Npq)^{\frac{3}{2}}} \left(\frac{p}{6} - \frac{p^2}{2} + \frac{p^3}{3} \right) + \mathcal{O}(t^4 N^{-2}) \right] \\
&= \frac{1}{2} t^2 + \mathcal{O}(t^3 N^{-\frac{1}{2}})
\end{aligned}$$

ove si è sviluppato in serie di McLaurin prima

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

e poi

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

e si sono svolti i prodotti tenendo solo i termini dei primi ordini.

Chiaramente quando N viene fatto tendere all'infinito tutti i termini eccetto il primo tendono a zero, per cui

$$\lim_{N \rightarrow \infty} M_z(t) = e^{\frac{t^2}{2}}$$

e $M_z(t)$ tende quindi alla funzione generatrice dei momenti di una distribuzione normale standardizzata; in conseguenza si è effettivamente provato, visto uno dei teoremi citati nel paragrafo 6.4, che la distribuzione binomiale tende ad una distribuzione normale.

In pratica, quando il numero di prove N è elevato e la probabilità p non è troppo vicina ai valori estremi 0 ed 1, la distribuzione binomiale è bene approssimata da una distribuzione normale; in generale si ritiene che l'approssimazione sia accettabile quando entrambi i prodotti Np e Nq hanno valore non inferiore a 5.

8.4.1 Applicazione: decadimenti radioattivi

Se la probabilità Λ_t per un singolo nucleo instabile di decadere in un intervallo di tempo t è costante, la probabilità di avere un numero prefissato di decadimenti nel tempo t in un insieme di N nuclei è data dalla distribuzione binomiale; in particolare, il numero medio di decadimenti in N nuclei e nel tempo t è $N\Lambda_t$.

Se si ammette poi che Λ_t sia proporzionale al tempo⁸ t , indicando con λ la probabilità di decadimento nell'unità di tempo avremo $\Lambda_t = \lambda t$; in un tempo infinitesimo dt , il numero di atomi $N(t)$ presente al tempo t varia mediamente di

$$dN = -N \lambda dt .$$

Separando le variabili ed integrando, il numero *medio* di atomi presenti al tempo t , dopo il decadimento di una parte di quelli N_0 presenti all'istante iniziale $t = 0$, è dato dalla

$$N(t) = N_0 e^{-\lambda t} \quad (8.9)$$

ed il numero medio di decadimenti dalla

$$N_0 - N(t) = N_0(1 - e^{-\lambda t}) .$$

La *vita media* τ di una sostanza radioattiva si può definire come il tempo necessario perché il numero originario di nuclei si riduca mediamente di un fattore $1/e$; quindi $\tau = 1/\lambda$, e la (8.9) si può riscrivere

$$N(t) = N_0 e^{-\frac{t}{\tau}} . \quad (8.10)$$

⁸Questa ipotesi può evidentemente essere soddisfatta solo in prima approssimazione: basta pensare al fatto che Λ_t deve raggiungere l'unità solo dopo un tempo infinito. In particolare, la probabilità per un nucleo di decadere in un tempo $2t$ vale $\Lambda_{2t} = \Lambda_t + (1 - \Lambda_t) \cdot \Lambda_t = 2\Lambda_t - \Lambda_t^2$; e l'ipotesi fatta è in effetti valida solo se Λ_t è infinitesimo, o (in pratica) se l'osservazione riguarda un lasso di tempo trascurabile rispetto alla vita media.

8.4.2 Applicazione: il rapporto di asimmetria

Frequentemente, nella fisica, si devono considerare esperimenti in cui si cerca di mettere in evidenza delle *asimmetrie*; ovvero, la non invarianza della funzione di frequenza di una qualche variabile casuale per riflessione rispetto ad un piano. Supponiamo, come esempio, di considerare la possibilità che un dato fenomeno abbia una diversa probabilità di presentarsi *in avanti* o *all'indietro* rispetto ad una opportuna superficie di riferimento; e di raccogliere N eventi sperimentali dividendoli in due sottoinsiemi (mutualmente esclusivi ed esaurienti) collegati a queste due categorie, indicando con F e B (iniziali delle due parole inglesi *forward* e *backward*) il loro numero: ovviamente dovrà risultare $N = F + B$.

Il cosiddetto *rapporto di asimmetria*, R , si definisce come

$$R = \frac{F - B}{F + B} = \frac{F - B}{N} = \frac{2F}{N} - 1 \quad : \quad (8.11)$$

è ovvio sia che $-1 \leq R \leq 1$, sia che soltanto due dei quattro valori N , F , B ed R sono indipendenti; e, volendo, dall'ultima forma della (8.11) si possono ricavare le espressioni di F e B in funzione di N ed R , ovvero

$$F = \frac{N(1 + R)}{2} \quad \text{e} \quad B = \frac{N(1 - R)}{2} .$$

Se indichiamo con p la probabilità di un evento in avanti (e con $q = 1 - p$ quella di uno all'indietro), il trovare esattamente F eventi in avanti su un totale di N ha probabilità data dalla distribuzione binomiale: ovvero

$$\Pr(F) = \binom{N}{F} p^F (1 - p)^{N-F}$$

con, inoltre,

$$E(F) = Np \quad \text{e} \quad \text{Var}(F) = Np(1 - p) .$$

Ma, per quanto detto, c'è una corrispondenza biunivoca tra i valori di N ed F da una parte, e quello di R ; così che

$$\Pr(R) = \binom{N}{\frac{N(1+R)}{2}} p^{\frac{N(1+R)}{2}} (1 - p)^{\frac{N(1-R)}{2}} ,$$

$$E(R) = \frac{2}{N} E(F) - 1 = 2p - 1$$

e

$$\text{Var}(R) = \frac{4}{N^2} \text{Var}(F) = \frac{4p(1-p)}{N} .$$

Se il numero di eventi nel campione è elevato e p lontano dai valori estremi, così da potere sia sfruttare l'approssimazione normale alla distribuzione di Bernoulli, sia pensare che risulti

$$p \simeq \frac{F}{N} \quad \text{che} \quad q \simeq \frac{B}{N} ,$$

come conseguenza anche la distribuzione di R sarà approssimativamente normale; e con i primi due momenti dati da

$$E(R) \simeq 2 \frac{F}{N} - 1 \quad \text{e} \quad \text{Var}(R) \simeq 4 \frac{FB}{N^3} .$$

Del rapporto di asimmetria parleremo ancora più avanti, nel corso di questo stesso capitolo: più esattamente nel paragrafo 8.5.2.

8.4.3 La distribuzione binomiale negativa

Consideriamo ancora un evento casuale E ripetibile, avente probabilità costante p di presentarsi (e quindi probabilità $q = 1 - p$ di non presentarsi) in una singola prova; in più prove successive l'evento seguirà dunque la statistica di Bernoulli. Vogliamo ora calcolare la probabilità $f(x; N, p)$ che, prima che si verifichi l' N -esimo successo, si siano avuti esattamente x insuccessi; o, se si preferisce, la probabilità che l' N -simo successo si presenti nella $(N + x)$ -sima prova.

L'evento casuale considerato si realizza se e solo se nelle prime $N + x - 1$ prove si è presentata una delle possibili sequenze di $N - 1$ successi e x fallimenti; e se poi, nella prova successiva, si ha un ulteriore successo. La prima condizione, applicando la (8.7), ha probabilità

$$\binom{N + x - 1}{N - 1} p^{N-1} q^x ;$$

e, vista l'indipendenza statistica delle prove tra loro, risulta dunque

$$f(x; N, p) = \binom{N + x - 1}{N - 1} p^N q^x = \binom{N + x - 1}{x} p^N q^x . \quad (8.12)$$

(nell'ultimo passaggio si è sfruttata la proprietà dei coefficienti binomiali $\binom{N}{K} \equiv \binom{N}{N-K}$; vedi in proposito il paragrafo A.6).

Questa distribuzione di probabilità prende il nome di *distribuzione binomiale negativa*⁹; il motivo di tale nome è che l'equazione (8.12) può essere riscritta in forma compatta sfruttando una “estensione” dei coefficienti binomiali che permette di definirli anche per valori negativi di N . La funzione generatrice dei momenti è

$$M_x(t) = E(e^{tx}) = \left(\frac{p}{1 - qe^t} \right)^N ;$$

da questa si possono poi ricavare la speranza matematica

$$E(x) = \frac{Nq}{p} ,$$

e la varianza

$$\text{Var}(x) = N \frac{q}{p^2} .$$

La distribuzione binomiale negativa con $N = 1$ prende il nome di *distribuzione geometrica*; la probabilità (8.12) diventa

$$f(x; p) = pq^x ,$$

ed è quella dell'evento casuale consistente nell'ottenere il primo successo dopo esattamente x insuccessi; ponendo $N = 1$ nelle formule precedenti, speranza matematica e varianza della distribuzione geometrica sono rispettivamente

$$E(x) = \frac{q}{p} \quad \text{e} \quad \text{Var}(x) = \frac{q}{p^2} .$$

8.5 La distribuzione di Poisson

Sia E un evento casuale che avvenga rispettando le seguenti ipotesi:

1. La probabilità del verificarsi dell'evento E in un intervallo di tempo¹⁰ molto piccolo (al limite infinitesimo) dt è proporzionale alla durata di tale intervallo;

⁹La distribuzione binomiale negativa è talvolta chiamata anche *distribuzione di Pascal* o *di Pólya*.

¹⁰Sebbene 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

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 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

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

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

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

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

$$P(x; t) = \frac{(\lambda t)^x}{x!} e^{-\lambda t} . \quad (8.13)$$

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).

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

$$P(x; \alpha) = \frac{\alpha^x}{x!} e^{-\alpha} . \quad (8.14)$$

Questa distribuzione di probabilità per una variabile casuale (discreta) x prende il nome di *distribuzione di Poisson*¹¹; da essa si può ottenere, ad esempio, la probabilità di ottenere x decadimenti in una massa nota di sostanza radioattiva nel tempo t : infatti per questo tipo di processi fisici risultano soddisfatte le tre ipotesi di partenza.

Più esattamente, la probabilità di avere precisamente x decadimenti radioattivi nel tempo t è data dalla distribuzione binomiale; la distribuzione di Poisson è una approssimazione alla distribuzione binomiale che si può ritenere valida qualora si considerino eventi casuali di probabilità estremamente piccola, e che ci è possibile vedere solo perché si compiono osservazioni su un numero molto elevato di essi: in formula, quando

$$p^2 \ll p \quad \text{e} \quad p \ll Np \ll N \quad (8.15)$$

(*eventi rari su larga base statistica*).

Anche se la distribuzione di Poisson è, come nel caso dei decadimenti radioattivi, una approssimazione di quella binomiale, si preferisce però sempre usarla nella pratica al posto di quest'ultima quando le (8.15) siano approssimativamente verificate: infatti se N è grande i fattoriali e le potenze presenti nella (8.7) rendono generalmente l'espressione difficile da calcolare. 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 McLaurin della funzione esponenziale). Calcoliamo poi la speranza mate-

¹¹ Siméon Denis Poisson visse in Francia dal 1781 al 1840; matematico e fisico di valore, si occupò della teoria degli integrali e delle serie, di meccanica, elettricità, magnetismo ed astronomia. Gli studi sulla distribuzione che porta il suo nome compaiono nel trattato del 1837 "Recherches sur la probabilité des jugements...".

matica di x :

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

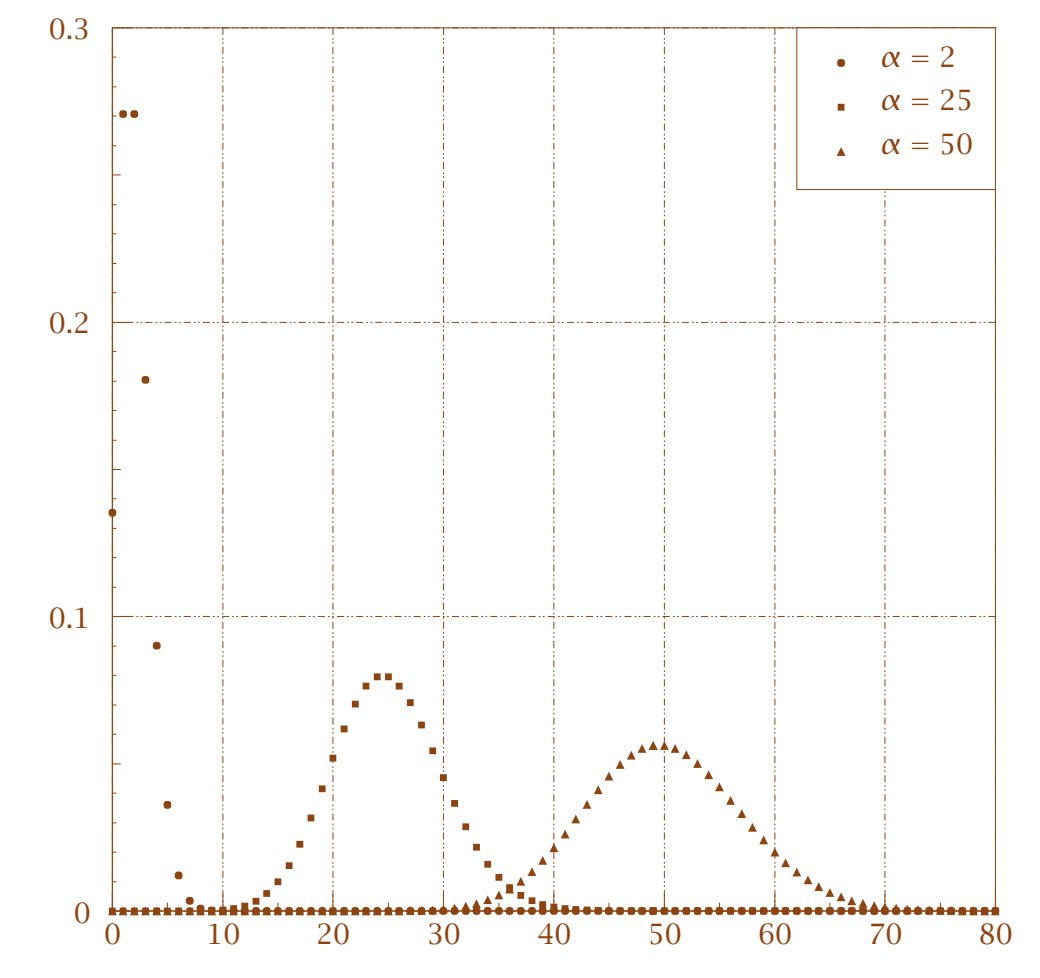
Nei passaggi si è prima osservato che il primo termine della sommatoria ($x = 0$) è nullo, e si è poi introdotta la nuova variabile $y = x - 1$. Troviamo ora la speranza matematica di x^2 : con passaggi analoghi, si ottiene

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

e la varianza di x risulta allora anch'essa data da

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

FIGURA 8f - La distribuzione di Poisson, per tre diversi valori del parametro α .



La funzione generatrice dei momenti, come si potrebbe facilmente ottenere dalla definizione, è la

$$M_x(t) = e^{-\alpha} e^{\alpha e^t} = e^{\alpha(e^t-1)} ; \quad (8.16)$$

la funzione caratteristica di variabile reale

$$\phi_x(t) = e^{\alpha(e^{it}-1)} ,$$

e la funzione caratteristica di variabile complessa

$$\phi_x(z) = e^{\alpha(z-1)} . \quad (8.17)$$

Da esse potrebbero essere ricavati tutti i momenti successivi; i primi quattro valgono

$$\left\{ \begin{array}{l} \lambda_1 \equiv E(x) = \alpha \\ \lambda_2 \equiv E(x^2) = \alpha(\alpha + 1) \\ \lambda_3 \equiv \alpha[(\alpha + 1)^2 + \alpha] \\ \lambda_4 \equiv \alpha[\alpha^3 + 6\alpha^2 + 7\alpha + 1] \end{array} \right. \quad \left\{ \begin{array}{l} \mu_1 \equiv 0 \\ \mu_2 \equiv \text{Var}(x) = \alpha \\ \mu_3 \equiv \alpha \\ \mu_4 \equiv \alpha(3\alpha + 1) \end{array} \right.$$

Un'altra conseguenza della (8.16) è che la somma $w = x + y$ di due variabili casuali indipendenti che seguano la distribuzione di Poisson (con valori medi ξ ed η) segue anch'essa tale distribuzione (con valore medio pari a $\xi + \eta$):

$$M_w(t) = e^{\xi(e^t-1)} e^{\eta(e^t-1)} = e^{(\xi+\eta)(e^t-1)} . \quad (8.18)$$

Anche la distribuzione di Poisson, come si vede dai grafici di figura 8f, è bene approssimata da una distribuzione normale quando α è abbastanza elevato; questo non deve stupire, visto lo stretto legame che esiste tra la distribuzione di Poisson e quella di Bernoulli — il cui limite per grandi N è appunto la funzione di Gauss. Volendo, si potrebbe ripetere per la funzione generatrice (8.16) una analisi analoga a quella a suo tempo compiuta per la distribuzione binomiale; in questo modo si proverebbe rigorosamente il fatto che anche la distribuzione di Poisson, per grandi α , tende a quella normale.

In genere si ritiene che, per valori medi $\alpha \gtrsim 8$, si possa ritenere soddisfacente l'approssimazione normale alla distribuzione di Poisson.

8.5.1 Applicazione: esperimenti “negativi”

Si osserva un numero N_0 di protoni per un tempo t , e non si registra alcun decadimento. Quale è il limite inferiore che si può dare sulla vita media del protone, τ , con una probabilità (*livello di confidenza*) del 95%?

L'evento casuale consistente nel non avere osservato alcun decadimento è somma logica di altri due eventi mutuamente esclusivi: o il protone è stabile (e non può quindi decadere); o il protone è instabile, e si sono inoltre verificati 0 decadimenti nel tempo di osservazione (supponiamo per semplicità che ognuno di essi abbia poi probabilità 1 di essere osservato).

In questa seconda eventualità, dalla (8.10) si può ricavare il numero medio di decadimenti attesi nel tempo t , che è

$$\alpha = N_0 \left(1 - e^{-\frac{t}{\tau}}\right) \approx N_0 \frac{t}{\tau}$$

(supponendo che τ sia molto maggiore del periodo di osservazione t); e da esso la probabilità di osservare 0 eventi sempre nel tempo t , che è data dalla statistica di Poisson e vale

$$P(0) = \frac{\alpha^0}{0!} e^{-\alpha} = e^{-\alpha}.$$

Quello che si domanda è di calcolare, assumendo come certa l'ipotesi che il protone sia instabile, il valore minimo che deve avere la sua vita media perché la probabilità di non osservare nulla sia almeno del 95%: e questo avviene quando

$$P(0) = e^{-\alpha} \approx e^{-N_0 \frac{t}{\tau}} \geq 0.95$$

$$-N_0 \frac{t}{\tau} \geq \ln 0.95$$

$$\tau \geq -\frac{N_0 t}{\ln 0.95}$$

(abbiamo invertito il segno della disuguaglianza nell'ultimo passaggio perché $\ln 0.95 \approx -0.0513$ è un numero negativo).

8.5.2 Applicazione: ancora il rapporto di asimmetria

Nel paragrafo 8.4.2 abbiamo supposto che il numero N di osservazioni effettuate sia noto a priori e costante: però questo non è in generale corretto; e, nella realtà, il numero di volte in cui un certo fenomeno fisico si presenterà è di norma esso stesso una variabile casuale. Continuiamo la nostra analisi

supponendo che si tratti di un fenomeno nel quale N segua la distribuzione di Poisson con valore medio ν .

Continuando ad usare gli stessi simboli del paragrafo 8.4.2, la probabilità congiunta di osservare N eventi dei quali F in avanti è in realtà

$$\begin{aligned}\Pr(F, N) &= \frac{\nu^N}{N!} e^{-\nu} \cdot \frac{N!}{F! (N-F)!} p^F q^{N-F} \\ &= \frac{p^F q^{N-F}}{F! (N-F)!} \nu^N e^{-\nu} ;\end{aligned}$$

o anche, cambiando coppia di variabili casuali passando da $\{F, N\}$ a $\{F, B\}$:

$$\begin{aligned}\Pr(F, B) &= \frac{p^F q^B}{F! B!} \nu^{F+B} e^{-\nu} \\ &= \frac{(\nu p)^F (\nu q)^B}{F! B!} e^{-\nu} \\ &= \frac{(\nu p)^F}{F!} e^{-\nu p} \cdot \frac{(\nu q)^B}{B!} e^{-\nu q} .\end{aligned}$$

che è il *prodotto di due funzioni di frequenza di Poisson*.

In definitiva abbiamo scoperto che la composizione di un processo di Poisson e di un processo di Bernoulli equivale al prodotto di due Poissoniane: il numero N di eventi osservato segue la statistica di Poisson; la scelta dello stato finale F o B quella binomiale; ma tutto avviene come se i decadimenti dei due tipi, in avanti ed all'indietro, si verificassero *separatamente* ed *indipendentemente* secondo la statistica di Poisson.

Accettato questo fatto appena dimostrato (ancorché inaspettato), e pensando sia ad F che a B come variabili casuali statisticamente indipendenti tra loro e che seguono singolarmente la statistica di Poisson, per il rapporto di asimmetria *asintoticamente* (ovvero per grandi N) si ricava:

$$\text{Var}(F) = E(F) \simeq F$$

$$\text{Var}(B) = E(B) \simeq B$$

e, per il rapporto di asimmetria R :

$$\begin{aligned}R &= \frac{F - B}{F + B} \\ &\simeq \frac{E(F) - E(B)}{E(F) + E(B)} + \frac{\partial R}{\partial F} [F - E(F)] + \frac{\partial R}{\partial B} [B - E(B)] ;\end{aligned}$$

visto che la speranza matematica degli ultimi due termini è nulla,

$$\begin{aligned}
 E(R) &\simeq \frac{E(F) - E(B)}{E(F) + E(B)} \\
 &= \frac{2F}{N} - 1 ; \\
 \text{Var}(R) &\simeq \left(\frac{\partial R}{\partial F} \right)^2 \text{Var}(F) + \left(\frac{\partial R}{\partial B} \right)^2 \text{Var}(B) \\
 &= \frac{4}{(F+B)^4} \left[B^2 \text{Var}(F) + F^2 \text{Var}(B) \right] \\
 &= \frac{4FB}{(F+B)^3} \\
 &= 4 \frac{FB}{N^3} ,
 \end{aligned}$$

e le cose, di fatto, non cambiano (almeno nel limite dei grandi N) rispetto alla precedente analisi del paragrafo 8.4.2.

8.5.3 La distribuzione esponenziale

Alla distribuzione di Poisson ne è strettamente legata un'altra, quella *esponenziale*: sia infatti un fenomeno casuale qualsiasi che segua la distribuzione di Poisson, ovvero tale che la probabilità di osservare x eventi nell'intervallo finito di tempo t sia data dalla (8.13); definiamo una nuova variabile casuale, δ , come l'intervallo di tempo che intercorre tra due eventi successivi.

Visto che in un tempo δ nessun evento deve venire osservato, la probabilità che δ risulti maggiore di un valore predeterminato d coincide con la probabilità di osservare zero eventi nel tempo d :

$$\Pr(\delta > d) \equiv \Pr(0; d) = e^{-\lambda d}$$

e quindi la *funzione di distribuzione* di δ è la

$$F(d) = \Pr(\delta \leq d) = 1 - e^{-\lambda d}$$

Come conseguenza, la funzione di frequenza è *esponenziale*:

$$f(\delta) = \frac{dF(\delta)}{d\delta} = \lambda e^{-\lambda \delta} ; \quad (8.19)$$

e, volendo, da essa si può ricavare la funzione caratteristica — che vale

$$\phi_{\delta}(t) = \frac{\lambda}{\lambda - it} \ .$$

I momenti successivi della distribuzione esponenziale si possono ottenere o integrando direttamente la funzione densità di probabilità (moltiplicata per potenze opportune di δ) o derivando successivamente la funzione caratteristica; troviamo i primi due momenti, speranza matematica e varianza, usando questo secondo metodo:

$$\frac{d\phi_{\delta}(t)}{dt} = \frac{d}{dt} \left(\frac{\lambda}{\lambda - it} \right) = \frac{i\lambda}{(\lambda - it)^2}$$

$$\left. \frac{d\phi_{\delta}(t)}{dt} \right|_{t=0} = \frac{i}{\lambda} \equiv i \cdot E(\delta)$$

per cui la speranza matematica di δ vale

$$E(\delta) = \frac{1}{\lambda} \ ;$$

poi

$$\frac{d^2\phi_{\delta}(t)}{dt^2} = \frac{-i\lambda \cdot 2(\lambda - it)(-i)}{(\lambda - it)^4} = -\frac{2\lambda}{(\lambda - it)^3}$$

$$\left. \frac{d^2\phi_{\delta}(t)}{dt^2} \right|_{t=0} = -\frac{2}{\lambda^2} \equiv i^2 E(\delta^2) = -E(\delta^2) \ ,$$

ed infine la varianza è

$$\text{Var}(\delta) = E(\delta^2) - [E(\delta)]^2 = \frac{1}{\lambda^2} \ .$$

Se una variabile casuale t rappresenta il tempo trascorso tra due eventi casuali successivi che seguono una distribuzione di Poisson, t necessariamente ha una distribuzione di probabilità di tipo esponenziale data dalla (8.19); vogliamo ora calcolare la probabilità che t sia maggiore di una quantità $t_0 + \Delta t$, *condizionata* però dal sapere in anticipo che t è sicuramente

maggiore di t_0 . Sfruttando la (3.3), abbiamo:

$$\begin{aligned}
 \Pr(t > t_0 + \Delta t \mid t > t_0) &= \frac{\Pr(t > t_0 + \Delta t)}{\Pr(t > t_0)} \\
 &= \frac{\int_{t_0+\Delta t}^{+\infty} e^{-\lambda t} dt}{\int_{t_0}^{+\infty} e^{-\lambda t} dt} \\
 &= \frac{[-e^{-\lambda t}]_{t_0+\Delta t}^{+\infty}}{[-e^{-\lambda t}]_{t_0}^{+\infty}} \\
 &= \frac{e^{-\lambda(t_0+\Delta t)}}{e^{-\lambda t_0}} \\
 &= e^{-\lambda \Delta t} \\
 &\equiv \Pr(t > \Delta t) \quad .
 \end{aligned}$$

In conclusione, la distribuzione esponenziale (ovvero la cadenza temporale di eventi casuali che seguono la statistica di Poisson) *non ricorda la storia precedente*: il presentarsi o meno di uno di tali eventi in un tempo Δt non dipende in alcun modo da quello che è accaduto nell'arbitrario intervallo di tempo t_0 precedente; così come ci dovevamo aspettare, vista l'ipotesi numero 2 formulata a pagina 117.

8.5.4 La distribuzione di Erlang

La funzione di frequenza esponenziale (8.19) si può considerare come un caso particolare di un'altra funzione di frequenza, detta *di Erlang*¹². Supponiamo di voler trovare la densità di probabilità $f_n(t; \lambda)$ dell'evento casuale consistente nel presentarsi, dopo un tempo t , dell' n -esimo di una serie di altri eventi che seguano la statistica di Poisson con costante di tempo λ ; la (8.19) è ovviamente la prima di esse, $f_1(t; \lambda) = \lambda e^{-\lambda t}$.

Il secondo evento si manifesta dopo un tempo t con densità di probabilità

¹²Agner Krarup Erlang fu un matematico danese vissuto dal 1878 al 1929; si occupò di analisi e di fisica oltre che di statistica. Dette notevoli contributi alla tabulazione di varie funzioni, ed applicò in particolare la statistica a numerosi problemi relativi al traffico telefonico.

data da

$$\begin{aligned}
 f_2(t; \lambda) &= \int_0^t f_1(x; \lambda) f_1(t-x; \lambda) dx \\
 &= \lambda^2 \int_0^t e^{-\lambda x} e^{-\lambda(t-x)} dx \\
 &= \lambda^2 e^{-\lambda t} \int_0^t dx \\
 &= \lambda^2 t e^{-\lambda t} ;
 \end{aligned}$$

si è infatti supposto che il primo dei due eventi si sia presentato dopo un tempo x (con $0 < x < t$), si è sfruttata l'indipendenza statistica degli eventi casuali tra loro ed infine si è sommato su tutti i possibili valori di x . Allo stesso modo

$$\begin{aligned}
 f_3(t; \lambda) &= \int_0^t f_2(x; \lambda) f_1(t-x; \lambda) dx \\
 &= \lambda^3 \int_0^t x e^{-\lambda x} e^{-\lambda(t-x)} dx \\
 &= \lambda^3 e^{-\lambda t} \int_0^t x dx \\
 &= \frac{t^2}{2} \lambda^3 e^{-\lambda t} ;
 \end{aligned}$$

la formula generale (appunto la *funzione di frequenza di Erlang*) è la

$$f_n(t; \lambda) = \frac{t^{n-1}}{(n-1)!} \lambda^n e^{-\lambda t} ,$$

con speranza matematica

$$E(t) = \frac{n}{\lambda}$$

e varianza

$$\text{Var}(t) = \frac{n}{\lambda^2} .$$

8.5.5 La distribuzione composta di Poisson

La *distribuzione composta di Poisson* è quella seguita da una variabile che sia somma di un numero casuale N di valori di un'altra variabile casuale x , quando sia N che x seguono singolarmente delle distribuzioni di Poisson. Indichiamo con ν e ξ i valori medi delle popolazioni delle variabili N e x rispettivamente; le funzioni caratteristiche (di variabile complessa) ad esse associate sono date, come sappiamo, dalla (8.17):

$$\phi_N(z) = e^{\nu(z-1)} \quad \text{e} \quad \phi_x(z) = e^{\xi(z-1)} ;$$

ricordando la (6.14), la variabile casuale

$$S = \sum_{i=1}^N x_i$$

ha funzione caratteristica di variabile complessa

$$\phi_S(z) = \phi_N[\phi_x(z)] = e^{\nu[e^{\xi(z-1)} - 1]}$$

e funzione caratteristica di variabile reale

$$\phi_S(t) = e^{\nu[e^{\xi(e^{it}-1)} - 1]} ;$$

da quest'ultima poi si ricava

$$\frac{d\phi_S(t)}{dt} = \phi_S(t) \cdot \nu e^{\xi(e^{it}-1)} \cdot \xi e^{it} \cdot i$$

ed infine

$$\left. \frac{d\phi_S(t)}{dt} \right|_{t=0} = i\nu\xi .$$

La speranza matematica di una variabile che segua la distribuzione composta di Poisson vale quindi

$$E(S) = \nu\xi ,$$

e, similmente, si potrebbero ottenere

$$\text{Var}(S) = \nu\xi(1 + \xi)$$

per la varianza, e

$$\Pr(S) = \sum_{N=0}^{\infty} \left[\frac{(N\xi)^S}{S!} e^{-N\xi} \cdot \frac{\nu^N}{N!} e^{-\nu} \right]$$

per la funzione di frequenza.

Quest'ultima formula non sorprende: è la somma (su tutti i valori ammissibili) della probabilità di ottenere un determinato N , moltiplicata per la probabilità di ottenere il valore di S condizionato da quello di N ; infatti la somma di N variabili indipendenti distribuite secondo Poisson con valore medio ξ è ancora, in base a quanto dedotto dall'equazione (8.18), una variabile distribuita secondo Poisson e con valore medio $N\xi$.

8.5.6 Esempio: l'osservazione di un quark isolato

Un esempio classico di applicazione delle formule precedenti è la discussione di un articolo del 1969¹³ in cui veniva annunciata l'osservazione di un quark isolato; l'esperienza ivi descritta consisteva nell'analizzare foto esposte in una camera a nebbia, attraversata da un fascio di particelle (aventi carica unitaria) che generavano tracce con un numero medio di gocce per unità di lunghezza $\alpha = 229$: su 55'000 tracce osservate ce ne era una con un numero di gocce per unità di lunghezza $n = 110$.

Questa traccia apparteneva indiscutibilmente al fascio di particelle; la probabilità che venisse osservato un numero di gocce per unità di lunghezza pari (o inferiore) a 110, se il fenomeno è descritto da una distribuzione di Poisson con media 229, è data da

$$\Pr(n \leq 110) = \sum_{k=0}^{110} \frac{229^k}{k!} e^{-229} \approx 1.6 \times 10^{-18}$$

e risulta ben inferiore (per 13 ordini di grandezza!) alla frequenza osservata $f = 1/55'000 \approx 2 \times 10^{-5}$. Per questo motivo gli autori sostenevano di avere osservato una particella con carica frazionaria (un quark), e che causava in conseguenza una ionizzazione assai inferiore alle altre.

Una prima obiezione¹⁴ fu che, in ogni urto elementare tra le particelle del fascio e le molecole di gas della camera, vengono generati in media $\nu = 4$ prodotti ionizzati indipendenti: e quindi 4 gocce. Il numero medio effettivo

¹³McCusker e Cairns: Evidence of quarks in air-shower cores; Phys. Rev. Lett. **23** (1969), pagg. 658-659.

¹⁴Adair e Kasha: Analysis of some results of quark searches; Phys. Rev. Lett. **23** (1969), pagg. 1355-1358.

di urti per unità di lunghezza era $229/4 = 57.25$, mentre la traccia osservata ne aveva invece $\lambda = 110/4 = 27.5$; la probabilità di ottenere, per motivi puramente casuali, una fluttuazione almeno pari a quella osservata doveva essere quindi calcolata come probabilità di avere meno di 28 eventi da una distribuzione di Poisson con valore medio 57.25, che vale

$$\Pr(n \leq 110) = \sum_{k=0}^{27} \frac{57.25^k}{k!} e^{-57.25} \approx 6.7 \times 10^{-6}$$

ed è quindi assai maggiore di quanto venisse ipotizzato all'inizio dai due autori (pur mantenendosi più di 33 volte superiore alla frequenza osservata).

L'analisi del fenomeno è però ancora più complessa¹⁵: il numero u di urti elementari per unità di lunghezza segue la distribuzione di Poisson con valore medio $\lambda = 57.25$, ed ogni urto genera un numero di gocce che non è costante, ma segue anch'esso la distribuzione di Poisson con valore medio $\nu = 4$; quindi il numero complessivo di gocce segue una legge di distribuzione che è quella *composta di Poisson*.

La probabilità di osservare k gocce per unità di lunghezza è quindi

$$\Pr(k) = \sum_{u=0}^{\infty} \left[\frac{(u\nu)^k}{k!} e^{-u\nu} \cdot \frac{\lambda^u}{u!} e^{-\lambda} \right],$$

e la probabilità cercata vale

$$\Pr(k \leq 110) = \sum_{k=0}^{110} \Pr(k) \approx 4.7 \times 10^{-5}$$

(ben compatibile quindi con quanto osservato).

8.5.7 Applicazione: segnale e fondo

Supponiamo di osservare sperimentalmente un processo fisico, per il quale il numero di eventi s che potrebbero presentarsi in un intervallo temporale prefissato (eventi di *segnale*) segua una distribuzione di Poisson con valore medio S , e che indicheremo col simbolo $P(s; S)$;

$$\Pr(s) = P(s; S) = \frac{S^s}{s!} e^{-S} :$$

in generale S è ignoto, e ci si propone appunto di determinarlo dall'esperimento. Questo problema verrà poi trattato anche nel paragrafo 11.2.1 a

¹⁵Eadie, Drijard, James, Roos e Sadoulet: Statistical Methods in Experimental Physics; North-Holland Publishing Co. (1971), pag. 53.

pagina 173, usando altri metodi; qui vogliamo solo vedere come, partendo dal dato di fatto consistente nell'osservazione effettiva di N eventi in un singolo esperimento, si possa ricavare un *limite superiore* sui possibili valori di S .

Fissato arbitrariamente un valore della probabilità ϵ , si tratta di trovare il valore S_u per il quale la probabilità di osservare un numero di eventi non superiore ad N vale proprio ϵ : vale a dire, risolvere rispetto a S_u l'equazione

$$\sum_{s=0}^N P(s; S_u) = \epsilon ;$$

e diremo poi di poter affermare che $S \leq S_u$ con un *livello di confidenza* ϵ . Il significato esatto della frase è che, se risultasse realmente $S \leq S_u$, in una frazione pari almeno ad ϵ di esperimenti analoghi a quello i cui risultati stiamo esaminando ci aspetteremmo di ottenere al massimo N eventi come in esso.

Le cose si complicano in presenza di processi fisici che possono produrre risultati che simulano il segnale: processi fisici che indicheremo col nome complessivo di *fondo*. Se il fondo è presente, se è inoltre indipendente dal segnale e se segue la distribuzione di Poisson con valore medio *noto* F , già sappiamo che la probabilità di osservare N eventi in tutto *tra fondo e segnale* segue ancora la distribuzione di Poisson, con valore medio $F + S$:

$$\Pr(N) \equiv P(N; F + S) = \frac{(F + S)^N}{N!} e^{-(F+S)} .$$

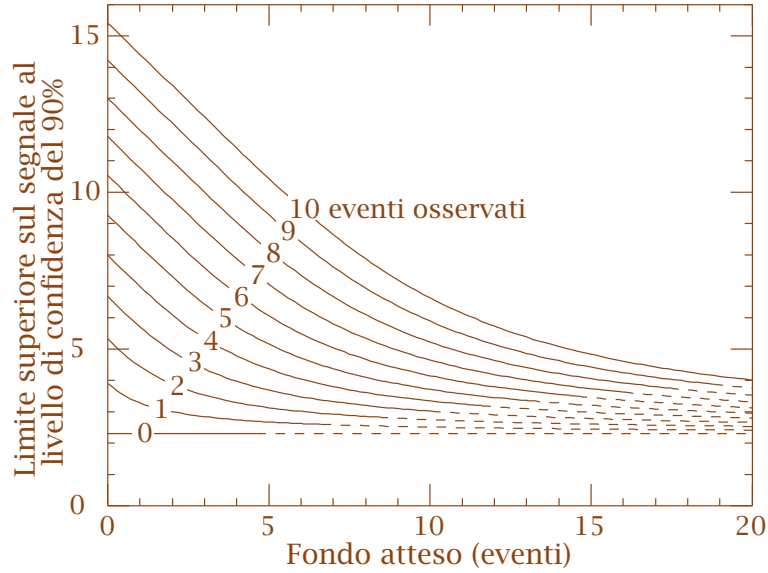
Se sono stati realmente osservati N eventi, si può ancora determinare un *limite superiore per S* ; questo calcolando il valore S_u per il quale la probabilità di osservare un numero di eventi (tra fondo e segnale) non superiore a N e *condizionato all'avere ottenuto un numero di eventi di fondo che non può superare N* vale una quantità predeterminata ϵ . Insomma, si tratta di risolvere, rispetto a S_u , l'equazione

$$\frac{\sum_{n=0}^N P(n; F + S_u)}{\sum_{f=0}^N P(f; F)} = \epsilon ; \quad (8.20)$$

e, con lo steso significato della frase prima evidenziato, potremo allora affermare che risulta $S \leq S_u$ con un libello di confidenza ϵ .

Nella figura 8g, che è tratta dal libretto pubblicato annualmente dal "Particle Data Group" (vedi la bibliografia), si possono trovare già calcolate e rappresentate da curve continue le soluzioni dell'equazione (8.20) relative ad un livello di confidenza fisso del 90%.

FIGURA 8g - I limiti superiori sul segnale, ad un livello di confidenza fisso $\epsilon = 90\%$, in presenza di fondo noto.



8.6 La distribuzione log-normale

Sia una variabile casuale y avente distribuzione normale con media μ e varianza σ^2 : la sua densità di probabilità sia insomma data dalla funzione

$$g(y) = N(y; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2};$$

definiamo poi una nuova variabile casuale x attraverso la relazione

$$x = e^y \quad \Longleftrightarrow \quad y = \ln x$$

(la corrispondenza tra x ed y è ovviamente biunivoca). Il dominio di definizione della x è il semiasse $x > 0$ e, in base alla (6.15), la sua densità di probabilità $f(x)$ sarà data da

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2}.$$

Questa funzione di frequenza si chiama *log-normale*; sfruttando l'identi-

tà

$$\begin{aligned}
 \frac{(y - \mu - k\sigma^2)^2}{2\sigma^2} - k\mu - \frac{1}{2}k^2\sigma^2 &= \frac{1}{2\sigma^2} (y^2 + \mu^2 + k^2\sigma^4 - 2\mu y \\
 &\quad - 2k\sigma^2 y + 2k\mu\sigma^2 - 2k\mu\sigma^2 - k^2\sigma^4) \\
 &= \frac{1}{2\sigma^2} [(y^2 - 2\mu y + \mu^2) - 2k\sigma^2 y] \\
 &= \frac{(y - \mu)^2}{2\sigma^2} - ky
 \end{aligned}$$

se ne possono facilmente calcolare i momenti rispetto all'origine, che valgono

$$\begin{aligned}
 \lambda_k &= \int_0^{+\infty} x^k f(x) dx \\
 &= \int_{-\infty}^{+\infty} e^{ky} g(y) dy \\
 &= \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(y-\mu)^2}{2\sigma^2} - ky\right]} dy \\
 &= e^{(k\mu + \frac{1}{2}k^2\sigma^2)} \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu-k\sigma^2)^2}{2\sigma^2}} dy \\
 &= e^{(k\mu + \frac{1}{2}k^2\sigma^2)}
 \end{aligned}$$

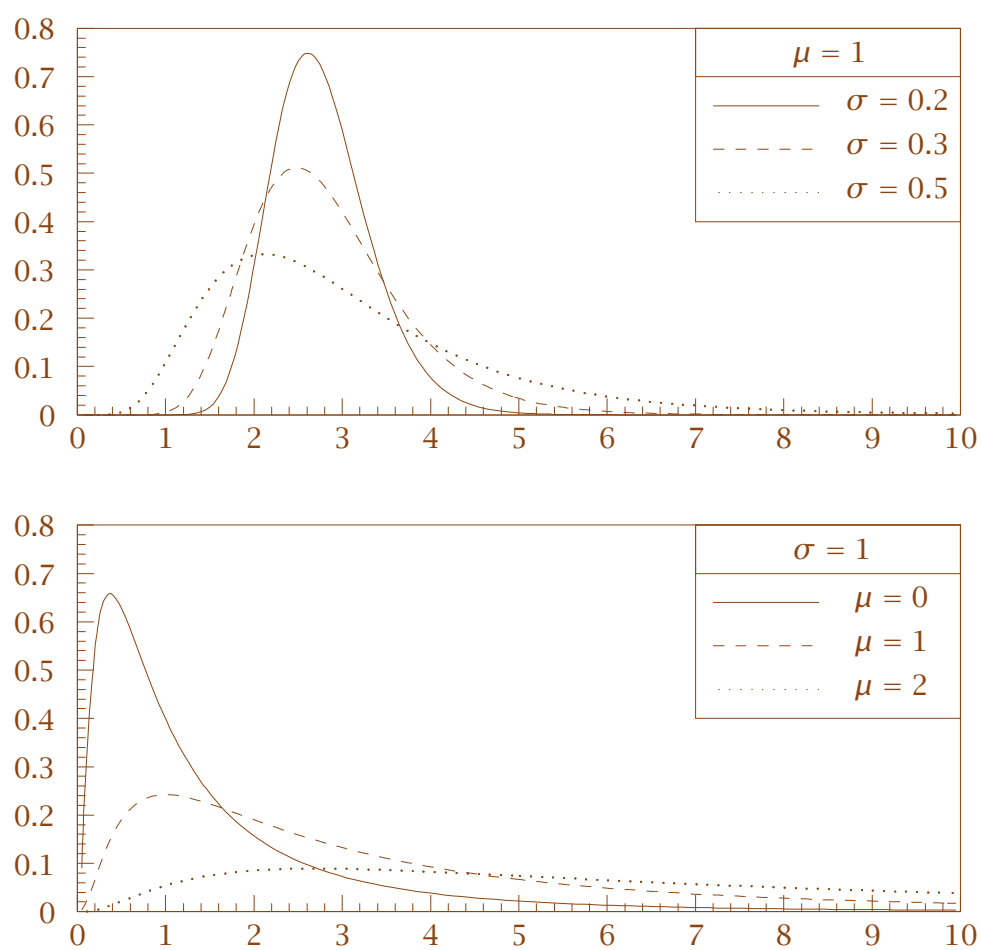
(infatti l'integrale è quello di una distribuzione normale avente $\mu + k\sigma^2$ come valore medio e σ come varianza — e vale dunque uno).

In particolare

$$\begin{aligned}
 E(x) &\equiv \lambda_1 = e^{(\mu + \frac{\sigma^2}{2})} \\
 E(x^2) &\equiv \lambda_2 = e^{(2\mu + 2\sigma^2)} \\
 \text{Var}(x) &= \lambda_2 - \lambda_1^2 = e^{(2\mu + \sigma^2)} (e^{\sigma^2} - 1)
 \end{aligned}$$

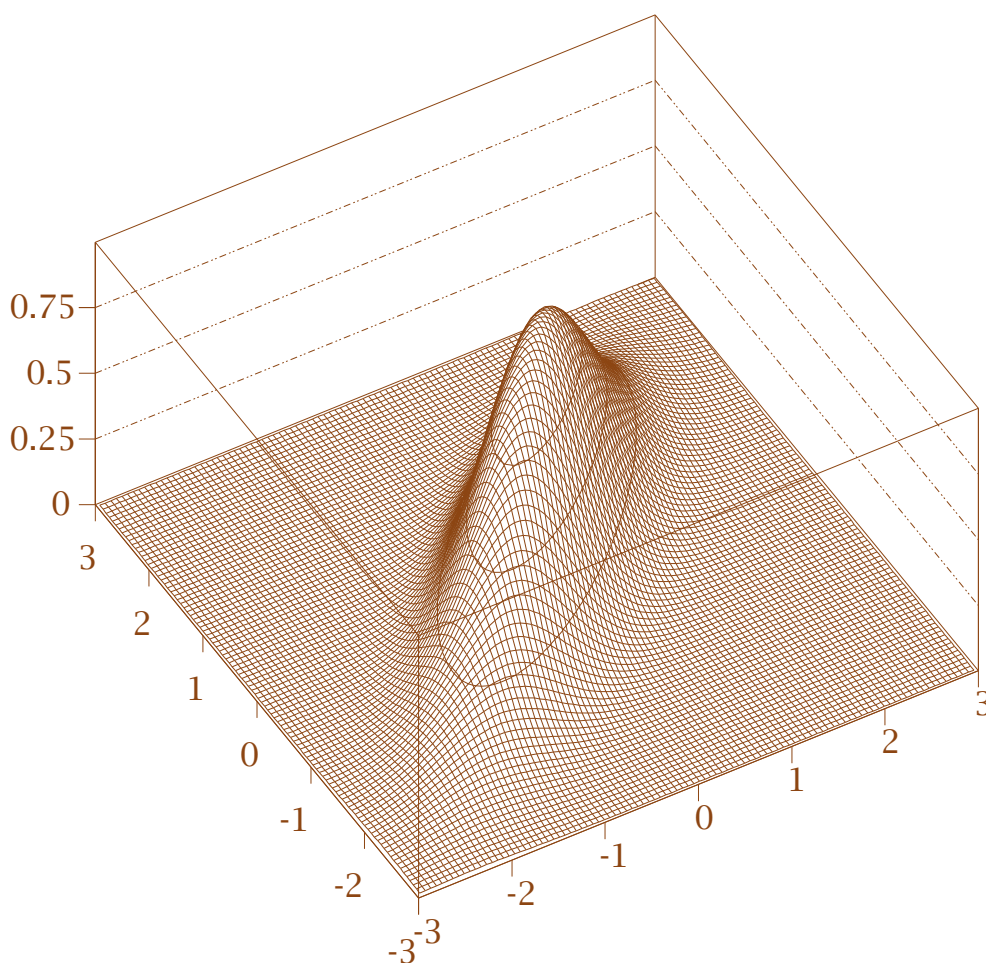
Nella figura 8h ci sono i grafici di alcune distribuzioni log-normali corrispondenti a vari valori dei parametri μ e σ della funzione normale di partenza (*non* della funzione di frequenza esaminata); per finire notiamo che, analogamente a quanto ricavato nel teorema di pagina 103 per quanto attiene alle somme, si può dimostrare che il *prodotto* di variabili casuali *log-normali* ed indipendenti debba seguire una *distribuzione log-normale*.

FIGURA 8h - La distribuzione log-normale, per vari valori dei parametri (μ e σ) della funzione normale di partenza.



8.7 La distribuzione normale in più dimensioni

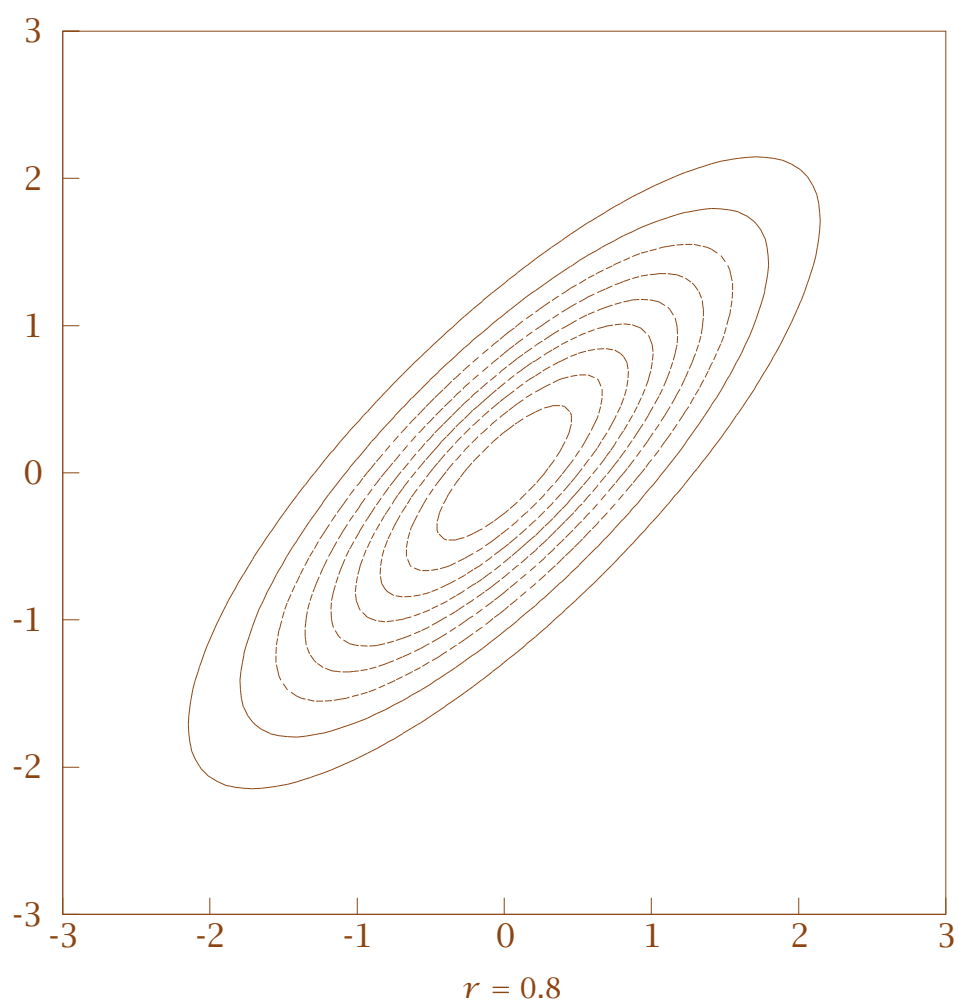
FIGURA 8i - La funzione normale in due dimensioni, nel piano $\{u, v\}$ delle variabili standardizzate e per un coefficiente di correlazione $r = 0.8$.



Accenniamo solo brevemente alla cosiddetta *distribuzione normale bidimensionale*: per essa la densità di probabilità congiunta delle due variabili x ed y è, nel caso generale, data dalla

$$f(x, y) = \frac{e^{\left\{ -\frac{1}{2(1-r^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 - 2r \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x \sigma_y} + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right] \right\}}}{2\pi \sigma_x \sigma_y \sqrt{1-r^2}}$$

FIGURA 8j - Le sole curve di livello della stessa funzione rappresentata nella figura 8i.



o, espressa in funzione delle *variabili standardizzate*

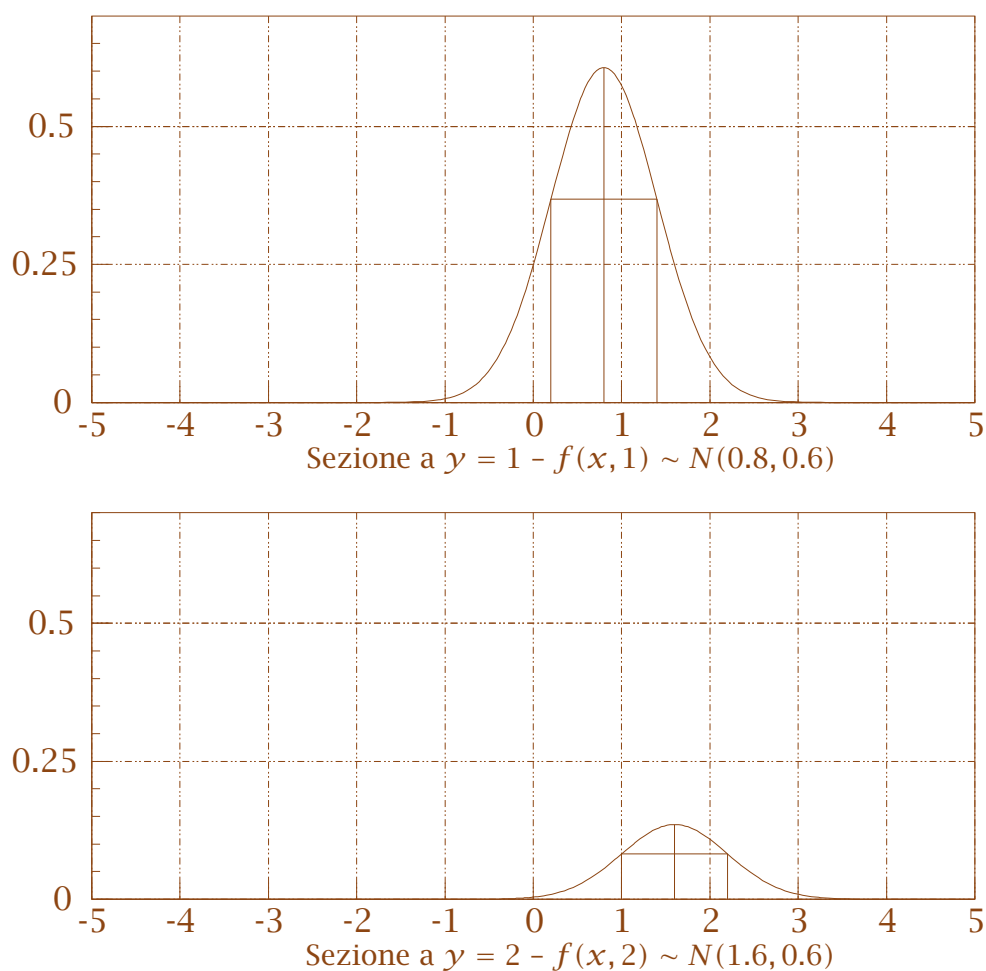
$$u = \frac{x - E(x)}{\sigma_x} \quad \text{e} \quad v = \frac{y - E(y)}{\sigma_y},$$

dalla

$$f(u, v) = \frac{e^{-\frac{1}{2} \frac{u^2 - 2ruv + v^2}{1-r^2}}}{2\pi\sqrt{1-r^2}};$$

un esempio è dato dalle figure 8i e 8j.

FIGURA 8k - Le sezioni della figura 8i con due piani di equazione $y = 1$ e $y = 2$ rispettivamente: ovvero (a parte la normalizzazione) le densità di probabilità condizionate $\pi(x|y = 1)$ e $\pi(x|y = 2)$.



r è il coefficiente di correlazione lineare tra x ed y ; $r = 0$ è condizione *necessaria e sufficiente* perché le due variabili siano statisticamente indipendenti tra loro. Le due distribuzioni marginali $g(x)$ e $h(y)$ sono due funzioni normali, aventi speranza matematica e varianza μ_x e σ_x^2 la prima, e μ_y e σ_y^2 la seconda:

$$g(x) = N(x; \mu_x, \sigma_x) \quad \text{e} \quad h(y) = N(y; \mu_y, \sigma_y) .$$

Le densità di probabilità condizionate sono anch'esse sempre delle funzioni normali; come esempio, nella figura 8k si mostrano due di queste funzioni per la stessa distribuzione di figura 8i.

Nel caso più generale, la densità di probabilità di una distribuzione di Gauss N -dimensionale è del tipo

$$f(x_1, x_2, \dots, x_N) = K e^{-\frac{1}{2}H(x_1, x_2, \dots, x_N)} , \quad (8.21)$$

ove H è una *forma quadratica* nelle variabili standardizzate

$$t_i = \frac{x_i - E(x_i)}{\sigma_i}$$

nella quale però i coefficienti dei termini quadratici sono *tutti uguali*; le t_i non sono generalmente indipendenti tra loro, e quindi H contiene anche i termini rettangolari del tipo $t_i \cdot t_j$.

K , nella (8.21), è un fattore di normalizzazione che vale

$$K = \sqrt{\frac{\Delta}{(2\pi)^N}} ;$$

a sua volta, Δ è il determinante della matrice (*simmetrica*) dei coefficienti della forma quadratica H . La condizione, poi, che la f di equazione (8.21) debba essere integrabile implica che le ipersuperfici di equazione $H = \text{cost.}$ siano tutte al finito, e siano quindi *iperellissi* nello spazio N -dimensionale dei parametri; le funzioni di distribuzione marginali e condizionate di *qualsiasi* sottoinsieme delle variabili sono ancora *sempre normali*.

Si può inoltre dimostrare che *esiste sempre* un cambiamento di variabili $x_i \rightarrow y_k$ che muta H nella cosiddetta *forma canonica* (senza termini rettangolari); in tal caso

$$\Delta = \left[\prod_{k=1}^N \text{Var}(y_k) \right]^{-1}$$

e

$$f(y_1, \dots, y_k) = \frac{1}{\sqrt{\prod_{k=1}^N [2\pi \text{Var}(y_k)]}} e^{-\frac{1}{2} \sum_{k=1}^N \frac{[y_k - E(y_k)]^2}{\text{Var}(y_k)}} :$$

e le y_k sono quindi tutte *statisticamente indipendenti* (oltre che normali).

Capitolo 9

La legge di Gauss

Vogliamo ora investigare sulla distribuzione dei risultati delle misure ripetute di una grandezza fisica, nell'ipotesi che esse siano affette da errori esclusivamente casuali¹.

9.1 La funzione di Gauss

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

- La probabilità di ottenere un certo scarto dal valore 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 uguali probabilità; in definitiva la distribuzione degli scarti deve essere *simmetrica rispetto allo zero*.
- La probabilità di ottenere un certo scarto dal valore vero (in modulo) deve essere *decrescente* al crescere di tale scarto e *tendere a zero* quando esso tende all'infinito; questo perché deve essere più probabile commettere errori piccoli che errori grandi, ed infinitamente improbabile commettere errori infinitamente grandi.

¹Il primo ad intuire la forma e l'equazione della distribuzione normale fu Abraham de Moivre nel 1733, che la derivò dalla distribuzione binomiale facendo uso della formula di Stirling per il fattoriale; fu poi studiata da Laplace, ma la teoria completa è dovuta a Gauss.

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 condizione di normalizzazione, ossia l'equazione (6.1) di cui abbiamo già parlato prima:

- L'area compresa tra la curva densità di probabilità dello scarto e l'asse delle ascisse, da $-\infty$ a $+\infty$, deve valere 1.

Da queste tre ipotesi e dal principio della media aritmetica, Gauss² dimostrò in modo euristico che la distribuzione degli scarti z delle misure affette da errori casuali è data dalla funzione

$$f(z) = \frac{h}{\sqrt{\pi}} e^{-h^2 z^2} \quad (9.1)$$

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 D.

La funzione di Gauss ha una caratteristica forma a campana: simmetrica rispetto all'asse delle ordinate (di equazione $z = 0$), decrescente man mano che ci si allontana da esso sia nel senso delle z positive che negative, e tendente a 0 per z che tende a $\pm\infty$; così come richiesto dalle ipotesi di partenza.

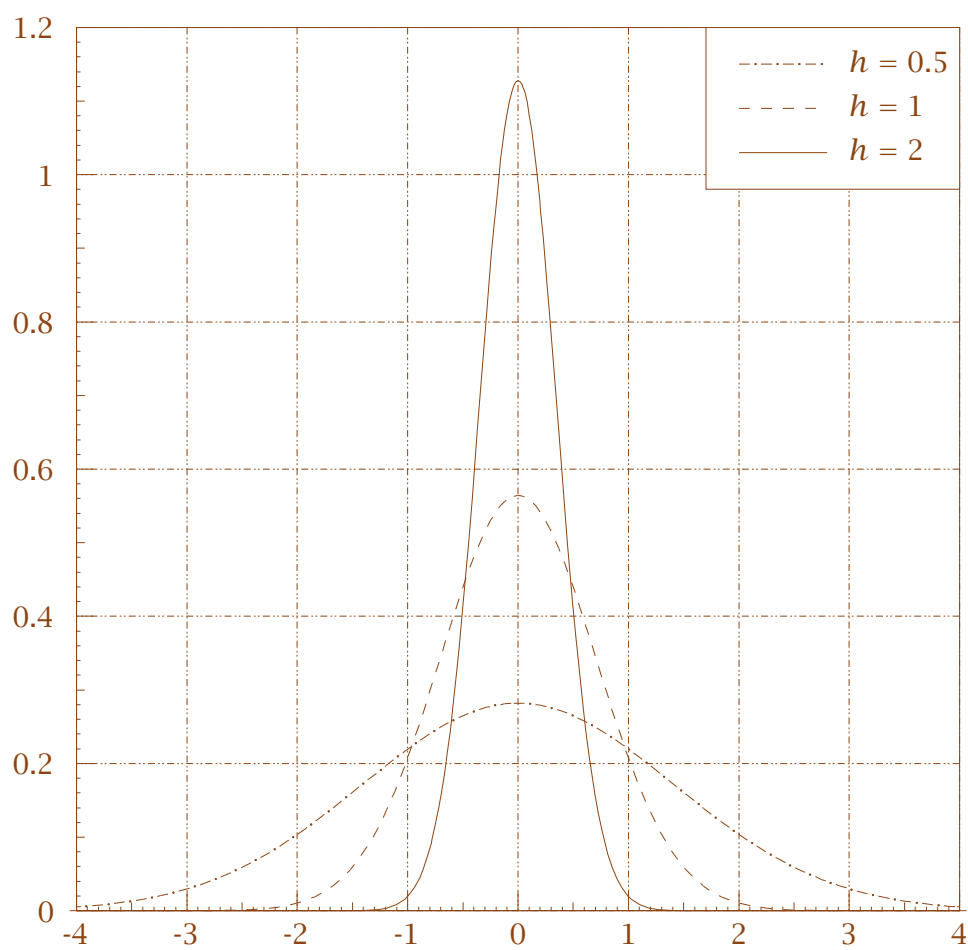
Essa dipende da un parametro $h > 0$ che prende il nome di *modulo di precisione* della misura: infatti quando h è piccolo la funzione è sensibilmente diversa da zero in una zona estesa dell'asse delle ascisse; mentre al crescere di h l'ampiezza di tale intervallo diminuisce, e la curva si stringe sull'asse delle ordinate (come si può vedere nella figura 9a).

9.2 Proprietà della legge normale

Possiamo ora, come già anticipato nei paragrafi 4.2.6 e 6.2, determinare il valore medio di una qualsiasi grandezza $W(z)$ legata alle misure, nel limite

²Karl Friedrich Gauss fu senza dubbio la maggiore personalità del primo 800 nel campo della fisica e della matematica; si occupò di teoria dei numeri, analisi, geometria analitica e differenziale, statistica e teoria dei giochi, geodesia, elettricità e magnetismo, astronomia e ottica. Visse a Göttingen dal 1777 al 1855 e, nel campo di cui ci stiamo occupando, teorizzò (tra le altre cose) la funzione normale ed il metodo dei minimi quadrati, quest'ultimo (studiato anche da Laplace) all'età di soli 18 anni.

FIGURA 9a - La funzione di Gauss per tre diversi valori di h .



di un numero infinito di misure effettuate; questo valore medio sappiamo dall'equazione (6.2) che si dovrà ricavare calcolando l'integrale

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

dove per $f(z)$ si intende la funzione densità di probabilità dello scarto dal valore vero, che supporremo qui essere la distribuzione normale (9.1).

Se vogliamo ad esempio calcolare il valore 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 prodotto $z f(z) dz$ assume valori uguali in modulo ma di segno opposto; così che la loro somma sia zero. Essendo poi

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

abbiamo così *dimostrato* quanto assunto nel paragrafo 5.2, ossia che

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

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

$$\begin{aligned} E(|z|) &= \int_{-\infty}^{+\infty} |z| f(z) dz \\ &= 2 \int_0^{+\infty} z f(z) dz \\ &= \frac{2h}{\sqrt{\pi}} \int_0^{+\infty} z e^{-h^2 z^2} dz \\ &= \frac{h}{\sqrt{\pi}} \int_0^{+\infty} e^{-h^2 t} dt \\ &= -\frac{1}{h\sqrt{\pi}} \left[e^{-h^2 t} \right]_0^{+\infty} \\ &= \frac{1}{h\sqrt{\pi}} \end{aligned}$$

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

Il valore medio del modulo degli scarti è quella grandezza che abbiamo definito come “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 valore medio del quadrato degli scarti:

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

Per giungere al risultato, per prima cosa si è effettuata la sostituzione di variabile $t = hz$; poi si è integrato per parti; ed 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 sono invece legati dalla

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

Sostituendo nella (9.1) il valore di h in funzione di σ , la legge di Gauss si può quindi anche scrivere nella forma equivalente

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}} \quad (9.2)$$

9.3 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 (9.2) 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 $\varphi(t)$ 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

$$\Pr(x \in [x_1, x_2]) = \int_{x_1}^{x_2} f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2} \left(\frac{x-x^*}{\sigma} \right)^2} dx . \quad (9.3)$$

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

$$\Pr(t \in [t_1, t_2]) = \int_{t_1}^{t_2} \varphi(t) dt . \quad (9.4)$$

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 identicamente uguale alla probabilità che t sia compresa in $[t_1, t_2]$.

Eseguiamo sull'espressione (9.3) della probabilità per x un cambiamento di variabile, sostituendovi la t :

$$\Pr(x \in [x_1, x_2]) = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{1}{2}t^2} dt . \quad (9.5)$$

Confrontando le due espressioni (9.4) e (9.5) (che, ricordiamo, devono assumere lo stesso valore per *qualunque* coppia di valori x_1 e x_2), si ricava immediatamente che deve essere

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \quad (9.6)$$

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 (9.6) 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; ad esempio si ha

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

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

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

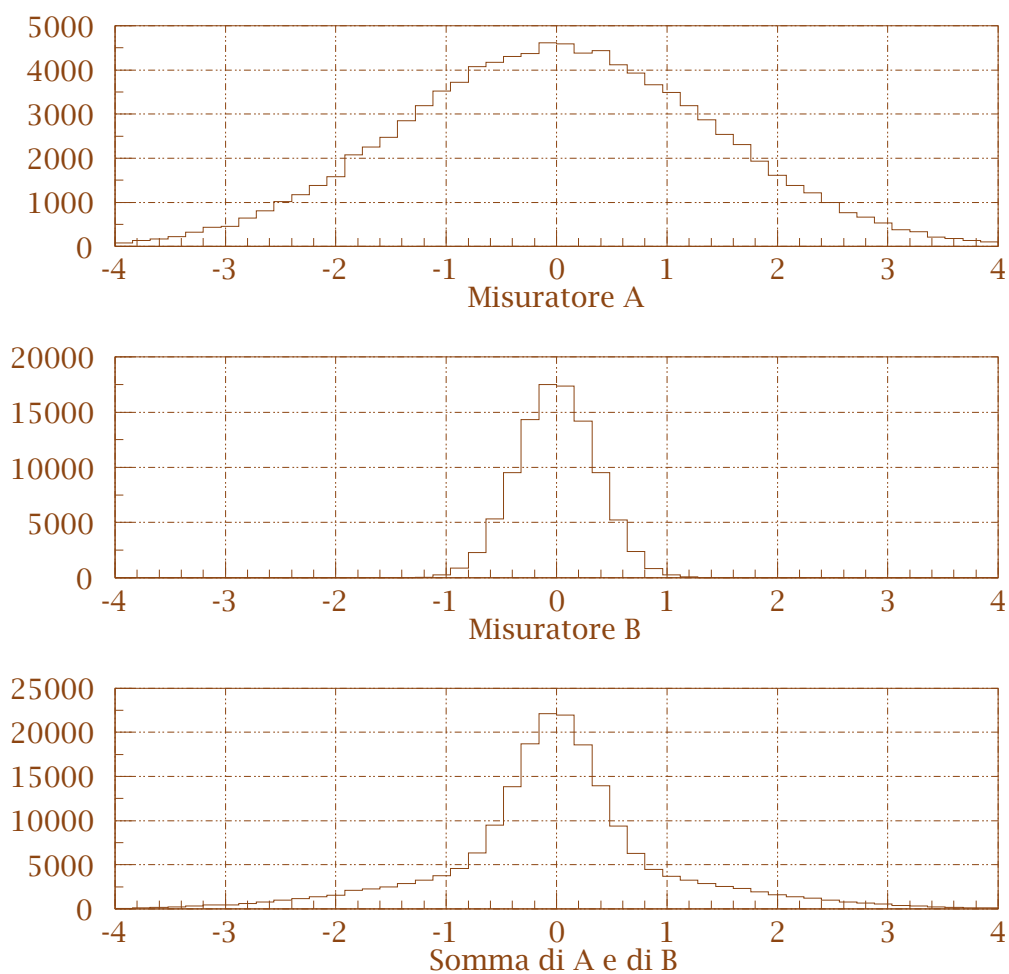
e, ricordando la relazione che intercorre tra z e t , questo implica che risulti anche

$$\Pr(z \in [-\sigma, +\sigma]) \equiv \Pr(t \in [-1, +1]) \approx 0.6827$$

$$\Pr(z \in [-2\sigma, +2\sigma]) \equiv \Pr(t \in [-2, +2]) \approx 0.9545$$

$$\Pr(z \in [-3\sigma, +3\sigma]) \equiv \Pr(t \in [-3, +3]) \approx 0.9973 .$$

FIGURA 9b - Gli istogrammi relativi a due campioni di misure aventi differente precisione, e quello relativo ai dati di entrambi i campioni.



Possiamo quindi far uso di una qualsiasi di queste relazioni per dare una *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 di un campione ha pertanto una probabilità del 68% di contenere il valore vero, sempreché gli errori siano casuali e normali.*

9.4 Il significato geometrico di σ

Calcoliamo ora la derivata prima della funzione di Gauss, nella sua forma (9.2):

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

La funzione $f(z)$ è crescente ($f'(z) > 0$) quando z è negativa, e viceversa; ha quindi un massimo per $z = 0$, come d'altronde richiesto dalle ipotesi fatte nel paragrafo 9.1 per ricavarne la forma analitica. La derivata seconda invece vale

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

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.

9.5 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 B).

Per N abbastanza grande³ 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 nella misura 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 valore 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 conforme alla funzione di Gauss (o meglio, di stimare con quale livello di probabilità possa provenire da una distribuzione normale).

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 che consiste nel 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*; si veda in proposito il paragrafo 12.2.1) 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.

³Cosa si debba intendere esattamente per "abbastanza grande" risulterà chiaro dall'analisi dell'appendice B; normalmente si richiedono almeno 30 misure, dimensione del campione che corrisponde per s ad un errore relativo di poco superiore al 10%.

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

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

Cioè l'altezza dell'istogramma, in ognuna delle classi di frequenza, tende al valore 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

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

Sempre per quel che riguarda le implicazioni “pratiche” della legge normale di distribuzione degli errori, un altro punto sul quale gli studenti hanno frequentemente dei dubbi riguarda l'applicazione della funzione di Gauss a grandezze misurate sì commettendo errori casuali, ma che siano per loro natura *limitate*. Ad esempio, una lunghezza è una grandezza fisica implicitamente non negativa: quindi la densità di probabilità associata ai particolari valori ottenibili x dovrebbe essere identicamente nulla quando $x < 0$, mentre la funzione normale si annulla soltanto quando $x = \pm\infty$. Affermare che i risultati della misura seguono la legge di Gauss sembra dunque una contraddizione.

La risposta a questa obiezione è che la funzione di distribuzione della x effettivamente *non può essere normale*: ma che la reale differenza tra la vera funzione di distribuzione e quella di Gauss è assolutamente trascurabile.

Facciamo un esempio pratico: supponiamo di voler misurare la dimensione del lato di un quaderno (di valore vero 20 cm) con un regolo graduato, e di commettere un errore di misura $\sigma = 1$ mm; la probabilità di trovare un risultato in un intervallo ampio 1 mm appena alla sinistra dello zero secondo la legge normale vale

$$p \simeq \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} 200^2} \approx 0.4 e^{-2 \times 10^4} ;$$

quindi $\ln(p) \sim -2 \times 10^4$, e $\log_{10}(p) = \ln(p) \cdot \log_{10}(e) \sim -10^4$, mentre dovrebbe essere rigorosamente $p \equiv 0$.

Per valutare le reali implicazioni di un valore di p come quello che stiamo considerando, attualmente il numero di atomi presenti nell'universo si stima essere dell'ordine di 10^{79} ; mentre l'età dell'universo stesso si stima in circa 10^{10} anni, ovvero dell'ordine di 10^{18} secondi; se pensiamo ad un gruppo di misuratori in numero pari al numero di atomi nell'universo, ed ognuno dei quali esegua una misura al secondo, dovrebbe passare un tempo pari circa a 7 volte l'età dell'universo stesso per ottenere un valore illegale qualora le misure seguissero veramente la legge di Gauss: quindi la differenza tra la funzione di distribuzione reale e quella ipotizzata è effettivamente trascurabile.

9.6 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 differiscano 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 ad esempio di 60, ci si attende che (per fluttuazioni dovute esclusivamente al caso) solo 0.18 misure (praticamente: nessuna) differiscano dal valore medio, in modulo, per più di 3σ ; se troviamo una (o più) misure di questo tipo, possiamo attribuire la loro presenza, piuttosto che ad una fluttuazione casuale, a cause d'errore del tipo di quelle considerate, quindi etichettarle come *sospette* ed infine scartarle.

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.

9.7 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 cambiando gli strumenti, o modificando le modalità delle operazioni da compiere, o correggendo opportunamente i risultati.*
- *Potendo, bisogna effettuare misure ripetute: perché in questo caso sappiamo stimare ragionevolmente l'errore commesso a partire dalle misure stesse (se non è possibile effettuare misure ripetute, si assumerà convenzionalmente come errore l'inverso della sensibilità dello strumento, ovvero sia la più piccola variazione della grandezza indicata sulla scala di lettura); e bisogna effettuarne quante più possibile per aumentare in corrispondenza la validità statistica dei nostri risultati.*
- *Se il numero di misure effettuate è basso⁴ si scartano quei dati che differiscano dal valore 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 $s_{\bar{x}} = s/\sqrt{N}$.*

⁴“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.

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

9.8 Il teorema del limite centrale

Fino ad ora abbiamo più volte sottolineato il fatto che un preciso significato (quello statistico) dell'errore quadratico medio può essere enunciato solo se la distribuzione delle misure effettuate è quella normale.

Con riguardo alla media aritmetica delle misure, se queste seguono la legge normale e se, inoltre, sono statisticamente indipendenti tra loro, il teorema di pagina 103 ci assicura che qualunque loro combinazione lineare (ed in particolare la media aritmetica) è ancora distribuita secondo la legge normale; ed all'errore della media $\sigma_{\bar{x}}$ si può quindi attribuire lo stesso significato statistico.

Vogliamo ora ampliare questo discorso dimostrando un importantissimo teorema della statistica e discutendone le implicazioni:

TEOREMA (DEL LIMITE CENTRALE): *siano N variabili casuali x_i , statisticamente indipendenti tra loro e provenienti da una distribuzione avente densità di probabilità ignota, della quale esistano finite sia la media μ che la varianza σ^2 ; sotto questa ipotesi, la distribuzione della media aritmetica del campione, \bar{x} , tende asintoticamente alla distribuzione normale con media μ e varianza σ^2/N al crescere di N .*

Dimostreremo questo teorema facendo l'ipotesi, più restrittiva, che esistano i momenti della funzione di frequenza delle x_i di qualunque ordine k (esso può essere dimostrato, come si vede dall'enunciato, anche se esistono solamente i primi due); e partiamo dal fatto che, sotto le ipotesi su dette, la somma S delle N variabili casuali

$$S = \sum_{i=1}^N x_i$$

ha valore medio e varianza date dalle

$$E(S) = N\mu \quad \text{e} \quad \sigma_S^2 = N\sigma^2 .$$

Inoltre, visto che i valori x_i sono tra loro statisticamente indipendenti, possiamo applicare l'equazione (6.11) per trovare la funzione caratteristica della S , che vale

$$\begin{aligned}\phi_S(t) &= \prod_{i=1}^N \phi_{x_i}(t) \\ &= [\phi_x(t)]^N\end{aligned}$$

visto che le x_i hanno tutte la stessa distribuzione (e quindi la stessa funzione caratteristica). Se consideriamo invece gli scarti $z_i = x_i - \mu$ delle x_i dalla media, dalla (6.17) possiamo ricavare la funzione caratteristica della z :

$$\phi_z(t) = e^{-i\mu t} \phi_x(t) \quad (9.7)$$

e, se esistono tutti i momenti fino a qualsiasi ordine della x (e in conseguenza anche della z), la (6.8) implica

$$\phi_z(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \lambda_k = 1 + 0 - \frac{1}{2} t^2 \sigma^2 + \mathcal{O}(t^3) \quad (9.8)$$

in cui i λ_k sono i momenti della funzione di frequenza della z , i primi due dei quali valgono 0 e σ^2 .

Introduciamo infine la nuova variabile

$$y = \frac{S - E(S)}{\sigma_S} = \frac{1}{\sigma\sqrt{N}} S - \frac{N\mu}{\sigma\sqrt{N}} = \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^N (x_i - \mu)$$

e indichiamo poi con $\phi_y(t)$ la funzione caratteristica della y ; essendo quest'ultima lineare in S abbiamo dalla (6.17) che

$$\begin{aligned}\phi_y(t) &= e^{-it \frac{N\mu}{\sigma\sqrt{N}}} \cdot \phi_S\left(\frac{t}{\sigma\sqrt{N}}\right) \\ &= e^{-it \frac{N\mu}{\sigma\sqrt{N}}} \left[\phi_x\left(\frac{t}{\sigma\sqrt{N}}\right) \right]^N \\ &= \left[e^{-i\mu \frac{t}{\sigma\sqrt{N}}} \cdot \phi_x\left(\frac{t}{\sigma\sqrt{N}}\right) \right]^N \\ &= \left[\phi_z\left(\frac{t}{\sigma\sqrt{N}}\right) \right]^N\end{aligned}$$

ricordando la (9.7). Da qui, introducendo l'espressione (9.8) prima ottenuta per lo sviluppo di $\phi_z(t)$,

$$\begin{aligned}\phi_y(t) &= \left[1 - \frac{1}{2} \frac{t^2}{N\sigma^2} \sigma^2 + \mathcal{O}\left(\frac{t^3}{N^{\frac{3}{2}}}\right) \right]^N \\ &= \left[1 - \frac{t^2}{2N} + \mathcal{O}\left(N^{-\frac{3}{2}}\right) \right]^N\end{aligned}$$

e quando N tende all'infinito

$$\lim_{N \rightarrow \infty} \phi_y(t) = e^{-\frac{t^2}{2}}$$

sfruttando il limite notevole

$$\lim_{x \rightarrow +\infty} \left(1 + \frac{k}{x} \right)^x = e^k \quad (9.9)$$

(qui, appunto, $k = -t^2/2$). Insomma la funzione caratteristica della y tende a quella di una distribuzione normale di media zero e varianza 1: quindi la S tende asintoticamente ad una distribuzione normale di media $N\mu$ e varianza $N\sigma^2$; e $\bar{x} = S/N$ tende asintoticamente ad una distribuzione normale di media μ e varianza σ^2/N .

Il teorema è di fondamentale importanza perché non fa alcuna ipotesi sulla distribuzione delle variabili che compongono il campione (all'infuori del requisito dell'esistenza di media e varianza). Con riguardo alle misure ripetute di una stessa grandezza fisica esso ci dice che, se anche la loro distribuzione non segue la legge di Gauss, *purché se ne abbia un numero sufficiente* il nostro risultato finale (la media aritmetica) tuttavia la segue ugualmente in modo approssimato: così che l'errore della media conserva il consueto significato statistico (di semiampiezza dell'intervallo, centrato su \bar{x} , che contiene il valore vero con probabilità costante prefissata del 68%) anche se questo non è verificato per le singole misure.

Da notare che il teorema del limite centrale implica una convergenza asintoticamente normale del valore medio del campione al valore medio della popolazione delle misure; per attribuire a quest'ultimo, come si è fatto nell'ultima frase, il significato di valore vero della grandezza misurata, si sottintende che le misure abbiano distribuzione, ancorché di forma non specificata, simmetrica rispetto al valore vero x^* ; insomma che errori per difetto e per eccesso siano ugualmente probabili.

Incidentalmente, notiamo qui come il *prodotto* di molte variabili casuali indipendenti debba avere un comportamento, indipendentemente dal tipo di distribuzione, asintoticamente tendente a quello di una distribuzione *log-normale*.

9.8.1 Applicazione: numeri casuali normali

Siano gli u_i (con $i = 1, \dots, N$) dei numeri provenienti da una popolazione u distribuita uniformemente nell'intervallo $[0, 1]$; abbiamo visto nel paragrafo 8.1 che $E(u) = \frac{1}{2}$ e $\text{Var}(u) = \frac{1}{12}$. La loro media aritmetica \bar{u} , in conseguenza del teorema del limite centrale, tende asintoticamente (al crescere di N) alla distribuzione normale con media $\frac{1}{2}$ e varianza $\frac{1}{12 \cdot N}$; quindi la loro somma $N\bar{u}$ è asintoticamente normale con media $\frac{N}{2}$ e varianza $\frac{N}{12}$; e, infine, la variabile casuale

$$x = \frac{\sum_{i=1}^N u_i - \frac{N}{2}}{\sqrt{\frac{N}{12}}} \quad (9.10)$$

è asintoticamente normale con media 0 e varianza 1.

Di questa proprietà si può far uso per ottenere da un computer dei numeri pseudo-casuali con distribuzione (approssimativamente) normale, a partire da altri numeri pseudo-casuali con distribuzione uniforme; in pratica l'approssimazione è già buona quando $N \gtrsim 10$, e scegliendo $N = 12$ possiamo, ad esempio, porre semplicemente

$$x = \sum_{i=1}^{12} u_i - 6 \quad .$$

È da notare, comunque, che **non** è buona pratica servirsi di questo metodo: anche se la parte centrale della distribuzione normale è approssimata abbastanza bene, le code mancano totalmente (essendo impossibile che risulti $|x| > \sqrt{3N}$); l'effetto di questa mancanza, quando (come nelle analisi fisiche basate su metodi di Montecarlo) vengano richiesti numeri pseudo casuali per generare eventi simulati in quantità dell'ordine di milioni almeno, è tale da invalidare completamente i risultati.

Soprattutto, poi, generare numeri pseudo-casuali normali usando il teorema del limite centrale non è solo sbagliato, ma inutile: esistono altri metodi (come ad esempio quello di Box-Muller che discuteremo ora) che sono in grado di generare numeri pseudo-casuali con una *vera* distribuzione normale usando, per il calcolo, un tempo non molto superiore a quello richiesto dalla (9.10).

Siano x ed y due variabili casuali *statisticamente indipendenti*, ed aventi distribuzione uniforme nell'intervallo $[0, 1]$; consideriamo le altre due variabili casuali u e v definite attraverso le

$$u = \sqrt{-2 \ln x} \cdot \cos(2\pi y) \quad \text{e} \quad v = \sqrt{-2 \ln x} \cdot \sin(2\pi y) \quad . \quad (9.11)$$

Queste funzioni si possono invertire, e risulta

$$x = e^{-\frac{1}{2}(u^2+v^2)} \quad \text{e} \quad y = \frac{1}{2\pi} \arctan\left(\frac{v}{u}\right)$$

con derivate parziali prime date da

$$\begin{cases} \frac{\partial x}{\partial u} = -u \cdot e^{-\frac{1}{2}(u^2+v^2)} \\ \frac{\partial x}{\partial v} = -v \cdot e^{-\frac{1}{2}(u^2+v^2)} \end{cases}$$

e da

$$\begin{cases} \frac{\partial y}{\partial u} = \frac{1}{2\pi} \frac{1}{1 + \frac{v^2}{u^2}} \left(-\frac{v}{u^2}\right) \\ \frac{\partial y}{\partial v} = \frac{1}{2\pi} \frac{1}{1 + \frac{v^2}{u^2}} \frac{1}{u} \end{cases}$$

Il determinante Jacobiano delle (x, y) rispetto alle (u, v) vale

$$\begin{aligned} \frac{\partial(x, y)}{\partial(u, v)} &= \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial y}{\partial u} \\ &= -\frac{1}{2\pi} e^{-\frac{1}{2}(u^2+v^2)} \end{aligned}$$

per cui, essendo la densità di probabilità congiunta delle due variabili casuali x ed y data dalla

$$f(x, y) = 1$$

e applicando la (7.9), la densità di probabilità congiunta della u e della v è data da

$$f(u, v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2}$$

e quindi, in conseguenza della (7.8), la u e la v sono due variabili casuali *st statisticamente indipendenti* tra loro ed entrambe aventi funzione di frequenza data dalla *distribuzione normale standardizzata*; questo è appunto il metodo cosiddetto “di Box-Muller” per la generazione di numeri pseudo-casuali con distribuzione normale, a partire da numeri pseudo-casuali con distribuzione uniforme.

Una variante che consente di sveltire questo metodo (lento, perché l'esecuzione delle funzioni logaritmo, seno e coseno consuma molto tempo

di CPU) consiste nel generare dapprima due numeri pseudo-casuali x' e y' distribuiti uniformemente tra i limiti -1 e $+1$; e nell'accettarli se $S = R^2 = x'^2 + y'^2 \leq 1$, in modo che il punto P le cui coordinate essi rappresentano nel piano $\{x', y'\}$ sia uniformemente distribuito entro il cerchio avente centro nell'origine O e raggio unitario — o nel rigettarli in caso contrario, ripetendo il passo precedente.

Questa prima condizione in realtà *rallenta* il procedimento, perché la coppia di numeri a caso viene accettata con probabilità $\frac{\pi}{4} \approx 78.5\%$; ma se, a questo punto, si usa al posto della x nella (9.11) il valore di S (che, come non è difficile dimostrare, è anch'esso distribuito uniformemente nell'intervallo $[0, 1]$); e se si prende poi in luogo dell'angolo $2\pi y$ l'angolo polare θ tra \overline{OP} e l'asse delle x' , il calcolo risulta in definitiva molto più rapido: perché il seno ed il coseno di θ si possono valutare come y'/R ed x'/R rispettivamente, eseguendo il calcolo di una radice quadrata e due divisioni soltanto.

Capitolo 10

Le misure indirette

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

Supponiamo per semplicità che queste altre 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 dai valori 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 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 ; infine, nell'ipotesi che le variabili siano anche tra loro statisticamente indipendenti, vedremo come si può valutare l'errore collegato a tale stima.

10.1 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:

$$F^* = F(x^*, y^*, z^*, \dots) \quad . \quad (10.1)$$

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:

$$\begin{aligned}
 F(x, y, z, \dots) &= F(x_0, y_0, z_0, \dots) + && \text{(ordine zero)} \\
 &+ \frac{\partial F}{\partial x} (x - x_0) + \frac{\partial F}{\partial y} (y - y_0) + \dots && \text{(primo ordine)} \\
 &+ \frac{\partial^2 F}{\partial x^2} \frac{(x - x_0)^2}{2!} + \dots && \text{(secondo ordine)} \\
 &+ \mathcal{O}(3)
 \end{aligned}$$

(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 è possibile trascurare i termini di ordine superiore al primo, possiamo in particolare ricavare:

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

Prendendo poi il valore medio di entrambi i membri e tenendo presente nei passaggi sia la (10.1), sia che risulta $E(\bar{x} - x^*) = E(\bar{x}) - x^* \equiv 0$ in assenza di errori sistematici (e similmente per le altre variabili), si ottiene

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

cioè

$$E(\bar{F}) \approx F^*$$

e, in definitiva:

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

(ossia \bar{F} è una stima *imparziale* di F^*).

Ricordiamo che questa conclusione è valida *solo approssimativamente*, perché 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*.

Un caso particolare è poi quello in cui la F è una funzione lineare in ognuna delle variabili da cui dipende; in questo caso, ovviamente, tutte le derivate di ordine successivo al primo sono identicamente nulle, e le conclusioni precedenti sono valide *esattamente*.

10.2 Combinazioni lineari di misure dirette

Supponiamo che le misure dirette delle variabili indipendenti da cui dipende la F siano esenti da errori sistematici, e che siano pertanto distribuite secondo la legge normale; consideriamo dapprima quelle particolari funzioni che sono le combinazioni lineari di più variabili:

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

Abbiamo già visto, nell'equazione (5.2) a pagina 51, che il valore medio di una tale funzione è la combinazione lineare, con gli stessi coefficienti, delle medie delle variabili; e, se supponiamo inoltre che le variabili siano tutte *statisticamente indipendenti* tra loro, sappiamo anche che la varianza di F è poi data (equazione (5.5) a pagina 54) dalla combinazione lineare delle loro varianze con coefficienti pari ai quadrati dei rispettivi coefficienti:

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

e

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

Abbiamo inoltre dimostrato, a pagina 103, un teorema secondo il quale una qualsiasi combinazione lineare a coefficienti costanti di variabili casuali

aventi distribuzione normale, ed inoltre tra loro statisticamente indipendenti, è anch'essa distribuita secondo la legge normale.

Ora, sapendo che la distribuzione della F è data dalla funzione di Gauss, 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 medio $E(F) = F^*$, che contiene un qualsiasi valore della F (ed in particolare la nostra miglior stima \bar{F}) con una probabilità del 68%; o, di converso, la semiampiezza di un intervallo centrato sulla nostra migliore stima \bar{F} e che contiene l'ignoto valore 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.

10.3 La formula di propagazione degli errori

Ora, qualsiasi funzione di più variabili si può considerare in prima approssimazione lineare; questo 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

$$\sigma_F^2 \approx \left(\frac{\partial F}{\partial x} \right)^2 \sigma_x^2 + \left(\frac{\partial F}{\partial y} \right)^2 \sigma_y^2 + \left(\frac{\partial F}{\partial z} \right)^2 \sigma_z^2 + \dots \quad (10.2)$$

(le derivate vanno calcolate per i valori $x = \bar{x}$, $y = \bar{y}$, $z = \bar{z}$,... delle variabili indipendenti).

Questa formula è nota sotto il nome di *formula di propagazione degli errori*: ripetiamo che si tratta di una formula *approssimata*; che è valida solo se non si commettono errori troppo grandi nelle misure dirette delle variabili; e che presuppone che le variabili stesse siano tra loro statisticamente

indipendenti¹. La formula di propagazione è invece *esatta* nel caso particolare (esaminato nel paragrafo precedente) di una combinazione lineare di variabili casuali indipendenti, caso questo nel quale tutte le derivate parziali di ordine superiore al primo sono identicamente nulle.

10.4 Errore dei prodotti di potenze

Applichiamo ora la formula (10.2) di propagazione degli errori a quella particolare classe di funzioni costituita dai prodotti di potenze delle variabili indipendenti: cioè alle funzioni del tipo

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

Calcoliamo innanzi tutto le derivate parziali di F ; risulta

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

(ammettendo che nessuna delle variabili sia nulla; questo implica che anche la F abbia valore diverso da zero). Introducendo questi valori delle derivate nella formula di propagazione degli errori, avremo

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

ed in definitiva

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

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; e possiamo ritenerla valida se le variabili indipendenti sono misurate con errori piccoli.

¹Una formula di propagazione degli errori per variabili qualsiasi (che ossia non ne presupponga l'indipendenza statistica) verrà ricavata più avanti, nel paragrafo C.3.

10.5 Errori massimi

Quando si parla di errori di misura senza specificare null'altro, si sottintende di norma che i numeri riportati si riferiscono ad errori quadratici medi; talvolta però si è in grado di indicare un intervallo all'interno del quale si sa con assoluta certezza che deve trovarsi il valore vero della grandezza misurata: in questi casi si può ovviamente attribuire un errore in termini *assoluti* (sia in difetto che in eccesso) al valore indicato. Supponendo per semplicità che i valori limite siano simmetrici rispetto al risultato trovato, che si potrà quindi ancora esprimere nella forma

$$x = x_0 \pm \Delta x ,$$

vogliamo ora determinare la legge secondo la quale si propagano questi *errori massimi* nelle misure indirette.

È immediato riconoscere che, se $F = Kx$ e $x \in [x_0 - \Delta x, x_0 + \Delta x]$, necessariamente F deve appartenere ad un intervallo di semiampiezza ΔF , con $\Delta F = |K| \Delta x$. Similmente, sommando (o sottraendo) due grandezze indipendenti di cui si conoscano gli errori massimi, il risultato $F = x \pm y$ dovrà essere compreso in un intervallo di semiampiezza $\Delta F = \Delta x + \Delta y$. Usando entrambe queste conclusioni, nel caso di una combinazione lineare

$$F = \sum_{i=1}^N a_i x_i$$

l'errore massimo su F vale

$$\Delta F = \sum_{i=1}^N |a_i| \Delta x_i .$$

Per una relazione funzionale qualsiasi

$$F = F(x_1, x_2, \dots, x_N) ,$$

e nei limiti in cui si possano trascurare i termini di ordine superiore al primo in uno sviluppo in serie di Taylor, la formula di propagazione per gli errori massimi è dunque

$$\Delta F \approx \sum_{i=1}^N \left| \frac{\partial F}{\partial x_i} \right| \Delta x_i ;$$

e, per i prodotti di potenze del tipo $F = K \cdot x^\alpha \cdot y^\beta \dots$,

$$\frac{\Delta F}{|F|} \approx |\alpha| \frac{\Delta x}{|x|} + |\beta| \frac{\Delta y}{|y|} + \dots .$$

Capitolo 11

Stime di parametri

In questo capitolo prenderemo in considerazione due speciali tecniche di elaborazione dei dati che sono utilizzate per stimare il valore di parametri ignoti dai quali le distribuzioni teoriche dipendono: la media pesata di determinazioni sperimentali aventi diversa precisione; e la valutazione 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).

Il metodo usato per la soluzione è, in entrambi i casi, quello della *massima verosimiglianza* (introdotto originariamente da Fisher¹ nel 1921); la prima parte del capitolo riguarderà appunto il problema della stima del valore dei parametri in generale, e questo metodo in particolare.

11.1 Stime e loro caratteristiche

Supponiamo che la densità di probabilità $f(x; \theta)$ di una variabile casuale continua x (che possa assumere tutti i valori dell'asse reale) dipenda da un parametro θ , il cui valore vero θ^* ci sia ignoto; se si hanno a disposizione N determinazioni sperimentali indipendenti x_i della grandezza x , vogliamo

¹Sir Ronald Fisher nacque a Londra nel 1890 e morì ad Adelaide (in Australia) nel 1962. È considerato, per l'importanza dei suoi lavori, uno dei fondatori della moderna statistica: oltre al concetto di verosimiglianza (*likelihood* in inglese), introdusse per primo l'analisi delle varianze e scoprì la forma analitica delle funzioni di distribuzione di molte importanti variabili casuali; dette poi importanti contributi ai metodi per i piccoli campioni ed a quelli per la verifica delle ipotesi.

trovare una funzione $\bar{\theta} = \bar{\theta}(x_1, x_2, \dots, x_N)$ che, a partire da esse, ci permetta di ricavare, nella maniera migliore possibile, un valore numerico da attribuire a θ^* : le funzioni $\bar{\theta}$ di questo tipo si chiamano appunto *stime*.

Una stima è dunque una funzione di variabili casuali, e, pertanto, una variabile casuale essa stessa; potremo in conseguenza parlare del valore medio o della varianza di una particolare stima, intendendo così riferirci alle caratteristiche della popolazione dei possibili valori restituiti dalla stima stessa in corrispondenza di tutti i possibili campioni che possono essere usati per calcolarla.

Nella statistica, alle stime si possono associare svariate caratteristiche; la prima di esse (e la più importante) è la **consistenza**. Una stima si dice *consistente* quando converge (probabilisticamente) al valore vero del parametro, ossia quando

$$\lim_{N \rightarrow \infty} \bar{\theta}(x_1, x_2, \dots, x_N) = \theta^* .$$

Ad esempio, il teorema di Čebyšef si può enunciare sinteticamente affermando che “il valore medio di un campione è una stima consistente del valore medio della popolazione”.

Una seconda caratteristica delle stime è la **distorsione**: una stima si dice *indistorta*, o *imparziale*, se mediamente coincide col valore vero del parametro; insomma se

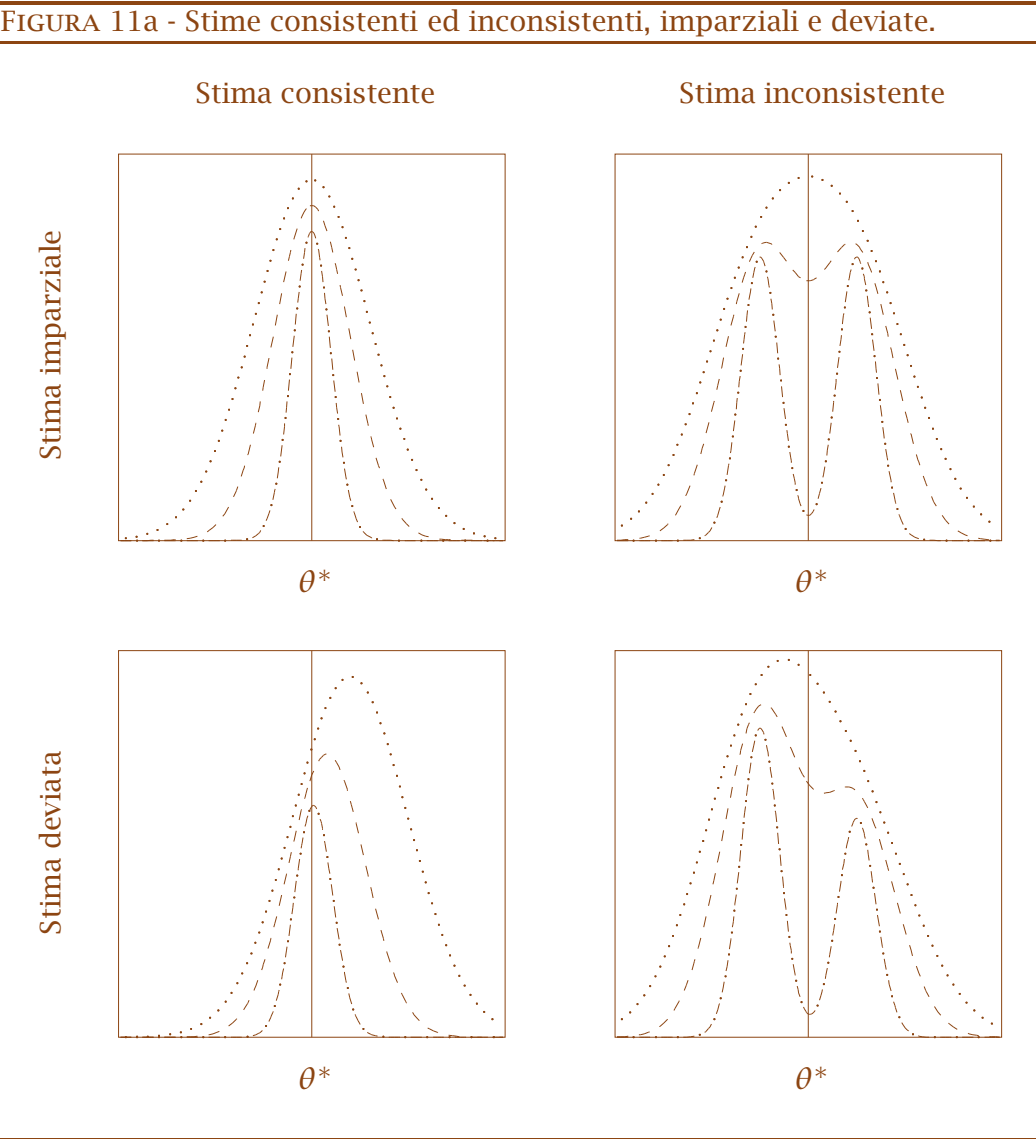
$$E(\bar{\theta}) = \theta^* .$$

Già sappiamo, dai paragrafi 5.3 e 5.8 rispettivamente, che la media dei campioni è una stima indistorta del valore medio della popolazione; mentre la varianza del campione è una stima distorta, ancorché consistente, della varianza della popolazione (a meno che non sia opportunamente corretta moltiplicandola per un fattore $N/(N-1)$).

Nella figura 11a sono riportati esempi di stime consistenti ed inconsistenti, distorte ed indistorte, per dare un'idea dell'andamento della densità di probabilità delle stime stesse all'aumentare delle dimensioni del campione.

Una terza caratteristica delle stime è l'**efficienza**: diremo che una prima stima è più *efficiente* di una seconda se la sua varianza è inferiore, e, quindi, se mediamente essa è più vicina al valore centrale $E(\bar{\theta})$; che coincide con θ^* se la stima è anche imparziale. Esiste un teorema (teorema di Cramér-Rao) del quale ci occuperemo sia più avanti nel corso di questo capitolo, sia in particolare nell'appendice E; questo teorema dimostra l'esistenza di un limite inferiore per la varianza delle stime, e quindi di un limite superiore per la loro efficienza.

Se abbiamo a disposizione, poi, M stime differenti θ_j dello stesso parametro θ , ogni campione di N valori x_i produrrà, attraverso l'applicazione



di ognuna di tali stime, M diversi valori per θ . Se ora indichiamo con f la densità di probabilità congiunta di questi M valori, risulterà in generale

$$f(\theta_1, \theta_2, \dots, \theta_M; \theta^*) = f^M(\theta_1; \theta^*) \cdot \varphi(\theta_2, \theta_3, \dots, \theta_M; \theta^* | \theta_1)$$

dove con $f^M(\theta_1; \theta^*)$ abbiamo, al solito, indicato la funzione densità di probabilità marginale della sola θ_1 (ovvero la densità di probabilità collegata al presentarsi di un certo valore per θ_1 indipendentemente da quello ottenuto per le altre stime); mentre $\varphi(\theta_2, \theta_3, \dots, \theta_M; \theta^* | \theta_1)$ è la densità di probabilità di queste ulteriori $M - 1$ stime condizionata dal valore della prima.

Nel caso che φ risulti *indipendente* da θ^* , la conseguenza che da questo fatto si deduce è che, una volta calcolata θ_1 , le altre stime sarebbero distribuite comunque nello stesso modo per *qualunque* valore di θ^* ; esse non potrebbero quindi aggiungere nulla alla conoscenza già ottenuta sul valore del parametro θ : ovverosia θ_1 *sfrutta tutta l'informazione* sul parametro ignoto che è contenuta nei dati, ed in questo caso la stima θ_1 si dice **sufficiente**. Non è detto che una stima sufficiente per un certo parametro θ esista; ma se ne esiste una, $\bar{\theta}$, allora ne esistono infinite: si può dimostrare infatti che ogni funzione monotona in senso stretto di $\bar{\theta}$ gode della stessa proprietà.

11.2 La stima di massima verosimiglianza

Dato un campione di N determinazioni *indipendenti* x_i , l'espressione

$$\prod_{i=1}^N f(x_i; \theta^*)$$

rappresenta la densità di probabilità da associare all'evento casuale consistente nell'ottenere una determinata N -pla di valori, essendo θ^* il valore del parametro da cui la f dipende.

Se in questa espressione si sostituisce al valore vero (che avevamo supposto noto) θ^* il generico valore θ ; e se le x_i non vengono considerate più variabili casuali, ma costanti che sono state determinate dalle nostre operazioni di misura, la funzione

$$\mathcal{L}(x_1, x_2, \dots, x_N; \theta) = \prod_{i=1}^N f(x_i; \theta) \quad (11.1)$$

(*funzione di verosimiglianza*) rappresenta la densità di probabilità da associare all'evento casuale consistente nell'essere un certo θ il valore vero del

nostro parametro, nell'ipotesi di avere già ottenuto la particolare N -pla di valori sperimentali x_1, x_2, \dots, x_N .

Il metodo della massima verosimiglianza consiste nell'adottare, come stima del parametro θ , quel valore $\hat{\theta}$ che rende massima la funzione di verosimiglianza (11.1); ovvero la soluzione delle

$$\frac{d\mathcal{L}}{d\theta} = 0 \qquad \frac{d^2\mathcal{L}}{d\theta^2} < 0 \qquad (11.2)$$

(nel caso che le (11.2) abbiano più di una soluzione, si sceglie quella che corrisponde al massimo assoluto).

Visto che il logaritmo naturale è (essendo la base, e , maggiore di uno) una funzione monotona strettamente crescente dell'argomento, trovare il massimo di $\ln \mathcal{L}$ condurrebbe ancora a tutti e soli i valori che rendono massima \mathcal{L} ; questo corrisponde al sostituire (essendo $\mathcal{L} > 0$), alla prima delle (11.2), l'equivalente

$$\frac{1}{\mathcal{L}} \frac{d\mathcal{L}}{d\theta} = \frac{d(\ln \mathcal{L})}{d\theta} = 0 .$$

Enunciamo qui, senza dimostrarle, alcune proprietà fondamentali della stima di massima verosimiglianza:

1. La stima di massima verosimiglianza è una stima *asintoticamente consistente* al crescere della dimensione del campione.
2. La stima di massima verosimiglianza ha una densità di probabilità *asintoticamente normale* al crescere della dimensione del campione.
3. La stima di massima verosimiglianza è asintoticamente, al crescere della dimensione del campione, anche *la stima più efficiente possibile* (ossia quella di minima varianza).
4. Se esiste una stima sufficiente di θ , essa può sempre essere espressa come funzione della sola stima di massima verosimiglianza $\hat{\theta}$.

Le ipotesi sotto le quali si riesce a dimostrare che la stima di massima verosimiglianza gode asintoticamente delle proprietà su dette sono estremamente generali: per la normalità basta che esistano i primi due momenti della $f(x; \theta)$; per la consistenza e la massima efficienza basta che $f(x; \theta)$ sia continua, dotata di derivata prima e seconda rispetto al parametro, e che l'operazione di integrazione rispetto a x commuti con quella di derivazione rispetto a θ (ovvero, in pratica, che il dominio di definizione della x non dipenda dal parametro).

Il teorema di Cramér-Rao (cui si è prima accennato) permette di dimostrare, sotto ipotesi del tutto generali, che esiste *un estremo inferiore* per le

varianze delle stime *imparziali* di una qualsiasi grandezza dipendente dal parametro θ ; non solo, ma che, se una stima di varianza minima esiste, essa *rende massima la funzione di verosimiglianza*.

Più in dettaglio: nell'ipotesi che la densità di probabilità $f(x; \theta)$ sia una funzione definita in una regione dell'asse x avente estremi indipendenti dal parametro θ ; che esista ovunque la derivata rispetto a θ di $\ln f(x; \theta)$; e, infine, che esista finito il valore medio del quadrato di questa derivata

$$E \left\{ \left[\frac{\partial}{\partial \theta} \ln f(x; \theta) \right]^2 \right\} = \frac{1}{N} \cdot E \left\{ \left[\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right]^2 \right\}$$

il teorema di Cramér-Rao afferma che una *qualsiasi* stima *imparziale* $\bar{\theta}$ di θ ha una varianza che non può essere inferiore ad un valore (*limite di Cramér-Rao*) dato dalla

$$\text{Var}(\bar{\theta}) \geq \frac{1}{N \cdot E \left\{ \left[\frac{\partial}{\partial \theta} \ln f(x; \theta) \right]^2 \right\}}. \quad (11.3)$$

Inoltre questo estremo inferiore viene raggiunto, e vale il segno di uguaglianza nella (11.3), *se e solo se* esiste una funzione $R(\theta)$ per la quale risulti

$$\frac{\partial (\ln \mathcal{L})}{\partial \theta} = \sum_{i=1}^N \frac{\partial}{\partial \theta} \ln f(x_i; \theta) = \frac{\bar{\theta}(x_1, x_2, \dots, x_N) - \theta}{R(\theta)} \quad (11.4)$$

e, in tal caso, la stima di minima varianza *rende anche massima la funzione di verosimiglianza*.

La condizione (11.4) è assai restrittiva, potendosi tra l'altro dimostrare che essa implica che la densità di probabilità $f(x; \theta)$ deve essere una funzione di tipo esponenziale: nel caso generale non è quindi affatto certo che una stima di varianza minima esista, essendo questo subordinato alla validità della (11.4).

In ogni caso la stima di massima verosimiglianza deve, come prima detto, tendere *asintoticamente* a questo comportamento al crescere di N ; però nulla si può dire sulla rapidità di tale convergenza. Così, per un numero di misure finito, non c'è alcuna garanzia che la funzione di verosimiglianza abbia un solo massimo; e, se essa ne ammette più d'uno, non esiste modo di sapere quale di essi corrisponde (asintoticamente) alla stima di minima varianza, né esiste modo di sapere quale di questi massimi rappresenti la stima corretta del valore vero.

Come abbiamo detto, la funzione di verosimiglianza (11.1) può essere interpretata come densità di probabilità del parametro una volta che si

sia ottenuto un certo insieme di valori misurati; sfruttando la seconda delle proprietà su elencate, la densità di probabilità di θ deve anche essere (asintoticamente) data da

$$\mathcal{L}(\theta) = \frac{1}{\sigma_\theta \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\theta - \hat{\theta}}{\sigma_\theta} \right)^2}$$

quindi, nell'intorno di $\hat{\theta}$, deve essere

$$\ln \mathcal{L} = -\ln(\sigma_\theta \sqrt{2\pi}) - \frac{1}{2} \left(\frac{\theta - \hat{\theta}}{\sigma_\theta} \right)^2$$

e, derivando due volte rispetto al parametro,

$$\frac{d^2(\ln \mathcal{L})}{d\theta^2} = -\frac{1}{\sigma_\theta^2}$$

ed infine si giunge alla

$$\text{Var}(\hat{\theta}) \equiv \sigma_\theta^2 = -\frac{1}{\left. \frac{d^2(\ln \mathcal{L})}{d\theta^2} \right|_{\theta=\hat{\theta}}} \quad (11.5)$$

frequentemente usata per il calcolo dell'errore della stima di massima verosimiglianza.

11.2.1 Un esempio di stima sufficiente

Supponiamo di avere un campione di N determinazioni indipendenti x_k di una variabile che segua la distribuzione di Poisson; le probabilità dei differenti valori sono date dalla (8.14), e dipendono da un unico parametro: il valore medio della distribuzione, α . La funzione di verosimiglianza è la

$$\begin{aligned} \text{Pr}(x_1, \dots, x_N; \alpha) &= \frac{\alpha^{x_1}}{x_1!} e^{-\alpha} \cdot \frac{\alpha^{x_2}}{x_2!} e^{-\alpha} \dots \frac{\alpha^{x_N}}{x_N!} e^{-\alpha} \\ &= \frac{\alpha^{\sum_k x_k} \cdot e^{-N\alpha}}{x_1! x_2! \dots x_N!} \cdot \frac{(N\bar{x})!}{(N\bar{x})!} \\ &= \frac{\alpha^{N\bar{x}}}{(N\bar{x})!} e^{-N\alpha} \cdot \frac{(N\bar{x})!}{x_1! x_2! \dots x_N!} \cdot \frac{N^{N\bar{x}}}{N^{N\bar{x}}} \\ &= \left\{ \frac{(N\alpha)^{N\bar{x}}}{(N\bar{x})!} e^{-N\alpha} \right\} \left\{ \frac{(N\bar{x})!}{x_1! x_2! \dots x_N!} \frac{1}{N^{N\bar{x}}} \right\} \end{aligned} \quad (11.6)$$

Nei passaggi, per due volte si è moltiplicato e diviso per una stessa quantità non nulla: prima per $(N\bar{x})!$ e poi per $N^{N\bar{x}}$.

La stima di massima verosimiglianza si trova annullando la derivata della (11.6); che, a meno di un fattore costante, è della forma

$$f(\alpha) = \alpha^{N\bar{x}} e^{-N\alpha}$$

per cui

$$\frac{df}{d\alpha} = N\bar{x}\alpha^{N\bar{x}-1}e^{-N\alpha} - N\alpha^{N\bar{x}}e^{-N\alpha} = N\alpha^{N\bar{x}-1}e^{-N\alpha}(\bar{x} - \alpha)$$

e quindi la stima cercata è

$$\hat{\alpha} = \bar{x}$$

Il primo termine dell'espressione finale (11.6) per la funzione di verosimiglianza è la probabilità $\Pr(S)$ che la variabile casuale

$$S = \sum_{k=1}^N x_k = N\bar{x}$$

abbia un determinato valore: $\Pr(S)$ infatti, come già sappiamo dal paragrafo 8.5, segue la distribuzione di Poisson con valore medio $N\alpha$. Notiamo anche che, avendo N un valore costante noto a priori, $\Pr(S)$ coincide con $\Pr(\bar{x})$: il secondo termine *deve* quindi essere la probabilità che i dati osservati valgano x_1, x_2, \dots, x_N *condizionata* dal fatto che la loro somma vale $N\bar{x}$; ma, non dipendendo questo termine da α , tale probabilità è la stessa qualunque sia il parametro.

Qualunque sia \bar{x} , una volta noto il suo valore le x_k sono distribuite allo stesso modo: \bar{x} riassume insomma tutta l'informazione contenuta nei dati, ed è quindi per definizione una stima *sufficiente* del parametro. In effetti, se la probabilità dei valori x_k una volta nota \bar{x} non dipende dal parametro, questo implica che *qualunque* funzione dei dati ha probabilità (condizionata) che gode della stessa proprietà. Citiamo senza dimostrarlo, in proposito, il seguente

TEOREMA: $\bar{\theta}$ è una stima sufficiente di θ se e solo se la funzione di verosimiglianza è fattorizzabile nella forma

$$\mathcal{L}(x_1, x_2, \dots, x_N; \theta) = f(\bar{\theta}, \theta) \cdot \phi(x_1, x_2, \dots, x_N)$$

11.3 Media pesata

Quando si abbiano a disposizione più determinazioni ripetute di una stessa grandezza fisica, sappiamo che da esse si può ricavare un valore unico da usare come risultato finale attraverso il calcolo della media aritmetica; questa (come già anticipato senza dimostrazione nel paragrafo 4.4) è la funzione dei dati con la distribuzione più stretta attorno al valore 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 indipendenti, 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.

La funzione di verosimiglianza è la

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

(cioè la densità di probabilità di cui sopra, nella quale il valore vero x^* è sostituito dal parametro variabile x); e ricordiamo che essa rappresenta la densità di probabilità associata all'evento casuale consistente nell'essere il numero x il valore 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 è quella che, rendendo massima \mathcal{L} , individua quel numero che, sulla base delle osservazioni disponibili, possiede la *massima probabilità* di coincidere con il valore vero: vedremo tra poco che la

soluzione è unica. Prendendo il logaritmo naturale di \mathcal{L} ,

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

e ricordando, come prima detto, che il logaritmo naturale è una funzione monotona strettamente crescente dell'argomento, si vede che il massimo di \mathcal{L} corrisponde al minimo di $-2 \ln \mathcal{L}$; 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:

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

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

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

$$\bar{x} = \frac{1}{K} \sum_{i=1}^N \frac{x_i}{\sigma_i^2} . \quad (11.7)$$

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 (10.2) sia in questo caso *esatta* e non approssimata (dando insomma risultati sempre validi, indipendentemente dall'entità degli errori commessi).

Applichiamo direttamente l'equazione (5.5) per la varianza delle combinazioni lineari di variabili tra loro indipendenti, invece della più complicata (10.2): \bar{x} è calcolata come combinazione lineare delle x_i con coefficienti $1/(K \sigma_i^2)$, e quindi avremo

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

cioè

$$\sigma_{\bar{x}}^2 = \frac{1}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} . \quad (11.8)$$

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'analoga interpretazione di semiampiezza dell'intervallo con centro in \bar{x} avente probabilità pari al 68% di contenere il valore vero x^* .

Per quanto concerne le proprietà della media pesata \bar{x} come stima del valore vero, la derivata del logaritmo della funzione di verosimiglianza rispetto al parametro incognito (che è x) vale

$$\frac{d(\ln \mathcal{L})}{dx} = \sum_{i=1}^N \frac{x_i}{\sigma_i^2} - x \sum_{i=1}^N \frac{1}{\sigma_i^2} = K(\bar{x} - x)$$

ed è soddisfatta la condizione (11.4) sotto la quale il teorema di Cramér-Rao (che esamineremo in dettaglio nell'appendice E) ci permette di affermare che la stima di massima verosimiglianza è *anche quella di varianza minima*: ovvero, tra tutte le possibili funzioni dei dati che si potrebbero definire per stimare il valore vero x^* dal campione, quella mediamente più vicina ad esso.

È 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 non considerare 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 (11.7) si riduce allora alla media aritmetica (4.1) (ed il suo errore (11.8) alla già nota espressione (5.6)).

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 valore vero sulla base del nostro campione di misure, e che rappresenta la stima di minima varianza.

11.4 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 dai 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 (o 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 si presuppone esistente 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.

11.4.1 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:

$$y = a + bx \text{ .}$$

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 $\{x, y\}$ punti con queste coordinate, essi si dovranno disporre approssimativamente lungo una linea retta.

Ora, 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 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 con la *distanza complessiva* tra i nostri punti sperimentali 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;

dimosteremo ora che per “distanza complessiva” si deve intendere *la somma dei quadrati delle lunghezze dei segmenti di retta parallela all’asse y compresi tra i punti misurati e la retta esaminata*.

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

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

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 valore vero assunto in corrispondenza dalla y è $a + bx_i$; visto che è $y = a + bx$ la legge fisica che lega le due variabili tra loro.

Questa funzione di verosimiglianza rappresenta allora 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.

I valori più verosimili del parametro saranno quelli che rendono massima \mathcal{L} : vedremo ora che la soluzione è unica; e, ancora, il teorema di Cramér-Rao ci permetterebbe di dimostrare che la stima, appunto, più verosimile (la retta che corrisponde al massimo della probabilità) è anche la stima di minima varianza (ovvero la più precisa possibile). Prendendo il logaritmo naturale di entrambi i membri, risulta

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

I valori più verosimili dei parametri a e b sono quelli per cui è massima \mathcal{L} , ovvero è minima $-2 \ln \mathcal{L}$: 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 però *parallelamente all'asse y* , dall'insieme dei punti misurati; queste "distanze" sono anche comunemente chiamate "residui". 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

$$\begin{cases} a \cdot N + b \cdot \sum_i x_i = \sum_i y_i \\ a \cdot \sum_i x_i + b \cdot \sum_i x_i^2 = \sum_i x_i y_i \end{cases} \quad (11.9)$$

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

$$\begin{cases} a = \frac{1}{\Delta} \left[(\sum_i x_i^2) \cdot (\sum_i y_i) - (\sum_i x_i) \cdot (\sum_i x_i y_i) \right] \\ b = \frac{1}{\Delta} \left[N \cdot (\sum_i x_i y_i) - (\sum_i x_i) \cdot (\sum_i y_i) \right] \end{cases} \quad (11.10)$$

in cui si è posto per brevità

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

(le formule (11.10) 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 (10.2), che è in questo caso esatta; oppure la più semplice (5.5). Possiamo esprimere a e b in funzione delle y_i come

$$a = \sum_{i=1}^N a_i y_i \quad \text{e} \quad b = \sum_{i=1}^N b_i y_i$$

una volta posto

$$\begin{cases} a_i = \frac{1}{\Delta} \left(\sum_j x_j^2 - x_i \sum_j x_j \right) \\ b_i = \frac{1}{\Delta} \left(N x_i - \sum_j x_j \right) \end{cases}$$

e, se indichiamo con σ_y^2 la varianza comune a tutte le y_i , si ottiene, per l'errore di a :

$$\begin{aligned}
\sigma_a^2 &= \sum_i a_i^2 \sigma_y^2 \\
&= \sigma_y^2 \sum_i \left[\frac{1}{\Delta} \left(\sum_j x_j^2 - x_i \sum_j x_j \right) \right]^2 \\
&= \frac{\sigma_y^2}{\Delta^2} \sum_i \left[\left(\sum_j x_j^2 \right)^2 + x_i^2 \left(\sum_j x_j \right)^2 - 2 x_i \left(\sum_j x_j^2 \right) \left(\sum_j x_j \right) \right] \\
&= \frac{\sigma_y^2}{\Delta^2} \left[N \left(\sum_j x_j^2 \right)^2 + \left(\sum_i x_i^2 \right) \left(\sum_j x_j \right)^2 - 2 \left(\sum_j x_j^2 \right) \left(\sum_j x_j \right)^2 \right] \\
&= \frac{\sigma_y^2}{\Delta^2} \left[N \left(\sum_j x_j^2 \right)^2 - \left(\sum_j x_j \right)^2 \left(\sum_j x_j^2 \right) \right] \\
&= \frac{\sigma_y^2}{\Delta^2} \left(\sum_j x_j^2 \right) \left[N \left(\sum_j x_j^2 \right) - \left(\sum_j x_j \right)^2 \right] \\
&= \sigma_y^2 \frac{\sum_j x_j^2}{\Delta}
\end{aligned}$$

e, similmente, per b :

$$\begin{aligned}
\sigma_b^2 &= \sum_i b_i^2 \sigma_y^2 \\
&= \sigma_y^2 \sum_i \left[\frac{1}{\Delta} \left(N x_i - \sum_j x_j \right) \right]^2 \\
&= \frac{\sigma_y^2}{\Delta^2} \sum_i \left[N^2 x_i^2 + \left(\sum_j x_j \right)^2 - 2 N x_i \sum_j x_j \right] \\
&= \frac{\sigma_y^2}{\Delta^2} \left[N^2 \left(\sum_i x_i^2 \right) + N \left(\sum_j x_j \right)^2 - 2 N \left(\sum_j x_j \right)^2 \right] \\
&= \frac{N \sigma_y^2}{\Delta^2} \left[N \left(\sum_i x_i^2 \right) - \left(\sum_j x_j \right)^2 \right] \\
&= \sigma_y^2 \frac{N}{\Delta} .
\end{aligned}$$

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

$$\begin{cases} \sigma_a = \sigma_y \sqrt{\frac{\sum_i x_i^2}{\Delta}} \\ \sigma_b = \sigma_y \sqrt{\frac{N}{\Delta}} \end{cases} \quad (11.11)$$

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 statistico.

11.4.2 Stima a posteriori degli errori di misura

È da osservare come nelle formule (11.10) dei minimi quadrati non compaia il valore di σ_y : la soluzione del problema dell'interpolazione lineare è indipendente dall'entità degli errori di misura, nel senso che i coefficienti della retta interpolante possono essere calcolati anche se gli errori sulle y non sono noti (purché 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 valore 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 valore 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 ed alla retta vera che corrisponde alla legge fisica reale che collega le due variabili.

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'analoga somma per $N - 2$; in definitiva, che la corretta stima di σ_y è data dalla formula

$$\sigma_y^2 = \frac{\sum_{i=1}^N [(a + bx_i) - y_i]^2}{N - 2} .$$

In essa a numeratore compare la somma dei quadrati dei residui, cioè delle “distanze” dei punti misurati (x_i, y_i) dalla retta interpolante di equazione $a + bx$ calcolate secondo la direzione parallela all'asse delle ordinate. Questa formula² 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.

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 valore 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 $N - M$.

11.4.3 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

²Una formula equivalente (ma più semplice) per il calcolo di σ_y si può trovare nell'equazione (C.8) alla pagina 264.

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 uguali, 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 massima la funzione di verosimiglianza

$$\mathcal{L}(x_1, y_1, x_2, y_2, \dots, x_N, y_N; m) = \prod_{i=1}^N \frac{1}{\sigma_y \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{mx_i - y_i}{\sigma_y} \right)^2}.$$

Passando al logaritmo naturale di \mathcal{L} , è facile vedere che la soluzione ricercata è sempre quella che rende minima la somma dei quadrati dei residui dai punti misurati: che cioè minimizza la funzione

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

La derivata prima di f vale

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

e, 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 è esatta anche in questo caso, perché m è una combinazione lineare delle variabili affette da errore (le y_i); il coefficiente di y_i nella combinazione vale

$$\frac{x_i}{\sum_k x_k^2}$$

e quindi

$$\sigma_m^2 = \sum_{i=1}^N \left(\frac{x_i}{\sum_k x_k^2} \right)^2 \sigma_y^2 = \frac{\sigma_y^2}{(\sum_k x_k^2)^2} \sum_{i=1}^N x_i^2 = \frac{\sigma_y^2}{\sum_k x_k^2}$$

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 .

11.4.4 Interpolazione lineare nel caso generale

Le condizioni 1) e 2) sugli errori delle grandezze misurate x e y date nel paragrafo 11.4.1 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 fattore 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 e preventivamente, 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 differenti; 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 valori misurati (x_i, y_i) non già 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 $v = v(t) = v_0 + gt$ che descrive la caduta di un grave, e pensiamo di misurare la sua velocità in un certo istante: nell'ipotesi originale t sarebbe determinabile esattamente, ma l'imprecisione nella misura delle velocità ci darebbe valori di v compresi in un intervallo di ampiezza non nulla (dipendente dall'errore quadratico medio σ_v).

Se delle due grandezze, al contrario, fosse la velocità ad essere conoscibile esattamente, l'impossibilità di determinare con precisione l'istante t in cui essa deve essere misurata ci darebbe ugualmente valori di v distribuiti in un intervallo di ampiezza non nulla (legata stavolta a $g \cdot \sigma_t$).

Indicando, insomma, con σ_x e σ_y gli errori (sempre supposti costanti) di ognuna delle determinazioni (sempre supposte indipendenti) x_i e y_i , la formula dell'errore a posteriori ci permette di ricavare dai dati una ragionevole stima non tanto del solo σ_y quanto, piuttosto, della combinazione (quadratica) dell'errore intrinseco delle ordinate e di quello intrinseco delle ascisse propagato sulle ordinate:

$$\sigma^2 \approx \sigma_y^2 + (b^* \sigma_x)^2$$

(ove b^* è il valore vero della pendenza della retta).

11.4.5 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 naturale ne avrà una di tipo lineare:

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

11.5 Altre applicazioni della stima di massima verosimiglianza

Per concludere il capitolo, presentiamo altre tre applicazioni del metodo della massima verosimiglianza: la stima delle probabilità ignote di un insieme di modalità esclusive ed esaurienti cui può dar luogo un fenomeno casuale; la stima sia della media che della varianza di una popolazione normale; e la stima del range di una popolazione uniforme.

11.5.1 Stima di probabilità

Supponiamo che un fenomeno casuale possa dare origine ad un numero finito M di eventualità, ognuna delle quali sia associata ad un valore p_i ignoto della probabilità; se, eseguite N prove indipendenti, indichiamo con n_i la frequenza assoluta con cui ognuna delle M eventualità si è presentata nel corso di esse, quale è la stima di massima verosimiglianza, \hat{p}_i , per le incognite probabilità p_i ?

La funzione di verosimiglianza è, visto che la generica delle M eventualità, di probabilità p_i , si è presentata n_i volte, data³ da

$$\mathcal{L}(\mathbf{n}; \mathbf{p}) = \prod_{i=1}^M p_i^{n_i}$$

³A meno di un fattore moltiplicativo costante, corrispondente al numero di modi in cui N oggetti si possono ripartire tra M gruppi in modo che ogni gruppo sia composto da n_i oggetti; numero delle *partizioni ordinate* (vedi in proposito il paragrafo A.7).

(in cui abbiamo indicato sinteticamente con due vettori, \mathbf{n} e \mathbf{p} , entrambi di dimensione M , l'insieme degli M valori n_i e quello degli M valori p_i rispettivamente); ed il suo logaritmo da

$$\ln \mathcal{L}(\mathbf{n}; \mathbf{p}) = \sum_{i=1}^M n_i \ln p_i . \quad (11.12)$$

Il problema della ricerca del massimo della (11.12) è complicato dal fatto che i valori delle p_i non sono liberi, ma *vincolati* dalla condizione

$$\sum_{i=1}^M p_i = 1 . \quad (11.13)$$

Usiamo quindi il metodo dei moltiplicatori di Lagrange, costruendo la funzione

$$f(\mathbf{n}; \mathbf{p}) = \sum_{i=1}^M n_i \ln p_i - \lambda \left(\sum_{i=1}^M p_i - 1 \right) \quad (11.14)$$

e risolvendo il sistema delle $M + 1$ equazioni, nelle $M + 1$ incognite p_i e λ , composto dalla (11.13) e dalle altre M ottenute derivando la (11.14) rispetto ad ognuna delle p_k :

$$\frac{\partial f}{\partial p_k} = n_k \frac{1}{p_k} - \lambda = 0 \quad (k = 1, 2, \dots, M) .$$

Da quest'ultima si ricava

$$\hat{p}_k = \frac{n_k}{\lambda}$$

e, sostituendo nella (11.13),

$$\sum_{i=1}^M \hat{p}_i = \frac{1}{\lambda} \sum_{i=1}^M n_i = \frac{N}{\lambda} = 1$$

si ottiene

$$\lambda = N$$

per cui in definitiva la soluzione di massima verosimiglianza è (cosa non sorprendente) data dalle

$$\hat{p}_i = \frac{n_i}{N} .$$

11.5.2 Media e varianza di una popolazione normale

Abbiamo già visto nel paragrafo 11.3 che, *ammessa nota la varianza* σ^2 di una popolazione normale, il suo valore medio μ ha come stima di massima verosimiglianza la media aritmetica \bar{x} di un campione di stime indipendenti; vogliamo ora stimare *contemporaneamente sia μ che σ* dai dati, usando sempre il metodo della massima verosimiglianza.

La densità di probabilità vale

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

ed il suo logaritmo

$$\ln f(x; \mu, \sigma) = -\ln \sigma - \ln \sqrt{2\pi} - \frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 .$$

Il logaritmo della funzione di verosimiglianza è

$$\ln \mathcal{L}(\mathbf{x}; \mu, \sigma) = \sum_{i=1}^N \ln f(x_i; \mu, \sigma)$$

e dunque

$$\ln \mathcal{L}(\mathbf{x}; \mu, \sigma) = -N \ln \sigma - N \ln \sqrt{2\pi} - \frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 ;$$

e le sue derivate parziali prime sono

$$\frac{\partial}{\partial \mu} \ln \mathcal{L} = \frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \mu) = \frac{1}{\sigma^2} \left(\sum_{i=1}^N x_i - N\mu \right)$$

e

$$\frac{\partial}{\partial \sigma} \ln \mathcal{L} = -\frac{N}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^N (x_i - \mu)^2 = \frac{1}{\sigma^3} \left[\sum_{i=1}^N (x_i - \mu)^2 - N\sigma^2 \right] .$$

Il sistema ottenuto annullando le due derivate parziali prime ha l'unica soluzione (in effetti un massimo) data da

$$\hat{\mu} = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{e} \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2 .$$

Questo era già noto: entrambe le stime, come sappiamo, sono consistenti; però la seconda non è imparziale (ma può essere resa tale moltiplicandola per un opportuno fattore di correzione).

In sostanza il fatto che la varianza della popolazione abbia un determinato valore (come assunto nel paragrafo 11.3) non cambia il fatto che la nostra migliore stima del valore medio della popolazione sia comunque data dalla media aritmetica del campione: vedremo poi nel paragrafo 12.1 che il valore medio del campione e la sua varianza sono variabili casuali *statisticamente indipendenti tra loro*.

11.5.3 Range di una popolazione uniforme

Sia una variabile casuale x distribuita uniformemente tra un estremo inferiore noto, che senza perdere in generalità possiamo supporre sia lo zero, ed un estremo superiore ignoto A ; in questo caso dobbiamo innanzi tutto osservare sia che il dominio di definizione della funzione di frequenza $f(x)$ della x *dipende dal parametro* che dobbiamo stimare, sia che $f(x)$ e la sua derivata prima hanno dei punti di discontinuità: e non possiamo in conseguenza garantire *a priori* né la consistenza, né la massima efficienza asintotica del metodo usato.

Comunque, introducendo la cosiddetta *funzione gradino* (o *step function*) $S(x)$, definita attraverso la

$$\begin{cases} S(x) = 0 & (x < 0) \\ S(x) = 1 & (x \geq 0) \end{cases}$$

la funzione di frequenza $f(x)$ si può anche scrivere

$$f(x) = \frac{1}{A} S(x) S(A - x)$$

e la funzione di verosimiglianza

$$\mathcal{L}(x_1, x_2, \dots, x_N; A) = \frac{1}{A^N} S(x_{\min}) S(A - x_{\max}) .$$

Come sappiamo, ammesso noto il valore del parametro A essa rappresenta la densità di probabilità di ottenere gli N valori $x_i \in [-\infty, +\infty]$; se invece si considera A come l'unica variabile e si ammettono noti gli N valori x_i , rappresenta la densità di probabilità che un dato A abbia prodotto i dati osservati. Ma in quest'ultimo caso $S(x_{\min}) \equiv 1$, e la funzione di verosimiglianza si riduce alla

$$\mathcal{L}(A) = \frac{1}{A^N} S(A - x_{\max}) \quad (11.15)$$

che è nulla per $A < x_{\max}$ e monotona strettamente decrescente per $A \geq x_{\max}$; ed ammette quindi un unico massimo all'estremo del dominio, che vale

$$\hat{A} = x_{\max} . \quad (11.16)$$

Valore medio e varianza della stima valgono, come già sappiamo dal paragrafo 8.1.3,

$$E(\hat{A}) = A^* - \frac{A^*}{N+1} = \frac{N}{N+1} A^*$$

e

$$\text{Var}(\hat{A}) = \frac{N}{(N+1)^2 (N+2)} (A^*)^2$$

e quindi la stima è consistente, ma non imparziale; una stima imparziale è invece

$$\bar{A} = \frac{N+1}{N} x_{\max} = x_{\max} \left(1 + \frac{1}{N} \right) ,$$

di varianza ovviamente superiore per un fattore $(1 + 1/N)^2$. È anche ovvio, dalla forma sia della (11.15) che della (11.16), che \hat{A} è una stima sufficiente di A^* .

11.5.4 Stima della vita media di una particella

Nel processo di decadimento di una particella instabile, indichiamo con τ l'incognita vita media e con t i tempi (propri) di decadimento osservati; tempi che (come sappiamo) seguono la distribuzione esponenziale:

$$f(t; \tau) = \frac{1}{\tau} e^{-\frac{t}{\tau}} \quad \Rightarrow \quad \begin{cases} E(t) = \tau \\ \text{Var}(t) = \tau^2 \end{cases}$$

Ammettendo per semplicità che l'osservazione avvenga con una efficienza unitaria, o, in altre parole, che tutti i decadimenti vengano osservati, la funzione di verosimiglianza si scrive

$$\mathcal{L} = \prod_{k=1}^N f(t_k; \tau) = \frac{1}{\tau^N} e^{-\frac{1}{\tau} \sum_k t_k} ,$$

ed il suo logaritmo vale

$$\ln(\mathcal{L}) = -N \ln \tau - \frac{1}{\tau} \sum_{k=1}^N t_k = -N \left(\frac{\bar{t}}{\tau} + \ln \tau \right) .$$

Derivando rispetto al parametro e cercando gli estremanti,

$$\frac{d \ln(\mathcal{L})}{d\tau} = \frac{N}{\tau^2} (\bar{t} - \tau) = 0 ;$$

e quindi l'unico estremante della funzione di verosimiglianza si ha per

$$\hat{\tau} = \bar{t} .$$

Se calcoliamo la derivata seconda,

$$\frac{d^2 \ln(\mathcal{L})}{d\tau^2} = -\frac{N}{\tau^3} (2\bar{t} - \tau)$$

essa, calcolata per $t = \bar{t}$ è negativa; quindi l'unico estremante è effettivamente un punto di massimo.

La soluzione di massima verosimiglianza $\hat{\tau} = \bar{t}$ è *consistente ed imparziale* (essendo il valore medio del campione); *di varianza minima* (per il teorema di Cramér–Rao); inoltre la stima è *sufficiente* (riassume insomma tutta l'informazione del campione).

Normalmente l'efficienza non è però unitaria; ad esempio il nostro rivelatore può avere dimensioni confrontabili col cammino medio delle particelle, che possono quindi uscirne prima di decadere. In questo caso, visto che i decadimenti possono essere stati osservati solo essendo avvenuti all'interno di un intervallo compreso tra un valore temporale minimo (eventualmente nullo) ed uno massimo (ad esempio dipendente dalla posizione del decadimento, dalla direzione di emissione dei suoi prodotti, dalle dimensioni del rivelatore, ...) — intervallo differente per ognuno dei decadimenti — dovremo costruire la funzione di verosimiglianza considerando le probabilità di osservazione *condizionate* dall'essere il decadimento i -esimo avvenuto tra un certo $(t_{\min})_i$ ed un certo $(t_{\max})_i$:

$$\mathcal{L} = \prod_{i=1}^N \left[\frac{\frac{1}{\tau} e^{-\frac{t_i}{\tau}}}{e^{-\frac{(t_{\min})_i}{\tau}} - e^{-\frac{(t_{\max})_i}{\tau}}} \right] \quad (11.17)$$

Il denominatore della (11.17) rappresenta infatti la probabilità di decadere tra il tempo t_{\min} e quello t_{\max} , come è immediato ricavare dalla funzione di distribuzione della densità di probabilità esponenziale, che vale

$$F(t) = \int_0^t \frac{1}{\tau} e^{-\frac{x}{\tau}} dx = 1 - e^{-\frac{t}{\tau}} ;$$

dalla (11.17) si ricava poi

$$\ln(\mathcal{L}) = -N \ln \tau + \sum_{i=1}^N \left\{ -\frac{t_i}{\tau} - \ln \left[e^{-\frac{(t_{\min})_i}{\tau}} - e^{-\frac{(t_{\max})_i}{\tau}} \right] \right\}$$

e, posto per brevità

$$\varphi_i(\tau) = \frac{(t_{\min})_i^2 \cdot e^{-\frac{(t_{\min})_i}{\tau}} - (t_{\max})_i^2 \cdot e^{-\frac{(t_{\max})_i}{\tau}}}{e^{-\frac{(t_{\min})_i}{\tau}} - e^{-\frac{(t_{\max})_i}{\tau}}}$$

si arriva alla

$$\frac{d \ln(\mathcal{L})}{d\tau} = \frac{1}{\tau^2} \left\{ \sum_{i=1}^N [t_i - \varphi_i(\tau)] - N\tau \right\} = 0$$

che bisogna risolvere in modo numerico. Non si può inoltre in questo caso garantire che le proprietà precedentemente delineate per $\hat{\tau}$ (consistenza, normalità, efficienza, ...) siano ancora valide, almeno per N finito. Può darsi che la funzione di verosimiglianza ammetta più di un massimo, e non si sa a priori quale di essi convergerà verso τ^* ; e, per finire, l'errore della stima deve essere ricavato dalla concavità della funzione di verosimiglianza, supposta approssimativamente normale.

Capitolo 12

La verifica delle ipotesi (I)

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 valore vero sono sempre possibili (anche se con probabilità sempre più piccole al crescere dello scarto), una qualunque ipotesi sulla grandezza fisica misurata potrà essere confermata o rigettata dai dati solo ad un certo livello di probabilità.

Qui ci occuperemo inoltre di alcune funzioni di frequenza collegate a quella di Gauss, ossia della distribuzione del χ^2 , di quella di Student¹ e di quella di Fisher; e dell'uso che di esse si può fare per la verifica di ipotesi statistiche: quali ad esempio quella che un campione di dati sperimentali provenga da una popolazione descritta da una densità di probabilità nota a priori; o quella che il valore vero della grandezza misurata coincida con un valore determinato, noto anch'esso a priori.

12.1 La distribuzione del χ^2

Se le N variabili casuali x_i , tra loro statisticamente indipendenti, sono variabili normali standardizzate (ovverosia distribuite secondo la legge normale con media 0 e varianza 1), si può dimostrare che la nuova variabile

¹“Student” è lo pseudonimo con cui vennero pubblicati i lavori statistici di William Gosset, scienziato inglese vissuto dal 1876 al 1937. Uno dei pionieri di questo ramo della matematica, svolse le sue ricerche essendo dipendente (prima come chimico, poi come dirigente) della Guinness Brewery di Dublino.

casuale

$$X = \sum_{i=1}^N x_i^2$$

(ovviamente non negativa) è distribuita con una densità di probabilità data dalla

$$\frac{dp}{dX} = f(X; N) = K_N X^{\left(\frac{N}{2}-1\right)} e^{-\frac{X}{2}} \quad (12.1)$$

(*distribuzione del chi quadro*); la costante K_N viene fissata dalla condizione di normalizzazione, ed il parametro N prende il nome di *numero di gradi di libertà* della distribuzione.

La funzione caratteristica della X si può trovare facilmente considerando che, se la x è una variabile normale standardizzata, il suo quadrato $y = x^2$ ha una funzione caratteristica

$$\begin{aligned} \phi_y(t) &= E(e^{ity}) \\ &= E(e^{itx^2}) \\ &= \int_{-\infty}^{+\infty} e^{itx^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}(1-2it)} dx \\ &= \frac{1}{\sqrt{1-2it}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \\ &= (1-2it)^{-\frac{1}{2}} \end{aligned}$$

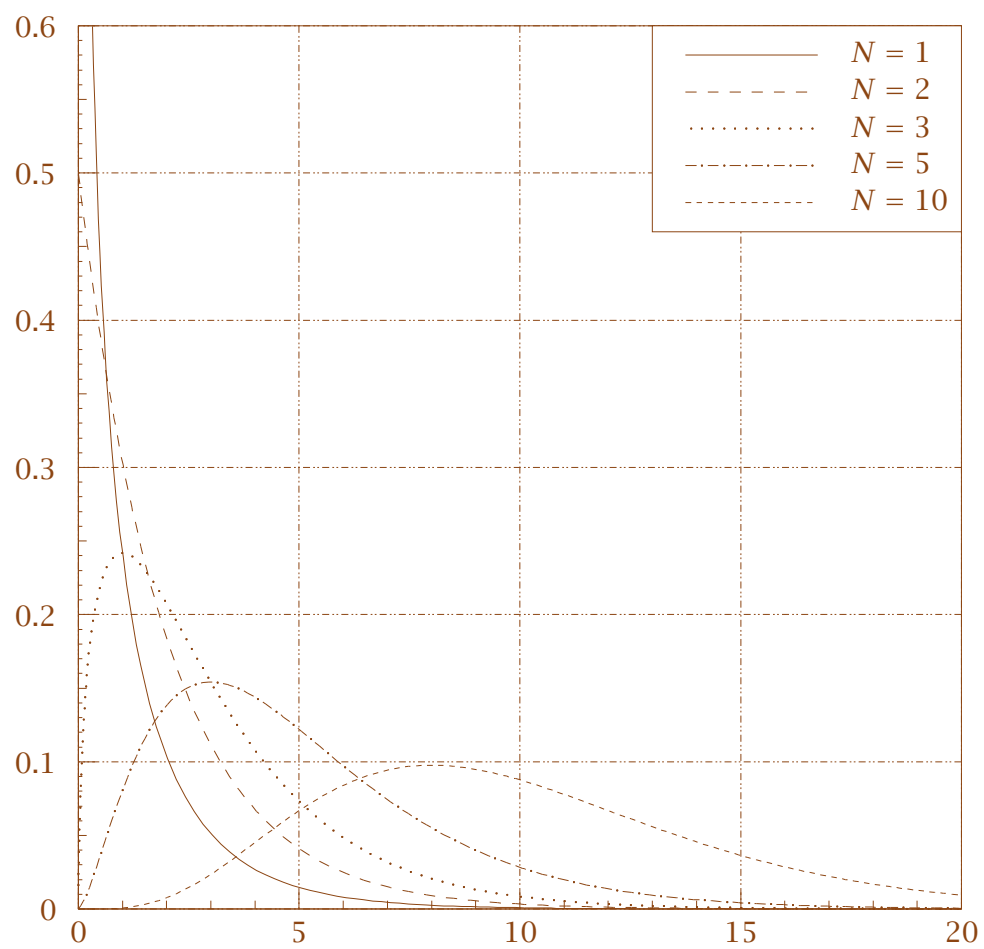
(si è eseguita la sostituzione di variabile $u = x\sqrt{1-2it}$; l'integrale definito è quello di una distribuzione normale $N(u; 0, 1)$ e vale dunque 1). Di conseguenza, applicando l'equazione (6.11), la funzione caratteristica della X vale

$$\phi_X(t) = (1-2it)^{-\frac{N}{2}}. \quad (12.2)$$

Per dimostrare che la funzione di frequenza della X è effettivamente la (12.1), si parte poi dall'espressione (12.2) della funzione caratteristica e le si applica la trasformazione inversa di Fourier già definita nella (6.10).

Con simili passaggi si potrebbe ricavare la funzione generatrice dei momenti, che vale

$$M_X(t) = (1-2t)^{-\frac{N}{2}}$$

FIGURA 12a - La distribuzione del χ^2 per alcuni valori del parametro N .

e, da queste, si ottiene infine che il valore medio e la varianza di una variabile casuale distribuita come il χ^2 a N gradi di libertà sono

$$E(X) = N \quad \text{e} \quad \text{Var}(X) = 2N$$

mentre i coefficienti di asimmetria e di curtosi valgono

$$y_1 = 2\sqrt{\frac{2}{N}} \quad \text{e} \quad y_2 = \frac{12}{N}.$$

La distribuzione del χ^2 tende asintoticamente ad una distribuzione normale con la stessa media N e la stessa varianza $2N$; infatti la funzione caratteristica della variabile standardizzata

$$y = \frac{X - N}{\sqrt{2N}} = \frac{X}{\sqrt{2N}} - \frac{N}{\sqrt{2N}}$$

vale, ricordando la (6.17),

$$\phi_y(t) = e^{-\frac{iNt}{\sqrt{2N}}} \left[1 - \frac{2it}{\sqrt{2N}} \right]^{-\frac{N}{2}}.$$

Passando ai logaritmi naturali,

$$\ln \phi_y(t) = -\frac{iNt}{\sqrt{2N}} - \frac{N}{2} \ln \left(1 - \frac{2it}{\sqrt{2N}} \right)$$

e, sviluppando in serie di McLaurin il logaritmo,

$$\begin{aligned} \ln \phi_y(t) &= -\frac{iNt}{\sqrt{2N}} - \frac{N}{2} \left[-\frac{2it}{\sqrt{2N}} - \frac{1}{2} \left(\frac{2it}{\sqrt{2N}} \right)^2 + \mathcal{O}(N^{-\frac{3}{2}}) \right] \\ &= -\frac{t^2}{2} + \mathcal{O}(N^{-\frac{1}{2}}) \end{aligned}$$

da cui

$$\lim_{N \rightarrow \infty} \phi_y(t) = e^{-\frac{t^2}{2}}$$

che è appunto la funzione caratteristica di una distribuzione normale standardizzata.

In definitiva:

- 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.

- Inoltre si potrebbe analogamente dimostrare che la variabile casuale $\sqrt{2X}$, anche per valori relativamente piccoli di N , ha una distribuzione che è assai bene approssimata da una funzione normale con media $\sqrt{2N-1}$ e varianza 1; l'approssimazione è già buona per $N \gtrsim 8$.

Dalla definizione (o dalla funzione caratteristica (12.2)) discende immediatamente la cosiddetta *regola di somma del χ^2* : ossia che, se X ed Y sono due variabili casuali statisticamente indipendenti entrambe 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.

Ovviamente, se le x_i (con $i = 1, \dots, N$) sono N variabili casuali statisticamente indipendenti tra loro e provenienti da una stessa distribuzione normale con media μ e varianza σ^2 , discende da quanto detto che la nuova variabile casuale

$$X' = \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^2$$

è distribuita come il χ^2 a N gradi di libertà. Indichiamo ora, al solito, con \bar{x} la media aritmetica delle x_i : vogliamo dimostrare che la variabile casuale

$$X'' = \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^2$$

è distribuita *ancora come il χ^2 , ma con $N - 1$ gradi di libertà*.

A questo scopo facciamo dapprima alcune considerazioni, indipendenti dalle ipotesi prima fatte sulle x_i e che risultano quindi valide per variabili casuali qualunque: supponiamo di definire N nuove variabili y_i come generiche combinazioni lineari delle x_j , con coefficienti che indicheremo col simbolo A_{ij} ; in modo insomma che risulti

$$y_i = \sum_{j=1}^N A_{ij} x_j .$$

La somma dei quadrati delle y_i è data da

$$\sum_{i=1}^N y_i^2 = \sum_{i=1}^N \left(\sum_{j=1}^N A_{ij} x_j \right) \left(\sum_{k=1}^N A_{ik} x_k \right) = \sum_{jk} x_j x_k \sum_i A_{ij} A_{ik} ;$$

è possibile che questa somma risulti uguale alla somma dei quadrati delle x_i *qualunque* sia il valore di queste ultime? Ovviamente questo avviene se e

solo se vale la

$$\sum_i A_{ij} A_{ik} = \delta_{jk} = \begin{cases} 0 & \text{per } j \neq k \\ 1 & \text{per } j = k \end{cases} \quad (12.3)$$

(il simbolo δ_{jk} , che assume il valore 1 quando gli indici sono uguali e 0 quando sono invece diversi, si chiama *simbolo di Kronecker* o *delta di Kronecker*).

Consideriamo gli A_{ij} come gli elementi di una matrice quadrata \mathbf{A} di ordine N ; gli x_j e le y_i si possono invece considerare come le componenti di due *vettori* \mathbf{X} ed \mathbf{Y} definiti in uno spazio N -dimensionale — ossia come gli elementi di due matrici rettangolari con N righe ed 1 colonna.

La trasformazione che muta \mathbf{X} in \mathbf{Y} si può scrivere, in forma matriciale, come $\mathbf{Y} = \mathbf{A}\mathbf{X}$; la somma dei quadrati delle x_j o delle y_i altro non è se non il prodotto scalare, di \mathbf{X} ed \mathbf{Y} rispettivamente, per loro stessi: ovverosia la loro *norma*, il quadrato della loro lunghezza nello spazio a N dimensioni. Quella che abbiamo ricavato adesso è la condizione perché una *trasformazione lineare* applicata ad un vettore ne conservi la lunghezza: occorre e basta che la matrice \mathbf{A} sia *ortogonale*. Infatti la (12.3) si può scrivere

$$\tilde{\mathbf{A}}\mathbf{A} = \mathbf{1} \quad \text{ossia} \quad \tilde{\mathbf{A}} = \mathbf{A}^{-1}$$

($\tilde{\mathbf{A}}$ è la matrice trasposta di \mathbf{A} , di elementi $\tilde{A}_{ij} = A_{ji}$; $\mathbf{1}$ è la matrice unità, di elementi $\mathbf{1}_{ij} = \delta_{ij}$; \mathbf{A}^{-1} è la matrice inversa di \mathbf{A} ; ed una matrice per cui $\tilde{\mathbf{A}} = \mathbf{A}^{-1}$ si dice, appunto, ortogonale).

Consideriamo adesso una trasformazione lineare definita dalle seguenti relazioni:

$$\begin{cases} y_1 = \frac{1}{\sqrt{N}} (x_1 + x_2 + \cdots + x_N) \\ y_2 = \frac{1}{\sqrt{2}} (x_1 - x_2) \\ y_3 = \frac{1}{\sqrt{6}} (x_1 + x_2 - 2x_3) \\ \cdots \\ y_N = \frac{1}{\sqrt{N(N-1)}} [x_1 + x_2 + \cdots + x_{N-1} - (N-1)x_N] \end{cases} \quad (12.4)$$

e per la quale la matrice di trasformazione abbia, insomma, elementi A_{ij}

definiti come

$$A_{ij} \equiv \begin{cases} i = 1: & \frac{1}{\sqrt{N}} \\ i > 1: & \begin{cases} j < i: & \frac{1}{\sqrt{i(i-1)}} \\ j = i: & -\frac{i-1}{\sqrt{i(i-1)}} \\ j > i: & 0 \end{cases} \end{cases}$$

Non è difficile controllare che la matrice A è ortogonale; inoltre la prima riga è stata scelta in modo tale che

$$y_1 = \sum_{i=1}^N \frac{1}{\sqrt{N}} x_i = \frac{1}{\sqrt{N}} \cdot N\bar{x} = \sqrt{N} \bar{x}$$

e quindi

$$\sum_{i=1}^N x_i^2 = \sum_{i=1}^N y_i^2 = N\bar{x}^2 + \sum_{i=2}^N y_i^2 .$$

Inoltre risulta (per $i > 1$)

$$\sum_{j=1}^N A_{ij} = \sum_{j=1}^{i-1} \frac{1}{\sqrt{i(i-1)}} - \frac{i-1}{\sqrt{i(i-1)}} = 0 \quad (12.5)$$

e, per ogni i ,

$$\sum_{j=1}^N A_{ij}^2 = (\mathbf{A}\tilde{\mathbf{A}})_{ii} = \delta_{ii} = 1 . \quad (12.6)$$

Tornando al nostro problema, supponiamo ora che tutte le x_j siano variabili aventi distribuzione normale; che abbiano tutte valore medio μ e varianza σ^2 ; ed inoltre che siano tra loro tutte statisticamente indipendenti. Una qualsiasi loro combinazione lineare, quindi anche ognuna delle y_i legate alle x_j da quella particolare matrice di trasformazione (12.4) che abbiamo prima definita, è anch'essa distribuita secondo la legge normale; inoltre risulta

$$\begin{aligned}
\frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \bar{x})^2 &= \frac{1}{\sigma^2} \left(\sum_{i=1}^N x_i^2 - N\bar{x}^2 \right) \\
&= \frac{1}{\sigma^2} \left(N\bar{x}^2 + \sum_{i=2}^N y_i^2 - N\bar{x}^2 \right) \\
&= \sum_{i=2}^N \frac{y_i^2}{\sigma^2} .
\end{aligned}$$

Applicando alle $y_i = \sum_j A_{ij}x_j$ le formule per la media e la varianza delle combinazioni lineari di variabili casuali statisticamente indipendenti già ricavate nel capitolo 5, si trova facilmente (tenendo presenti la (12.5) e la (12.6)) che la varianza di ognuna di esse è ancora σ^2 ; e che, per $i \neq 1$, il loro valore medio è 0. Di conseguenza, per $i \geq 2$ le y_i/σ sono variabili casuali normali aventi media 0 e varianza 1: e questo implica che

$$X'' = \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 \quad (12.7)$$

sia effettivamente distribuita come il χ^2 a $N - 1$ gradi di libertà.

È interessante confrontare questo risultato con quello precedentemente ricavato, 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 invece alla media aritmetica 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 contributi *indipendenti*: ovvero il numero di termini con distribuzione normale sommati in quadratura (qui N , uno per ogni determinazione x_i) diminuito del numero di parametri che compaiono nella formula e che sono stati stimati dai dati stessi (qui uno: appunto la media della popolazione, stimata usando la media aritmetica delle misure).

Un'ultima notevole conseguenza del fatto che la variabile casuale X'' definita dalla (12.7) sia distribuita come il χ^2 a $N - 1$ gradi di libertà è la seguente: la stima della varianza della popolazione ottenuta dal campione, s^2 , vale

$$s^2 = X'' \frac{\sigma^2}{N - 1} \quad (12.8)$$

e, essendo proporzionale a X'' , è anch'essa distribuita come il χ^2 a $N - 1$ gradi di libertà; quindi la sua densità di probabilità è data dalla (12.1) e dipende

solamente da N ; non dipende, in particolare, dalla media del campione \bar{x} . Quindi:

*Il valore medio \bar{x} e la varianza campionaria s^2 , calcolati su valori estratti a caso da una stessa popolazione normale, sono due variabili casuali **statisticamente indipendenti** tra loro.*

Questo risulta anche intuitivamente comprensibile; se infatti ci è noto che un certo campione di dati ha una dispersione più o meno grande, questo non deve alterare la probabilità che il suo valore medio abbia un valore piuttosto che un altro; né, viceversa, il fatto che il campione sia centrato attorno ad un certo valore deve permetterci di prevedere in qualche modo la sua dispersione.

12.2 Verifiche basate sulla distribuzione del χ^2

12.2.1 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, dalla 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 valore 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 deve obbedire poi alla distribuzione binomiale: il loro valore 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 di frequenza che governa il presentarsi, su un grande numero di osservazioni, di eventi aventi probabilità trascurabile di verificarsi singolarmente in ognuna: distribuzione nella quale l'errore quadratico medio è effettivamente dato dalla radice quadrata del valore 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 relativa funzione di distribuzione con

la funzione normale, la quantità

$$X = \sum_{i=1}^M \frac{(n_i - Np_i)^2}{Np_i} = \sum_{i=1}^M \frac{(O_i - A_i)^2}{A_i} \quad (12.9)$$

cioè la somma, su tutte le classi di frequenza (il cui numero abbiamo supposto sia M), 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 *approssimativamente* la distribuzione del χ^2 , con $M - 1$ gradi di libertà; il motivo di 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 (12.9), mediante la quale abbiamo definito X , per calcolare il numero A_i di misure attese in ogni intervallo).

La condizione enunciata si può in pratica supporre verificata se le A_i in ogni intervallo sono 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 10%). In realtà, se le classi di frequenza si possono scegliere arbitrariamente, la cosa migliore consiste nel definirle di ampiezze differenti: in modo tale che quegli intervalli dove cadono poche misure vengano riuniti assieme in un'unica classe più ampia, ove n_i valga almeno 5 (ma nemmeno troppo ampia, per soddisfare al vincolo di avere $p_i^2 \ll p_i$; in genere si cerca di riunire assieme più classi in modo da avere degli $n_i \sim 5 \div 10$).

Tornando al problema iniziale, per la verifica dell'ipotesi statistica che i dati vengano dalla distribuzione usata per il calcolo delle A_i basta:

- fissare arbitrariamente un livello di probabilità che rappresenti il confine tra eventi ammissibili nell'ipotesi della pura casualità ed eventi invece tanto improbabili da far supporre che il loro verificarsi sia dovuto non a fluttuazioni statistiche, ma al non essere verificate le ipotesi fatte in partenza (il provenire i dati dalla distribuzione nota a priori): ad esempio il 95% o il 99%.
- Cercare nelle apposite tabelle² il valore di taglio corrispondente alla coda superiore della distribuzione del χ^2 ad $M - 1$ gradi di libertà avente area pari al livello di confidenza desiderato; ossia quell'ascissa

²Alcuni valori numerici di questo tipo sono tabulati nell'appendice G. È bene anche ricordare che quando il numero di gradi di libertà N è superiore a 30 si può far riferimento alla distribuzione normale con media N ed errore quadratico medio $\sqrt{2N}$; e che, già per piccoli N , $\sqrt{2\chi^2}$ è approssimativamente normale con media $\sqrt{2N-1}$ e varianza 1.

ξ che lascia alla propria sinistra, sotto la curva della distribuzione del χ^2 ad $M - 1$ gradi di libertà, un'area pari a tale valore.

- Calcolare X ; ed infine rigettare l'ipotesi (al livello di confidenza prescelto) perché incompatibile con i dati raccolti, se X risultasse superiore a ξ (o, altrimenti, considerare l'ipotesi compatibile con i dati al livello di confidenza prescelto e quindi accettarla).

Quanto detto a proposito della particolare distribuzione del χ^2 da usare per la verifica della nostra ipotesi, però, è valido solo se le caratteristiche della distribuzione teorica con cui confrontare i nostri dati sono note a priori; se, invece, R parametri da cui essa dipende fossero stati stimati a partire dai dati, il numero di gradi di libertà sarebbe inferiore e pari ad $M - R - 1$.

Così se le p_i sono state ricavate integrando sulle classi di frequenza una distribuzione normale la cui media e la cui varianza siano state a loro volta ottenute dal campione istogrammato, il numero di gradi di libertà, essendo $R = 2$, sarebbe pari a $M - 3$.

Per dare un'idea dei valori del χ^2 che corrispondono al rigetto di una ipotesi (ad un certo livello di confidenza), e senza ricorrere alle tabelle numeriche, nella figura 12b sono riportati in grafico i valori P dell'integrale da x a $+\infty$ della funzione di frequenza del χ^2 (ovvero il complemento ad uno della funzione di distribuzione), per alcuni valori del parametro N .

Le curve della figura 12c permettono invece di identificare (per differenti scelte del livello di confidenza ε) i corrispondenti valori di taglio del χ^2 *ridotto* — ovvero del rapporto χ^2/N tra esso ed il numero di gradi di libertà N . Insomma, ogni punto di queste curve al di sopra di un'ascissa (intera) N ha come ordinata un numero X/N tale che l'integrale da X a $+\infty$ della funzione di frequenza del χ^2 ad N gradi di libertà sia uguale ad ε .

12.2.2 Il metodo del minimo χ^2

Supponiamo di sapere a priori che i nostri dati istogrammati debbano seguire una data distribuzione, ma che essa dipenda da R parametri incogniti che dobbiamo stimare a partire dai dati stessi; visto che l'accordo tra i dati e la distribuzione è dato dalla X definita nella (12.9), ed è tanto migliore quanto più il valore ottenuto per essa è basso, un metodo plausibile di stima potrebbe essere quello di trovare per quali valori dei parametri stessi la X è minima (*metodo del minimo χ^2*).

Indicando con α_k ($k = 1, \dots, R$) i parametri da stimare, ognuna delle p_i sarà esprimibile in funzione delle α_k ; ed imponendo che le derivate prime della X rispetto ad ognuna delle α_k siano tutte nulle contemporaneamente,

FIGURA 12b - L'integrale da x a $+\infty$ della funzione di frequenza del χ^2 , per alcuni valori del parametro N .

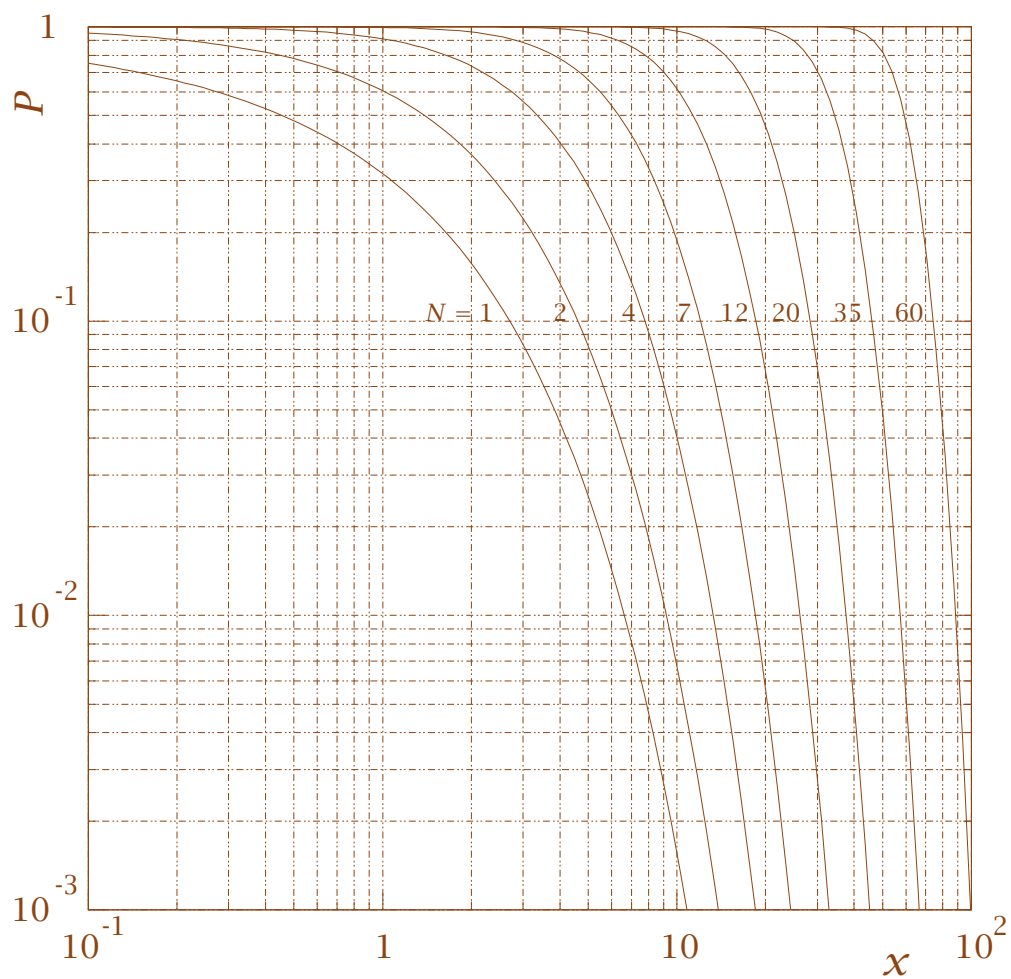
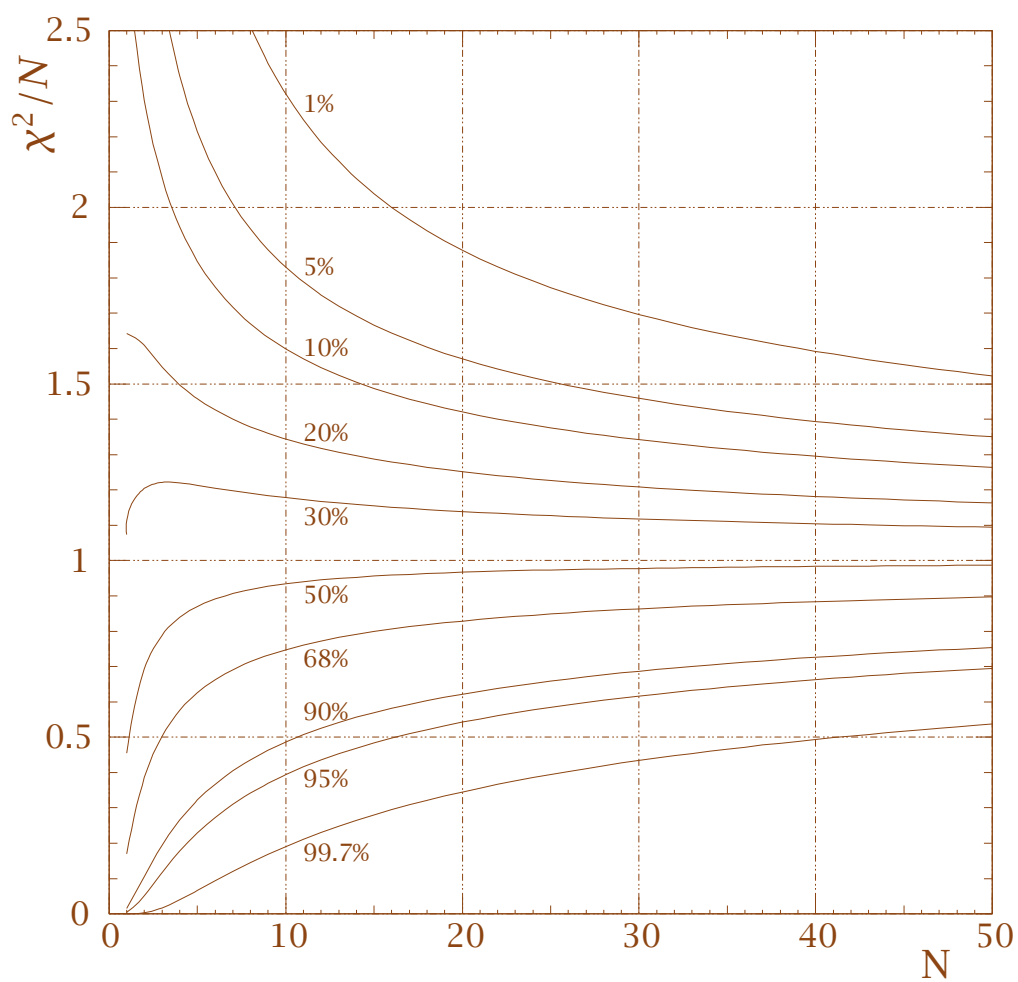


FIGURA 12c - I valori del χ^2 ridotto (χ^2/N) che corrispondono, per differenti gradi di libertà N , ad un certo livello di confidenza.



otteniamo

$$\frac{\partial X}{\partial \alpha_k} = \sum_{i=1}^M \frac{-2(n_i - Np_i) N^2 p_i - N(n_i - Np_i)^2}{N^2 p_i^2} \frac{\partial p_i}{\partial \alpha_k} = 0 ,$$

ossia

$$-\frac{1}{2} \frac{\partial X}{\partial \alpha_k} = \sum_{i=1}^M \left[\frac{n_i - Np_i}{p_i} + \frac{(n_i - Np_i)^2}{2Np_i^2} \right] \frac{\partial p_i}{\partial \alpha_k} = 0 . \quad (12.10)$$

L'insieme delle (12.10) costituisce un sistema di R equazioni, nelle R incognite α_k , che ci permetterà di stimarne i valori (salvo poi, nel caso il sistema delle (12.10) abbia più di una soluzione, controllare quali di esse corrispondono in effetti ad un minimo e quale tra queste ultime corrisponde al minimo assoluto); le condizioni sotto le quali il metodo è applicabile sono quelle già enunciate in precedenza³, ossia $p_i^2 \ll p_i$ e $n_i \gtrsim 5$.

In genere però si preferisce servirsi, in luogo delle equazioni (12.10), di una forma semplificata, ottenuta trascurando il secondo termine nella parentesi quadra: che, si può dimostrare, è molto inferiore al primo per grandi N (infatti il rapporto tra i due termini vale

$$\frac{(n_i - Np_i)^2}{2Np_i^2} \frac{p_i}{n_i - Np_i} = \frac{n_i - Np_i}{2Np_i} = \frac{1}{2p_i} \left(\frac{n_i}{N} - p_i \right)$$

e converge ovviamente a zero all'aumentare di N); e risolvere, insomma, il sistema delle

$$\sum_{i=1}^M \left(\frac{n_i - Np_i}{p_i} \right) \frac{\partial p_i}{\partial \alpha_k} = 0 \quad (12.11)$$

(metodo *semplificato* del minimo χ^2).

Si può dimostrare che le soluzioni $\bar{\alpha}_k$ del sistema delle (12.11) tendono stocasticamente ai valori veri α_k^* (in assenza di errori sistematici) al crescere di N ; inoltre il valore di X calcolato in corrispondenza dei valori ricavati $\bar{\alpha}_k$ dà, se rapportato alla distribuzione del χ^2 con $M - R - 1$ gradi di libertà, una misura della bontà della soluzione stessa.

Ora, le equazioni (12.11) si possono scrivere anche

$$\sum_{i=1}^M \left(\frac{n_i - Np_i}{p_i} \right) \frac{\partial p_i}{\partial \alpha_k} = \sum_{i=1}^M \frac{n_i}{p_i} \frac{\partial p_i}{\partial \alpha_k} - N \sum_{i=1}^M \frac{\partial p_i}{\partial \alpha_k}$$

³Se la prima di esse non si può ritenere accettabile, delle equazioni ancora valide ma più complesse si possono ottenere dalla (12.9) sostituendo $Np_i(1 - p_i)$ al posto di Np_i nel denominatore.

e si possono ulteriormente semplificare, visto che l'ultimo termine si annulla, essendo

$$\sum_{i=1}^M \frac{\partial p_i}{\partial \alpha_k} = \frac{\partial}{\partial \alpha_k} \sum_{i=1}^M p_i = \frac{\partial}{\partial \alpha_k} 1 \equiv 0$$

se si fa l'ulteriore ipotesi che l'intervallo dei valori indagati copra, anche approssimativamente, tutti quelli in pratica permessi; per cui il sistema di equazioni da risolvere è in questo caso quello delle

$$\sum_{i=1}^M \frac{n_i}{p_i} \frac{\partial p_i}{\partial \alpha_k} = 0 . \quad (12.12)$$

Per la stima di parametri incogniti a partire da dati misurati abbiamo già affermato che teoricamente è da preferire il metodo della massima verosimiglianza, le cui soluzioni sono quelle affette, come sappiamo, dal minimo errore casuale (almeno asintoticamente); in questo caso particolare (dati in istogramma), come lo si dovrebbe applicare? Se le misure sono indipendenti, la probabilità di avere n_i eventi nella generica classe di frequenza è data da $p_i^{n_i}$; la funzione di verosimiglianza⁴ da

$$\mathcal{L}(\alpha_1, \dots, \alpha_R) = \prod_{i=1}^M p_i^{n_i} \quad (12.13)$$

ed il suo logaritmo da

$$\ln \mathcal{L} = \sum_{i=1}^M (n_i \cdot \ln p_i) .$$

La soluzione di massima verosimiglianza (e quindi di minima varianza) si trova cercando il massimo di $\ln \mathcal{L}$: e risolvendo quindi il sistema delle

$$\frac{\partial}{\partial \alpha_k} \ln \mathcal{L} = \sum_{i=1}^M n_i \frac{1}{p_i} \frac{\partial p_i}{\partial \alpha_k} = 0 ;$$

in questo caso, vista l'equazione (12.12) in precedenza ricavata, i due metodi (della massima verosimiglianza e del minimo χ^2 semplificato) conducono dunque *alla stessa soluzione*.

⁴Per essere precisi, la probabilità che n_1 misure si trovino nella prima classe di frequenza, n_2 nella seconda e così via, è dato dalla espressione (12.13) moltiplicata per il numero di modi differenti in cui N oggetti possono essere suddivisi in M gruppi composti da n_1, n_2, \dots, n_M oggetti rispettivamente (numero delle *partizioni ordinate*); questo vale, come mostrato nel paragrafo A.7, $N!/(n_1! n_2! \dots n_M!)$, e rappresenta un fattore costante che non incide nella ricerca del massimo della (12.13).

12.2.3 Test di omogeneità per dati raggruppati

Supponiamo di avere a disposizione Q campioni di dati, indipendenti l'uno dall'altro e composti da n_1, n_2, \dots, n_Q elementi rispettivamente; e, all'interno di ognuno di tali campioni, i dati siano suddivisi nei medesimi P gruppi: indichiamo infine col simbolo v_{ij} il numero di dati appartenenti al gruppo i -esimo all'interno del campione j -esimo.

Per fare un esempio, i campioni si potrebbero riferire alle regioni italiane e i gruppi al livello di istruzione (licenza elementare, media, superiore, laurea): così che i v_{ij} rappresentino il numero di persone, per ogni livello di istruzione, residenti in ogni data regione; oppure (e questo è un caso che si presenta frequentemente nelle analisi fisiche) si abbiano vari istogrammi all'interno di ognuno dei quali i dati siano stati raggruppati secondo le medesime classi di frequenza: allora i v_{ij} saranno il numero di osservazioni che cadono in una determinata classe in ogni istogramma.

Il problema che ci poniamo è quello di verificare l'ipotesi che tutti i campioni provengano dalla stessa popolazione e siano perciò compatibili tra loro (*test di omogeneità*). Indichiamo con il simbolo N il numero totale di dati a disposizione; e con m_i (con $i = 1, \dots, P$) il numero totale di dati che cadono nell' i -esimo gruppo in tutti i campioni a disposizione.

TABELLA 12.1 - Un esempio delle cosiddette *tabelle delle contingenze*.

Gruppi	Campioni					
	v_{11}	v_{12}	v_{13}	\dots	v_{1Q}	
	v_{21}	v_{22}	\dots	\dots	v_{2Q}	
	v_{31}	\dots	\dots	\dots	\dots	
	\dots	\dots	\dots	\dots	\dots	
	v_{P1}	v_{P2}	\dots	\dots	v_{PQ}	
	n_1	n_2	n_3	\dots	n_Q	

È consuetudine che dati di questo genere siano rappresentati in una tabella del tipo della 12.1, che si chiama *tabella delle contingenze*; e risulta

ovviamente

$$n_j = \sum_{i=1}^P v_{ij} \quad (j = 1, 2, \dots, Q) ;$$

$$m_i = \sum_{j=1}^Q v_{ij} \quad (i = 1, 2, \dots, P) ;$$

$$N = \sum_{j=1}^Q n_j = \sum_{i=1}^P m_i = \sum_{i,j} v_{ij} .$$

Vogliamo ora dimostrare che la variabile casuale

$$X = N \left[\sum_{i,j} \frac{(v_{ij})^2}{m_i n_j} - 1 \right] \quad (12.14)$$

è distribuita come il χ^2 a $(P - 1)(Q - 1)$ gradi di libertà: a questo scopo supponiamo innanzi tutto sia valida l'ipotesi che i dati provengano tutti dalla medesima popolazione, ed indichiamo con i simboli p_i e q_j le probabilità che un componente di tale popolazione scelto a caso cada rispettivamente nel gruppo i -esimo o nel campione j -esimo; e sappiamo inoltre che (ammessa però vera l'ipotesi che *tutti* i campioni provengano dalla stessa distribuzione) questi due eventi sono statisticamente indipendenti: per cui ognuno dei dati ha probabilità complessiva $p_i q_j$ di cadere in una delle caselle della tabella delle contingenze.

Possiamo stimare i P valori p_i a partire dai dati sperimentali: si tratta in realtà solo di $P - 1$ stime *indipendenti*, perché, una volta ricavate le prime $P - 1$ probabilità, l'ultima di esse risulterà univocamente determinata dalla condizione che la somma complessiva valga 1. Analogamente possiamo anche stimare i Q valori q_j dai dati sperimentali, e si tratterà in questo caso di effettuare $Q - 1$ stime indipendenti.

Le stime di cui abbiamo parlato sono ovviamente

$$p_i = \frac{m_i}{N} \quad \text{e} \quad q_j = \frac{n_j}{N} \quad (12.15)$$

e, applicando le conclusioni del paragrafo precedente (l'equazione (12.9)), la

variabile

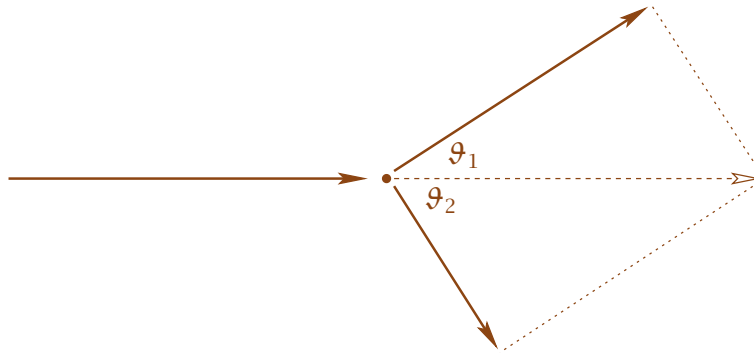
$$\begin{aligned}
 X &= \sum_{i,j} \frac{(v_{ij} - Np_iq_j)^2}{Np_iq_j} \\
 &= \sum_{i,j} \left[\frac{(v_{ij})^2}{Np_iq_j} - 2v_{ij} + Np_iq_j \right] \\
 &= \sum_{i,j} \frac{(v_{ij})^2}{Np_iq_j} - 2N + N \\
 &= \sum_{i,j} \frac{(v_{ij})^2}{Np_iq_j} - N
 \end{aligned}$$

deve essere distribuita come il χ^2 .

Sostituendo in quest'ultima espressione i valori (12.15) per p_i e q_j , essa si riduce alla (12.14); il numero di gradi di libertà è pari al numero di contributi sperimentali indipendenti, $PQ - 1$ (c'è il vincolo che la somma totale sia N), diminuito del numero $(P - 1) + (Q - 1)$ di parametri stimato sulla base dei dati: ovverosia proprio $(P - 1)(Q - 1)$ come anticipato.

12.2.4 Un esempio: diffusione elastica protone-protone

FIGURA 12d - Urto elastico protone-protone.



Nella figura 12d è schematicamente rappresentato un processo di urto elastico tra due particelle, una delle quali sia inizialmente ferma; dopo l'urto esse si muoveranno lungo traiettorie rettilinee ad angoli ϑ_1 e ϑ_2 rispetto alla direzione originale della particella urtante.

Gli angoli ϑ_i vengono misurati; supponendo che il processo di misura introduca errori che seguono la distribuzione normale ed abbiano una entità che (per semplificare le cose) assumiamo sia costante, nota ed indipendente dall'ampiezza dell'angolo, vogliamo verificare l'ipotesi che le due particelle coinvolte nel processo d'urto siano di massa uguale (ad esempio che siano entrambe dei protoni).

La prima cosa da fare è quella di ricavare dai dati misurati ϑ_i , che per ipotesi hanno una funzione di frequenza

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

una stima dei valori veri ϑ^* . Il logaritmo della funzione di verosimiglianza è dato da

$$\ln \mathcal{L} = -2 \ln(\sigma\sqrt{2\pi}) - \frac{1}{2} \left(\frac{\vartheta_1 - \vartheta_1^*}{\sigma} \right)^2 - \frac{1}{2} \left(\frac{\vartheta_2 - \vartheta_2^*}{\sigma} \right)^2 ;$$

ma le variabili ϑ_1 e ϑ_2 *non sono indipendenti*, visto che il processo deve conservare sia energia che quantità di moto. Ammessa vera l'ipotesi che le due particelle abbiano uguale massa (e restando nel limite non-relativistico), le leggi di conservazione impongono il vincolo che l'angolo tra le due particelle dopo l'urto sia di 90° (o, in radianti, $\pi/2$); usando il metodo dei moltiplicatori di Lagrange, la funzione da massimizzare è

$$\varphi(\vartheta_1^*, \vartheta_2^*, \lambda) = -\frac{1}{2} \left(\frac{\vartheta_1 - \vartheta_1^*}{\sigma} \right)^2 - \frac{1}{2} \left(\frac{\vartheta_2 - \vartheta_2^*}{\sigma} \right)^2 + \lambda \left(\vartheta_1^* + \vartheta_2^* - \frac{\pi}{2} \right)$$

e, annullando contemporaneamente le sue derivate rispetto alle tre variabili, si giunge al sistema

$$\begin{cases} \frac{\partial \varphi}{\partial \lambda} = \vartheta_1^* + \vartheta_2^* - \frac{\pi}{2} = 0 \\ \frac{\partial \varphi}{\partial \vartheta_1^*} = \frac{1}{\sigma^2} (\vartheta_1 - \vartheta_1^*) + \lambda = 0 \\ \frac{\partial \varphi}{\partial \vartheta_2^*} = \frac{1}{\sigma^2} (\vartheta_2 - \vartheta_2^*) + \lambda = 0 \end{cases}$$

Eliminando λ dalle ultime due equazioni otteniamo

$$\vartheta_1 - \vartheta_1^* = \vartheta_2 - \vartheta_2^*$$

e, sostituendo l'espressione per ϑ_2^* ricavata dalla prima equazione,

$$\vartheta_1 - \vartheta_1^* = \vartheta_2 - \left(\frac{\pi}{2} - \vartheta_1^* \right)$$

per cui le due stime di massima verosimiglianza sono

$$\begin{cases} \hat{\vartheta}_1^* &= \vartheta_1 + \frac{1}{2} \left(\frac{\pi}{2} - \vartheta_1 - \vartheta_2 \right) \\ \hat{\vartheta}_2^* &= \vartheta_2 + \frac{1}{2} \left(\frac{\pi}{2} - \vartheta_1 - \vartheta_2 \right) \end{cases}$$

Ammettendo che queste soluzioni siano buone stime dei valori veri, la variabile casuale

$$X = \left(\frac{\vartheta_1 - \vartheta_1^*}{\sigma} \right)^2 + \left(\frac{\vartheta_2 - \vartheta_2^*}{\sigma} \right)^2 = \frac{1}{2\sigma^2} \left(\frac{\pi}{2} - \vartheta_1 - \vartheta_2 \right)^2$$

è distribuita come il χ^2 ad un grado di libertà (due contributi, un vincolo); ed il valore di X confrontato con le tabelle del χ^2 può essere usato per la verifica dell'ipotesi.

12.3 Compatibilità con un valore prefissato

Un altro caso che frequentemente si presenta è il seguente: si vuole 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 9.3 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 $-1 \leq t \leq 1$ corrisponde al 68.27% dell'area della curva normale, e quello $-2 \leq t \leq 2$ al 95.45%.

Se consideriamo poi un campione di N misure indipendenti, avente valore 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

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

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 ξ il valore vero della grandezza misurata sarà compatibile o incompatibile con i nostri dati a seconda che lo scarto normalizzato

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

relativo a tale numero sia, in valore assoluto, 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 :

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

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 perché 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%.

TABELLA 12.2 - Alcuni valori della probabilità P e dei corrispondenti limiti τ sullo scarto normalizzato, per verifiche two-tailed (τ_2) o one-tailed (τ_1).

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

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

Se l'ipotesi da verificare riguarda l'essere *differenti tra loro* due entità (il

presupposto valore 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% (usando il termine inglese, stiamo effettuando un *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 12.2 e 12.3; altri si possono facilmente ricavare dalle tabelle dell'appendice G.

12.4 I piccoli campioni e la distribuzione di Student

Cosa si può fare riguardo alla verifica di ipotesi statistiche come quella (considerata nel paragrafo precedente) della compatibilità del risultato delle misure con un valore noto a priori, quando si abbiano a disposizione solamente piccoli campioni? Ci riferiamo, più esattamente, a 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).

Sia X una variabile casuale distribuita come il χ^2 ad N gradi di libertà, ed u una seconda variabile casuale, indipendente dalla prima, e avente distribuzione normale standardizzata $N(u; 0, 1)$; consideriamo la nuova variabile casuale t definita attraverso la

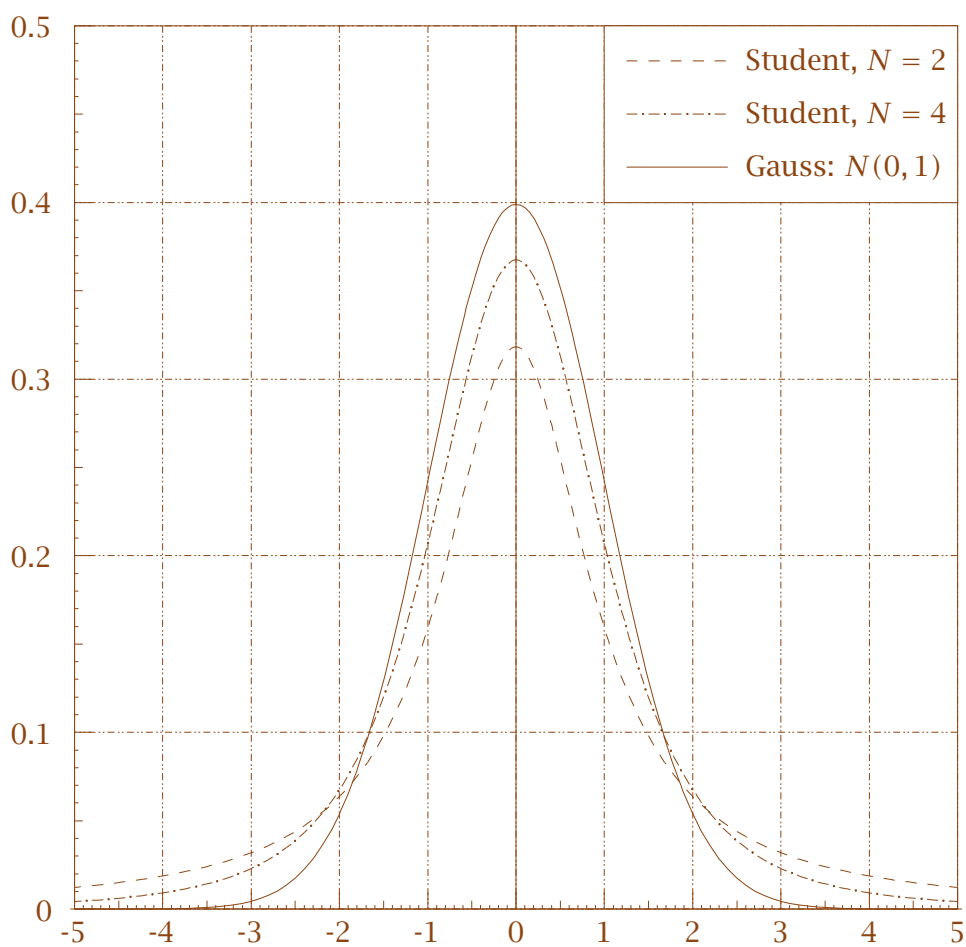
$$t = \frac{u}{\sqrt{\frac{X}{N}}} . \quad (12.16)$$

Si può dimostrare che la funzione densità di probabilità relativa alla variabile casuale t è data dalla

$$f(t; N) = \frac{T_N}{\left(1 + \frac{t^2}{N}\right)^{\frac{N+1}{2}}}$$

che si chiama *distribuzione di Student ad N gradi di libertà*.

FIGURA 12e - 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 (come si potrebbe facilmente provare partendo dal limite notevole (9.9)) 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 di Student ricorda molto quella della funzione di Gauss, come appare evidente dalla figura 12e; soltanto, rispetto a dati che seguano la distribuzione normale, valori elevati dello scarto sono relativamente più probabili⁵.

La distribuzione di Student è simmetrica, quindi tutti i momenti di ordine dispari (compreso il valore medio λ_1) sono nulli; mentre la varianza della distribuzione è

$$\text{Var}(t) = \frac{N}{N-2}$$

(se $N > 2$); ed il coefficiente di curtosi vale

$$\gamma_2 = \frac{6}{N-4}$$

(se $N > 4$).

Indicando con \bar{x} la media aritmetica di un campione di dimensione N , estratto a caso da una popolazione normale avente valore medio $E(x)$ e varianza σ^2 ; e con s la stima della deviazione standard della popolazione ottenuta dal campione stesso, cioè

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

sappiamo, ricordando l'equazione (12.8), che la variabile casuale

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

è distribuita come il χ^2 ad $N-1$ gradi di libertà; inoltre, ovviamente,

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

segue la legge normale, con media 0 e varianza 1. Di conseguenza la variabile casuale

$$t = \frac{u}{\sqrt{\frac{X''}{N-1}}} = \frac{\bar{x} - E(x)}{\frac{s}{\sqrt{N}}} \quad (12.17)$$

⁵Per valori di $N \gtrsim 35$ la distribuzione di Student si può approssimare con la distribuzione normale a media 0 e varianza 1.

segue la distribuzione di Student ad $N - 1$ gradi di libertà.

Insomma: 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 occorrerà confrontare il suo valore con i limiti degli intervalli di confidenza relativi alla distribuzione di Student e non alla distribuzione normale⁶.

12.5 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 valore 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; 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^2 e σ_y^2 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

⁶Per taluni più usati valori del livello di confidenza, i limiti rilevanti si possono trovare tabulati anche nell'appendice G.

campioni provengano da popolazioni aventi la stessa media si traduce nella verifica dell'ipotesi che δ abbia valore 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 deviazioni standard 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.

In caso contrario, presupponendo cioè di avere a disposizione piccoli campioni per almeno una delle due variabili, limitiamo la nostra analisi al caso in cui si sappia con sicurezza che le due popolazioni x ed y *abbiano la stessa varianza*,

$$\sigma_x^2 = \sigma_y^2 \equiv \sigma^2$$

e definiamo la grandezza S^2 (*varianza globale* dei campioni) come

$$\begin{aligned} S^2 &= \frac{1}{N+M-2} \cdot \left\{ \sum_{i=1}^N [x_i - \bar{x}]^2 + \sum_{j=1}^M [y_j - \bar{y}]^2 \right\} \\ &= \frac{(N-1)s_x^2 + (M-1)s_y^2}{N+M-2}. \end{aligned}$$

Sapendo, dall'equazione (12.8), che le due variabili

$$(N-1) \frac{s_x^2}{\sigma^2} \quad \text{e} \quad (M-1) \frac{s_y^2}{\sigma^2}$$

sono entrambe distribuite come il χ^2 , con $N-1$ ed $M-1$ gradi di libertà rispettivamente, sfruttando la regola di somma enunciata a pagina 199 si ricava che la variabile casuale

$$X = \frac{(N-1)s_x^2 + (M-1)s_y^2}{\sigma^2} = (N+M-2) \frac{S^2}{\sigma^2}$$

è distribuita come il χ^2 ad $N+M-2$ gradi di libertà; essendo inoltre $\delta = \bar{x} - \bar{y}$ una variabile normale con media e varianza date da

$$E(\delta) = E(x) - E(y) \quad \text{e} \quad \sigma_\delta^2 = \frac{\sigma^2}{N} + \frac{\sigma^2}{M}$$

la variabile casuale

$$u = \frac{\delta - E(\delta)}{\sigma_\delta} = \frac{(\bar{x} - \bar{y}) - [E(x) - E(y)]}{\sqrt{\sigma^2 \left(\frac{1}{N} + \frac{1}{M} \right)}}$$

è normale con media 0 e varianza 1. Per concludere,

$$t = \frac{u}{\sqrt{\frac{X}{N+M-2}}} = \frac{(\bar{x} - \bar{y}) - [E(x) - E(y)]}{\sqrt{S^2 \left(\frac{1}{N} + \frac{1}{M} \right)}} \quad (12.18)$$

deve seguire la distribuzione di Student con $N+M-2$ gradi di libertà; di conseguenza, per verificare l'ipotesi che le due popolazioni *normali* da cui i campioni provengono abbiano la stessa media *ammesso già che posseggano la stessa varianza*, basta confrontare con le apposite tabelle della distribuzione di Student il valore della t ottenuta dalla (12.18) ponendovi $E(x) - E(y) = 0$:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{S^2 \left(\frac{1}{N} + \frac{1}{M} \right)}} .$$

12.6 La distribuzione di Fisher

Sia X una variabile casuale distribuita come il χ^2 ad M gradi di libertà; ed Y una seconda variabile casuale, indipendente dalla prima, distribuita ancora come il χ^2 , ma con N gradi di libertà.

La variabile casuale w (sempre positiva) definita in funzione di esse attraverso la relazione

$$w = \frac{\frac{X}{M}}{\frac{Y}{N}}$$

ha una densità di probabilità che segue la cosiddetta *funzione di frequenza di Fisher* con M ed N gradi di libertà. La forma analitica della funzione di Fisher è data dalla

$$F(w; M, N) = K_{MN} \frac{w^{\frac{M}{2}-1}}{(Mw + N)^{\frac{M+N}{2}}} \quad (12.19)$$

(nella quale K_{MN} è un fattore costante determinato dalla condizione di normalizzazione).

Il valore medio e la varianza della funzione di frequenza di Fisher sono dati poi rispettivamente da

$$E(F) = \frac{N}{N-2} \quad (\text{se } N > 2)$$

e da

$$\text{Var}(F) = \frac{2 N^2 (M + N - 2)}{M(N-2)^2(N-4)} \quad (\text{se } N > 4) .$$

Si può dimostrare che, se Y è una variabile casuale distribuita come il χ^2 ad N gradi di libertà,

$$\lim_{N \rightarrow +\infty} \frac{Y}{N} = 1$$

in senso statistico (ovverosia la probabilità che il rapporto Y/N sia differente da 1 tende a zero quando N viene reso arbitrariamente grande); per cui, indicando con $f(x; M)$ la funzione di frequenza del χ^2 ad M gradi di libertà,

$$F(w; M, \infty) \equiv \lim_{N \rightarrow +\infty} F(w; M, N) = \frac{f(w; M)}{M} .$$

Allo stesso modo

$$F(w; \infty, N) \equiv \lim_{M \rightarrow +\infty} F(w; M, N) = \frac{N}{f(w; N)}$$

e quindi esiste una stretta relazione tra le distribuzioni di Fisher e del chi quadro.

Inoltre, ricordando che, se u è una variabile casuale distribuita secondo la legge normale standardizzata $N(u; 0, 1)$, l'altra variabile casuale u^2 è distribuita come il χ^2 ad un grado di libertà, il rapporto

$$w = \frac{u^2}{\frac{Y}{N}}$$

deve essere distribuito secondo $F(w; 1, N)$; ma, se definiamo

$$t = \frac{u}{\sqrt{\frac{Y}{N}}}$$

sappiamo anche dalla (12.16) che la t segue la distribuzione di Student ad N gradi di libertà. La conclusione è che il quadrato di una variabile t che segua

la distribuzione di Student ad N gradi di libertà è a sua volta distribuito con una densità di probabilità data da $F(t^2; 1, N)$.

Per terminare, quando i due parametri M ed N (da cui la funzione di frequenza di Fisher (12.19) dipende) vengono resi arbitrariamente grandi, essa tende ad una distribuzione normale; ma la convergenza è lenta, e l'approssimazione normale alla distribuzione di Fisher si può pensare in pratica usabile quando sia M che N sono superiori a 50.

12.6.1 Confronto tra varianze

Supponiamo di avere a disposizione due campioni di misure, che ipotizziamo provenire da due differenti popolazioni che seguano delle distribuzioni normali.

Siano M ed N le dimensioni di tali campioni, e siano σ_1^2 e σ_2^2 le varianze delle rispettive popolazioni di provenienza; indichiamo poi con s_1^2 ed s_2^2 le due stime delle varianze delle popolazioni ricavate dai campioni. Vogliamo ora capire come si può verificare l'ipotesi statistica che le due popolazioni abbiano *la stessa varianza*, ossia che $\sigma_1 = \sigma_2$.

Ora sappiamo già dalla equazione (12.8) che le due variabili casuali

$$X = (M - 1) \frac{s_1^2}{\sigma_1^2} \quad \text{e} \quad Y = (N - 1) \frac{s_2^2}{\sigma_2^2}$$

sono entrambe distribuite come il χ^2 , con $M - 1$ ed $N - 1$ gradi di libertà rispettivamente; quindi la quantità

$$w = \frac{X}{M - 1} \frac{N - 1}{Y} = \frac{s_1^2}{\sigma_1^2} \frac{\sigma_2^2}{s_2^2}$$

ha densità di probabilità data dalla funzione di Fisher con $M - 1$ ed $N - 1$ gradi di libertà.

Assunta a priori vera l'ipotesi statistica $\sigma_1 = \sigma_2$, la variabile casuale

$$w = \frac{s_1^2}{s_2^2}$$

ha densità di probabilità data dalla funzione di Fisher prima menzionata, $F(w; M - 1, N - 1)$; per cui, fissato un livello di confidenza al di là del quale rigettare l'ipotesi, e ricavato dalle apposite tabelle⁷ il valore W che lascia alla propria sinistra, al di sotto della funzione $F(w; M - 1, N - 1)$, un'area pari al livello di confidenza prescelto, si può escludere che i due campioni provengano da popolazioni con la stessa varianza se $w > W$.

⁷Per un livello di confidenza pari a 0.95 o 0.99, e per alcuni valori dei due parametri M ed N , ci si può riferire ancora alle tabelle dell'appendice G; in esse si assume che sia $s_1 > s_2$, e quindi $w > 1$.

12.7 Il metodo di Kolmogorov e Smirnov

Il *test di Kolmogorov e Smirnov* è un metodo di analisi statistica che permette di confrontare tra loro un campione di dati ed una distribuzione teorica (oppure due campioni di dati) allo scopo di verificare l'ipotesi statistica che la popolazione da cui i dati provengono sia quella in esame (oppure l'ipotesi che entrambi i campioni provengano dalla stessa popolazione).

Una caratteristica interessante di questo metodo è che esso non richiede la preventiva, e più o meno arbitraria, suddivisione dei dati in classi di frequenza; definendo queste ultime in modo diverso si ottengono ovviamente, dal metodo del χ^2 , differenti risultati per gli stessi campioni.

Il test di Kolmogorov e Smirnov si basa infatti sulla *frequenza cumulativa relativa* dei dati, introdotta nel paragrafo 4.1 a pagina 33; e sull'analogo concetto di *funzione di distribuzione* di una variabile continua definito nel paragrafo 6.1 a pagina 68. Per la compatibilità tra un campione ed una ipotetica legge che si ritiene possa descriverne la popolazione di provenienza, e collegata ad una funzione di distribuzione $\Phi(x)$, bisogna confrontare la frequenza cumulativa relativa $F(x)$ del campione con $\Phi(x)$ per ricavare il *valore assoluto del massimo scarto tra esse*,

$$\delta = \max \left\{ |F(x) - \Phi(x)| \right\} .$$

Si può dimostrare che, se l'ipotesi da verificare fosse vera, la probabilità di ottenere casualmente un valore di δ non inferiore ad una prefissata quantità (positiva) δ_0 sarebbe data da

$$\Pr (\delta \geq \delta_0) = F_{KS} (\delta'_0)$$

ove F_{KS} è la serie

$$F_{KS}(x) = 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2 x^2} \quad (12.20)$$

e δ'_0 vale

$$\delta'_0 = \left(\sqrt{N} + 0.12 + \frac{0.11}{\sqrt{N}} \right) \delta_0 . \quad (12.21)$$

La legge ora enunciata è approssimata, ma il test di Kolmogorov e Smirnov può essere usato già per dimensioni del campione N uguali a 5. Attenzione però che, se qualche parametro da cui la distribuzione teorica dipende è stato stimato sulla base dei dati, l'integrale della densità di probabilità

per la variabile δ di Kolmogorov e Smirnov *non segue più* la legge (12.20): non solo, ma non è più possibile ricavare teoricamente una funzione che ne descriva il comportamento in generale (in questi casi, nella pratica, la distribuzione di δ viene studiata usando metodi di Montecarlo).

Se si vogliono invece confrontare tra loro due campioni indipendenti per verificarne la compatibilità, bisogna ricavare dai dati il massimo scarto (in valore assoluto), δ , tra le due frequenze cumulative relative; e ricavare ancora dalla (12.20) la probabilità che questo possa essere avvenuto (ammessa vera l'ipotesi) per motivi puramente casuali. L'unica differenza è che la funzione (12.20) va calcolata in un'ascissa δ'_0 data dalla (12.21), nella quale N vale

$$N = \frac{1}{\frac{1}{N_1} + \frac{1}{N_2}} = \frac{N_1 N_2}{N_1 + N_2}$$

(N_1 ed N_2 sono le dimensioni dei due campioni).

Oltre al già citato vantaggio di non richiedere la creazione di più o meno arbitrarie classi di frequenza per raggrupparvi i dati, un'altra caratteristica utile del test di Kolmogorov e Smirnov è quella di essere, entro certi limiti, *indipendente dalla variabile usata* nella misura: se al posto di x si usasse, per caratterizzare il campione, $\ln(x)$ o \sqrt{x} , il massimo scarto tra frequenza cumulativa e funzione di distribuzione rimarrebbe invariato.

Un altrettanto ovvio svantaggio è collegato al fatto che per valori molto piccoli (o molto grandi) della variabile casuale usata, qualsiasi essa sia, *tutte* le funzioni di distribuzione e tutte le frequenze cumulative *hanno lo stesso valore* (0, o 1 rispettivamente). Per questo motivo il test di Kolmogorov e Smirnov è assai sensibile a differenze nella zona centrale dei dati (attorno al valore medio), mentre non è affatto efficace per discriminare tra due distribuzioni che differiscano significativamente tra loro solo nelle code; ad esempio che abbiano lo stesso valore medio e differente ampiezza.

Capitolo 13

La verifica delle ipotesi (II)

Nel precedente capitolo 12 abbiamo esaminato varie tecniche che ci permettono di decidere se una caratteristica del processo fisico che ha prodotto un campione di dati è o non è confermata dai dati stessi; tutte queste tecniche non sono che casi particolari di una teoria generale, di cui ora ci occuperemo, senza però scendere in profondità nei dettagli.

In sostanza, nei vari casi del capitolo 12, abbiamo formulato una certa ipotesi H_0 sulla natura di un fenomeno casuale; e, ammesso per assurdo che questa ipotesi fosse vera, abbiamo associato un ben definito valore della densità di probabilità ad ogni punto E dello spazio S degli eventi.

Se indichiamo con K un valore (arbitrariamente scelto) della probabilità, *livello di confidenza* nel linguaggio statistico, abbiamo in sostanza diviso S in due sottoinsiemi esclusivi ed esaurienti: uno \mathcal{R} di eventi con probabilità complessiva $1 - K$, ed uno $\mathcal{A} = S - \mathcal{R}$ di eventi con probabilità complessiva K .

Per verificare l'ipotesi H_0 occorre scegliere a priori un valore di K da assumere come il confine che separi, da una parte, eventi che riteniamo ragionevole si possano presentare nell'ambito di pure fluttuazioni casuali se è vera H_0 ; e, dall'altra, eventi così improbabili (sempre ammesso che H_0 sia vera) da far sì che la loro effettiva realizzazione debba implicare la falsità dell'ipotesi.

Normalmente si sceglie $K = 0.95$ o $K = 0.997$, i valori della probabilità che corrispondono a scarti di due o tre errori quadratici medi per la distribuzione di Gauss, anche se altri valori (come ad esempio $K = 0.999$ o $K = 0.99$) sono abbastanza comuni; e, una volta fatto questo, si *rigetta l'ipotesi* H_0 se il

dato a disposizione (un evento E ottenuto dall'effettivo studio del fenomeno in esame) appartiene ad \mathcal{R} ; e la si accetta se appartiene ad \mathcal{A} .

In realtà nella pratica si presenta in generale la necessità di discriminare tra *due* ipotesi, sempre mutuamente esclusive, che indicheremo con i simboli H_0 ed H_a e che, usando la terminologia della statistica, si chiamano rispettivamente *ipotesi nulla* ed *ipotesi alternativa*; i casi precedenti corrispondono al caso particolare in cui l'ipotesi alternativa coincida con il *non realizzarsi* di H_0 .

Ipotesi nulla ed ipotesi alternativa possono essere entrambe eventi semplici, oppure composti (ossia somma logica di più eventualità semplici); e lo scopo di questo capitolo è quello di mostrare dei criteri sulla base dei quali si possa opportunamente definire nello spazio degli eventi una *regione di rigetto* \mathcal{R} per l'ipotesi nulla (e, in corrispondenza, ovviamente, una regione $\mathcal{A} = S - \mathcal{R}$ nella quale tale ipotesi viene accettata).

È chiaro che si corre sempre il rischio di sbagliare: o rigettando erroneamente ipotesi in realtà vere (*errori di prima specie*) o accettando invece ipotesi in realtà false (*errori di seconda specie*); e che, allargando o restringendo la regione di rigetto, si può diminuire la probabilità di uno di questi due tipi di errori solo per aumentare la probabilità di quelli dell'altra categoria. Se indichiamo con P_I e P_{II} le probabilità degli errori di prima e seconda specie rispettivamente, sulla base della definizione risulta

$$P_I = \Pr(E \in \mathcal{R} | H_0) \quad \text{e} \quad P_{II} = \Pr(E \in \mathcal{A} | H_a) .$$

Quello che abbiamo finora chiamato “livello di confidenza” non è altro che $1 - P_I$; P_I viene anche indicato col simbolo α e chiamato *significanza* del criterio adottato. Infine, la probabilità di *non* commettere un errore di seconda specie, ovvero la probabilità di rigettare H_0 quando l'ipotesi nulla è falsa (e quindi quella alternativa è vera) si indica col simbolo β e si chiama *potenza* del criterio adottato; essa vale quindi

$$\beta = \Pr(E \in \mathcal{R} | H_a) = 1 - P_{II} .$$

Per fare un esempio concreto, il fisico si trova spesso ad esaminare “eventi” sperimentali e deve decidere se essi sono del tipo desiderato (segnale) o no (fondo): in questo caso l'ipotesi nulla H_0 consiste nell'appartenenza di un evento al segnale, mentre l'ipotesi alternativa H_a corrisponde invece all'appartenenza dello stesso evento al fondo; che in genere non è l'intero insieme di eventi complementare all'ipotesi nulla, \bar{H}_0 , ma si sa restringere ad una classe ben definita di fenomeni.

Gli errori di prima specie consistono in questo caso nello scartare eventi buoni (errori di *impoverimento* del segnale), e quelli di seconda specie nell'introduzione nel segnale di eventi di fondo (errori di *contaminazione*).

I criteri da seguire per definire una regione \mathcal{R} nella quale rigettare H_0 sono dettati dalle caratteristiche del processo di generazione: se gli eventi di fondo sono preponderanti rispetto al segnale, ad esempio, bisognerà evitare gli errori di seconda specie per quanto possibile; anche al prezzo di scartare in questo modo una parte consistente del segnale.

Estendendo al caso generale il metodo seguito nei vari casi del capitolo 12 e prima delineato, se si è in grado di associare ad ogni punto dello spazio degli eventi *due* valori della probabilità (o della densità di probabilità nel caso di variabili continue), sia ammessa vera l'ipotesi nulla che ammessa invece vera l'ipotesi alternativa, si può pensare di usare *il loro rapporto* per definire la regione di rigetto.

Limitandoci al caso delle variabili continue, insomma, dopo aver definito una nuova variabile casuale λ attraverso la

$$\lambda = \frac{\mathcal{L}(\mathbf{x}|H_0)}{\mathcal{L}(\mathbf{x}|H_a)} ,$$

possiamo scegliere arbitrariamente un numero reale k e decidere di accettare l'ipotesi H_0 se $\lambda \geq k$ o di rifiutarla se $\lambda < k$; in definitiva ad ogni k ammissibile è associata una differente regione di rigetto \mathcal{R}_k definita da

$$\mathcal{R}_k \equiv \left\{ \lambda = \frac{\mathcal{L}(\mathbf{x}|H_0)}{\mathcal{L}(\mathbf{x}|H_a)} < k \right\} .$$

\mathcal{L} , nelle espressioni precedenti, è la funzione di verosimiglianza; che rappresenta appunto la densità di probabilità corrispondente all'ottenere (sotto una certa ipotesi) un campione di N valori x_1, x_2, \dots, x_N (qui indicato sinteticamente come un vettore \mathbf{x} a N componenti). Ma in base a quale criterio dobbiamo scegliere k ?

13.1 Un primo esempio

Cominciamo con un esempio didattico: supponiamo che i valori x_i si sappiano provenienti da una popolazione normale $N(x; \mu, \sigma)$ di varianza σ^2 nota: e che il nostro scopo consista nel discriminare tra due possibili valori μ_1 e μ_2 per μ ; valori che, senza perdere in generalità, supponiamo siano 0 e 1 (potendosi sempre effettuare un opportuno cambiamento di variabile casuale che ci porti in questa situazione). Riassumendo: siano

$$\left\{ \begin{array}{l} x \sim N(x; \mu, \sigma) \\ H_0 \equiv \{\mu = 0\} \\ H_a \equiv \{\mu = 1\} \end{array} \right. \quad (\text{con } \sigma > 0 \text{ noto})$$

le nostre ipotesi.

La densità di probabilità della x vale

$$N(x; \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2}$$

e, quindi, la funzione di verosimiglianza ed il suo logaritmo valgono

$$\mathcal{L}(\mathbf{x}; \mu, \sigma) = \frac{1}{(\sigma \sqrt{2\pi})^N} \prod_{i=1}^N e^{-\frac{1}{2\sigma^2} (x_i - \mu)^2}$$

e, rispettivamente,

$$\begin{aligned} \ln \mathcal{L}(\mathbf{x}; \mu, \sigma) &= -N \ln(\sigma \sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \\ &= -N \ln(\sigma \sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_{i=1}^N (x_i^2 - 2\mu x_i + \mu^2) ; \end{aligned}$$

per cui

$$\ln \mathcal{L}(\mathbf{x}; \mu, \sigma) = -N \ln(\sigma \sqrt{2\pi}) - \frac{1}{2\sigma^2} \left(\sum_{i=1}^N x_i^2 - 2N\bar{x}\mu + N\mu^2 \right) . \quad (13.1)$$

Dalla (13.1) si ricava immediatamente

$$\ln \lambda = \ln \mathcal{L}(\mathbf{x}; 0, \sigma) - \ln \mathcal{L}(\mathbf{x}; 1, \sigma) = \frac{N}{2\sigma^2} (1 - 2\bar{x})$$

e la regione di rigetto \mathcal{R}_k è definita dalla

$$\mathcal{R}_k \equiv \left\{ \ln \lambda = \frac{N}{2\sigma^2} (1 - 2\bar{x}) < \ln k \right\}$$

da cui consegue, con facili passaggi,

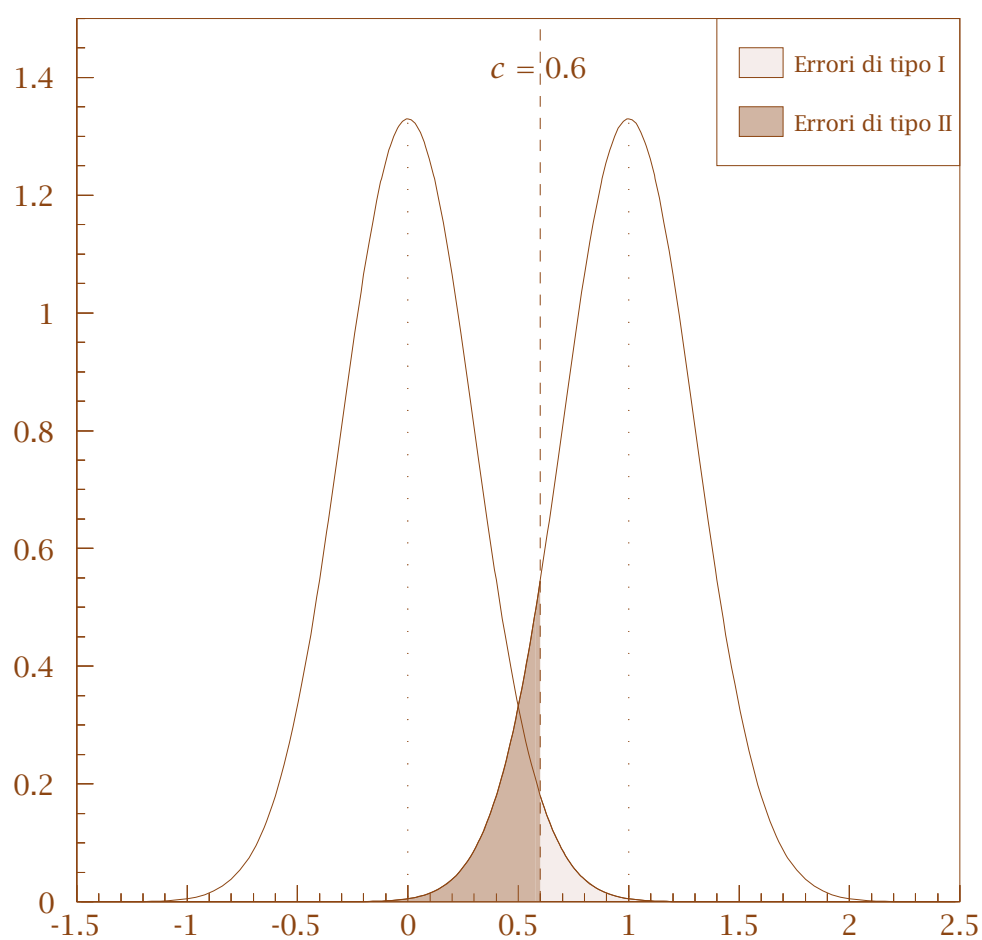
$$\mathcal{R}_k \equiv \left\{ \bar{x} > \frac{N - 2\sigma^2 \ln k}{2N} = c \right\} ;$$

ed insomma H_0 va rigettata se la media aritmetica del campione \bar{x} risulta superiore a c ; ed accettata altrimenti.

Come si può scegliere un valore opportuno di k (e quindi di c)? Gli errori di prima specie (si faccia riferimento anche alla figura 13a) hanno probabilità

$$P_I = 1 - \alpha = \Pr(\bar{x} > c | H_0) = \int_c^{+\infty} N\left(x; 0, \frac{\sigma}{\sqrt{N}}\right) dx \quad (13.2)$$

FIGURA 13a - L'esempio del paragrafo 13.1, con delineate (in corrispondenza ad un particolare valore di c) le probabilità degli errori di prima e seconda specie; le due curve sono $N(0, \sigma/\sqrt{N})$ e $N(1, \sigma/\sqrt{N})$.



e quelli di seconda specie

$$P_{II} = 1 - \beta = \Pr(\bar{x} < c | H_a) = \int_{-\infty}^c N\left(x; 1, \frac{\sigma}{\sqrt{N}}\right) dx \quad (13.3)$$

per cui si hanno svariate possibilità: ad esempio, se interessa contenere gli errori di prima specie e la dimensione del campione è nota, si fissa un valore opportunamente grande per α e dalla (13.2) si ricava c ; o, se interessa contenere gli errori di seconda specie e la dimensione del campione è nota, si fissa β e si ricava c dalla (13.3); o, infine, se si vogliono contenere gli errori di entrambi i tipi, si fissano sia α che β e si ricava la dimensione minima del campione necessaria per raggiungere lo scopo utilizzando entrambe le equazioni (13.2) e (13.3).

13.2 Il lemma di Neyman-Pearson

L'essere ricorsi per la definizione della regione di rigetto \mathcal{R} al calcolo del rapporto delle funzioni di verosimiglianza non è stato casuale; esiste infatti un teorema (il cosiddetto *lemma di Neyman-Pearson*) il quale afferma che

Se si ha a disposizione un campione di N valori indipendenti x_i da utilizzare per discriminare tra un'ipotesi nulla ed un'ipotesi alternativa entrambe semplici, e se è richiesto un livello fisso α di significanza, la massima potenza del test (ovvero la minima probabilità di errori di seconda specie) si raggiunge definendo la regione di rigetto \mathcal{R}_k attraverso una relazione del tipo

$$\mathcal{R}_k \equiv \left\{ \lambda = \frac{\mathcal{L}(\mathbf{x}|H_0)}{\mathcal{L}(\mathbf{x}|H_a)} < k \right\} . \quad (13.4)$$

Infatti, indicando con $f = f(x; \theta)$ la densità di probabilità della variabile x (che supponiamo dipenda da un solo parametro θ), siano $H_0 \equiv \{\theta = \theta_0\}$ e $H_a \equiv \{\theta = \theta_a\}$ le due ipotesi (semplici) tra cui decidere; la funzione di verosimiglianza vale, come sappiamo,

$$\mathcal{L}(\mathbf{x}; \theta) = \prod_{i=1}^N f(x_i; \theta) .$$

Indichiamo con \mathfrak{R} l'insieme di tutte le regioni \mathcal{R} per le quali risulti

$$P_I = \int_{\mathcal{R}} \mathcal{L}(\mathbf{x}; \theta_0) d\mathbf{x} = 1 - \alpha \quad (13.5)$$

con α costante prefissata (nella (13.5) abbiamo sinteticamente indicato con $d\mathbf{x}$ il prodotto $dx_1 dx_2 \cdots dx_N$). Vogliamo trovare quale di queste regioni rende massima

$$\beta = 1 - P_{II} = \int_{\mathcal{R}} \mathcal{L}(\mathbf{x}; \theta_a) d\mathbf{x} .$$

Ora, per una qualsiasi regione $\mathcal{R} \neq \mathcal{R}_k$, valgono sia la

$$\mathcal{R}_k = (\mathcal{R}_k \cap \mathcal{R}) \cup (\mathcal{R}_k \cap \overline{\mathcal{R}})$$

che la

$$\mathcal{R} = (\mathcal{R} \cap \mathcal{R}_k) \cup (\mathcal{R} \cap \overline{\mathcal{R}_k}) ;$$

e quindi, per una qualsiasi funzione $\phi(\mathbf{x})$, risulta sia

$$\int_{\mathcal{R}_k} \phi(\mathbf{x}) d\mathbf{x} = \int_{\mathcal{R}_k \cap \mathcal{R}} \phi(\mathbf{x}) d\mathbf{x} + \int_{\mathcal{R}_k \cap \overline{\mathcal{R}}} \phi(\mathbf{x}) d\mathbf{x}$$

che

$$\int_{\mathcal{R}} \phi(\mathbf{x}) d\mathbf{x} = \int_{\mathcal{R} \cap \mathcal{R}_k} \phi(\mathbf{x}) d\mathbf{x} + \int_{\mathcal{R} \cap \overline{\mathcal{R}_k}} \phi(\mathbf{x}) d\mathbf{x}$$

e, sottraendo membro a membro,

$$\int_{\mathcal{R}_k} \phi(\mathbf{x}) d\mathbf{x} - \int_{\mathcal{R}} \phi(\mathbf{x}) d\mathbf{x} = \int_{\mathcal{R}_k \cap \overline{\mathcal{R}}} \phi(\mathbf{x}) d\mathbf{x} - \int_{\mathcal{R} \cap \overline{\mathcal{R}_k}} \phi(\mathbf{x}) d\mathbf{x} . \quad (13.6)$$

Applicando la (13.6) alla funzione $\mathcal{L}(\mathbf{x}|\theta_a)$ otteniamo:

$$\begin{aligned} \int_{\mathcal{R}_k} \mathcal{L}(\mathbf{x}|\theta_a) d\mathbf{x} - \int_{\mathcal{R}} \mathcal{L}(\mathbf{x}|\theta_a) d\mathbf{x} &= \\ &= \int_{\mathcal{R}_k \cap \overline{\mathcal{R}}} \mathcal{L}(\mathbf{x}|\theta_a) d\mathbf{x} - \int_{\mathcal{R} \cap \overline{\mathcal{R}_k}} \mathcal{L}(\mathbf{x}|\theta_a) d\mathbf{x} ; \end{aligned} \quad (13.7)$$

ma, nel primo integrale del secondo membro, essendo la regione di integrazione contenuta in \mathcal{R}_k , deve valere la (13.4); e quindi risultare ovunque

$$\mathcal{L}(\mathbf{x}|\theta_a) > \frac{1}{k} \cdot \mathcal{L}(\mathbf{x}|\theta_0)$$

mentre, per lo stesso motivo, nel secondo integrale

$$\mathcal{L}(\mathbf{x}|\theta_a) \leq \frac{1}{k} \cdot \mathcal{L}(\mathbf{x}|\theta_0)$$

e quindi la (13.7) implica che

$$\begin{aligned} \int_{\mathcal{R}_k} \mathcal{L}(\mathbf{x}|\theta_a) d\mathbf{x} - \int_{\mathcal{R}} \mathcal{L}(\mathbf{x}|\theta_a) d\mathbf{x} &> \\ &> \frac{1}{k} \cdot \left[\int_{\mathcal{R}_k \cap \overline{\mathcal{R}}} \mathcal{L}(\mathbf{x}|\theta_0) d\mathbf{x} - \int_{\mathcal{R} \cap \overline{\mathcal{R}_k}} \mathcal{L}(\mathbf{x}|\theta_0) d\mathbf{x} \right]. \end{aligned}$$

Ricordando la (13.6),

$$\int_{\mathcal{R}_k} \mathcal{L}(\mathbf{x}|\theta_a) d\mathbf{x} - \int_{\mathcal{R}} \mathcal{L}(\mathbf{x}|\theta_a) d\mathbf{x} > \frac{1}{k} \cdot \left[\int_{\mathcal{R}_k} \mathcal{L}(\mathbf{x}|\theta_0) d\mathbf{x} - \int_{\mathcal{R}} \mathcal{L}(\mathbf{x}|\theta_0) d\mathbf{x} \right]$$

e, se $\mathcal{R} \in \mathfrak{R}$ e quindi soddisfa anch'essa alla (13.5),

$$\int_{\mathcal{R}_k} \mathcal{L}(\mathbf{x}|\theta_a) d\mathbf{x} - \int_{\mathcal{R}} \mathcal{L}(\mathbf{x}|\theta_a) d\mathbf{x} > \frac{1}{k} \cdot [P_I - P_I] = 0$$

che era la nostra tesi.

13.3 Tests di massima potenza uniforme

Consideriamo ora un esempio del tipo di quello del paragrafo 13.1; e sia sempre disponibile un campione di N misure indipendenti derivante da una popolazione normale di varianza nota. Assumiamo ancora come ipotesi nulla quella che la popolazione abbia un certo valore medio, che supponiamo essere 0, ma sostituiamo alla vecchia ipotesi alternativa H_a una nuova ipotesi *composta*; ovvero quella che il valore medio della popolazione sia positivo:

$$\begin{cases} x \sim N(x; \mu, \sigma) & (\text{con } \sigma > 0 \text{ noto}) \\ H_0 \equiv \{\mu = 0\} \\ H_a \equiv \{\mu > 0\} \end{cases}$$

(l'ipotesi alternativa è dunque somma logica di infinite ipotesi semplici del tipo $\mu = \mu_a$ con $\mu_a > 0$).

Dalla (13.1) ricaviamo immediatamente le

$$\mathcal{L}(\mathbf{x}; 0, \sigma) = -N \ln(\sigma \sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_{i=1}^N x_i^2$$

e

$$\mathcal{L}(\mathbf{x}; \mu_a, \sigma) = -N \ln(\sigma \sqrt{2\pi}) - \frac{1}{2\sigma^2} \left(\sum_{i=1}^N x_i^2 - 2N\bar{x}\mu_a + N\mu_a^2 \right)$$

(sempre con $\mu_a > 0$); e, sostituendole nella (13.4), che definisce la generica regione di rigetto \mathcal{R}_k , otteniamo

$$\ln \lambda = \ln \mathcal{L}(\mathbf{x}; 0, \sigma) - \mathcal{L}(\mathbf{x}; \mu_a, \sigma) = \frac{N\mu_a}{2\sigma^2} (\mu_a - 2\bar{x}) < \ln k$$

equivalente alla

$$\mathcal{R}_k \equiv \left\{ \bar{x} > \frac{N\mu_a^2 - 2\sigma^2 \ln k}{2N\mu_a} = c \right\}.$$

Si rigetta quindi H_0 se la media aritmetica del campione è superiore a c e la si accetta altrimenti: la probabilità di commettere errori di prima specie vale

$$P_I = 1 - \alpha = \int_c^{+\infty} N\left(x; 0, \frac{\sigma}{\sqrt{N}}\right) dx$$

ed è ben definita; ma, al contrario, la probabilità di commettere errori di seconda specie dipende dal particolare valore di μ_a , e non può quindi essere calcolata.

Se interessa solo contenere gli errori di prima specie e la dimensione del campione è nota, si fissa α e si ricava il corrispondente valore di c dall'equazione precedente; altrimenti occorre fare delle ulteriori ipotesi sulla funzione di frequenza dei differenti valori di μ_a , e, ad esempio, calcolare la probabilità degli errori di seconda specie con tecniche di Montecarlo.

In ogni caso, però, osserviamo che la regione di rigetto è sempre dello stesso tipo (13.4) per *qualsiasi* $\mu_a > 0$; e quindi un confronto separato tra H_0 ed ognuna delle differenti ipotesi semplici che costituiscono H_a è comunque del tipo per cui il lemma di Neyman-Pearson garantisce la massima potenza.

Tests di questo tipo, per i quali *la significanza è costante e la potenza è massima per ognuno dei casi semplici che costituiscono l'ipotesi alternativa*, si dicono “tests di massima potenza uniforme”.

13.4 Il rapporto delle massime verosimiglianze

Nel caso generale in cui sia l'ipotesi nulla che quella alternativa siano composte, la situazione è più complicata: non esiste normalmente un test di massima potenza uniforme, e, tra i vari criteri possibili per decidere tra le due ipotesi, bisogna capire quali abbiano caratteristiche (significanza e potenza) adeguate; un metodo adatto a costruire una regione di rigetto dotata asintoticamente (per grandi campioni) di caratteristiche, appunto, desiderabili, è quello seguente (*metodo del rapporto delle massime verosimiglianze*).

Sia una variabile casuale x , la cui densità di probabilità supponiamo sia una funzione $f(x; \theta_1, \theta_2, \dots, \theta_M)$ dipendente da M parametri: indicando sinteticamente la M -pla dei valori dei parametri come un vettore θ in uno spazio a M dimensioni (*spazio dei parametri*), consista H_0 nell'essere θ compreso all'interno di una certa regione Ω_0 di tale spazio; mentre H_a consista nell'appartenere θ alla regione Ω_a complementare a Ω_0 : $\Omega_a \equiv \bar{\Omega}_0$, così che $(\Omega_0 \cup \Omega_a)$ coincida con l'intero spazio dei parametri S .

In particolare, Ω_0 può estendersi, in alcune delle dimensioni dello spazio dei parametri, da $-\infty$ a $+\infty$; e, in tal caso, il vincolo sulle θ_i cui corrisponde l'ipotesi nulla riguarderà un numero di parametri minore di M .

Scritta la funzione di verosimiglianza,

$$\mathcal{L}(\mathbf{x}; \theta) = \prod_{i=1}^N f(x_i; \theta) \quad (13.8)$$

indichiamo con $\mathcal{L}(\hat{S})$ il suo massimo valore nell'intero spazio dei parametri; e con $\mathcal{L}(\hat{R})$ il massimo valore assunto sempre della (13.8), ma con i parametri vincolati a trovarsi nella regione Ω_0 (quindi limitatamente a quei casi nei quali H_0 è vera). Il rapporto

$$\lambda = \frac{\mathcal{L}(\hat{R})}{\mathcal{L}(\hat{S})} \quad (13.9)$$

deve essere un numero appartenente all'intervallo $[0, 1]$; se si fissa un arbitrario valore k ($0 < k < 1$), esso definisce una generica regione di rigetto, \mathcal{R}_k , attraverso la

$$\mathcal{R}_k \equiv \left\{ \lambda = \frac{\mathcal{L}(\hat{R})}{\mathcal{L}(\hat{S})} < k \right\}$$

(ovvero si accetta H_0 quando $\lambda \geq k$ e la si rigetta quando $\lambda < k$). Nel caso si sappia determinare la densità di probabilità di λ condizionata all'assunzione che H_0 sia vera, $g(\lambda|H_0)$, la probabilità di un errore di prima specie è data ovviamente da

$$P_I = \alpha = \Pr(\lambda \in [0, k]|H_0) = \int_0^k g(\lambda|H_0) d\lambda .$$

L'importanza del metodo sta nel fatto che si può dimostrare il seguente

TEOREMA: *se l'ipotesi nulla H_0 consiste nell'appartenenza di un insieme di $P \leq M$ dei parametri θ_i ad una determinata regione Ω_0 , e se l'ipotesi alternativa H_a consiste nel fatto che essi non vi appartengano ($H_a \equiv \bar{\Omega}_0$), allora $-2 \ln \lambda$, ove λ è definito dalla (13.9), ha densità di probabilità che, ammessa vera l'ipotesi nulla, converge in probabilità (all'aumentare di N) alla distribuzione del χ^2 a P gradi di libertà.*

che, dicendoci quale è (almeno nel limite di grandi campioni) la forma di $g(\lambda|H_0)$, ci mette comunque in grado di calcolare la significanza del test.

Illustriamo il metodo con un esempio: disponendo ancora di un campione di N determinazioni indipendenti, provenienti da una popolazione normale di varianza nota, vogliamo applicarlo per discriminare tra l'ipotesi nulla che il valore medio abbia valore 0 ($H_0 \equiv \{\mu = 0\}$) e quella che esso abbia valore differente ($H_a \equiv \{\mu \neq 0\}$).

Il logaritmo della funzione di verosimiglianza è ancora dato dalla (13.1); e già sappiamo, dal paragrafo 11.3, che \mathcal{L} assume il suo massimo valore quando $\mu = \bar{x}$, per cui

$$\ln \mathcal{L}(\hat{S}) = -N \ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2} \left(\sum_{i=1}^N x_i^2 - N\bar{x}^2 \right) .$$

Inoltre Ω_0 si riduce ad un unico punto, $\mu = 0$; per cui

$$\ln \mathcal{L}(\hat{R}) = -N \ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_{i=1}^N x_i^2 .$$

Dalla (13.9) si ricava

$$\ln \lambda = \ln \mathcal{L}(\hat{R}) - \ln \mathcal{L}(\hat{S}) = -\frac{1}{2\sigma^2} N\bar{x}^2$$

e la regione di rigetto è definita dalla $\ln \lambda < \ln k$; ovvero (ricordando che $\ln k < 0$) da

$$\mathcal{R}_k \equiv \left\{ \bar{x}^2 > -\frac{2\sigma^2 \ln k}{N} \right\}$$

e, posto

$$c = \sigma \sqrt{-\frac{2 \ln k}{N}}$$

si accetterà H_0 se $|\bar{x}| \leq c$ (e la si rigetterà se $|\bar{x}| > c$).

In questo caso il teorema precedentemente citato afferma che

$$-2 \ln \lambda = \frac{\bar{x}^2}{\frac{\sigma^2}{N}}$$

è distribuito asintoticamente come il χ^2 ad un grado di libertà (cosa che del resto già sapevamo, vista l'espressione di $-2 \ln \lambda$); per cui, indicando

con $F(t; N)$ la densità di probabilità della distribuzione del χ^2 a N gradi di libertà, avremo

$$P_I = \alpha = \int_0^k g(\lambda | H_0) d\lambda = \int_{-2 \ln k}^{+\infty} F(t; 1) dt$$

della quale ci possiamo servire per ricavare k se vogliamo che la significanza del test abbia un certo valore: ad esempio un livello di confidenza del 95% corrisponde ad $\alpha = 0.05$ e, dalle tabelle della distribuzione del χ^2 , ricaviamo

$$-2 \ln k = 3.84 \quad \text{e quindi} \quad c = 1.96 \frac{\sigma}{\sqrt{N}} .$$

Anche senza dover ricorrere al teorema sul comportamento asintotico di $-2 \ln \lambda$, allo stesso risultato si può pervenire per altra via: in questo caso si conosce infatti esattamente α , che vale

$$P_I = \alpha = \Pr(|\bar{x}| > c | H_0) = 2 \int_c^{+\infty} N\left(t; 0, \frac{\sigma}{\sqrt{N}}\right) dt$$

e, dalle tabelle della distribuzione normale standardizzata, si ricava che un'area two-tailed del 5% corrisponde ad un valore assoluto dello scarto normalizzato $t_0 = 1.96$; per cui, ancora, si ricaverebbe $|\bar{x}| > 1.96(\sigma/\sqrt{N})$ come test per un livello di confidenza del 95%.

13.5 Applicazione: ipotesi sulle probabilità

Nel paragrafo 11.5 abbiamo preso in considerazione il caso di un evento casuale che si può manifestare in un numero finito M di modalità, aventi ognuna probabilità incognita p_i ; la stima di massima verosimiglianza delle p_i è data dal rapporto tra la frequenza assoluta di ogni modalità, n_i , ed il numero totale di prove, N .

Vogliamo ora applicare il metodo del rapporto delle massime verosimiglianze per discriminare, sulla base di un campione di determinazioni indipendenti, l'ipotesi nulla che le probabilità abbiano valori noti a priori e l'ipotesi alternativa complementare, $H_a \equiv \bar{H}_0$:

$$\begin{cases} H_0 \equiv \{p_i = \pi_i\} & (\forall i \in \{1, 2, \dots, M\}) \\ H_a \equiv \{p_i \neq \pi_i\} & (\exists i \in \{1, 2, \dots, M\}) \end{cases}$$

Ricordiamo che la funzione di verosimiglianza, a meno di un fattore moltiplicativo costante, è data da

$$\mathcal{L}(\mathbf{n}; \mathbf{p}) = \prod_{i=1}^M p_i^{n_i}$$

e che, essendo la stima di massima verosimiglianza data da

$$\hat{p}_i = \frac{n_i}{N}$$

il massimo assoluto di \mathcal{L} è

$$\mathcal{L}(\hat{S}) = \prod_{i=1}^M \left(\frac{n_i}{N} \right)^{n_i} = \frac{1}{N^N} \prod_{i=1}^M n_i^{n_i} .$$

Inoltre, nell'unico punto dello spazio dei parametri che corrisponde ad H_0 ,

$$\mathcal{L}(\hat{R}) = \prod_{i=1}^M \pi_i^{n_i}$$

per cui

$$\lambda = \frac{\mathcal{L}(\hat{R})}{\mathcal{L}(\hat{S})} = N^N \prod_{i=1}^M \left(\frac{\pi_i}{n_i} \right)^{n_i}$$

dalla quale si può, come sappiamo, derivare una generica regione di rigetto attraverso la consueta $\mathcal{R}_k \equiv \{\lambda < k\}$.

$$-2 \ln \lambda = -2 \left[N \ln N + \sum_{i=1}^M n_i (\ln \pi_i - \ln n_i) \right]$$

è inoltre asintoticamente distribuita come il χ^2 a $M - 1$ gradi di libertà (c'è un vincolo: che le n_i abbiano somma N), e questo può servire a scegliere un k opportuno (nota la dimensione del campione) una volta fissata α .

Il criterio di verifica dell'ipotesi dato in precedenza consisteva nel calcolo del valore della variabile casuale

$$X = \sum_{i=1}^M \frac{(n_i - N\pi_i)^2}{N\pi_i}$$

e nel suo successivo confronto con la distribuzione del χ^2 a $M - 1$ gradi di libertà; lo studio del rapporto delle massime verosimiglianze porta dunque ad un criterio *differente* e, senza sapere nulla della probabilità di commettere errori di seconda specie, non è possibile dire quale dei due risulti migliore (a parità di significanza).

13.6 Applicazione: valore medio di una popolazione normale

Ancora un esempio: sia una popolazione normale $N(x; \mu, \sigma)$ dalla quale vengano ottenuti N valori indipendenti x_i , ma questa volta *la varianza σ sia ignota*; vogliamo discriminare, sulla base del campione, tra l'ipotesi nulla che il valore medio della popolazione abbia un valore prefissato e l'ipotesi alternativa complementare,

$$\begin{cases} H_0 \equiv \{\mu = \mu_0\} \\ H_a \equiv \{\mu \neq \mu_0\} \end{cases}$$

Il logaritmo della funzione di verosimiglianza è

$$\ln \mathcal{L}(\mathbf{x}; \mu, \sigma) = -N \ln \sigma - \frac{N}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \quad (13.10)$$

ed essendo le stime di massima verosimiglianza date, come avevamo trovato nel paragrafo 11.5, da

$$\hat{\mu} = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{e} \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2$$

ne deriva, sostituendo nella (13.10), che

$$\ln \mathcal{L}(\hat{S}) = -\frac{N}{2} \ln \left[\sum_{i=1}^N (x_i - \bar{x})^2 \right] + \frac{N}{2} \ln N - \frac{N}{2} \ln(2\pi) - \frac{N}{2} .$$

D'altra parte, ammessa vera H_0 , abbiamo che

$$\ln \mathcal{L}(\mathbf{x}|H_0) = -N \ln \sigma - \frac{N}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu_0)^2$$

e, derivando rispetto a σ ,

$$\frac{d}{d\sigma} \ln \mathcal{L}(\mathbf{x}|H_0) = -\frac{N}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^N (x_i - \mu_0)^2 .$$

Annullando la derivata prima, si trova che l'unico estremante di $\mathcal{L}(\mathbf{x}|H_0)$ si ha per

$$\sigma_0 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_0)^2$$

mentre la derivata seconda vale

$$\frac{d^2}{d\sigma^2} \ln \mathcal{L}(\mathbf{x}|H_0) = \frac{N}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^N (x_i - \mu_0)^2$$

e, calcolata per $\sigma = \sigma_0$,

$$\left. \frac{d^2(\ln \mathcal{L})}{d\sigma^2} \right|_{\sigma=\sigma_0} = -\frac{2N^2}{\sum_i (x_i - \mu_0)^2} < 0$$

per cui l'estremante è effettivamente un massimo. Sostituendo,

$$\begin{aligned} \ln \mathcal{L}(\hat{R}) &= -\frac{N}{2} \ln \left[\sum_{i=1}^N (x_i - \mu_0)^2 \right] + \frac{N}{2} \ln N - \frac{N}{2} \ln(2\pi) - \frac{N}{2} \\ \ln \lambda &= \ln \mathcal{L}(\hat{R}) - \ln \mathcal{L}(\hat{S}) = -\frac{N}{2} \left\{ \ln \left[\sum_{i=1}^N (x_i - \mu_0)^2 \right] - \ln \left[\sum_{i=1}^N (x_i - \bar{x})^2 \right] \right\} \end{aligned}$$

ed infine

$$\begin{aligned} \ln \lambda &= -\frac{N}{2} \ln \left[\frac{\sum_i (x_i - \mu_0)^2}{\sum_i (x_i - \bar{x})^2} \right] \\ &= -\frac{N}{2} \ln \left[1 + \frac{N(\bar{x} - \mu_0)^2}{\sum_i (x_i - \bar{x})^2} \right] \\ &= -\frac{N}{2} \ln \left(1 + \frac{t^2}{N-1} \right) \end{aligned}$$

tenendo conto dapprima del fatto che $\sum_i (x_i - \mu_0)^2 = \sum_i (x_i - \bar{x})^2 + N(\bar{x} - \mu_0)^2$, e definendo poi una nuova variabile casuale

$$t = (\bar{x} - \mu_0) \sqrt{\frac{N(N-1)}{\sum_i (x_i - \bar{x})^2}} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{N}}} .$$

Un qualunque metodo per il rigetto di H_0 definito confrontando λ con un prefissato valore k si traduce, in sostanza, in un corrispondente confronto da eseguire per t :

$$\mathcal{R}_k \equiv \{\ln \lambda < \ln k\}$$

che porta alla

$$-\frac{N}{2} \ln \left(1 + \frac{t^2}{N-1} \right) < \ln k$$

ed alla condizione

$$t^2 > (N-1) \left(k^{-\frac{2}{N}} - 1 \right) ;$$

ovvero si rigetta l'ipotesi nulla se $|t|$ è maggiore di un certo t_0 (derivabile dall'equazione precedente), e la si accetta altrimenti.

Ma t (vedi anche l'equazione (12.17)) segue la distribuzione di Student a $N-1$ gradi di libertà, e quindi accettare o rigettare H_0 sotto queste ipotesi si riduce ad un test relativo a quella distribuzione: come già si era concluso nel capitolo 12. Il livello di significanza α è legato a t_0 dalla

$$\frac{\alpha}{2} = \int_{t_0}^{+\infty} F(t; N-1) dt$$

(indicando con $F(t; N)$ la funzione di frequenza di Student a N gradi di libertà), tenendo conto che abbiamo a che fare con un two-tailed test ($\mathcal{R}_k \equiv \{|t| > t_0\}$).

Insomma non c'è differenza, in questo caso, tra quanto esposto nel capitolo precedente e la teoria generale discussa in quello presente: nel senso che i due criteri di verifica dell'ipotesi portano per questo problema allo stesso metodo di decisione (ma, come abbiamo visto nel paragrafo precedente, non è sempre così).

Appendice A

Cenni di calcolo combinatorio

Il *calcolo combinatorio* è una branca della matematica orientata alla discussione ed allo sviluppo di formule che permettano di ottenere il numero di casi distinti che si possono presentare in un esperimento, od il numero di elementi che compongono un insieme, senza ricorrere alla loro enumerazione esplicita.

Il calcolo combinatorio trova importanti applicazioni nella teoria della probabilità e nella statistica: alcune formule, specificatamente quelle per le permutazioni e le combinazioni, vengono usate nel corso del testo; qui se ne dà una breve giustificazione.

A.1 Il lemma fondamentale del calcolo combinatorio

LEMMA FONDAMENTALE DEL CALCOLO COMBINATORIO: *dati due insiemi I_1 ed I_2 , composti da N_1 ed N_2 elementi distinti rispettivamente, l'insieme $I = I_1 \otimes I_2$ di tutte le coppie ordinate che si possono costruire associando un elemento di I_1 con un elemento di I_2 è composto da $N_1 \cdot N_2$ elementi.*

Questo lemma si può immediatamente generalizzare (per induzione completa) a K insiemi I_1, \dots, I_K composti da N_1, \dots, N_K elementi distinti rispettivamente: l'insieme $I = I_1 \otimes I_2 \otimes \dots \otimes I_K$, costituito da tutte le possibili associazioni ordinate di K elementi ognuno dei quali provenga da un differente insieme I_j , con $j = 1, \dots, K$, è composto da $N_1 \cdot N_2 \cdot \dots \cdot N_K$ elementi.

A.2 Fattoriale di un numero intero

Si definisce come *fattoriale* di un numero intero positivo N , e si indica con il simbolo $N!$, il prodotto dei primi N numeri interi:

$$N! = 1 \cdot 2 \cdot 3 \cdots N ;$$

per motivi che appariranno chiari più avanti¹, si definisce poi il fattoriale di zero come $0! = 1$.

A.3 Disposizioni

Se N e K sono due numeri interi positivi tali che sia $K \leq N$, si definisce come numero delle disposizioni di N oggetti di classe K (che si indica con il simbolo D_K^N) il numero dei gruppi *distinti* di K oggetti che è possibile formare a partire dagli N originali; definendo come *distinti* due gruppi se essi differiscono o per qualche elemento o per l'ordine.

Come esempio, le disposizioni di classe 2 che si possono formare con le 21 lettere dell'alfabeto italiano sono le seguenti:

$$\left\{ \begin{array}{cccccc} AB & AC & AD & \cdots & AV & AZ \\ BA & & BC & BD & \cdots & BV & BZ \\ & & & & \cdots & & \\ ZA & ZB & ZC & ZD & \cdots & ZV & \end{array} \right.$$

Il valore di D_K^N si può facilmente trovare sfruttando il lemma fondamentale del calcolo combinatorio: il primo elemento di una disposizione si può infatti scegliere in N modi distinti, il secondo in $N - 1$, e così via. Di conseguenza D_K^N è il prodotto di K numeri interi decrescenti a partire da N :

$$D_K^N = N \cdot (N - 1) \cdot (N - 2) \cdots (N - K + 1) = \frac{N!}{(N - K)!} \quad (\text{A.1})$$

(nel caso dell'esempio fatto, le disposizioni sono $D_2^{21} = 21 \cdot 20 = 420$; nella tabella in cui sono state elencate vi sono 21 righe di 20 elementi ciascuna).

L'espressione (A.1) è verificata anche se $K = N$, però purché (come prima detto) si ponga $0! = 1$.

¹La "definizione" $0! = 1$ non è così arbitraria come può sembrare: in realtà si comincia definendo una certa funzione di variabile complessa $\Gamma(z)$ che, quando l'argomento z è un numero intero positivo, coincide con il suo fattoriale; e per la quale si vede che $\Gamma(0) = 1$.

A.4 Permutazioni

Se N è un numero intero positivo, si definisce come numero delle permutazioni di N oggetti, e si indica con P_N , il numero di maniere distinte in cui si possono ordinare gli N oggetti stessi. Evidentemente risulta

$$P_N \equiv D_N^N = N! .$$

A.5 Permutazioni con ripetizione

Se gli N oggetti che si hanno a disposizione sono tali da poter essere divisi in M gruppi (composti da N_1, N_2, \dots, N_M oggetti rispettivamente; ovviamente $N_1 + N_2 + \dots + N_M = N$), tali che gli oggetti in ognuno di questi gruppi siano *indistinguibili* tra loro, il numero di permutazioni che con essi si possono realizzare è inferiore a P_N ; più precisamente, visto che gli oggetti di ogni gruppo si possono scambiare tra loro in qualsiasi modo senza per questo dare luogo a una sequenza distinta, il numero di *permutazioni con ripetizione* è dato da

$$\frac{N!}{N_1! \cdot N_2! \cdot \dots \cdot N_M!} . \quad (\text{A.2})$$

A.6 Combinazioni

Se N e K sono due numeri interi positivi tali che sia $K \leq N$, si definisce come numero delle combinazioni di classe K di N oggetti il numero dei sottoinsiemi *distinti* composti da K oggetti che è possibile formare a partire dagli N originali; definendo come *distinti* due sottoinsiemi se essi differiscono per qualche elemento. Il numero delle combinazioni di classe K di N oggetti si indica con uno dei due simboli

$$C_K^N \quad \text{o} \quad \binom{N}{K}$$

(l'ultimo dei quali si chiama *coefficiente binomiale*).

Consideriamo l'insieme composto da tutte le disposizioni di classe K di N oggetti, e pensiamo di raggruppare i suoi elementi in sottoinsiemi in modo

che ciascuno di essi contenga tutte e sole quelle disposizioni che differiscano esclusivamente per l'ordine ma siano composte dagli stessi oggetti; ovviamente il numero di questi sottoinsiemi è C_K^N : ed ognuno di essi contiene un numero di elementi che è P_K .

Da qui ricaviamo

$$C_K^N \equiv \binom{N}{K} = \frac{D_K^N}{P_K} = \frac{N \cdot (N-1) \cdots (N-K+1)}{K \cdot (K-1) \cdots 1} = \frac{N!}{K! (N-K)!} \quad (\text{A.3})$$

O, in altre parole, il numero di combinazioni di classe K di N oggetti è uguale al rapporto tra il prodotto di K numeri interi decrescenti a partire da N ed il prodotto di K numeri interi crescenti a partire dall'unità.

Si dimostrano poi facilmente, a partire dalla definizione, due importanti proprietà dei coefficienti binomiali:

$$\binom{N}{K} = \binom{N}{N-K}$$

e

$$\binom{N+1}{K} = \binom{N}{K-1} + \binom{N}{K}.$$

È da osservare che, così come sono stati ricavati (dalla definizione delle possibili combinazioni di N oggetti), i coefficienti binomiali hanno senso solo se N e K sono numeri interi; ed inoltre se risulta sia $N > 0$ che $0 \leq K \leq N$. La definizione (A.3) può comunque essere estesa a valori interi qualunque, ed anche a valori reali di N — ma questo esula dal nostro interesse.

A.7 Partizioni ordinate

Consideriamo un insieme di N oggetti; vogliamo calcolare il numero di maniere in cui essi possono essere divisi in M gruppi che siano composti da N_1, N_2, \dots, N_M oggetti rispettivamente (essendo $N_1 + N_2 + \cdots + N_M = N$).

Gli N_1 oggetti che compongono il primo gruppo possono essere scelti in $C_{N_1}^N$ modi differenti; quelli del secondo gruppo in $C_{N_2}^{N-N_1}$ modi; e così via. Per il lemma fondamentale del calcolo combinatorio, il numero delle *partizioni*

ordinate deve essere uguale a

$$\begin{aligned}
 & \binom{N}{N_1} \binom{N-N_1}{N_2} \binom{N-N_1-N_2}{N_3} \cdots \binom{N-N_1-\cdots-N_{M-1}}{N_M} = \\
 &= \frac{N!}{N_1! (N-N_1)!} \cdot \frac{(N-N_1)!}{N_2! (N-N_1-N_2)!} \cdots \frac{(N-N_1-\cdots-N_{M-1})!}{N_M! (N-N_1-\cdots-N_M)!} = \\
 &= \frac{N!}{N_1! N_2! \cdots N_M!}
 \end{aligned}$$

(sfruttando il fatto che tutti i numeratori dei termini dal secondo in poi si semplificano con uno dei fattori del denominatore del termine precedente; inoltre, nell'ultimo termine, $N-N_1-\cdots-N_M \equiv 0$). Si può notare che l'ultimo termine della prima espressione, essendo $N-N_1-\cdots-N_{M-1} = N_M$, vale sempre uno; cosa non sorprendente visto che, quando i primi $M-1$ gruppi sono stati scelti, anche l'ultimo risulta univocamente determinato.

Insomma il numero delle partizioni ordinate è uguale al numero delle permutazioni con ripetizione di N oggetti raggruppabili in M insiemi, composti rispettivamente da N_1, N_2, \dots, N_M oggetti indistinguibili tra loro, dato dalla formula (A.2)

Appendice B

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 sperimentalmente 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'adequatezza 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, perché 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 di pagina 52 ed i ragionamenti conseguenti. La varianza di una qualsiasi variabile casuale x , indicata di seguito come $\text{Var}(x)$, si può scrivere come

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

e, usando questa formula per calcolare la varianza della varianza di un campione di N misure s^2 , avremo

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

Ora

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

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

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

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

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

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

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

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

Il valore 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,

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

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

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

e, riprendendo la formule di partenza,

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

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

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

inoltre

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

(essendo $x^* = 0$) e

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

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:

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

Questa relazione ha validità generale. *Nel caso poi che la popolazione ubbidisca alla legge normale*, potremo calcolare il valore medio di x^4 usando la forma analitica della funzione di Gauss: per distribuzioni normali qualsiasi, i momenti di ordine pari rispetto alla media sono dati dalla formula (8.5), che qui ricordiamo:

$$\mu_{2k} = E \{ [x - E(x)]^{2k} \} = \frac{(2k)!}{2^k k!} \mu_2^k = \frac{(2k)!}{2^k k!} \sigma^2 .$$

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

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

ed infine *l'errore quadratico medio della stima della varianza della popolazione* ricavata dal campione 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 valore 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

$$\sigma_\sigma \approx \frac{\sigma}{\sqrt{2(N-1)}} \quad (\text{B.1})$$

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

Una conseguenza dell'equazione (B.1) è che l'errore relativo di σ dipende *solo dal numero di misure*; diminuisce poi all'aumentare di esso, ma questa diminuzione è inversamente proporzionale alla radice quadrata di N e risulta perciò lenta.

In altre parole, per diminuire l'errore relativo di σ di un ordine di grandezza occorre aumentare il numero delle misure di *due* ordini di grandezza; σ_σ/σ è (circa) il 25% per 10 misure, il 7% per 100 misure ed il 2% per 1000 misure effettuate: e questo è sostanzialmente il motivo per cui, di norma, si scrive l'errore quadratico medio *dando per esso una sola cifra significativa*.

Due cifre significative *reali* per σ corrisponderebbero infatti ad un suo errore relativo compreso tra il 5% (se la prima cifra significativa di σ è 1, ad esempio $\sigma = 10 \pm 0.5$) e lo 0.5% ($\sigma = 99 \pm 0.5$); e presupporrebbero quindi che siano state effettuate almeno 200 misure nel caso più favorevole e quasi 20'000 in quello più sfavorevole.

Appendice C

Covarianza e correlazione

C.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:

$$\begin{aligned}\text{Cov}(x, y) &= E\{[x - E(x)][y - E(y)]\} \\ &= E(xy) - E(x) \cdot E(y) \quad .\end{aligned}$$

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

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

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

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.

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

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

Non è però vero l'inverso: consideriamo ad esempio le due variabili casuali x ed $y = x^2$, ovviamente dipendenti l'una dall'altra: la loro covarianza vale

$$\text{Cov}(x, y) = E(xy) - E(x) \cdot E(y) = E(x^3) - E(x) \cdot E(x^2)$$

ed è chiaramente nulla per qualunque variabile casuale x con distribuzione simmetrica rispetto allo zero; quindi l'annullarsi della covarianza è condizione *necessaria ma non sufficiente* per l'indipendenza statistica di due variabili casuali.

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:

$$E(z) = a E(x) + b E(y) = 0$$

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

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

ed infine

$$\text{Var}(z) = a^2 \text{Var}(x) + b^2 \text{Var}(y) + 2ab \text{Cov}(x, y) . \quad (\text{C.1})$$

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 vale la

$$E(\xi) = E(\eta) = 0$$

con le

$$\text{Var}(\xi) = \text{Var}(x) \quad \text{e} \quad \text{Var}(\eta) = \text{Var}(y) ;$$

basta osservare infatti che vale anche la

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

La (C.1) si può poi generalizzare, per induzione completa, ad una variabile z definita come combinazione lineare di un numero qualsiasi N di variabili casuali: si trova che, se

$$z = \sum_{i=1}^N a_i x_i$$

risulta

$$\text{Var}(z) = \sum_i a_i^2 \text{Var}(x_i) + \sum_{\substack{i,j \\ j>i}} 2 a_i a_j \text{Cov}(x_i, x_j) . \quad (\text{C.2})$$

Per esprimere in modo compatto la (C.2), si ricorre in genere ad una notazione che usa la cosiddetta *matrice delle covarianze* delle variabili x ; ovvero sia una matrice quadrata V di ordine N , in cui il generico elemento V_{ij} è uguale alla covarianza delle variabili casuali x_i e x_j :

$$V_{ij} = \text{Cov}(x_i, x_j) = E(x_i \cdot x_j) - E(x_i) \cdot E(x_j) . \quad (\text{C.3})$$

La matrice è ovviamente *simmetrica* ($V_{ij} = V_{ji}$); e, in particolare, gli elementi diagonali V_{ii} valgono

$$V_{ii} = E(x_i^2) - [E(x_i)]^2 \equiv \text{Var}(x_i) .$$

Consideriamo poi le a_i come le N componenti di un *vettore* A di dimensione N (che possiamo concepire come una matrice rettangolare di N righe ed una colonna); ed introduciamo la matrice *trasposta* di A , \tilde{A} , che è una matrice rettangolare di una riga ed N colonne i cui elementi valgono

$$\tilde{A}_i = A_i .$$

Possiamo allora scrivere la (C.2) nella forma

$$\text{Var}(z) = \sum_{i,j} \tilde{A}_i V_{ij} A_j$$

(la simmetria di \mathbf{V} e quella tra $\tilde{\mathbf{A}}$ ed \mathbf{A} produce, nello sviluppo delle somme, il fattore 2 che moltiplica le covarianze); o anche, ricordando le regole del prodotto tra matrici,

$$\text{Var}(z) = \tilde{\mathbf{A}} \mathbf{V} \mathbf{A}$$

Si può poi facilmente dimostrare il seguente teorema, che ci sarà utile più avanti:

TEOREMA: *due differenti combinazioni lineari delle stesse variabili casuali sono sempre correlate.*

Infatti, dette A e B le due combinazioni lineari:

$$\begin{aligned} A &= \sum_{i=1}^N a_i x_i & \Rightarrow & E(A) = \sum_{i=1}^N a_i E(x_i) \\ B &= \sum_{j=1}^N b_j x_j & \Rightarrow & E(B) = \sum_{j=1}^N b_j E(x_j) \end{aligned}$$

abbiamo che la covarianza di A e B vale

$$\begin{aligned} \text{Cov}(A, B) &= E \left\{ [A - E(A)] [B - E(B)] \right\} \\ &= E \left\{ \sum_{i,j} a_i b_j [x_i - E(x_i)] [x_j - E(x_j)] \right\} \\ &= \sum_{i,j} a_i b_j E \left\{ [x_i - E(x_i)] [x_j - E(x_j)] \right\} \\ &= \sum_i a_i b_i \text{Var}(x_i) + \sum_{\substack{i,j \\ i \neq j}} a_i b_j \text{Cov}(x_i, x_j) \end{aligned}$$

e non è in genere nulla. In forma matriciale e con ovvio significato dei simboli,

$$\text{Cov}(A, B) = \tilde{\mathbf{A}} \mathbf{V} \mathbf{B} .$$

È da notare come A e B siano di norma sempre correlate *anche se le variabili di partenza x_i sono tutte tra loro statisticamente indipendenti*: in questo caso infatti tutti i termini non diagonali della matrice delle covarianze si annullano, e risulta

$$\text{Cov}(A, B) = \sum_{i=1}^N a_i b_i \text{Var}(x_i) . \quad (\text{C.4})$$

C.2 La correlazione lineare

Per due variabili casuali qualunque si definisce poi il *coefficiente di correlazione lineare* $\text{Corr}(x, y)$ (anche indicato col simbolo r_{xy} , o semplicemente come r) nel modo seguente:

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

Il coefficiente di correlazione di due variabili è ovviamente adimensionale; è nullo quando le variabili stesse sono statisticamente indipendenti (visto che è zero la loro covarianza); ed è comunque compreso tra i due limiti -1 e $+1$. Che valga quest'ultima proprietà si può dimostrare calcolando dapprima la varianza di una variabile casuale ausiliaria z definita attraverso la relazione $z = \sigma_y x - \sigma_x y$, 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

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

Poi, compiendo analoghi passaggi su un'altra variabile definita stavolta come $z = \sigma_y x + \sigma_x y$, si troverebbe che deve essere anche $\text{Corr}(x, y) \geq -1$.

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

$$z = \sigma_y x \mp \sigma_x y = \text{costante}$$

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

Vale anche l'inverso: partendo infatti dall'ipotesi che le due variabili siano legate da una relazione lineare data da $y = a + bx$, con b finito e non nullo, ne consegue che:

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

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 ($x = \text{costante}$ o $y = \text{costante}$) 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 $E(xy) \equiv E(x) \cdot E(y)$ 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 più propriamente “coefficiente di correlazione *lineare*”.

C.3 Propagazione degli errori per variabili correlate

Vediamo ora come si può ricavare una formula di propagazione per gli errori (da usare in luogo dell'equazione (10.2) che abbiamo incontrato a pagina 164) se le grandezze fisiche misurate direttamente non sono tra loro statisticamente indipendenti; nel corso di questo paragrafo continueremo ad usare la notazione già introdotta nel capitolo 10.

Consideriamo una funzione F di N variabili, $F = F(x_1, x_2, \dots, x_N)$; ed ammettiamo che sia lecito svilupparla in serie di Taylor nell'intorno del punto $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)$ trascurando i termini di ordine superiore al primo (questo avviene, come sappiamo, o se gli errori di misura sono piccoli o se F è lineare rispetto a tutte le variabili). Tenendo presente il teorema di pagina 52, ed applicando alla formula dello sviluppo

$$F(x_1, x_2, \dots, x_N) \approx F(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N) + \sum_{i=1}^N \frac{\partial F}{\partial x_i} (x_i - \bar{x}_i)$$

l'equazione (C.2), otteniamo

$$\text{Var}(F) \approx \sum_i \left(\frac{\partial F}{\partial x_i} \right)^2 \text{Var}(x_i) + 2 \sum_{\substack{i,j \\ j>i}} \frac{\partial F}{\partial x_i} \frac{\partial F}{\partial x_j} \text{Cov}(x_i, x_j) . \quad (\text{C.5})$$

Per esprimere in modo compatto la (C.5), si può ricorrere ancora alla matrice delle covarianze \mathbf{V} delle variabili x_i ; ricordandone la definizione (data dall'equazione (C.3) a pagina 257) ed introducendo poi un vettore \mathbf{F} di dimensione N di componenti

$$F_i = \frac{\partial F}{\partial x_i}$$

ed il suo trasposto $\tilde{\mathbf{F}}$, la (C.5) si può riscrivere nella forma

$$\text{Var}(F) = \sum_{i,j} \tilde{F}_i V_{ij} F_j$$

ossia

$$\text{Var}(F) = \tilde{\mathbf{F}} \mathbf{V} \mathbf{F}$$

C.4 Applicazioni all'interpolazione lineare

Riprendiamo adesso il problema dell'interpolazione lineare, già discusso nel capitolo 11: 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 11.4.1; 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.

C.4.1 Riscrittura delle equazioni dei minimi quadrati

Sebbene i valori della x siano scelti dallo sperimentatore e privi di errore, e non siano pertanto variabili casuali in senso stretto; 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

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

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

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

Queste grandezze permettono di riscrivere le equazioni (11.9) risolutive del problema dell'interpolazione lineare per un insieme di dati, che abbiamo già incontrato a pagina 181, nella forma

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

La prima equazione intanto implica che la retta interpolante deve passare per il punto (\bar{x}, \bar{y}) le cui coordinate sono le medie dei valori misurati delle due variabili in gioco; poi, ricavando da essa $a = \bar{y} - b\bar{x}$ e sostituendo nella seconda equazione, dopo aver semplificato alcuni termini si ottiene la soluzione per l'altra incognita:

$$b = \frac{\text{Cov}(x, y)}{\text{Var}(x)} \equiv \text{Corr}(x, y) \sqrt{\frac{\text{Var}(y)}{\text{Var}(x)}} \quad (\text{C.6})$$

e la retta interpolante ha quindi equazione

$$y = a + bx = (\bar{y} - b\bar{x}) + bx$$

o anche

$$(y - \bar{y}) = b(x - \bar{x})$$

(in cui b ha il valore (C.6)). Introduciamo ora le due variabili casuali ausiliarie $\xi = x - \bar{x}$ e $\eta = y - \bar{y}$, per le quali valgono le

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

(con le analoghe per η ed y), ed inoltre la

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

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

$$\hat{y}_i = a + bx_i = \bar{y} + b(x_i - \bar{x}) \quad (\text{C.7})$$

e con δ_i la differenza $\hat{y}_i - y_i$. Le differenze δ_i prendono il nome di *residui*, e di essi ci occuperemo ancora più avanti; risulta che

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

in cui r è il coefficiente di correlazione lineare calcolato usando, sempre solo formalmente, i campioni dei valori misurati delle x e delle 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 11.4.2):

$$\mu_y = \sqrt{\frac{\sum_i \delta_i^2}{N-2}} = \sqrt{\frac{N}{N-2} \text{Var}(y) (1-r^2)} \quad (\text{C.8})$$

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.

È 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.

C.4.2 Verifica di ipotesi sulla correlazione lineare

Non è pensabile di poter svolgere una teoria completa della correlazione in queste pagine; tuttavia talvolta un fisico si trova nella necessità di verificare delle ipotesi statistiche sul coefficiente di correlazione lineare ricavato sperimentalmente da un insieme di N coppie di valori misurati $\{x_i, y_i\}$. Nel seguito riassumiamo, senza darne alcuna dimostrazione, alcune proprietà di questi coefficienti:

- Se si vuole verificare che la correlazione tra x ed y sia significativamente differente da zero, si può calcolare il valore della variabile casuale

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

che è distribuita secondo Student con $(N-2)$ gradi di libertà, e controllare la sua compatibilità con lo zero.

- Rispetto al valore vero ρ della correlazione lineare tra le due variabili, il valore r ricavato da un insieme di N coppie $\{x_i, y_i\}$ estratte a caso dalle rispettive popolazioni è tale che la variabile casuale

$$Z(r) = \ln \sqrt{\frac{1+r}{1-r}} = \frac{1}{2} [\ln(1+r) - \ln(1-r)] \quad (\text{C.9})$$

(detta *variabile di Fisher*) segue una distribuzione approssimativamente normale con valore medio e varianza date da

$$E(Z) = Z(\rho) \quad \text{e} \quad \text{Var}(Z) = \sigma_Z^2 = \frac{1}{N-3}$$

rispettivamente; la trasformazione inversa della (C.9) è la

$$r = \frac{e^{2Z} - 1}{e^{2Z} + 1}.$$

Quindi:

- per verificare se il valore vero della correlazione può essere una quantità prefissata ρ , si controlla la compatibilità con la distribuzione normale $N(Z(\rho), \sigma_Z)$ del valore ottenuto $Z(r)$;
- per calcolare un intervallo di valori corrispondente ad un certo livello di confidenza, si usano i corrispondenti intervalli per la distribuzione normale con deviazione standard σ_Z ;
- per verificare se due coefficienti di correlazione lineare r_1 ed r_2 , ricavati da N_1 ed N_2 coppie di valori $\{x_i, y_i\}$ rispettivamente, siano o meno significativamente differenti, si calcola la variabile casuale

$$\delta = \frac{Z_1 - Z_2}{\sqrt{\text{Var}(Z_1) + \text{Var}(Z_2)}} = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}$$

(ove Z_1 e Z_2 sono le variabili di Fisher ricavate dalla (C.9) per i due campioni; δ segue asintoticamente la distribuzione normale con media $E(Z_1) - E(Z_2)$ e varianza 1) e si verifica se il risultato ottenuto è compatibile con lo zero.

C.4.3 La correlazione tra i coefficienti della retta

Notiamo anche che l'intercetta e la pendenza a e b della retta interpolante un insieme di punti sperimentali, essendo ottenute come combinazioni

lineari delle stesse variabili casuali, sono tra loro correlate (come osservato nel paragrafo C.1); le equazioni (11.10) possono essere riscritte (l'abbiamo già visto nel paragrafo 11.4.1) come

$$a = \sum_i a_i y_i \quad \text{e} \quad b = \sum_i b_i y_i$$

una volta che si sia posto

$$\begin{cases} a_i = \frac{1}{\Delta} \left[\sum_j x_j^2 - \left(\sum_j x_j \right) x_i \right] \\ b_i = \frac{1}{\Delta} \left[N x_i - \sum_j x_j \right] \end{cases}$$

Ricordando la definizione di Δ , possiamo esprimerlo come

$$\begin{aligned} \Delta &= N \left(\sum_j x_j^2 \right) - \left(\sum_j x_j \right)^2 \\ &= N^2 \left[\frac{\sum_j x_j^2}{N} - \left(\frac{\sum_j x_j}{N} \right)^2 \right] \end{aligned}$$

ed infine come

$$\Delta = N^2 \text{Var}(x) \quad . \quad (\text{C.10})$$

Visto che le y_i sono per ipotesi statisticamente indipendenti tra loro, possiamo applicare la (C.4); ne ricaviamo, sostituendovi la (C.10) e ricordando che gli errori sono tutti uguali, che

$$\begin{aligned} \text{Cov}(a, b) &= \sum_i a_i b_i \text{Var}(y_i) \\ &= \frac{\sigma_y^2}{\Delta^2} \sum_i \left[\sum_j x_j^2 - \left(\sum_j x_j \right) x_i \right] \left[N x_i - \sum_j x_j \right] \\ &= \frac{\sigma_y^2}{\Delta^2} \sum_i \left[\left(\sum_j x_j^2 \right) N x_i - \left(\sum_j x_j^2 \right) \left(\sum_j x_j \right) - \right. \\ &\quad \left. - \left(\sum_j x_j \right) N x_i^2 + \left(\sum_j x_j \right)^2 x_i \right] \\ &= \frac{\sigma_y^2}{\Delta^2} \left[N \left(\sum_j x_j^2 \right) \left(\sum_i x_i \right) - N \left(\sum_j x_j^2 \right) \left(\sum_j x_j \right) - \right. \\ &\quad \left. - N \left(\sum_j x_j \right) \left(\sum_i x_i^2 \right) + \left(\sum_j x_j \right)^3 \right] \\ &= \frac{\sigma_y^2}{\Delta^2} \left(\sum_j x_j \right) \left[\left(\sum_j x_j \right)^2 - N \left(\sum_j x_j^2 \right) \right] \end{aligned}$$

ed infine, vista la definizione di Δ ,

$$\text{Cov}(a, b) = - \frac{\sum_j x_j}{\Delta} \sigma_y^2 ; \quad (\text{C.11})$$

o anche

$$\text{Cov}(a, b) = - \frac{\bar{x}}{N} \frac{\sigma_y^2}{\text{Var}(x)} ,$$

diversa da zero se $\bar{x} \neq 0$; inoltre il segno della correlazione tra a e b è opposto a quello di \bar{x} . Questo non sorprende: la retta interpolante deve necessariamente passare per il punto (\bar{x}, \bar{y}) , come abbiamo notato nel paragrafo C.4.1: se \bar{x} è positivo, aumentando la pendenza della retta deve diminuire il valore della sua intercetta; e viceversa per $\bar{x} < 0$.

C.4.4 Stima puntuale mediante l'interpolazione lineare

Bisogna tener presente che le formule dei minimi quadrati ci danno l'equazione della retta che meglio approssima la relazione tra le due variabili nella parte di piano in cui esistono punti misurati; ma non ci dicono nulla sulla dipendenza tra le variabili stesse in zone in cui non siano state effettuate delle osservazioni.

In altre parole, non si può mai escludere sulla base delle misure che la $y = f(x)$ sia una funzione comunque complessa ma approssimativamente lineare *solamente* nell'intervallo in cui abbiamo investigato; per questo motivo bisogna evitare per quanto possibile di usare l'equazione della retta interpolante per ricavare valori stimati \hat{y} della variabile indipendente (dalla $\hat{y} = a + bx$) in corrispondenza di valori della x non compresi nell'intorno delle misure, e questo tanto più rigorosamente quanto più x è distante da tale intorno: *non è lecito usare l'interpolazione per estrapolare* su regioni lontane da quelle investigate.

A questo proposito, se lo scopo primario dell'interpolazione non è tanto quello di ottenere una stima di a o b quanto quello di ricavare il valore \hat{y} assunto dalla y in corrispondenza di un particolare valore \hat{x} della variabile indipendente x , applicando la formula di propagazione degli errori (C.5) alla $\hat{y} = a + b\hat{x}$ ricaviamo:

$$\text{Var}(\hat{y}) = \text{Var}(a) + \hat{x}^2 \text{Var}(b) + 2 \hat{x} \text{Cov}(a, b) .$$

Sostituendo nell'equazione precedente le espressioni (11.11) per le varianze di a e b e quella (C.11) della loro covarianza, si ha poi

$$\text{Var}(\hat{y}) = \frac{\sum_i x_i^2}{\Delta} \sigma_y^2 + \hat{x}^2 \frac{N}{\Delta} \sigma_y^2 - 2 \hat{x} \frac{\sum_i x_i}{\Delta} \sigma_y^2$$

e, introducendo nell'equazione precedente sia l'espressione (C.10) per Δ che la

$$\sum_i x_i^2 = N [\text{Var}(x) + \bar{x}^2]$$

e la

$$\sum_i x_i = N\bar{x}$$

otteniamo

$$\begin{aligned} \text{Var}(\hat{y}) &= \frac{\sigma_y^2}{\Delta} \left[\sum_i x_i^2 - 2\hat{x} \sum_i x_i + N\hat{x}^2 \right] \\ &= \frac{\sigma_y^2}{N \text{Var}(x)} \left[\text{Var}(x) + \bar{x}^2 - 2\hat{x}\bar{x} + \hat{x}^2 \right] \end{aligned}$$

ed infine la

$$\text{Var}(\hat{y}) = \frac{\sigma_y^2}{N} \left[1 + \frac{(\hat{x} - \bar{x})^2}{\text{Var}(x)} \right]. \quad (\text{C.12})$$

Si vede immediatamente dalla (C.12) che l'errore sul valore stimato \hat{y} (che è funzione di \hat{x}) è minimo quando $\hat{x} = \bar{x}$: quindi, per ricavare una stima della y con il più piccolo errore casuale possibile, bisogna che il corrispondente valore della x sia nel centro dell'intervallo in cui si sono effettuate le misure (questo in accordo con le considerazioni qualitative precedenti a riguardo di interpolazione ed estrapolazione).

C.4.5 Verifica di ipotesi nell'interpolazione lineare

Nel paragrafo 11.4.1 abbiamo ricavato le formule (11.10) dei minimi quadrati e le formule (11.11) per l'errore dei coefficienti della retta interpolata: queste ultime richiedono la conoscenza dell'errore comune sulle ordinate σ_y che, di consueto, viene ricavato a posteriori dai dati attraverso l'equazione (C.8).

Assai di frequente è necessario verificare delle ipotesi statistiche sui risultati dell'interpolazione lineare; una volta ricavata, ad esempio, la pendenza della retta interpolante, si può o voler confrontare la stima ottenuta con un valore noto a priori, o voler costruire attorno ad essa un intervallo corrispondente ad un certo livello di confidenza; o, ancora, si può voler effettuare il confronto (o calcolare l'ampiezza dell'intervallo di confidenza) per un valore $\hat{y} = a + b\hat{x}$ della y stimato sulla base dell'interpolazione, e la cui varianza è data dalla (C.12).

È naturale pensare di sfruttare per questo scopo le tabelle della distribuzione normale; ma questo implicitamente richiede che il numero di coppie di dati a disposizione *sia sufficientemente elevato* perché la stima di σ_y ottenuta a posteriori dai dati si possa considerare esatta. In realtà quando l'errore è ricavato a posteriori *tutte* le grandezze precedentemente citate *non* seguono la distribuzione normale *ma la distribuzione di Student* con $(N - 2)$ gradi di libertà.

C.4.6 Adeguatezza dell'interpolazione lineare o polinomiale in genere

Talvolta non si sa a priori che la relazione tra due variabili x ed y è di tipo lineare; ma, una volta eseguite le misure, si nota un loro approssimativo disporsi lungo una linea retta. In questo caso si può cercare di inferire una legge fisica dai dati sperimentali; ma come si può essere ragionevolmente sicuri che la relazione tra le due grandezze sia effettivamente lineare, anche limitandoci all'intervallo in cui sono distribuite le osservazioni?

Una possibilità è quella di eseguire interpolazioni successive con più curve polinomiali di grado M crescente, del tipo

$$y = P_M(x) = \sum_{k=0}^M a_k x^k \quad (C.13)$$

e di osservare l'andamento del *valore* dei residui in funzione di M : al crescere del grado del polinomio questi diminuiranno, dapprima in modo rapido per poi assestarsi su valori, sempre decrescenti, ma più o meno dello stesso ordine di grandezza; ed infine si annulleranno quando $M = N - 1$. Il valore di M che segna la transizione tra questi due comportamenti ci dà il grado della curva polinomiale che descrive in modo soddisfacente la relazione tra le variabili senza per questo seguire le fluttuazioni casuali di ogni singola misura.

Nel caso in cui ognuno degli N valori y_i ha errore noto σ_i (e le y_i non sono correlate), la somma dei quadrati dei residui pesati in maniera inversamente proporzionale ai quadrati degli errori,

$$S_M = \sum_{i=1}^N \frac{\delta_i^2}{\sigma_i^2} = \sum_{i=1}^N \frac{(y_i - \hat{y}_i)^2}{\sigma_i^2}$$

(le \hat{y}_i si intendono calcolate usando il polinomio interpolante $y = P_M(x)$ (C.13), i cui $M + 1$ coefficienti siano stati stimati dai dati) è distribuita come il χ^2 a $N - M - 1$ gradi di libertà; la probabilità di ottenere un determinato valore

di S_M sotto l'ipotesi $y = P_M(x)$ può dunque essere stimata dalle tabelle della distribuzione, e ci dà una misura statisticamente corretta dell'“accordo complessivo” tra i punti misurati ed un polinomio di grado M .

Per valutare numericamente la significatività della variazione di questo “accordo complessivo” dovuta ad un aumento di grado del polinomio interpolante, la regola di somma del χ^2 ci dice che la variazione della somma pesata dei quadrati dei residui, $S_{M-1} - S_M$, deve essere distribuita come il χ^2 ad un grado di libertà; normalmente, in luogo di usare direttamente $S_{M-1} - S_M$, si considera l'altra variabile casuale

$$F = \frac{S_{M-1} - S_M}{\frac{S_M}{N - M - 1}}$$

(che rappresenta una sorta di “variazione relativa dell'accordo complessivo”), e la si confronta con la funzione di frequenza di Fisher a 1 e $(N - M - 1)$ gradi di libertà; un valore di F elevato, e che con piccola probabilità potrebbe essere ottenuto da quella distribuzione, implicherà che l'aumento di grado è significativo.

C.4.7 Il *run test* per i residui

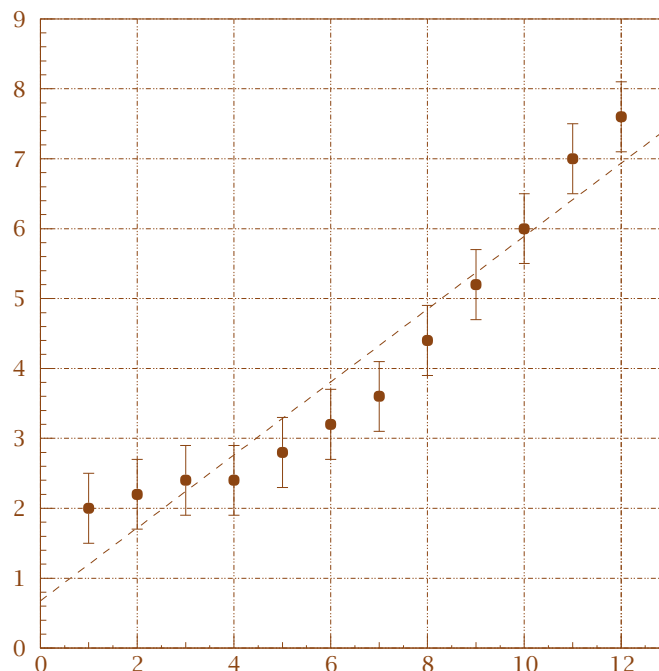
Un'altra tecnica che ci permette di capire se una funzione di primo grado è o meno adeguata a rappresentare un insieme di dati è quella che consiste nell'osservare l'andamento, in funzione della x , del solo segno dei residui δ_i differenza tra i valori misurati e quelli stimati della y :

$$\delta_i = y_i - \hat{y}_i = y_i - (a + b x_i) .$$

Per meglio chiarire questo concetto, osserviamo la figura C1 tratta dal paragrafo 8.3.2 del testo di Barlow citato nella bibliografia (appendice H, a pagina 323). È evidente come l'andamento dei dati sperimentali non suggerisca affatto l'ipotesi di una dipendenza lineare del tipo $y = A + Bx$; questo anche se l'entità degli errori assunti sulle y_i fa sì che l'accordo tra i dati stessi e la retta interpolante, se valutato con il calcolo del χ^2 , risulti comunque accettabile: infatti il metodo citato usa come stima la somma dei quadrati dei *rapporti tra i residui e gli errori stimati*, ovviamente piccola se questi ultimi sono stati sopravvalutati.

Quello che è in grado di suggerire il sospetto di un andamento non lineare della legge $y = y(x)$, in casi come questo, è un altro tipo di controllo basato appunto *sul solo segno dei residui e non sul loro valore* (come il calcolo del χ^2 , o dell'errore a posteriori, o del coefficiente di correlazione lineare, o

FIGURA C1 - Esempio di interpolazione lineare per un insieme di 12 punti.



dalla somma pesata dei quadrati dei residui). Segni che siano (come nell'esempio di figura C1) per piccole x tutti positivi, poi tutti negativi, ed infine tutti positivi per i valori più grandi delle x suggeriranno che si sta tentando di approssimare con una retta una funzione che in realtà è una curva più complessa (ad esempio una parabola) avente concavità rivolta verso l'alto.

Cominciamo con l'osservare che il valore medio $\bar{\delta}$ dei residui è identicamente nullo: infatti dalla (C.7) ricaviamo immediatamente

$$\bar{\delta} = \frac{1}{N} \sum_{i=1}^N \left[(y_i - \bar{y}) - b(x_i - \bar{x}) \right] \equiv 0 .$$

Questo è dovuto al fatto che sia la somma degli scarti $x_i - \bar{x}$ che quella degli scarti $y_i - \bar{y}$ sono identicamente nulle in conseguenza della (4.2), ed è vero quindi indipendentemente dal valore di b : questa proprietà vale insomma per residui calcolati rispetto a *qualunque* retta del fascio di centro (\bar{x}, \bar{y}) cui sappiamo che la retta interpolante *deve* appartenere.

Continuiamo osservando che i residui δ_i e le coordinate x_i hanno tra loro

covarianza nulla:

$$\begin{aligned}
 \text{Cov}(\delta, x) &= \frac{1}{N} \sum_{i=1}^N \left[(\delta_i - \bar{\delta}) (x_i - \bar{x}) \right] \\
 &= \frac{1}{N} \sum_{i=1}^N \left[(y_i - \bar{y}) (x_i - \bar{x}) - b (x_i - \bar{x})^2 \right] \\
 &= \text{Cov}(x, y) - b \text{Var}(x) \\
 &\equiv 0
 \end{aligned}$$

(si è sfruttata, alla fine, la (C.6)). Questa condizione non è sufficiente, come si sa, ad assicurare l'indipendenza statistica tra i residui δ e le coordinate x dei punti interpolati; in effetti queste due variabili casuali *non* sono tra loro indipendenti, essendo ad esempio impossibile che i residui si presentino in una sequenza crescente all'aumentare delle ascisse x_i .

Però, quando il numero N dei dati è grande, delle $N!$ sequenze possibili di residui assai poche sono quelle escluse a priori dalla natura della loro origine; mentre la probabilità delle altre sequenze (che decresce in maniera inversamente proporzionale a $N!$) è comunque assai piccola: e si può in prima approssimazione assumere che *tutte* le sequenze di residui siano *possibili ed equiprobabili*.

Tornando alla figura C1, i residui sono (muovendosi nel senso delle x crescenti) dapprima positivi, poi negativi, poi ancora positivi; sono composti insomma da una sequenza di tre sottoinsiemi di valori aventi tutti lo stesso segno (o, con parola anglosassone, da una sequenza di tre *runs*). Se possiamo assumere equiprobabili tutte le possibili sequenze dei residui è intuitivo capire come un numero così basso di runs si debba presentare con piccola probabilità sulla base di fluttuazioni unicamente casuali; per cui l'osservazione di tale evento può essere attribuita invece alla falsità dell'ipotesi che ha prodotto i residui, ovvero alla non linearità della dipendenza funzionale $y = y(x)$.

Nell'ipotesi di avere un insieme composto da N_+ numeri positivi (corrispondenti, nel caso dei residui, a punti al di sopra della retta interpolante) e da $N_- = N - N_+$ numeri negativi (residui di punti al di sotto della retta interpolante), è possibile calcolare quante delle loro permutazioni producono un prefissato numero di runs N_r ; se è accettabile l'approssimazione cui abbiamo appena accennato, il rapporto tra N_r ed il numero totale di permutazioni possibili ci darà la probabilità di ottenere un certo N_r .

I calcoli dettagliati si possono trovare nel citato paragrafo 8.3.2 del Bar-

low o, più in breve, nel paragrafo 11.3.1 di Eadie *et al.* (testo sempre citato nella bibliografia); qui ricordiamo solo come:

1. se $N_+ = 0$ o $N_- = 0$ (caso impossibile questo per i residui di un'interpolazione lineare) l'unico valore possibile è $N_r = 1$.
2. Se $N_+ > 0$ e $N_- > 0$, ed indicando con $m = \min(N_+, N_-)$ il più piccolo di questi due valori, il numero di runs N_r è compreso tra i seguenti estremi:

$$\begin{cases} 2 \leq N_r \leq 2m & \text{se } N_+ = N_- = m ; \\ 2 \leq N_r \leq 2m + 1 & \text{se } N_+ \neq N_- . \end{cases}$$

Il massimo valore di N_r corrisponde, nel primo caso, ad un alternarsi di valori positivi e negativi; nel secondo, a singoli valori del segno che si è presentato con minore frequenza che separino sequenze di uno o più valori dell'altro segno.

3. Se N_r è pari, con $N_r = 2s$, la probabilità di avere N_r runs è data da

$$\Pr(N_r) = 2 \frac{\binom{N_+ - 1}{s - 1} \binom{N_- - 1}{s - 1}}{\binom{N}{N_+}} ; \quad (\text{C.14})$$

4. se N_r è dispari, con $N_r = 2s - 1$ (ed ovviamente $s \geq 2$ essendo $N_+ > 0$ e $N_- > 0$), la probabilità di avere N_r runs è data da

$$\Pr(N_r) = \frac{\binom{N_+ - 1}{s - 2} \binom{N_- - 1}{s - 1} + \binom{N_+ - 1}{s - 1} \binom{N_- - 1}{s - 2}}{\binom{N}{N_+}} . \quad (\text{C.15})$$

5. In ogni caso, valore medio e varianza di N_r valgono rispettivamente

$$E(N_r) = 1 + 2 \frac{N_+ N_-}{N}$$

e

$$\text{Var}(N_r) = 2 \frac{N_+ N_- (2N_+ N_- - N)}{N^2 (N - 1)} .$$

Nel caso della figura C1 ($N_+ = N_- = 6$) la probabilità di ottenere casualmente $N_r \leq 3$, calcolata applicando direttamente le formule (C.14) e (C.15), vale appena l'1.3%; è insomma lecito (almeno ad un livello di confidenza del 98.7%) rigettare l'ipotesi di un andamento lineare della y in funzione di x .

C.5 Applicazioni alla stima di parametri

Se la densità di probabilità $f(x; \theta)$ di una variabile casuale x dipende da un parametro di valore vero ignoto θ^* , abbiamo visto nel capitolo 11 che una stima $\hat{\theta}$ di tale valore può essere ottenuta col metodo della massima verosimiglianza; e che è possibile ricavare dalla derivata seconda della funzione di verosimiglianza (che ne misura la concavità nel punto di massimo) l'errore di questa stima, attraverso l'equazione (11.5).

Il metodo si può ovviamente estendere alla stima contemporanea di più parametri (e lo abbiamo in effetti usato, ad esempio, per ricavare i due coefficienti della retta interpolante nel paragrafo 11.4.1): se x è la variabile misurata, di densità di probabilità $f(x; \theta_1, \theta_2, \dots, \theta_M)$ dipendente da M parametri θ_k , le stime di massima verosimiglianza $\hat{\theta}_k$ si ricavano risolvendo il sistema ottenuto annullando contemporaneamente ognuna delle derivate parziali prime (ed esaminando poi ognuna delle eventuali soluzioni per controllare se corrisponde ad un massimo).

I punti enunciati nel paragrafo 11.2 continuano a rimanere validi anche nel caso multidimensionale: in particolare, ognuna delle stime $\hat{\theta}_k$ è asintoticamente normale; e si troverebbe che le derivate seconde della funzione di verosimiglianza nel punto di minimo sono legate all'inversa della matrice delle covarianze delle M stime attraverso la

$$(V^{-1})_{ij} = -N \cdot E \left\{ \frac{\partial^2 \ln f(x; \theta_1, \theta_2, \dots, \theta_M)}{\partial \theta_i \partial \theta_j} \right\} \quad (C.16)$$

che si può pensare come la generalizzazione a più stime contemporanee dell'equazione (11.5).

Come esempio, abbiamo già visto nel paragrafo 11.5 come stimare *contemporaneamente* i due parametri μ e σ di una popolazione normale da un campione di N determinazioni indipendenti col metodo della massima verosimiglianza; e vogliamo ora ricavare gli errori di quelle stime dalla (C.16).

Ricordiamo dal paragrafo 11.5 che il logaritmo della densità di probabilità vale

$$\ln f(x; \mu, \sigma) = -\ln \sigma - \ln \sqrt{2\pi} - \frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 ;$$

e che le due stime di massima verosimiglianza per μ e σ sono

$$\hat{\mu} = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{e} \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2$$

rispettivamente. Le derivate prime di $\ln f$ sono

$$\frac{\partial}{\partial \mu} \ln f = \frac{x - \mu}{\sigma^2} \quad \frac{\partial}{\partial \sigma} \ln f = -\frac{1}{\sigma} + \frac{(x - \mu)^2}{\sigma^3}$$

e le derivate seconde

$$\begin{aligned}\frac{\partial^2}{\partial \mu^2} \ln f &= -\frac{1}{\sigma^2} & \frac{\partial^2}{\partial \mu \partial \sigma} \ln f &= -2 \frac{x - \mu}{\sigma^3} \\ \frac{\partial^2}{\partial \sigma \partial \mu} \ln f &= -2 \frac{x - \mu}{\sigma^3} & \frac{\partial^2}{\partial \sigma^2} \ln f &= \frac{1}{\sigma^2} - 3 \frac{(x - \mu)^2}{\sigma^4}\end{aligned}$$

di valori medi

$$\begin{aligned}E\left(\frac{\partial^2}{\partial \mu^2} \ln f\right) &= -\frac{1}{\sigma^2} & E\left(\frac{\partial^2}{\partial \mu \partial \sigma} \ln f\right) &= 0 \\ E\left(\frac{\partial^2}{\partial \sigma \partial \mu} \ln f\right) &= 0 & E\left(\frac{\partial^2}{\partial \sigma^2} \ln f\right) &= -\frac{2}{\sigma^2} ;\end{aligned}$$

per cui, dalla (C.16), l'inverso della matrice delle covarianze (che è diagonale) è

$$\mathbf{V}^{-1} = \begin{vmatrix} \frac{N}{\hat{\sigma}^2} & 0 \\ 0 & \frac{2N}{\hat{\sigma}^2} \end{vmatrix}$$

e la matrice \mathbf{V} stessa vale

$$\mathbf{V} = \begin{vmatrix} \frac{\hat{\sigma}^2}{N} & 0 \\ 0 & \frac{\hat{\sigma}^2}{2N} \end{vmatrix}.$$

Insomma, oltre alla consueta espressione della varianza della media

$$\text{Var}(\hat{\mu}) = \frac{\hat{\sigma}^2}{N}$$

abbiamo ottenuto quella della varianza di $\hat{\sigma}$

$$\text{Var}(\hat{\sigma}) = \frac{\hat{\sigma}^2}{2N}$$

da confrontare con la (B.1), in cui però $\hat{\sigma}$ era già stato corretto, moltiplicandolo per un fattore $N/(N-1)$, per eliminare la distorsione della stima; e la riconferma del fatto, già visto nel paragrafo 12.1 a pagina 203, che valore medio e varianza di un campione di stime indipendenti sono variabili casuali statisticamente indipendenti tra loro (le covarianze infatti sono nulle).

Appendice D

Il modello di Laplace e la funzione di Gauss

Pensiamo di eseguire una misura di una grandezza fisica (il cui valore 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*¹ per gli errori di misura):

1. *Ognuna delle singole cause di disturbo presenti introdurrà nella misura una variazione rispetto al valore 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.*

¹Pierre Simon de Laplace visse in Francia dal 1749 al 1827; famoso matematico, fisico ed astronomo, provò la stabilità del sistema solare, sviluppò la teoria delle equazioni differenziali e dei potenziali, contribuì allo studio del calore e dei fenomeni capillari oltre a gettare le basi matematiche per una teoria dell'elettromagnetismo. Durante la rivoluzione francese fu uno degli ideatori del sistema metrico decimale; per quel che riguarda la statistica, nel 1812 pubblicò il trattato "Théorie Analytique des Probabilités" che contiene, tra l'altro, studi sulla distribuzione normale e la derivazione della regola dei minimi quadrati.

Ognuna delle N cause indipendenti di disturbo produce quindi la variazione $+\epsilon$ con probabilità $p = 0.5$ oppure $-\epsilon$ con probabilità $q = 1 - p = 0.5$; se M tra le N perturbazioni sono positive (e le altre $N - M$ negative), il valore osservato sarà

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

La probabilità di un dato valore di M sulle N prove è data dalla distribuzione binomiale (vedi il paragrafo 8.4, ed in particolare l'equazione (8.7)), e vale

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

Il valore 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 valore 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

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

Valore medio e varianza di λ valgono poi

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

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 :

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

e $P(M, N)$ 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 $\mu = \lceil Np - q \rceil$ 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*² per il fattoriale:

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

con

$$0 \leq \epsilon_N < \frac{1}{11 \cdot N}.$$

È 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,

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

²Per la dimostrazione, vedi ad esempio: G. Castelnuovo - Calcolo delle probabilità (Zanichelli), in appendice. La formula è dovuta al solo Stirling, che la pubblicò nel suo libro "Methodus Differentialis" del 1730; ma non divenne nota nel mondo scientifico fino a quando de Moivre non la usò — da qui il nome comunemente adottato.

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

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

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 valore massimo $P(0, N)$, e se ne prenda il logaritmo naturale:

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

Ora, poiché 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 McLaurin:

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

Il primo termine di $\ln \kappa$ 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

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

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

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

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, $p = q = 0.5$ 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

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

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

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

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

e possiede varianza

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

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

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

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 :

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

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, purché ciascuno abbia una varianza dello stesso ordine di grandezza degli altri ed infinitesima al divergere del numero delle cause di errore.

Appendice E

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 data da una funzione nota $f_i(x_i; \theta)$; funzione che supponiamo dipenda da un parametro θ di valore vero θ^* ignoto, e 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 valore 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 valore vero:

$$E(t) = \tau(\theta) .$$

Il caso particolare della stima *del parametro stesso* corrisponde alla funzione $\tau(\theta) = \theta$, che soddisfa evidentemente alla richiesta di possedere derivata prima non nulla $\tau'(\theta) = 1$.

Una importante proprietà della stima t è la sua varianza, data (se essa è imparziale) da

$$\begin{aligned}\sigma_t^2 &= \int_{-\infty}^{+\infty} [t - \tau(\theta)]^2 g(t; \theta) dt \\ &= \int_{-\infty}^{+\infty} dx_1 f_1(x_1; \theta) \cdots \int_{-\infty}^{+\infty} dx_N f_N(x_N; \theta) [t(x_1, x_2, \dots, x_N) - \tau(\theta)]^2\end{aligned}$$

perché la *minima varianza* sarà il nostro criterio di scelta fra diverse stime 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 valore vero θ^* si pone il parametro variabile θ , si ottiene la *funzione di verosimiglianza*

$$\mathcal{L}(x_1, x_2, \dots, x_N; \theta) = \prod_{i=1}^N f_i(x_i; \theta) \quad .$$

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 detta variabile 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

$$\begin{aligned}
 \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \cdots \int_{-\infty}^{+\infty} dx_N \mathcal{L} \left(\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta} \right) &= \\
 &= \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \cdots \int_{-\infty}^{+\infty} dx_N \mathcal{L} \frac{\partial (\ln \mathcal{L})}{\partial \theta} \\
 &= \int_{-\infty}^{+\infty} dx_1 f_1(x_1; \theta) \cdots \int_{-\infty}^{+\infty} dx_N f_N(x_N; \theta) \frac{\partial (\ln \mathcal{L})}{\partial \theta} \\
 &= 0
 \end{aligned}$$

ossia

$$E \left\{ \frac{\partial (\ln \mathcal{L})}{\partial \theta} \right\} = 0 \quad (\text{E.1})$$

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

$$\begin{aligned}
 \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \cdots \int_{-\infty}^{+\infty} dx_N t \frac{\partial \mathcal{L}}{\partial \theta} &= \\
 &= \int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_N t \mathcal{L} \left(\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta} \right) \\
 &= \int_{-\infty}^{+\infty} dx_1 f_1(x_1; \theta) \cdots \int_{-\infty}^{+\infty} dx_N f_N(x_N; \theta) t \frac{\partial (\ln \mathcal{L})}{\partial \theta} \\
 &= E \left\{ t \frac{\partial (\ln \mathcal{L})}{\partial \theta} \right\}
 \end{aligned}$$

e, in definitiva,

$$E \left\{ t \frac{\partial (\ln \mathcal{L})}{\partial \theta} \right\} = \tau'(\theta)$$

Infine, sottraendo membro a membro da questa equazione la precedente (E.1) moltiplicata per $\tau(\theta)$, si ottiene

$$E \left\{ t \frac{\partial (\ln \mathcal{L})}{\partial \theta} \right\} - \tau(\theta) \cdot E \left\{ \frac{\partial (\ln \mathcal{L})}{\partial \theta} \right\} = \tau'(\theta)$$

ovvero

$$E \left\{ [t - \tau(\theta)] \cdot \frac{\partial (\ln \mathcal{L})}{\partial \theta} \right\} = \tau'(\theta) .$$

Se ora si definiscono il rapporto

$$R(\theta) = \frac{E \left\{ [t - \tau(\theta)] \cdot \frac{\partial (\ln \mathcal{L})}{\partial \theta} \right\}}{E \left\{ \left[\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right]^2 \right\}} = \frac{\tau'(\theta)}{E \left\{ \left[\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right]^2 \right\}}$$

(che è una costante dipendente da θ ; osserviamo anche che deve risultare $R(\theta) \neq 0$) e la variabile casuale

$$z = [t - \tau(\theta)] - R(\theta) \frac{\partial (\ln \mathcal{L})}{\partial \theta}$$

il cui quadrato risulta essere

$$z^2 = [t - \tau(\theta)]^2 - 2 R(\theta) \cdot [t - \tau(\theta)] \frac{\partial (\ln \mathcal{L})}{\partial \theta} + R^2(\theta) \cdot \left[\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right]^2$$

prendendo il valore medio di z^2 si ottiene

$$\begin{aligned} E(z^2) = E \left\{ [t - \tau(\theta)]^2 \right\} - 2 R(\theta) \cdot E \left\{ [t - \tau(\theta)] \cdot \frac{\partial (\ln \mathcal{L})}{\partial \theta} \right\} + \\ + R^2(\theta) \cdot E \left\{ \left[\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right]^2 \right\} \end{aligned}$$

ossia

$$\begin{aligned} E(z^2) = \sigma_t^2 - 2 \frac{\tau'(\theta)}{E \left\{ \left[\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right]^2 \right\}} \tau'(\theta) + \\ + \left\{ \frac{\tau'(\theta)}{E \left\{ \left[\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right]^2 \right\}} \right\}^2 E \left\{ \left[\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right]^2 \right\} \end{aligned}$$

ed infine

$$\begin{aligned} E(z^2) &= \sigma_t^2 - 2 \frac{[\tau'(\theta)]^2}{E \left\{ \left[\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right]^2 \right\}} + \frac{[\tau'(\theta)]^2}{E \left\{ \left[\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right]^2 \right\}} \\ &= \sigma_t^2 - \frac{[\tau'(\theta)]^2}{E \left\{ \left[\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right]^2 \right\}} . \end{aligned}$$

Ma il valore medio del quadrato di una qualsiasi variabile casuale non può essere negativo, e dunque

$$0 \leq E(z^2) = \sigma_t^2 - \frac{[\tau'(\theta)]^2}{E \left\{ \left[\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right]^2 \right\}}$$

ed infine

$$\sigma_t^2 \geq \frac{[\tau'(\theta)]^2}{E \left\{ \left[\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right]^2 \right\}} = [\tau'(\theta)]^2 \frac{R(\theta)}{\tau'(\theta)} = \tau'(\theta) \cdot R(\theta)$$

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

$$z = t - \tau(\theta) - R(\theta) \frac{\partial (\ln \mathcal{L})}{\partial \theta} \equiv 0$$

o, altrimenti detto, se la derivata logaritmica della verosimiglianza è proporzionale alla variabile casuale $t - \tau(\theta)$:

$$\boxed{\frac{\partial (\ln \mathcal{L})}{\partial \theta} = \frac{t - \tau(\theta)}{R(\theta)}} \quad (\text{E.2})$$

Nel caso particolare che tutte le x_i provengano dalla stessa popolazione, e che quindi abbiano la stessa densità di probabilità $f(x; \theta)$,

$$\begin{aligned}\frac{\partial(\ln \mathcal{L})}{\partial \theta} &= \frac{\partial}{\partial \theta} \sum_{i=1}^N \ln f(x_i; \theta) = \sum_{i=1}^N \frac{\partial}{\partial \theta} \ln f(x_i; \theta) \\ E \left\{ \frac{\partial(\ln \mathcal{L})}{\partial \theta} \right\} &= \sum_{i=1}^N E \left\{ \frac{\partial}{\partial \theta} \ln f(x_i; \theta) \right\} = N \cdot E \left\{ \frac{\partial}{\partial \theta} \ln f(x; \theta) \right\}\end{aligned}$$

e, tenuto conto della (E.1), questo implica che

$$E \left\{ \frac{\partial}{\partial \theta} \ln f(x; \theta) \right\} = 0. \quad (\text{E.3})$$

Ora

$$\begin{aligned}E \left\{ \left[\frac{\partial(\ln \mathcal{L})}{\partial \theta} \right]^2 \right\} &= E \left\{ \left[\sum_{i=1}^N \frac{\partial}{\partial \theta} \ln f(x_i; \theta) \right] \left[\sum_{k=1}^N \frac{\partial}{\partial \theta} \ln f(x_k; \theta) \right] \right\} \\ &= \sum_{i=1}^N E \left\{ \left[\frac{\partial}{\partial \theta} \ln f(x_i; \theta) \right]^2 \right\} + \sum_{\substack{i,k \\ i \neq k}} E \left\{ \frac{\partial}{\partial \theta} \ln f(x_i; \theta) \cdot \frac{\partial}{\partial \theta} \ln f(x_k; \theta) \right\} \\ &= N \cdot E \left\{ \left[\frac{\partial}{\partial \theta} \ln f(x; \theta) \right]^2 \right\} + \sum_{\substack{i,k \\ i \neq k}} E \left\{ \frac{\partial}{\partial \theta} \ln f(x_i; \theta) \right\} \cdot E \left\{ \frac{\partial}{\partial \theta} \ln f(x_k; \theta) \right\}\end{aligned}$$

(tenendo conto del fatto che le x_i sono indipendenti); l'ultimo termine si annulla in conseguenza della (E.3), ed infine, in questo caso, il minorante della varianza della stima si può scrivere

$$\sigma_t^2 \geq \frac{[\tau'(\theta)]^2}{N \cdot E \left\{ \left[\frac{\partial}{\partial \theta} \ln f(x; \theta) \right]^2 \right\}}$$

Col metodo della massima verosimiglianza si assume, come stima del valore 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 $\tau(\theta)$* , tenendo conto della (E.2) la condizione perché la funzione di verosimiglianza abbia un estremante diviene

$$\frac{\partial(\ln \mathcal{L})}{\partial \theta} = \frac{t - \tau(\theta)}{R(\theta)} = 0$$

e le soluzioni $\hat{\theta}$ sono tutte e sole quelle dell'equazione

$$\tau(\theta) = t(x_1, x_2, \dots, x_N) \quad .$$

La derivata seconda di $\ln \mathcal{L}$ è in tal caso

$$\begin{aligned} \frac{\partial^2 (\ln \mathcal{L})}{\partial \theta^2} &= - \frac{\tau'(\theta) \cdot R(\theta) + R'(\theta) \cdot [t - \tau(\theta)]}{R^2(\theta)} \\ &= - \frac{\sigma_t^2 + R'(\theta) \cdot [t - \tau(\theta)]}{R^2(\theta)} \end{aligned}$$

ma se $\theta = \hat{\theta}$ è anche $t - \tau(\hat{\theta}) = 0$ e risulta

$$\left[\frac{\partial^2 (\ln \mathcal{L})}{\partial \theta^2} \right]_{\theta=\hat{\theta}} = - \frac{\sigma_t^2}{R^2(\hat{\theta})} < 0 \quad ;$$

cioè *per tutte le soluzioni* $\theta = \hat{\theta}$ *la verosimiglianza è massima.*

Ora, se la funzione $\ln \mathcal{L}$ 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 $t(x_1, x_2, \dots, x_N)$:

$$\hat{\theta} = \tau^{-1}[t(x_1, x_2, \dots, x_N)] \quad .$$

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 valore 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

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{i=1}^N \frac{(x_i - \lambda)^2}{\sigma_i^2} - \sum_{i=1}^N \ln(\sigma_i \sqrt{2\pi})$$

e la sua derivata rispetto al parametro λ

$$\frac{\partial (\ln \mathcal{L})}{\partial \lambda} = \sum_{i=1}^N \frac{x_i - \lambda}{\sigma_i^2} = \left(\sum_i \frac{1}{\sigma_i^2} \right) \left(\frac{\sum_i \frac{x_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}} - \lambda \right) .$$

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

$$\frac{\partial (\ln \mathcal{L})}{\partial \lambda} = \frac{1}{\sigma^2} \left[\left(\sum_{i=1}^N x_i \right) - N\lambda \right] = \frac{N}{\sigma^2} (\bar{x} - \lambda) = \frac{\bar{x} - \lambda}{R}$$

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 d'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 valore medio λ , la stima della varianza σ^2 si ottiene cercando lo zero della derivata logaritmica

$$\frac{\partial (\ln \mathcal{L})}{\partial \sigma} = \frac{1}{\sigma^3} \left[\sum_{i=1}^N (x_i - \lambda)^2 \right] - \frac{N}{\sigma} = \frac{N}{\sigma^3} \left\{ \left[\frac{1}{N} \sum_{i=1}^N (x_i - \lambda)^2 \right] - \sigma^2 \right\}$$

la quale ha la forma richiesta perché 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 B.

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 .

Appendice F

La licenza GNU GPL (General Public License)

Questo capitolo contiene la licenza GNU GPL, sotto la quale questo libro viene distribuito, sia nella versione originale inglese¹ (la sola dotata di valore legale) che in una traduzione non ufficiale in italiano² che aiuti chi ha difficoltà con l'inglese legale a comprendere meglio il significato della licenza stessa.

F.1 The GNU General Public License

Version 2, June 1991

Copyright © 1989, 1991 Free Software Foundation, Inc.

59 Temple Place - Suite 330, Boston, MA 02111-1307, USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General

¹<http://www.gnu.org/licenses/gpl.html>

²La traduzione è dovuta al gruppo *Pluto* (PLUTO Linux/Lumen Utentibus Terrarum Orbis, <http://www.pluto.linux.it/>).

Public License is intended to guarantee your freedom to share and change free software — to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Library General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid anyone to deny you these rights or to ask you to surrender the rights. These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

Also, for each author's protection and ours, we want to make certain that everyone understands that there is no warranty for this free software. If the software is modified by someone else and passed on, we want its recipients to know that what they have is not the original, so that any problems introduced by others will not reflect on the original authors' reputations.

Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.

The precise terms and conditions for copying, distribution and modification follow.

GNU GENERAL PUBLIC LICENSE

TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

0. This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License. The “Program”, below, refers to any such program or work, and a “work based on the Program” means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term “modification”.) Each licensee is addressed as “you”.

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

1. You may copy and distribute verbatim copies of the Program’s source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program. You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.
2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:
 - (a) You must cause the modified files to carry prominent notices stating that you changed the files and the date of any change.

- (b) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.
- (c) If the modified program normally reads commands interactively when run, you must cause it, when started running for such interactive use in the most ordinary way, to print or display an announcement including an appropriate copyright notice and a notice that there is no warranty (or else, saying that you provide a warranty) and that users may redistribute the program under these conditions, and telling the user how to view a copy of this License. (Exception: if the Program itself is interactive but does not normally print such an announcement, your work based on the Program is not required to print an announcement.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program.

In addition, mere aggregation of another work not based on the Program with the Program (or with a work based on the Program) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may copy and distribute the Program (or a work based on it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you also do one of the following:
 - (a) Accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

- (b) Accompany it with a written offer, valid for at least three years, to give any third party, for a charge no more than your cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,
- (c) Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for noncommercial distribution and only if you received the program in object code or executable form with such an offer, in accord with Subsection b above.)

The source code for a work means the preferred form of the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.

If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.

4. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.
5. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate

your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.

6. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.
7. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

8. If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add

an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

9. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and “any later version”, you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.

10. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

NO WARRANTY

11. BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.
12. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO

MAY MODIFY AND/OR REDISTRIBUTE THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

END OF TERMS AND CONDITIONS

Appendix: How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively convey the exclusion of warranty; and each file should have at least the “copyright” line and a pointer to where the full notice is found.

```
<one line to give the program's name and a brief idea of what it does.>  
Copyright (C) <year> <name of author>
```

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111-1307, USA.

Also add information on how to contact you by electronic and paper mail. If the program is interactive, make it output a short notice like this when it starts in an interactive mode:

Gnomovision version 69, Copyright (C) <year> <name of author>
Gnomovision comes with ABSOLUTELY NO WARRANTY; for details type
‘show w’.

This is free software, and you are welcome to redistribute it under
certain conditions; type ‘show c’ for details.

The hypothetical commands `show w` and `show c` should show the appropriate parts of the General Public License. Of course, the commands you use may be called something other than `show w` and `show c`; they could even be mouse-clicks or menu items — whatever suits your program.

You should also get your employer (if you work as a programmer) or your school, if any, to sign a “copyright disclaimer” for the program, if necessary. Here is a sample; alter the names:

Yoyodyne, Inc., hereby disclaims all copyright interest in the program
‘Gnomovision’ (which makes passes at compilers) written by James Hacker.

<signature of Ty Coon>, 1 April 1989
Ty Coon, President of Vice

This General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Library General Public License instead of this License.

F.2 Licenza pubblica generica del progetto GNU

Versione 2, Giugno 1991
Copyright © 1989, 1991 Free Software Foundation, Inc.

59 Temple Place - Suite 330, Boston, MA 02111-1307, USA

Tutti possono copiare e distribuire copie letterali di questo documento di
licenza, ma non è permesso modificarlo.

Preambolo

Le licenze per la maggioranza dei programmi hanno lo scopo di togliere all’utente la libertà di condividerlo e di modificarlo. Al contrario, la Licenza Pubblica Generica GNU è intesa a garantire

la libertà di condividere e modificare il free software, al fine di assicurare che i programmi siano “liberi” per tutti i loro utenti. Questa Licenza si applica alla maggioranza dei programmi della Free Software Foundation e a ogni altro programma i cui autori hanno scelto questa Licenza. Alcuni altri programmi della Free Software Foundation sono invece coperti dalla Licenza Pubblica Generica per Librerie (LGPL). Chiunque può usare questa Licenza per i propri programmi.

Quando si parla di free software, ci si riferisce alla libertà, non al prezzo. Le nostre Licenze (la GPL e la LGPL) sono progettate per assicurare che ciascuno abbia la libertà di distribuire copie del software libero (e farsi pagare per questo, se vuole), che ciascuno riceva il codice sorgente o che lo possa ottenere se lo desidera, che ciascuno possa modificare il programma o usarne delle parti in nuovi programmi liberi e che ciascuno sappia di potere fare queste cose.

Per proteggere i diritti dell'utente, abbiamo bisogno di creare delle restrizioni che vietino a chiunque di negare questi diritti o di chiedere di rinunciarvi. Queste restrizioni si traducono in certe responsabilità per chi distribuisce copie del software e per chi lo modifica.

Per esempio, chi distribuisce copie di un Programma coperto da GPL, sia gratuitamente sia facendosi pagare, deve dare agli acquirenti tutti i diritti che ha ricevuto. Deve anche assicurarsi che gli acquirenti ricevano o possano ricevere il codice sorgente. E deve mostrar loro queste condizioni di Licenza, in modo che conoscano i loro diritti.

Proteggiamo i diritti dell'utente attraverso due azioni: (1) proteggendo il software con un diritto d'autore (una nota di copyright), e (2) offrendo una Licenza che concede il permesso legale di copiare, distribuire e/o modificare il Programma.

Infine, per proteggere ogni autore e noi stessi, vogliamo assicurarci che ognuno capisca che non ci sono garanzie per i programmi coperti da GPL. Se il Programma viene modificato da qualcun altro e ridistribuito, vogliamo che gli acquirenti sappiano che ciò che hanno non è l'originale, in modo che ogni problema introdotto da altri non si rifletta sulla reputazione degli autori originari.

Infine, ogni programma libero è costantemente minacciato dai

brevetti sui programmi. Vogliamo evitare il pericolo che chi ridistribuisce un Programma libero ottenga brevetti personali, rendendo perciò il Programma una cosa di sua proprietà. Per prevenire questo, abbiamo chiarito che ogni prodotto brevettato debba essere reso disponibile perché tutti ne usufruiscano liberamente; se l'uso del prodotto deve sottostare a restrizioni allora tale prodotto non deve essere distribuito affatto.

Seguono i termini e le condizioni precisi per la copia, la distribuzione e la modifica.

LICENZA PUBBLICA GENERICA GNU

TERMINI E CONDIZIONI PER LA COPIA, LA DISTRIBUZIONE E LA MODIFICA

0. Questa Licenza si applica a ogni Programma o altra opera che contenga una nota da parte del detentore del diritto d'autore che dica che tale opera può essere distribuita nei termini di questa Licenza Pubblica Generica. Il termine "Programma" nel seguito indica ognuno di questi programmi o lavori, e l'espressione "lavoro basato sul Programma" indica sia il Programma sia ogni opera considerata "derivata" in base alla legge sul diritto d'autore: cioè un lavoro contenente il Programma o una porzione di esso, sia letteralmente sia modificato e/o tradotto in un'altra lingua; da qui in avanti, la traduzione è in ogni caso considerata una "modifica". Vengono ora elencati i diritti dei detentori di licenza.

Attività diverse dalla copiatura, distribuzione e modifica non sono coperte da questa Licenza e sono al di fuori della sua influenza. L'atto di eseguire il programma non viene limitato, e l'output del programma è coperto da questa Licenza solo se il suo contenuto costituisce un lavoro basato sul Programma (indipendentemente dal fatto che sia stato creato eseguendo il Programma). In base alla natura del Programma il suo output può essere o meno coperto da questa Licenza.

1. È lecito copiare e distribuire copie letterali del codice sorgente del Programma così come viene ricevuto, con qualsiasi mezzo, a condizione che venga riprodotta chiaramente su ogni copia un'appropriata nota di diritto d'autore e di assenza di garanzia; che si mantengano intatti tutti i riferimenti a questa Licenza e all'assenza di ogni garanzia; che si

dia a ogni altro acquirente del Programma una copia di questa Licenza insieme al Programma.

È possibile richiedere un pagamento per il trasferimento fisico di una copia del Programma, è anche possibile a propria discrezione richiedere un pagamento in cambio di una copertura assicurativa.

2. È lecito modificare la propria copia o copie del Programma, o parte di esso, creando perciò un lavoro basato sul Programma, e copiare o distribuire queste modifiche e questi lavori secondo i termini del precedente comma 1, a patto che vengano soddisfatte queste condizioni:
 - (a) Bisogna indicare chiaramente nei file che si tratta di copie modificate e la data di ogni modifica.
 - (b) Bisogna fare in modo che ogni lavoro distribuito o pubblicato, che in parte o nella sua totalità derivi dal Programma o da parti di esso, sia utilizzabile gratuitamente da terzi nella sua totalità, secondo le condizioni di questa licenza.
 - (c) Se di solito il programma modificato legge comandi interattivamente quando viene eseguito, bisogna fare in modo che all'inizio dell'esecuzione interattiva usuale, stampi un messaggio contenente un'appropriata nota di diritto d'autore e di assenza di garanzia (oppure che specifichi che si offre una garanzia). Il messaggio deve inoltre specificare agli utenti che possono ridistribuire il programma alle condizioni qui descritte e deve indicare come consultare una copia di questa licenza. Se però il programma di partenza è interattivo ma normalmente non stampa tale messaggio, non occorre che un lavoro derivato lo stampi.

Questi requisiti si applicano al lavoro modificato nel suo complesso. Se sussistono parti identificabili del lavoro modificato che non siano derivate dal Programma e che possono essere ragionevolmente considerate lavori indipendenti, allora questa Licenza e i suoi termini non si applicano a queste parti quando vengono distribuite separatamente. Se però queste parti vengono distribuite all'interno di un prodotto che è un lavoro basato sul Programma, la distribuzione di questo prodotto nel suo complesso deve avvenire nei termini di questa Licenza, le cui norme nei confronti di altri utenti si estendono a tutto il prodotto, e quindi a ogni sua parte, chiunque ne sia l'autore.

Sia chiaro che non è nelle intenzioni di questa sezione accampare diritti su lavori scritti interamente da altri, l'intento è piuttosto quello di

esercitare il diritto di controllare la distribuzione di lavori derivati o dal Programma o di cui esso sia parte.

Inoltre, se il Programma o un lavoro derivato da esso viene aggregato a un altro lavoro non derivato dal Programma su di un mezzo di memorizzazione o di distribuzione, il lavoro non derivato non ricade nei termini di questa licenza.

3. È lecito copiare e distribuire il Programma (o un lavoro basato su di esso, come espresso al comma 2) sotto forma di codice oggetto o eseguibile secondo i termini dei precedenti commi 1 e 2, a patto che si applichi una delle seguenti condizioni:
 - (a) Il Programma sia corredato dal codice sorgente completo, in una forma leggibile dal calcolatore e tale sorgente deve essere fornito secondo le regole dei precedenti commi 1 e 2 su di un mezzo comunemente usato per lo scambio di programmi.
 - (b) Il Programma sia accompagnato da un'offerta scritta, valida per almeno tre anni, di fornire a chiunque ne faccia richiesta una copia completa del codice sorgente, in una forma leggibile dal calcolatore, in cambio di un compenso non superiore al costo del trasferimento fisico di tale copia, che deve essere fornita secondo le regole dei precedenti commi 1 e 2 su di un mezzo comunemente usato per lo scambio di programmi.
 - (c) Il Programma sia accompagnato dalle informazioni che sono state ricevute riguardo alla possibilità di ottenere il codice sorgente. Questa alternativa è permessa solo in caso di distribuzioni non commerciali e solo se il programma è stato ricevuto sotto forma di codice oggetto o eseguibile in accordo al precedente punto b).

Per “codice sorgente completo” di un lavoro si intende la forma preferenziale usata per modificare un lavoro. Per un programma eseguibile, “codice sorgente completo” significa tutto il codice sorgente di tutti i moduli in esso contenuti, più ogni file associato che definisca le interfacce esterne del programma, più gli script usati per controllare la compilazione e l'installazione dell'eseguibile. In ogni caso non è necessario che il codice sorgente fornito includa nulla che sia normalmente distribuito (in forma sorgente o in formato binario) con i principali componenti del sistema operativo sotto cui viene eseguito il Programma (compilatore, kernel, e così via), a meno che tali componenti accompagnino l'eseguibile.

Se la distribuzione dell'eseguibile o del codice oggetto è effettuata indicando un luogo dal quale sia possibile copiarlo, permettere la copia del codice sorgente dallo stesso luogo è considerata una valida forma di distribuzione del codice sorgente, anche se copiare il sorgente è facoltativo per l'acquirente.

4. Non è lecito copiare, modificare, sublicenziare, o distribuire il Programma in modi diversi da quelli espressamente previsti da questa Licenza. Ogni tentativo contrario di copiare, modificare, sublicenziare o distribuire il Programma è legalmente nullo, e farà cessare automaticamente i diritti garantiti da questa Licenza. D'altra parte ogni acquirente che abbia ricevuto copie, o diritti, coperti da questa Licenza da parte di persone che violano la Licenza come qui indicato non vedranno invalidare la loro Licenza, purché si comportino conformemente a essa.
5. L'acquirente non è obbligato ad accettare questa Licenza, poiché non l'ha firmata. D'altra parte nessun altro documento garantisce il permesso di modificare o distribuire il Programma o i lavori derivati da esso. Queste azioni sono proibite dalla legge per chi non accetta questa Licenza; perciò, modificando o distribuendo il Programma o un lavoro basato sul programma, si accetta implicitamente questa Licenza e quindi di tutti i suoi termini e le condizioni poste sulla copia, la distribuzione e la modifica del Programma o di lavori basati su di esso.
6. Ogni volta che il Programma o un lavoro basato su di esso vengono distribuiti, l'acquirente riceve automaticamente una licenza d'uso da parte del licenziatario originale. Tale licenza regola la copia, la distribuzione e la modifica del Programma secondo questi termini e queste condizioni. Non è lecito imporre restrizioni ulteriori all'acquirente nel suo esercizio dei diritti qui garantiti. Chi distribuisce programmi coperti da questa Licenza non è comunque responsabile per la conformità alla Licenza da parte di terzi.
7. Se, come conseguenza del giudizio di un tribunale, o di un'imputazione per la violazione di un brevetto o per ogni altra ragione (anche non relativa a questioni di brevetti), vengono imposte condizioni che contraddicono le condizioni di questa licenza, che queste condizioni siano dettate dal tribunale, da accordi tra le parti o altro, queste condizioni non esimono nessuno dall'osservazione di questa Licenza. Se non è possibile distribuire un prodotto in un modo che soddisfi simultaneamente gli obblighi dettati da questa Licenza e altri obblighi pertinenti,

il prodotto non può essere distribuito affatto. Per esempio, se un brevetto non permettesse a tutti quelli che lo ricevono di ridistribuire il Programma senza obbligare al pagamento di diritti, allora l'unico modo per soddisfare contemporaneamente il brevetto e questa Licenza è di non distribuire affatto il Programma.

Se parti di questo comma sono ritenute non valide o inapplicabili per qualsiasi circostanza, deve comunque essere applicata l'idea espressa da questo comma; in ogni altra circostanza invece deve essere applicato il comma 7 nel suo complesso.

Non è nello scopo di questo comma indurre gli utenti a violare alcun brevetto né ogni altra rivendicazione di diritti di proprietà, né di contestare la validità di alcuna di queste rivendicazioni; lo scopo di questo comma è solo quello di proteggere l'integrità del sistema di distribuzione del software libero, che viene realizzato tramite l'uso della licenza pubblica. Molte persone hanno contribuito generosamente alla vasta gamma di programmi distribuiti attraverso questo sistema, basandosi sull'applicazione consistente di tale sistema. L'autore/donatore può decidere di sua volontà se preferisce distribuire il software avvalendosi di altri sistemi, e l'acquirente non può imporre la scelta del sistema di distribuzione.

Questo comma serve a rendere il più chiaro possibile ciò che crediamo sia una conseguenza del resto di questa Licenza.

8. Se in alcuni paesi la distribuzione e/o l'uso del Programma sono limitati da brevetto o dall'uso di interfacce coperte da diritti d'autore, il detentore del copyright originale che pone il Programma sotto questa Licenza può aggiungere limiti geografici espliciti alla distribuzione, per escludere questi paesi dalla distribuzione stessa, in modo che il programma possa essere distribuito solo nei paesi non esclusi da questa regola. In questo caso i limiti geografici sono inclusi in questa Licenza e ne fanno parte a tutti gli effetti.
9. All'occorrenza la Free Software Foundation può pubblicare revisioni o nuove versioni di questa Licenza Pubblica Generica. Tali nuove versioni saranno simili a questa nello spirito, ma potranno differire nei dettagli al fine di coprire nuovi problemi e nuove situazioni.

Ad ogni versione viene dato un numero identificativo. Se il Programma asserisce di essere coperto da una particolare versione di questa Licenza e "da ogni versione successiva", l'acquirente può scegliere se seguire

le condizioni della versione specificata o di una successiva. Se il Programma non specifica quale versione di questa Licenza deve applicarsi, l'acquirente può scegliere una qualsiasi versione tra quelle pubblicate dalla Free Software Foundation.

10. Se si desidera incorporare parti del Programma in altri programmi liberi le cui condizioni di distribuzione differiscano da queste, è possibile scrivere all'autore del Programma per chiederne l'autorizzazione. Per il software il cui copyright è detenuto dalla Free Software Foundation, si scriva alla Free Software Foundation; talvolta facciamo eccezioni alle regole di questa Licenza. La nostra decisione sarà guidata da due scopi: preservare la libertà di tutti i prodotti derivati dal nostro software libero e promuovere la condivisione e il riutilizzo del software in generale.

NESSUNA GARANZIA

11. POICHÉ IL PROGRAMMA È CONCESSO IN USO GRATUITAMENTE, NON C'È ALCUNA GARANZIA PER IL PROGRAMMA, NEI LIMITI PERMESSI DALLE VIGENTI LEGGI. SE NON INDICATO DIVERSAMENTE PER ISCRITTO, IL DETENTORE DEL COPYRIGHT E LE ALTRE PARTI FORNISCONO IL PROGRAMMA "COSÌ COM'È", SENZA ALCUN TIPO DI GARANZIA, NÉ ESPlicitA NÉ IMPLICITa; CIÒ COMPRENDE, SENZA LIMITARSI A QUESTO, LA GARANZIA IMPLICITa DI COMMERCIALIZZABILITÀ E UTILIZZABILITÀ PER UN PARTICOLARE SCOPO. L'INTERO RISCHIO CONCERNENTE LA QUALITÀ E LE PRESTAZIONI DEL PROGRAMMA È DELL'ACQUIRENTE. SE IL PROGRAMMA DOVESSE RIVELARSI DIFETTOSO, L'ACQUIRENTE SI ASSUME IL COSTO DI OGNI MANUTENZIONE, RIPARAZIONE O CORREZIONE NECESSARIA.
12. NÉ IL DETENTORE DEL COPYRIGHT NÉ ALTRE PARTI CHE POSSONO MODIFICARE O RIDISTRIBUIRE IL PROGRAMMA COME PERMESSO IN QUESTA LICENZA SONO RESPONSABILI PER DANNI NEI CONFRONTI DELL'ACQUIRENTE, A MENO CHE QUESTO NON SIA RICHIESTO DALLE LEGGI VIGENTI O APPAIA IN UN ACCORDO SCRITTO. SONO INCLUSI DANNI GENERICI, SPECIALI O INCIDENTALI, COME PURE I DANNI CHE CONSEGUONO DALL'USO O DALL'IMPOSSIBILITÀ DI USARE IL PROGRAMMA; CIÒ COMPRENDE, SENZA LIMITARSI A QUESTO, LA PERDITA DI DATI, LA CORRUZIONE DEI DATI, LE PERDITE SOSTENUTE DALL'ACQUIRENTE O DA TERZE PARTI E L'INABILITÀ DEL PROGRAMMA A LAVORARE INSIEME AD ALTRI PROGRAMMI, ANCHE SE IL DETENTORE O ALTRE PARTI SONO STATE AVVISATE DELLA POSSIBILITÀ DI QUESTI DANNI.

FINE DEI TERMINI E DELLE CONDIZIONI

Appendice: come applicare questi termini ai nuovi programmi

Se si sviluppa un nuovo programma e lo si vuole rendere della maggiore utilità possibile per il pubblico, la cosa migliore da fare è fare sì che divenga software libero, cosicché ciascuno possa ridistribuirlo e modificarlo secondo questi termini.

Per fare questo, si inserisca nel programma la seguente nota. La cosa migliore da fare è mettere la nota all'inizio di ogni file sorgente, per chiarire nel modo più efficace possibile l'assenza di garanzia; ogni file dovrebbe contenere almeno la nota di diritto d'autore e l'indicazione di dove trovare l'intera nota.

<una riga per dire in breve il nome del programma e cosa fa>
Copyright (C) <anno> <nome dell'autore>

Questo programma è software libero; è lecito ridistribuirlo e/o modificarlo secondo i termini della Licenza Pubblica Generica GNU come pubblicata dalla Free Software Foundation; o la versione 2 della licenza o (a scelta) una versione successiva.

Questo programma è distribuito nella speranza che sia utile, ma SENZA ALCUNA GARANZIA; senza neppure la garanzia implicita di COMMERCIALIZZABILITÀ o di APPLICABILITÀ PER UN PARTICOLARE SCOPO. Si veda la Licenza Pubblica Generica GNU per avere maggiori dettagli.

Ognuno dovrebbe avere ricevuto una copia della Licenza Pubblica Generica GNU insieme a questo programma; in caso contrario, la si può ottenere dalla Free Software Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111-1307, Stati Uniti.

Si aggiungano anche informazioni su come si può essere contattati tramite posta elettronica e cartacea.

Se il programma è interattivo, si faccia in modo che stampi una breve nota simile a questa quando viene usato interattivamente:

Gnomovision versione 69, Copyright (C) <anno> <nome dell'autore>
Gnomovision non ha ALCUNA GARANZIA; per i dettagli digitare 'show w'.

Questo è software libero, e ognuno è libero di ridistribuirlo sotto certe condizioni; digitare 'show c' per dettagli.

Gli ipotetici comandi `show w` e `show c` mostreranno le parti appropriate della Licenza Pubblica Generica. Chiaramente, i comandi usati possono essere chiamati diversamente da `show w` e `show c` e possono anche essere selezionati con il mouse o attraverso un menù; in qualunque modo pertinente al programma.

Se necessario, si dovrebbe anche far firmare al proprio datore di lavoro (se si lavora come programmatore) o alla propria scuola, se si è studente, una "rinuncia ai diritti" per il programma. Ecco un esempio con nomi fittizi:

Yoyodyne, Inc. rinuncia con questo documento a ogni rivendicazione di
diritti d'autore sul programma
'Gnomovision' (che fa il primo passo con i compilatori) scritto da James
Hacker.

<Firma di Ty Coon>, 1 Aprile 1989
Ty Coon, Presidente di Yoyodyne, Inc.

I programmi coperti da questa Licenza Pubblica Generica non possono essere incorporati all'interno di programmi non liberi. Se il proprio programma è una libreria di funzioni, può essere più utile permettere di collegare applicazioni proprietarie alla libreria. In questo caso consigliamo di usare la Licenza Generica Pubblica GNU per Librerie (LGPL) al posto di questa Licenza.

Appendice G

Tabelle

Nelle pagine seguenti sono riportati alcuni valori tabulati relativi alle distribuzioni normale, di Student, del χ^2 e di Fisher.

Per le tabelle della distribuzione normale, per i valori dell'ascissa compresi tra 0 e 4 sono state calcolate sia l'ordinata della funzione di Gauss standardizzata

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

che i valori I_1 ed I_2 di due differenti funzioni integrali:

$$I_1 = \int_{-x}^x f(t) dt \quad \text{e} \quad I_2 = \int_{-\infty}^x f(t) dt .$$

Per la distribuzione di Student, facendo variare il numero di gradi di libertà N (nelle righe della tabella) da 1 a 40, sono riportati i valori dell'ascissa x che corrispondono a differenti aree P (nelle colonne della tabella): in modo che, indicando con $S(t)$ la funzione di frequenza di Student,

$$P = \int_{-\infty}^x S(t) dt .$$

Per la distribuzione del χ^2 , poi, e sempre per diversi valori di N , sono riportati i valori dell'ascissa x corrispondenti ad aree determinate P , così che (indicando con $C(t)$ la funzione di frequenza del χ^2) risulti

$$P = \int_0^x C(t) dt .$$

Per la distribuzione di Fisher, infine, per i soli due valori 0.95 e 0.99 del livello di confidenza P , sono riportati (per differenti gradi di libertà M ed

N) le ascisse x che corrispondono ad aree uguali al livello di confidenza prescelto; ossia (indicando con $F(w)$ la funzione di Fisher) tali che

$$P = \int_0^x F(w) dw . \quad (G.1)$$

Per calcolare i numeri riportati in queste tabelle si è usato un programma in linguaggio C che si serve delle costanti matematiche e delle procedure di calcolo numerico della *GNU Scientific Library* (GSL); chi volesse maggiori informazioni al riguardo le può trovare sul sito web della *Free Software Foundation*, sotto la URL <http://www.gnu.org/software/gsl/>.

La GSL contiene procedure per il calcolo numerico sia delle funzioni di frequenza che di quelle cumulative per tutte le funzioni considerate in questa appendice; e per tutte, meno che per la funzione di Fisher, anche procedure per invertire le distribuzioni cumulative. Per trovare l'ascissa x per cui l'integrale (G.1) raggiunge un valore prefissato si è usato il pacchetto della GSL che permette di trovare gli zeri di una funzione definita dall'utente in un intervallo arbitrario.

x	y	I_1	I_2	x	y	I_1	I_2
0.00	0.39894	0.00000	0.50000	0.40	0.36827	0.31084	0.65542
0.01	0.39892	0.00798	0.50399	0.41	0.36678	0.31819	0.65910
0.02	0.39886	0.01596	0.50798	0.42	0.36526	0.32551	0.66276
0.03	0.39876	0.02393	0.51197	0.43	0.36371	0.33280	0.66640
0.04	0.39862	0.03191	0.51595	0.44	0.36213	0.34006	0.67003
0.05	0.39844	0.03988	0.51994	0.45	0.36053	0.34729	0.67364
0.06	0.39822	0.04784	0.52392	0.46	0.35889	0.35448	0.67724
0.07	0.39797	0.05581	0.52790	0.47	0.35723	0.36164	0.68082
0.08	0.39767	0.06376	0.53188	0.48	0.35553	0.36877	0.68439
0.09	0.39733	0.07171	0.53586	0.49	0.35381	0.37587	0.68793
0.10	0.39695	0.07966	0.53983	0.50	0.35207	0.38292	0.69146
0.11	0.39654	0.08759	0.54380	0.51	0.35029	0.38995	0.69497
0.12	0.39608	0.09552	0.54776	0.52	0.34849	0.39694	0.69847
0.13	0.39559	0.10343	0.55172	0.53	0.34667	0.40389	0.70194
0.14	0.39505	0.11134	0.55567	0.54	0.34482	0.41080	0.70540
0.15	0.39448	0.11924	0.55962	0.55	0.34294	0.41768	0.70884
0.16	0.39387	0.12712	0.56356	0.56	0.34105	0.42452	0.71226
0.17	0.39322	0.13499	0.56749	0.57	0.33912	0.43132	0.71566
0.18	0.39253	0.14285	0.57142	0.58	0.33718	0.43809	0.71904
0.19	0.39181	0.15069	0.57535	0.59	0.33521	0.44481	0.72240
0.20	0.39104	0.15852	0.57926	0.60	0.33322	0.45149	0.72575
0.21	0.39024	0.16633	0.58317	0.61	0.33121	0.45814	0.72907
0.22	0.38940	0.17413	0.58706	0.62	0.32918	0.46474	0.73237
0.23	0.38853	0.18191	0.59095	0.63	0.32713	0.47131	0.73565
0.24	0.38762	0.18967	0.59483	0.64	0.32506	0.47783	0.73891
0.25	0.38667	0.19741	0.59871	0.65	0.32297	0.48431	0.74215
0.26	0.38568	0.20514	0.60257	0.66	0.32086	0.49075	0.74537
0.27	0.38466	0.21284	0.60642	0.67	0.31874	0.49714	0.74857
0.28	0.38361	0.22052	0.61026	0.68	0.31659	0.50350	0.75175
0.29	0.38251	0.22818	0.61409	0.69	0.31443	0.50981	0.75490
0.30	0.38139	0.23582	0.61791	0.70	0.31225	0.51607	0.75804
0.31	0.38023	0.24344	0.62172	0.71	0.31006	0.52230	0.76115
0.32	0.37903	0.25103	0.62552	0.72	0.30785	0.52848	0.76424
0.33	0.37780	0.25860	0.62930	0.73	0.30563	0.53461	0.76730
0.34	0.37654	0.26614	0.63307	0.74	0.30339	0.54070	0.77035
0.35	0.37524	0.27366	0.63683	0.75	0.30114	0.54675	0.77337
0.36	0.37391	0.28115	0.64058	0.76	0.29887	0.55275	0.77637
0.37	0.37255	0.28862	0.64431	0.77	0.29659	0.55870	0.77935
0.38	0.37115	0.29605	0.64803	0.78	0.29431	0.56461	0.78230
0.39	0.36973	0.30346	0.65173	0.79	0.29200	0.57047	0.78524

Tabelle della distribuzione normale (I)

x	y	I_1	I_2	x	y	I_1	I_2
0.80	0.28969	0.57629	0.78814	1.20	0.19419	0.76986	0.88493
0.81	0.28737	0.58206	0.79103	1.21	0.19186	0.77372	0.88686
0.82	0.28504	0.58778	0.79389	1.22	0.18954	0.77754	0.88877
0.83	0.28269	0.59346	0.79673	1.23	0.18724	0.78130	0.89065
0.84	0.28034	0.59909	0.79955	1.24	0.18494	0.78502	0.89251
0.85	0.27798	0.60467	0.80234	1.25	0.18265	0.78870	0.89435
0.86	0.27562	0.61021	0.80511	1.26	0.18037	0.79233	0.89617
0.87	0.27324	0.61570	0.80785	1.27	0.17810	0.79592	0.89796
0.88	0.27086	0.62114	0.81057	1.28	0.17585	0.79945	0.89973
0.89	0.26848	0.62653	0.81327	1.29	0.17360	0.80295	0.90147
0.90	0.26609	0.63188	0.81594	1.30	0.17137	0.80640	0.90320
0.91	0.26369	0.63718	0.81859	1.31	0.16915	0.80980	0.90490
0.92	0.26129	0.64243	0.82121	1.32	0.16694	0.81316	0.90658
0.93	0.25888	0.64763	0.82381	1.33	0.16474	0.81648	0.90824
0.94	0.25647	0.65278	0.82639	1.34	0.16256	0.81975	0.90988
0.95	0.25406	0.65789	0.82894	1.35	0.16038	0.82298	0.91149
0.96	0.25164	0.66294	0.83147	1.36	0.15822	0.82617	0.91309
0.97	0.24923	0.66795	0.83398	1.37	0.15608	0.82931	0.91466
0.98	0.24681	0.67291	0.83646	1.38	0.15395	0.83241	0.91621
0.99	0.24439	0.67783	0.83891	1.39	0.15183	0.83547	0.91774
1.00	0.24197	0.68269	0.84134	1.40	0.14973	0.83849	0.91924
1.01	0.23955	0.68750	0.84375	1.41	0.14764	0.84146	0.92073
1.02	0.23713	0.69227	0.84614	1.42	0.14556	0.84439	0.92220
1.03	0.23471	0.69699	0.84849	1.43	0.14350	0.84728	0.92364
1.04	0.23230	0.70166	0.85083	1.44	0.14146	0.85013	0.92507
1.05	0.22988	0.70628	0.85314	1.45	0.13943	0.85294	0.92647
1.06	0.22747	0.71086	0.85543	1.46	0.13742	0.85571	0.92785
1.07	0.22506	0.71538	0.85769	1.47	0.13542	0.85844	0.92922
1.08	0.22265	0.71986	0.85993	1.48	0.13344	0.86113	0.93056
1.09	0.22025	0.72429	0.86214	1.49	0.13147	0.86378	0.93189
1.10	0.21785	0.72867	0.86433	1.50	0.12952	0.86639	0.93319
1.11	0.21546	0.73300	0.86650	1.51	0.12758	0.86896	0.93448
1.12	0.21307	0.73729	0.86864	1.52	0.12566	0.87149	0.93574
1.13	0.21069	0.74152	0.87076	1.53	0.12376	0.87398	0.93699
1.14	0.20831	0.74571	0.87286	1.54	0.12188	0.87644	0.93822
1.15	0.20594	0.74986	0.87493	1.55	0.12001	0.87886	0.93943
1.16	0.20357	0.75395	0.87698	1.56	0.11816	0.88124	0.94062
1.17	0.20121	0.75800	0.87900	1.57	0.11632	0.88358	0.94179
1.18	0.19886	0.76200	0.88100	1.58	0.11450	0.88589	0.94295
1.19	0.19652	0.76595	0.88298	1.59	0.11270	0.88817	0.94408

Tabelle della distribuzione normale (II)

x	y	I_1	I_2	x	y	I_1	I_2
1.60	0.11092	0.89040	0.94520	2.00	0.05399	0.95450	0.97725
1.61	0.10915	0.89260	0.94630	2.01	0.05292	0.95557	0.97778
1.62	0.10741	0.89477	0.94738	2.02	0.05186	0.95662	0.97831
1.63	0.10567	0.89690	0.94845	2.03	0.05082	0.95764	0.97882
1.64	0.10396	0.89899	0.94950	2.04	0.04980	0.95865	0.97932
1.65	0.10226	0.90106	0.95053	2.05	0.04879	0.95964	0.97982
1.66	0.10059	0.90309	0.95154	2.06	0.04780	0.96060	0.98030
1.67	0.09893	0.90508	0.95254	2.07	0.04682	0.96155	0.98077
1.68	0.09728	0.90704	0.95352	2.08	0.04586	0.96247	0.98124
1.69	0.09566	0.90897	0.95449	2.09	0.04491	0.96338	0.98169
1.70	0.09405	0.91087	0.95543	2.10	0.04398	0.96427	0.98214
1.71	0.09246	0.91273	0.95637	2.11	0.04307	0.96514	0.98257
1.72	0.09089	0.91457	0.95728	2.12	0.04217	0.96599	0.98300
1.73	0.08933	0.91637	0.95818	2.13	0.04128	0.96683	0.98341
1.74	0.08780	0.91814	0.95907	2.14	0.04041	0.96765	0.98382
1.75	0.08628	0.91988	0.95994	2.15	0.03955	0.96844	0.98422
1.76	0.08478	0.92159	0.96080	2.16	0.03871	0.96923	0.98461
1.77	0.08329	0.92327	0.96164	2.17	0.03788	0.96999	0.98500
1.78	0.08183	0.92492	0.96246	2.18	0.03706	0.97074	0.98537
1.79	0.08038	0.92655	0.96327	2.19	0.03626	0.97148	0.98574
1.80	0.07895	0.92814	0.96407	2.20	0.03547	0.97219	0.98610
1.81	0.07754	0.92970	0.96485	2.21	0.03470	0.97289	0.98645
1.82	0.07614	0.93124	0.96562	2.22	0.03394	0.97358	0.98679
1.83	0.07477	0.93275	0.96638	2.23	0.03319	0.97425	0.98713
1.84	0.07341	0.93423	0.96712	2.24	0.03246	0.97491	0.98745
1.85	0.07206	0.93569	0.96784	2.25	0.03174	0.97555	0.98778
1.86	0.07074	0.93711	0.96856	2.26	0.03103	0.97618	0.98809
1.87	0.06943	0.93852	0.96926	2.27	0.03034	0.97679	0.98840
1.88	0.06814	0.93989	0.96995	2.28	0.02965	0.97739	0.98870
1.89	0.06687	0.94124	0.97062	2.29	0.02898	0.97798	0.98899
1.90	0.06562	0.94257	0.97128	2.30	0.02833	0.97855	0.98928
1.91	0.06438	0.94387	0.97193	2.31	0.02768	0.97911	0.98956
1.92	0.06316	0.94514	0.97257	2.32	0.02705	0.97966	0.98983
1.93	0.06195	0.94639	0.97320	2.33	0.02643	0.98019	0.99010
1.94	0.06077	0.94762	0.97381	2.34	0.02582	0.98072	0.99036
1.95	0.05959	0.94882	0.97441	2.35	0.02522	0.98123	0.99061
1.96	0.05844	0.95000	0.97500	2.36	0.02463	0.98173	0.99086
1.97	0.05730	0.95116	0.97558	2.37	0.02406	0.98221	0.99111
1.98	0.05618	0.95230	0.97615	2.38	0.02349	0.98269	0.99134
1.99	0.05508	0.95341	0.97670	2.39	0.02294	0.98315	0.99158

Tabelle della distribuzione normale (III)

x	y	I_1	I_2	x	y	I_1	I_2
2.40	0.02239	0.98360	0.99180	2.80	0.00792	0.99489	0.99744
2.41	0.02186	0.98405	0.99202	2.81	0.00770	0.99505	0.99752
2.42	0.02134	0.98448	0.99224	2.82	0.00748	0.99520	0.99760
2.43	0.02083	0.98490	0.99245	2.83	0.00727	0.99535	0.99767
2.44	0.02033	0.98531	0.99266	2.84	0.00707	0.99549	0.99774
2.45	0.01984	0.98571	0.99286	2.85	0.00687	0.99563	0.99781
2.46	0.01936	0.98611	0.99305	2.86	0.00668	0.99576	0.99788
2.47	0.01888	0.98649	0.99324	2.87	0.00649	0.99590	0.99795
2.48	0.01842	0.98686	0.99343	2.88	0.00631	0.99602	0.99801
2.49	0.01797	0.98723	0.99361	2.89	0.00613	0.99615	0.99807
2.50	0.01753	0.98758	0.99379	2.90	0.00595	0.99627	0.99813
2.51	0.01709	0.98793	0.99396	2.91	0.00578	0.99639	0.99819
2.52	0.01667	0.98826	0.99413	2.92	0.00562	0.99650	0.99825
2.53	0.01625	0.98859	0.99430	2.93	0.00545	0.99661	0.99831
2.54	0.01585	0.98891	0.99446	2.94	0.00530	0.99672	0.99836
2.55	0.01545	0.98923	0.99461	2.95	0.00514	0.99682	0.99841
2.56	0.01506	0.98953	0.99477	2.96	0.00499	0.99692	0.99846
2.57	0.01468	0.98983	0.99492	2.97	0.00485	0.99702	0.99851
2.58	0.01431	0.99012	0.99506	2.98	0.00470	0.99712	0.99856
2.59	0.01394	0.99040	0.99520	2.99	0.00457	0.99721	0.99861
2.60	0.01358	0.99068	0.99534	3.00	0.00443	0.99730	0.99865
2.61	0.01323	0.99095	0.99547	3.01	0.00430	0.99739	0.99869
2.62	0.01289	0.99121	0.99560	3.02	0.00417	0.99747	0.99874
2.63	0.01256	0.99146	0.99573	3.03	0.00405	0.99755	0.99878
2.64	0.01223	0.99171	0.99585	3.04	0.00393	0.99763	0.99882
2.65	0.01191	0.99195	0.99598	3.05	0.00381	0.99771	0.99886
2.66	0.01160	0.99219	0.99609	3.06	0.00370	0.99779	0.99889
2.67	0.01130	0.99241	0.99621	3.07	0.00358	0.99786	0.99893
2.68	0.01100	0.99264	0.99632	3.08	0.00348	0.99793	0.99896
2.69	0.01071	0.99285	0.99643	3.09	0.00337	0.99800	0.99900
2.70	0.01042	0.99307	0.99653	3.10	0.00327	0.99806	0.99903
2.71	0.01014	0.99327	0.99664	3.11	0.00317	0.99813	0.99906
2.72	0.00987	0.99347	0.99674	3.12	0.00307	0.99819	0.99910
2.73	0.00961	0.99367	0.99683	3.13	0.00298	0.99825	0.99913
2.74	0.00935	0.99386	0.99693	3.14	0.00288	0.99831	0.99916
2.75	0.00909	0.99404	0.99702	3.15	0.00279	0.99837	0.99918
2.76	0.00885	0.99422	0.99711	3.16	0.00271	0.99842	0.99921
2.77	0.00861	0.99439	0.99720	3.17	0.00262	0.99848	0.99924
2.78	0.00837	0.99456	0.99728	3.18	0.00254	0.99853	0.99926
2.79	0.00814	0.99473	0.99736	3.19	0.00246	0.99858	0.99929

Tabelle della distribuzione normale (IV)

x	y	I_1	I_2	x	y	I_1	I_2
3.20	0.00238	0.99863	0.99931	3.60	0.00061	0.99968	0.99984
3.21	0.00231	0.99867	0.99934	3.61	0.00059	0.99969	0.99985
3.22	0.00224	0.99872	0.99936	3.62	0.00057	0.99971	0.99985
3.23	0.00216	0.99876	0.99938	3.63	0.00055	0.99972	0.99986
3.24	0.00210	0.99880	0.99940	3.64	0.00053	0.99973	0.99986
3.25	0.00203	0.99885	0.99942	3.65	0.00051	0.99974	0.99987
3.26	0.00196	0.99889	0.99944	3.66	0.00049	0.99975	0.99987
3.27	0.00190	0.99892	0.99946	3.67	0.00047	0.99976	0.99988
3.28	0.00184	0.99896	0.99948	3.68	0.00046	0.99977	0.99988
3.29	0.00178	0.99900	0.99950	3.69	0.00044	0.99978	0.99989
3.30	0.00172	0.99903	0.99952	3.70	0.00042	0.99978	0.99989
3.31	0.00167	0.99907	0.99953	3.71	0.00041	0.99979	0.99990
3.32	0.00161	0.99910	0.99955	3.72	0.00039	0.99980	0.99990
3.33	0.00156	0.99913	0.99957	3.73	0.00038	0.99981	0.99990
3.34	0.00151	0.99916	0.99958	3.74	0.00037	0.99982	0.99991
3.35	0.00146	0.99919	0.99960	3.75	0.00035	0.99982	0.99991
3.36	0.00141	0.99922	0.99961	3.76	0.00034	0.99983	0.99992
3.37	0.00136	0.99925	0.99962	3.77	0.00033	0.99984	0.99992
3.38	0.00132	0.99928	0.99964	3.78	0.00031	0.99984	0.99992
3.39	0.00127	0.99930	0.99965	3.79	0.00030	0.99985	0.99992
3.40	0.00123	0.99933	0.99966	3.80	0.00029	0.99986	0.99993
3.41	0.00119	0.99935	0.99968	3.81	0.00028	0.99986	0.99993
3.42	0.00115	0.99937	0.99969	3.82	0.00027	0.99987	0.99993
3.43	0.00111	0.99940	0.99970	3.83	0.00026	0.99987	0.99994
3.44	0.00107	0.99942	0.99971	3.84	0.00025	0.99988	0.99994
3.45	0.00104	0.99944	0.99972	3.85	0.00024	0.99988	0.99994
3.46	0.00100	0.99946	0.99973	3.86	0.00023	0.99989	0.99994
3.47	0.00097	0.99948	0.99974	3.87	0.00022	0.99989	0.99995
3.48	0.00094	0.99950	0.99975	3.88	0.00021	0.99990	0.99995
3.49	0.00090	0.99952	0.99976	3.89	0.00021	0.99990	0.99995
3.50	0.00087	0.99953	0.99977	3.90	0.00020	0.99990	0.99995
3.51	0.00084	0.99955	0.99978	3.91	0.00019	0.99991	0.99995
3.52	0.00081	0.99957	0.99978	3.92	0.00018	0.99991	0.99996
3.53	0.00079	0.99958	0.99979	3.93	0.00018	0.99992	0.99996
3.54	0.00076	0.99960	0.99980	3.94	0.00017	0.99992	0.99996
3.55	0.00073	0.99961	0.99981	3.95	0.00016	0.99992	0.99996
3.56	0.00071	0.99963	0.99981	3.96	0.00016	0.99993	0.99996
3.57	0.00068	0.99964	0.99982	3.97	0.00015	0.99993	0.99996
3.58	0.00066	0.99966	0.99983	3.98	0.00014	0.99993	0.99997
3.59	0.00063	0.99967	0.99983	3.99	0.00014	0.99993	0.99997

Tabelle della distribuzione normale (V)

N	Probabilità (in percentuale)									
	99.9	99.8	99.5	99.0	98.0	95.0	90.0	80.0	75.0	60.0
1	318.309	159.153	63.657	31.821	15.895	6.314	3.078	1.376	1.000	0.325
2	22.327	15.764	9.925	6.965	4.849	2.920	1.886	1.061	0.816	0.289
3	10.215	8.053	5.841	4.541	3.482	2.353	1.638	0.978	0.765	0.277
4	7.173	5.951	4.604	3.747	2.999	2.132	1.533	0.941	0.741	0.271
5	5.893	5.030	4.032	3.365	2.757	2.015	1.476	0.920	0.727	0.267
6	5.208	4.524	3.707	3.143	2.612	1.943	1.440	0.906	0.718	0.265
7	4.785	4.207	3.499	2.998	2.517	1.895	1.415	0.896	0.711	0.263
8	4.501	3.991	3.355	2.896	2.449	1.860	1.397	0.889	0.706	0.262
9	4.297	3.835	3.250	2.821	2.398	1.833	1.383	0.883	0.703	0.261
10	4.144	3.716	3.169	2.764	2.359	1.812	1.372	0.879	0.700	0.260
11	4.025	3.624	3.106	2.718	2.328	1.796	1.363	0.876	0.697	0.260
12	3.930	3.550	3.055	2.681	2.303	1.782	1.356	0.873	0.695	0.259
13	3.852	3.489	3.012	2.650	2.282	1.771	1.350	0.870	0.694	0.259
14	3.787	3.438	2.977	2.624	2.264	1.761	1.345	0.868	0.692	0.258
15	3.733	3.395	2.947	2.602	2.249	1.753	1.341	0.866	0.691	0.258
16	3.686	3.358	2.921	2.583	2.235	1.746	1.337	0.865	0.690	0.258
17	3.646	3.326	2.898	2.567	2.224	1.740	1.333	0.863	0.689	0.257
18	3.610	3.298	2.878	2.552	2.214	1.734	1.330	0.862	0.688	0.257
19	3.579	3.273	2.861	2.539	2.205	1.729	1.328	0.861	0.688	0.257
20	3.552	3.251	2.845	2.528	2.197	1.725	1.325	0.860	0.687	0.257
21	3.527	3.231	2.831	2.518	2.189	1.721	1.323	0.859	0.686	0.257
22	3.505	3.214	2.819	2.508	2.183	1.717	1.321	0.858	0.686	0.256
23	3.485	3.198	2.807	2.500	2.177	1.714	1.319	0.858	0.685	0.256
24	3.467	3.183	2.797	2.492	2.172	1.711	1.318	0.857	0.685	0.256
25	3.450	3.170	2.787	2.485	2.167	1.708	1.316	0.856	0.684	0.256
26	3.435	3.158	2.779	2.479	2.162	1.706	1.315	0.856	0.684	0.256
27	3.421	3.147	2.771	2.473	2.158	1.703	1.314	0.855	0.684	0.256
28	3.408	3.136	2.763	2.467	2.154	1.701	1.313	0.855	0.683	0.256
29	3.396	3.127	2.756	2.462	2.150	1.699	1.311	0.854	0.683	0.256
30	3.385	3.118	2.750	2.457	2.147	1.697	1.310	0.854	0.683	0.256
31	3.375	3.109	2.744	2.453	2.144	1.696	1.309	0.853	0.682	0.256
32	3.365	3.102	2.738	2.449	2.141	1.694	1.309	0.853	0.682	0.255
33	3.356	3.094	2.733	2.445	2.138	1.692	1.308	0.853	0.682	0.255
34	3.348	3.088	2.728	2.441	2.136	1.691	1.307	0.852	0.682	0.255
35	3.340	3.081	2.724	2.438	2.133	1.690	1.306	0.852	0.682	0.255
36	3.333	3.075	2.719	2.434	2.131	1.688	1.306	0.852	0.681	0.255
37	3.326	3.070	2.715	2.431	2.129	1.687	1.305	0.851	0.681	0.255
38	3.319	3.064	2.712	2.429	2.127	1.686	1.304	0.851	0.681	0.255
39	3.313	3.059	2.708	2.426	2.125	1.685	1.304	0.851	0.681	0.255
40	3.307	3.055	2.704	2.423	2.123	1.684	1.303	0.851	0.681	0.255

Percentili della distribuzione di Student

N	Probabilità (in percentuale)												
	99.9	99.5	99.0	95.0	90.0	75.0	50.0	25.0	10.0	5.0	1.0	0.5	0.1
1	10.83	7.88	6.63	3.84	2.71	1.32	0.45	0.10	0.02	0.00	0.00	0.00	0.00
2	13.82	10.60	9.21	5.99	4.61	2.77	1.39	0.58	0.21	0.10	0.02	0.01	0.00
3	16.27	12.84	11.34	7.81	6.25	4.11	2.37	1.21	0.58	0.35	0.11	0.07	0.02
4	18.47	14.86	13.28	9.49	7.78	5.39	3.36	1.92	1.06	0.71	0.30	0.21	0.09
5	20.52	16.75	15.09	11.07	9.24	6.63	4.35	2.67	1.61	1.15	0.55	0.41	0.21
6	22.46	18.55	16.81	12.59	10.64	7.84	5.35	3.45	2.20	1.64	0.87	0.68	0.38
7	24.32	20.28	18.48	14.07	12.02	9.04	6.35	4.25	2.83	2.17	1.24	0.99	0.60
8	26.12	21.95	20.09	15.51	13.36	10.22	7.34	5.07	3.49	2.73	1.65	1.34	0.86
9	27.88	23.59	21.67	16.92	14.68	11.39	8.34	5.90	4.17	3.33	2.09	1.73	1.15
10	29.59	25.19	23.21	18.31	15.99	12.55	9.34	6.74	4.87	3.94	2.56	2.16	1.48
11	31.26	26.76	24.72	19.68	17.28	13.70	10.34	7.58	5.58	4.57	3.05	2.60	1.83
12	32.91	28.30	26.22	21.03	18.55	14.85	11.34	8.44	6.30	5.23	3.57	3.07	2.21
13	34.53	29.82	27.69	22.36	19.81	15.98	12.34	9.30	7.04	5.89	4.11	3.57	2.62
14	36.12	31.32	29.14	23.68	21.06	17.12	13.34	10.17	7.79	6.57	4.66	4.07	3.04
15	37.70	32.80	30.58	25.00	22.31	18.25	14.34	11.04	8.55	7.26	5.23	4.60	3.48
16	39.25	34.27	32.00	26.30	23.54	19.37	15.34	11.91	9.31	7.96	5.81	5.14	3.94
17	40.79	35.72	33.41	27.59	24.77	20.49	16.34	12.79	10.09	8.67	6.41	5.70	4.42
18	42.31	37.16	34.81	28.87	25.99	21.60	17.34	13.68	10.86	9.39	7.01	6.26	4.90
19	43.82	38.58	36.19	30.14	27.20	22.72	18.34	14.56	11.65	10.12	7.63	6.84	5.41
20	45.31	40.00	37.57	31.41	28.41	23.83	19.34	15.45	12.44	10.85	8.26	7.43	5.92
21	46.80	41.40	38.93	32.67	29.62	24.93	20.34	16.34	13.24	11.59	8.90	8.03	6.45
22	48.27	42.80	40.29	33.92	30.81	26.04	21.34	17.24	14.04	12.34	9.54	8.64	6.98
23	49.73	44.18	41.64	35.17	32.01	27.14	22.34	18.14	14.85	13.09	10.20	9.26	7.53
24	51.18	45.56	42.98	36.42	33.20	28.24	23.34	19.04	15.66	13.85	10.86	9.89	8.08
25	52.62	46.93	44.31	37.65	34.38	29.34	24.34	19.94	16.47	14.61	11.52	10.52	8.65
26	54.05	48.29	45.64	38.89	35.56	30.43	25.34	20.84	17.29	15.38	12.20	11.16	9.22
27	55.48	49.64	46.96	40.11	36.74	31.53	26.34	21.75	18.11	16.15	12.88	11.81	9.80
28	56.89	50.99	48.28	41.34	37.92	32.62	27.34	22.66	18.94	16.93	13.56	12.46	10.39
29	58.30	52.34	49.59	42.56	39.09	33.71	28.34	23.57	19.77	17.71	14.26	13.12	10.99
30	59.70	53.67	50.89	43.77	40.26	34.80	29.34	24.48	20.60	18.49	14.95	13.79	11.59

Percentili della distribuzione del χ^2

N	M											
	1	2	3	4	5	6	8	12	24	36	48	∞
1	161.45	199.50	215.71	224.58	230.16	233.99	238.88	243.91	249.05	250.79	251.67	254.31
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.47	19.47	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.85	8.74	8.64	8.60	8.58	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.73	5.70	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.47	4.45	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.79	3.76	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.35	3.32	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	3.06	3.02	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.84	2.81	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.67	2.64	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.54	2.51	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.85	2.69	2.51	2.44	2.41	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.77	2.60	2.42	2.35	2.32	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.28	2.24	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.22	2.18	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.17	2.13	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	2.12	2.08	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	2.08	2.04	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	2.04	2.00	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	2.01	1.97	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.98	1.94	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.95	1.91	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.37	2.20	2.01	1.93	1.89	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.91	1.87	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.89	1.85	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.81	1.77	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.71	1.67	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.10	1.92	1.70	1.61	1.57	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.02	1.83	1.61	1.52	1.46	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	1.94	1.75	1.52	1.42	1.36	1.00

Percentili della distribuzione di Fisher per $P = 0.95$

N	M											
	1	2	3	4	5	6	8	12	24	36	48	∞
1	4052.02	4999.47	5403.35	5624.58	5763.65	5858.99	5981.32	6106.44	6234.68	6278.10	6299.92	6365.88
2	98.50	99.00	99.17	99.25	99.30	99.33	99.37	99.42	99.46	99.47	99.48	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.49	27.05	26.60	26.44	26.36	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37	13.93	13.78	13.70	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.29	9.89	9.47	9.32	9.25	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.31	7.17	7.10	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.47	6.07	5.94	5.87	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	5.14	5.07	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.73	4.59	4.53	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.33	4.19	4.12	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.89	3.82	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.65	3.58	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.30	3.96	3.59	3.45	3.38	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.14	3.80	3.43	3.29	3.22	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	3.16	3.09	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	3.05	2.98	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.46	3.08	2.95	2.88	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.86	2.79	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.79	2.72	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.72	2.65	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.66	2.59	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.75	2.61	2.54	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.70	2.56	2.49	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.66	2.52	2.45	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.32	2.99	2.62	2.48	2.41	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.33	2.25	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	2.14	2.07	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.97	1.89	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.80	1.71	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.63	1.54	1.00

Percentili della distribuzione di Fisher per $P = 0.99$

Appendice H

Bibliografia

Per approfondire:

1. Roger J. Barlow: *Statistics: a guide to the use of statistical methods in the physical sciences* – J. Wiley & Sons, 1997
2. G. Cowan: *Statistical data analysis* – Oxford University Press, 1998 (ISBN 0-19-850155-2)
3. H. Cramér: *Mathematical methods of statistics* – Princeton University Press, 1946
4. W.T. Eadie, D. Drijard, F.E. James, M. Roos e B. Sadoulet: *Statistical methods in experimental physics* – North-Holland Publishing Company, 1971 (ISBN 0-7204-0239-5)
5. W. Feller: *An introduction to probability theory and its applications (3rd Ed.)* – J. Wiley & Sons, 1970 (ISBN 0-471-25711-7)
6. R.A. Fisher: *Statistical methods for research workers* – Oliver & Boyd, 1954
7. H. Freeman: *Introduction to statistical inference* – Addison-Wesley, 1963
8. M.G. Kendall e A. Stuart: *The advanced theory of statistics* – Griffin & Co., 1958
9. W.H. Press, S.A. Teukolsky, W.T. Vetterling e B.P. Flannery: *Numerical recipes in C* – Cambridge University Press, 1992 (ISBN 0-521-43108-5)

10. M.R. Spiegel: *Statistica* – Collana “Schaum” – McGraw-Hill, 1961 (ISBN 88-386-5000-4)
11. J.R. Taylor: *Introduzione all’analisi degli errori* – Zanichelli, 1986 (ISBN 88-08-03292-2)
12. Particle Data Group: *Review of particle physics: reviews, tables, and plots - Mathematical tools* – <http://pdg.web.cern.ch/pdg/pdg.html>

*In realtà un lavoro simile non termina mai.
Lo si deve dichiarare concluso quando,
a seconda del tempo e delle circostanze,
si è fatto il possibile.*

*Johann Wolfgang von Goethe
Italienische Reise (1789)*

Indice analitico

— A —

arrotondamenti, *vedi* cifre
 significative
asimmetria, *vedi* coefficiente di
 asimmetria
asimmetria, rapporto di, 114-115,
 122-124

— B —

Bayes, teorema di, 25-26
Bernoulli
 Jacob (o Jacques), 109
 teorema di, 57-58
Bienaymé, Irénée-Jules, 55
Bienaymé-Čebyšef, disuguaglianza
 di, 55-56
binomiale negativa, distribuzione,
 vedi distribuzione binomiale
 negativa
binomiale, distribuzione, *vedi*
 distribuzione di Bernoulli
Box-Muller, metodo di, 157-159
branching ratio, 86

— C —

calcolo combinatorio, 243-247
 lemma fondamentale, 243
cambiamento di variabile casuale,
 77-78, 84-85, 90-91
campione, 47
 valori estremi, 78-79, 87-88,
 99-100

casuali
 eventi, 19-20
 variabili, 20, 48
Čebyšef
 Pafnuty Lvovič, 55
 teorema di, 57, 69
cifre significative, 17-18, 254
classi di frequenza, 32, 203-204, 210
coefficiente
 di asimmetria, 71
 di curtosi, 71
coefficienti binomiali, 245-246
combinazioni, 245-246
combinazioni lineari
 di variabili di Cauchy, 107
 di variabili normali, 103-104
 e loro correlazione, 258
 speranza matematica, 50-51
 varianza
 di variabili correlate, 256-258
 di variabili indipendenti, 51-54
compatibilità
 con un valore, 214-217
 con una distribuzione, 203-205,
 225-226
 tra dati sperimentali, 210-212,
 225-226
 tra valori misurati, 220-222
 tra varianze, 224
complementare, evento, 22
contingenze, tabella delle, 210

correlazione lineare, coefficiente di,
83, **259-260**, 263-265
covarianza, 83, **255-259**
 dei coefficienti della retta
 interpolante, 267
 matrice di, **257**, 261
Cramér-Rao, teorema di, 168,
 171-172, 177, 180, **284-289**
curtosi, *vedi* coefficiente di curtosi

— D —

de Moivre
 Abraham, 24, 141, 279
 e Stirling, formula di, 279
decili, *vedi* quantili
deviazione media assoluta, *vedi*
 errore medio
dimensioni (delle grandezze fisiche),
 6-7
dispersione massima, *vedi*
 semidispersione massima
disposizioni, 244
distribuzione
 binomiale negativa, 115-116
 del χ^2 , 195-203
 e distribuzione normale,
 198-199
 regola di somma, 199
 di Bernoulli, 108-113
 e distribuzione normale,
 110-113
 di Breit-Wigner, *vedi* distribuzione
 di Cauchy
 di Cauchy, 49, **104-107**
 di Erlang, 126-127
 di Fisher, 222-224
 di Maxwell-Boltzmann, 40
 di Poisson, 116-121
 composta, 128-129
 e distribuzione di Bernoulli, 118
 e distribuzione normale, 121
 di Student, **217-220**, 222, 264, 269
 e distribuzione normale, 219
 esponenziale, 124-126

geometrica, 116
log-normale, 132-133
normale, **101-104**, **141-149**,
 277-282
normale bidimensionale, 135-138
uniforme, **93-94**
 range, 100, 191-192

— E —

Erlang
 Agner Krarup, 126
 distribuzione di, *vedi* distribuzione
 di Erlang
errore
 a posteriori, **183-184**, 185, 264
 dell'errore stimato, 254
 della media, 45
 della varianza stimata, 253
 di prima specie, 228
 di seconda specie, 228
 massimo, 166
 medio, 43
 della distribuzione normale,
 144-146
 quadratico medio, 43-44
 della distribuzione normale,
 145-146, 147-149
 relativo, 18
errori di misura
 casuali, 12, **16-17**, 58
 sistematici, **12-16**, 58, 150
esame dei dati, **152-153**, 177-178

— F —

Fisher, sir Ronald Aylmer, **167**, 265
Fourier, trasformata di, 73, 196
frequenza
 assoluta, 21
 cumulativa, 33
 relativa, 21
funzione
 caratteristica, 72-73, 84
 di somme di variabili, 74-75
 per trasformazioni lineari, 78

per variabili discrete, 75-76
 di distribuzione, **68**, 81, 96
 di verosimiglianza, **170-171**,
 179-180, 229-242
 generatrice dei momenti, 71-72, 84
 per trasformazioni lineari, 78

■ G ■

Gauss, Karl Friedrich, 44, 56, **142**
 Gosset, William ("Student"), 195
 grandi numeri, legge dei, 28-29, 33,
 50, **55-58**

■ I ■

interpolazione lineare, **179-183**,
 186-187, 262-273
 con una retta per l'origine,
 184-185
 ipotesi
 alternativa, 228
 nulla, 228
 istogrammi, **31-33**, 65-67
 e curva normale, 150-151

■ J ■

Jacobiano
 determinante, 84-85
 Jacobiano, determinante, **90**, 158

■ K ■

Kolmogorov e Smirnov, test di,
 225-226
 Kolmogorov, Andrei Nikolaevich, 27,
 225
 Kronecker, delta di, 200

■ L ■

Laplace
 modello di, 277-282
 Pierre Simon de, 141, 142, **277**
 limite centrale, teorema del, 154-156
 limite debole, *vedi* statistica,
 convergenza

■ M ■

massima verosimiglianza, metodo
 della, **170-171**, 188-192,
 209, 288-289
 media
 aritmetica
 come stima del valore vero,
 40-41, 44-45, 58-59, 144,
 156, **178**, 289-290
 come stima di tendenza centrale,
 37-38
 e varianza, 191, **203**, 275
 proprietà matematiche, 37-38,
 41-42
 armonica, 38
 geometrica, 38
 pesata, 41, **176-177**
 quadratica, 38
 mediana, 35-37
 metodo
 dei rigetti, 96-99
 del χ^2 , 150
 del minimo χ^2 , 205-209
 del rapporto delle massime
 verosimiglianze, 235-237
 scientifico, 2-3
 minimi quadrati, formule dei, **181**,
 262
 misure
 dirette, 5
 indirette, 5, 161-165
 moda, 35
 modulo di precisione della misura,
142, 145-146
 momenti, 69-71, 83
 funzione generatrice, *vedi*
 funzione generatrice dei
 momenti
 ■ N ■
 Neyman-Pearson, lemma di, 232-234
 normalizzazione
 condizione di, **68**, 82, 89, 142

della funzione normale agli istogrammi, 150-151

— O —

omogeneità, test di, *vedi*
compatibilità tra dati sperimentali
one-tailed test, 217

— P —

partizioni ordinate, 188, 209, **246-247**
pendolo, periodo del, 14-15
percentili, *vedi* quantili
permutazioni, 245
con ripetizione, 245
piccoli campioni, 217-222
Planck, Max Karl Ernst Ludwig, 7
Poisson
distribuzione di, *vedi* distribuzione di Poisson
Siméon Denis, 118
popolazione, 47
potenza, 228
probabilità
composta (teorema della), 24, 28
condizionata, 24, 28, 82, 89
definizione
assiomatica, 27
classica, 20
empirica, 21
densità di, **65-68**, 81, 88-89
funzione marginale, 82, **89**
totale (teorema della), 23, 28
propagazione degli errori, formula di, **164-165**, 261
per errori massimi, 166
per prodotti di potenze, 165
pseudo-casuali, numeri, 96-99
con distribuzione normale, 157-159

— Q —

quantili, 42-43

quantili, *vedi* quantili

— R —

range, 42
rapporto di asimmetria, *vedi*
asimmetria, rapporto di
rapporto di variabili, 85-86
normali, 107-108
residui, **180**, 263, 269-272
run test, 270-273

— S —

scarto normalizzato, 146-147
semidispersione massima, 42
significanza, 228
Sistema Internazionale, 6
somma di un numero casuale di variabili discrete, 76
speranza matematica, 49
della media aritmetica, 51
di combinazioni lineari, 50-51
per variabili continue, **69**, 83
statistica
convergenza, 28-29
indipendenza, 24-25, 82, 89-90, 256, 259
stima
imparziale, 59, 61, 163, **283**
sufficiente, **170**, 173-174
stime, 167-170
di dispersione, 42-44
di tendenza centrale, 33-38
Stirling, James, 279
strumenti di misura, 9-12
caratteristiche, 10-11
Student, *vedi* Gosset, William

— T —

two-tailed test, 217

— U —

unità di misura
fondamentali e derivate, 6-9
naturali, 7

■ V ■

- varianza, 43-44
 - della media aritmetica, 54
 - della popolazione, **49-50**, 290-291
 - della popolazione e di campioni, 59-63
 - di combinazioni lineari
 - di variabili correlate, 256-258
 - di variabili indipendenti, 51-54
 - e media aritmetica, 191, **203**, 275
 - proprietà matematiche, 44
- vita media, 86, 113, 192-194
- von Mises, Richard, 21