

Alberto Rotondi
Paolo Pedroni
Antonio Pievatolo

Probabilità, Statistica e Simulazione

Programmi applicativi
scritti con Scilab

3^a edizione

 Springer

UNITEXT

Probabilità, Statistica e Simulazione

Alberto Rotondi · Paolo Pedroni · Antonio Pievatolo

Probabilità, Statistica e Simulazione

Programmi applicativi scritti con Scilab

3^a edizione

 Springer

Alberto Rotondi

Dipartimento di Fisica Nucleare e Teorica,
Università di Pavia

Paolo Pedroni

Istituto Nazionale di Fisica Nucleare,
Sezione di Pavia

Antonio Pievatolo

IMATI-CNR, Milano

ISBN 978-88-470-2363-5

e-ISBN 978-88-470-2364-2

DOI 10.1007/978-88-470-2364-2

Springer Milan Dordrecht Heidelberg London New York

© Springer-Verlag Italia 2012

Quest'opera è protetta dalla legge sul diritto d'autore e la sua riproduzione è ammessa solo ed esclusivamente nei limiti stabiliti dalla stessa. Le fotocopie per uso personale possono essere effettuate nei limiti del 15% di ciascun volume dietro pagamento alla SIAE del compenso previsto dall'art. 68. Le riproduzioni per uso non personale e/o oltre il limite del 15% potranno avvenire solo a seguito di specifica autorizzazione rilasciata da AIDRO, Via Corso di Porta Romana n. 108, Milano 20122, e-mail segreteria@aidro.org e sito web www.aidro.org. Tutti i diritti, in particolare quelli relativi alla traduzione, alla ristampa, all'utilizzo di illustrazioni e tabelle, alla citazione orale, alla trasmissione radiofonica o televisiva, alla registrazione su microfilm o in database, o alla riproduzione in qualsiasi altra forma (stampata o elettronica) rimangono riservati anche nel caso di utilizzo parziale. La violazione delle norme comporta le sanzioni previste dalla legge.

L'utilizzo in questa pubblicazione di denominazioni generiche, nomi commerciali, marchi registrati, ecc. anche se non specificatamente identificati, non implica che tali denominazioni o marchi non siano protetti dalle relative leggi e regolamenti.

Layout copertina: Simona Colombo, Milano

Impaginazione: PTP-Berlin, Protago T_EX-Production GmbH, Germany (www.ptp-berlin.eu)

Stampa: Grafiche Porpora, Segrate (MI)

Stampato in Italia

Springer-Verlag Italia S.r.l., Via Decembrio 28, I-20137 Milano

Springer-Verlag fa parte di Springer Science+Business Media (www.springer.com)

Prefazione

Questo testo nasce dalla collaborazione tra due fisici sperimentali e uno statistico.

Tra i non statistici, i fisici sono forse quelli che più apprezzano e utilizzano il calcolo delle probabilità e la statistica, il più delle volte però in modo pragmatico e manualistico, avendo in mente la soluzione di problemi o applicazioni tecniche. D'altra parte, nel confronto cruciale fra teoria ed esperimento, occorre a volte utilizzare metodi sofisticati, che richiedono una conoscenza profonda dei principi, anche logici e matematici, che stanno alla base dello studio dei fenomeni casuali. Più in generale, anche chi non è statistico deve spesso affrontare, nell'ambito della ricerca, problemi che richiedono particolari doti di attenzione e competenza nel trattamento degli aspetti casuali o aleatori. Queste doti sono invece possedute in modo naturale dallo statistico, il quale fa delle leggi del caso l'oggetto delle proprie ricerche.

Questo testo è maturato con l'intento di cercare una sintesi tra queste esperienze diverse, per fornire al lettore non solo uno strumento utile ad affrontare i problemi, ma anche una guida ai metodi corretti per comprendere il complicato e affascinante mondo dei fenomeni aleatori.

Un tale obiettivo ha comportato ovviamente delle scelte, talvolta anche dolorose, sia nel tipo sia nella forma dei contenuti. Nella forma, abbiamo cercato di non rinunciare alla precisione necessaria per insegnare correttamente i concetti importanti; nelle applicazioni, abbiamo privilegiato i metodi che non richiedono eccessive elaborazioni concettuali preliminari. Abbiamo ad esempio cercato di utilizzare, quando è possibile, metodi approssimati per la stima intervallare, con le approssimazioni gaussiane alla distribuzione degli stimatori. Allo stesso modo, nel caso dei minimi quadrati, abbiamo usato in modo esteso l'approssimazione basata sulla distribuzione χ^2 per la verifica dell'adattamento di un modello ai dati. Abbiamo anche evitato di insistere nel trattamento formale di problemi complicati nei casi in cui si può trovare la soluzione utilizzando il computer e facili programmi di simulazione. Nel nostro testo la simulazione riveste quindi un ruolo importante nell'illustrazione di molti argomenti e nella verifica della bontà di molte tecniche ed approssimazioni.

Il libro si rivolge in primo luogo agli studenti dei primi anni dei corsi di indirizzo scientifico, come ingegneria, informatica e fisica. Pensiamo però che esso possa risultare utile anche a tutti quei ricercatori che devono risolvere problemi concreti che coinvolgono aspetti probabilistici, statistici e di simulazione. Per questo abbiamo dato spazio ad alcuni argomenti, come il metodo Monte Carlo e le sue applicazioni, le tecniche di minimizzazione e i metodi di analisi dei dati, che solitamente non vengono trattati nei testi di carattere introduttivo.

Le conoscenze matematiche richieste al lettore sono quelle impartite solitamente nell'insegnamento di Analisi Matematica I dei corsi di laurea ad indirizzo scientifico, con l'aggiunta di nozioni minime di Analisi Matematica II, come i fondamenti della derivazione ed integrazione delle funzioni di più variabili.

La struttura del testo consente diversi percorsi didattici e livelli di lettura. I primi 6 capitoli trattano tutti gli argomenti solitamente svolti in un corso istituzionale di statistica. A scelta del docente, questo programma può essere integrato con alcuni argomenti più avanzati tratti dagli altri capitoli. Ad esempio, in un corso orientato alle tecniche di simulazione, va senz'altro incluso il cap. 7. Le nozioni di probabilità e statistica impartite di solito agli studenti di fisica nei corsi di laboratorio del primo biennio sono contenute nei primi 3 capitoli, nel cap. 6 (statistica di base) e nel cap. 11, scritto esplicitamente per i fisici e per tutti coloro che hanno a che fare con il trattamento dei dati provenienti da esperienze di laboratorio.

Molte pagine sono dedicate alla risoluzione completa di numerosi esercizi inseriti direttamente nei capitoli ad illustrazione degli argomenti trattati. Raccomandiamo al lettore anche i problemi (tutti con soluzione) riportati alla fine di ogni capitolo.

Il testo è stato scritto nel corso di diversi anni, nel tempo "libero" dagli impegni dell'insegnamento e della ricerca. Molte parti del materiale che qui presentiamo sono state collaudate in alcuni corsi per fisici, istituzionali e di dottorato, sull'analisi statistica dei dati e sulle tecniche di simulazione Monte Carlo. Ringraziamo tutti gli studenti e i colleghi che ci hanno dato consigli, suggerimenti e corretto errori e imprecisioni.

Le principali modifiche introdotte in questa terza edizione sono la revisione della prima parte del cap. 6 (Statistica di base) e di tutto il cap. 10 (Minimi quadrati).

Nel capitolo 6 sono state riportate nuove formule per il calcolo degli intervalli di confidenza per la stima di efficienze e frequenze negli esperimenti di conteggio. Le formule proposte hanno una migliore probabilità di includere il valore vero (copertura), come risulta dalla verifica, con tecniche di simulazione, riportata in dettaglio nel par. 7.11 (Simulazione degli esperimenti di conteggio).

Il cap. 10 (Minimi quadrati), seguendo anche le indicazioni di alcuni lettori, è stato interamente rivisto e reso più chiaro e leggibile, almeno a nostro parere.

Anche in questa edizione è stato mantenuto l'utilizzo, in tutto il testo, del software libero SCILAB¹. I programmi che abbiamo scritto e utilizzato nel testo si possono ottenere dal sito web

<http://www.mi.imati.cnr.it/~marco/springer/>

Consigliamo quindi una lettura interattiva, che allo studio di un argomento faccia seguire l'uso delle routine secondo le modalità indicate nel testo e le istruzioni tecniche contenute in queste pagine web.

Ringraziamo ancora i lettori che ci hanno segnalato errori o imprecisioni e la Springer per la fiducia che ha continuato ad accordarci.

Pavia, giugno 2011

Alberto Rotondi
Paolo Pedroni
Antonio Piovatolo

¹ <http://www.scilab.org>

Come utilizzare il testo

Figure, equazioni, definizioni, teoremi, tabelle ed esercizi sono numerati progressivamente all'interno di ogni capitolo.

Le sigle delle citazioni (ad esempio [57]) si riferiscono all'elenco riportato nella bibliografia alla fine del libro.

La soluzione dei problemi è a pagina 499. Può anche essere utile la tabella dei simboli riportata nelle pagg. 493–494.

I codici di calcolo come `histplot` sono indicati su sfondo grigio. Le routine che iniziano con lettera minuscola sono quelle originali di SCILAB, che possono essere liberamente copiate da

`http://www.scilab.org,`

mentre quelle che cominciano in maiuscolo sono scritte dagli autori e si trovano in

`http://www.mi.imati.cnr.it/~marco/springer/`

In questo sito si trovano anche tutte le informazioni per la installazione e l'uso di SCILAB, una guida all'uso delle routine scritte dagli autori e materiale didattico complementare al testo.

Indice

Prefazione	V
1 La probabilità	1
1.1 Caso, caos e determinismo	1
1.2 Terminologia	8
1.3 Il concetto di probabilità	10
1.4 Probabilità assiomatica	13
1.5 Prove ripetute	20
1.6 Calcolo combinatorio	24
1.7 Teorema di Bayes	25
1.8 L'approccio bayesiano	31
Problemi	33
2 Rappresentazione dei fenomeni aleatori	37
2.1 Introduzione	37
2.2 Variabili aleatorie	38
2.3 Funzione cumulativa o di ripartizione	42
2.4 La rappresentazione dei dati	44
2.5 Variabili aleatorie discrete	47
2.6 La distribuzione binomiale	49
2.7 Variabili aleatorie continue	52
2.8 Media, Varianza e Deviazione Standard	55
2.9 Operatori	60
2.10 Il campione casuale	63
2.11 Criteri di convergenza	64
Problemi	68
3 Calcolo elementare delle probabilità	71
3.1 Introduzione	71
3.2 Proprietà della distribuzione binomiale	71
3.3 La distribuzione di Poisson	74

3.4	Densità di Gauss o normale	76
3.5	Calcolo di distribuzioni in SCILAB	81
3.6	Legge 3-sigma e gaussiana standard	82
3.7	Il teorema Limite Centrale: universalità della gaussiana	86
3.8	Processi stocastici poissoniani	87
3.9	La densità χ^2	94
3.10	La densità uniforme	100
3.11	Disuguaglianza di Tchebychev	104
3.12	Come utilizzare il calcolo delle probabilità	105
	Problemi	114
4	Calcolo delle probabilità per più variabili	117
4.1	Introduzione	117
4.2	Distribuzioni statistiche multidimensionali	117
4.3	Covarianza e correlazione	125
4.4	Densità gaussiana bidimensionale	129
4.5	Generalizzazione in più dimensioni	138
4.6	Insiemi di probabilità in più dimensioni	142
4.7	La distribuzione multinomiale	146
	Problemi	148
5	Funzioni di variabili aleatorie	149
5.1	Introduzione	149
5.2	Funzione di una variabile aleatoria	151
5.3	Funzioni di più variabili aleatorie	153
5.4	La trasformazione della media e della varianza	165
5.5	Medie e varianze per n variabili	171
	Problemi	176
6	Statistica di base	179
6.1	Introduzione	179
6.2	Intervalli di confidenza	181
6.3	Determinazione degli intervalli di confidenza	185
6.4	Cenno all'approccio bayesiano	186
6.5	Alcune notazioni	187
6.6	Stima della probabilità da grandi campioni	188
6.7	Stima della probabilità da piccoli campioni	193
6.8	Intervalli di stima per eventi poissoniani	200
6.9	Stima della media da grandi campioni	203
6.10	Stima della varianza da grandi campioni	205
6.11	Stima di media e varianza da piccoli campioni	209
6.12	Come utilizzare la teoria della stima	212
6.13	Stime da una popolazione finita	218
6.14	Verifica di una ipotesi	220
6.15	Verifica di compatibilità tra due valori	223

6.16	Stima della densità di una popolazione	228
6.17	Verifica di compatibilità tra campione e popolazione	234
6.18	Verifica di ipotesi con test non parametrici	242
6.19	Stima della correlazione	249
	Problemi	257
7	Il metodo Monte Carlo	261
7.1	Introduzione	261
7.2	Cos'è il metodo Monte Carlo?	262
7.3	Fondamenti matematici	265
7.4	Generazione di variabili aleatorie discrete	267
7.5	Generazione di variabili aleatorie continue	270
7.6	Metodo del rigetto	274
7.7	Metodo di ricerca lineare	281
7.8	Metodi particolari di generazione casuale	282
7.9	Studio Monte Carlo di distribuzioni	288
7.10	Determinazione degli intervalli di confidenza	291
7.11	Simulazione degli esperimenti di conteggio	295
7.12	<i>Bootstrap</i> non parametrico	300
	Problemi	304
8	Applicazioni del metodo Monte Carlo	307
8.1	Introduzione	307
8.2	Studio dei fenomeni di diffusione	307
8.3	Simulazione dei processi stocastici	315
8.4	Il numero di addetti ad un impianto: simulazione sincrona	320
8.5	Il numero di addetti ad un impianto: simulazione asincrona	326
8.6	Algoritmo di Metropolis	327
8.7	Il modello di Ising	330
8.8	Calcolo di integrali definiti	334
8.9	Campionamento ad importanza	338
8.10	Campionamento stratificato	339
8.11	Integrali multidimensionali	343
	Problemi	344
9	Inferenza statistica e verosimiglianza	347
9.1	Introduzione	347
9.2	Il metodo della massima verosimiglianza	349
9.3	Proprietà degli stimatori	354
9.4	Teoremi sugli stimatori	357
9.5	Intervalli di confidenza	366
9.6	Il metodo dei minimi quadrati e la massima verosimiglianza	370
9.7	Adattamento di densità (<i>best fit</i>) ad istogrammi	372
9.8	La media pesata	376
9.9	Verifica delle ipotesi	380

9.10	Test di potenza massima	383
9.11	Funzioni test	386
9.12	Test sequenziali	390
	Problemi	395
10	Minimi quadrati	399
10.1	Introduzione	399
10.2	Predittori osservati senza errore	402
10.3	Errori sui predittori	405
10.4	Rette dei minimi quadrati	408
10.5	Minimi quadrati lineari: caso generale	414
10.6	Minimi quadrati lineari pesati	419
10.7	Fit con errori sistematici	421
10.8	Proprietà delle stime dei minimi quadrati	424
10.9	Verifica del modello e determinazione di forme funzionali	427
10.10	Ricerca delle correlazioni	434
10.11	Test di adattamento del modello	439
10.12	Minimizzazione non lineare	441
	Problemi	443
11	Analisi dei dati sperimentali	445
11.1	Introduzione	445
11.2	Terminologia	446
11.3	Grandezze fisiche costanti e variabili	447
11.4	Sensibilità e accuratezza degli strumenti	448
11.5	Incertezza nelle operazioni di misura	451
11.6	Misure indirette e propagazione degli errori	454
11.7	Definizione dei tipi di misura	465
11.8	Misure del tipo $M(0, 0, \Delta)$	466
11.9	Misure del tipo $M(0, \sigma, 0)$	467
11.10	Misure del tipo $M(0, \sigma, \Delta)$	470
11.11	Misure del tipo $M(f, 0, 0)$	472
11.12	Misure del tipo $M(f, \sigma, 0)$, $M(f, 0, \Delta)$, $M(f, \sigma, \Delta)$	477
11.13	Studio di un caso: gli esperimenti di Millikan	482
11.14	Alcune note sul metodo scientifico	485
	Problemi	491
Appendice A	Tabella dei simboli	493
Appendice B	Funzioni generatrici	495
Appendice C	Soluzioni dei problemi	499
Appendice D	Tabelle	517
D.1	Integrale della densità gaussiana	517
D.2	Valori quantili della densità di Student	518

D.3	Integrale della densità χ^2 ridotto	518
D.4	Valori quantili di χ^2 non ridotto	519
D.5	Valori quantili della densità F	519
Bibliografia		527
Indice analitico		533

La probabilità

Sembra non esserci alternativa all'accettare l'idea che nell'esistenza c'è un qualche elemento incomprensibile. La scelta è libera. Noi tutti fluttuiamo delicatamente fra una visione soggettiva e una visione oggettiva del mondo, e questo dilemma è al centro della natura umana.

Douglas R. Hofstadter, *L'io della mente*

1.1 Caso, caos e determinismo

Questo libro si propone di studiare in modo dettagliato i fenomeni detti aleatori, stocastici o casuali. Prima di immergerci nell'argomento, in questa introduzione analizzeremo brevemente la collocazione ed il ruolo di questi fenomeni nel contesto della realtà in cui viviamo.

Nella misura o nell'osservazione di un fenomeno naturale, inizialmente si cerca di identificare tutte le cause e le condizioni esterne che ne determinano l'evoluzione. Successivamente, si opera in modo da tenere fisse o sotto controllo, per quanto possibile, queste cause esterne e si procede a registrare i risultati delle osservazioni.

Ripetendo le osservazioni, si possono allora presentare due casi:

- si ottiene sempre lo stesso risultato. Come esempio, pensate alla misura di un tavolo con un metro commerciale;
- si ottiene ogni volta un risultato differente. Pensate ad un fenomeno naturale molto semplice: il lancio di una moneta.

Mentre nel primo caso per il momento non c'è molto da dire, nel secondo vogliamo qui chiederci a cosa sono dovute le variazioni osservate nei risultati. Possibili cause sono il non aver tenuto sotto controllo tutte le condizioni del fenomeno, oppure avere definito in modo scorretto la grandezza da osservare. Una volta effettuate queste correzioni, il fenomeno può diventare stabile, oppure continuare a presentare fluttuazioni.

Spieghiamo con un esempio: supponiamo di voler misurare l'ora del sorgere del sole all'orizzonte in una data località. Osserveremo che misure ripetute in

giorni successivi forniscono un risultato differente. Ovviamente, in questo caso la variazione del risultato è dovuta ad una cattiva definizione della misura. L'ora del sorgere del sole va misurata in un certo luogo e per un determinato giorno dell'anno, e va ripetuta a distanza di un anno in quello stesso giorno e luogo. Ridefinendo in questo modo l'osservazione, i risultati di misure ripetute coincidono. Come è noto, le leggi del moto dei pianeti forniscono in questo caso un modello che permette di prevedere (a meno di piccole correzioni che non vogliamo discutere) le ore dell'alba per tutti i giorni dell'anno¹.

Consideriamo ora la misura di un'altra grandezza, la temperatura di un giorno ad una determinata ora. In questo caso, anche considerando un certo giorno dell'anno e ripetendo le misure di anno in anno, si osservano risultati differenti. A differenza dell'ora del sorgere del sole, non siamo in questo caso nemmeno in possesso di un modello che ci permetta di prevedere con precisione il risultato della misura.

Come mai la temperatura, a differenza dell'ora dell'alba, sembra avere intrinsecamente un comportamento casuale? La ragione è che, mentre l'ora dell'alba dipende dall'interazione di pochi corpi (il sole e i pianeti), la temperatura dipende, oltre che dalle condizioni astronomiche, anche dallo stato dell'atmosfera, il quale è il risultato dell'interazione di innumerevoli fattori, che *nemmeno in linea di principio* possono essere determinati con assoluta precisione o comunque tenuti sotto controllo.

Questa distinzione è cruciale ed è la chiave per stabilire la differenza tra grandezze che presentano fluttuazioni e quelle che sembrano fisse o comunque prevedibili con precisione in base a modelli deterministici.

Storicamente, i sistemi deterministici sono stati per lungo tempo considerati privi di fluttuazioni ed il loro studio, nell'ambito della fisica classica, è proseguito in parallelo a quello dei sistemi detti stocastici, aleatori o casuali, nati con lo studio dei giochi d'azzardo: lancio di dadi, giochi di carte, roulette, slot machines, gioco del lotto e così via. Questi ultimi sono sistemi costruiti appositamente per assicurare la casualità del risultato.

Esistevano pertanto due mondi separati: quello privo di fluttuazioni dei fenomeni deterministici, legati alle leggi fondamentali della fisica, ai sistemi semplici (in genere a pochi corpi), ed il mondo dei fenomeni aleatori, soggetti a fluttuazioni, legato ai sistemi complessi (in genere a molti corpi).

Tuttavia, già all'inizio del secolo scorso il fisico-matematico francese H. Poincaré si accorse che in alcuni casi la conoscenza della legge deterministica non era sufficiente per fare predizioni esatte sulla dinamica del sistema a partire da condizioni iniziali note. Il problema, che oggi si chiama studio del caos, venne però ripreso e studiato a fondo solo ben più tardi, a partire dagli anni '70, grazie all'aiuto dei calcolatori. Oggi sappiamo che, nei sistemi macroscopici, l'origine delle fluttuazioni può essere duplice, cioè dovuta a leggi deterministi-

¹ In realtà non si sa esattamente quanto sia stabile il sistema solare. Alcuni modelli indicano che non sia possibile estendere le previsioni oltre un intervallo di tempo dell'ordine dei cento milioni di anni [BV03].

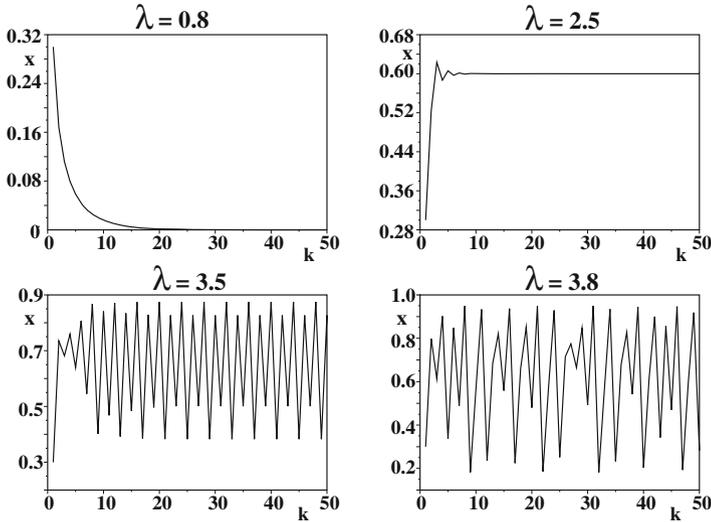


Fig. 1.1. Valori assunti da x in base alla equazione logistica (1.1) partendo dal valore iniziale $x = 0.3$ per differenti valori del parametro λ

che presentano elevata sensibilità rispetto alle condizioni iniziali (sistemi caotici) oppure alla impossibilità di definire in modo deterministico tutte le variabili del sistema (sistemi stocastici).

Uno dei paradigmi migliori per spiegare il caos è la mappa logistica, proposta fin dal 1838 dal matematico belga P.F. Verhulst e studiata in dettaglio dal biologo R. May nel 1976 e dal fisico M. Feigenbaum nel 1978:

$$x(k+1) = \lambda x(k) [1 - x(k)] , \quad (1.1)$$

dove k è il ciclo di crescita della popolazione, λ è legato al tasso di crescita e $0 \leq x(k) \leq 1$ è una variabile di stato proporzionale al numero di individui della popolazione. Affinché x si mantenga nei limiti prefissati, deve essere $0 \leq \lambda \leq 4$. La legge logistica descrive bene la dinamica di evoluzione delle popolazioni dove esiste un accrescimento per ciclo proporzionale a $\lambda x(k)$ con un termine correttivo negativo (reazione o *feedback*) $-\lambda x^2(k)$ proporzionale al quadrato della dimensione già raggiunta dalla popolazione.

Senza addentrarci troppo nello studio della mappa logistica, notiamo che il comportamento della popolazione evolve col numero dei cicli secondo le caratteristiche seguenti (riportate anche in Fig. 1.1):

- per $\lambda \leq 1$ il modello conduce sempre all'estinzione della popolazione;
- per $1 < \lambda \leq 3$ il sistema raggiunge un livello stabile, che dipende da λ *ma non dalla condizione iniziale* $x(0)$;
- per $3 < \lambda \leq 3.56994\dots$ il sistema oscilla periodicamente tra alcuni valori fissi. Ad esempio, come mostrato in Fig. 1.1 per $\lambda = 3.5$, i valori possibili

sono 4 (si noti che in figura i valori discreti di x sono uniti dalla linea continua). Anche in questo caso gli stati raggiunti dal sistema non dipendono dalla condizione iniziale;

- per $\lambda > 3.56994\dots$ il sistema è caotico: le fluttuazioni sembrano regolari, ma, come si vede guardando con attenzione la Fig. 1.1 per $\lambda = 3.8$, esse non sono né periodiche né sembrano del tutto casuali. Uno studio approfondito mostra anche che le fluttuazioni non sono nemmeno prevedibili con precisione, perché valori della condizione iniziale $x(0)$ molto vicini *portano ad evoluzioni completamente diverse della popolazione*. Questo fenomeno, che si chiama *dipendenza sensibile dalle condizioni iniziali o effetto farfalla*², è una delle caratteristiche principali del caos. Si noti che le fluttuazioni nei sistemi caotici sono *oggettive, intrinseche o essenziali*, poiché la riproducibilità dei risultati richiederebbe condizioni iniziali fissate ad un livello di precisione inferiore a quello della scala atomica, il che non è possibile *nemmeno in linea di principio*.

Potete fare esperienza numerica con la mappa logistica e verificare l'effetto farfalla utilizzando le nostre funzioni SCILAB `Logist` e `Logiplot`³, con cui abbiamo prodotto la Fig. 1.1.

I metodi per distinguere i sistemi caotici da quelli stocastici si basano essenzialmente sullo studio degli scarti, cioè della differenza tra i valori corrispondenti delle stesse variabili di stato nelle evoluzioni successive del sistema, in funzione del numero di variabili di stato.

In un sistema caotico, raggiunto un certo numero di variabili, gli scarti si stabilizzano o tendono a diminuire. Questo comportamento indica che si è raggiunto un numero di variabili di stato adeguato a descrivere il sistema e che si può ottenere la legge deterministica che ne regola la dinamica. Le fluttuazioni nei risultati di esperimenti ripetuti in questo caso vengono attribuite, come abbiamo visto, alla sensibilità del sistema rispetto alle condizioni iniziali

In un sistema stocastico, invece, il numero di variabili di stato necessario alla descrizione completa del sistema non viene mai raggiunto e la somma degli scarti, o delle grandezze ad essi connesse, continua a crescere col numero delle variabili di stato considerate [BV03]. Le fluttuazioni delle variabili appaiono *casuali* e seguono le distribuzioni del calcolo delle probabilità.

Lo studio del caos e delle transizioni dallo stato caotico a quello stocastico (e viceversa) è un'area di ricerca molto recente e tuttora aperta, dove molti problemi restano ancora irrisolti. Il lettore interessato può

² Riferendosi al caos nei sistemi meteorologici, spesso si dice: “un battito d'ali di farfalla ai tropici può scatenare un uragano in Florida”.

³ Le istruzioni per eseguire le nostre routine e a quelle del software SCILAB [SCI], utilizzato come codice di riferimento nel seguito del testo, si possono ottenere dal sito <http://www.mi.imati.cnr.it/~marco/springer/> [RPP]. D'ora in poi i nomi dei codici verranno evidenziati su sfondo grigio. Le routine originali SCILAB iniziano con lettera minuscola, le nostre con lettera maiuscola.

entrare in questo affascinante argomento attraverso le letture introduttive [BV03, Rue87, Ste93].

Nel seguito del libro non tratteremo il caos, ma ci dedicheremo invece allo studio dei sistemi aleatori o stocastici, cioè di tutti questi sistemi nei quali, in base a quanto detto, vi sono variabili che seguono, in linea di massima, l'affermazione⁴:

Affermazione 1.1 (Variabile aleatoria in senso lato). *Una variabile statistica, aleatoria o casuale è il risultato dell'interazione di molti fattori, ognuno dei quali non è preponderante sugli altri. Questi fattori (e le loro leggi dinamiche) non possono essere completamente individuati, fissati e comunque tenuti sotto controllo, nemmeno in linea di principio.*

Nel testo useremo prevalentemente il termine “variabile aleatoria”. Proviamo ora a identificare alcuni sistemi o processi stocastici che in natura producono variabili aleatorie o casuali.

In natura tutti i sistemi a molti corpi hanno un grado molto elevato di casualità: le osservabili dinamiche di sistemi di molecole, gas ideali e sistemi termodinamici in genere seguono molto bene l'affermazione 1.1. Questi sistemi sono oggetto di studio della fisica statistica.

Chiameremo per brevità *sistema stocastico puro* un sistema stocastico (o aleatorio) che produce variabili aleatorie.

Possiamo a questo punto precisare il significato del termine “*fattori e leggi dinamiche impossibile da determinare, nemmeno in linea di principio*” che abbiamo usato nell'affermazione 1.1. Supponiamo di lanciare un dado 100 volte. Per costruire un modello deterministico che possa prevedere il risultato dell'esperimento sarebbe necessario introdurre nelle equazioni del moto del dado tutte le condizioni iniziali relative al lancio, i vincoli dati dalle pareti della mano o del bicchiere entro cui si scuote il dado prima del lancio, i vincoli dati dal piano dove si getta il dado, e forse altro ancora. Avremmo così un insieme veramente formidabile di numeri, relativi alle condizioni iniziali e ai vincoli per ognuno dei cento lanci, enormemente più grande dei cento numeri che costituiscono il risultato finale dell'esperimento. È chiaro che il potere predittivo di una teoria del genere è nullo e la sua applicabilità pratica inesistente. Un modello deterministico, per essere tale, deve basarsi su un insieme compatto di equazioni e condizioni iniziali e deve essere in grado di prevedere un insieme vastissimo di fenomeni. È questo, ad esempio, il caso della legge logistica (1.1) o della semplice legge della caduta dei gravi, che collega lo spazio percorso s alla accelerazione gravitazionale g e al tempo di caduta t attraverso la formula $s = gt^2/2$. Questa sola formula consente di prevedere, assegnando s o t come condizione iniziale, il risultato di infiniti esperimenti.

Possiamo riassumere il significato delle osservazioni appena fatte dicendo che ogni volta che un modello deterministico dà luogo ad algoritmi che

⁴ Nel testo le definizioni operative, di tipo non matematico, verranno chiamate “affermazioni”.

richiedono un insieme numerico di condizioni iniziali, vincoli ed equazioni enormemente più numeroso dell'insieme costituito dal risultato che si intende prevedere, esso perde di senso. In alternativa, va usato l'approccio statistico che, basandosi sullo studio *a posteriori* dei risultati ottenuti, cerca di quantificare l'entità delle fluttuazioni ed estrarre da queste delle regolarità globali che possano essere utili nella previsione d'insieme (cioè probabilistica) dei risultati futuri.

Questa linea di pensiero, che si andata sviluppando nel corso degli ultimi tre secoli, è pervenuta, studiando i sistemi stocastici puri, ad identificare le leggi matematiche fondamentali per la descrizione dei fenomeni casuali. L'insieme di queste leggi è oggi noto come calcolo delle probabilità.

Tutti i libri dove si espone il calcolo delle probabilità (e il nostro non fa eccezione, come vedrete) fanno largo uso di esempi tratti dai giochi d'azzardo, come il lancio dei dadi. Queste esemplificazioni, lungi dall'essere una banalizzazione del problema, ne costituiscono invece l'essenza. Solo studiando sistemi stocastici puri si può pervenire ad una identificazione non ambigua delle leggi del caso. Grandi matematici e statistici, come Fermat (1601-1665), Laplace (1749-1827) e Bernoulli (1654-1705), discutono spesso, nelle loro memorie, esperimenti da loro eseguiti con lanci di dadi od altri meccanismi tratti dai giochi. Uno dei loro obiettivi era proprio quello di fornire strategie vincenti per i giochi d'azzardo, già allora molto diffusi e da loro stessi praticati assiduamente. In questo modo essi gettarono le fondamenta del calcolo delle probabilità e della statistica basandosi esclusivamente, come vuole il metodo scientifico, sui fatti sperimentali.

Accanto ai giochi tradizionali, oggi esiste un altro laboratorio "artificiale", per così dire, costituito dai processi casuali generati mediante calcolatore. Come vedremo, è infatti possibile, disponendo di un generatore di numeri casuali equiprobabili (una specie di roulette elettronica), simulare sistemi stocastici puri di qualunque tipo e comunque complessi: lanci di dadi, giochi di carte, sistemi fisici a molti corpi ed altro ancora. Queste tecniche, dette Monte Carlo (in omaggio alla patria dei giochi d'azzardo) o di simulazione, sono molto pratiche ed efficaci, perché permettono di ottenere in pochi secondi banche di dati che con un esperimento reale richiederebbero anni. È però importante sottolineare che dal punto di vista concettuale queste tecniche non introducono elementi nuovi. Il fine è sempre quello di ottenere variabili casuali da *modelli* costituiti da sistemi stocastici puri (in tutto o anche solo in parte, se il problema lo richiede). Questi dati servono poi per sviluppare ed ottimizzare gli strumenti logico-matematici da applicare allo studio dei sistemi reali.

E veniamo ora proprio ai sistemi reali in generale. Osservate la Fig. 1.2, che rappresenta la temperatura media della terra negli ultimi 130 anni. Come potete ben immaginare, sull'andamento di questa curva negli anni futuri si giocherà l'avvenire delle prossime generazioni. Paragonando "ad occhio" questa curva con quella di Fig. 1.3, che rappresenta un processo stocastico puro, sembra che ad un andamento casuale si sia sovrapposto, agli inizi del secolo scorso, un andamento (*trend*) parametrizzabile in modo più o meno esatto con

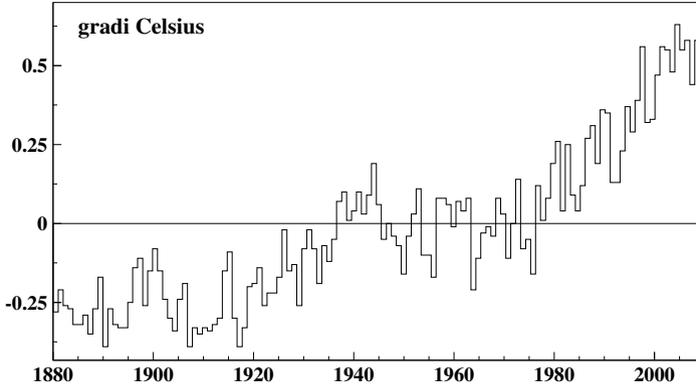


Fig. 1.2. Variazione della temperatura media della terra negli ultimi 130 anni. La linea di zero rappresenta la media degli anni 1951-1980 (dati tratti da: <http://www.giss.nasa.gov>)

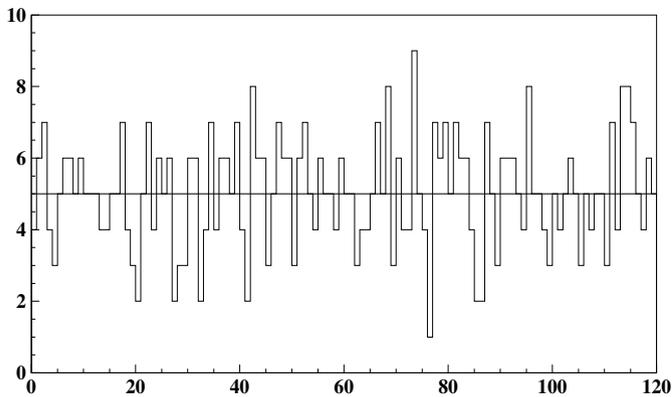


Fig. 1.3. Numero di teste ottenuto nel lancio di 10 monete nel corso di 120 prove simulate al computer. Il numero progressivo della prova è riportato in ascisse, il numero di teste ottenute in ordinate. La linea continua rappresenta il valore medio atteso (5 teste). Si confronti questa figura con la Fig. 1.1 per $\lambda = 3.8$, che riporta le fluttuazioni caotiche

una funzione matematica crescente. Non andiamo oltre con questo esempio, piuttosto preoccupante, che ci è servito solo per mostrare che nei casi reali la contemporanea presenza di effetti stocastici e deterministici è molto comune.

Per tenere conto di queste possibili complicazioni, lo studio di un sistema reale viene condotto con un approccio graduale, secondo dei passi che possiamo schematizzare come segue:

- a) identificare i processi puramente stocastici del sistema ed individuarne, in base alle leggi del calcolo delle probabilità, le leggi di evoluzione;
- b) separare le componenti stocastiche da quelle non stocastiche (dette anche sistematiche), se ve ne sono. Questo passo viene generalmente fatto utilizzando i metodi della statistica;
- c) se il problema è particolarmente difficile, eseguire al computer la simulazione del sistema in esame e confrontare i dati simulati con quelli reali.

Spesso è necessario ripetere i passi a)-c) fino a che i dati simulati sono in accordo soddisfacente con quelli reali. Questa tecnica ricorsiva costituisce un metodo di analisi molto potente ed è oggi applicata in molti campi della ricerca scientifica, dalla fisica all'economia.

Prima di chiudere questa introduzione dobbiamo fare cenno a ciò che avviene nel mondo microscopico. In questo caso le equazioni fondamentali della fisica forniscono una funzione di stato complessa $\psi(r)$ il cui modulo quadrato fornisce la probabilità di localizzazione della particella nello spazio: $P(r) = |\psi(r)|^2$. La probabilità così trovata *obbedisce alle leggi generali del calcolo delle probabilità che illustreremo nel seguito*.

Il fatto che le leggi fondamentali del mondo microscopico contengano una funzione di probabilità e che nessuno sia stato finora in grado di trovare leggi fondamentali più elementari fondate su grandezze diverse, porta correttamente ad affermare che la probabilità è una quantità fondamentale della natura. L'indeterminismo, essendo presente nelle leggi fondamentali che regolano la dinamica del mondo microscopico, assume in questo caso un carattere oggettivo, non legato alla ignoranza o alle capacità limitate di chi osserva.

1.2 Terminologia

Cominciamo ora ad introdurre, in modo poco rigoroso ma intuitivo, i termini fondamentali del linguaggio usato nello studio dei fenomeni stocastici. Nel seguito del testo questi termini verranno gradualmente ridefiniti in modo matematicamente rigoroso.

- *Spazio campionario o dei casi*: è l'insieme di tutti i possibili valori diversi (casi) che può assumere una variabile aleatoria. Ad esempio: la variabile aleatoria *carta di un mazzo da gioco* dà luogo a uno spazio di 52 elementi. La struttura dello spazio dipende dal modo utilizzato per definire la variabile aleatoria. Infatti lo spazio relativo alla variabile aleatoria *carta di un mazzo da gioco* è costituito dalle 52 carte, oppure da 52 numeri interi se creiamo una tabella di corrispondenza tra carte e numeri.
- *Evento*: è una particolare combinazione o un particolare sottoinsieme di casi. Ad esempio: nel caso delle carte da gioco, se si definisce come evento una carta dispari, l'insieme dei casi che si ottiene è 1, 3, 5, 7, 9, per ognuno

dei 4 colori. Questo evento dà luogo ad un sottoinsieme di 20 elementi di uno spazio che ne contiene 52 (tutte le carte del mazzo).

- *Spettro*: è l'insieme di tutti gli elementi diversi del sottoinsieme di casi che definisce l'evento. Nel caso delle carte da gioco dispari, lo spettro è dato da: 1, 3, 5, 7, 9. Ovviamente, lo spettro può coincidere con l'intero spazio della variabile aleatoria che si sta considerando (se, ad esempio, l'evento è definito come una carta qualunque di un mazzo).
- *Probabilità*: è la valutazione quantitativa della possibilità di ottenere un determinato evento. Essa viene fatta sulla base dell'esperienza, utilizzando modelli matematici oppure anche su base puramente soggettiva. Ad esempio, la probabilità che a questo punto voi continuiate a leggere il nostro libro è, secondo noi, del 95%.
- *Prova*: è l'insieme delle operazioni che realizzano l'evento.
- *Esperimento, Misura, Campionamento*: è un insieme di prove. Il termine usato dagli statistici è campionamento, mentre i fisici spesso usano il termine esperimento o misura. Ovviamente, in fisica un esperimento può consistere in un campionamento, ma non necessariamente.
- *Campione*: è il risultato di un esperimento (campionamento).
- *Popolazione*: è il risultato di un numero di prove, finito o infinito, tale da esaurire tutti gli eventi possibili. Ad esempio, nel gioco del lotto la popolazione si può pensare come l'insieme finito di tutte le cinque estratte da un'urna di 90 numeri; nel caso dell'altezza degli italiani, possiamo immaginare l'insieme delle misure delle altezze di ogni individuo. Quando la popolazione è pensata come un campione costituito da un numero infinito di eventi, va considerata come una astrazione matematica non raggiungibile in pratica.

Le idee ora introdotte possono essere riassunte nello schema di Fig. 1.4. Una volta assegnate le probabilità elementari per gli elementi dello spazio campionario (passo induttivo), utilizzando il calcolo delle probabilità si possono calcolare le probabilità degli eventi e dedurre i modelli matematici per la popolazione (passo deduttivo). Eseguendo invece una serie di misure si può ottenere un campione di eventi (spettro sperimentale) rappresentativo della popolazione in esame. Utilizzando poi l'analisi statistica dei dati (passo induttivo/deduttivo) si cerca di individuare, dall'analisi del campione, le proprietà della popolazione. Queste tecniche sono dette di *inferenza statistica*. Una volta assunto un modello, si può verificarne la congruenza con il campione dei dati raccolti. Questa tecnica si chiama *verifica di ipotesi*.

Nel testo verranno inizialmente illustrati i fondamenti del calcolo delle probabilità, con particolare riguardo alla assegnazione delle probabilità elementari agli elementi dello spazio campionario e all'utilizzo del calcolo combinatorio per ottenere alcuni modelli matematici fondamentali.

Successivamente verranno illustrati i metodi di analisi statistica dei dati che permettono di stimare, partendo da quantità misurate, i valori "veri" di

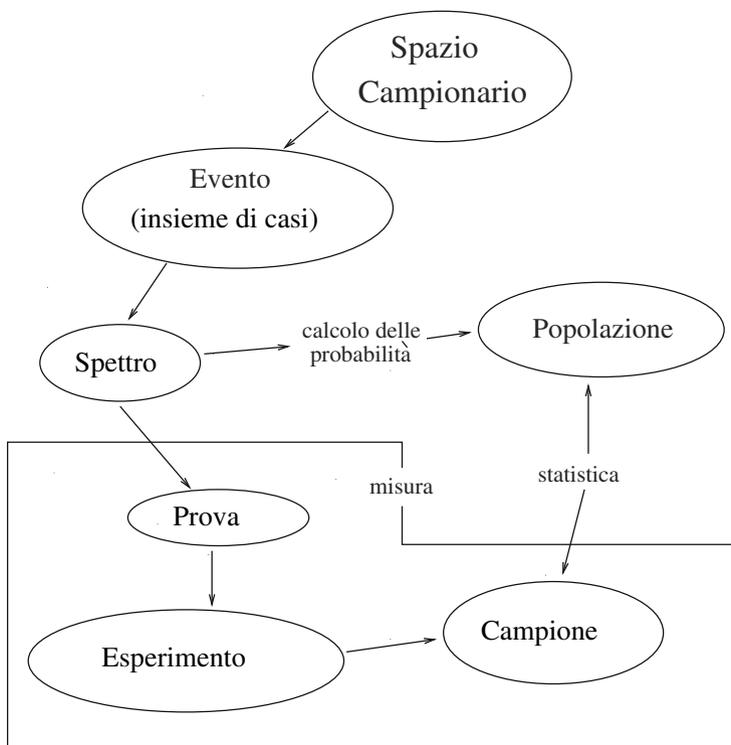


Fig. 1.4. Connessione tra calcolo delle probabilità, statistica e misura

grandezze fisiche o di verificare la congruenza dei campioni sperimentali con i modelli matematici.

Gli elementi di calcolo delle probabilità e statistica precedentemente acquisiti verranno poi applicati in modo esteso alle tecniche di simulazione.

1.3 Il concetto di probabilità

L'esperienza mostra che, quando un fenomeno stocastico o aleatorio è stabile nel tempo, alcuni valori dello spettro capitano più frequentemente di altri. Se lanciamo in aria 10 monete e contiamo il numero di teste, vediamo che 5 teste si ottengono più frequentemente di 8, mentre 10 teste è un evento veramente raro, quasi impossibile. Se consideriamo un esperimento costituito da 100 prove (dove la prova è il lancio di 10 monete), osserviamo che il numero di volte in cui si ottengono 5, 8 e 10 teste, pur variando di poco da esperimento a esperimento, si mantiene abbastanza regolare, perché i valori 5, 8 e 10 appaiono sempre (o quasi) con frequenza decrescente. Se immaginiamo tutti i possibili allineamenti di 10 monete, possiamo avere una spiegazione

intuitiva di questo fatto: all'evento 10 teste (o 10 croci) corrisponde un solo allineamento, mentre all'evento 5 croci-5 teste corrispondono molti allineamenti possibili (5 croci e poi 5 teste, croce-testa in modo alternato, e così via fino a ben 252 allineamenti diversi). Quando lanciamo 10 monete scegliamo a caso, *senza privilegiarne nessuno*, uno dei possibili allineamenti ed è intuitivo che quasi sempre otterremo risultati bilanciati (più o meno 5 teste) e quasi mai i casi estremi. Ragionamenti di questo tipo, comuni all'esperienza quotidiana di ognuno di noi, portano istintivamente a pensare che questa regolarità dei fenomeni stocastici sia dovuta all'esistenza di quantità fisse, dette *probabilità*, definibili, ad esempio, come il rapporto tra i casi (allineamenti) favorevoli e quelli possibili.

Queste considerazioni portarono J. Bernoulli alla formulazione della prima legge matematica con la quale oggi possiamo prevedere, a meno delle fluttuazioni dovute alle numerose interazioni tra gli elementi interni al sistema aleatorio, l'andamento d'insieme dei risultati in esperimenti come quello del lancio delle monete.

Nel caso delle monete, la probabilità viene introdotta per tenere conto della variabilità dei risultati sperimentali; tuttavia, ognuno di noi utilizza la probabilità anche per tenere conto dell'incertezza di molte situazioni che capitano nella vita reale, quantificando soggettivamente le possibilità che si presentano e scegliendo quelle di maggior probabilità in relazione ai costi o benefici che se ne possono ricavare.

Ad esempio, quando siamo alla guida dell'auto ed incontriamo un semaforo rosso, abbiamo due possibilità: fermarci o proseguire. Se siamo in un incrocio ad alto traffico e in un'ora di punta, ci fermiamo, perché sappiamo, in base all'esperienza, che la probabilità di entrare in collisione con altri veicoli è elevatissima. Se siamo invece in un incrocio a basso traffico e nel pieno della notte, siamo tentati di proseguire, perché sappiamo che la probabilità di una collisione è molto bassa.

Un altro esempio di probabilità soggettiva, di tipo discreto, è dato dal giudizio di colpevolezza o meno di un imputato in un processo da parte di una giuria. In questo caso la probabilità può essere espressa con due valori, 0 od 1, cioè colpevole o innocente. In genere le giurisprudenze attuali formulano il giudizio finale combinando le probabilità individuali soggettive espresse dai singoli giurati.

Attualmente l'approccio considerato più corretto, efficace e in fondo più economico per lo studio dei fenomeni aleatori è quello che considera la scelta della probabilità come un atto soggettivo, basato sull'esperienza.

Una prima possibile definizione operativa di probabilità è pertanto:

Affermazione 1.2 (Probabilità soggettiva). *La probabilità è il grado di convinzione (degree of belief) soggettivo circa il verificarsi di un evento.*

La probabilità soggettiva è libera, ma generalmente si assume che essa debba essere *coerente*, vale a dire espressa come un numero reale $0 \leq p \leq 1$ e pari a $p = 1$ per un evento ritenuto certo e $p = 0$ per un evento giudicato

impossibile. Considerando poi due o più eventi tra loro incompatibili (come ottenere 2 o 4 nel lancio di un dado) la coerenza impone che le probabilità siano additive. Queste ipotesi sono sufficienti per l'assiomatizzazione secondo lo schema di Kolmogorov, che illustreremo tra poco. La probabilità soggettiva è largamente usata nelle scienze generalmente denominate “non esatte” (giurisprudenza, economia, parte della medicina, ecc.). Nelle scienze esatte come la fisica (preciseremo meglio in seguito, a pagina 446, il significato del termine scienza esatta), la probabilità soggettiva viene generalmente evitata e ad essa vengono sostituite le definizioni di probabilità a priori (Laplace, 1749-1827) e probabilità frequentista (Von Mises, 1883-1953).

Definizione 1.3 (Probabilità a priori o classica). *Se N è il numero totale di casi dello spazio campionario di una variabile aleatoria ed n il numero di casi favorevoli per i quali si realizza l'evento A , la probabilità a priori di A è data da:*

$$P(A) = \frac{n}{N}. \quad (1.2)$$

Nel testo useremo prevalentemente il termine “probabilità classica”.

Ad esempio, la probabilità classica di una faccia nel lancio di un dado non truccato è:

$$P(A) = \frac{n}{N} = \frac{\text{numero di casi favorevoli}}{\text{numero di casi possibili}} = \frac{1}{6},$$

mentre la probabilità di estrarre l'asso di quadri da un mazzo di carte è $1/52$, la probabilità di estrarre un seme di quadri è $1/4$ e così via.

Definizione 1.4 (Probabilità frequentista). *Se m è il numero di prove in cui si è verificato l'evento A su un totale di M prove, la probabilità di A è data da:*

$$P(A) = \lim_{M \rightarrow \infty} \frac{m}{M}. \quad (1.3)$$

Il limite che compare in questa definizione non va inteso in senso matematico, ma in senso *sperimentale*: il valore vero della probabilità si trova solo effettuando un numero infinito di prove. Chiameremo nel seguito questa operazione, con il limite scritto in corsivo, *limite frequentista*.

La scelta delle probabilità elementari da assegnare agli eventi è dunque induttiva ed arbitraria. Il calcolo delle probabilità di eventi complessi a partire dalle probabilità elementari assegnate è invece di tipo deduttivo e viene fatto, come vedremo, senza uscire dal rigore matematico.

L'uso della probabilità soggettiva è detto anche *approccio bayesiano*, perché in questo caso le probabilità iniziali vengono spesso riaggiustate in funzione dei risultati ottenuti utilizzando la famosa formula di Bayes, che dedurremo tra poco nel par. 1.7.

L'approccio frequentista è quello nettamente prevalente in ambito fisico e tecnico. In base alla nostra esperienza, riteniamo che nei lavori di fisica sperimentale l'approccio frequentista sia seguito nel 99% dei casi, e questa è una valutazione ... soggettiva!

Nell'approccio frequentista si ritiene che la (1.3) permetta una valutazione "oggettiva" della probabilità per quei fenomeni naturali che si possono campionare facilmente o riprodurre numerose volte in laboratorio. In molti casi l'esperienza mostra che la probabilità frequentista tende a coincidere con quella classica:

$$\lim_{M \rightarrow \infty} \frac{m}{M} = \frac{n}{N} \quad (\text{dall'esperienza!}) . \quad (1.4)$$

Quando si verifica questa condizione, che è detta *legge dei grandi numeri* ed anche *legge empirica del caso*, si dice che i casi sono equiprobabili. Spesso, se su basi intuitive si ritiene fondata l'ipotesi dei casi equiprobabili, si utilizza la probabilità classica per costruire utili modelli matematici di riferimento.

Prendiamo, ad esempio, il lancio di un dado: se si è sicuri che esso non è truccato, è intuitivo assumere che la probabilità di ottenere in un lancio una certa faccia (poniamo la numero 3) sia pari a $1/6$. L'esperienza mostra che, eseguendo numerosi lanci, la probabilità frequentista (detta anche limite della frequenza, equazione (1.3)) tende effettivamente a $1/6$, in accordo con la (1.4). Se il dado è invece truccato, la probabilità di ottenere la faccia con il numero 3 può essere valutata solo eseguendo numerose prove. Poiché il limite per il numero di prove tendente a infinito non è praticamente raggiungibile, ci si ferma solitamente ad un numero di prove elevato ma finito e si stima la probabilità vera col metodo dell'intervallo di confidenza (vedi cap. 6).

La definizione frequentista (1.3) sembrerebbe dunque la più generale ed affidabile; tuttavia, essa non è esente da problemi:

- dato che non si può ripetere un esperimento infinite volte, la probabilità (1.3) non sarà mai osservabile;
- il limite che compare nella (1.3) non ha un senso matematico preciso.

La decisione su quale sia l'approccio migliore da usare (bayesiano o frequentista) in base al tipo di problema che si intende affrontare (incertezza in senso lato o variabilità dei risultati di un esperimento) è una questione tuttora aperta ed è fonte di continue controversie; ci sembra pertanto opportuno fermare qui la discussione sui fondamenti. Questo testo, che è dedicato a studenti e ricercatori in ambito tecnico-scientifico, è basato sull'approccio frequentista. Non mancheremo tuttavia di accennare in alcuni casi anche al punto di vista bayesiano, rimandando il lettore ai testi più specifici, come [Pic96].

1.4 Probabilità assiomatica

Per formalizzare in modo matematicamente corretto la probabilità è necessario ricorrere alla interpretazione insiemistica dei concetti fondamentali fin qui introdotti.

Se S è lo spazio campionario di una variabile aleatoria, consideriamo la famiglia \mathcal{F} di sottoinsiemi di S con le proprietà date dalla:

Definizione 1.5 (σ -algebra). Ogni famiglia \mathcal{F} di sottoinsiemi di S avente le proprietà:

- a) l'insieme vuoto appartiene ad \mathcal{F} : $\emptyset \in \mathcal{F}$;
 b) se una infinità numerabile di insiemi $A_1, A_2, \dots \in \mathcal{F}$, allora

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F} ;$$

- c) se $A \in \mathcal{F}$, lo stesso vale per l'insieme complementare: $\overline{A} \in \mathcal{F}$;
 si chiama σ -algebra.

Utilizzando le note proprietà degli insiemi:

$$\begin{aligned} \overline{\overline{A \cup B}} &= A \cap B , \\ A \cap \overline{B} &= A - B , \end{aligned}$$

è facile dimostrare che anche l'intersezione di un insieme numerabile di insiemi appartenenti ad \mathcal{F} e la differenza $A - B$ di due insiemi di \mathcal{F} appartengono a loro volta ad \mathcal{F} :

$$\bigcap_{i=n}^{\infty} A_i \in \mathcal{F} , \tag{1.5}$$

$$A - B \in \mathcal{F} . \tag{1.6}$$

La corrispondenza tra concetti insiemistici e probabilistici introdotta dal formalismo è riassunta in Tab. 1.1. Se, per fissare le idee, consideriamo un mazzo di carte da gioco e definiamo come evento A l'estrazione di un asso e come evento B l'estrazione di un seme quadri (Fig. 1.5), otteniamo la corrispondenza seguente tra insiemi (eventi) ed elementi di S :

Tabella 1.1. Corrispondenza tra concetti insiemistici e probabilistici

<i>Notazione</i>	<i>Significato insiemistico</i>	<i>Significato probabilistico</i>
S	insieme totale	spazio dei casi: evento certo
a	elemento di S	risultato di una prova
A	sottoinsieme di S	se $a \in A$ si realizza l'evento A
\emptyset	insieme vuoto	nessun evento realizzato
\overline{A}	insieme degli elementi di S che non appartengono ad A	evento A non realizzato
$A \cup B$	elementi di A o di B	si realizzano gli eventi A o B
$A \cap B$	elementi comuni ad A e B	si realizzano gli eventi A e B
$A - B$	elementi di A non appartenenti a B	si realizza l'evento A ma non l'evento B
$A \subseteq B$	gli elementi di A sono anche elementi di B	se si realizza A si realizza anche B

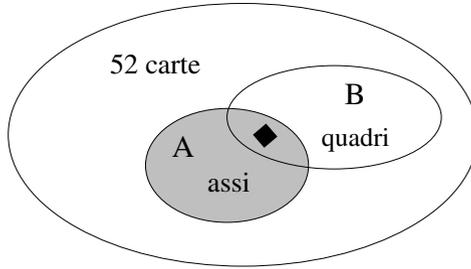


Fig. 1.5. Esempio di rappresentazione insiemistica della variabile casuale “carta da gioco” e degli eventi “estrazione di un asso” ed “estrazione di un seme di quadri”

- S : l’insieme delle 52 carte;
- a : una delle 52 carte del mazzo;
- $A \cup B$: seme di quadri o asso di cuori, fiori, picche;
- $A \cap B$: asso di quadri;
- $\overline{A} - B$: asso di cuori, fiori o picche;
- \overline{A} : tutte le carte tranne gli assi;
- \overline{B} : un seme non di quadri.

Sia ora $P(A)$ una applicazione o funzione che fa corrispondere ad un generico insieme A appartenente ad una σ -algebra \mathcal{F} un numero reale compreso nell’intervallo $[0, 1]$. In simboli,

$$P : \mathcal{F} \rightarrow [0, 1] . \tag{1.7}$$

Nella formulazione di Kolmogorov, la probabilità segue la:

Definizione 1.6 (Probabilità di Kolmogorov o assiomatica). *Una funzione $P(A)$ che soddisfa alla (1.7) ed alle proprietà:*

$$P(A) \geq 0 ; \tag{1.8}$$

$$P(S) = 1 ; \tag{1.9}$$

per ogni famiglia finita o numerabile A_1, A_2, \dots di insiemi di \mathcal{F} , tra loro mutuamente disgiunti:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \text{ se } A_i \cap A_j = \emptyset \ \forall i \neq j , \tag{1.10}$$

viene detta probabilità.

Definizione 1.7 (Spazio di probabilità). *La terna*

$$\mathcal{E} \equiv (S, \mathcal{F}, P) , \tag{1.11}$$

formata dallo spazio campionario, da una σ -algebra \mathcal{F} e da P viene detta spazio di probabilità.

La probabilità definita con gli assiomi di Kolmogorov gode delle seguenti importanti proprietà:

$$P(\bar{A}) = 1 - P(A) , \quad (1.12)$$

$$P(\emptyset) = 0 , \quad (1.13)$$

$$P(A) \leq P(B) \quad \text{se} \quad A \subseteq B . \quad (1.14)$$

La (1.12) è valida perché l'insieme complementare \bar{A} è definito in modo tale che $\bar{A} \cup A = S$, quindi $P(\bar{A}) + P(A) = P(S) = 1$ dalle (1.9, 1.10), dato che A e \bar{A} sono insiemi disgiunti. Inoltre,

$$P(S \cup \emptyset) = P(S) = 1 \quad \text{dalla (1.9)} , \quad (1.15)$$

$$P(S \cup \emptyset) = P(S) + P(\emptyset) = 1 \quad \text{dalla (1.10)} , \quad (1.16)$$

da cui la (1.13): $P(\emptyset) = 1 - P(S) = 1 - 1 = 0$. Infine, occorre tenere presente che se $A \subseteq B$ si può scrivere $B = (B - A) \cup A$, dove $B - A$ è l'insieme costituito dagli elementi di B che non appartengono ad A ; allora

$$P(B) = P[(B - A) \cup A] = P(B - A) + P(A)$$

e poiché $P(B - A) \geq 0$, anche la proprietà (1.14) è dimostrata.

È di notevole importanza il:

Teorema 1.1 (di additività). *La probabilità dell'evento dato dal verificarsi degli eventi elementari A oppure B , nel caso generale in cui $A \cap B \neq \emptyset$, è data da:*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) . \quad (1.17)$$

Dimostrazione. Sono di facile dimostrazione le proprietà (provate a disegnare gli insiemi...):

$$A \cup B = A \cup [B - (A \cap B)] ,$$

$$B = [B - (A \cap B)] \cup (A \cap B) ;$$

poiché $A \cup B$ e B sono ora espressi in termini di insiemi disgiunti, è possibile applicare la (1.10) ottenendo:

$$P(A \cup B) = P(A) + P[B - (A \cap B)] ,$$

$$P(B) = P[B - (A \cap B)] + P(A \cap B) .$$

Sottraendo membro a membro, si ottiene:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) . \quad \square$$

Sia la probabilità classica sia quella frequentista soddisfano agli assiomi (1.8-1.10). Infatti, per la probabilità classica, si ha:

$$\begin{aligned} P(A) &= (n_A/N) \geq 0 && \text{sempre, perché } n, N \geq 0, \\ P(S) &= N/N = 1, \\ P(A \cup B) &= \frac{n_A + n_B}{N} = \frac{n_A}{N} + \frac{n_B}{N} = P(A) + P(B). \end{aligned}$$

In modo del tutto analogo si può dimostrare la validità degli assiomi anche nel caso della probabilità frequentista, poiché il limite che vi compare va considerato come un operatore lineare.

Le probabilità classica e frequentista definite nel paragrafo precedente soddisfano quindi a tutte le proprietà (1.8-1.17). Ad esempio, la probabilità a priori di estrarre un asso oppure una carta rossa da un mazzo di carte, in base alla (1.17), è data da:

$$\begin{aligned} A &= \text{asso di cuori, asso di quadri, asso di fiori, asso di picche,} \\ B &= 13 \text{ carte di quadri, } 13 \text{ carte di cuori,} \\ P(A \cap B) &= \text{asso di cuori, asso di quadri,} \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) = 4/52 + 1/2 - 2/52 = 7/13. \end{aligned}$$

La probabilità associata all'insieme $A \cap B$ riveste, come vedremo, un ruolo particolarmente importante nell'algebra delle probabilità. Essa prende il nome di probabilità composta:

Definizione 1.8 (Probabilità composta). *Si dice probabilità composta e si indica con*

$$P(A \cap B) \quad \text{oppure con} \quad P(AB)$$

la probabilità che si verificano contemporaneamente gli eventi A e B .

Introduciamo ora un nuovo tipo di probabilità. Supponiamo di essere interessati alla probabilità che, estratto un seme di quadri, la carta sia un asso, oppure che, estratto un asso, esso sia di quadri. Indichiamo con A l'insieme degli assi, con B quello delle carte di quadri e con $P(A|B)$ la probabilità che si verifichi A dopo che si è verificato B , ovvero che, una volta estratto un seme di quadri, la carta sia un asso. Si ha ovviamente:

$$\begin{aligned} P(A|B) &= \frac{\text{casi favorevoli all'asso di quadri}}{\text{casi favorevoli al seme di quadri}} \\ &= \frac{1}{13} = \frac{1}{52} / \frac{13}{52} = \frac{P(A \cap B)}{P(B)}. \end{aligned} \quad (1.18)$$

Analogamente, la probabilità di ottenere un seme di quadri se si è estratto un asso è data da:

$$P(B|A) = \frac{\text{casi favorevoli all'asso di quadri}}{\text{casi favorevoli ad un asso}} = \frac{1}{4} = \frac{1}{52} / \frac{4}{52} = \frac{P(B \cap A)}{P(A)}.$$

Nell'esempio appena visto la probabilità condizionata $P(A|B)$ di ottenere un asso una volta estratto un seme di quadri è uguale alla probabilità non condizionata $P(A)$ di ottenere un asso; infatti:

$$P(A|B) = \frac{1}{13} = P(A) = \frac{4}{52} .$$

Si dice in questo caso che *gli eventi A e B sono indipendenti*. Tuttavia, se A rappresenta l'insieme [asso di quadri, asso di picche] e B è sempre l'insieme dei semi di quadri, abbiamo questa volta:

$$P(A|B) = \frac{1}{13} \neq P(A) = \frac{2}{52} = \frac{1}{26} .$$

In questo caso gli eventi A e B sono dipendenti, perché, se si estrae un seme di quadri, la previsione su A viene modificata. Si noti che però la (1.18) vale anche in questo caso:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{52} \frac{52}{13} = \frac{1}{13} .$$

Questi esempi suggeriscono la:

Definizione 1.9 (Probabilità condizionata). *Si definisce probabilità condizionata e si indica con $P(B|A)$ la probabilità che si verifichi l'evento B essendosi verificato l'evento A . Essa è espressa come:*

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{se } P(A) > 0 . \quad (1.19)$$

È facile mostrare (ve lo lasciamo come esercizio) che la definizione di probabilità condizionata (1.19) è in accordo con gli assiomi generali di Kolmogorov (1.8-1.10).

È importante anche notare che la formula della probabilità condizionata è stata introdotta *come definizione*. Tuttavia, per le probabilità trattate nel testo, sussiste il seguente:

Teorema 1.2 (del prodotto). *Per le probabilità classica e frequentista, la probabilità che si verifichi l'evento costituito dalla realizzazione degli eventi elementari A e B è data da:*

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) . \quad (1.20)$$

Dimostrazione. Per la probabilità classica, se N è il numero dei casi possibili e n_{AB} quello dei casi favorevoli ad A e B , si ha:

$$P(A \cap B) = \frac{n_{AB}}{N} = \frac{n_{AB}}{n_B} \frac{n_B}{N} = P(A|B)P(B) ,$$

dato che, per definizione, $n_{AB}/n_B \equiv P(A|B)$. Questa proprietà continua ovviamente a valere scambiando A con B , da cui la (1.20).

Per la probabilità frequentista, si procede come segue:

$$P(A \cap B) = \frac{\text{numero di volte che si sono ottenuti A e B}}{\text{numero di prove } (\rightarrow \infty)} = \lim_{N \rightarrow \infty} \frac{n_{AB}}{N},$$

$$P(A) = \frac{\text{numero di volte che si è ottenuto A}}{\text{numero di prove } (\rightarrow \infty)} = \lim_{N \rightarrow \infty} \frac{n_A}{N},$$

$$P(B) = \frac{\text{numero di volte che si è ottenuto B}}{\text{numero di prove } (\rightarrow \infty)} = \lim_{N \rightarrow \infty} \frac{n_B}{N},$$

$$P(B|A) = \frac{\text{n. di volte che si è ottenuto B insieme ad A}}{\text{n. di prove che hanno realizzato A } (\rightarrow \infty)} = \lim_{n_A \rightarrow \infty} \frac{n_{AB}}{n_A},$$

$$P(A|B) = \frac{\text{n. di volte che si è ottenuto A insieme a B}}{\text{n. di prove che hanno realizzato B } (\rightarrow \infty)} = \lim_{n_B \rightarrow \infty} \frac{n_{AB}}{n_B},$$

$$\begin{aligned} P(A \cap B) &= \frac{n_A}{N} \frac{n_{AB}}{n_A} = P(A)P(B|A) = \\ &= \frac{n_B}{N} \frac{n_{AB}}{n_B} = P(B)P(A|B) = \frac{n_{AB}}{N}, \quad N \rightarrow \infty. \quad \square \end{aligned}$$

Abbiamo introdotto, negli esempi trattati, la nozione di eventi indipendenti; in modo generale, possiamo introdurre la:

Definizione 1.10 (Eventi indipendenti). *Due eventi A e B sono detti indipendenti se e solo se*

$$P(A \cap B) = P(A)P(B).$$

Più in generale, gli eventi di una famiglia $(A_i, i = 1, 2, \dots, n)$ sono indipendenti se e solo se

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i), \quad (1.21)$$

per ogni sottoinsieme J di indici (tutti distinti) della famiglia.

Dalla (1.19) segue che per eventi indipendenti $P(A|B) = P(A)$ e $P(B|A) = P(B)$. Ancora una definizione:

Definizione 1.11 (Eventi incompatibili). *Due eventi sono incompatibili o disgiunti se quando si realizza A non si realizza B e viceversa:*

$$A \cap B = \emptyset.$$

In base alle (1.13) e (1.19) si ha allora:

$$P(A \cap B) = 0, \quad P(A|B) = P(B|A) = 0.$$

Ad esempio, se A è l'asso di picche e B il seme di quadri, A e B sono eventi incompatibili.

Con l'aiuto di queste definizioni l'essenza del calcolo delle probabilità può essere sintetizzata nelle formule:

- eventi incompatibili:

$$P(A \text{ oppure } B) \equiv P(A \cup B) = P(A) + P(B) ; \quad (1.22)$$

- eventi indipendenti:

$$P(A \text{ e } B) \equiv P(A \cap B) \equiv P(AB) = P(A) \cdot P(B) . \quad (1.23)$$

Ancora un punto ovvio ma fondamentale: $P(A|B)$ e $P(B|A)$ non sono la stessa cosa. Infatti, la probabilità che si verifichi A dopo B è diversa da quella che si verifichi B dopo A , come risulta subito evidente considerando eventi indipendenti con $P(A) \neq P(B)$, per i quali $P(A|B) = P(A)$ e $P(B|A) = P(B)$. Tra queste due probabilità condizionate esiste però una connessione molto importante, data dal teorema di Bayes, che illustreremo tra poco.

1.5 Prove ripetute

Nei casi considerati finora abbiamo sempre considerato esperimenti consistenti in una sola prova. Tuttavia, spesso ci si deve occupare di esperimenti consistenti in prove ripetute: due carte estratte da un mazzo, il punteggio che si ottiene lanciando 5 dadi, e così via. Affrontiamo questo problema considerando due prove ripetute, essendo ovvia, come vedremo, la generalizzazione ad un numero finito qualunque di prove.

Due prove ripetute possono essere viste come la realizzazione di due eventi relativi a due esperimenti $(S_1, \mathcal{F}_1, P_1)$ e $(S_2, \mathcal{F}_2, P_2)$ che soddisfano alle condizioni delle definizioni 1.6, 1.7. È allora spontaneo definire un nuovo spazio campionario $S = S_1 \times S_2$ come prodotto cartesiano dei due spazi campionari di partenza, in cui l'evento è costituito dalla coppia ordinata (x_1, x_2) , dove $x_1 \in S_1$ e $x_2 \in S_2$ e il nuovo spazio S contiene $n_1 n_2$ elementi, se n_1 ed n_2 sono gli elementi di S_1 ed S_2 rispettivamente. Ad esempio, [asso di cuori, donna di fiori] è un elemento dell'insieme S dello spazio di probabilità relativo all'estrazione di due carte da un mazzo. Si noti che il prodotto cartesiano può essere definito anche quando S_1 ed S_2 sono lo stesso insieme.

Avendo definito gli eventi in questo modo, dato che $A_1 \subseteq S_1$ e $A_2 \subseteq S_2$, è immediato rendersi conto che

$$A_1 \times A_2 = (A_1 \times S_2) \cap (S_1 \times A_2) . \quad (1.24)$$

Occorre adesso definire una probabilità P , definita in $S_1 \times S_2$, che goda delle proprietà di Kolmogorov (1.8-1.10) e che possa essere associata in modo univoco agli esperimenti consistenti in prove ripetute. La (1.24) e la (1.23),

valida per eventi indipendenti, portano in modo naturale alla definizione di probabilità prodotto:

$$P(A_1 \times A_2) = P_1(A_1) P_2(A_2), \quad A_1 \in \mathcal{F}_1, \quad A_2 \in \mathcal{F}_2. \quad (1.25)$$

In base a questa definizione, le probabilità degli eventi $A_1 \in \mathcal{F}_1$ e $A_2 \in \mathcal{F}_2$ possono essere calcolate anche nello spazio S tramite le uguaglianze:

$$\begin{aligned} P(A_1 \times S_2) &= P_1(A_1) P_2(S_2) = P_1(A_1), \\ P(S_1 \times A_2) &= P_1(S_1) P_2(A_2) = P_2(A_2), \end{aligned} \quad (1.26)$$

che risultano ovvie nel caso delle probabilità classica e frequentista. Ad esempio, nell'estrazione di due carte da gioco, le probabilità degli eventi A_1 =[estrazione di un asso la prima volta] e $A_1 \times S_2$ =[asso, carta qualunque] sono uguali, come quelle degli eventi A_2 =[estrazione di un seme di quadri la seconda volta] e $S_1 \times A_2$ =[carta qualunque, seme di quadri].

Come abbiamo detto, la relazione (1.25) viene ritenuta valida solo per eventi indipendenti, per i quali, in base alla (1.23), il verificarsi di un evento qualsiasi tra quelli considerati non altera la probabilità degli altri.

Per fissare meglio le idee con un esempio, supponiamo di estrarre due carte da un mazzo (con reimmissione nel mazzo della prima carta dopo la prima prova) e sia A_1 l'insieme degli assi e A_2 l'insieme dei semi di quadri. La (1.26) diventa:

$$\begin{aligned} P_1(A_1) &= \frac{4}{52} = P(A_1 \times S_2) = \frac{4}{52} \frac{52}{52}, \\ P_2(A_2) &= \frac{13}{52} = P(S_1 \times A_2) = \frac{52}{52} \frac{13}{52}, \end{aligned}$$

mentre la (1.25) fornisce:

$$P(A_1 \times A_2) = \frac{4}{52} \frac{13}{52}$$

in accordo con il rapporto tra il numero dei casi favorevoli ($4 \cdot 13$) e di quelli possibili ($52 \cdot 52$) nello spazio $S_1 \times S_2$. La famiglia di insiemi $\mathcal{F}_1 \times \mathcal{F}_2 = \{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$ non è in generale una σ -algebra, ma si può mostrare che esiste un'unica σ -algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ di sottoinsiemi di $S_1 \times S_2$ che contiene $\mathcal{F}_1 \times \mathcal{F}_2$ e che la (1.25) permette di estendere in modo univoco la probabilità di ogni evento $A \subset S_1 \times S_2$ dalla famiglia di insiemi $\mathcal{F}_1 \times \mathcal{F}_2$ alla σ -algebra prodotto $\mathcal{F}_1 \otimes \mathcal{F}_2$ [GS92]. Si può quindi scrivere lo spazio di probabilità prodotto come:

$$\mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2 \equiv (S_1 \times S_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P).$$

Un'estensione della (1.25) che useremo spesso viene utilizzata quando lo spazio S_2 non può essere assegnato a priori, ma dipende dai risultati dell'esperimento precedente \mathcal{E}_1 . Tipico è il caso del gioco del lotto, in cui vengono estratti

5 numeri, senza la reimmissione nell'urna del numero estratto. Nel caso di due prove, possiamo pensare alla estrazione di due carte da gioco: se si reinserisce la prima carta estratta nel mazzo, S_2 è costituito da 52 elementi, altrimenti da 51. In questi casi è necessario definire lo spazio $S = S_1 \times S_2$ non come prodotto cartesiano, ma come l'insieme di tutte le coppie ordinate possibili dei due insiemi, così come risultano dal particolare esperimento. Possiamo dire in questo caso che l'evento A_2 dipende dall'evento A_1 e generalizzare la (1.25) come:

$$P(A \times B) = P_2(B|A) P_1(A) , \quad (1.27)$$

ottenendo così un'estensione del teorema 1.2 del prodotto (eq. 1.20). È immediato infatti dimostrare che le probabilità a priori e frequentista soddisfano la (1.27). La dimostrazione per la probabilità frequentista è identica a quella del teorema 1.2, quella per la probabilità classica richiede di ridefinire N come l'insieme delle coppie possibili, n_{AB} come l'insieme delle coppie favorevoli all'evento e n_A come l'insieme delle coppie in cui, nella prima estrazione, si è verificato l'evento A .

È importante a questo punto, per evitare confusione, distinguere tra *esperimenti indipendenti* ed *eventi indipendenti*. L'ipotesi di esperimenti indipendenti, che utilizzeremo in tutto il testo, è del tutto generale ed implica che *le operazioni o prove che portano alla realizzazione di un evento qualsiasi sono invarianti rispetto alle operazioni che portano alla realizzazione degli altri eventi rimanenti*. Essa si riferisce alla *procedura sperimentale che si segue, non al numero di elementi dello spazio campionario su cui si agisce*. Al contrario, nello schema a prove ripetute saranno considerati dipendenti gli eventi per i quali la dimensione dello spazio i -esimo S_i dipende dalle $(i - 1)$ prove già effettuate. Questo è l'unico tipo di dipendenza che si assume, considerando prove ripetute, quando si scrive la probabilità condizionata $P(A|B)$.

Prendiamo confidenza con questi concetti con un semplice esempio. Sia $(B_1 \times N_2)$ l'evento consistente nell'estrarre nell'ordine, da un'urna contenente 2 palline bianche e 3 nere, una pallina bianca e una nera (senza reinserire nell'urna la pallina dopo la prima estrazione); si ha in questo caso (con ovvio significato dei simboli):

$$S_1 = [B1, B2, N1, N2, N3] ,$$

$$S_2 = [\text{insieme formato da 4 palline, 2 bianche e 2 nere} \\ \text{oppure 1 bianca e 3 nere}] ,$$

$$S_1 \times S_2 = \begin{array}{l} B1B2, B2B1, N1B1, N2B1, N3B1 \\ B1N1, B2N1, N1B2, N2B2, N3B2 \\ B1N2, B2N2, N1N2, N2N1, N3N1 \\ B1N3, B2N3, N1N3, N2N3, N3N2 \end{array} = 5 \times 4 = 20 \text{ elementi.}$$

Si noti che il mancato reinserimento della pallina dopo la prima estrazione esclude coppie del tipo $B1B1$, $N1N1$, ecc. Definiamo inoltre:

$$B_1 \times S_2 = [\text{pallina bianca, pallina qualunque}] ,$$

$$S_1 \times N_2 = [\text{pallina qualunque, pallina nera}] ,$$

$$B_1 \times N_2 = [\text{pallina bianca, pallina nera}] = \left(\begin{array}{l} B_1N_1, B_2N_1 \\ B_1N_2, B_2N_2 \\ B_1N_3, B_2N_3 \end{array} \right) = 6 \text{ elementi} .$$

Se i casi sono equiprobabili, la probabilità classica fornisce:

$$P(B_1 \times N_2) = \frac{\text{sei casi favorevoli}}{\text{venti coppie possibili}} = \frac{6}{20} = \frac{3}{10} .$$

Fin qui abbiamo usato la definizione di probabilità classica in modo del tutto generale. Ora notiamo che le probabilità (1.26) degli eventi B_1 e N_2 sono date da:

$$P_1(B_1) = 2/5 , \quad P_2(N_2|B_1) = 3/4 ,$$

perché nell'urna si hanno inizialmente B_1, B_2, N_1, N_2, N_3 (due palline bianche su un totale di 5) e, nella seconda estrazione, si hanno nell'urna 3 palline nere e una bianca se la prima pallina estratta era bianca, per un totale di 3 palline nere su 4. Se ora applichiamo la (1.27) otteniamo:

$$P(B_1 \times N_2) = P_2(N_2|B_1)P_1(B_1) = 3/4 \times 2/5 = 6/20 = 3/10 ,$$

in accordo con il calcolo diretto dei casi favorevoli su quelli possibili fatto in precedenza.

Se non diamo importanza all'ordine di estrazione e definiamo come evento la estrazione di una pallina bianca ed una nera o viceversa, dobbiamo definire gli insiemi:

$$B_1 \times S_2 = [\text{pallina bianca, pallina qualunque}] ,$$

$$N_1 \times S_2 = [\text{pallina nera, pallina qualunque}] ,$$

$$S_1 \times N_2 = [\text{pallina qualunque, pallina nera}] ,$$

$$S_1 \times B_2 = [\text{pallina qualunque, pallina bianca}]$$

ed applicare la (1.27):

$$P[(N_1 \times B_2) \cup (B_1 \times N_2)] = P_2(B_2|N_1)P_1(N_1) + P_2(N_2|B_1)P_1(B_1) = \frac{3}{5}$$

in accordo con il rapporto tra numero di coppie favorevoli (12) e quelle possibili (20).

La generalizzazione dello schema a prove ripetute per un numero finito di prove richiede la naturale estensione delle formule che qui abbiamo ricavato per due sole prove e non presenta difficoltà.

1.6 Calcolo combinatorio

Supponendo che conosciate già il calcolo combinatorio, ne riassumiamo qui brevemente le formule di base, che spesso sono di aiuto, nel calcolo delle probabilità, per contare il numero dei casi possibili o di quelli favorevoli. Per contare bene, occorre tenere presente che il numero delle coppie possibili (abbinamenti) $A \times B$ tra due insiemi A di a elementi e B di b elementi è dato dal prodotto ab e che il numero delle permutazioni possibili di n oggetti è dato dal fattoriale $n!$. Basandosi su queste due proprietà è possibile dimostrare facilmente le quattro formule fondamentali del calcolo combinatorio, che si riferiscono alle disposizioni senza ripetizione $D(n, k)$ di n oggetti in gruppi di k (n oggetti in classe k), a quelle con ripetizione $D^*(n, k)$ ed alle combinazioni senza e con ripetizione $C(n, k)$ e $C^*(n, k)$. Nelle disposizioni ha importanza l'ordine dei k oggetti, ma non nelle combinazioni. Le formule, come forse già sapete, sono:

$$D(n, k) = n(n-1) \cdots (n-k+1), \quad (1.28)$$

$$D^*(n, k) = n^k, \quad (1.29)$$

$$C(n, k) = \frac{n(n-1) \cdots (n-k+1)}{k!} = \frac{n!}{k!(n-k)!} \equiv \binom{n}{k}, \quad (1.30)$$

$$\begin{aligned} C^*(n, k) &= \frac{(n+k-1)(n+k-2) \cdots n}{k!} \\ &= \frac{(n+k-1)!}{k!(n-1)!} \equiv \binom{n+k-1}{k}, \end{aligned} \quad (1.31)$$

dove si è fatto uso del coefficiente binomiale. Per ricordare le formule, basta immaginare il gruppo di k oggetti come il prodotto cartesiano di k insiemi. In $D(n, k)$ il primo insieme contiene n elementi, il secondo, non potendo ripetere il primo elemento, ne conterrà $n-1$ fino ad arrivare, dopo k volte, a $(n-k+1)$. Se si può ripetere l'elemento, i k insiemi conterranno ognuno n elementi, dando luogo alla (1.29). La numerazione in base n è proprio un insieme $D^*(n, k)$: se ad esempio $n = 10$, $k = 6$, si hanno 10^6 numeri, da 000 000 a 999 999. La relazione (1.30) $C(n, k) = D(n, k)/k!$ evita di contare più volte i gruppi che contengono gli stessi oggetti, dato che in questo caso non ha importanza l'ordine. Per ottenere la (1.31) bisogna immaginare di scrivere, ad esempio, una combinazione $C^*(n, 5)$ del tipo a_1, a_2, a_2, a_2, a_7 come $a_1, *, a_2, *, *, *, a_3, a_4, a_5, a_6, a_7, *$, dove ogni elemento è seguito da un numero di asterischi pari al numero di volte che esso compare nella combinazione; esiste pertanto una corrispondenza biunivoca tra le combinazioni in esame e tutte le permutazioni possibili nell'allineamento di lettere e asterischi. Poichè ogni allineamento comincia con a_1 per costruzione, si possono permutare $n-1+k$ oggetti, di cui k (gli asterischi) ed $n-1$ (gli elementi a_i con $i = 2, \dots, n$) uguali tra loro, da cui la (1.31).

Una formula particolarmente utile è la *legge ipergeometrica*, che permette di calcolare la probabilità che da un'urna contenente a biglie di tipo A e b

biglie di tipo B si estraggano k biglie di tipo A avendone estratte $n \leq a + b$ senza reintroduzione nell'urna. Assumendo che tutte le biglie abbiano la stessa probabilità di essere estratte e che le estrazioni siano indipendenti, adottando la definizione a priori (1.2) e utilizzando i coefficienti binomiali, abbiamo:

$$P(k; a, b, n) = \frac{\binom{a}{k} \binom{b}{n-k}}{\binom{a+b}{n}}, \quad \max(0, n-b) \leq k \leq \min(n, a). \quad (1.32)$$

Infatti, il numero di casi possibili è dato dal coefficiente binomiale al denominatore, mentre al numeratore abbiamo il numero di casi favorevoli, dato dal numero di elementi dell'insieme prodotto cartesiano di due insiemi di $C(a, k)$ e $C(b, n-k)$ elementi rispettivamente.

Esercizio 1.1

Trovare la probabilità di vincere un ambo ed un terno al gioco del lotto.

Risposta. *La soluzione, se il gioco non è truccato, è data dalla legge ipergeometrica (1.32) con $a = k$ e $b = 90 - k$:*

$$P(2; 2, 88, 5) = \frac{\binom{88}{3}}{\binom{90}{5}} = \frac{2}{800} \quad (\text{ambo}),$$

$$P(3; 3, 87, 5) = \frac{\binom{87}{2}}{\binom{90}{5}} = \frac{1}{11\,748} \quad (\text{terno}).$$

La probabilità di un ambo è di circa 1 su 400, quella del terno di circa 1 su 12 000. Un gioco è onesto se la vincita è pari all'inverso della probabilità della puntata; nel gioco del lotto l'erario paga l'ambo 250 volte ed il terno 4250 volte ...

1.7 Teorema di Bayes

In linea di principio, qualunque problema che coinvolge l'uso di probabilità a priori può essere risolto con le due leggi fondamentali dell'additività e del prodotto. Tuttavia, l'algebra della probabilità porta rapidamente a formule complicate, anche nel caso di situazioni relativamente semplici. In questi casi sono di grande aiuto due formule fondamentali, la formula delle probabilità totali e il teorema di Bayes, che andiamo ad illustrare.

Se gli insiemi B_i ($i = 1, 2, \dots, n$) sono mutuamente disgiunti ed esaustivi dello spazio totale S ,

$$\bigcup_{i=1}^n B_i = S, \quad B_i \cap B_k = \emptyset \quad \forall i, k, \quad (1.33)$$

utilizzando la (1.20) è facile dimostrare che, per ogni insieme A contenuto in S :

$$\begin{aligned} P(A) &= P[A \cap (B_1 \cup B_2 \cup \dots \cup B_n)] \\ &= P[(A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_n)] \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n) \\ &= \sum_{i=1}^n P(A|B_i)P(B_i). \end{aligned} \quad (1.34)$$

La (1.34) è detta *formula delle probabilità totali o di partizione dell'evento certo*. Quando $B_1 = B$ e $B_2 = \overline{B}$, si ottiene la relazione:

$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B}). \quad (1.35)$$

Con queste formule è possibile risolvere problemi che capitano di frequente, come quelli mostrati nei due esempi che seguono.

Esercizio 1.2

Una certa malattia H colpisce in un anno il 10% degli uomini e il 5% delle donne. Sapendo che la popolazione è composta dal 45% di uomini e dal 55% di donne, trovare in numero N atteso di ammalati su una popolazione di 10 000 persone.

Risposta. Per prima cosa è necessario determinare la probabilità di ammalarsi per un generico individuo (uomo o donna) della popolazione. Essa è data dalla probabilità che l'individuo sia una donna, per la probabilità che una donna ha di ammalarsi, più la probabilità che l'individuo sia un uomo per la probabilità che un uomo ha di ammalarsi. Il tutto è condensato nelle formule (1.34, 1.35). Si ha pertanto:

$$P(H) = 0.45 \cdot 0.10 + 0.55 \cdot 0.05 = 0.0725 .$$

Il numero atteso di ammalati (speranza matematica) si calcola moltiplicando il numero di prove (individui) per la probabilità classica $P(H)$ appena trovata. Abbiamo allora:

$$N = 10\,000 \cdot 0.0725 = 725 .$$

Esercizio 1.3

Un'urna contiene 10 palline, 6 bianche e 4 nere. Se si effettuano delle estrazioni successive senza reinserimento delle palline estratte, qual è la probabilità di estrarre una pallina bianca alla seconda estrazione?

Risposta. Indicando con A e B l'uscita di una pallina bianca alla prima e alla seconda estrazione rispettivamente, applicando la (1.35) si ottiene, con ovvio significato dei simboli:

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) = \frac{5}{9} \frac{6}{10} + \frac{6}{9} \frac{4}{10} = 0.60 .$$

Se ora esprimiamo la probabilità $P(A)$ che compare nella (1.20) tramite la (1.34), otteniamo il famoso:

Teorema 1.3 (di Bayes). *Se gli eventi B_k soddisfano alla (1.33), la probabilità condizionata $P(B_k|A)$ si può scrivere come:*

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^n P(A|B_i)P(B_i)} , \quad P(A) > 0 . \quad (1.36)$$

Questo teorema, che è forse il risultato più importante dell'algebra elementare delle probabilità, viene spesso utilizzato per "riaggiustare", sulla base di una serie di dati reali A_k , delle probabilità $P(B_k)$ assegnate a priori in base ad ipotesi più o meno arbitrarie. La procedura da usare è mostrata negli esempi che seguono. Consigliamo agli studenti di fisica di risolvere anche il problema 1.8 alla fine del capitolo.

Esercizio 1.4

Un test per la diagnosi di una malattia è efficace nel 100% dei casi effettivi di malattia, ma risulta positivo anche nel 5% dei casi se effettuato su persone sane. Sapendo che la malattia è presente mediamente nell'1% della popolazione, qual è la probabilità di essere veramente malati se si risulta positivi al test?

Risposta. Poiché quello dei test diagnostici è un problema interessante, affrontiamo l'argomento in modo generale. Dai dati in nostro possesso possiamo definire le seguenti probabilità condizionate:

$$\begin{aligned} P(P|S) &= 0.05 \text{ probabilità di risultare positivi essendo sani ,} \\ P(N|S) &= 0.95 \text{ probabilità di risultare negativi essendo sani,} \\ P(P|M) &= 1. \text{ probabilità di risultare positivi essendo malati,} \\ P(N|M) &= 0. \text{ probabilità di risultare negativi essendo malati.} \end{aligned}$$

Fate bene attenzione, nel definire le probabilità condizionate, a non scambiare gli argomenti (ad esempio $P(M|S)$ con $P(S|M)$), altrimenti applicando il teorema di Bayes combinerete solo pasticci.

Le definizioni di probabilità implicano ovviamente:

$$P(P|S) + P(N|S) = 1.$$

$$P(P|M) + P(N|M) = 1.$$

È anche ovvio che per un test ideale si deve avere:

$$P(P|S) = 1, \quad P(N|S) = 0,$$

$$P(P|M) = 0, \quad P(N|M) = 1.$$

Il problema richiede il calcolo della probabilità $P(M|P)$ di essere malati condizionata dalla positività del test. Applicando il teorema di Bayes (1.36) e tenendo presente che dai dati risulta che le probabilità non condizionate di essere sani o malati valgono rispettivamente $P(S) = .99$ e $P(M) = 0.01$, otteniamo:

$$P(M|P) = \frac{P(P|M)P(M)}{P(P|M)P(M) + P(P|S)P(S)} = \frac{1 \times 0.01}{1 \times 0.01 + 0.05 \times 0.99} = 0.168,$$

pari ad una probabilità di circa il 17%.

Il risultato (una probabilità così bassa con il test positivo) può sembrare a prima vista paradossale. Per aiutare la vostra intuizione, vi invitiamo ad esaminare la Fig. 1.6, che mostra un metodo grafico equivalente alla applicazione del teorema di Bayes. Se 100 persone vengono sottoposte al test, mediamente 99 saranno sane ed una sola malata; il test, applicato ai 99 sani, fallirà nel 5% dei casi e segnalerà $0.05 \times 99 = 4.95 \simeq 5$ casi positivi; a questi va aggiunto il caso di malattia correttamente diagnosticato. In definitiva, avremo una sola persona effettivamente malata su un totale di 6 test positivi:

$$\frac{1}{6} = 16.67\%,$$

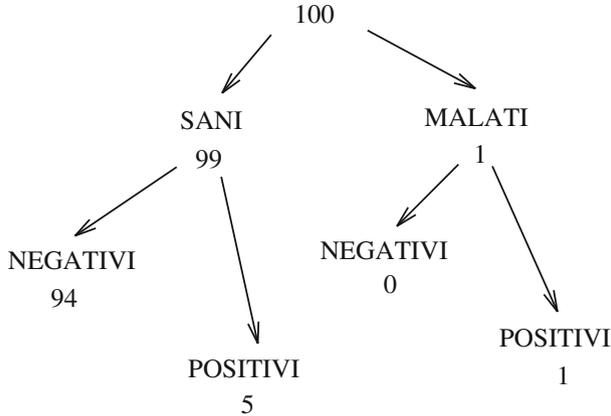
dove la piccola differenza con il calcolo esatto è dovuta solo agli arrotondamenti.

Spesso i medici in questi casi fanno ripetere il test. Se il risultato è negativo, allora la persona è sana, perché il test considerato in questo esercizio non può mai sbagliare su persone malate. Se il test risulta ancora positivo, allora occorre calcolare, in base alla (1.23), la probabilità di un test doppiamente positivo su un sano:

$$P(PP|S) = P(P|S)P(P|S) = (0.05)^2 = 0.0025,$$

che vale il 2.5 per mille, e quella $P(PP|M) = 1$ di un test doppiamente positivo su un malato (che ovviamente vale sempre 1) ed applicare ancora il teorema di Bayes:

$$\begin{aligned} P(M|PP) &= \frac{P(PP|M)P(M)}{P(PP|M)P(M) + P(PP|S)P(S)} \\ &= \frac{1 \times 0.01}{1 \times 0.01 + 0.0025 \times 0.99} = 0.802 \simeq 80\%. \end{aligned}$$



$$\text{malati/positivi} = 1/6 = 17\%$$

Fig. 1.6. Illustrazione grafica del teorema di Bayes nel caso di un test inefficiente al 5% sui sani per una malattia di frequenza pari all'1%

Come vedete, nemmeno due test positivi bastano in questo caso per essere ragionevolmente certi della malattia. Potete calcolare da voi che questa certezza, nel caso considerato, la si ha solo dopo ben tre test consecutivi (circa il 99%). L'esempio mostra quanto si debba essere cauti con i test che possono risultare positivi anche su persone sane. Il caso opposto si ha con i test che sono sempre negativi sui sani ma non sempre sono positivi sui malati. In questo caso il test positivo dà la certezza della malattia, mentre il test negativo crea incertezza. Vi sono anche i casi in cui i test hanno una efficienza limitata sia sui sani sia sui malati. In tutti questi casi il teorema di Bayes permette di calcolare con esattezza le probabilità di interesse. Provate ad inventarvi dei casi (oppure a chiedere a un medico qualche esempio concreto) e a risolvere da soli il problema.

Esercizio 1.5

Un insieme di sintomi A_1, A_2, A_3, A_4 può essere collegato a tre malattie H_1, H_2, H_3 , che, in base ai dati epidemiologici, hanno una frequenza relativa rispettivamente del 10%, 30% e 60%. Le probabilità relative sono pertanto:

$$P(H_1) = 0.1, \quad P(H_2) = 0.3, \quad P(H_3) = 0.6. \quad (1.37)$$

Sempre in base ai dati epidemiologici, risulta che l'occorrenza dei sintomi nelle tre malattie segue la tabella:

	A_1	A_2	A_3	A_4
H_1	.9	.8	.2	.5
H_2	.7	.5	.9	.99
H_3	.9	.9	.4	.7

dalla quale risulta, ad esempio, che il sintomo A_2 si presenta nell'80% dei casi nella malattia H_1 , il sintomo A_4 si presenta nel 70% dei casi nella malattia H_3 e così via.

Un paziente presenta solo i sintomi A_1 e A_2 . Quale tra le tre malattie considerate è la più probabile?

Risposta. Per applicare il teorema di Bayes occorre innanzitutto definire il paziente come un evento A tale che

$$A = A_1 \cap A_2 \cap \bar{A}_3 \cap \bar{A}_4$$

e calcolare le probabilità di questo evento, condizionate dalle tre malattie (ipotesi) H_1, H_2, H_3 come:

$$P(A|H_i) = P(A_1|H_i) P(A_2|H_i) P(\bar{A}_3|H_i) P(\bar{A}_4|H_i) \quad (i = 1, 2, 3) .$$

In base ai dati della tabella, facendo corrispondere ai sintomi assenti le probabilità complementari, otteniamo:

$$\begin{aligned} P(A|H_1) &= .9 \times .8 \times .8 \times .5 = 0.288 , \\ P(A|H_2) &= .7 \times .5 \times .1 \times .01 = 0.00035 , \\ P(A|H_3) &= .9 \times .9 \times .6 \times .3 = 0.1458 . \end{aligned}$$

La malattia più probabile sembra la H_1 , ma non abbiamo ancora tenuto conto delle frequenze epidemiologiche (1.37); qui sta il punto cruciale per la comprensione del teorema di Bayes!

Applichiamo pertanto la (1.36) e otteniamo finalmente le probabilità per ognuna delle tre malattie (notate che la somma dà 1, grazie al denominatore della formula di Bayes, che ha proprio la funzione di fattore di normalizzazione):

$$P(H_1|A) = \frac{0.288 \times 0.1}{0.288 \times 0.1 + 0.00035 \times 0.3 + 0.1458 \times 0.6} = 0.2455 ,$$

$$P(H_2|A) = \frac{0.00035 \times 0.3}{0.288 \times 0.1 + 0.00035 \times 0.3 + 0.1458 \times 0.6} = 0.0009 ,$$

$$P(H_3|A) = \frac{0.1458 \times 0.6}{0.288 \times 0.1 + 0.00035 \times 0.3 + 0.1458 \times 0.6} = 0.7456 .$$

Risulta pertanto una diagnosi che propende per la malattia H_3 , che ha una probabilità di circa il 75%.

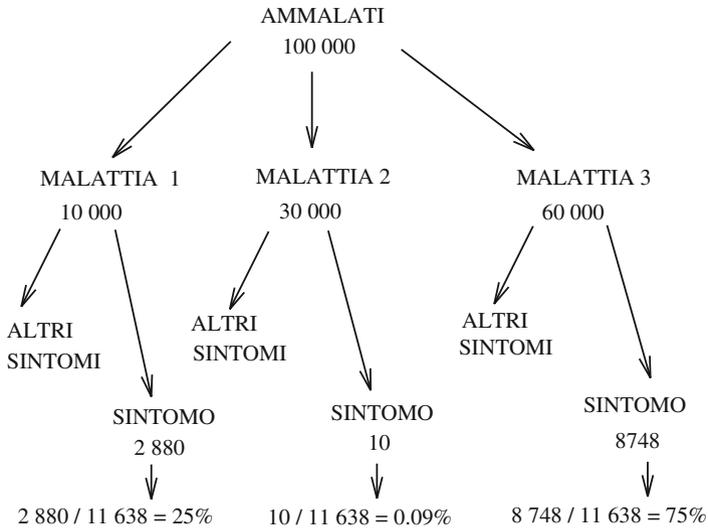


Fig. 1.7. Illustrazione grafica del teorema di Bayes nel caso di 3 malattie con dei sintomi comuni

I codici di calcolo per la diagnosi automatica fanno in genere largo uso del teorema di Bayes.

La soluzione del problema può anche essere trovata per via grafica come mostrato in Fig. 1.7: poiché vi sono probabilità piccole, in figura si considerano 100 000 ammalati, che vengono ripartiti secondo le tre malattie in base alle frequenze epidemiologiche 0.1, 0.3, 0.6; applicando a questi tre gruppi di ammalati le probabilità dell'insieme di sintomi A (0.288, 0.00035, 0.1458), si ottengono i numeri finali 2880, 10, 8748. Si arriva anche in questo modo, a meno degli arrotondamenti, al risultato fornito dalla formula di Bayes.

1.8 L'approccio bayesiano

I due esercizi del paragrafo precedente non si sarebbero potuti risolvere senza l'informazione *a priori* sulla percentuale sani/ammalati e sulle frequenze epidemiologiche delle tre malattie H_1 , H_2 , H_3 .

L'approccio bayesiano estende l'utilizzo della formula di Bayes anche ai casi in cui le probabilità iniziali non sono note con ragionevole certezza. In questi casi si utilizza la (1.36) per modificare, sulla base dei dati ottenuti, le probabilità iniziali di ipotesi $P(H_i)$ valutate in modo soggettivo in accordo con l'affermazione 1.2.

Indicando con l'evento generico "dati" il risultato di una o più prove (esperimento), nell'approccio bayesiano si applica la (1.36) come segue:

$$P(H_k|\text{dati}) = \frac{P(\text{dati}|H_k)P(H_k)}{\sum_{i=1}^n P(\text{dati}|H_i)P(H_i)}. \quad (1.38)$$

Le probabilità $P(H_k|\text{dati})$ così ricavate vengono poi sostituite alle $P(H_k)$ nel termine a destra della (1.38) ed il calcolo può essere ripetuto in modo iterativo:

$$P_n(H_k|E) = \frac{P(E_n|H_k)P_{n-1}(H_k)}{\sum_{i=1}^n P(E_n|H_i)P_{n-1}(H_i)}, \quad (1.39)$$

dove E_n indica l'evento n -esimo. Nell'esempio che segue, le probabilità $P(E_n|H_k)$ restano costanti.

Esercizio 1.6

Un'urna contiene 5 biglie, bianche e nere, in proporzione sconosciuta. Assumendo uguale probabilità $P(H_i) = 1/6$ per le 6 ipotesi di partenza possibili, scritte con ovvia notazione come:

$$H_i = (b, n) \equiv \begin{cases} H_1 = (5, 0), & H_2 = (4, 1), & H_3 = (3, 2), \\ H_4 = (2, 3), & H_5 = (1, 4), & H_6 = (0, 5), \end{cases}$$

calcolare le 6 probabilità $P(H_i|\text{dati})$ nel caso si siano estratte, con reimmisione nell'urna, $n = 1, 5, 10$ biglie nere consecutive.

Risposta. L'esercizio si risolve facilmente definendo l'evento "dati" = $E = E_n$ come l'estrazione di una biglia nera, utilizzando, per le probabilità a priori $P(E|H_k) = P(E_n|H_k)$, i valori:

$$\begin{aligned} P(\text{nera}|H_1) &= 0, & P(\text{nera}|H_2) &= 1/5, & P(\text{nera}|H_3) &= 2/5, \\ P(\text{nera}|H_4) &= 3/5, & P(\text{nera}|H_5) &= 4/5, & P(\text{nera}|H_6) &= 1 \end{aligned}$$

ed applicando in modo iterativo la (1.39). Si ottiene la Tab. 1.2, da cui si vede che, aumentando il numero di palline nere estratte consecutivamente, l'ipotesi H_6 (5 biglie nere) diventa sempre più probabile.

È cruciale notare che questo problema non è stato risolto in puro ambito frequentista, perché si è fatto ricorso alle probabilità a priori (sogettive e arbitrarie) $P(H_i) = 1/6$ ($i = 1, 2, \dots, 6$) per le ipotesi iniziali. Questa maggiore flessibilità si paga però con un certo margine di ambiguità, perché con ipotesi iniziali diverse, si sarebbero ottenuti risultati diversi, come nei problemi 1.12 e 1.13. Per fare esperimenti numerici col problema dell'urna, potete anche copiare dal nostro sito [RPP] la routine **Bayes**.

Tabella 1.2. Calcolo delle probabilità a posteriori $P_n(H|n\bullet)$ a partire da probabilità a priori tutte uguali nel caso di estrazioni consecutive di n biglie nere da un'urna contenente un insieme di 5 biglie bianche e nere

<i>Ipotesi</i>	H_1	H_2	H_3	H_4	H_5	H_6
<i>Tipo di urna</i>	○ ○ ○ ○ ○	○ ○ ○ ○ ●	○ ○ ○ ● ●	○ ○ ● ● ●	○ ● ● ● ●	● ● ● ● ●
$P(H_i)$ a priori	1/6	1/6	1/6	1/6	1/6	1/6
$P_n(H_i n = 1\bullet)$	0	0.07	0.13	0.20	0.27	0.33
$P_n(H_i n = 5\bullet)$	0.	0.00	0.01	0.05	0.23	0.71
$P_n(H_i n = 10\bullet)$	0.	0.00	0.00	0.00	0.11	0.89

Il dilemma “maggiore flessibilità di applicazione ma ambiguità dei risultati” è spesso al centro di accesi dibattiti, come in [JLPe00].

L'esempio appena visto è quindi fondamentale per comprendere la differenza tra l'approccio frequentista e quello bayesiano:

- approccio frequentista (adottato in questo testo): non si assumono probabilità arbitrarie di tipo soggettivo per le ipotesi. Non si possono quindi mai determinare probabilità di ipotesi del tipo $P(H|\text{dati}) \equiv P(\text{ipotesi}|\text{dati})$. Per una soluzione frequentista dell'esercizio appena visto potete andare a vedere più avanti, a pag. 199, l'esercizio 6.4;
- approccio bayesiano: si determinano probabilità del tipo $P(\text{ipotesi}|\text{dati})$ che dipendono, tramite la (1.38), sia dai dati ottenuti durante le prove sia dalle probabilità di ipotesi iniziali assunte in modo arbitrario.

Problemi

1.1. Il gioco di Monty Hall prende il nome dal conduttore di un gioco televisivo che nel 1990 fece discutere di probabilità milioni di americani. Il concorrente viene messo davanti a tre porte, dietro una delle quali c'è un'auto nuova. Il concorrente sceglie una porta, che rimane chiusa. Monty apre una delle due porte non scelte, dietro alla quale non c'è l'auto (operazione sempre possibile). A questo punto al concorrente viene chiesto se intende mantenere la prima porta scelta oppure cambiare scegliendo l'altra porta rimasta chiusa. Convieni cambiare, non cambiare o la scelta è indifferente?

1.2. Nel gioco del bridge un mazzo di 52 carte viene suddiviso in 4 gruppi di 13 carte e distribuito a 4 giocatori. Calcolare la probabilità che 4 giocatori che fanno 100 partite al giorno per 15 miliardi di anni (l'età dell'universo) possano giocare la stessa partita.

1.3. Una macchina è composta da tre elementi, che si possono guastare indipendentemente l'uno dall'altro. Le probabilità di funzionamento dei tre elementi durante

un tempo fissato T sono: $p_1 = 0.8$, $p_2 = 0.9$, $p_3 = 0.7$. La macchina si ferma per un guasto al primo elemento oppure per un guasto al secondo e al terzo elemento. Calcolare la probabilità P che la macchina funzioni entro T .

1.4. Una macchina è composta da quattro elementi aventi tutti la stessa probabilità $p = 0.8$ di funzionamento entro un tempo T . La macchina si ferma per un guasto contemporaneo agli elementi 1 e 2 oppure per un guasto contemporaneo agli elementi 3 e 4. a) Disegnate lo schema di flusso di funzionamento della macchina e b) calcolatene la probabilità P di funzionamento entro T .

1.5. Calcolare la probabilità di ottenere almeno un 6 lanciando 3 dadi.

1.6. Un lotto contenente 10 pezzi viene sottoposto ad un controllo di qualità che accetta l'intero lotto se tre pezzi scelti a caso sono tutti buoni. Calcolare la probabilità P che il lotto venga scartato nel caso vi siano a) un pezzo difettoso oppure b) quattro pezzi difettosi.

1.7. Il famoso problema dell'incontro: due amici X e Y decidono di incontrarsi in un certo luogo tra le 12 e le 13, scegliendo casualmente il tempo di arrivo. X arriva, aspetta dieci minuti, poi se ne va. Y fa come X , ma aspetta 12 minuti. Qual è la probabilità P che X e Y si incontrino?

1.8. Il problema del *trigger*, comune in fisica: un sistema fisico emette in modo aleatorio gli eventi A e B con probabilità del 90% e del 10% rispettivamente. Un apparato, che intende selezionare gli eventi "buoni" B , dà in consenso alla registrazione (innesco o *trigger*) il 5% delle volte per gli eventi A , nel 95% dei casi per gli eventi B . Calcolare la percentuale di eventi accettata dal trigger $P(T)$ e la percentuale $P(B|T)$ di eventi di tipo B tra quelli accettati.

1.9. Valutare la probabilità $P\{X \leq Y\}$ che in una prova, in cui si estraggono a caso in modo uniforme due coordinate $0 \leq X, Y \leq 1$, si ottengano i valori $x \leq y$.

1.10. Tre ditte elettroniche A, B e C forniscono di componenti identici un laboratorio. Le percentuali di fornitura sono il 20% per A, il 30% per B e il 50% per C. La percentuale di componenti difettosi dei tre fornitori è il 10% per A, il 15% per B ed il 20% per C. Qual è la probabilità che un componente scelto a caso risulti difettoso?

1.11. Un certo tipo di pilastro ha il carico di rottura R compreso uniformemente tra 150 e 170 kN. Sapendo che esso è sottoposto ad un carico aleatorio C distribuito uniformemente tra 140 e 155 kN, calcolare la probabilità di rottura del pilastro.

1.12. Risolvere l'esercizio 1.6 per $n = 5$ biglie nere estratte, con le seguenti probabilità iniziali (di tipo binomiale): $P(H_1) = 0.034$, $P(H_2) = 0.156$, $P(H_3) = 0.310$, $P(H_4) = 0.310$, $P(H_5) = 0.156$, $P(H_6) = 0.034$.

1.13. Assumendo che un amico sia baro od onesto con il 50% di probabilità, trovare la probabilità a posteriori che l'amico sia un baro dopo che egli è risultato vincitore a testa o croce per $n = 5, 10, 15$ volte consecutive.

1.14. Siano A, B e C tre eventi con probabilità diversa da zero. Dire se sono vere o false le seguenti affermazioni:

1) $P(ABC) = P(A|BC)P(B|C)P(C)$; 2) $P(AB) = P(A)P(B)$; 3) $P(A) = P(AB) + P(A\bar{B})$; 4) $P(A|BC) = P(AB|C)P(B|C)$.

1.15. Un termometro scelto a caso da un campione segna 21° Celsius. Dalle specifiche di produzione si sa che le probabilità che il termometro segni la temperatura diminuita di un grado, quella giusta e quella aumentata di un grado sono rispettivamente 0.2, 0.6, 0.2. La probabilità a priori soggettiva circa la temperatura dell'ambiente, in base ad un sondaggio, vale: $P(19^\circ) = 0.1$, $P(20^\circ) = 0.4$, $P(21^\circ) = 0.4$, $P(22^\circ) = 0.1$. Calcolare le probabilità a posteriori delle temperature.

(Suggerimento: indicare le temperature da trovare come $P(\text{vera}|\text{misurata}) \equiv P(\text{vera}|21^\circ)$.)

Rappresentazione dei fenomeni aleatori

La scienza è il corollario della convinzione che l'universo sia algoritmicamente comprimibile, e l'attuale ricerca di una teoria di ogni cosa è l'espressione ultima di questa convinzione che esista una rappresentazione abbreviata della logica sottostante alle proprietà dell'universo, e che gli esseri umani possano scrivere per esteso, in forma finita, questa rappresentazione.

John D. Barrow, *Teorie del tutto*

2.1 Introduzione

Inizieremo questo capitolo definendo meglio il formalismo, la notazione e la terminologia che ci accompagneranno per tutto il resto del testo. Questo è un passo molto importante, senza il quale si espone il lettore al rischio di fraintendere il significato di gran parte delle formule di base della probabilità e della statistica.

Proseguiremo illustrando la rappresentazione degli eventi in istogrammi, che è quella più conveniente per un adeguato sviluppo della teoria, sia probabilistica che statistica, e quella più vicina alle applicazioni.

La scelta della rappresentazione in istogrammi ci porterà a definire subito il primo tipo di distribuzione importante, quella binomiale, mentre le altre distribuzioni saranno studiate successivamente nel cap. 3.

Nel nostro sito si trova una breve guida all'uso del software libero SCILAB [SCI], nato in ambiente scientifico, che fornisce parecchie routine di utilità in campo statistico e matematico e che consente all'utente, con un linguaggio semplice ed intuitivo, di scrivere ed eseguire con estrema facilità le proprie routine (tutte quelle utilizzate nel testo si trovano nel nostro sito web).

Raccomandiamo quindi ai lettori di installare SCILAB sul proprio calcolatore, di copiare le nostre routine e di cercare, d'ora in poi, di affiancare allo studio del testo gli esercizi con SCILAB che verranno via via suggeriti.

Questo consentirà di sfruttare da subito le eccezionali possibilità che il calcolatore offre nella verifica e nella applicazione delle leggi fondamentali della probabilità, della statistica e della simulazione.

2.2 Variabili aleatorie

Avendo presente le definizioni 1.5, 1.6 e 1.7 del capitolo precedente, consideriamo uno spazio di probabilità $\mathcal{E} = (S, \mathcal{F}, P)$ e siano $a \in A \subseteq S$ i risultati di una generica prova che realizza l'evento A della σ -algebra \mathcal{F} . Ad ogni elemento a associamo un numero reale mediante una corrispondenza o funzione

$$X : S \rightarrow (-\infty, +\infty), \quad \text{ovvero} \quad X(a) = x. \quad (2.1)$$

Si dice variabile aleatoria o casuale una variabile che segue la:

Definizione 2.1 (Variabile aleatoria o casuale). *Una variabile aleatoria $X(a)$ è una funzione avente come dominio lo spazio S , come codominio la retta reale e tale che l'insieme degli elementi a per i quali vale la relazione*

$$X(a) \leq x \quad (2.2)$$

è un evento per ogni $x \in \mathbb{R}$.

Fate attenzione perché abbiamo compiuto un passo concettuale importante: la variabile aleatoria è definita come *una corrispondenza o funzione* che porta dallo spazio campionario all'asse reale. Ovviamente, sarebbe più appropriato parlare di funzione aleatoria invece che di variabile, ma questa infelice terminologia si è ormai definitivamente consolidata.

Se $a_1 \in A_1$ e $a_2 \in A_2$ sono elementi di due insiemi (eventi) di \mathcal{F} , in base alla (1.6) anche l'elemento $a \in (A_2 - A_1)$ apparterrà ad un insieme (evento) del campo. Se ora $X(a_1) \leq x_1$ e $X(a_2) \leq x_2$, $(A_2 - A_1)$ sarà l'evento corrispondente all'insieme numerico

$$x_1 < X(a) \leq x_2 \quad \text{ovvero} \quad x_1 < X \leq x_2. \quad (2.3)$$

Se la variabile aleatoria fa corrispondere agli eventi un insieme di numeri reali, dalla definizione 2.1 e dalla (1.5) discende che anche l'insieme

$$\bigcap_{n=k}^{\infty} \left\{ a : x_0 - \frac{1}{n} < X(a) \leq x_0 \right\} = \{ a : X(a) = x_0 \} \equiv \{ a : X = x_0 \} \quad (2.4)$$

dove x_0 è un numero reale, è un evento. Prendiamo come esempio l'esperimento consistente nella estrazione di una biglia numerata nel gioco del lotto. Secondo il formalismo appena introdotto la (2.2) diventa:

$$\text{ESTRAI E LEGGI (biglia)} \leq \text{numero intero}.$$

Il dominio di questa legge è l'insieme delle biglie, il codominio un sottoinsieme dell'asse reale e la variabile aleatoria è definita come "estrazione a caso di una biglia e lettura del numero".

In generale, se \mathbb{R}_0 è un qualunque sottoinsieme dell'asse reale, rappresentabile come unione (somma), intersezione (prodotto) o differenza di intervalli, è facile convincersi che anche l'insieme numerico

$$X(a) \in \mathbb{R}_0$$

può essere associato ad una σ -algebra \mathcal{F} . In breve, diremo che

$$\{a : X(a) \in \mathbb{R}_0\} \quad (2.5)$$

è un evento.

Riassumendo, nel testo useremo la notazione:

$$\{X \geq x_0\}, \quad \{X \leq x_0\}, \quad \{x_1 \leq X \leq x_2\}, \quad \{X \in \mathbb{R}_0\}, \quad (2.6)$$

per indicare un insieme $A \in \mathcal{F}$ di elementi $[a]$ ottenuti dalle prove sperimentali (non un insieme di numeri reali!) per i quali la funzione $X(a)$ soddisfa la condizione numerica espressa entro parentesi graffa. La probabilità

$$P(A) = P\{X(a) \in \mathbb{R}_0\} \quad (2.7)$$

è detta *legge o distribuzione* di X . Essa è un'applicazione $\mathbb{R}_0 \rightarrow P\{X(a) \in \mathbb{R}_0\}$ che ad ogni sottoinsieme $\mathbb{R}_0 \subseteq \mathbb{R}$ associa la probabilità che X prenda valori appartenenti ad \mathbb{R}_0 . La legge è una funzione definita su un insieme. Essa può essere messa in corrispondenza con combinazioni lineari di somme o integrali di funzioni, che sono più facili da trattare. Queste sono le funzioni cumulative o di ripartizione e le funzioni di densità, che definiremo tra poco.

Nel testo considereremo probabilità per le quali vale la condizione

$$P\{X = \pm\infty\} = 0.$$

Si noti che questa proprietà non esclude che la variabile assuma valori infiniti; tuttavia si richiede che il valore della probabilità di questi eventi sia nullo.

Queste considerazioni sono molto importanti e vanno tenute sempre presenti nella lettura del seguito. Su altri testi è possibile trovare, invece della (2.2), la notazione

$$x(a) \leq x,$$

o altre forme equivalenti. Noi useremo, per indicare le variabili aleatorie, le lettere maiuscole (in genere le ultime) e riserveremo le lettere minuscole per rappresentare i particolari valori numerici ottenuti in una prova; nei pochi casi in cui avremo conflitti di notazione useremo per le variabili aleatorie le lettere maiuscole in grassetto (\mathbf{X}). Inoltre, per capire a cosa ci riferiamo dovrete fare attenzione alle parentesi: l'espressione $\{ \dots \}$ indicherà sempre un insieme definito nello spazio campionario, corrispondente all'insieme di valori reali indicato all'interno delle parentesi graffe.

Le notazioni che utilizzeremo sono riportate in Tab. 2.1. È essenziale capire bene la differenza tra la notazione maiuscola e minuscola:

Tabella 2.1. Confronto tra la notazione usata nel testo ed altre notazioni correnti

<i>Significato</i>	<i>Notazioni del testo</i>	<i>Altre notazioni</i>
risultato di una prova	a	a
il risultato è un evento	$a \in A$	$a \in A$
variabile aleatoria	$X(a) = x$	$x(a) = x$
Se $a \in A$, $X \in \mathbb{R}_0$	$\{X \in \mathbb{R}_0\}$	$\{x \in \mathbb{R}_0\}$
Probabilità $P(A)$	$P\{X \in \mathbb{R}_0\}$	$P\{x \in \mathbb{R}_0\}$
codominio reale di $X(a)$	spettro	codominio, coinsieme
operatore su una variabile aleatoria	$O[X]$	$O[x], O(X)$
valore atteso	$E[X], \langle X \rangle$	$E(x), E(X)$
varianza	$\text{Var}[X]$	$\text{Var}[x], \sigma^2(X)$

- X indica la variabile aleatoria, che può assumere un insieme finito o infinito di valori numerici possibili, *prima* di una o più prove;
- x è un numero, cioè un particolare valore numerico assunto da X *dopo* una ben determinata prova. La n -upla di valori (x_1, x_2, \dots, x_n) è l'insieme di valori assunti da X in un campionamento.

È importante notare che la variabile aleatoria deve essere definita su tutti gli elementi dello spazio campionario S . Sia dato ad esempio uno spazio (S, \mathcal{F}, P) e sia A un evento con probabilità p , ovvero $P(A) = p$. È naturale allora definire la funzione X , detta funzione o variabile *indicatrice* tale che:

$$X(a) = 1 \text{ se } a \in A, \quad X(a) = 0 \text{ se } a \in \bar{A}. \quad (2.8)$$

X è una variabile casuale, perché:

- se $x < 0$ a $X \leq x$ corrisponde l'insieme vuoto \emptyset ;
- se $0 \leq x < 1$ a $X \leq x$ corrisponde l'insieme \bar{A} ;
- se $x \geq 1$ a $X \leq x$ corrisponde lo spazio campionario $A \cup \bar{A} = S$.

Poiché, per le proprietà della σ -algebra \mathcal{F} , se A è un evento lo sono anche gli insiemi \emptyset , \bar{A} ed S , la X soddisfa alla definizione 2.1 ed è pertanto una variabile aleatoria.

Su uno stesso spazio di probabilità si possono definire più variabili aleatorie. Ad esempio, se lo spazio campionario è formato da un insieme di persone, la variabile aleatoria $X(a)$ può essere il peso della persona e la variabile $Y(a)$ la statura. Avremo anche spesso a che fare con funzioni di una, due o più variabili aleatorie, del tipo

$$Z = f(X, Y). \quad (2.9)$$

Il dominio di Z sono gli elementi dello spazio campionario S , poiché per definizione $X \equiv X(a)$ e $Y \equiv Y(b)$ con $a, b \in S$, per cui si ha anche

$Z = f(X(a), Y(b))$. Dato che $X(a) = x$ e $Y(b) = y$ è possibile associare a Z anche una funzione “tradizionale” con dominio e codominio in campo reale:

$$z = f(x, y) .$$

Dalla discussione fatta finora dovrebbe esservi chiaro che la variabile aleatoria crea una corrispondenza tra unioni o intersezioni numerabili, differenze e complementi di insiemi di S (campo \mathcal{F}) ed intervalli del campo reale. È allora possibile dimostrare (ma è abbastanza intuitivo) che se, per ogni z reale, la disuguaglianza $f(x, y) \leq z$ è soddisfatta da unioni, intersezioni numerabili o differenze di intervalli reali delle variabili x e y , allora anche la variabile Z della (2.9) è una variabile aleatoria che obbedisce alla definizione 2.1 [Bal98], [Pap77]. Se le variabili di partenza X e Y sono variabili aleatorie definite sullo stesso spazio di probabilità, questo avviene per tutte le funzioni $f(X, Y, \dots)$ che tratteremo in seguito, così che supporremo sempre, da ora in poi, che funzioni di variabile aleatoria siano anch'esse variabili aleatorie.

È importante capire bene la differenza tra $Z = f(X, Y)$ e $z = f(x, y)$. Consideriamo ad esempio la somma

$$Z = X + Y$$

e i risultati di due prove

$$z_1 = x_1 + y_1 \quad \text{e} \quad z_2 = x_2 + y_2 .$$

La variabile aleatoria Z indica un possibile insieme di risultati (z_1, z_2, \dots) che sono gli esiti di Z in un esperimento consistente in una serie di prove in cui si fanno accadere X e Y e poi si sommano i risultati. Invece, z_1 rappresenta il valore di Z ottenuto nella prima prova, z_2 quello ottenuto nella seconda ed analogamente per X e Y .

L'indipendenza tra eventi comporta anche quella tra variabili aleatorie:

Definizione 2.2 (Variabili aleatorie indipendenti). *Se le variabili aleatorie X_i , ($i = 1, 2, \dots, n$) sono definite sullo stesso spazio di probabilità e gli eventi $\{X_i \in A_i\}$, ($i = 1, 2, \dots, n$) sono indipendenti secondo la definizione 1.10 per ogni scelta possibile degli intervalli $A_i \in \mathbb{R}_x$, dalla (1.21) si ha:*

$$P\{X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n\} = \prod_i P\{X_i \in A_i\} . \quad (2.10)$$

Si dice in questo caso che le variabili X_i sono stocasticamente indipendenti.

Nel seguito diremo semplicemente che le variabili sono indipendenti, avendo però cura di distinguere, quando necessario, questa indipendenza da quella di tipo matematico, come la dipendenza (o indipendenza) lineare, che implica l'esistenza (o meno) di equazioni lineari tra le variabili.

Per finire, un'ultima definizione:

Definizione 2.3 (Spettro). *Chiameremo spettro il codominio reale della variabile aleatoria $X(a)$. Se il codominio è un insieme numerabile, diremo che lo spettro è discreto, quando i valori possibili sono su tutto \mathbb{R} o su unioni di intervalli di \mathbb{R} , diremo che lo spettro è continuo.*

Una variabile X a spettro discreto verrà chiamata *variabile aleatoria discreta*, una variabile a spettro continuo *variabile aleatoria continua*. Sappiamo bene che il termine spettro in matematica si usa in altri contesti, come nella teoria delle trasformate; tuttavia, è abbastanza usuale, tra fisici e ingegneri, nella attività di ricerca o di laboratorio, dire che “si studia o si rappresenta lo spettro”, non “il codominio” della variabile casuale che si sta esaminando. Vi sono inoltre campi della fisica, come la spettroscopia atomica o nucleare, dove si studiano proprio le probabilità degli stati energetici, discreti o continui, che un sistema fisico può assumere. Poiché nel seguito l'uso del termine spettro non causerà conflitti di sorta, abbiamo deciso di mantenere questa dizione. La legge o distribuzione (2.7) associa una probabilità a valori (o insiemi di valori) dello spettro.

2.3 Funzione cumulativa o di ripartizione

La legge o distribuzione di una variabile aleatoria viene solitamente espressa in termini della *funzione cumulativa o di ripartizione*, che dà la probabilità che la variabile aleatoria assuma valori minori o uguali di un certo valore assegnato x .

Definizione 2.4 (Funzione cumulativa o di ripartizione). *Se X è una variabile aleatoria, continua o discreta, la funzione cumulativa o di ripartizione*

$$F(x) = P\{X \leq x\} \quad (2.11)$$

rappresenta la probabilità che X assuma un valore non superiore ad un valore assegnato x . Le funzioni cumulative verranno sempre indicate con lettere maiuscole.

Si noti che x non deve necessariamente fare parte dello spettro di X ; ad esempio, nel caso del lancio di un dado, dove $X = 1, 2, 3, 4, 5, 6$:

$$F(3.4) = P\{X \leq 3.4\} = P\{X \leq 3\} = F(3) .$$

Se $x_1 < x_2$, gli eventi $\{X \leq x_1\}$ e $\{x_1 < X \leq x_2\}$ sono incompatibili e la probabilità totale dell'evento $\{X \leq x_2\}$ è data dalla (1.10):

$$P\{X \leq x_2\} = P\{X \leq x_1\} + P\{x_1 < X \leq x_2\} ,$$

da cui, in base alla (2.11):

$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) . \quad (2.12)$$

Ricordando che la probabilità è una quantità non negativa, abbiamo quindi

$$F(x_2) \geq F(x_1) .$$

Se x_{\max} e x_{\min} rappresentano rispettivamente il massimo e il minimo dei valori assunti da X , dalla (1.13) e dalla definizione (1.9) risulta:

$$F(x) = 0 \quad \text{per } x < x_{\min} , \quad (2.13)$$

$$F(x) = \sum_{i=1}^{\infty} p(x_i) = 1 \quad \text{per } x \geq x_{\max} . \quad (2.14)$$

Risulta anche, per costruzione, che $F(x)$ è continua a destra, cioè continua in ogni punto x se ci si avvicina da destra. Questo fatto dipende dalla posizione del segno di uguale che compare nella (2.12):

$$\lim_{x_1 \rightarrow x_2} [F(x_2) - F(x_1)] = P\{X = x_2\} , \quad (2.15)$$

$$\lim_{x_2 \rightarrow x_1} [F(x_2) - F(x_1)] = 0 \quad (\text{continuità destra}) . \quad (2.16)$$

Si può dimostrare che ogni funzione cumulativa o di ripartizione soddisfa alle proprietà

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1 , \quad (2.17)$$

è non decrescente e continua a destra. Viceversa, se una funzione soddisfa queste proprietà, essa rappresenta la funzione cumulativa di una variabile aleatoria. Per dettagli matematici sulla dimostrazione potete consultare [Bal98].

Notiamo anche che dalla (2.11) si calcola facilmente la probabilità di ottenere valori maggiori di un limite assegnato x :

$$P\{X > x\} = 1 - F(x) . \quad (2.18)$$

La funzione cumulativa $F(x)$ permette di definire una grandezza che useremo spesso nel seguito:

Definizione 2.5 (Quantile). *Si chiama quantile di ordine α il più piccolo valore x_α che verifica la disuguaglianza:*

$$P\{X \leq x_\alpha\} = F(x_\alpha) \geq \alpha . \quad (2.19)$$

La disuguaglianza \geq nella (2.19) tiene conto del fatto che, per variabili discrete, $F(x_\alpha)$ può non coincidere con α quando questo valore viene assegnato a priori. Nel caso di variabili continue, $F(x_\alpha) = \alpha$ e il quantile è il valore x che ha come pedice la probabilità di osservare valori in $(-\infty, x]$.

Ad esempio, se $P\{X \leq x_\alpha\} = F(x_\alpha) = 0.25$, si può dire che $x_{0.25}$ è il quantile 0.25; se i valori α sono dati in percentuale, si dice anche che x è il 25-esimo percentile, oppure che x ha un valore tra il secondo e terzo decile, oppure tra il 20% e il 30%.

2.4 La rappresentazione dei dati

La rappresentazione più usata per trattare campioni di dati provenienti da un esperimento è detta *istogramma*. Nel caso di una variabile aleatoria discreta, l'istogramma si costruisce come segue:

- sull'asse x si riporta lo *spettro* di X ;
- sull'asse y si riporta il *numero di volte* che ogni valore dello spettro è comparso nel corso dell'esperimento.

Consideriamo un esperimento consistente in 100 prove, ogni prova essendo costituita dal lancio di 10 monete. Definiamo l'evento come il risultato di un lancio (fila ordinata di teste e croci) e sia X il numero di teste ottenuto; lo spettro di X è dato dagli 11 numeri interi $0, 1, 2, \dots, 10$. Ad ognuno di questi valori corrisponderà un numero, limitato tra 0 e 100, dato dal numero di volte che il valore (cioè quel numero di teste) si è presentato nelle 100 prove dell'esperimento. Se riportiamo in ascisse lo spettro dell'evento riportato nella prima colonna della Tab. 2.2 e in ordinate i risultati ottenuti in un esperimento reale, riportati nella seconda colonna della Tab. 2.2, otteniamo l'istogramma di Fig. 2.1. Il numero di eventi ottenuti con valore dello spettro x_i , si indica con $n(x_i)$. Poiché x_i viene detto *canale dell'istogramma (bin)*, si dice anche che $n(x_i)$ corrisponde al numero n di eventi caduti nel canale i -esimo dell'istogramma. Sussiste pure l'ovvia relazione

$$\sum_{i=1}^C n(x_i) = N, \quad (2.20)$$

Tabella 2.2. Risultati di un esperimento reale costituito da 100 prove, ognuna consistente nel lancio di 10 monete. La seconda colonna riporta il numero di prove in cui si è ottenuto il numero di teste riportato in prima colonna. La terza colonna riporta le frequenze empiriche ottenute semplicemente dividendo per 100 i valori della seconda colonna, la quarta colonna riporta le frequenze cumulative

<i>Spettro (n. teste)</i>	<i>N. di prove</i>	<i>Frequenza</i>	<i>Frequenze cumulative</i>
0	0	0.00	0.00
1	0	0.00	0.00
2	5	0.05	0.05
3	13	0.13	0.18
4	12	0.12	0.30
5	25	0.25	0.55
6	24	0.24	0.79
7	14	0.14	0.93
8	6	0.06	0.99
9	1	0.01	1.00
10	0	0.00	1.00

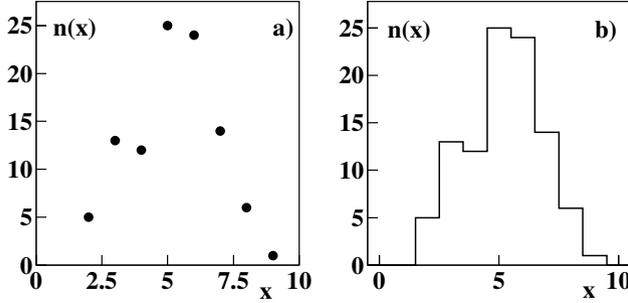


Fig. 2.1. Due modi di istogrammare il numero di prove in cui si sono ottenute x teste nel lancio di 10 monete: i valori x dello spettro, compresi tra 0 e 10, sono riportati in ascissa; il numero di successi $n(x)$, per ogni valore dello spettro, è riportato in ordinata, come un punto di coordinata x_i a) oppure come una barra larga quanto la distanza tra due valori dello spettro b). I dati, riportati anche in Tab. 2.2, si riferiscono ad un esperimento reale costituito da 100 prove

dove C è il numero dei canali dello spettro discreto ed N il numero totale di eventi, ovvero delle prove effettuate nell'esperimento. Si dice anche che l'istogramma rappresenta un *campione di N eventi*. L'istogramma così costruito ha lo svantaggio di avere le ordinate divergenti con il numero delle prove, poiché $n(x_i) \rightarrow \infty$ per $N \rightarrow \infty$. Si rimedia a questo inconveniente rappresentando il campione come un *istogramma normalizzato* dove al numero di eventi $n(x_i)$ si sostituisce la frequenza

$$f(x_i) = \frac{n(x_i)}{N} . \quad (2.21)$$

Le frequenze relative all'esempio di Fig. 2.1 sono riportate nella terza colonna di Tab. 2.2. In questo modo, quando $N \rightarrow \infty$, il contenuto dei canali resta finito e, in base alla (1.3), $f(x_i)$ tende alla probabilità $p(x_i)$ di cadere nel canale i -esimo dello spettro. L'equazione (2.20) diviene:

$$\sum_{i=1}^C f(x_i) = 1 , \quad (2.22)$$

che è detta condizione di normalizzazione.

Le frequenze dell'istogramma possono venire rappresentate anche come frequenza cumulativa, indicata solitamente con lettera maiuscola: per il k -esimo canale essa si ottiene, come in Fig. 2.2, sommando i contenuti del k -esimo canale e di tutti i canali a sinistra e dividendo per il numero totale di eventi:

$$F_k = \frac{\sum_{i=1}^k n_i}{N} = \sum_{i=1}^k f(x_i) . \quad (2.23)$$

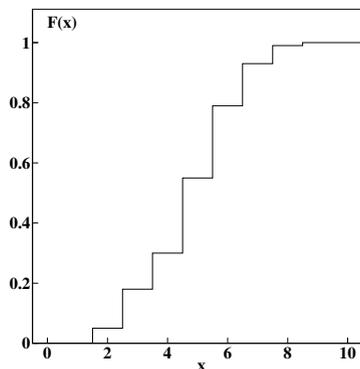


Fig. 2.2. Istogramma delle frequenze cumulative per i dati di Fig. 2.1

La frequenza cumulativa $F_k \equiv F(x_k)$ dà la percentuale di valori $x \leq x_k$ del campione sperimentale. Essa è il corrispondente “sperimentale” della funzione cumulativa (2.11).

Vediamo ora come sia possibile estendere la rappresentazione in istogrammi anche al caso di variabili continue, che possono assumere valori su un qualsiasi intervallo $[a, b]$ del campo \mathbb{R} dei numeri reali.

Siano x i valori dello spettro continuo in esame e siano compresi in $[a, b]$. Dividiamo questo intervallo in parti uguali $[x_1, x_2), [x_2, x_3), \dots, [x_{m-1}, x_m]$ di ampiezza Δx ed assegniamo ad ognuno di questi nuovi intervalli (canali o *bin*) un valore dato dal numero di eventi che hanno il valore dello spettro contenuto entro l’intervallo. Dividendo per il numero di prove effettuate, otteniamo l’analogo della (2.21):

$$f(\Delta x_k) = \frac{n(\Delta x_k)}{N}. \quad (2.24)$$

Per $N \rightarrow \infty$, in base alla (1.3) $f(\Delta x_k) \rightarrow p(\Delta x_k)$, che è la probabilità di ottenere valori dello spettro compresi entro $[x_{k-1}, x_k]$, cioè nel k -esimo canale. La rappresentazione grafica di dati di variabili continue è in genere quella dell’istogramma di Fig. 2.1b), per indicare che i valori sono distribuiti equamente entro tutto il canale; se si vuole utilizzare la rappresentazione dell’istogramma di Fig. 2.1a), l’ascissa del punto è il valore centrale del canale.

Per ottenere istogrammi al computer esiste la routine originale SCILAB `histplot(n,data)`, che esegue l’istogramma in n canali di un insieme di dati grezzi `data`. In alternativa, potete usare la nostra routine `Histplote(n,data)` che offre più opzioni statistiche (elencate nel commento al codice). Se i dati sono già raggruppati in canali, va usata la routine `Histfreqe(fre, xdata)`, dove `fre` è il vettore delle frequenze $n(x)/N$ o del numero di prove $n(x)$ e `xdata` è lo spettro della variabile. Se `fre` e `xdata` hanno la stessa dimensione, `xdata` è interpretato come il valor medio del canale oppure come lo spettro di una variabile discreta; se `xdata` ha una posizione in più di `fre`, esso viene

interpretato come il vettore degli estremi dei canali per una variabile continua. Ad esempio, per ottenere gli istogrammi di Fig. 2.1 e 2.2, basta scrivere¹ (`clf()` è una routine SCILAB che pulisce il display grafico):

```
fre = [0 0 5 13 12 25 24 14 6 1 0];
xdata= [0 1 2 3 4 5 6 7 8 9 10];
Histfreqe(fre,xdata,points=1); // istogramma a punti di Fig. 2.1a
clf(); Histfreqe(fre,xdata); // istogramma a gradino di Fig. 2.1b
clf(); Histfreqe(fre,xdata,cum=1); // istogramma di Fig. 2.2
```

2.5 Variabili aleatorie discrete

Come abbiamo visto nella definizione 2.3, una variabile aleatoria X discreta prende al più una infinità numerabile di valori (x_1, x_2, \dots, x_n) . Sappiamo anche, dalla (2.4), che gli insiemi $\{X = x_i\}$ sono degli eventi. Possiamo allora definire l'analogo probabilistico della rappresentazione in istogrammi di frequenza, la funzione di densità di probabilità:

Definizione 2.6 (Densità di probabilità per variabili discrete). *Data una variabile X discreta, la funzione $p(x)$, che vale*

$$p(x_i) = P\{X = x_i\} \quad (2.25)$$

in corrispondenza dei valori dello spettro di X e $p(x) = 0$ al di fuori di essi, è detta densità di probabilità, funzione di densità o, più semplicemente, densità.

Questa funzione gode della importante proprietà di *normalizzazione*:

$$\sum_{i=1}^{\infty} p(x_i) = \sum_{i=1}^{\infty} P\{X = x_i\} = P\left(\bigcup_{i=1}^{\infty} \{X = x_i\}\right) = P(S) = 1. \quad (2.26)$$

La conoscenza della densità permette il calcolo di leggi o distribuzioni statistiche del tipo (2.7):

$$P(A) = P\{X(a) \in \mathbb{R}_0\} = \sum_{x_i \in \mathbb{R}_0} P\{X = x_i\} = \sum_{x_i \in \mathbb{R}_0} p(x_i). \quad (2.27)$$

La funzione cumulativa o di ripartizione (2.11) si può quindi esprimere come:

$$P\{X \leq x_k\} = F(x_k) = \sum_{i=1}^k p(x_i). \quad (2.28)$$

¹ Per utilizzare le routine come `Histfreqe` occorre inizializzare SCILAB come indicato nel file `README` contenuto nella cartella dei programmi.

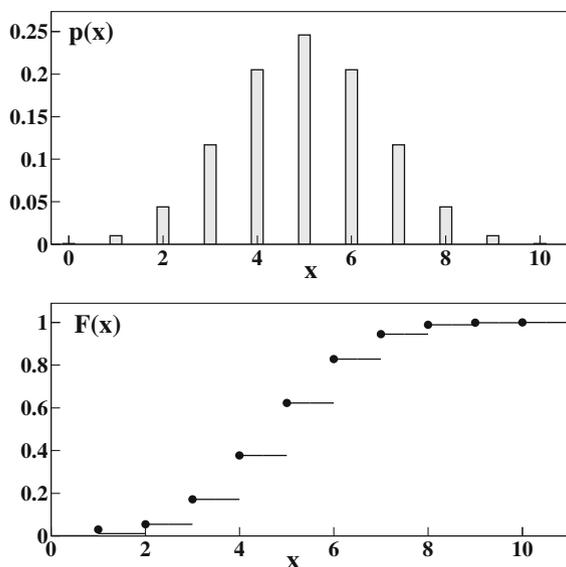


Fig. 2.3. Rappresentazione a barre della densità di probabilità binomiale $p(x) = b(x; 10, 1/2)$ e corrispondente funzione cumulativa o di ripartizione $F(x)$. Questa distribuzione è il modello probabilistico per l'esperimento delle 10 monete di Tab. 2.2. I dati sono riportati anche in Tab. 2.3

Essa può essere vista come è il corrispondente probabilistico dell'istogramma delle frequenze cumulative (2.23). Un esempio di funzioni di densità e cumulativa (o di ripartizione) è mostrato in Fig. 2.3.

Come sappiamo, la funzione cumulativa è definita per ogni x , non solo per i valori discreti assunti dalla variabile X . Quindi, nel caso di Fig. 2.3 e Tab. 2.3 abbiamo, ad esempio:

$$F(6.4) = P\{X \leq 6.4\} = P\{X = 0, 1, 2, 3, 4, 5 \text{ oppure } 6\} = 0.625 .$$

$F(x)$ presenta salti per i valori discreti di X di altezza pari a $P\{X = x_i\}$ e resta costante in $[x_k, x_{k+1})$. La continuità a destra è indicata graficamente in Fig. 2.3 con i punti in grassetto a sinistra dei valori costanti in ordinata.

L'importante relazione (2.12) può essere allora espressa in termini di densità con l'uguaglianza:

$$P\{X = x_k\} = P\{x_{k-1} < X \leq x_k\} = F(x_k) - F(x_{k-1}) = p(x_k) , \quad (2.29)$$

che permette di svolgere i calcoli probabilistici in termini di funzione di densità o di funzione cumulativa, come risulta più agevole.

Esiste un modello di distribuzione per esperimenti del tipo di quelli di Fig. 2.1 e Tab. 2.2? La risposta è affermativa: la distribuzione cercata, detta binomiale o di Bernoulli, è uno dei risultati più importanti del calcolo delle probabilità.

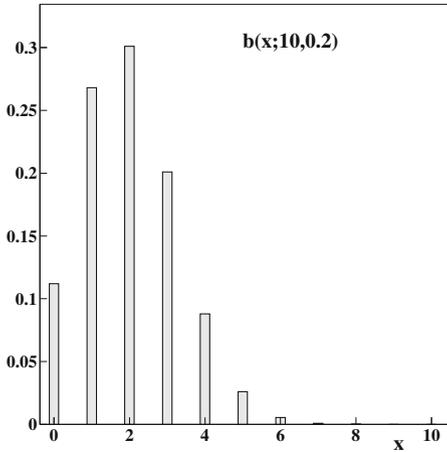


Fig. 2.4. Rappresentazione a barre della distribuzione binomiale $b(x; 10, 0.2)$

2.6 La distribuzione binomiale

Consideriamo una prova consistente in n tentativi e sia p la probabilità assegnata a priori di ottenere l'evento desiderato (successo) in ogni tentativo. Si vuole trovare la probabilità dell'evento costituito da x successi nella prova considerata. Il problema è pertanto ricondotto alla determinazione di una funzione $b(x; n, p)$ dove (fate bene attenzione!) n e p sono *parametri* assegnati e $b(x; n, p)$ è la probabilità dell'evento costituito da x successi.

Consideriamo una serie di prove tutte costituite da x successi e $n - x$ fallimenti, indicati rispettivamente con i simboli X e O :

```

X X O O O X ... X O
X O O X O O ... O X
. . . . .
X X O O O X ... X X .
    
```

In base alla legge (1.23) delle probabilità composte, se gli eventi sono *indipendenti*, la probabilità di ogni configurazione (riga) è la stessa ed è data dal prodotto della probabilità di ottenere x successi e $(n - x)$ fallimenti, vale a dire

$$p \cdot p \cdot \dots \cdot (1 - p) \cdot (1 - p) \cdot (1 - p) = p^x (1 - p)^{n-x}.$$

Quante configurazioni diverse (righe) sono possibili? Tante quante le combinazioni di n oggetti di cui x ed $n - x$ uguali tra loro. Dal calcolo combinatorio sappiamo che questo numero è dato dal coefficiente binomiale (1.30):

$$\frac{n!}{x!(n - x)!} \equiv \binom{n}{x}.$$

Dato che una prova che realizza l'evento desiderato fornisce una qualunque delle righe considerate, in base alla legge delle probabilità totali (1.17) le probabilità delle righe vanno tutte sommate per ottenere la probabilità finale dell'evento. Abbiamo pertanto l'espressione finale della densità binomiale:

$$\begin{aligned} P\{X = x \text{ successi}\} &= b(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= \binom{n}{x} p^x (1-p)^{n-x}. \end{aligned} \quad (2.30)$$

È importante ricordare sempre che questa distribuzione è valida *se e solo se* gli n tentativi sono *indipendenti* e la probabilità di successo in un tentativo è sempre *costante* e pari a p .

La distribuzione binomiale, quando $n = 1$, è detta anche di Bernoulli, a ricordo del matematico svizzero Jacques Bernoulli, che la formulò per primo, verso la fine del '600. Essa ha numerose applicazioni, come mostrato dagli esempi che seguono.

Esercizio 2.1

Calcolare la probabilità che in 10 tentativi con probabilità del 20% si ottengano 5 successi.

Risposta. Dalla (2.30) si ha immediatamente:

$$b(5; 10, 0.2) = \frac{10!}{5!5!} (0.2)^5 (0.8)^5 = 0.0264 \simeq 2.6\% .$$

Ripetendo questo conto per $0 \leq x \leq 10$ e riportando in grafico le corrispondenti probabilità, si ottiene la rappresentazione della distribuzione binomiale di Fig. 2.4.

Le probabilità della distribuzione binomiale possono anche essere ottenute con la routine SCILAB `binomial(p,n)`. Ad esempio, per ottenere il grafico per punti della distribuzione di Fig. 2.4, basta scrivere:

```
y=binomial(0.2,10); plot(y);
```

poiché, se il vettore delle ascisse non è specificato, esso viene assunto per difetto come formato da interi positivi.

Esercizio 2.2

Il 10% dei pezzi prodotti da una catena di montaggio è difettoso. Calcolare la probabilità che su 40 pezzi ne siano difettosi soltanto 2.

Risposta. Dato che $n = 40$, $p = 0.1$ e $x = 2$:

$$b(2; 40, 0.1) = \binom{40}{2} (0.1)^2 (0.9)^{38} = 0.142 .$$

Esercizio 2.3

Assumendo che la probabilità di avere figli maschi e femmine sia la medesima, calcolare la probabilità di avere 5 figlie femmine.

Risposta. Dato che $p = (1 - p) = 1/2$, $n = 5$, $x = 5$, si ha:

$$b(5; 5, 0.5) = \binom{5}{5} 2^{-5} = 3.12 \cdot 10^{-2} .$$

Esercizio 2.4

La probabilità che un bambino contragga la scarlattina in età prescolare è del 35%. Calcolare la probabilità che, in una prima classe elementare di 25 alunni, 10 alunni abbiano già fatto la scarlattina.

Risposta. Dato che $p = 0.35$, $n = 25$, $x = 10$, si ha:

$$b(10; 25, 0.35) = \binom{25}{10} (0.35)^{10} (0.65)^{15} = 0.141 .$$

Possiamo ora confrontare l'esperimento di Tab. 2.2 e Fig. 2.1 con le previsioni del calcolo delle probabilità. Per fare questo basta calcolare le probabilità $b(x; 10, 0.5)$ per $0 \leq x \leq 10$. Ovviamente, si suppone che le monete non siano truccate e che la probabilità di testa (o croce) sia costante e pari a $1/2$. I risultati sono riportati in Fig. 2.3 e nella Tab. 2.3, dove si nota che fra le frequenze sperimentali e le probabilità teoriche vi sono sensibili differenze. Se esse fossero dovute solo alla dimensione limitata dell'esperimento (100 prove), potremmo affermare che al limite, per un numero infinito di prove, si otterrebbero i valori dati dalla distribuzione binomiale. Si dice in questo caso che teoria ed esperimento sono in accordo entro le *fluttuazioni statistiche*. Se invece le differenze fossero dovute a un modello probabilistico sbagliato (il che avverrebbe nel caso di lanci correlati di monete, di monete con probabilità di testa o croce diversa da $1/2$, o di monete dotate di memoria...) dovremmo parlare di *differenze sistematiche* e non statistiche tra teoria e modello.

Chi ci può dire allora se le differenze in discussione sono fluttuazioni statistiche o sistematiche? La statistica. Ci spiace, ma dovrete aspettare fino al cap. 6 per una interpretazione definitiva e soddisfacente della Tab. 2.2. Come vedete, senza nozioni di statistica, le quali non sono affatto intuitive, *non si può interpretare correttamente nemmeno un esperimento veramente banale come il lancio delle monete!* Se ci credete sulla parola, possiamo anticiparvi che le differenze tra distribuzione binomiale ed esperimento delle 10 monete sono fluttuazioni statistiche e che l'accordo tra teoria ed esperimento è buono.

Tabella 2.3. Risultati dell'esperimento di Tab. 2.2 confrontate con le previsioni della legge binomiale. Le probabilità previste dalla distribuzione binomiale ponendo nella (2.30) $n = 10$, $p = 1/2$ sono riportate nella terza colonna, mentre la quarta e la quinta colonna riportano rispettivamente le frequenze cumulative dei dati e le probabilità cumulative calcolate inserendo nella (2.28) le probabilità della terza colonna

<i>Spettro (n. teste)</i>	<i>Frequenza</i>	<i>Probabilità</i>	<i>Frequenza cumulativa</i>	<i>Probabilità cumulativa</i>
0	0.00	0.001	0.00	0.001
1	0.00	0.010	0.00	0.011
2	0.05	0.044	0.05	0.055
3	0.13	0.117	0.18	0.172
4	0.12	0.205	0.30	0.377
5	0.25	0.246	0.55	0.623
6	0.24	0.205	0.79	0.828
7	0.14	0.117	0.93	0.945
8	0.06	0.044	0.99	0.989
9	0.01	0.010	1.00	0.999
10	0.00	0.001	1.00	1.000

2.7 Variabili aleatorie continue

Nel caso di variabili aleatorie che possono variare con continuità in un intervallo reale, finito o infinito, la definizione della funzione di densità va fatta con cautela e per passi successivi. Il procedimento è analogo a quello spesso fatto in fisica, quando, ad esempio, si passa da un insieme di cariche puntiformi entro un volume alla densità di carica, che fornisce la carica effettiva in una regione di spazio solo quando viene integrata sul volume corrispondente.

Si possono costruire variabili aleatorie continue la cui funzione di ripartizione abbia dei punti di discontinuità, ma nella pratica è molto più frequente incontrare variabili aleatorie continue con funzione di ripartizione continua. Ci limiteremo dunque allo studio di queste ultime.

Se allora vogliamo calcolare la probabilità che X assuma un certo valore, dato che $\{X = x\} \subset \{x - \varepsilon < X \leq x\}$ per ogni $\varepsilon > 0$, dalle (1.14, 2.12) risulta:

$$P\{X = x\} \leq P\{x - \varepsilon < X \leq x\} = F(x) - F(x - \varepsilon)$$

per ogni $\varepsilon > 0$. Quindi

$$\lim_{\varepsilon \rightarrow 0} [F(x) - F(x - \varepsilon)] = 0$$

per la continuità di $F(x)$. Risulta allora che *la probabilità di un valore assegnato x è sempre nulla*. Questo risultato è intuitivamente compatibile con il

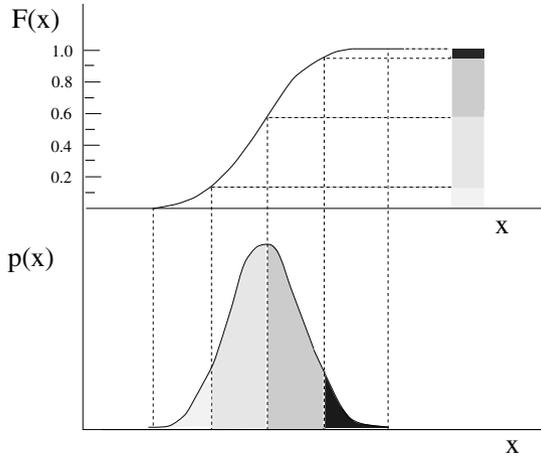


Fig. 2.5. Densità di probabilità (in basso) e corrispondente funzione cumulativa (in alto) per variabili aleatorie continue. L'ombreggiatura evidenzia come alle aree parziali della densità corrispondano gli incrementi nelle ordinate della cumulativa

concetto di probabilità classica e frequentista: lo spettro continuo comprende una infinità non numerabile di valori e la probabilità, in una prova, di ottenere *esattamente* il valore x su un insieme infinito di casi possibili (probabilità a priori) o di ottenere *esattamente* il valore x su numero infinito di prove (probabilità frequentista) deve essere nullo.

Per i valori di una variabile continua X hanno quindi senso solo le probabilità di localizzazione entro un intervallo; valgono pertanto le uguaglianze:

$$P\{a < X < b\} = P\{a \leq X < b\} = P\{a < X \leq b\} = P\{a \leq X \leq b\} ,$$

le quali mostrano che includere o meno gli estremi in un intervallo di probabilità è indifferente.

Consideriamo ora una variabile X continua che assume valori in $[a, b]$. Dividiamo questo intervallo in sottointervalli $[a = x_1, x_2], [x_2, x_3], \dots, [x_{n-1}, x_n = b]$ di ampiezza Δx_k . Definiamo una variabile aleatoria discreta X' che assume valori solo nei punti medi x'_k di questi intervalli e sia

$$p_{X'}(\Delta x_k) \equiv p_{X'}(x'_k) = P\{x_k \leq X \leq x_{k+1}\}$$

la densità di X' , che fornisce la probabilità che X sia compresa entro il k -esimo intervallo.

La densità $p_{X'}(\Delta x_k)$, essendo *funzione dell'ampiezza del canale*, è una quantità che dipende dalla scelta arbitraria di Δx , poiché lo spettro è continuo. Tuttavia, è possibile definire una funzione $p(x)$ nel modo seguente:

$$p_{X'}(\Delta x_k) = p(x_k)\Delta x_k , \tag{2.31}$$

dove x_k è un punto interno (ad esempio il punto medio) al canale k -esimo. In base alla definizione di integrale di Riemann, possiamo ora scrivere la (2.26) come:

$$\lim_{\substack{\Delta x \rightarrow 0 \\ k \rightarrow \infty}} \sum_k p_{X'}(x_k) = \lim_{\substack{\Delta x \rightarrow 0 \\ k \rightarrow \infty}} \sum_k p(x_k) \Delta x_k = \int p(x) dx = 1. \quad (2.32)$$

La funzione $p(x)$ così definita è la densità di probabilità cercata della variabile aleatoria continua X . Arriviamo pertanto alla:

Definizione 2.7 (Densità di probabilità per variabili continue). *La densità di probabilità di una variabile continua $X \in \mathbb{R}$ è una funzione $p(x) \geq 0$ che soddisfa, per ogni x , l'equazione:*

$$F(x) = \int_{-\infty}^x p(t) dt. \quad (2.33)$$

Dalla definizione risulta che la probabilità di ottenere valori in $[x_k, x_{k+1}]$ di ampiezza Δx è allora data da (vedere Fig. 2.5):

$$P\{x_k \leq X \leq x_{k+1}\} = F(x_{k+1}) - F(x_k) = \int_{x_k}^{x_{k+1}} p(x) dx, \quad (2.34)$$

e che la proprietà di normalizzazione (2.26) si scrive in questo caso come:

$$\int_{-\infty}^{+\infty} p(x) dx = 1. \quad (2.35)$$

Se l'intervallo è piccolo rispetto alla scala di variazione della funzione, si può approssimare $p(x)$ con un tratto di retta entro ogni Δx (approssimazione lineare) e la (2.34) viene sostituita dalla (2.31), dove x_k è il punto medio del canale.

Possiamo dunque indicare simbolicamente il passaggio dallo spettro discreto a quello continuo come

$$\sum_k p(x_k) \rightarrow \int p(x) dx, \quad p(x_k) \rightarrow p(x_k) dx,$$

da cui si vede che si passa da una funzione a valori discreti al prodotto di una funzione a valori continui per un differenziale. La quantità differenziale $p(x) dx$ fornisce la probabilità di ottenere valori di X contenuti in $[x, x + dx]$. Nei punti dove $F(x)$ è derivabile, dalla (2.33) si ottiene anche l'importante relazione:

$$p(x) = \frac{dF(x)}{dx}. \quad (2.36)$$

Le funzioni di densità e di ripartizione (o cumulative) hanno come dominio lo spettro della variabile aleatoria e come codominio un insieme di valori reali; esse possono essere messe in corrispondenza con leggi o distribuzioni del tipo (2.7), come mostrato in Fig. 2.6.

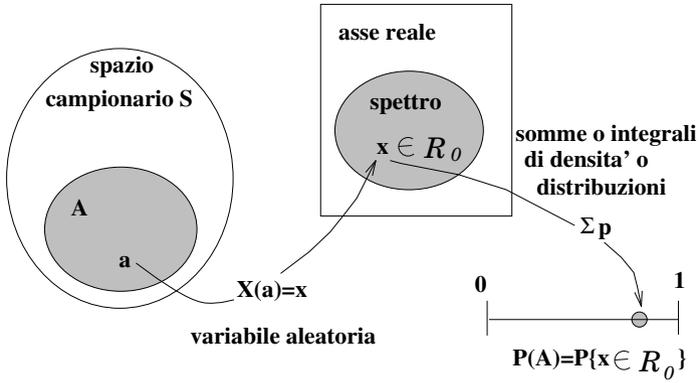


Fig. 2.6. Assegnazione di una probabilità $P(A)$ ad un evento A dello spazio campionario S . La variabile aleatoria $X(a) = x$ trasforma il risultato di una prova $a \in A$ in un codominio dell'asse reale $x \in \mathbb{R}_0$, che chiamiamo spettro; le funzioni di densità e di ripartizione permettono di calcolare, a partire dai valori dello spettro dell'evento, la probabilità $P(A)$

2.8 Media, Varianza e Deviazione Standard

La descrizione di un fenomeno aleatorio in termini di media e varianza è un modo meno completo di quello che fa uso della funzione di densità, ma che ha il pregio di essere semplice e di risultare spesso adeguato per molte applicazioni pratiche. In sostanza, la media individua dove è il baricentro dei valori della variabile X , mentre la deviazione standard, che è la radice quadrata della varianza, fornisce una stima della dispersione dei valori intorno alla media.

Nel trattare questi argomenti riteniamo utile, per aiutare la vostra intuizione, trattare prima la media e la varianza di un insieme di dati (anche se questo tema verrà approfondito più avanti, in statistica) e poi media e varianza di una distribuzione.

Introduciamo prima una notazione che ci accompagnerà per tutto il testo:

Definizione 2.8 (Parametri veri e Parametri campionari). *I parametri media, varianza e deviazione standard di una variabile aleatoria X verranno indicati con lettere greche; i parametri campionari o sperimentali, ottenuti da insiemi finiti di valori x_i , con lettere latine. La probabilità “vera” verrà invece sempre indicata con p e la frequenza relativa sperimentale con f , poiché questa è ormai la notazione standard della maggior parte dei testi di statistica.*

Definiamo ora media e varianza di un insieme di dati sperimentali:

Definizione 2.9 (Media e varianza di un insieme di dati). *Se x_i ($i = 1, 2, \dots, N$) sono le N realizzazioni numeriche di una variabile aleatoria, la*

media e la varianza campionarie sono definite rispettivamente come:

$$m = \frac{1}{N} \sum_{i=1}^N x_i, \quad (2.37)$$

$$s_\mu^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2, \quad (2.38)$$

dove μ è il valore della media assegnato a priori. La varianza rispetto alla media campionaria m della (2.37) è invece data da:

$$s_m^2 = \frac{\sum_{i=1}^N (x_i - m)^2}{(N - 1)} = \frac{N}{N - 1} \frac{\sum_{i=1}^N (x_i - m)^2}{N}. \quad (2.39)$$

Avrete notato che nel denominatore della varianza calcolata rispetto alla media campionaria compare il termine $N - 1$ invece di N : il motivo, concettualmente non banale, è di natura statistica e ve lo spiegheremo più avanti, nel cap. 6. Dal punto di vista pratico, il fattore $N/(N - 1)$ è importante solo per campioni molto piccoli. Quando si trascura questa differenza, scriveremo $s^2 \equiv s_m^2 \simeq s_\mu^2$. La radice quadrata $s = \sqrt{s^2}$ è detta deviazione standard campionaria o scarto quadratico medio. È necessario ricorrere alla radice quadrata per misurare la dispersione dei dati nella stessa unità di misura in cui è definita la media.

Dato che la differenza $(x_i - \mu)$ è detta *scarto del dato i -esimo dalla media*, la varianza può essere vista come la media dei quadrati degli scarti.

Quando i valori di una variabile X discreta sono rappresentati in istogrammi, la media e la varianza, invece che dalle (2.37-2.38), possono essere calcolate anche come:

$$m = \frac{\sum_{k=1}^C n_k x_k}{N} = \sum_{k=1}^C x_k f_k \quad (2.40)$$

$$s^2 = \frac{\sum_{k=1}^C n_k (x_k - \mu)^2}{N} = \sum_{k=1}^C (x_k - \mu)^2 f_k. \quad (2.41)$$

dove C è il numero dei canali dell'istogramma, n_k , f_k e x_k sono rispettivamente il contenuto (frequenza assoluta), la frequenza relativa e l'ascissa del canale. Nelle (2.37-2.38) la somma viene fatta direttamente su tutti i dati, come ad esempio $(3 + 3 + 4 + 2 + 4 + \dots)$, mentre nelle (2.40, 2.41) la stessa somma viene eseguita nel modo compatto $(2 + 2 \cdot 3 + 2 \cdot 4 + \dots)$. Pertanto, nel caso dello spettro discreto, i due procedimenti danno un risultato assolutamente

identico. Quando lo spettro è continuo, le formule (2.40, 2.41) possono ancora essere usate assegnando a x_k il valore del punto medio dell'ascissa del canale. Tuttavia in questo caso esse forniscono risultati approssimati.

Vediamo ora di definire media, varianza e deviazione standard quando si conosce a priori della densità di probabilità di una variabile X . Le formule, per una variabile aleatoria discreta, non sono altro che la spontanea generalizzazione delle (2.40, 2.41), ottenuta sostituendo alle frequenze misurate f_k le probabilità a priori p_k date dalla funzione di densità. Per una variabile aleatoria continua, le formule si trovano con un processo di limite analogo a quello fatto nella (2.32). Arriviamo pertanto alla seguente:

Definizione 2.10 (Media e varianza di una variabile aleatoria). *La media (o valore atteso) μ e la varianza σ^2 di una variabile X sono date da:*

$$\begin{aligned}\mu &= \sum_{k=1}^{\infty} x_k p_k \quad (\text{per variabili discrete}) , \\ &= \int_{-\infty}^{+\infty} x p(x) dx \quad (\text{per variabili continue}) ,\end{aligned}\tag{2.42}$$

$$\begin{aligned}\sigma^2 &= \sum_{k=1}^{\infty} (x_k - \mu)^2 p_k \quad (\text{per variabili discrete}) , \\ &= \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) dx \quad (\text{per variabili continue}) .\end{aligned}\tag{2.43}$$

Come nel caso di un campione, la deviazione standard è data da $\sigma = \sqrt{\sigma^2}$.

La definizione (2.42) si assume valida solo se esiste il limite della serie o dell'integrale dei valori assoluti. Dato che $p_k, p(x) \geq 0$, questa condizione equivale a porre:

$$\sum_{k=1}^{\infty} |x_k| p_k < \infty , \quad \int_{-\infty}^{+\infty} |x| p(x) dx < \infty .\tag{2.44}$$

Le (2.44) ovviamente implicano le (2.42). Perché si richieda anche l'inverso è una questione più sottile: nel calcolo delle probabilità si assume sempre che l'ordine dei termini nelle serie sia indifferente e che valgano relazioni del tipo $\sum_k x_k f(x_k) \sum_i y_i g(y_i) = \sum_{ik} x_k y_i f(x_k) g(y_i)$, e questo, come sappiamo dall'analisi matematica, è possibile solo se le serie sono assolutamente convergenti. Un discorso analogo vale nel caso degli integrali. Nel seguito l'esistenza dei valori medi implicherà sempre anche l'esistenza dei valori medi assoluti.

Per quanto riguarda la varianza, la definizione è ovviamente valida se le serie o gli integrali (2.43) (a valori sempre positivi) non sono divergenti. Vi sono tuttavia casi (per fortuna molto rari) di variabili aleatorie che non hanno varianza finita; in questo caso si deve rinunciare a caratterizzare la dispersione dei dati in termini di varianza e si deve ricorrere allo studio delle funzioni di densità o di ripartizione.

Si può mostrare che la successione di momenti calcolati dalla funzione generatrice (si veda l'Appendice B):

$$\sum_i (x_i)^k p_i, \quad \int x^k p(x) dx \quad (k = 1, 2 \dots \infty)$$

consente di determinare in modo univoco la funzione di ripartizione di una variabile casuale X [Cra51]. Il momento per $k = 1$ è la media e quindi, una volta noti i momenti, si possono determinare anche i momenti rispetto alla media:

$$\Delta_k = \sum_i (x_i - \mu)^k p_i, \quad \int (x - \mu)^k p(x) dx \quad (k = 1, 2 \dots \infty). \quad (2.45)$$

Si noti che $\Delta_2 \equiv \sigma^2$ e che il momento Δ_1 è sempre nullo. Infatti:

$$\Delta_1 = \sum_i (x_i - \mu) p_i = \sum_i x_i p_i - \mu = 0.$$

Le prime due quantità significative sono allora la media e la varianza. Come vedremo, per la quasi totalità delle applicazioni pratiche che considereremo esse sono sufficienti a definire univocamente o in modo accettabile le distribuzioni statistiche. Nel seguito utilizzeremo molto raramente i momenti di ordine superiore al secondo.

La media è forse il parametro più efficace per valutare il baricentro di una distribuzione, poiché *il momento del secondo ordine, se calcolato rispetto alla media, risulta minimo*. Infatti:

$$\frac{d\Delta_2}{d\mu} = \frac{d}{d\mu} \sum_k (x_k - \mu)^2 p_k = -2 \left[\sum_i x_k p_k - \mu \right] \quad (2.46)$$

e la derivata si annulla in un minimo solo se μ è la media (2.42), nel qual caso $\Delta_2 \equiv \sigma^2$.

La varianza, se si sviluppano i quadrati nelle (2.39, 2.43), può anche essere espressa come *differenza tra la media dei quadrati e il quadrato della media*:

$$s_\mu^2 = \sum_{k=1}^C (f_k x_k^2) - \mu^2, \quad (2.47)$$

$$s_m^2 = \frac{N}{N-1} \left[\frac{\sum_{i=1}^N x_i^2}{N} - m^2 \right] \quad (2.48)$$

$$= \frac{N}{N-1} \left[\sum_{k=1}^C (f_k x_k^2) - m^2 \right], \quad (2.49)$$

$$\sigma^2 = \sum_{k=1}^C (p_k x_k^2) - \mu^2 \rightarrow \int_{-\infty}^{+\infty} x^2 p(x) dx - \mu^2. \quad (2.50)$$

Queste relazioni sono a volte utili nei calcoli pratici, come vedremo nell'esercizio 2.6.

I momenti di una funzione di densità, varianza compresa, sono indipendenti dalla posizione della media, cioè invarianti per traslazioni lungo l'asse x . Questa proprietà è ovvia (ma importante), poiché la misura della larghezza intrinseca di una funzione non può dipendere dalla posizione di questa lungo l'asse delle ascisse. Si può comunque dimostrare questo fatto in modo formale, definendo una generica traslazione

$$x' = x + a, \quad \mu' = \mu + a,$$

$$dx' = dx, \quad p(x) \rightarrow p(x' - a) \equiv p'(x'),$$

e verificando che

$$\begin{aligned} \Delta_n(x) &= \int (x - \mu)^n p(x) dx = \int (x' - \mu')^n p(x' - a) dx' \\ &= \int (x' - \mu')^n p'(x') dx' \equiv \Delta_n(x'). \end{aligned} \quad (2.51)$$

Riassumendo, la media, la varianza ed i momenti della (2.45) nel caso di istogrammi di campioni sperimentali o di variabili aleatorie sono date, indicando con una freccia il passaggio da una variabile discreta ad una continua, da:

$$m = \sum_{k=1}^C f_k x_k, \quad (2.52)$$

$$\mu = \sum_{k=1}^{\infty} p_k x_k \rightarrow \int_{-\infty}^{+\infty} xp(x) dx, \quad (2.53)$$

$$s_{\mu}^2 = \sum_{k=1}^C f_k (x_k - \mu)^2, \quad (2.54)$$

$$s_m^2 = \frac{N}{N-1} \sum_{k=1}^C f_k (x_k - m)^2, \quad (2.55)$$

$$\sigma^2 = \sum_{k=1}^{\infty} p_k (x_k - \mu)^2 \rightarrow \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) dx, \quad (2.56)$$

$$D_n = \sum_{k=1}^C f_k (x_k - \mu)^n, \quad (2.57)$$

$$\Delta_n = \sum_{k=1}^{\infty} p_k (x_k - \mu)^n \rightarrow \int_{-\infty}^{+\infty} (x - \mu)^n p(x) dx. \quad (2.58)$$

È bene far rilevare ancora la notazione in lettere greche o latine, che ci accompagnerà nel seguito del testo.

Il calcolo di media e varianza da un insieme x di dati grezzi secondo le (2.37, 2.39), può essere effettuato con le funzioni SCILAB `mean(x)` e `variance(x)`. Se invece i dati sono in forma di istogrammi, il calcolo va fatto con le (2.53, 2.56) e vanno usate le funzioni `mediaf(xdata,fre)` e `variancef(xdata,fre)`, dove $xdata$ è lo spettro della variabile aleatoria e fre è il corrispondente vettore delle frequenze o del numero delle prove favorevoli.

2.9 Operatori

Come abbiamo visto, le formule per il calcolo della media e della varianza assumono forma diversa a seconda che l'insieme dei dati sia finito o infinito, le variabili aleatorie siano discrete o continue e i dati siano grezzi o raggruppati in istogrammi.

Per questo motivo è comodo considerare media e varianza non solo come numeri, ma anche come operatori su variabili aleatorie od insiemi di dati, con i quali si indica il tipo di operazione che si effettua, a prescindere dalla particolare rappresentazione utilizzata per i dati o le variabili.

Indicheremo con lettere latine e variabili minuscole:

$$m_x, m(x), \langle x \rangle, s_x^2, s^2(x), s_x, s(x),$$

medie, varianze e deviazioni standard ricavate da insiemi di valori sperimentali di una variabile X in una ben determinata prova o campionamento. Si noti il simbolo $\langle x \rangle$ per la media, che è una notazione molto usata da fisici e ingegneri.

L'operatore media di una variabile aleatoria viene solitamente indicato con $\langle \dots \rangle$ oppure con $E[\dots]$, l'operatore varianza con $\text{Var}[\dots]$. Le variabili su cui agiscono gli operatori saranno sempre indicate con lettere maiuscole:

$$\langle X \rangle, E[X], \text{Var}[X]. \quad (2.59)$$

Nel testo verrà usata preferenzialmente la notazione $\langle X \rangle$, $\text{Var}[X]$. Per l'operatore deviazione standard useremo la notazione $\sqrt{\text{Var}[X]} \equiv \sigma[X]$.

A questo punto vi consigliamo di riguardare ancora la Tab. 2.1, dalla quale risulta che le funzioni che operano sulle variabili aleatorie sono le probabilità e gli operatori: *la variabile aleatoria risulta quindi sempre racchiusa tra parentesi graffe, quadre o tra i simboli $\langle \dots \rangle$* . Useremo questa convenzione in tutto il seguito del testo.

Il mettere in evidenza in tutte le formule l'aspetto operatoriale di media e varianza, sebbene sia corretto, non è sempre indispensabile ed appesantisce notevolmente il formalismo. Per alleggerire la notazione, noi indicheremo talvolta media, varianza e deviazione standard come:

$$\langle X \rangle = \mu_x = \mu, \quad \text{Var}[X] = \sigma_x^2 = \sigma^2, \quad \sigma[X] = \sigma_x = \sigma \quad (2.60)$$

(notate i pedici scritti con lettere minuscole). La scrittura μ e σ^2 (oppure μ_x e σ_x^2 se è necessario specificare il tipo di variabile) va intesa come il risultato

numerico del corrispondente operatore statistico. La scrittura in lettere greche (valori veri), definisce comunque in modo univoco il tipo di operazione fatta, evitando possibili confusioni.

È facile ora verificare, ricorrendo alle (2.52-2.56), che gli operatori media e varianza godono delle proprietà seguenti:

$$\text{Var}[X] = \langle (X - \langle X \rangle)^2 \rangle, \quad (2.61)$$

$$\langle \alpha X \rangle = \alpha \langle X \rangle, \quad (2.62)$$

$$\text{Var}[\alpha X] = \alpha^2 \text{Var}[X], \quad (2.63)$$

$$\langle X + \alpha \rangle = \langle X \rangle + \alpha, \quad (2.64)$$

$$\begin{aligned} \text{Var}[X + \alpha] &= \langle (X + \alpha - \langle X + \alpha \rangle)^2 \rangle \\ &= \langle (X + \alpha - \langle X \rangle - \alpha)^2 \rangle = \text{Var}[X], \end{aligned} \quad (2.65)$$

dove α è una costante. Le ultime due relazioni mostrano, come è ovvio, che la media di una costante coincide con la costante e che la dispersione di una costante è nulla. Sfruttando le (2.61, 2.62), possiamo riscrivere in notazione operatoriale la (2.50), che esprime la varianza come media dei quadrati meno il quadrato della media:

$$\begin{aligned} \text{Var}[X] &= \langle (X - \mu)^2 \rangle \\ &= \langle (X - \langle X \rangle)^2 \rangle \\ &= \langle X^2 \rangle - 2 \langle X \rangle \langle X \rangle + \langle X \rangle^2 \\ &= \langle X^2 \rangle - \langle X \rangle^2. \end{aligned} \quad (2.66)$$

L'operatore di media permette la definizione in generale del valore medio vero di una qualunque funzione di variabile aleatoria del tipo (2.9):

Definizione 2.11 (Valore atteso). *Se $f(X)$ è una funzione di una variabile aleatoria X avente densità di probabilità $p(x)$, si definisce come valore di aspettazione o valore atteso (expected value) di $f(X)$ la quantità:*

$$\langle f(X) \rangle \equiv E[f(X)] = \sum_k f(x_k) p(x_k) \rightarrow \int f(x) p(x) dx, \quad (2.67)$$

dove con la freccia si è indicato il passaggio da una variabile X discreta ad una continua.

Le funzioni di variabili aleatorie verranno trattate in dettaglio più avanti, nel cap. 5. In base a questa definizione, la media vera μ può essere considerata come il valore atteso di X . Termini sinonimi di valore atteso sono aspettazione, valore di aspettazione, speranza matematica o semplicemente valore medio.

Come nel caso della definizione 2.10, perché la definizione di valore atteso di una variabile $f(X)$ abbia senso, si deve imporre che esista la somma $\sum_k |f(x_k)| p(x_k)$ o l'integrale $\int |f(x)| p(x) dx$. Per questo si assume sempre che una variabile o funzione di variabile, per la quale si definisce il valore atteso

(2.67), sia *assolutamente sommabile o integrabile* sulla corrispondente densità di probabilità. In pratica, si richiede, per una variabile continua, che la densità di probabilità tenda a zero per $x \rightarrow \pm\infty$ almeno come $1/|x|^\alpha$ con $\alpha > 2$. Tutte le densità che considereremo in seguito godono di questa proprietà.

Esercizio 2.5

Si consideri uno spazio \mathcal{E} in cui l'evento (A_1) può realizzarsi col valore x_1 con probabilità p_1 e l'evento (A_2) , indipendente dal primo, col valore x_2 con probabilità p_2 . Calcolare la media della somma $(X_1 + X_2)$ dove X_1 e X_2 sono due variabili indicatrici definite come $X_1(A_1) = x_1$, $X_2(A_2) = x_2$ e $X_1(\bar{A}_1) = 0$, $X_2(\bar{A}_2) = 0$.

Risposta. Lo spettro della somma è dato dai quattro valori $(0 + 0)$, $(0 + x_2)$, $(x_1 + 0)$, $(x_1 + x_2)$, aventi, in base al teorema delle probabilità composte (1.20), probabilità rispettivamente pari a $(1 - p_1)(1 - p_2)$, $(1 - p_1)p_2$, $p_1(1 - p_2)$ e p_1p_2 . In base alla (2.53) si ha allora:

$$\begin{aligned} \mu &= (1 - p_1)(1 - p_2)(0 + 0) + (1 - p_1)p_2(0 + x_2) \\ &\quad + p_1(1 - p_2)(x_1 + 0) + p_1p_2(x_1 + x_2) \\ &= p_1x_1 + p_2x_2 . \end{aligned}$$

Poiché $\langle X_1 \rangle = (1 - p_1) \cdot 0 + p_1x_1 = p_1x_1$, e lo stesso vale per X_2 , si può anche scrivere:

$$\langle X_1 + X_2 \rangle = \langle X_1 \rangle + \langle X_2 \rangle ,$$

da cui si vede che la media della somma delle due variabili aleatorie è uguale alla somma delle medie.

Esercizio 2.6

Calcolare media e deviazione standard dei dati di Tab. 2.2, e confrontarli con i valori dati dalla corrispondente distribuzione binomiale.

Risposta. Applicando la (2.52) ai dati della prima e terza colonna della tabella, si ottiene:

$$\langle x \rangle = m = (2 \cdot 0.05 + 3 \cdot 0.13 + \dots + 9 \cdot 0.01) = 5.21 ,$$

corrispondente ad un totale di 521 teste su 1000 lanci. La media dei quadrati è data da:

$$\langle x^2 \rangle = (4 \cdot 0.05 + 9 \cdot 0.13 + \dots + 81 \cdot 0.01) = 29.7 .$$

Dalla (2.49) abbiamo allora

$$\begin{aligned} s^2 &= \frac{N}{N-1} \left(\langle x^2 \rangle - \langle x \rangle^2 \right) = \frac{100}{99} (29.7 - 5.21^2) = 2.48 , \\ s &= \sqrt{2.48} = 1.57 . \end{aligned}$$

Applicando la stessa procedura alle probabilità vere della quarta colonna, otteniamo:

$$\begin{aligned}\langle X \rangle &= 5.00, & \langle X^2 \rangle &= 27.5. \\ \text{Var}[X] &= \sigma^2 = \langle X^2 \rangle - \langle X \rangle^2 = (27.5 - 25) = 2.5, \\ \sigma &= \sqrt{2.5} = 1.58.\end{aligned}$$

Si noti l'assenza del fattore 100/99 nel calcolo della varianza, poiché qui essa è calcolata rispetto alla media vera. Anche in questo caso, la differenza tra i valori sperimentali $m = 5.21$, $s = 1.57$ e teorici $\mu = 5.00$, $\sigma = 1.58$, verrà interpretata nel cap. 6.

2.10 Il campione casuale

Consideriamo un esperimento che coinvolge una variabile aleatoria X ed effettuiamo N prove indipendenti. Il risultato di questa operazione sia una N -upla di variabili (X_1, X_2, \dots, X_N) , per le quali valga la condizione di indipendenza (2.10). Arriviamo quindi alla:

Definizione 2.12 (Campione casuale). *L'insieme delle N variabili (X_1, X_2, \dots, X_N) indipendenti ed aventi densità di probabilità $p(x)$ viene detto campione casuale di dimensione N estratto dalla popolazione di densità $p(x)$.*

La dizione, corretta ma un po' lunga, di "popolazione di densità $p(x)$ ", è a volte abbreviata con "popolazione $p(x)$ ". Il concetto di popolazione, introdotto su basi intuitive nel par. 1.2, assume qui un significato preciso.

Nel caso di un campione di una variabile discreta, la probabilità di ottenere un certo insieme di valori campionati, in base alla (2.10), vale:

$$P\{X_1 = x_1, X_2 = x_2, \dots, X_N = x_N\} = \prod_{i=1}^N P\{X_i = x_i\} = \prod_{i=1}^N p(x_i). \quad (2.68)$$

Per capire il concetto di campione casuale ci si deve riferire ad una situazione in cui una variabile aleatoria X è campionata ripetutamente, secondo lo schema seguente:

$$\begin{array}{ll} \text{primo campione} & x'_1, x'_2, \dots, x'_N \rightarrow m_1, s_1^2, \\ \text{secondo campione} & x''_1, x''_2, \dots, x''_N \rightarrow m_2, s_2^2, \\ \text{terzo campione} & x'''_1, x'''_2, \dots, x'''_N \rightarrow m_3, s_3^2, \\ \dots & \dots \\ \text{tutti i campioni} & X_1, X_2, \dots, X_N \rightarrow M, S^2 \end{array} \quad (2.69)$$

X_1 rappresenta la variabile aleatoria “risultato della prima estrazione” o “primo elemento”, mentre x_1'' è il numero ottenuto alla prima estrazione nel secondo campione. I valori m_i ed s_i^2 sono rispettivamente la media e la varianza del campione i -esimo:

$$m = \sum_k \frac{x_k}{N}, \quad s^2 = \sum_k \frac{(x_k - \mu)^2}{N}.$$

Questi valori, che stimano con un campione di dimensione finita le corrispondenti quantità vere μ e σ^2 , sono dette *stime* della media e della varianza.

Se ripetiamo la prova o il campionamento, otterremo ad ogni prova media e varianza differenti: medie e varianze campionarie m ed s vanno quindi considerate come realizzazioni delle variabili aleatorie M ed S , indicate nell'ultima riga della (2.69), che sono funzioni, nel senso della (2.9), delle variabili X_i :

$$M = \sum_i \frac{X_i}{N}, \quad S^2 = \sum_i \frac{(X_i - \mu)^2}{N}. \quad (2.70)$$

Le variabili M ed S^2 sono funzioni del campione. In generale, queste quantità si chiamano *statistiche*:

Definizione 2.13 (Statistica). *Dato un campione (X_1, X_2, \dots, X_N) da una densità $p(x)$, una funzione*

$$T = t(X_1, X_2, \dots, X_N) \quad (2.71)$$

che non contiene alcun parametro incognito, è una variabile aleatoria che prende il nome di statistica.

Media e varianza campionarie sono quindi due *statistiche* che si possono ottenere dal campione. Come vedremo nel prossimo paragrafo, esse sono anche *stimatori* della media e della varianza.

2.11 Criteri di convergenza

Occorre a questo punto precisare bene il significato dei limiti ed i criteri di convergenza utilizzati nello studio delle variabili aleatorie.

Come abbiamo detto, il limite che compare nella definizione di probabilità frequentista (1.3) è fatto sulle realizzazioni numeriche della variabile aleatoria: esso non ha un significato matematico preciso ed indica *che le prove vanno ripetute un numero infinito o finito di volte, ma comunque fino ad esaurire la popolazione*. Questo è il significato da attribuire ad un limite ogni volta che esso è applicato ad una successione o somma di valori di una variabile (notazione in lettere minuscole). Abbiamo chiamato questa operazione con il nome di *limite frequentista*.

Va però subito chiarito che, in genere, si devono evitare scritte del tipo:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_k x_k n_k = \lim_{N \rightarrow \infty} \sum_k n_k f_k = \sum_k x_k p_k = \mu \quad (\text{da evitare!}),$$

perché nei termini più a destra compaiono quantità matematiche ben definite come somme, serie o funzioni, mentre a sinistra compare un limite matematicamente indefinito. Facciamo notare, tuttavia, che questo limite è giustificato nell'interpretazione frequentista (1.3) e che esso riproduce *qualitativamente* ciò che succede in molti fenomeni aleatori. Ad esempio, se consideriamo il lancio di un dado $\{X = 1, 2, 3, 4, 5, 6\}$ ed assegniamo ad ogni faccia la probabilità di $1/6$, secondo la (2.42) $\mu = (1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$. L'esperienza mostra che, lanciando un dado e facendo la media progressiva dei punteggi, secondo la (2.37), il risultato tende a 3.5 aumentando il numero dei lanci. Assegnando arbitrariamente ad ogni faccia la probabilità di $1/6$, abbiamo simulato mediante calcolatore il lancio di un dado ottenendo, per $N = 100, 1000, 100\,000, 1\,000\,000$, rispettivamente le medie campionarie $m = 3.46, 3.505, 3.50327, 3.500184$.

Per uno studio matematicamente rigoroso dei fenomeni aleatori, è però necessario stabilire il tipo di convergenza di successioni di variabili aleatorie e l'entità dello scostamento dal valore limite all'aumentare della dimensione del campione.

Una prima definizione rigorosa afferma che una successione di variabili X_N converge verso una variabile X se, prefissato un numero $\epsilon > 0$ comunque piccolo, la probabilità che X_N differisca da X per una quantità $> \epsilon$ tende a zero per $N \rightarrow \infty$:

$$\lim_{N \rightarrow \infty} P\{|X_N(a) - X(a)| \geq \epsilon\} = 0, \quad \text{o} \quad \lim_{N \rightarrow \infty} P\{|X_N(a) - X(a)| \leq \epsilon\} = 1. \quad (2.72)$$

Questo limite, che gode delle proprietà usuali dei limiti di successioni di numeri reali, viene detto *limite in probabilità*.

A questo punto dobbiamo farvi notare un aspetto sottile del problema: dire che la probabilità nella (2.72) tende a zero non assicura affatto che tutti i valori $|X_N(a) - X(a)|$, per ogni elemento a dello spazio campionario, si manterranno minori di ϵ al di sopra di un certo N , ma solo che l'insieme dei valori che superano ϵ ha una probabilità piccola, al limite nulla. Se si vuole che per la maggior parte delle *successioni* si abbia $|X_N(a) - X(a)| \leq \epsilon$ per ogni a , si deve introdurre il *limite quasi certo* o in probabilità 1, sull'insieme degli elementi a dello spazio di probabilità (S, \mathcal{F}, P) :

$$P\{\lim_{N \rightarrow \infty} |X_N(a) - X(a)| \leq \epsilon\} = 1; \quad (2.73)$$

in questo caso siamo sicuri che esiste un insieme di elementi a , che per $N \rightarrow \infty$ tende a coincidere con lo spazio campionario S , per i quali la successione di numeri reali $X_N(a)$ tende ad un limite nel senso ordinario dell'analisi matematica. I due tipi di convergenza in probabilità e quasi certa sono schematizzati in Fig. 2.7.

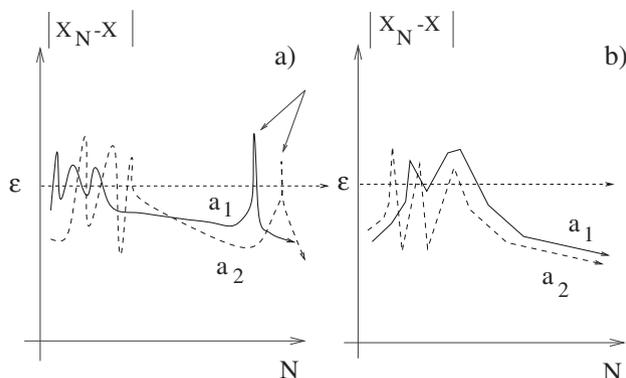


Fig. 2.7. a) Convergenza in probabilità: l'insieme dei valori $|X_N(a) - X(a)|$ che superano per $N \rightarrow \infty$ un valore ϵ assegnato tende ad avere probabilità nulla. Questo non impedisce però che vi siano singoli valori fuori dal limite (indicati dalle frecce). b) Convergenza quasi certa: la maggior parte delle successioni verificano la disuguaglianza $|X_N(a) - X(a)| \leq \epsilon$ a meno di un insieme di elementi $a \in S$ di probabilità nulla

Il terzo ed ultimo tipo di convergenza che si deve considerare è quello in legge o distribuzione. Una successione di variabili X_N converge in legge ad una variabile X se, essendo F_{X_N} e F_X le rispettive funzioni di ripartizione o cumulative,

$$\lim_{N \rightarrow \infty} F_{X_N}(x) = F_X(x) \quad (2.74)$$

per ogni punto x in cui $F_X(x)$ è continua. Questo limite è molto utilizzato in statistica quando si studia il tipo di distribuzione da assegnare, per grandi campioni, agli stimatori statistici fondamentali.

Si può dimostrare (come del resto è abbastanza intuitivo) che la convergenza quasi certa implica sia quella in probabilità sia quella in legge e che la convergenza in probabilità implica quella in legge, mentre non vale il contrario. Inoltre, un fondamentale teorema di Kolmogorov, che potete trovare dimostrato in [Fel47], assicura la convergenza quasi certa di successioni di variabili aleatorie indipendenti se vale la condizione

$$\sum_N \frac{\text{Var}[X_N]}{N^2} < +\infty. \quad (2.75)$$

Tutti gli stimatori statistici considerati nel testo soddisfano questa proprietà, per cui, quando affronteremo la statistica, non avremo problemi di convergenza: *tutte le variabili che considereremo convergeranno quasi certamente (e quindi anche in probabilità) ai limiti stabiliti*. I criteri di convergenza sono a volte importanti nella teoria delle funzioni stocastiche [Pap77], che noi però non considereremo.

Riassumendo, abbiamo considerato quattro differenti limiti: quello frequentista (che agisce sulle variabili “minuscole”), quello in probabilità e quello quasi certo (che agiscono entrambi sulle variabili “maiuscole”) e quello in legge, che riguarda le funzioni di ripartizione.

Le variabili aleatorie limite nelle (2.72, 2.73) possono anche semplicemente essere delle costanti, come spesso accade considerando i limiti degli stimatori statistici.

Infatti, una statistica $T_N(X)$, funzione di un campione casuale di dimensione N secondo la (2.71), che converge in probabilità ad un valore costante

$$\lim_{N \rightarrow \infty} P\{|T_N(X) - \mu| > \epsilon\} = 0, \quad (2.76)$$

è detta stimatore consistente di μ .

La teoria degli stimatori statistici, come quelli definiti nelle (2.70), verrà svolta in dettaglio più avanti, nei capitoli 6 e 9. Tuttavia, qui vogliamo illustrarvi subito il significato di media e varianza di uno stimatore. Dato che dalle (2.66) risulta che la varianza può essere ricondotta ad una operazione di media, sarà sufficiente discutere soltanto la media di uno stimatore.

Chiediamoci ad esempio cosa significa la media della varianza:

$$\langle T_N(X) \rangle \equiv \left\langle \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2 \right\rangle.$$

Ciò che sta tra parentesi $\langle \rangle$ è una variabile aleatoria costruita eseguendo N prove della variabile X , ottenendo altrettanti valori dello spettro x_i e calcolandone la varianza. La parentesi $\langle \rangle$ indica che si deve ripetere il procedimento un numero infinito di volte e fare la media delle infinite varianze così ottenute. Pertanto, nella visione frequentista, *prima* si ottiene da una serie di N' campioni di N elementi della stessa variabile X un campione della variabile $\sum_i (X_i - \mu)^2 / N$, *poi* si fa la media di quest'ultimo campione per $N' \rightarrow \infty$. Il procedimento è schematizzato in Fig. 2.8. Applicando le proprietà (2.66) dell'operatore media, si ottiene:

$$\left\langle \sum_{i=1}^N (X_i - \mu)^2 \right\rangle = \sum_{i=1}^N \langle (X_i - \mu)^2 \rangle = N\sigma^2, \quad (2.77)$$

ovvero:

$$\left\langle \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2 \right\rangle = \sigma^2.$$

Questa relazione può essere scritta come:

$$\langle T_N(X) \rangle = \sigma^2. \quad (2.78)$$

Gli stimatori consistenti che godono di questa proprietà sono detti corretti o non distorti. In sostanza, per questa classe di stimatori la media vera di una

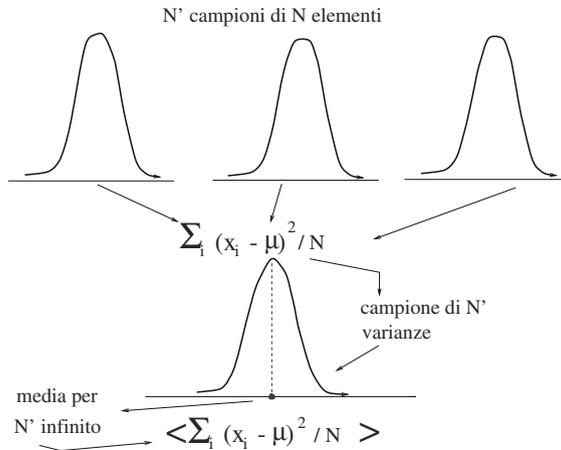


Fig. 2.8. Operazione di media di una varianza

popolazione costituita da elementi T_N , cioè da stimatori ottenuti da campioni di dimensione N , coincide con il limite all'infinito (valore vero) dello stimatore.

Mostreremo più avanti, nei parr. 6.10 e 6.11, che la media campionaria (2.52) gode delle proprietà (2.76) e (2.78), mentre la varianza rispetto alla media campionaria (2.55), privata del fattore $N/(N - 1)$, non soddisfa la (2.78).

In genere lo studio delle proprietà asintotiche delle variabili aleatorie e degli stimatori ad esse associati risulta più semplice se si applica l'operatore di media ad un insieme di stimatori $T_N(X)$, invece di studiare direttamente il limite dello stimatore (2.72) per $N \rightarrow \infty$.

Problemi

- 2.1.** Calcolare la probabilità di ottenere 2 volte 6 nel lancio di tre dadi.
- 2.2.** Assegnando all'evento testa (o croce) probabilità 1/2, calcolare la probabilità di ottenere 3 teste in 10 lanci come rapporto tra i casi favorevoli e quelli possibili.
- 2.3.** Un produttore sa che la percentuale di pezzi difettosi messi in vendita è del 10%. In un appalto accetta di pagare una penale se, in una scatola di dieci pezzi, più di due risultano difettosi. Trovare la probabilità di pagare la penale.
- 2.4.** Un giocatore scommette 60 euro alla roulette puntando 50 euro sul rosso (strategia X) e 10 euro sul numero nero 22 (strategia Y). Tenendo presente che i numeri vanno da 0 a 36 e che il banco vince se esce lo 0 mentre paga una quota pari alla posta se esce il rosso e a 36 volte la posta se esce il numero 22, calcolare il valor medio del capitale dopo una puntata. Che significato ha questo risultato?

- 2.5.** Una variabile X può assumere i valori $\{X = 2, 4, 6, 8, 10\}$ con uguale probabilità. Trovare media e deviazione standard delle variabili X e $Y = 2X + 1$.
- 2.6.** Da un'urna contenente 3 biglie rosse e 7 nere vengono estratte, senza reimmissione, tre biglie. Determinare la densità di probabilità discreta, la media e la deviazione standard del numero R di biglie nere dopo le tre estrazioni.
- 2.7.** Trovare la densità di probabilità $p(x)$, la funzione di ripartizione $F(x)$, media e deviazione standard di una variabile continua X , definita in $[0, 1]$, avente densità linearmente crescente e tale che $p(0) = 0$.
- 2.8.** Trovare il 25-esimo percentile della distribuzione del problema precedente.
- 2.9.** Per coprire la distanza di andata e ritorno tra due punti A e B , distanti 100 Km, un'auto viaggia alla velocità di 25 Km/h all'andata e di 50 Km/h al ritorno. Calcolare la velocità media del viaggio.
- 2.10.** Dai dati di Tab. 2.2 trovare il numero totale di teste sul totale di mille lanci.

Calcolo elementare delle probabilità

La questione non è tanto se Dio giochi a dadi, ma come lo faccia.

Ian Stewart, *Dio gioca a dadi?*

3.1 Introduzione

In questo capitolo approfondiremo le proprietà della distribuzione binomiale e dedurremo via via, utilizzando il calcolo delle probabilità, tutte le distribuzioni statistiche fondamentali.

Nello sviluppare questi argomenti, ci sforzeremo di sottolinearne più gli aspetti concettuali che quelli formali.

Non eviteremo comunque nessun passaggio matematico importante, poiché riteniamo che questo aiuti ad avere quella visione coerente delle cose che vi permetterà, analizzando un qualunque problema, di capirne immediatamente gli aspetti statistici e probabilistici e di trovarne la soluzione utilizzando le distribuzioni statistiche più appropriate.

Termineremo il capitolo accennando a come utilizzare il calcolo delle probabilità nella verifica delle ipotesi. Questo argomento, che sarà sviluppato in modo completo più avanti, in statistica, vi farà apprezzare meglio ciò che avrete imparato e vi consentirà di compiere un salto di qualità notevole nella comprensione di molti fenomeni naturali.

3.2 Proprietà della distribuzione binomiale

La densità binomiale, introdotta nel par. 2.6, è la legge di una variabile aleatoria X che rappresenta il numero x di successi che si possono ottenere in n tentativi indipendenti, ognuno dei quali ha una probabilità di successo costante.

Le proprietà di questa densità ci permetteranno di costruire per passi successivi lo schema di base del calcolo delle probabilità per variabili aleatorie monodimensionali.

La densità binomiale è normalizzata, poiché dalla formula del binomio di Newton e dalla (2.30), si ha:

$$\sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1. \quad (3.1)$$

La media della distribuzione binomiale è data dalla (2.42), dove la somma va estesa a tutti i valori $0 \leq x \leq n$, e la probabilità p_k è data dalla (2.30):

$$\begin{aligned} \mu &= \sum_{x=0}^n x b(x; n, p) = \sum_{x=0}^n x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= \sum_{x=1}^n x \frac{n(n-1)!}{x(x-1)!(n-x)!} p p^{x-1} (1-p)^{n-x} \end{aligned} \quad (3.2)$$

dove va notato, nell'ultimo membro, il cambiamento della somma su x da 0 a 1, poiché per $x = 0$ il termine della sommatoria è nullo. Ponendo ora nella (3.2)

$$x' = x - 1, \quad n' = n - 1, \quad n' - x' = n - x \quad (3.3)$$

otteniamo, utilizzando la (3.1):

$$\langle X \rangle = np \sum_{x'=0}^{n'} \frac{n!}{x'!(n'-x')!} p^{x'} (1-p)^{n'-x'} = np, \quad (3.4)$$

un risultato in accordo con l'intuizione, che considera la media come valore atteso, cioè il prodotto dei tentativi per la probabilità di ogni tentativo.

Niente affatto intuitivo e, come vedremo, estremamente importante, è il valore della varianza, alla quale si può arrivare facilmente calcolando la media dei quadrati con lo stesso procedimento della (3.2) ed utilizzando ancora la notazione (3.3):

$$\begin{aligned} \langle X^2 \rangle &= \sum_{x=1}^n x^2 \binom{n}{x} p^x (1-p)^{n-x} = np \sum_{x'=0}^{n'} (x'+1) \binom{n'}{x'} p^{x'} (1-p)^{n'-x'} \\ &= np(n'p + 1) = np[(n-1)p + 1]. \end{aligned}$$

In base alle (2.66) e (3.4), si può allora scrivere:

$$\text{Var}[X] = np[(n-1)p + 1] - n^2 p^2 = np(1-p). \quad (3.5)$$

In definitiva, si ottengono le formule fondamentali:

$$\mu = np, \quad \sigma^2 = np(1-p), \quad \sigma = \sqrt{np(1-p)} \quad (3.6)$$

le quali stanno alla base del calcolo delle probabilità. Il valore della media è intuitivo, quelli della varianza e della deviazione standard non lo sono, ed è utile che li mandiate a memoria fin da ora.

Esercizio 3.1

Trovare media, varianza e deviazione standard vere per l'esperimento delle 10 monete riportato in Tab. 2.2.

Risposta. Poiché nell'esperimento delle 10 monete si è adottata la distribuzione binomiale con $n = 10$ e $p = 0.5$, dalle (3.6) si ha immediatamente:

$$\begin{aligned}\mu &= 10 \cdot 0.5 = 5 \\ \sigma^2 &= 10 \cdot 0.5 (1 - 0.5) = 2.5 \\ \sigma &= \sqrt{2.5} = 1.58 .\end{aligned}$$

I risultati sono identici a quelli ottenuti nell'esercizio 2.6, dove si sono applicate le formule di base (2.53, 2.50) alle probabilità binomiali riportate nella quarta colonna della Tab. 2.2.

Esercizio 3.2

La probabilità di centrare un bersaglio è pari all'80%. Trovare la densità di probabilità del numero n di prove necessarie per effettuare un centro. Quanto valgono la media e la deviazione standard di n ?

Risposta. Si chiede di trovare la probabilità del numero di tentativi n da effettuare per avere successo, quando la probabilità è p .

In questo caso la variabile aleatoria da considerare non è più costituita dal numero X di successi in n tentativi fissati (che porta alla distribuzione binomiale), ma dal numero di tentativi n essendo il numero di successi fissato e pari a $x = 1$. Affinché il tentativo n -esimo sia coronato da successo, è necessario avere prima $x = 0$ successi in $n - 1$ prove; la probabilità di questo evento è data dalla densità binomiale (2.30):

$$b(x = 0; n - 1, p) = \frac{(n - 1)!}{(n - 1)!} (1 - p)^{n-1} = (1 - p)^{n-1} .$$

La densità che cerchiamo, che è la probabilità di avere un successo alla n -esima prova, sarà quindi data dalla probabilità composta:

$$g(n) = p(1 - p)^{n-1} , \quad (3.7)$$

che prende il nome di densità geometrica. Essa può quindi essere vista come una semplice applicazione della legge delle probabilità composte (1.23), dove le due probabilità si riferiscono ad $n - 1$ insuccessi consecutivi seguiti da un successo alla n -esima prova.

La densità è normalizzata, perché è noto dalla teoria delle serie che

$$\sum_{n=0}^{\infty} p^n = \frac{1}{1 - p} , \quad \sum_{n=1}^{\infty} p^n = \frac{p}{1 - p} , \quad 0 \leq p < 1 , \quad (3.8)$$

e quindi:

$$\sum_{k=1}^{\infty} p(1-p)^{k-1} = p \sum_{k=1}^{\infty} (1-p)^{k-1} = p \frac{1}{p} = 1, \quad 0 < p \leq 1.$$

Media e varianza si trovano dalle (2.42) e (2.66) sfruttando le proprietà della serie geometrica. Differenziando due volte la seconda delle (3.8), dopo facili calcoli si ottiene:

$$\sum_{k=1}^{\infty} k p^{k-1} = \frac{1}{(1-p)^2}, \quad \sum_{k=1}^{\infty} k^2 p^{k-1} = \frac{1+p}{(1-p)^3}. \quad (3.9)$$

Da queste formule segue che

$$\langle n \rangle = \sum_{k=1}^{\infty} k p (1-p)^{k-1} = p \sum_{k=1}^{\infty} k (1-p)^{k-1} = \frac{1}{p} \quad (3.10)$$

$$\langle n^2 \rangle = \sum_{k=1}^{\infty} k^2 p (1-p)^{k-1} = \frac{2-p}{p^2}. \quad (3.11)$$

La varianza è allora data da:

$$\sigma_n^2 = \langle n^2 \rangle - \langle n \rangle^2 = \frac{1-p}{p^2}. \quad (3.12)$$

Probabilità, media e deviazione standard nel caso considerato valgono:

$$\begin{aligned} g(n) &= 0.8 \cdot (0.2)^{n-1} \\ \mu_n &= 1.25 \\ \sigma_n &= \sqrt{(1-p)/p^2} = 0.31. \end{aligned}$$

Una media $\mu = 1.25$ sta a significare che in 12-13 prove mediamente si avranno 10 prove in cui il successo capita al primo tentativo.

3.3 La distribuzione di Poisson

Il calcolo della densità binomiale non è agevole se n è molto grande, a causa dei fattoriali che compaiono dalla formula (2.30). Se, oltre alla condizione $n \gg 1$, la probabilità dell'evento è piccola ($p \ll 1$), allora si avranno pochi successi ($x \ll n$) e si può scrivere:

$$\frac{n!}{(n-x)!} = n(n-1)(n-2)\dots(n-x+1) \stackrel{x \ll n}{\simeq} n^x.$$

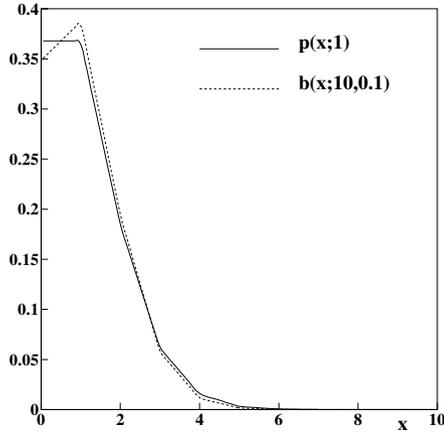


Fig. 3.1. Densità poissoniana (linea piena) e binomiale (linea tratteggiata) per $n = 10, p = 0.1, \mu = np = 1$. Per facilitare il confronto i valori discreti delle curve sono uniti da tratti di retta

Ponendo poi $y = (1 - p)^{n-x}$ e passando ai logaritmi, si ha:

$$\ln y = (n - x) \ln(1 - p) \xrightarrow{p \ll 1} -p(n - x) \xrightarrow{x \ll n} -np$$

$$y = (1 - p)^{n-x} = e^{\ln y} \rightarrow e^{-np} .$$

Il limite della densità binomiale per $n \gg 1$ e $p \ll 1$ diventa allora:

$$\lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0}} b(x; n, p) = \lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0}} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$= \frac{n^x}{x!} p^x e^{-np} = \frac{(np)^x}{x!} e^{-np} , \tag{3.13}$$

da cui, ricordando le (3.6) e ponendo $\mu = np$, otteniamo la densità discreta di Poisson:

$$p(x; \mu) = \frac{\mu^x}{x!} e^{-\mu} , \tag{3.14}$$

che rappresenta la probabilità di *ottenere un valore x quando la media dei valori è μ* . La poissoniana viene ritenuta in pratica una approssimazione accettabile della binomiale già a partire da $\mu > 10$ e $p < 0.1$, come mostrato in Fig. 3.1.

Esercizio 3.3

In un città di 50 000 abitanti si verificano in media 5 suicidi l'anno. Calcolare la probabilità che se ne verifichino 10.

Risposta. *Applicando la formula binomiale (2.30) abbiamo:*

$$b(10; 50000, \frac{5}{50000}) = \binom{50000}{10} \frac{5^{10}}{50000^{10}} \left(1 - \frac{5}{50000}\right)^{49990} = ??.$$

Dalla poissoniana otteniamo invece:

$$p(10; 5) = \frac{5^{10} e^{-5}}{10!} \simeq 0.018 = 1.8\%.$$

La densità di Poisson è normalizzata. Infatti:

$$\sum_{x=0}^{\infty} \frac{\mu^x}{x!} e^{-\mu} = e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^x}{x!} = e^{-\mu} e^{\mu} = 1. \quad (3.15)$$

La media e la varianza della distribuzione di Poisson possono essere trovate con lo stesso metodo usato per la distribuzione binomiale nelle (3.2-3.5). Lasciamo come esercizio il calcolo della media, che fornisce, come deve essere, il parametro μ che compare in chiaro nella (3.14). Per la varianza, utilizzando le (2.66, 3.3), si ottiene:

$$\begin{aligned} \text{Var}[X] &= \sum_{x=0}^{\infty} \left[x^2 \frac{\mu^x e^{-\mu}}{x!} \right] - \mu^2 \\ &= \mu \sum_{x'=0}^{\infty} \left[(x'+1) \frac{\mu^{x'} e^{-\mu}}{x'!} \right] - \mu^2 = \mu^2 + \mu - \mu^2 = \mu. \end{aligned}$$

Abbiamo pertanto le formule:

$$\mu \equiv np, \quad \sigma^2 = \mu, \quad \sigma = \sqrt{\mu}, \quad (3.16)$$

da cui risulta che la poissoniana ha media uguale alla varianza.

3.4 Densità di Gauss o normale

Un altro importante caso limite della densità binomiale è la densità normale o di Gauss. Questa approssimazione è possibile quando il numero di prove è molto grande: in questo caso solo i valori dello spettro in un intorno del massimo della densità sono importanti. Pensate, ad esempio, a 1000 lanci di una moneta: i valori dello spettro dell'evento *testa* vanno da 0 a 1000, ma i valori importanti saranno concentrati intorno al valore atteso 500, che presumibilmente è il valore di più alta probabilità. Non si otterranno praticamente mai valori come 5, 995 e così via. Consideriamo allora il caso in cui

$$n \gg 1, \quad 0 < p < 1, \quad x \gg 1 \quad (3.17)$$

ed approssimiamo il fattoriale con la formula di Stirling, valida per $x \gg 1$:

$$x! = \sqrt{2\pi x} x^x e^{-x} e^{-\varepsilon_x} \simeq \sqrt{2\pi x} x^x e^{-x}, \quad (3.18)$$

dove $|\varepsilon_x| \leq 1/(12x)$. Questa è già una approssimazione molto buona per $x \geq 10$, dato che $\exp[-1/120] = 0.992$, corrispondente ad un errore relativo dell'8 per mille, come potete facilmente verificare col vostro calcolatore tascabile. Notate pure che ora x è, come richiesto, una *variabile continua*, che approssima il fattoriale per valori interi.

Utilizzando la formula di Stirling, la densità binomiale (2.30) assume la forma:

$$\begin{aligned} b(x; n, p) &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &\simeq \frac{1}{\sqrt{2\pi}} \frac{\sqrt{nn^n}}{\sqrt{x(n-x)} x^x (n-x)^{n-x}} p^x (1-p)^{n-x}. \end{aligned} \quad (3.19)$$

Sviluppiamo ora la densità binomiale nell'intorno del suo punto di massimo. Passando ai logaritmi, otteniamo:

$$\ln b(x; n, p) = \ln n! - \ln x! - \ln(n-x)! + x \ln p + (n-x) \ln(1-p).$$

Il punto di massimo si ottiene annullando la derivata prima di $\ln b$, utilizzando la formula di Stirling e considerando quindi x come variabile continua:

$$\frac{d}{dx}(\ln b) = -\frac{d}{dx}(\ln x!) - \frac{d}{dx}[\ln(n-x)!] + \ln p - \ln(1-p) = 0. \quad (3.20)$$

Con le approssimazioni fatte, la derivata di $\ln x!$, vale:

$$\begin{aligned} \frac{d}{dx} \ln x! &\simeq \frac{d}{dx} \ln \left[(2\pi x)^{1/2} x^x e^{-x} \right] = \frac{d}{dx} \left[\frac{1}{2} \ln 2\pi + \frac{1}{2} \ln x + x \ln x - x \right] \\ &= \frac{1}{2x} + 1 + \ln x - 1 = \frac{1}{2x} + \ln x \xrightarrow{x \gg 1} \ln x. \end{aligned} \quad (3.21)$$

Poiché $d(\ln x!)/dx \simeq \ln x$, la (3.20) diventa:

$$\frac{d}{dx} \ln b \simeq -\ln x + \ln(n-x) + \ln p - \ln(1-p) = 0, \quad (3.22)$$

da cui:

$$\ln \frac{p(n-x)}{x(1-p)} = 0 \implies \frac{p(n-x)}{x(1-p)} = 1 \implies x = \mu = np.$$

Abbiamo ottenuto un primo risultato: per $\mu \gg 1$, ovvero $x \gg 1$, il valor medio diventa anche il punto di massimo.

Eseguiamo ora lo sviluppo in serie di Taylor al second'ordine nell'intorno di questo valore. Ponendo sempre $y(x) = \ln b$ otteniamo:

$$\begin{aligned} y(x) &= y(np) + \frac{1}{2} \left[\frac{d^2 y(x)}{dx^2} \right]_{x=np} (x - np)^2 + O\left(\frac{1}{n^2}\right) \\ &\simeq y(np) + \frac{1}{2} \left[-\frac{1}{x} - \frac{1}{n-x} \right]_{x=np} (x - np)^2, \end{aligned} \quad (3.23)$$

dove la derivata seconda è stata ottenuta derivando la (3.22) e il termine in $(1/n^2)$ derivando ancora la derivata seconda. Dopo facili calcoli si ottengono le equazioni:

$$y(x) \simeq y(np) - \frac{1}{2} \frac{(x - np)^2}{np(1-p)},$$

$$b(x; n, p) = e^{y(x)} \simeq e^{y(np)} \exp \left[-\frac{1}{2} \frac{(x - np)^2}{np(1-p)} \right]. \quad (3.24)$$

Dato che $e^{y(np)}$ non è altro che la densità binomiale calcolata nel punto di massimo $x = np$ con la formula di Stirling, dalla (3.19) otteniamo facilmente:

$$b(x = np; n, p) \simeq \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{np(1-p)}}, \quad (3.25)$$

da cui:

$$b(x; n, p) \simeq g(x; n, p) = \frac{1}{\sqrt{2\pi} \sqrt{np(1-p)}} \exp \left[-\frac{1}{2} \frac{(x - np)^2}{np(1-p)} \right]. \quad (3.26)$$

Questa è una forma approssimata della binomiale che vale per valori di x nell'intorno di $\mu = np$, a patto di trascurare i termini di ordine superiore a $1/n$ nello sviluppo di Taylor (3.23). La funzione $g(x; n, p)$, che approssima la binomiale per valori interi di x , può essere considerata come funzione continua di x . Infatti, il fattore

$$\Delta t \equiv \frac{1}{\sqrt{np(1-p)}},$$

che compare nella (3.26), è una quantità indipendente da x , che $\rightarrow 0$ per $n \rightarrow \infty$ e che assume il significato di differenziale della variabile t data da:

$$t = \frac{x - np}{\sqrt{np(1-p)}}.$$

È allora possibile, passando al limite di Riemann, calcolare la somma delle probabilità di un insieme di valori di X , per valori grandi di n , come:

$$\begin{aligned} P\{x_1 \leq X \leq x_2\} &= \lim_{n \rightarrow \infty} \sum_{x_1}^{x_2} b(x; n, p) = \lim_{\Delta t \rightarrow 0} \sum_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \frac{(x - np)^2}{np(1-p)} \right] \Delta t \\ &= \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{1}{2} t^2} dt. \end{aligned} \quad (3.27)$$

Questo fondamentale risultato è detto teorema di DeMoivre-Laplace.

Introducendo μ e σ definiti nelle (3.6) otteniamo infine:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}t^2\right] dt = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] dx = 1 ,$$

dove si è fatto uso dell'integrale notevole (vedere esercizio 3.4) $\int \exp(-z^2) dz = \sqrt{\pi}$.

La densità gaussiana o normale:

$$g(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] , \quad (3.28)$$

è quindi normalizzata e permette di calcolare, quando è integrata, la probabilità che un valore x cada in un intervallo vicino alla media μ quando la varianza è σ^2 . Questa è di gran lunga la più importante funzione di densità del calcolo delle probabilità e andrebbe pertanto ricordata a memoria.

Esercizio 3.4

Dimostrare che

$$\int_{-\infty}^{+\infty} \exp(-x^2) dx = \sqrt{\pi} . \quad (3.29)$$

Risposta. *Posto*

$$I = \int_{-\infty}^{+\infty} e^{-x^2} dx = \int_{-\infty}^{+\infty} e^{-y^2} dy ,$$

si ha

$$I^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(x^2+y^2)} dx dy ,$$

e, in coordinate polari:

$$I^2 = \int_0^{+\infty} \int_0^{2\pi} e^{-\rho^2} \rho d\rho d\theta = 2\pi \int_0^{+\infty} e^{-\rho^2} \rho d\rho = -\pi \left[e^{-\rho^2} \right]_0^{\infty} = \pi ,$$

da cui $I = \sqrt{\pi}$. Analogamente, si può dimostrare che

$$\int_{-\infty}^{+\infty} x^2 \exp(-x^2) dx = \sqrt{\frac{\pi}{2}} . \quad (3.30)$$

La media e la varianza della distribuzione binomiale sono ovviamente date, per costruzione, dai parametri μ e σ che compaiono in chiaro nella (3.28). Tuttavia, utilizzando gli integrali (3.29, 3.30), si può anche dimostrare direttamente che:

$$\langle X \rangle = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] dx = \mu \quad (3.31)$$

$$\text{Var}[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (x-\mu)^2 \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] dx = \sigma^2 . \quad (3.32)$$

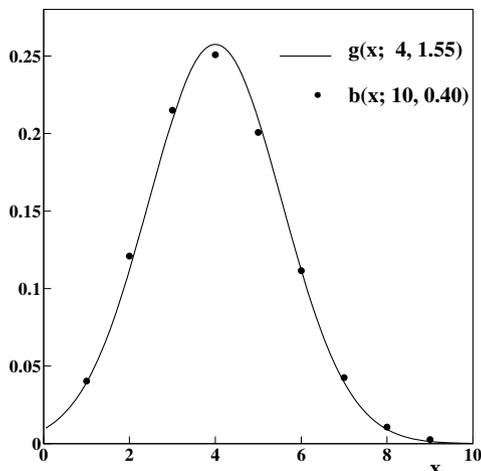


Fig. 3.2. Densità gaussiana (linea piena) e binomiale (punti) per $n = 10$, $p = 0.4$

In statistica dovremo utilizzare il momento di ordine 4 della gaussiana; tramite integrali del tipo (3.29, 3.30), si può mostrare che:

$$\Delta_4 = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (x - \mu)^4 \exp \left[-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} \right] dx = 3\sigma^4. \quad (3.33)$$

La densità di Gauss approssima la binomiale per $\mu \gg 1$. Grazie alla rapida convergenza dell'approssimazione di Stirling (3.18), in molti casi si commette un errore minore di qualche per cento già per $\mu \geq 10$. Come regola pratica, la funzione di Gauss può essere sostituita alla binomiale quando sono valide le condizioni

$$\mu = np \geq 10, \quad n(1-p) \geq 10 \quad (\text{in pratica!}), \quad (3.34)$$

che giustificano in modo ragionevole l'approssimazione dello sviluppo (3.23). Se p e $1-p$ non sono molto vicine agli estremi dell'intervallo $[0, 1]$, basta la condizione $\mu \geq 10$. Se invece le probabilità sono molto sbilanciate occorre verificare entrambe le condizioni della (3.34). Si veda la Fig. 3.2, dove viene riportata l'approssimazione gaussiana di una densità binomiale $b(x; 10, 0.4)$. In questo caso media e varianza sono date da $\mu = np = 10 \cdot 0.4 = 4$ e $\sigma^2 = np(1-p) = 2.4$. Pur essendo $\mu = np = 4$, $n(1-p) = 6 < 10$, l'approssimazione è già più che accettabile.

Esercizio 3.5

Utilizzando la distribuzione di Gauss, calcolare le probabilità di ottenere un dato numero di teste nell'esperimento delle 10 monete di Tab. 2.2.

Risposta. *Nell'esperimento delle 10 monete, di cui ci stiamo servendo ripetutamente per affinare la vostra sensibilità statistica, i valori di media, varianza e deviazione standard sono dati da:*

$$\begin{aligned}\mu &= np = 10 \cdot 0.5 = 5 , \\ \sigma^2 &= np(1-p) = 10 \cdot 0.5(1-0.5) = 2.50 , \\ \sigma &= \sqrt{2.5} = 1.58 .\end{aligned}$$

Questi valori, inseriti nella (3.28), permettono il calcolo delle probabilità gaussiane richieste:

$$p(x) = 0.252 \exp\left(-\frac{1}{2} \frac{(x-5)^2}{2.5}\right) , \quad (x = 0, 1, \dots, 10) .$$

Se ricopiamo la Tab. 2.2 e riportiamo in una colonna le probabilità così calcolate, otteniamo il risultato:

<i>Spettro (n. teste)</i>	<i>N. di prove</i>	<i>Frequenza</i>	<i>Probabilità binom.</i>	<i>Probabilità gauss.</i>
0	0	0.00	0.001	0.002
1	0	0.00	0.010	0.010
2	5	0.05	0.044	0.042
3	13	0.13	0.117	0.113
4	12	0.12	0.205	0.206
5	25	0.25	0.246	0.252
6	24	0.24	0.205	0.206
7	14	0.14	0.117	0.113
8	6	0.06	0.044	0.042
9	1	0.01	0.010	0.010
10	0	0.00	0.001	0.002

da cui si vede che, pur essendo $np = n(1-p) = 5 < 10$, la densità gaussiana fornisce risultati già in buon accordo con la densità binomiale. Per questo motivo spesso la condizione (3.34) viene ulteriormente allargata ponendo

$$\mu \geq 5 , \quad n(1-p) \geq 5 .$$

3.5 Calcolo di distribuzioni in SCILAB

In ambito SCILAB è possibile calcolo di tutte le principali distribuzioni di probabilità. Esistono infatti delle funzioni di sistema `cdfxxx` per il calcolo delle cumulative; i caratteri `xxx` sono riservati al tipo di distribuzione richiesta: ad esempio, se `xxx` vale `bin`, `nor`, `poi`, si ottiene rispettivamente il calcolo delle cumulative delle densità binomiale, normale o poissoniana.

Per tutte queste funzioni, sono possibili i seguenti tipi di chiamata (per fissare le idee, consideriamo il caso della distribuzione binomiale con parametri n, p):

$$\begin{aligned} [P, 1-P] &= \text{cdfbin}(PQ, X, n, p, 1-p) \\ [X] &= \text{cdfbin}(S, n, P, 1-P, p, 1-p) \\ [n] &= \text{cdfbin}(Xn, p, 1-p, P, 1-P, X) \\ [p, 1-p] &= \text{cdfbin}(PrOmp, P, 1-P, X, n) \end{aligned}$$

In modalità “PQ” si calcola in P il valore cumulativo da zero ad X; la modalità “S” calcola il valore X corrispondente al valore cumulativo P; con “Xn” viene fornito il numero n di tentativi relativo ad una binomiale con probabilità elementare p quando il valore cumulativo tra zero e X è pari a P; infine, con la chiamata “PrOmp” viene trovata la probabilità p a partire da n e dal valore cumulativo P tra zero e X.

Questa modalità di chiamata sono comuni a tutte le distribuzioni e possono subire alcune variazioni a seconda dei parametri matematici in gioco, che variano da caso a caso. È bene quindi, prima di utilizzare queste routine, consultare il manuale in linea di SCILAB.

3.6 Legge 3-sigma e gaussiana standard

La densità gaussiana, oltre ad essere una ottima approssimazione delle distribuzioni binomiale e poissoniana è anche, come vedremo nel prossimo paragrafo, la densità limite della combinazione lineare di più variabili aleatorie. Per queste ragioni essa rappresenta di certo la distribuzione più importante del calcolo delle probabilità.

Vogliamo qui discutere una proprietà matematica molto importante di questa funzione, detta *legge 3-sigma* (o 3σ):

$$\begin{aligned} P\{|X - \mu| \leq +k\sigma\} &\equiv \frac{1}{\sqrt{2\pi}\sigma} \int_{\mu-k\sigma}^{\mu+k\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \\ &= \begin{cases} 0.683 & \text{per } k = 1 \\ 0.954 & \text{per } k = 2 \\ 0.997 & \text{per } k = 3 \end{cases} \end{aligned} \quad (3.35)$$

che, espressa a parole, suona così : *se X è una variabile gaussiana di media $\langle X \rangle = \mu$ e deviazione standard $\sigma[X] = \sigma$, la probabilità di ottenere un valore x compreso in un intervallo centrato sulla media μ e di ampiezza $\pm\sigma$ è di circa il 68%, mentre se l'intervallo è ampio $\pm 2\sigma$ essa è di circa il 95%; inoltre il valore x può cadere fuori da un intervallo di ampiezza $\pm 3\sigma$ solo con una probabilità di 3 casi su mille*

In molti casi pratici si assume che i valori dello spettro siano tutti contenuti in un intervallo entrato su μ e largo $\pm 3\sigma$ (da cui il nome della legge (3.35)). Le probabilità definite dalla (3.35) sono dette anche *livelli o valori di probabilità gaussiani*.

Il calcolo dell'integrale (3.35) è difficile in pratica, perché la primitiva $E(x)$ della gaussiana:

$$E(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_0^x \exp \left[-\frac{1}{2} \left(\frac{t - \mu}{\sigma} \right)^2 \right] dt \quad (3.36)$$

non è nota analiticamente (strano, ma vero!). L'integrale viene di solito valutato con metodi numerici, sviluppando in serie l'esponenziale e calcolando numericamente il limite della serie integrata. Si ottengono in questo modo i livelli di probabilità della (3.35). Sembra a questo punto necessario, per ogni gaussiana di media e deviazione standard assegnate, procedere per via numerica al calcolo delle probabilità cumulative negli intervalli di interesse. Questa complicazione può essere però evitata ricorrendo ad una gaussiana universale o gaussiana standard, che si ottiene definendo una variabile fondamentale in statistica, la variabile standard

$$T = \frac{X - \mu}{\sigma} \quad \text{che assume i valori} \quad t = \frac{x - \mu}{\sigma}, \quad (3.37)$$

la quale misura *lo scarto di un valore x dalla sua media in unità di deviazione standard*. Applicando le (2.62, 2.64, 2.65) è facile vedere che la variabile standard ha media nulla e varianza unitaria.

Se ora operiamo il cambio di variabile (3.37) nell'integrale (3.36) otteniamo:

$$E(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp \left(-\frac{t^2}{2} \right) dt. \quad (3.38)$$

Questa primitiva, come è facile verificare, è legata alla nota funzione degli errori (*error function*) $\text{Erf}(x)$ dalla relazione:

$$\text{Erf}(x) \equiv \frac{1}{\sqrt{\pi}} \int_{-x}^{+x} \exp(-t^2) dt = 2 \frac{1}{\sqrt{2\pi}} \int_0^{\sqrt{2}x} \exp \left(-\frac{t^2}{2} \right) dt \equiv 2 E(\sqrt{2}x). \quad (3.39)$$

Essa può essere calcolata una volta per tutte, essendo indipendente da μ e σ .

Poiché la gaussiana è una curva simmetrica, vale anche la condizione

$$E(x) = -E(-x). \quad (3.40)$$

Considerando ora la trasformazione $t = \sqrt{2}z$ ed integrando la serie esponenziale, si ottiene:

$$\begin{aligned} E(x) &= \frac{1}{\sqrt{\pi}} \int_0^{x/\sqrt{2}} (1 - z^2 + \frac{z^4}{2!} - \frac{z^6}{3!} + \dots) dz \\ &= \frac{1}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k (x/\sqrt{2})^{2k+1}}{k! (2k+1)}. \end{aligned} \quad (3.41)$$

La somma di questa serie si trova tabulata in tutti i testi di statistica (nel nostro la trovate in Appendice D nella Tab. D.1 a pagina 521), ed è presente

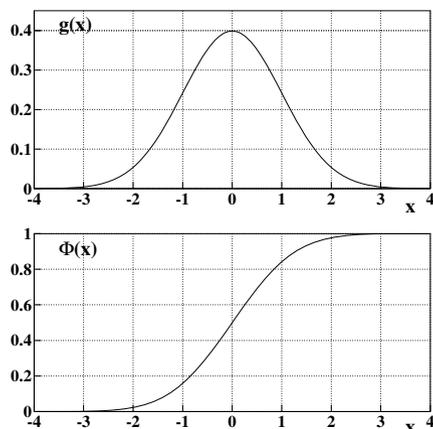


Fig. 3.3. Gaussiana standard $g(x; 0, 1)$ (3.42) e corrispondente cumulativa $\Phi(x)$ (3.43)

come routine di libreria in molti prodotti software di uso scientifico. In SCI-LAB è disponibile la funzione degli errori `erf` con 7 cifre decimali. Invertendo la (3.39) l'integrale $E(x)$, può essere calcolato con l'istruzione:

$$0.5 * \text{erf}(x / \text{sqrt}(2))$$

che può essere usata in alternativa alla Tab. D.1, ottenendo anche un risultato più accurato con 7 cifre significative.

La primitiva (3.38) corrisponde ad una densità gaussiana standard

$$g(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (3.42)$$

di media zero e deviazione standard unitaria. Una distribuzione avente densità gaussiana (3.28) viene chiamata *normale* ed indicata spesso con $N(\mu, \sigma^2)$; la variabile aleatoria corrispondente a volte viene indicata con $X \sim N(\mu, \sigma^2)$, o $T \sim N(0, 1)$, se è standard. Il simbolo \sim significa “distribuito come”.

La funzione cumulativa della gaussiana standard, rappresentata in Fig. 3.3, viene invece indicata solitamente con $\Phi(x)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx . \quad (3.43)$$

Essa è legata alla funzione (3.38) dalla relazione:

$$\Phi(x) = 0.5 + E(x) , \quad (3.44)$$

che vale anche per $x < 0$ grazie alla (3.40).

In sostanza, il calcolo delle probabilità integrate, con qualunque gaussiana, può sempre essere ricondotto al caso della gaussiana standard, di media nulla e varianza unitaria, per il quale esiste una tabella universale.

Esercizio 3.6

Trovare la probabilità $P\{\mu - \sigma \leq X \leq \mu + \sigma\}$ con quattro cifre decimali.

Risposta. La variabile standard (3.37), nei due estremi dell'intervallo da considerare, vale in questo caso ± 1 ; dalla Tab. D.1, riportata in Appendice D, per $t = 1.00$ (1.0 si legge nella riga e l'ultimo zero nella colonna in alto) si ha $P = 0.3413$. Poiché la tabella dà l'integrale tra zero e t , dobbiamo raddoppiare questo valore. Otteniamo pertanto $P\{\mu - \sigma \leq X \leq \mu + \sigma\} = 2 \times 0.3413 = 0.6826$, in accordo con la (3.35). Si noti che la tabella fornisce valori solo per $0 \leq t \leq 0.5$, dato che la gaussiana standard è simmetrica rispetto all'origine.

Esercizio 3.7

Qual è la probabilità di estrarre, da una gaussiana $N(5, 9)$, cioè di media 5 e deviazione standard 3, valori di X più grandi di 12 oppure $-2 \leq X \leq 12$?

Risposta. La variabile standard (3.37) vale in questo caso

$$t = \frac{12 - 5}{3} = 2.33, \quad t = \frac{-2 - 5}{3} = -2.33.$$

Dalla cumulativa $\Phi(x)$ di Fig. 3.3 si vede che la probabilità di ottenere valori $x > 2.33$ è molto piccola. Un valore preciso lo si può ottenere dalla Tab. D.1, dove in corrispondenza di questo valore si legge il valore 0.4901, che rappresenta la probabilità di ottenere valori tra 0 e 2.33. Le probabilità cercate sono date allora da:

$$\begin{aligned} P\{X > 12\} &= 0.5000 - 0.4901 = 0.0099 \simeq 1\%, \\ P\{-2 \leq X \leq 12\} &= 1 - 2 \times 0.0099 = 0.9802 \simeq 98\%. \end{aligned}$$

Esercizio 3.8

Quali sono gli estremi corrispondenti ad un intervallo di probabilità gaussiano del 95%?

Risposta. Si chiede di trovare un numero reale t tale che $P\{\mu - t\sigma \leq X \leq \mu + t\sigma\} = 0.95$. Dalla Tab. D.1 si legge, in corrispondenza di una probabilità pari a $0.95/2 = 0.475$, il valore $t = 1.96$. La risposta è pertanto: la probabilità del 95% si ottiene integrando la gaussiana in un intervallo ampio 1.96σ e centrato sulla media. Infatti

$$t = 1.96 = \left| \frac{x - \mu}{\sigma} \right| \quad \text{implica} \quad x = \mu \pm 1.96\sigma,$$

e quindi

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{\mu - 1.96\sigma}^{\mu + 1.96\sigma} e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2} dx = 0.95.$$

3.7 Il teorema Limite Centrale: universalità della gaussiana

La densità gaussiana appare fin qui come limite per medie elevate della densità binomiale. Tuttavia, un teorema fondamentale, proposto da Gauss e da Laplace nel 1809-1812 e dimostrato in modo generale agli inizi del '900, assegna ad essa un ruolo cruciale nel calcolo delle probabilità, sia per le variabili continue sia discrete. Questo è il:

Teorema 3.1 (Limite Centrale o Centrale Limite). *Sia Y una variabile aleatoria combinazione lineare di N variabili aleatorie X_i (si veda la(2.9)) con la relativa discussione)*

$$Y_N = \sum_{i=1}^N a_i X_i \quad (3.45)$$

dove a_i sono coefficienti costanti. Se

- a) le variabili X_i sono tra loro indipendenti (si veda la (2.10));
- b) le variabili X_i hanno varianza finita;
- c) le varianze (o deviazioni standard) sono tutte dello stesso ordine di grandezza:

$$\frac{\sigma[X_i]}{\sigma[X_j]} = O(1) \quad \text{per ogni } i, j ; \quad (3.46)$$

allora, per $N \rightarrow \infty$, la variabile aleatoria Y_N converge in legge, secondo la (2.74), verso la distribuzione gaussiana.

Si ha pertanto, in termini di funzioni di ripartizione:

$$P \left\{ \frac{Y_N - \langle Y_N \rangle}{\sigma[Y_N]} \leq x \right\} \xrightarrow{N \gg 1} \Phi(x) ,$$

dove $\Phi(x)$ è la funzione di ripartizione della gaussiana standard (3.43).

Dimostrazione. La dimostrazione del teorema, per variabili aventi tutte la stessa distribuzione, fa uso delle funzioni generatrici ed è riportata in Appendice B. Anche la formula di DeMoivre-Laplace (3.27) va vista come un caso particolare del teorema nel caso della somma di variabili di Bernoulli.

Si può anche dimostrare, più in generale, che il teorema vale per somme di variabili, sia discrete sia continue, aventi ognuna una distribuzione diversa [Cra51]. Tuttavia, è essenziale che le variabili X_i siano tra loro indipendenti (condizione a)), e che esse, se prese singolarmente, abbiano debole influenza sul risultato finale (condizioni b) e c)). Va però subito notato un fatto piuttosto strabiliante e di estrema importanza, che si verifica in pratica: alla condizione $N \rightarrow \infty$ si può sostituire con buona approssimazione la condizione $N \geq 10$. □

In sostanza il teorema afferma che una variabile aleatoria tende ad avere densità di Gauss se essa è la sovrapposizione lineare di più variabili aleatorie indipendenti, numerose, e, se prese singolarmente, di debole influenza sul risultato finale.

Parlando un po' liberamente, possiamo dire che il teorema assegna alla gaussiana il ruolo di una funzione universale a cui tendono le variabili di sistemi ad "alto equilibrio statistico".

Ancora più interessante è il fatto che la pratica e le simulazioni al calcolatore mostrano che N non deve essere poi tanto grande. Come già evidenziato nella dimostrazione del teorema, è lecito utilizzare la condizione $N \geq 10$.

Spesso si dice: "gli sperimentatori usano la gaussiana perché pensano che i matematici abbiano dimostrato che essa è la curva universale; i matematici la usano perché pensano che gli sperimentatori ne abbiano in pratica dimostrato l'universalità...". Questa affermazione, un po' semplicistica, andrebbe sostituita con la seguente: spesso le variabili aleatorie seguono la distribuzione gaussiana perché sono verificate abbastanza bene le condizioni del teorema Limite Centrale. Esistono tuttavia numerose ed importanti eccezioni a questa regola generale, la quale va considerata spesso come un indispensabile punto di riferimento, ma non va applicata alla cieca.

Ecco alcuni esempi di variabili aleatorie che, in accordo col teorema, si distribuiscono in modo gaussiano:

- l'altezza o il peso di una popolazione di individui etnicamente omogenei;
- il peso dei fagioli contenuti in una scatola (provare per credere...);
- i valori dei quozienti di intelligenza di un gruppo di persone;
- la media di un campione con numero di eventi maggiore di una decina: in questo caso le condizioni a)-c) del teorema sono senz'altro soddisfatte, perché le variabili sono indipendenti e le loro varianze sono tutte uguali;
- le componenti della velocità delle molecole di un gas perfetto.

Le grandezze seguenti *non* sono invece di tipo gaussiano:

- l'energia delle molecole di un gas perfetto: essa è proporzionale al modulo quadrato del vettore velocità (che ha componenti gaussiane) e quindi viene meno la condizione di linearità (3.45). Dimosteremo tra breve che il modulo quadro di un vettore di componenti gaussiane indipendenti segue una distribuzione particolare, detta di χ^2 (chi-quadrato o chi-quadro). In tre dimensioni, questa distribuzione è la famosa distribuzione di Maxwell, ben nota ai fisici;
- come vedremo tra breve, i tempi di arrivo di eventi poissoniani non si distribuiscono secondo una gaussiana.

3.8 Processi stocastici poissoniani

Nei processi stocastici le variabili aleatorie dipendono da un parametro continuo o discreto. In questo paragrafo tratteremo un caso molto frequente, quello in cui il parametro continuo che si considera è il tempo.

Per questa classe di fenomeni la distribuzione di Poisson riveste un ruolo fondamentale: essa non va pertanto considerata solo come una semplice ap-

prossimazione matematica della distribuzione binomiale, ma, analogamente al caso della gaussiana, come una legge universale della natura.

Consideriamo un processo stocastico in cui una sorgente genera eventi:

- discreti, in modo tale che in un intervallo infinitesimo di tempo dt si generi *al più* un evento;
- con probabilità per unità di tempo λ costante e uguale per tutti gli eventi generati. Questa quantità, di dimensione s^{-1} , rappresenta la frequenza di emissione per unità di tempo (per esempio, il numero di eventi al secondo). Questa proprietà implica che il numero medio di eventi contati in un intervallo dipenda solo dalla larghezza dell'intervallo, e non dalla sua posizione lungo l'asse dei tempi;
- indipendenti tra loro. In altre parole, i conteggi degli eventi in intervalli di tempo disgiunti sono variabili aleatorie indipendenti.

Un fenomeno stocastico che obbedisce alle condizioni a)-c) si chiama *processo stazionario di Poisson*.

Quale sarà la legge statistica del numero di eventi generati entro un intervallo di tempo misurabile Δt ? La risposta è facile se si definiscono le quantità seguenti:

- il numero di tentativi $N = \Delta t/dt$. Se Δt è una quantità macroscopica misurabile e dt è un differenziale, si ha sempre $N \gg 1$. Notate che si considera come durata di un tentativo utile alla generazione di un evento la quantità differenziale o comunque molto piccola dt ; è questa l'ipotesi cruciale di tutto il ragionamento che stiamo sviluppando;
- la probabilità $p = \lambda dt$ di osservare un evento in un tentativo di durata dt . Si ha sempre $p \ll 1$ se il processo di emissione è discreto e dt è un differenziale;
- il numero medio $\mu = \lambda \Delta t$ di eventi generati entro Δt .

Il processo di emissione degli eventi può allora essere visto come la probabilità binomiale di avere $\{X = n\}$ successi in $N = (\Delta t/dt) \gg 1$ tentativi quando la probabilità elementare di un successo è $p = \lambda dt \ll 1$. Tenendo presente i risultati del par. 3.3, è immediato dedurre che il conteggio $X(\Delta t)$ in un intervallo Δt in un processo di emissione discreto non può che essere una variabile aleatoria che segue la legge di Poisson, con $\mu = \lambda \Delta t$. Dalla (3.14) otteniamo:

$$P\{X(\Delta t) = n\} = p_n(\Delta t) = \frac{(\lambda \Delta t)^n}{n!} e^{-\lambda \Delta t}. \quad (3.47)$$

È facile mostrare che, se vale la (3.47), i conteggi in intervalli di tempo disgiunti sono variabili aleatorie indipendenti (si veda il problema 3.13).

Alla legge di Poisson obbedisce una classe veramente vasta di fenomeni: il numero di pesci pescati in un'ora in condizioni ambientali costanti, il numero di auto che si fermano a un semaforo a parità di giorno ed ora nel corso, poniamo, di un mese (se non vi sono eventi eccezionali), il numero di fotoni emessi da un insieme di atomi eccitati, il numero di particelle nucleari emesse

da una sorgente radioattiva, il numero di stelle cadenti osservate in 10 minuti in una sera di agosto, il numero di incidenti stradali in un anno se non vengono presi nuovi provvedimenti per la sicurezza, ... insomma, tutti quei casi in cui una sorgente *stabile* (λ costante) emette eventi *discreti* ($\lambda dt \ll 1$) ed *indipendenti*.

Fin qui abbiamo considerato le proprietà della sorgente degli eventi; qual è il ruolo dell'osservatore o dell'apparato di conteggio? L'osservazione del fenomeno non altera la statistica poissoniana se l'apparato di rivelazione o conteggio (l'occhio, uno strumento o un apparato comunque complicato) *ha un tempo morto e un tempo di risoluzione piccoli rispetto al tempo medio di arrivo degli eventi*. Per tempo morto si intende il tempo durante il quale lo strumento resta inattivo dopo il conteggio di un evento (in un contatore Geiger, ad esempio, questo tempo è dell'ordine del millisecondo); per tempo di risoluzione si intende il tempo al di sotto del quale l'apparato non registra più tutti i conteggi (per l'occhio questo tempo è di poco inferiore al secondo, per gli strumenti meccanici si va dalla frazione al millesimo di secondo, per gli strumenti elettronici si arriva al miliardesimo di secondo o anche meno). In sostanza, un evento che arriva entro il tempo morto non viene registrato, due o più eventi che arrivano entro il tempo di risoluzione vengono registrati come un solo evento. È chiaro allora che la statistica poissoniana originale non viene alterata se la probabilità di arrivo di un evento entro il tempo morto e la probabilità di arrivo di più di un evento entro il tempo di risoluzione sono entrambe trascurabili.

Il numero di eventi emessi e/o contati in un tempo finito Δt non è l'unica variabile aleatoria importante di un processo stocastico: fondamentale è anche la variabile aleatoria *tempo di arrivo* T dell'evento successivo rispetto al precedente. Dimostriamo ora un'ulteriore e importante legge statistica fondamentale della natura: la legge esponenziale negativa dei tempi di arrivo.

Azzeriamo l'orologio ad un tempo arbitrario $t = 0$ oppure sull'ultimo evento registrato: il tempo T che scorre prima che arrivi un altro evento è una variabile aleatoria che è determinata dalla probabilità composta di non registrare nulla fino a un tempo t e che venga registrato un evento in $(t, t + dt]$. Ponendo nella (3.47) $n = 0$, abbiamo $p_0(t) = e^{-\lambda t}$, mentre, in base alla definizione di λ , la probabilità di arrivo in $(t, t + dt]$ è data semplicemente da λdt . La variabile T ha dunque una densità di probabilità di tipo esponenziale negativo:

$$e(t) dt = p_0(t) \lambda dt = \lambda e^{-\lambda t} dt, \quad t \geq 0. \quad (3.48)$$

È facile vedere che essa è normalizzata e che media e varianza sono date da:

$$\mu = \int t e(t) dt = \frac{1}{\lambda}, \quad \sigma^2 = \int \left(t - \frac{1}{\lambda}\right)^2 e(t) dt = \frac{1}{\lambda^2}. \quad (3.49)$$

L'equivalente nel discreto della densità esponenziale è la densità geometrica (3.7). È facile infatti vedere che, per $p \ll 1$, il logaritmo di questa equazione

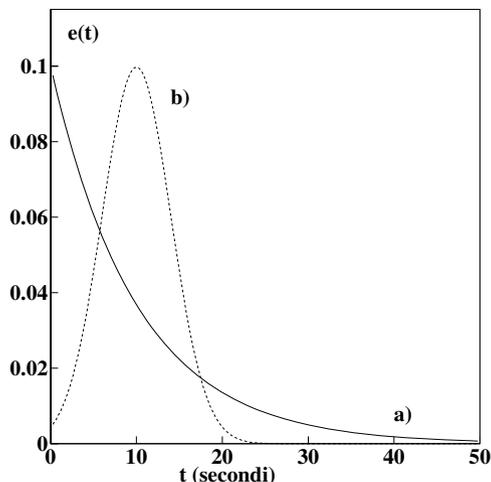


Fig. 3.4. La densità dei tempi di arrivo è di tipo esponenziale negativo come mostrato nella curva a) della figura, che si riferisce ad un flusso medio di 0.1 eventi/s ($\lambda = 0.1 \text{ s}^{-1}$, 1 evento ogni 10 secondi). La curva b) è una gaussiana di media 10 secondi e deviazione standard arbitraria di 4 secondi, che costituisce un esempio della rappresentazione intuitiva (ma errata!) che spesso si ha dei fenomeni stocastici

diventa:

$$\ln g(n) = \ln[p(1-p)^{n-1}] = \ln p + (n-1) \ln(1-p) \rightarrow \ln p - (n-1)p ,$$

e quindi:

$$g(n) \xrightarrow{p \ll 1} p e^{-(n-1)p} . \quad (3.50)$$

Il risultato ottenuto si presta ad alcune considerazioni interessanti. La rappresentazione intuitiva che spesso si ha dei tempi di arrivo di un flusso di eventi (10 eventi al secondo, poniamo) è di tipo gaussiano: gli eventi dovrebbero arrivare ad intervalli di circa 1/10 di secondo, con piccole fluttuazioni simmetriche intorno a questo valore. Questa rappresentazione intuitiva è *completamente sbagliata*: l'arrivo degli eventi è regolato dalla legge esponenziale, secondo la quale *eventi ravvicinati nel tempo sono molto più probabili di eventi separati da lunghi intervalli di attesa* (si veda la Fig. 3.4). Questo effetto è ben visibile con i contatori che spesso si usano in laboratorio: gli eventi sembrano arrivare “a grappoli” separati da lunghe attese, dando l'impressione ai non esperti di malfunzionamenti strumentali. In realtà, questa apparente correlazione temporale è proprio dovuta alla assoluta mancanza di correlazione tra gli eventi!

Eventi rari come gli incidenti o le calamità naturali spesso avvengono a breve distanza di tempo, generando la (falsa) credenza di misteriose correlazioni tra le sciagure. In realtà si tratta quasi sempre di effetti dovuti al caso, che si

riflettono in alcuni detti popolari quali “non c’è il due senza il tre . . .”, oppure “le disgrazie non vengono mai sole . . .”. Forse questi sono da annoverare tra i pochi proverbi che hanno un qualche fondamento sperimentale . . .

Dato che la densità esponenziale negativa non ha la classica forma “a campana”, la media in questo caso non dice molto su dove i dati si localizzano. È allora più utile rifarsi alla funzione cumulativa della (3.48), che si ottiene facilmente dalla (2.34):

$$P\{0 \leq T \leq t\} \equiv F(t) = \int_0^t \lambda e^{-\lambda\tau} d\tau = 1 - e^{-\lambda t} . \quad (3.51)$$

Questa funzione fornisce la probabilità di arrivo di *almeno un evento* in $[0, t]$. La probabilità di non osservare eventi fino a t è ovviamente data da:

$$P\{T > t\} = 1 - F(t) = 1 - P\{0 \leq T \leq t\} = e^{-\lambda t} = p_0(t) , \quad (3.52)$$

che coincide con la densità di Poisson (3.47) quando $n = 0$.

Poiché dalla (3.49) sappiamo che la media dei tempi vale $1/\lambda$, da queste formule possiamo calcolare la percentuale di arrivi prima e dopo la media:

$$P\{0 \leq T \leq 1/\lambda\} = 1 - e^{-1} = 0.63 , \quad (3.53)$$

$$P\{1/\lambda \leq T\} = e^{-1} = 0.37 . \quad (3.54)$$

Esercizio 3.9

Tra le 23 e le 23,30 lungo un certo tratto di strada transitano in media 40 veicoli. Qual è la probabilità di osservare, tra un veicolo ed il successivo, un intervallo di tempo minore di 10 secondi ed uno di più lungo di 5 minuti?

Risposta. Il tempo medio tra due veicoli successivi è dato da

$$\frac{60 \times 30}{40} = \frac{1800}{40} = 45 \text{ s} ,$$

cui corrisponde una probabilità o frequenza media di arrivo pari a

$$\lambda = \frac{1}{45} = 0.022 \text{ s}^{-1} .$$

Applicando la (3.51) con $t = 10 \text{ s}$ si ottiene:

$$P\{0 \leq T \leq 10 \text{ s}\} = 1 - e^{-0.022 \cdot 10} = 0.199 \simeq 20\% .$$

La probabilità di osservare intervalli più lunghi di $t = 5 \text{ min} = 300 \text{ s}$ è data invece da:

$$\begin{aligned} P\{T > 300 \text{ s}\} &= 1 - P\{0 \leq T \leq t\} = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t} \\ &= e^{-0.022 \cdot 300} = 1.27 \cdot 10^{-3} \simeq 0.13\% . \end{aligned}$$

Vogliamo ora mostrare che l'esponenziale negativa è l'unica forma funzionale che assicura l'indipendenza temporale degli eventi. Siano $A = \{T > t + \Delta t\}$ e $B = \{T > t\}$ gli eventi: “non si sono registrati arrivi fino a $t + \Delta t$ ” e “fino a t ” rispettivamente. Ovviamente $\{T > t + \Delta t\} \subset \{T > t\}$, perché se si verifica A si è verificato anche B , per cui $P(A \cap B) = P(A)$ (l'intersezione degli eventi *non* corrisponde a quella degli intervalli di tempo!). La probabilità condizionata (1.19) vale in questo caso:

$$\begin{aligned} P\{T > t + \Delta t | T > t\} &= \frac{P\{\{T > t + \Delta t\} \cap \{T > t\}\}}{P\{T > t\}} = \frac{P\{T > t + \Delta t\}}{P\{T > t\}} \\ &= \frac{e^{-\lambda(t+\Delta t)}}{e^{-\lambda t}} = e^{-\lambda \Delta t} = P\{T > \Delta t\}. \end{aligned} \quad (3.55)$$

Come si vede, l'informazione sull'assenza di eventi fino a t non ha alcuna influenza sulla legge di probabilità dei tempi successivi. La legge esponenziale rimane invariata, qualunque sia l'istante t in cui si decide di azzerare l'orologio; essa descrive pertanto i sistemi detti *senza memoria*. Basta però combinare tra loro più sistemi senza memoria per avere sistemi con memoria le cui densità di probabilità temporali non soddisfano la (3.55); si vedano, nei prossimi due capitoli, l'esercizio 4.1 ed il problema 5.8.

L'indipendenza tra i conteggi in intervalli di tempo separati (condizione c) di pagina 88) implica anche l'indipendenza degli eventi “c'è un arrivo al tempo t ” e “non ci sono arrivi in $(t, t + \Delta t]$ ”. Vale pertanto la relazione:

$$P\{T > \Delta t | T = t\} = P\{T > \Delta t\} = e^{-\lambda \Delta t},$$

da cui si vede che i tempi tra gli arrivi (intertempi) *sono variabili aleatorie indipendenti di legge esponenziale negativa*.

Per descrivere sistemi biologici o dispositivi dotati di memoria, che tendono a “invecchiare”, di solito si usa una variante della funzione esponenziale, detta densità di Weibull:

$$w(t) = a t^{b-1} e^{-(a/b) \cdot t^b}, \quad t \geq 0, \quad a, b > 0. \quad (3.56)$$

I parametri a e b vengono spesso determinati empiricamente dall'analisi dei dati. La curva di Weibull, per $a = b = 1$, è l'esponenziale negativa; tende invece ad assumere una forma a campana per $b > 1$. La corrispondente funzione cumulativa vale $F(t) = \int_0^t w(t) dt = 1 - \exp(-at^b/b)$. Se ad esempio $a = 2$, $b = 2$, allora $w(t) = 2t \exp(-t^2)$ e dalla (3.55) risulta $P\{T > t + \Delta t | t\} = \exp[-(\Delta t)^2 - 2t\Delta t]$. La probabilità di non avere eventi fino a $t + \Delta t$ (ovvero la probabilità di vita, se l'evento è la rottura del sistema) diminuisce se aumenta t .

Generalizziamo ora la legge esponenziale ad eventi poissoniani non contigui, dove interessa il tempo del k -esimo evento. Gli eventi registrati sono in questo caso separati da $k - 1$ eventi, come in Fig. 3.5.

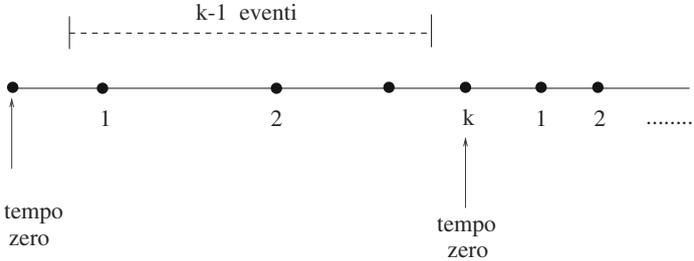


Fig. 3.5. Definizione di evento k -esimo nella densità gamma, relativa ai tempi di arrivo di eventi non contigui

Ponendo nella (3.48) $p_{k-1}(t)$ al posto di $p_0(t)$, otteniamo facilmente la densità erlanghiana di ordine k o *densità gamma* di Fig. 3.6:

$$e_k(t) = \lambda \frac{(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} = \frac{\lambda^k}{\Gamma(k)} t^{k-1} e^{-\lambda t}, \quad t \geq 0, \quad (k-1) \geq 0, \quad (3.57)$$

dove $\Gamma(k) = (k-1)!$ per k intero (andate a vedere, più avanti, la 3.65). La media e la varianza di questa densità sono date da:

$$\mu = \frac{k}{\lambda}, \quad \sigma^2 = \frac{k}{\lambda^2}. \quad (3.58)$$

Per eventi contigui, $k-1 = 0$ e riotteniamo le (3.48-3.49). Quando k aumenta, la media e la varianza aumentano e la densità del flusso $\lambda_k = 1/\mu$ diminuisce. La densità gamma è per costruzione la densità della somma di k variabili aleatorie esponenziali negative indipendenti. Si può ottenere questo risultato anche con la teoria delle funzioni di variabili aleatorie, che svilupperemo nel cap. 5 (si veda il problema 5.5).

Tra le densità di Poisson e di Erlang è facile mostrare che sussiste l'utile relazione:

$$\frac{dp_k(t)}{dt} = e_k(t) - e_{k+1}(t). \quad (3.59)$$

Abbiamo visto che dalle condizioni a)-c) di pagina 88, che definiscono il processo stazionario di Poisson, discende la (3.47), che implica la densità esponenziale negativa dei tempi di arrivo. Mostriamo ora che vale anche l'inverso se gli intertempi sono indipendenti e di legge esponenziale negativa. Infatti, sviluppando in serie le (3.51, 3.52) segue che la probabilità di osservare un evento in un intervallo infinitesimo vale λdt e quella di osservare più eventi è trascurabile (un infinitesimo di ordine superiore). Inoltre, se si verificano k conteggi entro t , significa che t sta tra i tempi di arrivo T_k e T_{k+1} , dell'evento k -esimo e $(k+1)$ -esimo rispettivamente. Dato che

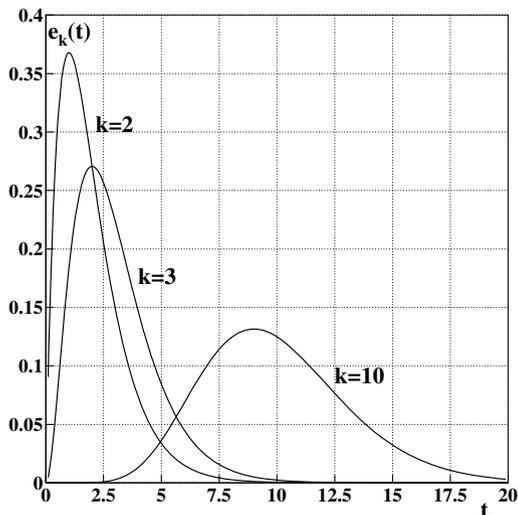


Fig. 3.6. Densità gamma per $\lambda = 1$ e vari valori di k

$P\{T_k \leq t\} = P\{T_k \leq t < T_{k+1}\} + P\{T_{k+1} \leq t\}$, vale la relazione:

$$\begin{aligned} P\{X(t) = k\} &= P\{T_k \leq t < T_{k+1}\} \\ &= P\{T_k \leq t\} - P\{T_{k+1} \leq t\} \\ &= \int_0^t [e_k(t) - e_{k+1}(t)] dt = p_k(t), \end{aligned}$$

dove nell'ultimo passaggio si è fatto uso della (3.59).

Avendo ritrovato la legge di Poisson, che implica l'indipendenza dei conteggi in intervalli di tempo disgiunti, in definitiva si arriva al seguente:

Teorema 3.2 (Indipendenza stocastica). *Condizione necessaria e sufficiente perché un processo sia stazionario di Poisson è che gli intertempi siano indipendenti ed abbiano legge esponenziale negativa.*

3.9 La densità χ^2

Se un vettore ha come componenti delle variabili aleatorie gaussiane, quale sarà la densità di probabilità del suo modulo quadrato? Questo problema riguarda la determinazione della densità di probabilità di una variabile aleatoria funzione di più variabili aleatorie e verrà affrontato in generale nel cap. 5. Tuttavia, qui anticipiamo un risultato particolare, relativo alla densità χ^2 , la quale ha un ruolo importante nell'ambito delle densità di probabilità monodimensionali, che stiamo qui illustrando.

Vogliamo trovare la funzione di densità della variabile

$$Q = \sum_{i=1}^n X_i^2 \quad (3.60)$$

somma dei quadrati di n variabili standard gaussiane tra loro indipendenti. Notiamo subito che la trattazione vale in generale per variabili gaussiane indipendenti, perché una variabile aleatoria può sempre essere resa standard attraverso la trasformazione (3.37).

Indichiamo con $F_Q(q)$ la cumulativa della funzione di densità che stiamo cercando: essa dà la probabilità $P\{Q \leq q\}$ che la variabile Q sia compresa entro una ipersfera di raggio \sqrt{q} . Poiché le variabili X_i provengono dalla densità gaussiana standard (3.42) e sono indipendenti, la probabilità cercata sarà data, in base alle (1.22, 1.23, 2.10), dal prodotto delle probabilità di ottenere un insieme di valori (x_1, x_2, \dots, x_n) sommato (integrato) su tutti gli insiemi che soddisfano alla condizione $\sum_i x_i^2 \leq q$. Si ha quindi:

$$\begin{aligned} P\left\{\sum X_i^2 \leq q\right\} &\equiv F_Q(q) = \int_{\sum x_i^2 \leq q} \dots \int \left(\frac{1}{\sqrt{2\pi}}\right)^n \prod_i e^{-\frac{x_i^2}{2}} dx_1 dx_2 \dots dx_n \\ &= \int_{\sum x_i^2 \leq q} \dots \int \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\sum_i \frac{x_i^2}{2}} dx_1 dx_2 \dots dx_n \quad (3.61) \end{aligned}$$

dove la produttoria nell'integrando viene dal teorema delle probabilità composte e dalle (1.22, 1.23) ed il legame funzionale tra Q e le variabili X_i compare solo nella definizione del dominio di integrazione. Se passiamo alle coordinate sferiche ponendo $\sqrt{\sum_i x_i^2} = r$, l'integrando risulta indipendente dagli angoli e il vincolo funzionale (3.60) dà luogo ad una integrazione sul raggio dell'ipersfera.

Per un noto risultato della geometria differenziale, sappiamo che l'elemento di volume integrato sugli angoli si trasforma, passando da una sfera tridimensionale ad una ipersfera di uno spazio ad n dimensioni, come:

$$\int_{d\Omega} dV = \int_{d\Omega} r^2 \sin \theta dr d\theta d\phi = 4\pi r^2 dr \rightarrow D r^{n-1} dr$$

dove D è una costante ottenuta integrando sulle variabili angolari. L'integrale (3.61) diventa allora:

$$F_Q(q) = F_0 \int_0^{\sqrt{q}} e^{-\frac{r^2}{2}} r^{n-1} dr$$

dove F_0 comprende tutti i fattori costanti apparsi nel calcolo. Se ora operiamo il cambio di variabile

$$r = \sqrt{q}, \quad dr = \frac{1}{2} \frac{1}{\sqrt{q}} dq$$

e conglobiamo ancora tutti i fattori costanti in F_0 , possiamo esprimere $F_Q(q)$ come:

$$F_Q(q) = F_0 \int_0^q e^{-\frac{q}{2}} q^{\frac{n}{2}-1} dq , \quad (3.62)$$

che è la primitiva della funzione di densità χ^2 che stiamo cercando. Abbiamo pertanto, dalla (2.36):

$$p(q) = \frac{dF_Q(q)}{dq} = F_0 e^{-\frac{q}{2}} q^{\frac{n}{2}-1} . \quad (3.63)$$

Ovviamente, questa funzione è definita solo per $q \geq 0$. La costante F_0 si trova imponendo la condizione di normalizzazione

$$F_0 \int_0^\infty e^{-\frac{q}{2}} q^{\frac{n}{2}-1} dq = 1 .$$

Per la valutazione di questo integrale si deve ricorrere all'integrale notevole (lo potete trovare su molte tavole standard di matematica):

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx , \quad (3.64)$$

dove $\Gamma(p)$, detta funzione gamma, si può ottenere integrando ripetutamente per parti la (3.64):

$$\begin{aligned} \Gamma(1) &= 1 \\ \Gamma(1/2) &= \sqrt{\pi} \\ \Gamma(p+1) &= p \Gamma(p) \\ \Gamma(p+1) &= p! \quad \text{per } p \text{ intero} \\ \Gamma(p+1) &= p(p-1)(p-2) \dots \left(\frac{3}{2}\right) \left(\frac{1}{2}\right) \sqrt{\pi} \quad \text{per } p \text{ semi-intero.} \end{aligned} \quad (3.65)$$

Dalle (3.63-3.64) otteniamo infine la densità:

$$p_n(q) dq = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} e^{-\frac{1}{2}q} q^{\frac{1}{2}(n-2)} dq , \quad q \geq 0 , \quad (3.66)$$

la quale individua la distribuzione statistica del *modulo quadro di un vettore di n componenti gaussiane indipendenti*. Se le componenti linearmente indipendenti fossero pari a $\nu < n$, allora nella (3.60) la somma può essere trasformata in una combinazione lineare dei quadrati di ν variabili gaussiane indipendenti e la funzione di densità è sempre data dalla (3.66), se si opera la sostituzione $n \rightarrow \nu$. Le variabili linearmente indipendenti di una distribuzione χ^2 si chiamano *gradi di libertà*.

Nella letteratura fisica internazionale si usa quasi sempre la notazione unica χ^2 per indicare la distribuzione, la variabile aleatoria e le sue realizzazioni

numeriche (!). Poiché questa ci sembra una disinvoltura eccessiva, per restare (parzialmente) coerenti con il nostro testo, noi indicheremo con Q una variabile aleatoria che ha una distribuzione χ^2 di densità (3.66) e con χ^2 (e non con q) i valori numerici di Q ottenuti in prove specifiche. Diremo quindi che $Q(\nu) \sim \chi^2(\nu)$ assume valori χ^2 .

Con questa notazione la densità (3.66) con ν gradi di libertà della variabile Q viene scritta come:

$$p_\nu(\chi^2) d\chi^2 \equiv p(\chi^2; \nu) d\chi^2 = \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} (\chi^2)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}} d\chi^2, \quad (3.67)$$

la quale, come è facile vedere dalle (3.64, 3.65), ha media e varianza pari a

$$\langle Q \rangle = \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} \int_0^\infty x(x)^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} dx = \nu \quad (3.68)$$

$$\text{Var}[Q] = \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} \int_0^\infty (x - \mu)^2 (x)^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} dx = 2\nu. \quad (3.69)$$

Spesso si usa la distribuzione χ^2 ridotto della variabile

$$Q_R(\nu) = \frac{Q(\nu)}{\nu}, \quad (3.70)$$

la quale, in base alle (2.62, 2.63, 3.68, 3.69), ha media e varianza date da:

$$\langle Q_R(\nu) \rangle = 1, \quad \text{Var}[Q_R(\nu)] = \frac{2}{\nu}. \quad (3.71)$$

Nella Tab. D.3 dell'Appendice D sono riportati i valori dell'integrale della densità χ^2 ridotto (vedere Fig. 3.7):

$$P\{Q_R(\nu) \geq \chi_R^2(\nu)\} = \int_{\chi_R^2(\nu)}^\infty \frac{\nu^{\frac{\nu}{2}}}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu x}{2}\right) dx, \quad (3.72)$$

ottenuto applicando la trasformazione (3.70) alla (3.67). La tabella fornisce la probabilità di superare un valore assegnato di χ_R^2 . Di essa si farà ampio uso in statistica.

Possiamo riassumere questi risultati, trovati da Helmer nel 1876 e generalizzati da Pearson nel 1900, con il:

Teorema 3.3 (di Pearson). *La somma dei quadrati di ν variabili gaussiane indipendenti è una variabile aleatoria che ha densità chi-quadrato (3.67) con ν gradi di libertà, detta $\chi^2(\nu)$.*

Sarà utile nel seguito anche il seguente:

Teorema 3.4 (Additività della variabile χ^2). *Siano Q_1 e Q_2 due variabili aleatorie indipendenti. Se esse hanno densità χ^2 con ν_1 e ν_2 gradi di libertà rispettivamente, la variabile*

$$Q = Q_1 + Q_2 \quad (3.73)$$

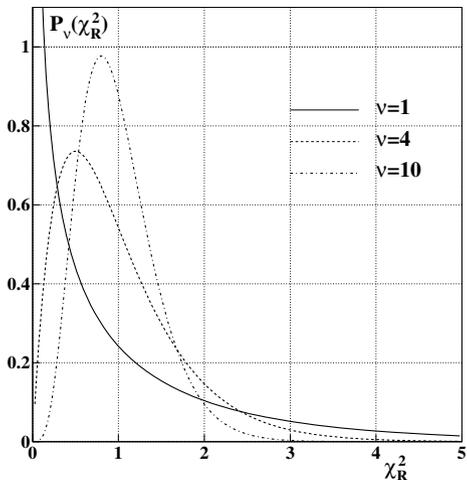


Fig. 3.7. Densità χ^2 ridotto per vari gradi di libertà

ha una densità $\chi^2(\nu)$ con $\nu = \nu_1 + \nu_2$ gradi di libertà: $Q \sim \chi^2(\nu_1 + \nu_2)$. Inoltre, se $Q \sim \chi^2(\nu)$ e $Q_1 \sim \chi^2(\nu_1)$, allora $Q_2 \sim \chi^2(\nu - \nu_1)$.

Dimostrazione. La dimostrazione della prima parte del teorema è immediata e può essere vista come un lemma del teorema di Pearson 3.3: poiché la somma dei quadrati di ν_1 variabili standard indipendenti si distribuisce come χ^2 , se a queste ν_1 se ne aggiungono, tramite la (3.73), altre ν_2 indipendenti, la somma risultante è una somma dei quadrati di $\nu_1 + \nu_2$ variabili gaussiane standard indipendenti, da cui l'enunciato. Anche la dimostrazione della seconda parte è facile se si utilizzano le funzioni generatrici dell'Appendice B e la (B.10).

Nell'applicare il teorema occorre ricordarsi sempre che esso vale per variabili $Q \sim \chi^2$, non per le variabili ridotte $Q_R \sim \chi_R^2$ della (3.70). \square

In SCILAB le probabilità della distribuzione χ^2 non ridotto con DF gradi di libertà vengono calcolate con la funzione `cdfchi` nei modi seguenti:

```
[P, 1-P]=cdfchi (PQ, X, DF)
[X]=cdfchi (X, DF, P, 1-P)
[DF]=cdfchi (DF, P, 1-P, X)
```

Nel modo “PQ” il risultato P contiene il valore cumulativo da zero ad X; la modalità “X” calcola il valore X corrispondente al valore cumulativo P; con “DF” viene calcolato il numero DF di gradi di libertà per il quantile X_p , cioè quando ad un valore X dello spettro corrisponde una probabilità cumulativa P.

Come vedremo nei prossimi esercizi, la densità χ^2 è di importanza fondamentale non solo in statistica, ma anche in fisica.

Esercizio 3.10

Trovare la distribuzione statistica del modulo R di un vettore tridimensionale avente componenti gaussiane indipendenti (X, Y, Z) di media nulla e varianza σ^2 .

Risposta. Supponiamo per un momento le componenti siano standard con varianza unitaria. La densità del modulo quadrato del vettore è data allora dalla (3.66) con $n = 3$:

$$p_3(q) dq = \frac{1}{\sqrt{2\pi}} e^{-\frac{q}{2}} q^{\frac{1}{2}} dq ,$$

poiché $\Gamma(3/2) = \Gamma(1/2 + 1) = \sqrt{\pi}/2$ dalle (3.65). Questa densità rappresenta la probabilità che il quadrato del modulo del vettore sia compreso in $(q, q + dq)$. Dato che è richiesto il modulo del vettore (non il suo quadrato), operiamo la trasformazione

$$q = (x^2 + y^2 + z^2) = r^2 , \quad r = \sqrt{(x^2 + y^2 + z^2)} = \sqrt{q} , \quad 2r dr = dq$$

ottenendo:

$$m(r) dr = \sqrt{\frac{2}{\pi}} r^2 e^{-\frac{r^2}{2}} dr . \quad (3.74)$$

Finora abbiamo considerato le variabili (X, Y, Z) come gaussiane standard. Per tenere conto della varianza non unitaria, operiamo la trasformazione

$$x \rightarrow \frac{x}{\sigma} , \quad y \rightarrow \frac{y}{\sigma} , \quad z \rightarrow \frac{z}{\sigma} ,$$

che porta a ridefinire r come:

$$r \rightarrow \frac{r}{\sigma} \quad dr \rightarrow \frac{dr}{\sigma} .$$

Sappiamo che R è il modulo di un vettore di componenti gaussiane non standard (X, Y, Z) aventi varianza finita σ^2 . Inserendo la trasformazione nella (3.74) otteniamo:

$$m(r) dr = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma^3} r^2 e^{-\frac{r^2}{2\sigma^2}} dr \quad (3.75)$$

che è la famosa densità di Maxwell o maxwelliana. Poiché R^2/σ^2 è una variabile χ^2 con 3 gradi di libertà (cioè $\chi^2(3)$), essa ha come media e varianza, in base alle (2.63, 3.68, 3.69), i valori

$$\left\langle \frac{R^2}{\sigma^2} \right\rangle = 3 , \quad \text{Var} \left[\frac{R^2}{\sigma^2} \right] = \frac{1}{\sigma^4} \text{Var}[R^2] = 6$$

e quindi

$$\langle R^2 \rangle = 3\sigma^2 ; \quad \text{Var}[R^2] = 6\sigma^4 . \quad (3.76)$$

Esercizio 3.11

Trovare la densità di energia delle molecole di un gas perfetto alla temperatura assoluta T .

Risposta. In un gas perfetto le componenti (V_x, V_y, V_z) della velocità delle molecole sono variabili aleatorie che soddisfano alle condizioni del teorema Limite Centrale 3.1. Il modulo della velocità sarà pertanto distribuito secondo la maxwelliana (3.75). Se teniamo conto delle note relazioni tra l'energia cinetica e la velocità di una molecola di massa m :

$$E = \frac{1}{2}mv^2, \quad v = \sqrt{\frac{2E}{m}}, \quad mv \, dv = dE,$$

possiamo scrivere:

$$m(E) \, dE = \frac{2}{\sqrt{\pi}} \frac{1}{\sigma^3} \frac{1}{m} \sqrt{\frac{E}{m}} e^{-\frac{E}{m\sigma^2}} \, dE.$$

Se ora usiamo la nota relazione tra varianza e temperatura,

$$m\sigma^2 = KT, \quad \sigma = \sqrt{\frac{KT}{m}}, \quad (3.77)$$

dove K è la costante di Boltzmann, otteniamo infine:

$$m(E) \, dE = \frac{2}{\sqrt{\pi}} \frac{1}{KT} \sqrt{\frac{E}{KT}} e^{-\frac{E}{KT}} \, dE, \quad (3.78)$$

che è la famosa funzione di Boltzmann.

Dalle (3.76, 3.77) si ottiene un altro fondamentale risultato della termodinamica, quello che lega l'energia cinetica media $\langle U \rangle$ della molecola alla temperatura assoluta:

$$\langle U \rangle = \frac{1}{2}m \langle V^2 \rangle = \frac{3}{2}m\sigma^2 = \frac{3}{2}KT.$$

A questo punto vorremmo farvi notare che abbiamo ottenuto, in questo esercizio e nel precedente, alcuni risultati fondamentali della fisica statistica utilizzando come ipotesi fisiche solo il teorema Limite Centrale e le relazioni (3.77).

3.10 La densità uniforme

Una variabile aleatoria X continua, che assume valori nell'intervallo finito $[a, b]$, si dice uniforme in $[a, b]$ e si denota con $X \sim U(a, b)$ quando ha una densità di probabilità costante, detta densità uniforme o piatta (riportata in Fig. 3.8), data da:

$$u(x) = \begin{cases} \frac{1}{b-a} & \text{per } a \leq x \leq b \\ 0 & \text{per } x < a, x > b \end{cases}. \quad (3.79)$$

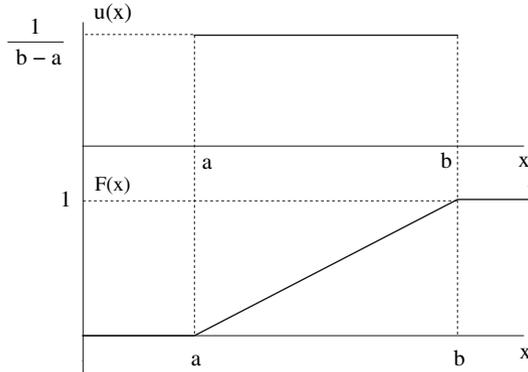


Fig. 3.8. Densità uniforme $u(x)$ e corrispondente funzione cumulativa $F(x)$ di una variabile aleatoria che assume valori compresi in $[a, b]$

È facile vedere che la condizione di normalizzazione

$$P\{a \leq X \leq b\} = \int_a^b u(x) dx = 1$$

è soddisfatta e che media e varianza sono date da:

$$\mu = \frac{1}{b-a} \int_a^b x dx = \frac{b+a}{2}, \quad (3.80)$$

$$\sigma^2 = \frac{1}{b-a} \int_a^b \left(x - \frac{b+a}{2}\right)^2 dx = \frac{(b-a)^2}{12}, \quad (3.81)$$

dove l'ultimo integrale può essere facilmente valutato operando la sostituzione $y = x - (b+a)/2$, $dy = dx$ e integrando tra gli estremi $-(b-a)/2$ e $(b-a)/2$.

Spesso si considera la densità uniforme per valori compresi tra gli estremi $a = 0$ e $b = \Delta$. Ovviamente in questo caso media e varianza sono date da:

$$\mu = \frac{\Delta}{2}, \quad \sigma^2 = \frac{\Delta^2}{12}. \quad (3.82)$$

La densità esiste per $\mu - \Delta/2 \leq x \leq \mu + \Delta/2$ e la deviazione standard vale $\sigma = \Delta/\sqrt{12}$.

Per una variabile aleatoria $a \leq X \leq b$, avente densità uniforme, la probabilità di localizzazione in (x_1, x_2) è proporzionale all'ampiezza dell'intervallo:

$$P\{x_1 \leq X \leq x_2\} = \frac{1}{b-a} \int_{x_1}^{x_2} dx = \frac{x_2 - x_1}{b-a}. \quad (3.83)$$

Viceversa, se una variabile aleatoria continua soddisfa alla (3.83), essa è distribuita in modo uniforme.

Veniamo ora ad un teorema molto semplice, ma importante ed estremamente generale.

Teorema 3.5 (Variabili aleatorie cumulative). *Se X è una variabile aleatoria avente densità $p(x)$ continua, la variabile aleatoria cumulativa C :*

$$C(X) = \int_{-\infty}^X p(x) dx \quad (3.84)$$

è uniforme in $[0, 1]$, cioè $C \sim U(0, 1)$.

Dimostrazione. La probabilità che un valore X sia compreso entro $[x_1, x_2]$ è identica, per costruzione, alla probabilità che la variabile cumulativa C sia compresa entro i valori $[c_1 \equiv c(x_1), c_2 \equiv c(x_2)]$. Utilizzando la (2.34) abbiamo allora:

$$\begin{aligned} P\{c_1 \leq C \leq c_2\} &= P\{x_1 \leq X \leq x_2\} = \int_{x_1}^{x_2} p(x) dx \\ &= \int_{-\infty}^{x_2} p(x) dx - \int_{-\infty}^{x_1} p(x) dx = c_2 - c_1, \end{aligned} \quad (3.85)$$

da cui si vede che $C \sim U(0, 1)$, poiché soddisfa la relazione (3.83) con $(b - a) = 1$. \square

L'importanza concettuale e pratica del teorema non vi deve sfuggire: la variabile cumulativa è *sempre uniforme, qualunque sia la distribuzione d'origine*.

Se l'integrale (3.84) è noto per via analitica, allora i valori della variabile cumulativa C sono esprimibili come una funzione nota $c = F(x)$. Se questa funzione è invertibile, allora la variabile

$$X = F^{-1}(C) \quad (3.86)$$

ha densità $p(x)$. Se si dispone di un generatore di variabili uniformi in $[0, 1]$ (una roulette o una funzione **random** su calcolatore), *si possono generare variabili aventi una densità qualsiasi* utilizzando l'equazione:

$$X = F^{-1}(\mathbf{random}) . \quad (3.87)$$

L'estensione del teorema 3.5 a variabili discrete richiede un minimo di attenzione e l'utilizzo della (2.29) di pagina 48. Se C è una variabile uniforme, tenendo presente che la cumulativa $F(x)$ è definita in $[0, 1]$, dalla (3.83) otteniamo:

$$\begin{aligned} P\{X = x_k\} &= P\{x_{k-1} < X \leq x_k\} \\ &= F(x_k) - F(x_{k-1}) = P\{F(x_{k-1}) < C \leq F(x_k)\} . \end{aligned} \quad (3.88)$$

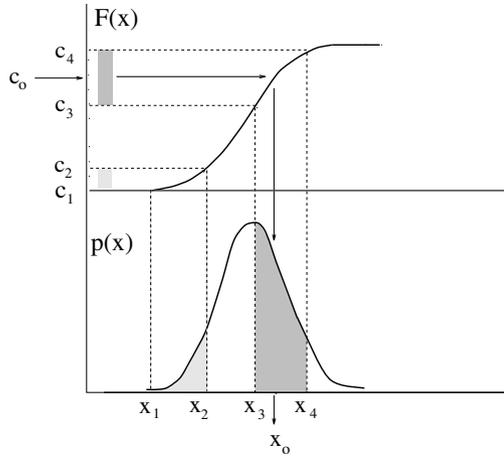


Fig. 3.9. Interpretazione grafica del teorema sulle variabili cumulative. Le funzioni $p(x)$ e $F(x)$ sono rispettivamente una generica densità di probabilità e la sua corrispondente cumulativa

L'equivalente nel discreto della (3.87), se si dispone di un generatore **random**, diventa allora:

$$\{X = x_k\} \text{ se } \{F(x_{k-1}) < \text{random} \leq F(x_k)\} \quad (\text{se } k = 1, F(x_0) = 0) . \tag{3.89}$$

Sulle (3.87, 3.89) si fondano tutti i metodi di simulazione Monte Carlo, su cui torneremo in dettaglio nel cap. 7.

La formula (3.86) ha una convincente interpretazione grafica, che riportiamo in Fig. 3.9. Consideriamo le probabilità $P\{x_1 \leq X \leq x_2\}$ e $P\{x_3 \leq X \leq x_4\}$ definite entro i valori $[x_1, x_2]$ e $[x_3, x_4]$: esse sono rappresentate dalle aree sottese dalla funzione di densità $p(x)$ ed ombreggiate in Fig. 3.9. L'area relativa ai valori meno probabili compresi in $[x_1, x_2]$ è minore di quella relativa ai valori più probabili compresi in $[x_3, x_4]$. Il valore di queste due aree è identico, per costruzione, alla lunghezza degli intervalli $[c_1, c_2]$ e $[c_3, c_4]$ di Fig. 3.9, ottenuti dalla funzione cumulativa $F(x)$. Se ora consideriamo sull'asse delle ordinate della cumulativa una variabile $C \sim U(0, 1)$, essa cadrà più frequentemente in $[c_3, c_4]$ che in $[c_1, c_2]$ con probabilità date *esattamente dalla ampiezza di questi intervalli*. Pertanto, se in corrispondenza di un valore c_0 assunto dalla variabile uniforme C ricaviamo (per via grafica o analitica) il corrispondente valore x_0 indicato in Fig. 3.9 e ripetiamo il procedimento più volte, otterremo un campione di valori della variabile X di densità $p(x)$.

Esercizio 3.12

Disponendo su calcolatore di un valore **random** uniformemente compreso in $[0, 1]$, estrarre a caso i tempi di arrivo di eventi stocasticamente indipendenti.

Risposta. Se gli eventi sono stocasticamente indipendenti, la distribuzione dei tempi di arrivo è esponenziale negativa e la corrispondente cumulativa, che dà la probabilità di un arrivo in $[0, t]$, è data dalla (3.51). In base al teorema 3.5 questa funzione è una variabile aleatoria uniforme. Abbiamo pertanto:

$$\mathbf{random} = 1 - e^{-\lambda t} .$$

Invertendo questa equazione otteniamo:

$$t = -\frac{1}{\lambda} \ln(1 - \mathbf{random}) , \quad (3.90)$$

Poiché, analogamente alla variabile \mathbf{random} , anche $(1 - \mathbf{random})$ è $\sim U(0, 1)$, la relazione precedente risulta equivalente a:

$$t = -\frac{1}{\lambda} \ln(\mathbf{random}) . \quad (3.91)$$

Le (3.90, 3.91) sono un esempio delle formule generali (3.86, 3.87).

3.11 Disuguaglianza di Tchebychev

La deviazione standard, che è un indicatore della dispersione dei valori di una variabile aleatoria intorno alla media, gode di un'importante e generale proprietà, che andiamo a illustrare.

Consideriamo una generica distribuzione statistica di media μ e di di varianza σ^2 finita e consideriamo l'intervallo dei valori $[\mu - K\sigma, \mu + K\sigma]$, dove K è un numero reale positivo. Ovviamente, i punti fuori da questo intervallo sono definiti dalla condizione $|x - \mu| > K\sigma$. Considerando l'espressione della varianza di variabili continue (2.56), possiamo scrivere:

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) dx \\ &= \int_{\mu - K\sigma}^{\mu + K\sigma} (x - \mu)^2 p(x) dx + \int_{|x - \mu| > K\sigma} (x - \mu)^2 p(x) dx \\ &\geq \int_{|x - \mu| > K\sigma} (x - \mu)^2 p(x) dx \geq K^2 \sigma^2 \int_{|x - \mu| > K\sigma} p(x) dx \\ &= K^2 \sigma^2 \left(1 - \int_{\mu - K\sigma}^{\mu + K\sigma} p(x) dx \right) . \end{aligned}$$

Dall'ultima relazione otteniamo:

$$\int_{\mu - K\sigma}^{\mu + K\sigma} p(x) dx \geq 1 - \frac{1}{K^2} \quad (3.92)$$

che è la disuguaglianza di Tchebychev. Questa è una relazione dimostrabile facilmente anche per variabili discrete ed è del tutto generale, perché l'unica condizione che è stata imposta su $p(x)$ è la varianza finita.

Vediamo ora quali informazioni sono contenute in questa legge generale. Analogamente alla (3.35), essa può essere messa nella forma:

$$P\{|X - \mu| \leq K\sigma\} = \int_{\mu - K\sigma}^{\mu + K\sigma} p(x) dx \geq 1 - \frac{1}{K^2} = \begin{cases} 0 & \text{per } K = 1 \\ 0.75 & \text{per } K = 2 \\ 0.89 & \text{per } K = 3 \end{cases} \quad (3.93)$$

da cui si vede che *negli intervalli centrati sulla media e ampi 2σ e 3σ sono compresi rispettivamente almeno il 75% e il 90% della probabilità totale.*

La (3.93) dà luogo a una legge 3σ generalizzata, che consiste nel ritenere trascurabili *per qualunque distribuzione statistica*, le probabilità di avere valori fuori da $[\mu - 3\sigma, \mu + 3\sigma]$.

Generalmente questa è una approssimazione molto buona, perché la disuguaglianza di Tchebychev, che prevede non più di un 10% di probabilità al di fuori di $\pm 3\sigma$, è quasi sempre una notevole sovrastima dei valori effettivi. Ad esempio, nel caso della densità di Gauss, “fuori da 3σ ” (così si dice in gergo) abbiamo solo lo 0.3% dei valori, nel caso della densità uniforme tutti i valori sono compresi entro $\pm 2\sigma$ (controllate come esercizio).

L'uso di considerare solo i valori entro 3σ è molto diffuso, ed in genere porta, come si è detto, a dei risultati corretti. Tuttavia, in casi particolari, se si ha a che fare con densità “molto larghe”, è bene ricordarsi che con questo metodo si può compiere un errore sensibile, fino al 10%, come mostrato dalla (3.93).

3.12 Come utilizzare il calcolo delle probabilità

In Fig. 3.10 ed in Tab. 3.1 sono riportate le distribuzioni statistiche fondamentali che abbiamo fin qui ottenuto col calcolo delle probabilità. Il punto di partenza è la distribuzione binomiale (2.30), relativa agli eventi generati in modo discreto, indipendente e con probabilità p costante. Il limite per np , $n(1-p) \gg 1$ (in pratica np , $n(1-p) > 10$), dove n è il numero dei tentativi effettuati, porta alla densità gaussiana o normale, mentre il limite per $n \gg 1$, $p \ll 1$ (in pratica $n > 10$, $p < 0.1$) porta alla densità poissoniana. Quest'ultima densità, quando $np > 10$ (e quindi anche $n(1-p) > 10$, dato che $p \ll 1$), evolve anch'essa verso la densità gaussiana, mantenendo però sempre la relazione (3.16) $\mu = \sigma^2$, tipica dei processi poissoniani.

Abbiamo anche visto che la gaussiana e la poissoniana, lungi dall'essere solo casi limite della densità binomiale, sono le distribuzioni di riferimento di molti fenomeni naturali di notevole importanza.

La gaussiana è la densità limite della sovrapposizione lineare di variabili aleatorie indipendenti, nessuna delle quali è preponderante sulle altre. Questo

risultato viene normalmente chiamato teorema Limite Centrale (o Centrale Limite) ed è stato da noi riportato come teorema 3.1.

La distribuzione statistica del modulo quadrato di un vettore di componenti gaussiane è individuata dalla densità χ^2 , che in tre dimensioni prende il nome di densità di Maxwell o maxwelliana.

La distribuzione poissoniana è invece la distribuzione universale del numero di eventi indipendenti generati in modo discreto con probabilità costante nel tempo (generazione stocastica). Il tempo di arrivo tra questi eventi segue anch'esso delle distribuzioni universali, l'esponenziale negativa e la distribuzione gamma o erlanghiana.

Infine, anche la densità uniforme è di importanza generale, perché, come mostrato nel teorema 3.5, essa individua la distribuzione di tutte le variabili aleatorie cumulative. Come vedremo, questo principio è la base del metodo di simulazione Monte Carlo.

Crediamo che vi sia già chiaro, e lo sarà comunque sempre di più nel seguito, come le distribuzioni statistiche dedotte col calcolo delle probabilità elementare forniscano uno schema coerente per l'interpretazione di una quantità veramente notevole di fenomeni naturali. Questo schema dovrebbe essere parte essenziale della formazione culturale dello scienziato, del ricercatore e, in genere, di chiunque voglia osservare la realtà con occhio scientifico. Cercheremo ora di spiegare meglio questo punto, con una serie di esempi ed alcuni approfondimenti.

La sola intuizione non è sufficiente per interpretare i fenomeni aleatori: se lanciamo in aria una moneta 1000 volte intuiamo che mediamente avremo testa (o croce) 500 volte, ma se otteniamo 450 teste e 550 croci come facciamo a sapere se il risultato è compatibile col caso o se invece la moneta è truccata o i lanci non sono stati regolari? (A tra poco la risposta, nell'esercizio 3.13).

Utilizzando il calcolo delle probabilità possiamo affrontare in modo quantitativo questi problemi, utilizzando funzioni di densità $p(x)$ (continue o discrete) e risolvendo somme (per variabili discrete) o integrali (per variabili continue) del tipo:

$$P\{a \leq X \leq b\} = \sum_{x=a}^b p_x \rightarrow \int_a^b p(x) dx, \quad (3.94)$$

che forniscono il *livello di probabilità*, cioè la probabilità di ottenere, se il modello rappresentato da $p(x)$ è valido, eventi con valori compresi nell'*intervallo di probabilità* $[a, b]$.

È possibile quindi giudicare se un certo scostamento dal valore atteso è compatibile col caso oppure no. Come abbiamo già accennato al par. 2.6, nel primo caso si dice che si è in presenza di una normale *fluttuazione statistica* e che il modello rappresentato da $p(x)$ è valido (o meglio, che non è stato *falsificato* dall'esperienza); nel secondo caso si interpreta il risultato come una *deviazione significativa* e si rigetta il modello. In questo tipo di analisi si possono in molti casi utilizzare, in alternativa agli integrali del tipo (3.94),

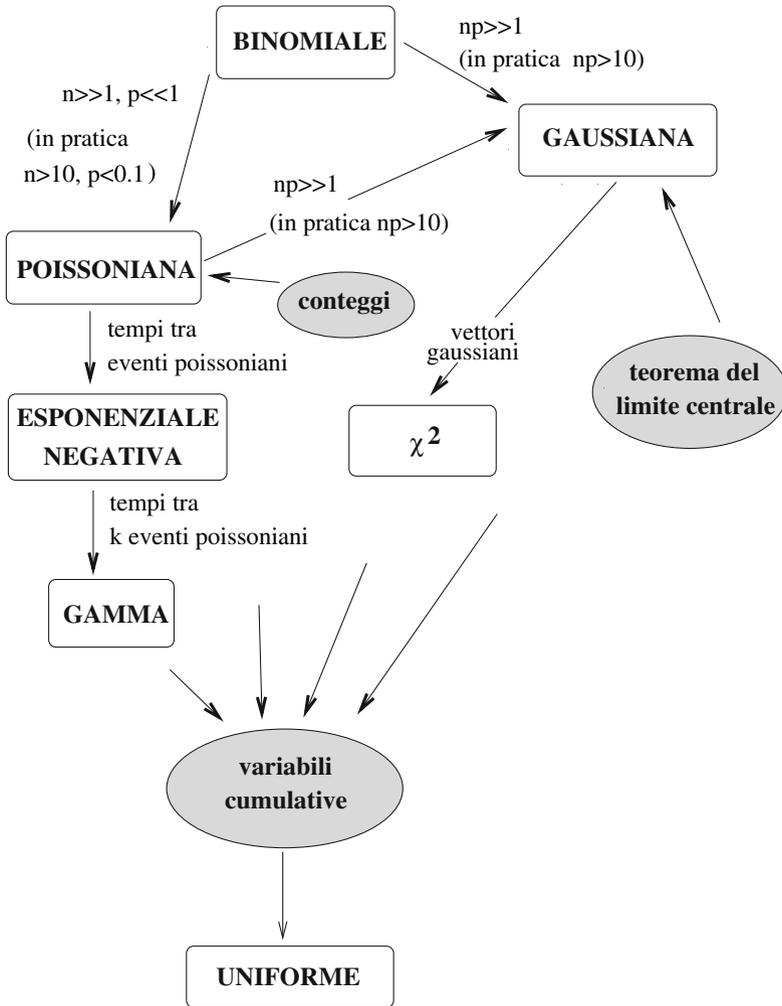


Fig. 3.10. Connessione tra le distribuzioni statistiche fondamentali del calcolo delle probabilità

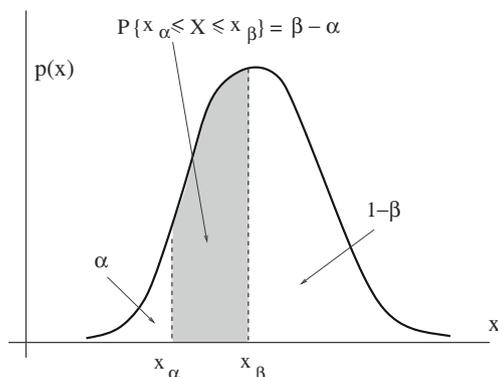
i valori quantili (2.19) oppure la media e la deviazione standard della distribuzione modello. In questo modo l’analisi risulta più rapida e permette di acquisire utili automatismi mentali, facili da ritenere a memoria, che consentono di giudicare anche “a occhio”, ma in modo rapido e corretto, i risultati sperimentali.

Se sono noti i valori quantili della distribuzione, ponendo $a = x_\alpha$ e $b = x_\beta$, l’intervallo di probabilità, rappresentato anche in Fig. 3.11, diventa:

$$P\{x_\alpha \leq X \leq x_\beta\} = \beta - \alpha . \tag{3.95}$$

Tabella 3.1. Le densità fondamentali del calcolo delle probabilità

Nome	Densità	Media	Deviazione standard	Commenti
binomiale	$\frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$	np	$\sqrt{np(1-p)}$	successi in prove indipendenti con probabilità costante
gaussiana	$\frac{\exp[-(x-\mu)^2/2\sigma^2]}{\sqrt{2\pi}\sigma}$	μ	σ	combinazione lineare di variabili indipendenti
chi-quadrato	$\frac{(\chi^2)^{\frac{\nu}{2}-1} \exp(-\frac{\chi^2}{2})}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})}$	ν	$\sqrt{2\nu}$	modulo di un vettore gaussiano
poissoniana	$\frac{\mu^x}{x!} e^{-\mu}$	μ	$\sqrt{\mu}$	conteggi
esponenziale	$\lambda e^{-\lambda t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda}$	tempi tra variabili poissoniane
gamma	$\frac{\lambda^k}{\Gamma(k)} t^{k-1} e^{-\lambda t}$	$\frac{k}{\lambda}$	$\frac{\sqrt{k}}{\lambda}$	somma di k variabili esponenziali negative
uniforme	$\frac{1}{\Delta} \quad (0 \leq x \leq \Delta)$	$\frac{\Delta}{2}$	$\frac{\Delta}{\sqrt{12}}$	variabili cumulative

**Fig. 3.11.** Rappresentazione grafica del livello di probabilità $\beta - \alpha$

Se X ha una densità simmetrica, spesso si usa la scrittura:

$$P \left\{ -t_{1-\alpha/2} \leq \frac{X - \mu}{\sigma} \leq t_{1-\alpha/2} \right\} = 1 - \alpha . \quad (3.96)$$

In molti casi t_α sono i quantili della curva normale standard (che si ricavano con facili passaggi dalla Tab. D.1), oppure della densità di Student (che vedremo più avanti). I quantili gaussiani su molti testi sono scritti come z_α .

Livelli usati di solito in statistica sono $1 - \alpha = 0.90, 0.95, 0.99, 0.999$, cui corrispondono i valori quantili gaussiani $t_{1-\alpha/2} = 1.64, 1.96, 2.58, 3.29$. In alternativa, come faremo spesso in questo testo, si può adottare la convenzione, usuale tra i fisici, che parametrizza l'intervallo di probabilità secondo la legge 3σ (3.35). *La deviazione standard è considerata come l'unità di misura internazionale delle fluttuazioni statistiche.* Per calcolarne il valore, non si può procedere per via intuitiva, ma si deve utilizzare il calcolo delle probabilità.

Allora, un valore di 450 successi in 1000 lanci, è una fluttuazione statistica rispetto a 500 oppure una deviazione significativa? La risposta è nell'esempio che segue.

Esercizio 3.13

In mille lanci di una moneta si sono ottenute 450 teste. È compatibile questo risultato con l'ipotesi di lanci casuali di una moneta non truccata?

Risposta. *Il modello assunto come riferimento, che in statistica verrà chiamato ipotesi nulla (null hypothesis), prevede che la probabilità a priori di ottenere testa in un singolo lancio sia $p = 0.50$ e che le probabilità dei valori possibili di mille lanci, che vanno da 0 a 1000, si possano calcolare dalla distribuzione binomiale con media e deviazione standard date dalle (3.6):*

$$\begin{aligned} \mu &= np = 500 \\ \sigma &= \sqrt{np(1-p)} = \sqrt{500 \cdot 0.5} = \sqrt{250} = 15.8 . \end{aligned}$$

Il valore della frequenza empirica, cioè dell'evento osservato, è invece $f = 0.45$, che corrisponde ad un valore della variabile standard pari a

$$t = \frac{500 - 450}{15.8} = 3.16 .$$

Il risultato si scosta dal valore medio della distribuzione prevista di 3.16 deviazioni standard. Dato che $np = n(1-p) = 1000 \cdot 0.5 = 500 \gg 10$, possiamo utilizzare l'approssimazione gaussiana per la variabile standard, ed utilizzare la Tab. D.1 della funzione gaussiana degli errori, riportata in Appendice D. Da questa tabella leggiamo, in corrispondenza di $t = 3.16$, il valore 0.4992. L'area della coda alla sinistra di t è data da:

$$P\{T < t = -3.16\} = 0.5000 - 0.4992 = 8 \cdot 10^{-4} ,$$

che è la probabilità di ottenere per puro caso, se il modello è valido, valori ≤ 450 .

Fate ora bene attenzione a questo passo cruciale: se rifiutiamo il modello quando è vero, la probabilità che abbiamo di sbagliare non è più grande di 8 su 10000. Si dice anche che il dato è in accordo col modello con un livello di significatività di $8 \cdot 10^{-4}$.

In conclusione, dato il livello di significatività molto piccolo, possiamo dire, con una piccola probabilità di sbagliare, che 450 successi su mille lanci rappresenta un evento in disaccordo con il modello binomiale con $p = 1/2$, che assume lanci indipendenti di una moneta non truccata.

In genere si ritiene falsificata l'ipotesi quando il livello di significatività osservato è al di sotto dell'1-5%. Tuttavia, la decisione dipende dal tipo di problema che si sta considerando, ed è di tipo soggettivo. Se invece della bontà o meno di una moneta stessimo considerando la sicurezza di un aereo e un certo risultato sperimentale ci fornisse un livello di significatività dell'1 per mille rispetto all'ipotesi nulla di un difetto di progettazione, probabilmente non ce la sentiremmo di rigettare l'ipotesi e concludere che l'aereo è ben progettato. Infatti, il test indica in questo caso che un aereo su mille potrebbe cadere. Il calcolo delle probabilità permette di quantificare le possibilità che vengono prese in considerazione nello studio di un problema, ma spesso nella decisione finale occorre fare una analisi costi/benefici che tenga in considerazione anche fattori non strettamente matematici e statistici. Vi sono comunque casi in cui la decisione è facile, perché si raggiunge un livello di significatività così piccolo da coincidere in pratica con la certezza. Ad esempio, se nei mille lanci della moneta si ottenessero 420 teste, la variabile standard varrebbe 5.1 e la probabilità di sbagliare rigettando l'ipotesi "moneta buona" sarebbe praticamente nulla.

Nell'esperimento reale di Tab. 2.2, come risulta dal problema 2.10 di pagina 69, si sono ottenute 479 teste su mille lanci. Questo risultato corrisponde ad un valore della variabile standard $t = |500 - 479|/15.8 = 1.39$, alla quale la Tab. D.1 assegna un valore di 0.4177, corrispondente ad un livello di significatività dell' 8.2%. Abbiamo qui un primo punto fermo nell'analisi dell'esperimento di Tab. 2.2 (sul quale siamo tornati e torneremo più volte nel corso del testo): il numero globale di teste ottenute è in ragionevole accordo col modello binomiale ed una probabilità a priori di 1/2.

Nell'esercizio precedente si è fatto ricorso al cosiddetto test a una coda. Nell'esempio che segue si effettua invece il test a due code, in cui si scartano i valori al di fuori di un intervallo, cioè quelli troppo piccoli o troppo grandi.

Esercizio 3.14

Qual è la probabilità di sbagliare adottando come regola di decisione per definire una moneta buona (cioè con probabilità 1/2) un numero di successi in mille lanci compreso tra 450 e 550?

Risposta. *Utilizzando i dati dell'esercizio precedente otteniamo immediatamente il risultato:*

$$P\{|T| > t = 3.16\} = 2(0.5000 - 0.4992) = 1.6 \cdot 10^{-3} ,$$

il cui significato è che si scartano circa 2 monete buone su 1000 nell'ipotesi che tutte le monete provate siano buone.

Chiediamoci ora, riferendoci all'ultimo esercizio, quante monete cattive si accettano. Qual è, in altri termini, la probabilità di accettare come giusta una ipotesi sbagliata? In genere trovare questa probabilità non è facile, perché si dovrebbero conoscere tutte le probabilità a priori delle variabili coinvolte nel problema. Nel caso delle monete, dovrebbero essere note le probabilità vere di tutte le monete utilizzate nel test, come chiarito nell'esercizio che segue.

Esercizio 3.15

Nello stock di monete dell'esercizio precedente, ce n'è una con probabilità a priori $p = 0.6$ per la faccia di interesse. Qual è la probabilità di accettarla come buona adottando ancora come regola di decisione $450 \leq x \leq 550$?

Risposta. *Dobbiamo calcolare $P\{450 \leq X \leq 550\}$ per una gaussiana di media e deviazione standard pari a:*

$$\begin{aligned} \mu &= 0.6 \cdot 1000 = 600 \\ \sigma &= \sqrt{1000 \cdot 0.6 \cdot (1 - 0.6)} = 15.5 . \end{aligned}$$

Considerando i due valori standard:

$$\begin{aligned} t_{450} &= \frac{450 - 600}{15.5} = -9.68 \\ t_{550} &= \frac{550 - 600}{15.5} = -3.22 , \end{aligned}$$

dalla Tab. D.1 in Appendice D otteniamo la probabilità:

$$\begin{aligned} P\{450 \leq X \leq 550\} &= P\{t_{450} \leq T \leq t_{550}\} \\ &= P\{t_{450} \leq T \leq 0\} - P\{t_{550} \leq T \leq 0\} \\ &= 0.5000 - 0.4993 = 7 \cdot 10^{-4} , \end{aligned}$$

pari a una probabilità di sbagliare, accettando la moneta, di circa 7 su 10 000.

Negli esempi trattati finora si è sempre usata l'approssimazione gaussiana della distribuzione binomiale. Ecco ora un caso diverso.

Esercizio 3.16

In una popolazione di 10 000 abitanti i dati storici riguardanti una malattia rara fanno registrare 4 casi l'anno. Se in un anno si verificano 10 casi, siamo in presenza di un incremento della malattia o si tratta semplicemente di una fluttuazione statistica?

Risposta. L'ipotesi di lavoro più ragionevole sta nell'assumere come modello del fenomeno la distribuzione di Poisson di media $\mu = 4$. In questo caso non possiamo utilizzare l'approssimazione gaussiana, che richiede $\mu > 10$ e siamo costretti ad utilizzare la (3.94). La probabilità di osservare, per puro caso, almeno 10 casi di malattia quando la media annuale è 4 vale allora:

$$\begin{aligned} P\{X \geq 10\} &= \sum_{x=10}^{\infty} \frac{(4)^x}{x!} e^{-4} = 1 - e^{-4} \sum_{x=0}^9 \frac{(4)^x}{x!} \\ &= 1 - (0.0183 + 0.0733 + 0.1465 + 0.1954 + 0.1954 \\ &\quad + 0.1563 + 0.1042 + 0.0595 + 0.0298 + 0.0132) \\ &= 1 - 0.9919 \simeq 8 \cdot 10^{-3}. \end{aligned}$$

In alternativa, si può usare l'istruzione SCILAB:

```
x= 1-cdfpoi(PQ,9,4)
```

che dà come risultato `x=0081322`.

Questo numero rappresenta il livello di significatività osservato dell'ipotesi, cioè la probabilità di scartare una ipotesi giusta. Possiamo allora ragionevolmente dire che siamo in presenza di un incremento della malattia, con una probabilità di sbagliarci di circa l'8 per mille.

Da ultimo, consideriamo un esempio molto istruttivo, che si può considerare come il paradigma del modo di operare della scienza per ciò che riguarda il rigetto (falsificazione) o meno di una teoria.

Esercizio 3.17

Una commissione di astrologi interroga una persona, di cui ignora le generalità anagrafiche, cercando di identificarne il segno zodiacale. Alla fine dell'interrogatorio vengono sottoposti alla commissione 3 segni zodiacali, uno corretto, gli altri due errati. Se la commissione ottenesse almeno 50 successi su 100 prove, si potrebbe ragionevolmente attribuire all'astrologia una qualche validità scientifica?

Risposta. Assumiamo come ipotesi nulla la pura casualità. In questo caso la probabilità che la commissione ha di dare la risposta giusta tirando a indovinare è pari a $1/3$ per ogni persona interrogata. I termini del problema sono allora:

- *funzione di densità: binomiale;*
- *numero di tentativi: $n = 100$;*
- *probabilità di successo in un tentativo: $p = 1/3$;*
- *numero di successi ottenuti: $x = 50$;*
- *media attesa della distribuzione: $\mu = np = 33.3$;*
- *deviazione standard: $\sigma = \sqrt{np(1-p)} = 4.7$.*

Dato che valgono le condizioni $np > 10$, $n(1-p) > 10$, possiamo utilizzare le probabilità gaussiane e trovare il livello di significatività del valore standard

$$t = \frac{50 - 33.3}{4.7} = 3.55 .$$

La probabilità di ottenere almeno 50 successi è calcolabile con la Tab. D.1 dell'Appendice D:

$$P\{X \geq 50\} = P\{T \geq 3.55\} = 0.5000 - 0.4998 \simeq 2 \cdot 10^{-4} ,$$

oppure con l'istruzione SCILAB $0.5-0.5\text{erf}(3.55/\text{sqrt}(2))=0.0001926$ spiegata a pagina 84. Possiamo pertanto rigettare l'ipotesi della casualità, con una probabilità di sbagliare di circa 2 su 10 000.*

Se si ottenesse una serie di risultati di questo genere, l'astrologia acquisterebbe dignità scientifica e potrebbe essere a buon titolo insegnata nelle scuole. Tuttavia, la realtà è ben diversa: nel 1985 la rivista Nature [Car85] ha riportato i risultati di questo test, compiuto da una commissione mista di scienziati e astrologi. La percentuale dei successi, in 120 prove, è stata del 34%. A tutt'oggi, tanto per l'astrologia come per molte altre pseudoscienze come la telepatia e la chiaroveggenza, non è mai stato pubblicato, su nessuna rivista accreditata presso la comunità scientifica internazionale, alcun risultato che si discosti significativamente dalle leggi del caso.

In altre parole, possiamo dire che in questo genere di esperimenti l'ipotesi della pura casualità non è mai stata falsificata, ovvero che astrologia, telepatia e chiaroveggenza non hanno validità scientifica

In questi ultimi esercizi abbiamo discusso se accettare o meno un valore sperimentale avendo adottato a priori un certo modello probabilistico. Questi argomenti, che sono oggetto della statistica, verranno discussi dettagliatamente più avanti, a partire dal cap. 6. Crediamo comunque che vi sia stato utile avere dato uno sguardo preliminare a questi interessanti problemi.

Problemi

3.1. Risolvere il problema 1.9 (valutare la probabilità $P\{X \leq Y\}$ per due variabili uniformi $0 \leq X, Y \leq 1$) senza ricorrere ad argomenti geometrici.

3.2. Passeggiata a caso: una particella procede per passi e ad ogni passo può stare ferma $x_0 = 0$ oppure effettuare una deviazione di $x_1 = +1$, con uguale probabilità. Calcolare, dopo 500 passi, il valor medio e la deviazione standard del percorso X .

3.3. A differenza del problema precedente, questa volta la particella sceglie ad ogni passo, con uguale probabilità $p = 1/2$, le deviazioni $x_0 = -1$ e $x_1 = +1$.

3.4. La probabilità di trasmettere un bit sbagliato è 10^{-3} . Calcolare la probabilità che a) in un numero di 16 bit vi sia un bit sbagliato e b) il valor medio o atteso dei bit sbagliati.

3.5. La probabilità che in una certa stagione una persona abbia il virus dell'influenza è del 20%. Trovare la probabilità che in una sala di 200 persone i portatori dell'influenza siano tra 30 e 50.

3.6. Trovare, per $n \rightarrow \infty$, la funzione di densità della variabile $Y = X_1 X_2 \dots X_n$, dove le X_i sono variabili aleatorie positive che soddisfano le condizioni del teorema Limite Centrale.

3.7. A volte, per caratterizzare la larghezza di una funzione di densità, si usa la “semi-larghezza a metà altezza”, detta anche FWHM (dall'inglese *Full Width at Half Maximum*), definita come $\text{FWHM} = |x_2 - x_1|$, dove $p(x_1) = p(x_2) = p_{\max}/2$. Trovare la relazione tra FWHM e σ per una gaussiana.

3.8. In un ospedale si registra in media un parto gemellare ogni tre mesi. Supponendo una distribuzione esponenziale negativa, determinare: a) la probabilità di non avere parti gemellari per almeno 8 mesi; b) la probabilità di non avere parti gemellari entro un mese se sono passati 8 mesi senza parti gemellari.

3.9. La somma dei quadrati di 10 variabili standard vale 7. Trovare la probabilità di sbagliare affermando che le variabili non sono gaussiane ed indipendenti.

3.10. I dati storici fanno registrare 8500 morti l'anno per incidenti stradali. L'anno successivo all'introduzione delle cinture di sicurezza i morti calano a 8100. Considerando il dato storico come valore medio vero e il dato annuale come variabile aleatoria, valutare se le cinture di sicurezza hanno fatto diminuire il numero dei morti in modo significativo, ad un livello dell'1%.

3.11. Se la frequenza media di eventi poissoniani è di 100 al secondo, calcolare la frazione di intervalli di tempo tra due eventi minori di 1 millisecondo.

3.12. Una certa quantità di sostanza radioattiva emette una particella ogni due secondi. Eseguendo un test su un campione, non si registrano conteggi per 10 secondi e si conclude che la sostanza è assente. Qual è la probabilità che la conclusione sia errata?

3.13. Dimostrare che, se vale la legge di Poisson (3.47), i conteggi in intervalli di tempo disgiunti sono variabili aleatorie indipendenti.

3.14. Trovare l'intervallo, centrato sulla media, che contiene con probabilità del 50% i valori di una variabile gaussiana standard.

3.15. Trovare media e deviazione standard di una gaussiana sapendo che la probabilità in una prova di ottenere valori maggiori di 4.41 vale il 21% e quella di ottenere valori maggiori di 6.66 è il 6%.

3.16. Un fenomeno poissoniano fa registrare 100 eventi in 5 giorni. Calcolare quanti giorni in media devono trascorrere prima di registrare 4 eventi in un'ora.

3.17. Utilizzando il teorema 3.5, trovare un algoritmo di generazione casuale di variabili aleatorie dalla densità $p(x) = 2x - 2, 1 \leq x \leq 2$.

3.18. Un campione di resistenze ha un valore medio di 100Ω , una deviazione standard di 5Ω e presenta una distribuzione gaussiana.
 a) Qual è la probabilità che un valore di resistenza si discosti di più del 10% dal valore atteso? b) Qual è la probabilità che 10 resistenze in serie abbiano un valore $\geq 1050 \Omega$? b) Quale limite si può ricavare, per il caso a), abbandonando l'ipotesi gaussiana?

3.19. In un processo raro un contatore può registrare da 0 a 3 conteggi, con le seguenti probabilità:

X	0	1	2	3
probabilità	0.1	0.4	0.4	0.1

Considerando questa come distribuzione vera, e utilizzando l'approssimazione gaussiana, calcolare la probabilità che in un mese (30 giorni) venga registrato un numero totale di conteggi maggiore di 80, supponendo conteggi giornalieri indipendenti.

3.20. Una ditta di componenti elettronici produce un particolare componente con una percentuale del 5% di pezzi difettosi. Avendo venduto numerosi lotti da 200 pezzi dichiarando al massimo 15 pezzi difettosi, trovare la percentuale di lotti che non rispettano la specifica di vendita.

Calcolo delle probabilità per più variabili

Qui, in effetti, la geometria dei picchi e delle valli è a n dimensioni ... ed è intrattabile.

David Ruelle, *Caso e caos*

4.1 Introduzione

In questo capitolo affronteremo argomenti un po' più complessi di quelli del capitolo precedente. Questa fatica sarà però ripagata dai risultati che otterremo, che saranno indispensabili per capire e trattare i problemi che coinvolgono più variabili aleatorie.

Per non complicare inutilmente il formalismo matematico, i problemi saranno discussi inizialmente per il caso di due variabili aleatorie. Successivamente, poiché l'ipotesi di due sole variabili non entrerà mai nelle dimostrazioni dei teoremi e nella discussione, i risultati saranno facilmente estesi ad un numero qualunque di variabili.

I risultati saranno inoltre prevalentemente ottenuti in forma integrale, considerando variabili continue. Il passaggio al caso di variabili discrete è immediatamente ottenuto con trasformazioni del tipo

$$\int_a^b (\dots)p(x) dx \rightarrow \sum_{x=a}^b (\dots)p(x) \quad (4.1)$$

dove (...) indica una qualunque espressione contenente variabili aleatorie e parametri. La (4.1) è immediatamente estendibile anche al caso di più variabili.

4.2 Distribuzioni statistiche multidimensionali

Se X e Y sono due variabili casuali definite sullo stesso spazio di probabilità, definiamo come A l'evento in cui un valore x è compreso tra due valori x_1 e

$x_2 \{x_1 \leq X \leq x_2\}$ e analogamente l'evento B come $\{y_1 \leq Y \leq y_2\}$. Scriviamo la probabilità composta $P(AB)$ come:

$$P(AB) \equiv P\{x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\}, \quad (4.2)$$

che è la probabilità che il valore di (x, y) cada in $[x_1, x_2] \times [y_1, y_2]$. Se esiste una funzione $p(x, y) \geq 0$ tale che

$$P\{x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\} = \int_{x_1}^{x_2} \int_{y_1}^{y_2} p(x, y) \, dx \, dy \quad (4.3)$$

si dice che $p(x, y)$ è la *densità congiunta* di probabilità delle due variabili. Essa è l'estensione bidimensionale della funzione di densità definita nel par. 2.7.

Più in generale, se $A \in \mathbb{R}^2$, si può dimostrare che, se $p(x, y)$ soddisfa la (4.3), la probabilità che $(X, Y) \in A$ (ad esempio $x + y \leq a$, con a costante) è data da [Bal98]:

$$P\{(X, Y) \in A\} = \int_A p(x, y) \, dx \, dy, \quad (4.4)$$

la quale diventa l'analogo del generico integrale (3.94) e della funzione cumulativa o di ripartizione (2.34) quando

$$P\{(X, Y) \in A\} = P\{-\infty \leq X \leq a, -\infty \leq Y \leq b\}.$$

La condizione di normalizzazione, le medie e le varianze delle variabili coinvolte si ottengono come immediata generalizzazione delle corrispondenti formule monodimensionali:

$$\begin{aligned} \int \int p(x, y) \, dx \, dy &= 1, \\ \langle X \rangle &= \int \int x p(x, y) \, dx \, dy = \mu_x, \\ \langle Y \rangle &= \int \int y p(x, y) \, dx \, dy = \mu_y, \\ \text{Var}[X] &= \int \int (x - \mu_x)^2 p(x, y) \, dx \, dy = \sigma_x^2, \\ \text{Var}[Y] &= \int \int (y - \mu_y)^2 p(x, y) \, dx \, dy = \sigma_y^2, \end{aligned} \quad (4.5)$$

dove l'integrazione si intende estesa in $(-\infty, +\infty)$, ovvero in tutto l'intervallo di esistenza della funzione di densità. Come nel caso monodimensionale (si veda la def. 2.10), per l'esistenza dei valori medi si richiede l'assoluta integrabilità o sommabilità. Nel seguito l'integrazione multipla verrà indicata semplicemente col simbolo di integrale singolo.

Se X e Y sono stocasticamente indipendenti secondo la (2.10), il teorema delle probabilità composte e le (1.20, 1.23, 2.10) permettono di scrivere, per ogni coppia (x_1, x_2) e (y_1, y_2) :

$$P\{x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\} = P\{x_1 \leq X \leq x_2\} P\{y_1 \leq Y \leq y_2\},$$

cioè:

$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} p(x, y) \, dx \, dy = \int_{x_1}^{x_2} p_X(x) \, dx \int_{y_1}^{y_2} p_Y(y) \, dy$$

e quindi di definire la densità congiunta come:

$$p(x, y) = p_X(x) p_Y(y) \quad (\text{se } X \text{ e } Y \text{ sono indipendenti}) , \quad (4.6)$$

dove $p_X(x)$ e $p_Y(y)$ sono le densità di probabilità delle variabili X e Y .

È possibile anche definire medie e varianze di combinazioni di variabili. Ad esempio:

$$\begin{aligned} \langle XY \rangle &= \int xy p(x, y) dx dy = \mu_{xy} , \\ \langle X + Y \rangle &= \int (x + y) p(x, y) dx dy = \mu_{x+y} , \\ \text{Var}[XY] &= \int (xy - \mu_{x \cdot y})^2 p(x, y) dx dy , \\ \text{Var}[X + Y] &= \int (x + y - \mu_{x+y})^2 p(x, y) dx dy . \end{aligned} \quad (4.7)$$

È facile mostrare, dalla seconda e dalla terza delle (4.5) e dalla seconda delle (4.7), che

$$\langle X + Y \rangle = \langle X \rangle + \langle Y \rangle \quad (4.8)$$

e che, se X e Y sono stocasticamente indipendenti, dalla prima delle (4.7) e dalla (4.6) segue:

$$\langle XY \rangle = \langle X \rangle \langle Y \rangle . \quad (4.9)$$

Fin qui non sembra esserci qualcosa di nuovo, se non la ovvia generalizzazione delle formule monodimensionali. Invece, le distribuzioni multidimensionali riservano subito delle sorprese, perché è possibile la definizione di almeno tre nuove importanti quantità, la *densità marginale*, la *densità condizionata* e la *covarianza* tra le variabili.

Definizione 4.1 (Densità marginale). *Se X e Y sono due variabili aleatorie con densità $p(x, y)$ ed A è un intervallo dell'asse reale, la densità marginale $p_X(x)$ è definita dalla relazione*

$$P\{X \in A\} \equiv \int_A p_X(x) dx = \int_A dx \int_{-\infty}^{+\infty} p(x, y) dy , \quad (4.10)$$

da cui:

$$p_X(x) = \int_{-\infty}^{+\infty} p(x, y) dy . \quad (4.11)$$

La densità marginale $p_Y(y)$ di y si ottiene dalla formula precedente scambiando x con y .

È facile vedere che le densità marginali sono normalizzate:

$$\int p_X(x) dx = \int p(x, y) dx dy = 1 .$$

Le densità marginali danno la probabilità di eventi del tipo $\{X \in A\}$ per ogni valore di Y (e viceversa); esse rappresentano pertanto le densità di probabilità monodimensionali delle variabili X e Y .

Poiché spesso è importante stabilire se due o più variabili sono indipendenti, risulta molto utile il seguente:

Teorema 4.1 (Indipendenza di variabili). *Due variabili aleatorie (X, Y) , di densità congiunta $p(x, y)$, sono stocasticamente indipendenti se e solo se esistono due funzioni $g(x)$ ed $h(y)$ tali che, per ogni $x, y \in \mathbb{R}$, si abbia:*

$$p(x, y) = g(x) h(y) . \quad (4.12)$$

Dimostrazione. Se le variabili sono indipendenti, la (4.6) dimostra la prima parte del teorema con $g(x) = p_X(x)$ e $h(y) = p_Y(y)$. Nel caso invece valga la (4.12), in generale si avrà:

$$\int_{-\infty}^{+\infty} g(x) dx = G , \quad \int_{-\infty}^{+\infty} h(y) dy = H .$$

Dato che $p(x, y)$ è una densità normalizzata, dalla prima delle (4.5) risulta (sottointendendo i limiti di integrazione):

$$HG = \int g(x) dx \int h(y) dy = \int p(x, y) dx dy = 1 .$$

È possibile allora definire le due densità marginali normalizzate:

$$p_X(x) = \int g(x)h(y) dy = Hg(x) , \quad p_Y(y) = \int g(x)h(y) dx = Gh(y) ,$$

da cui, dato che $HG = 1$, si ottiene la (4.6):

$$p(x, y) = g(x) h(y) = g(x) h(y)HG = p_X(x) p_Y(y) . \quad \square$$

In corrispondenza di un valore fissato x_0 la funzione $p(x_0, y)$ deve rappresentare la densità di probabilità monodimensionale della variabile Y . Tuttavia, poiché $p(x_0, y)$ non è normalizzata, tenendo presente la (4.11), si può definire la:

Definizione 4.2 (Densità condizionata). *Se X e Y sono due variabili aleatorie con densità $p(x, y)$, la densità condizionata $p(y|x_0)$ di y per ogni $x = x_0$ fissato e tale che $p_X(x_0) > 0$, è data da:*

$$p(y|x_0) = \frac{p(x_0, y)}{p_X(x_0)} = \frac{p(x_0, y)}{\int_{-\infty}^{+\infty} p(x_0, y) dy} . \quad (4.13)$$

Anche in questo caso la densità condizionata $p(x|y)$ di x rispetto ad y si ottiene scambiando le due variabili nella formula precedente.

Le densità condizionate così definite sono normalizzate. Infatti:

$$\int p(y|x_0) dy = \frac{\int p(x_0, y) dy}{\int p(x_0, y) dy} = 1 .$$

È importante notare che la densità condizionata $p(y|x)$ è funzione della sola y , essendo x un valore fissato (parametro). Lo stesso vale per $p(x|y)$ scambiando le due variabili. È quindi sbagliato scrivere:

$$P\{Y \in B|X \in A\} = \int_A \int_B p(y|x) dx dy = \int_A \int_B \frac{p(x, y)}{p_X(x)} dx dy \quad (\text{errato!}) ,$$

in quanto $p(y|x)$ è funzione solo di y . Per trovare la scrittura corretta occorre rifarsi alla definizione di probabilità condizionata data dalla (1.19):

$$\begin{aligned} P\{Y \in B|X \in A\} &= \frac{P\{X \in A, Y \in B\}}{P\{X \in A\}} \\ &= \frac{\int_A \int_B p(x, y) dx dy}{\int_A dx \int_{-\infty}^{+\infty} p(x, y) dy} . \end{aligned}$$

Sempre per un determinato valore fissato a priori, è possibile definire gli operatori media e varianza condizionati. Ad esempio, il valore atteso di Y condizionato da un valore x è dato da:

$$\langle Y|x \rangle = \int_{-\infty}^{+\infty} y p(y|x) dy = \frac{\int_{-\infty}^{+\infty} y p(x, y) dy}{p_X(x)} , \quad (4.14)$$

dove $p_X(x) = \int p(x, y) dy$ è un numero essendo x fissato.

Le densità marginale e condizionata discendono dal teorema 1.2 delle probabilità composte. Infatti, invertendo la (4.13), possiamo scrivere:

$$p(x, y) = p_Y(y) p(x|y) = p_X(x) p(y|x) , \quad (4.15)$$

che corrisponde alla formula delle probabilità composte (1.20) per variabili continue: la densità di (X, Y) è data dalla densità in Y per la densità di X per ogni Y fissato (o viceversa).

Quando le variabili sono indipendenti, dalle (4.6, 4.13) si ottiene:

$$p(x|y) = p_X(x) , \quad p(y|x) = p_Y(y) , \quad (X \text{ ed } Y \text{ indipendenti}). \quad (4.16)$$

La differenza tra densità marginale e condizionata può essere ben compresa con l'aiuto di Fig. 4.1; la densità marginale si ottiene proiettando semplicemente la curva in una dimensione, mentre la densità condizionata è la proiezione della curva sull'asse y (cioè la densità probabilità di y) per un determinato valore x (o viceversa).

Notiamo anche che le medie e le varianze definite nelle (4.5) possono essere espresse utilizzando le densità marginali:

$$\mu_x = \int x p_X(x) dx , \quad \mu_y = \int y p_Y(y) dy . \quad (4.17)$$

$$\sigma_x^2 = \int (x - \mu_x)^2 p_X(x) dx , \quad \sigma_y^2 = \int (y - \mu_y)^2 p_Y(y) dy . \quad (4.18)$$

Occupiamoci ora della quarta delle (4.7); poiché dalla (4.8) risulta che la media della somma è pari alla somma delle medie, otteniamo:

$$\begin{aligned} \text{Var}[X + Y] &= \int [(x - \mu_x) + (y - \mu_y)]^2 p(x, y) dx dy \\ &= \text{Var}[X] + \text{Var}[Y] + 2 \text{Cov}[X, Y] , \end{aligned} \quad (4.19)$$

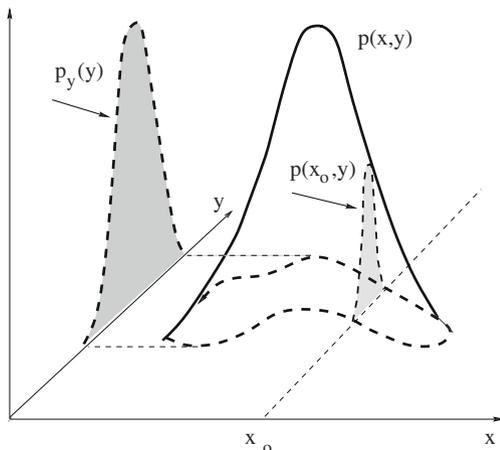


Fig. 4.1. La densità marginale $p_Y(y)$ di y rappresenta la proiezione sull'asse y della densità dei punti (x, y) per qualunque valore di x . La funzione $p(x_0, y)$ rappresenta invece la proiezione sull'asse y della densità dei punti in corrispondenza di un determinato valore x_0 (linea tratteggiata). Questa funzione, normalizzata con la (4.13), rappresenta la densità condizionata della variabile Y in corrispondenza del valore x_0

dove

$$\text{Cov}[X, Y] \equiv \int \int (x - \mu_x)(y - \mu_y)p(x, y) dx dy .$$

Introduciamo pertanto la:

Definizione 4.3 (Covarianza di due variabili). *La covarianza di due variabili aleatorie X e Y è definita come*

$$\text{Cov}[X, Y] = \int \int (x - \mu_x)(y - \mu_y) p(x, y) dx dy \equiv \sigma_{xy} . \quad (4.20)$$

Questa grandezza è una sorta di definizione “incrociata” rispetto alle due varianze possibili σ_x^2 e σ_y^2 definite nelle (4.5). Essa è una grandezza quadratica, avente come dimensione il prodotto delle dimensioni di X e Y . A titolo di esercizio, trascriviamo la definizione generale (4.20) per i vari casi possibili, analogamente a quanto fatto per la varianza nelle (2.38, 2.41, 2.66). Quando si ha una densità discreta p_{XY} la covarianza diventa:

$$\text{Cov}[X, Y] = \sum_x \sum_y (x - \mu_x)(y - \mu_y) p_{XY}(x, y) , \quad (4.21)$$

dove la somma si intende estesa a tutti i valori dello spettro delle due variabili. Il calcolo diretto da un insieme di dati va invece eseguito sulla somma di tutte

le realizzazioni numeriche delle variabili:

$$s_{xy} = \frac{1}{N} \sum_i (x_i - \mu_x)(y_i - \mu_y) , \quad (4.22)$$

$$s_{xy} = \frac{1}{N-1} \sum_i (x_i - m_x)(y_i - m_y) . \quad (4.23)$$

Qui la somma va fatta solo su un indice, cioè su tutte le coppie (x_i, y_i) che si sono ottenute, dividendo per $N - 1$ quando si calcolano gli scarti dalle medie campionarie e non dalle medie “vere” assegnate a priori. L’insieme di queste coppie esaurisce, per $N \rightarrow \infty$, l’insieme dato dalle coppie (x_i, y_k) quando gli indici i e k rappresentano i valori dello spettro delle due variabili. Questo è un punto formale ma importante che va tenuto presente per eseguire calcoli corretti: *quando la covarianza è calcolata attraverso la funzione di densità, la sommatoria è doppia e va fatta sulle probabilità delle coppie dei valori possibili delle due variabili; quando la covarianza è calcolata da un insieme sperimentale di dati, la somma è singola e va fatta su tutte le coppie ottenute nel campionamento.*

Il software SCILAB fornisce la funzione `covar(x,y, fre)` che calcola la covarianza dalla (4.21), dove \mathbf{x} e \mathbf{y} sono due vettori e \mathbf{fre} è la matrice p_{ij} delle probabilità o frequenze di tutte le coppie possibili (x_i, y_j) . Per calcolare la covarianza da un insieme di dati (x_i, y_i) con le (4.22, 4.23), abbiamo scritto la routine `Corcov(x,y)`, dove \mathbf{x} e \mathbf{y} sono i vettori dei valori misurati. Il risultato, come descritto nei commenti della funzione, è un vettore contenente la covarianza ed il coefficiente di correlazione con il relativo errore statistico, che verrà introdotto più avanti, nel cap. 6.

Per finire, notiamo che in notazione operatoriale abbiamo:

$$\text{Cov}[X, Y] = \langle (X - \mu_x)(Y - \mu_y) \rangle . \quad (4.24)$$

Vale anche la relazione notevole, analoga alla (2.50):

$$\begin{aligned} \text{Cov}[X, Y] &= \langle (X - \mu_x)(Y - \mu_y) \rangle = \langle XY - \mu_x Y - \mu_y X + \mu_x \mu_y \rangle \\ &= \langle XY \rangle - \mu_x \langle Y \rangle - \mu_y \langle X \rangle + \mu_x \mu_y \\ &= \langle XY \rangle - \mu_x \mu_y , \end{aligned} \quad (4.25)$$

dalla quale si vede che la covarianza è data dalla media dei prodotti meno il prodotto delle medie. La covarianza gode di proprietà particolarmente interessanti che verranno discusse nel prossimo paragrafo.

Esercizio 4.1

Un componente elettronico ha un tempo di vita esponenziale di media

$$\mu = 1/\lambda = 1000 \text{ ore}$$

(si veda la 3.49). Viene realizzato un dispositivo costituito dai due elementi in parallelo, il che significa che esso funziona finché almeno uno dei due dispositivi è funzionante. Quanto è maggiore il tempo di vita medio del dispositivo in

parallelo rispetto a quello del singolo dispositivo? Dopo un tempo pari a 2000 ore, qual è la probabilità che il dispositivo singolo e quello in parallelo siano ancora funzionanti?

Risposta. Se t_1 e t_2 sono i tempi di guasto dei due dispositivi, il parallelo cesserà di funzionare ad un tempo t tale che $t = \max(t_1, t_2)$. La probabilità di funzionamento fino a t , cioè di un guasto al tempo t , sarà data dalla probabilità composta che il primo dispositivo e il secondo cessino di funzionare. Dato che i due dispositivi sono indipendenti, basta allora integrare in $[0, t]$ la densità congiunta dei tempi (4.6), come previsto dalle (4.4, 4.12):

$$P\{\max(T_1, T_2) \leq t\} = P\{T_1 \leq t, T_2 \leq t\} = \int_0^t e_1(t) dt \int_0^t e_2(t) dt$$

poiché le due distribuzioni esponenziali e_1 ed e_2 sono uguali, dalla (3.51) si ha immediatamente:

$$P\{\max(T_1, T_2) \leq t\} \equiv P(t) = (1 - e^{-\lambda t})^2,$$

la quale rappresenta la funzione di ripartizione monodimensionale dei tempi di guasto del dispositivo in parallelo. Derivando la funzione di ripartizione otterremo, in base alla (2.36), la densità di probabilità $e_p(t)$ corrispondente. Si ha pertanto:

$$e_p(t) = 2\lambda e^{-\lambda t} (1 - e^{-\lambda t}),$$

la quale non è una esponenziale semplice. La media dei tempi di guasto:

$$\langle T \rangle = 2\lambda \int_0^\infty t (e^{-\lambda t} - e^{-2\lambda t}) dt = \frac{3}{2} \frac{1}{\lambda} = \frac{3}{2} \mu = 1500 \text{ ore},$$

corrisponde ad un incremento nella vita media del 50%.

Veniamo alla seconda parte del problema. La probabilità di funzionamento dopo un tempo $t_0 = 2000$ ore è data, per il dispositivo singolo, dalle (3.51, 3.52):

$$P\{T \geq t_0\} = 1 - \int_0^{t_0} \lambda e^{-\lambda \tau} d\tau = e^{-\lambda t_0} = e^{-2} = 0.135 = 13.5\%.$$

La stessa probabilità, per il parallelo, è data da:

$$\begin{aligned} P_p\{T \geq t_0\} &= 1 - \int_0^{t_0} e_p(t) dt = 1 - 2\lambda \int_0^{t_0} (e^{-\lambda t} - e^{-2\lambda t}) dt \\ &= 2e^{-\lambda t_0} - e^{-2\lambda t_0} = 2e^{-2} - e^{-4} = 0.252 \simeq 25\%. \end{aligned}$$

Dopo 2000 ore l'aumento di affidabilità è quasi del 100%.

4.3 Covarianza e correlazione

La covarianza di due variabili aleatorie gode dell'importante proprietà di annullarsi quando le variabili sono indipendenti. Infatti in questo caso la densità $p(x, y)$ che compare nella (4.20) si può scrivere, in accordo con la (4.6), come il prodotto $p_X(x)p_Y(y)$. Si ottiene allora immediatamente, dato che $p_X(x)$ e $p_Y(y)$ sono normalizzate:

$$\begin{aligned} \text{Cov}[X, Y] &= \iint (x - \mu_x)(y - \mu_y) p(x, y) dx dy \\ &= \int (x - \mu_x) p_X(x) dx \int (y - \mu_y) p_Y(y) dy \\ &= \left[\int x p_X(x) dx - \mu_x \int p_X(x) dx \right] \left[\int y p_Y(y) dy - \mu_y \int p_Y(y) dy \right] \\ &= \mu_x - \mu_x + \mu_y - \mu_y = 0 . \end{aligned} \quad (4.26)$$

Segue immediatamente dalla (4.19) che, per variabili X_i tra loro indipendenti, la varianza della somma è pari alla somma delle varianze:

$$\text{Var} \left[\sum_i X_i \right] = \sum_i \text{Var}[X_i] . \quad (4.27)$$

La covarianza si presta quindi ad essere un indicatore statistico della correlazione: tanto più essa è diversa da zero tanto più le variabili sono correlate. Come valutare però il grado massimo di correlazione? Come renderlo indipendente dall'insieme di dati che si sta esaminando? Questa difficoltà si può superare grazie al:

Teorema 4.2 (di Cauchy-Schwarz). *Se le variabili X e Y hanno varianze finite, allora vale la disuguaglianza:*

$$|\text{Cov}[X, Y]| \leq \sigma[X] \sigma[Y] . \quad (4.28)$$

Dimostrazione. La (4.28) può essere scritta come:

$$|\langle (X - \mu_x)(Y - \mu_y) \rangle| \leq \sqrt{\langle (X - \mu_x)^2 \rangle \langle (Y - \mu_y)^2 \rangle} . \quad (4.29)$$

Se definiamo le variabili centrate $X_0 = X - \mu_x$ e $Y_0 = Y - \mu_y$, la (4.29) diventa:

$$\langle X_0 Y_0 \rangle^2 \leq \langle X_0^2 \rangle \langle Y_0^2 \rangle . \quad (4.30)$$

Indichiamo con t un numero reale qualunque e consideriamo la variabile $(tX_0 - Y_0)$. Sfruttando le proprietà di linearità della media otteniamo facilmente:

$$0 \leq \langle (tX_0 - Y_0)^2 \rangle = t^2 \langle X_0^2 \rangle - 2t \langle X_0 Y_0 \rangle + \langle Y_0^2 \rangle .$$

Questo polinomio di secondo grado in t può essere sempre ≥ 0 se e solo se il discriminante rimane ≤ 0 . Si ha pertanto:

$$4 \langle X_0 Y_0 \rangle^2 - 4 \langle X_0^2 \rangle \langle Y_0^2 \rangle \leq 0 ,$$

che è proprio la (4.30). La (4.28) risulta di conseguenza verificata.

L'uguaglianza in questa equazione si ha per

$$\langle (tX_0 - Y_0)^2 \rangle = 0 \implies Y_0 = tX_0 \implies Y = tX - t\mu_x + \mu_y \equiv aX + b ,$$

cioè quando tra X e Y sono legate da una equazione lineare. \square

La definizione di covarianza e il teorema di Cauchy-Schwarz portano spontaneamente a definire il coefficiente di correlazione tra due variabili come:

$$\rho_{xy} \equiv \rho[X, Y] = \frac{\text{Cov}[X, Y]}{\sigma[X] \sigma[Y]} . \quad (4.31)$$

Questo coefficiente gode della proprietà

$$-1 \leq \rho_{xy} \leq 1 , \quad (4.32)$$

è nullo se le variabili sono indipendenti, è positivo se le variabili sono correlate e quando una di esse aumenta anche l'altra tende ad aumentare, è negativo (anticorrelazione) se la crescita di una variabile tende a far diminuire l'altra. Infine, i limiti $\rho = \pm 1$ si verificano quando esiste tra le due variabili una relazione lineare del tipo $Y = aX + b$. Siamo in questo caso in presenza in pratica di una sola variabile aleatoria, scritta in due modi matematicamente diversi.

Dobbiamo discutere ora un punto molto delicato: la connessione tra indipendenza statistica, correlazione e causalità. L'indipendenza statistica, in accordo con la def. 1.10, si ha quando la probabilità dell'evento può essere fattorizzata nel prodotto delle singole probabilità di ogni variabile, come nella (2.10) per variabili discrete e nella (4.6) per quelle continue. Per la correlazione, occorre fare uso della seguente:

Definizione 4.4 (Correlazione tra variabili). *Due variabili aleatorie si dicono incorrelate quando il loro coefficiente di correlazione è nullo, si dicono correlate in caso contrario.*

Evidentemente, la definizione dipende dal coefficiente di correlazione che viene usato nella analisi dei dati. Per il momento, usiamo il coefficiente di correlazione definito nella (4.31); nel seguito, nella (10.10), daremo la definizione più generale di questo coefficiente. Notiamo anche che qui il significato di correlazione è molto tecnico e non coincide necessariamente con il senso che il linguaggio comune assegna a questo termine.

Fatte queste premesse, analizziamo in dettaglio il problema. Notiamo subito che, se le variabili sono correlate (cioè il coefficiente ρ_{xy} della (4.31) non è nullo), allora tra di esse deve esistere un certo grado di dipendenza. Tuttavia, l'inverso non è vero. Per dimostrare questo fatto, basta un controesempio: consideriamo le variabili $X = U + V$ e $Y = U - V$, dove U e V sono variabili uniformi in $[0, 1]$; X e Y sono dipendenti perchè non vale la (4.12), ma hanno coefficiente di correlazione nullo perchè è nulla la covarianza. Infatti, dato che

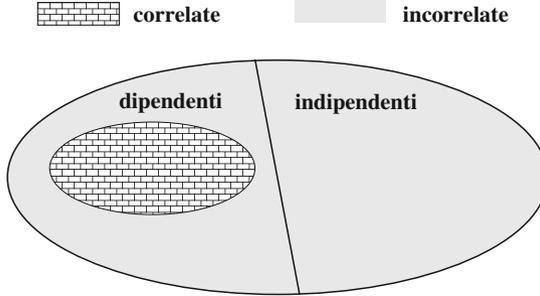


Fig. 4.2. Legame tra correlazione e dipendenza statistica. Per variabili gaussiane gli insiemi delle variabili correlate e dipendenti e quelli delle variabili indipendenti e incorrelate coincidono

$\langle U + V \rangle = 1$ e $\langle U - V \rangle = 0$:

$$\begin{aligned} \text{Cov}[X, Y] &= \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle \\ &= \langle (U + V - \langle U + V \rangle)(U - V - \langle U - V \rangle) \rangle \\ &= \langle (U + V - 1)(U - V) \rangle = \langle U^2 - V^2 - U + V \rangle = \\ &= \langle U^2 \rangle - \langle V^2 \rangle - \langle U \rangle + \langle V \rangle = 0, \end{aligned}$$

poiché U e V hanno la stessa distribuzione.

Un altro punto chiave, dimostrato nella (4.26), è che, se due variabili sono statisticamente indipendenti, allora esse sono anche incorrelate.

Possiamo riassumere la discussione nei punti seguenti:

- condizione sufficiente perché vi sia dipendenza statistica è la presenza di correlazione;
- condizione necessaria per l'indipendenza statistica è la mancanza di correlazione ($\rho_{xy} = 0$).

Queste proprietà sono riassunte in Fig. 4.2.

Esercizio 4.2

Se T, U e $V \sim U(0, 1)$, calcolare il coefficiente di correlazione lineare tra la variabile:

$$X = T$$

e le variabili:

$$Y = U, \quad Y_1 = X + V, \quad Y_2 = -X + V.$$

Fare inoltre una verifica con dati simulati al calcolatore.

Risposta. Tenendo presente le (3.82), per una generica variabile uniforme U valgono le relazioni:

$$\langle U \rangle = 1/2, \quad \langle U^2 \rangle = 1/3, \quad \langle (U - 1/2)^2 \rangle = 1/12,$$

mentre per una coppia (U, V) di variabili uniformi indipendenti si ha:

$$\langle UV \rangle = \int r u(r) dr \int s v(s) ds = 1/4 .$$

Dalle (4.24, 4.31) otteniamo facilmente:

$$\rho[X, Y] = \frac{\langle (T - 1/2)(U - 1/2) \rangle}{[\langle (T - 1/2)^2 \rangle \langle (U - 1/2)^2 \rangle]^{1/2}} = 0 ,$$

$$\rho[X, Y_1] = \frac{\langle (T - 1/2)(T + V - 1) \rangle}{[\langle (T - 1/2)^2 \rangle \langle (T + V - 1)^2 \rangle]^{1/2}} = \frac{1}{\sqrt{2}} \simeq 0.7071 ,$$

$$\rho[X, Y_2] = \frac{\langle (T - 1/2)(-T + V) \rangle}{[\langle (T - 1/2)^2 \rangle \langle (-T + V)^2 \rangle]^{1/2}} = -\frac{1}{\sqrt{2}} \simeq -0.7071 .$$

I diagrammi di dispersione delle tre coppie di variabili sono riportati in Fig. 4.3, dove è possibile vedere chiaramente la differenza tra due variabili incorrelate (X, Y) , due variabili con correlazione positiva (X, Y_1) e due variabili con correlazione negativa (X, Y_2) .

Si può verificare l'esercizio con una simulazione utilizzando SCILAB e la nostra funzione `Corcov`. Per generare $N = 20000$ variabili uniformi x , y , y_1 , y_2 e ottenerne la correlazione basta porre:

```
x=rand(1,20000); y=rand(1,20000); v=rand(1,20000);
y1= x + v;
y2= -x + v;
Corcov(x,y); // vengono mostrati i dati ...
Corcov(x,y1);
Corcov(x,y2);
```

Il coefficiente di correlazione $\rho[X, Y]$ (e analogamente i coefficienti $\rho[X, Y_1]$ e $\rho[X, Y_2]$) sono stimati dalla routine con le formule;

$$m_x = \frac{1}{N} \sum_i x_i , \quad m_y = \frac{1}{N} \sum_i y_i ,$$

$$s_x = \left[\frac{1}{N-1} \sum_i (x_i - m_x)^2 \right]^{1/2} ,$$

$$s_y = \left[\frac{1}{N-1} \sum_i (y_i - m_y)^2 \right]^{1/2} ,$$

$$s_{xy} = \frac{1}{N-1} \sum_i (x_i - m_x)(y_i - m_y) ,$$

$$r(x, y) = \frac{s_{xy}}{s_x s_y} = \frac{\sum_i (x_i - m_x)(y_i - m_y)}{[\sum_i (y_i - m_y)^2 \sum_i (x_i - m_x)^2]^{1/2}} ,$$

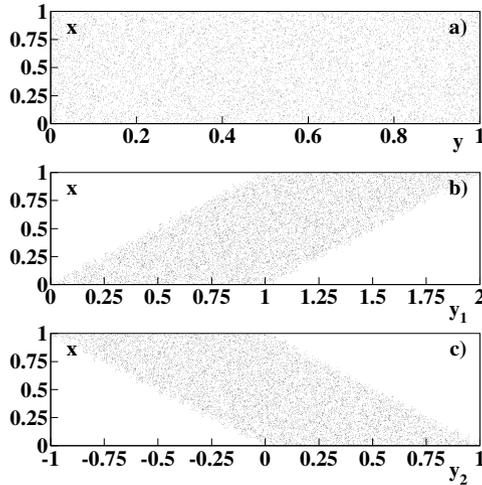


Fig. 4.3. Diagrammi di dispersione delle coppie di variabili considerate nell'esercizio 4.2: assenza di correlazione (a), correlazione positiva (b) e correlazione negativa (c)

dove si sono usati i simboli latini perché si tratta di valori campionari (si veda la def. 2.8). La covarianza viene calcolata dai dati con una sommatoria unica, utilizzando la (4.22). Inoltre, si usa il fattore $N-1$ al posto di N perché le varianze e covarianze sono state calcolate rispetto alle medie empiriche e non rispetto a quelle vere. Questa misteriosa correzione, già introdotta nella (2.39), verrà finalmente spiegata nel cap. 6.

Dalla simulazione abbiamo ottenuto i valori:

$$r(x, y) = 0.007, \quad r(x, y_1) = 0.712, \quad r(x, y_2) = -0.705.$$

Imparerete a decidere se la differenza tra questi valori “sperimentali” e quelli teorici $0, 0.7071, -0.7071$ è significativa o meno nel cap. 6, quando vi introdurremo alla statistica.

La routine `Corcov` fornisce anche gli errori sulle stime della covarianza e della correlazione, che saranno discussi più avanti, nel cap. 6.

4.4 Densità gaussiana bidimensionale

Anche nel caso multidimensionale moltissimi problemi possono essere trattati in modo esatto o approssimato supponendo di avere a che fare con variabili normali o gaussiane. Risulta pertanto quasi indispensabile studiare a questo punto la densità gaussiana bidimensionale.

Per prima cosa, notiamo subito che la densità $g(x, y)$ di due variabili gaussiane indipendenti sarà data, in base alle (3.28, 4.6), dalla semplice espressione:

$$g(x, y) = g_X(x) g_Y(y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[-\frac{1}{2} \left(\frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y - \mu_y)^2}{\sigma_y^2} \right) \right]. \quad (4.33)$$

Operando la sostituzione

$$U = \frac{X - \mu_x}{\sigma_x}, \quad V = \frac{Y - \mu_y}{\sigma_y} \quad (4.34)$$

l'analogo bidimensionale della densità gaussiana standard (3.42) sarà dato allora da:

$$g(u, v; 0, 1) = \frac{1}{2\pi} \exp \left[-\frac{1}{2}(u^2 + v^2) \right]. \quad (4.35)$$

La densità generale che vogliamo trovare deve però riferirsi a due variabili normali tra loro dipendenti e quindi, in generale, correlate con coefficiente di correlazione lineare $\rho_{uv} \equiv \rho \neq 0$. La densità standard corrispondente dovrà perciò soddisfare le proprietà:

- non essere fattorizzabile in due termini dipendenti rispettivamente solo da u e solo da v ;
- avere distribuzioni marginali gaussiane standard;
- soddisfare l'equazione $\sigma_{uv} = \rho$;
- soddisfare la condizione di normalizzazione.

Verifichiamo se esiste una forma funzionale del tipo

$$g(u, v; 0, 1) = \frac{1}{d} e^{-(au^2 + buv + cv^2)} \quad (4.36)$$

che soddisfa a tutte queste condizioni. Esse si traducono nel sistema di equazioni:

$$\begin{aligned} 1 &= \frac{1}{d} \int \int u^2 e^{-(au^2 + buv + cv^2)} du dv, \\ 1 &= \frac{1}{d} \int \int v^2 e^{-(au^2 + buv + cv^2)} du dv, \\ \rho &= \frac{1}{d} \int \int uv e^{-(au^2 + buv + cv^2)} du dv, \\ 1 &= \frac{1}{d} \int \int e^{-(au^2 + buv + cv^2)} du dv. \end{aligned}$$

Questi integrali possono essere risolti col metodo indicato nell'esercizio 3.4 e forniscono un sistema di 4 equazioni nelle 4 incognite a, b, c, d . Poiché il procedimento dettagliato, per quanto laborioso, non è particolarmente difficile né

istruttivo, qui ne riportiamo solo il risultato finale. Se volete, potete facilmente controllare a posteriori la correttezza del risultato, che, se $\rho \neq \pm 1$, risulta essere:

$$a = c = \frac{1}{2(1 - \rho^2)}, \quad b = -\frac{\rho}{(1 - \rho^2)}, \quad d = 2\pi\sqrt{1 - \rho^2}. \quad (4.37)$$

La (4.36) risulta allora:

$$g(u, v; 0, 1) = \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp \left[-\frac{1}{2(1 - \rho^2)}(u^2 - 2\rho uv + v^2) \right]. \quad (4.38)$$

È facile vedere che questa densità, come richiesto, dà luogo a due densità marginali gaussiane standard, *per ogni* $\rho \neq \pm 1$. Infatti, aggiungendo e sottraendo nell'esponente il termine $\rho^2 u^2$ ed integrando in y otteniamo:

$$\begin{aligned} g_u(u; 0, 1) &= \frac{e^{-\frac{1}{2}u^2}}{2\pi\sqrt{1 - \rho^2}} \int \exp \left[-\frac{(v - \rho u)^2}{2(1 - \rho^2)} \right] dv \\ &= \frac{e^{-\frac{1}{2}u^2}}{2\pi} \int e^{-\frac{1}{2}z^2} dz = \frac{e^{-\frac{1}{2}u^2}}{\sqrt{2\pi}}, \end{aligned} \quad (4.39)$$

dove si è posto $z = (v - \rho u)/\sqrt{1 - \rho^2}$, $dv = \sqrt{1 - \rho^2} dz$ e si è utilizzato l'integrale notevole (3.29) dell'esercizio 3.4. Lo stesso risultato si ottiene ovviamente per la densità marginale in v .

Tornando dalle variabili ridotte a quelle normali, dalla (4.38) otteniamo infine la densità:

$$\begin{aligned} g(x, y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho^2}} e^{-\frac{1}{2}\gamma(x, y)}, \quad (4.40) \\ \gamma(x, y) &= \frac{1}{1 - \rho^2} \left[\frac{(x - \mu_x)^2}{\sigma_x^2} - 2\rho \frac{(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y} + \frac{(y - \mu_y)^2}{\sigma_y^2} \right] \end{aligned}$$

che rappresenta la forma più generale di gaussiana in due dimensioni.

A questo punto possiamo enunciare una notevole proprietà delle variabili gaussiane: *quando il coefficiente di correlazione lineare è nullo, esse sono indipendenti*. Infatti, ponendo $\rho = 0$ nella (4.40), si ottiene una densità corrispondente al prodotto di due gaussiane in x e y . In base al teorema 4.1, questa è la densità di due variabili aleatorie indipendenti. La condizione $\rho = 0$, che in generale è solo necessaria per l'indipendenza, in questo caso diventa anche sufficiente. Vale pertanto il:

Teorema 4.3 (Indipendenza di variabili gaussiane). *Condizione necessaria e sufficiente affinché due variabili aleatorie congiuntamente gaussiane (normali) siano indipendenti, è che il loro coefficiente di correlazione lineare sia nullo.*

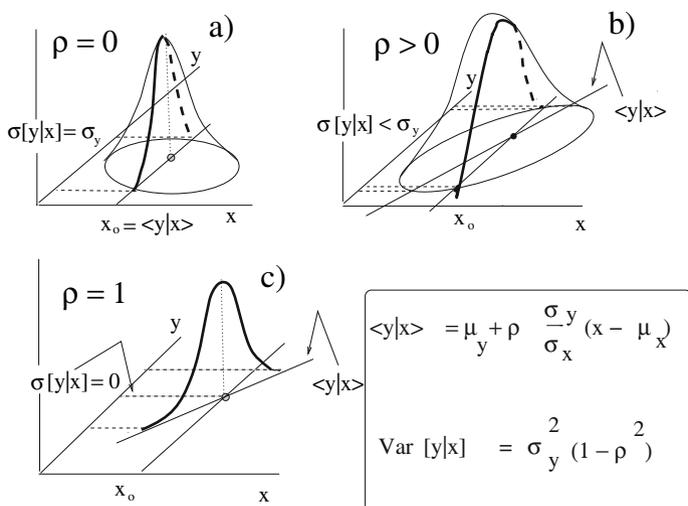


Fig. 4.4. Gaussiana in due dimensioni nel caso di variabili incorrelate a), parzialmente correlate b), totalmente correlate c). Sono anche mostrate le distribuzioni marginali per un determinato valore x_0 , la retta di regressione $\langle Y|x \rangle$ e la corrispondente dispersione $\sigma[Y|x]$

Vediamo quindi che l'esistenza di una covarianza nulla (che implica $\rho = 0$) tra variabili gaussiane assicura la indipendenza statistica e quindi l'assenza di un legame di causa-effetto tra di esse (si veda la Fig. 4.2).

Riconsiderando ora le distribuzioni marginali ridotte (4.39), notiamo che, passando dalle variabili standard a quelle normali si ottengono in x e y due distribuzioni gaussiane monodimensionali del tipo (3.28), nelle quali non compare il coefficiente di correlazione. Emerge qui un altro fatto importante: la conoscenza delle due distribuzioni marginali, cioè delle proiezioni della gaussiana sui due assi x e y , non basta per una conoscenza completa della densità bidimensionale, perché variabili correlate o incorrelate danno in entrambi i casi proiezioni gaussiane. La conoscenza della covarianza o del coefficiente di correlazione è quindi essenziale per la determinazione completa della distribuzione statistica delle variabili.

Facciamo ora qualche considerazione sulla forma della gaussiana bidimensionale. La Fig. 4.4 riporta tre casi: variabili incorrelate, parzialmente correlate e totalmente correlate. Se immaginiamo di tagliare la curva con piani di quota costante, l'intersezione dà luogo ad una curva di equazione:

$$u^2 - 2\rho uv + v^2 = \text{costante} \quad (4.41)$$

per variabili standard, e ad una curva

$$\frac{(x - \mu_x)^2}{\sigma_x^2} - 2\rho \frac{(x - \mu_x)(y - \mu_y)}{\sigma_x \sigma_y} + \frac{(y - \mu_y)^2}{\sigma_y^2} = \text{costante} \quad (4.42)$$

per variabili normali. Queste curve in generale rappresentano delle ellissi, dette *ellissi di concentrazione*, con centro nel punto (μ_x, μ_y) . Se $\rho = 0$, l'ellisse ha gli assi principali paralleli agli assi di riferimento e degenera in una circonferenza se $\sigma_x = \sigma_y$. Infine, se $\rho = \pm 1$, le variabili sono completamente correlate e l'ellisse degenera in una retta di equazione

$$\frac{(x - \mu_x)}{\sigma_x} \pm \frac{(y - \mu_y)}{\sigma_y} = \text{costante} . \quad (4.43)$$

In questo caso la curva normale si schiaccia completamente su un piano, come mostrato in Fig. 4.4c). Si ha allora a che fare con una sola variabile aleatoria (X o Y) essendo l'una completamente dipendente dall'altra attraverso una relazione matematica. A questo punto è necessario soffermarsi su di un'importante proprietà della gaussiana bidimensionale: quando gli assi principali dell'ellisse sono paralleli agli assi di riferimento, allora $\rho = 0$, le variabili sono incorrelate e quindi, per il teorema 4.3, anche indipendenti. È allora sempre possibile, operando una rotazione di assi, *trasformare una coppia di variabili gaussiane dipendenti in variabili gaussiane indipendenti*. Questa proprietà che, come vedremo, è generalizzabile ad un numero qualunque di dimensioni, sta a fondamento del test χ^2 applicato alla verifica di ipotesi in statistica.

L'espressione dell'angolo di rotazione la si trova in modo abbastanza semplice operando sulle variabili standard (4.34) una generica rotazione di un angolo α degli assi di riferimento:

$$\begin{aligned} A &= U \cos \alpha + V \sin \alpha \\ B &= -U \sin \alpha + V \cos \alpha . \end{aligned} \quad (4.44)$$

Se non vi ricordate questa formula potete consultare un qualunque testo di fisica o geometria. Le nuove variabili A e B per essere incorrelate e indipendenti devono avere la covarianza (e quindi il coefficiente di correlazione) uguale a zero. Imponiamo dunque:

$$\begin{aligned} \langle AB \rangle &= \langle (U \cos \alpha + V \sin \alpha)(-U \sin \alpha + V \cos \alpha) \rangle \\ &= -\sin \alpha \cos \alpha (\langle U^2 \rangle - \langle V^2 \rangle) + (\cos^2 \alpha - \sin^2 \alpha) \langle UV \rangle \\ &= -\frac{1}{2} \sin 2\alpha (\langle U^2 \rangle - \langle V^2 \rangle) + \cos 2\alpha \langle UV \rangle = 0 . \end{aligned}$$

Poiché per le variabili standard $\langle U^2 \rangle = \langle V^2 \rangle = \sigma^2 = 1$, si ottiene la condizione

$$\cos 2\alpha \langle UV \rangle = 0 \implies \cos 2\alpha = 0 \implies 2\alpha = \pm \frac{\pi}{2} \implies \alpha = \pm \frac{\pi}{4} . \quad (4.45)$$

Per questo valore di α il sistema (4.44) diventa:

$$A = \frac{1}{\sqrt{2}}(U + V) , \quad B = \frac{1}{\sqrt{2}}(-U + V) . \quad (4.46)$$

Sono naturalmente valide le relazioni inverse:

$$U = \frac{1}{\sqrt{2}}(A - B) , \quad V = \frac{1}{\sqrt{2}}(A + B) , \quad (4.47)$$

ed anche la conservazione della norma in una rotazione, come è facile verificare:

$$A^2 + B^2 = U^2 + V^2 . \quad (4.48)$$

Sostituendo le (4.46-4.48) nella (4.41) otteniamo:

$$U^2 - 2\rho UV + V^2 = (1 - \rho)A^2 + (1 + \rho)B^2 . \quad (4.49)$$

Se ora consideriamo l'esponente della gaussiana standard (4.38), passando alla notazione minuscola, vediamo che esso si trasforma come:

$$-\frac{1}{2(1 - \rho^2)}(u^2 - 2\rho uv + v^2) = -\frac{1}{2} \left[\frac{a^2}{1 + \rho} + \frac{b^2}{1 - \rho} \right] .$$

Poiché lo jacobiano della trasformazione (4.47) vale:

$$\begin{vmatrix} \frac{\partial u}{\partial a} & \frac{\partial u}{\partial b} \\ \frac{\partial v}{\partial a} & \frac{\partial v}{\partial b} \end{vmatrix} = \begin{vmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{vmatrix} = \frac{1}{2} + \frac{1}{2} = 1 ,$$

otteniamo:

$$\begin{aligned} g(u, v; 0, 1) du dv &= \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp \left[-\frac{1}{2(1 - \rho^2)}(u^2 - 2\rho uv + v^2) \right] \\ &= \frac{1}{2\pi} \exp \left[-\frac{1}{2} \left(\frac{a^2}{1 + \rho} + \frac{b^2}{1 - \rho} \right) \right] \frac{da db}{\sqrt{1 - \rho^2}} \\ &= \frac{1}{2\pi} \exp \left[-\frac{1}{2} (a_\rho^2 + b_\rho^2) \right] da_\rho db_\rho , \end{aligned} \quad (4.50)$$

dove:

$$a_\rho = \frac{a}{\sqrt{1 + \rho}} , \quad b_\rho = \frac{b}{\sqrt{1 - \rho}} . \quad (4.51)$$

Le (4.50, 4.51) contengono due risultati di fondamentale importanza.

Il primo è che, date due variabili gaussiane con $\rho \neq \pm 1$, è sempre possibile, passando alle variabili standard ed operando una rotazione di un angolo $\alpha = \pm 45^\circ$, definire due nuove variabili (A, B) tra loro incorrelate. L'angolo corrisponde ad una rotazione che porta gli assi principali dell'ellisse di concentrazione paralleli agli assi di riferimento. Il doppio segno nell'angolo indica semplicemente che la rotazione può consistere indifferentemente nel portare un determinato asse dell'ellisse parallelo all'asse x oppure all'asse y del sistema di riferimento.

Il secondo fatto importante è che dalle variabili incorrelate (A, B) è possibile ottenere due variabili (A_ρ, B_ρ) che sono *gaussiane standard indipendenti*. Poiché la somma dei quadrati di variabili gaussiane standard indipendenti,

in base al teorema di Pearson 3.3, si distribuisce come χ^2 , se ne deduce che due variabili gaussiane con coefficiente di correlazione $\rho \neq \pm 1$, con legame funzionale dato dall'esponente della gaussiana bidimensionale (4.40), danno luogo ad una variabile $\chi^2(2)$ con due gradi di libertà:

$$Q = \gamma(X, Y) . \quad (4.52)$$

Se $\rho = \pm 1$, tra A e B esiste una relazione deterministica e i gradi di libertà scendono a uno soltanto. Mostriamo tra poco che la relazione fondamentale (4.52) vale per un numero qualunque di dimensioni.

Il fatto che la (4.52) sia una variabile χ^2 è un fatto importante in sé, e non richiede di conoscere la forma esplicita delle variabili gaussiane indipendenti che vi compaiono: è sufficiente sapere che esse possono sempre essere trovate, se $\rho \neq 1$. Il calcolo del χ^2 viene fatto di solito con le variabili di partenza del problema, anche se esse sono correlate.

Per finire, studiamo le densità normali condizionate. La densità $g(y|x)$, per un certo x fissato, si trova facilmente applicando la definizione (4.13). Nel nostro caso questo comporta la divisione della densità (4.40) per la densità marginale $g_X(x)$, che, in base a quanto già detto, non è altro che una densità gaussiana monodimensionale $g(x; \mu_x, \sigma_x)$ data dalla (3.28). Dopo una facile rielaborazione, si ottiene:

$$\begin{aligned} g(y|x) &= \frac{1}{\sigma_y \sqrt{2\pi(1-\rho^2)}} \exp \left[-\frac{1}{2(1-\rho^2)} \left(\frac{y - \mu_y}{\sigma_y} - \rho \frac{x - \mu_x}{\sigma_x} \right)^2 \right] \\ &= \frac{1}{\sigma_y \sqrt{2\pi(1-\rho^2)}} \exp \left[-\frac{\left[y - \mu_y - \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x) \right]^2}{2\sigma_y^2(1-\rho^2)} \right] . \end{aligned} \quad (4.53)$$

Dato che x è costante, questa curva corrisponde ad una gaussiana monodimensionale di media e varianza date da:

$$\langle Y|x \rangle = \mu_y + \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x) \quad (4.54)$$

$$\text{Var}[Y|x] = \sigma_y^2 (1 - \rho^2) . \quad (4.55)$$

Le formule appena trovate mostrano che la media di Y condizionata ad x si dispone, nel piano (x, y) ed al variare di x , lungo una retta, detta *retta di regressione*. Si dice anche che tra le medie condizionate della gaussiana bivariata esiste un legame di *regressione lineare*. La varianza di Y , al contrario, si mantiene costante e dipende da x solo attraverso il coefficiente di correlazione. Uno sperimentatore che esegue una serie di esperimenti consistenti nel campionare Y per diversi valori x mantenuti costanti, osserva una media che si sposta lungo la retta (4.54) ed una varianza costante (4.55). La varianza condizionata $\text{Var}[Y|x]$ misura la dispersione dei dati intorno alla retta di regressione ed è sempre \leq di quella proiettata σ_y^2 . Essa dipende dall'angolo

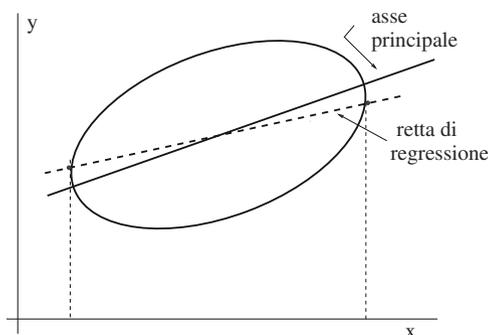


Fig. 4.5. Differenza tra retta di regressione e assi principali dell'ellisse di concentrazione

che gli assi principali dell'ellisse di concentrazione formano con gli assi di riferimento. Se $\rho = 0$ le due coppie di assi sono parallele, le rette di regressione coincidono con gli assi dell'ellisse, non c'è correlazione e le due varianze sono uguali; se $\rho = 1$, $\text{Var}[Y|x] = 0$ e lungo la retta di regressione non si osserva dispersione (Fig. 4.4).

La densità condizionata $g(x|y)$, per un prefissato valore di y , e le corrispondenti media, varianza e retta di regressione si ottengono dalle formule precedenti scambiando y con x .

È possibile, tra variabili gaussiane, una correlazione non lineare? I calcoli appena svolti mostrano che, se le densità marginali (proiettate) sono entrambe gaussiane, *la relazione deve essere lineare*: una relazione non lineare *distorce* la forma gaussiana su almeno uno dei due assi. Vi anticipiamo anche che la dipendenza non lineare verrà trattata in modo dettagliato più avanti, nel cap. 10.

Notiamo infine che le rette di regressione sono, per costruzione, il luogo dei punti di tangenza all'ellisse delle rette parallele agli assi e quindi non coincidono con gli assi principali dell'ellisse (Fig. 4.5).

Esercizio 4.3

Sono date due variabili gaussiane, X di media $\mu = 25$ e deviazione standard $\sigma_x = 6$ e

$$Y = 10 + X + Y_R, \quad (4.56)$$

dove Y_R è gaussiana di parametri $\mu_{y_R} = 0$ e $\sigma_{y_R} = 6$.

Determinare la covarianza e il coefficiente di correlazione tra le due variabili. Verificare i risultati con una simulazione.

Risposta. Media e deviazione standard di Y valgono:

$$\langle Y \rangle = \mu_y = 10 + \langle X \rangle = 35, \quad \sigma[Y] = \sigma_y = \sqrt{\sigma_x^2 + \sigma_{y_R}^2} = 8.48.$$

Definendo $\Delta X = X - \mu_x$, $\Delta Y = Y - \mu_y$, dalla (4.24) possiamo calcolare la covarianza tra le variabili come:

$$\text{Cov}[X, Y] = \langle \Delta X \Delta Y \rangle = \langle \Delta X (\Delta X + \Delta Y_R) \rangle = \langle \Delta^2 X \rangle = \sigma_x^2 = 36, \quad (4.57)$$

dove si è sfruttato il fatto che $\langle \Delta X \Delta Y_R \rangle = 0$, essendo le variabili X ed Y_R incorrelate per costruzione. Il coefficiente di correlazione vale:

$$\rho = \frac{\text{Cov}[X, Y]}{\sigma[X]\sigma[Y]} = \frac{36}{6 \cdot 8.48} = 0.707.$$

Si noti che, dalla (4.55), risulta $\text{Var}[Y|x] = \sigma_y^2(1 - \rho^2) = 8.48^2(1 - 0.707^2) = 6^2 = \sigma_{yR}^2$. Verifichiamo ora i risultati con una simulazione.

Come vi spiegheremo in dettaglio nel cap. 7, disponendo del generatore random di variabili uniformi, è possibile “produrre” attraverso un calcolatore delle variabili casuali gaussiane di media e deviazione standard prefissate.

Utilizzando il codice SCILAB:

```
x=grand(1,200000,'nor',25,6); // 200000 variabili normali
y=grand(1,200000,'nor',0,6); // di media 0 e 25 e std dev 6
y = 10. + x + y ;
clf(); // viene pulita la finestra grafica
Hist3d(x,y,20,20,plot4=1); // grafico di Fig. 4.6
```

si ottengono 200 000 valori simulati della coppia (X, Y) in accordo con la (4.56). Dato che SCILAB non possiede routine efficaci per la rappresentazione grafica di istogrammi bidimensionali, abbiamo fatto uso della nostra routine **Hist3d** che fornisce un risultato come quello di Fig. 4.6, corredato del calcolo di medie, deviazioni standard, covarianza e coefficiente di correlazione. Poiché la correlazione tra X e Y è di tipo lineare, gli istogrammi marginali $g(x)$ e $g(y)$ hanno una forma di tipo gaussiano, come si può vedere facilmente dai grafici di Fig. 4.6. Le medie e le deviazioni standard di questi istogrammi sono una stima delle corrispondenti quantità vere delle densità $g_X(x)$ e $g_Y(y)$. Esse si possono calcolare tramite le (2.37, 2.39, 4.23). Dalle 200 000 coppie simulate abbiamo ottenuto:

$$m_x = \frac{1}{N} \sum_{k=1}^N x_k = 24.99, \quad m_y = \frac{1}{N} \sum_{k=1}^N y_k = 34.98,$$

$$s_x = \left[\frac{1}{N-1} \sum_{k=1}^N (x_k - m_x)^2 \right]^{1/2} = 5.99,$$

$$s_y = \left[\frac{1}{N-1} \sum_{k=1}^N (y_k - m_y)^2 \right]^{1/2} = 8.46,$$

$$s_{xy} = \frac{1}{N-1} \sum_i (x_i - m_x)(y_i - m_y) = 35.81.$$

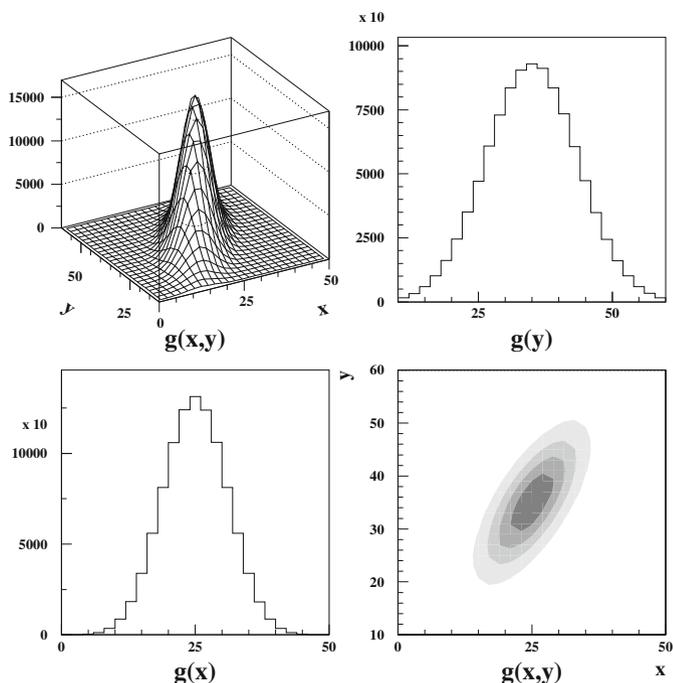


Fig. 4.6. Campione di 200 000 eventi, simulato al computer utilizzando una popolazione avente la densità gaussiana bivariata dell'esercizio 4.3. Istogramma bidimensionale $g(x, y)$, istogrammi marginali $g(x)$ e $g(y)$ e vista dall'alto con le curve di densità (in basso a destra)

Le dispersioni sono state calcolate dividendo per $(N - 1) = 199\,999$ poiché si sono utilizzate le medie empiriche. Queste quantità sono state indicate con lettere latine perché sono stime campionarie delle (4.17, 4.18). Il coefficiente di correlazione stimato vale allora:

$$r = \frac{s_{xy}}{s_x s_y} = 0.707 . \quad (4.58)$$

Dato che il campione simulato contiene un grande numero di eventi, tutti i valori che abbiamo ricavato dai dati coincidono, entro "qualche per mille", con quelli veri. Questi aspetti di convergenza saranno trattati nel cap. 6.

4.5 Generalizzazione in più dimensioni

La generalizzazione delle formule di questo capitolo al caso di più di due variabili è abbastanza immediata e non presenta difficoltà particolari.

La densità di probabilità congiunta di n variabili aleatorie è data da una funzione non negativa:

$$p(x_1, x_2, \dots, x_n), \quad (4.59)$$

tale che, se $A \subseteq \mathbb{R}^n$,

$$P\{(x_1, x_2, \dots, x_n) \in A\} = \int_A p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n.$$

Nel caso di variabili indipendenti si ha:

$$p(x_1, x_2, \dots, x_n) = p_1(x_1)p_2(x_2) \dots p_n(x_n). \quad (4.60)$$

La media e la varianza della variabile X_k si ottengono generalizzando le (4.5):

$$\begin{aligned} \langle X_k \rangle &= \int x_k p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = \mu_k, \\ \text{Var}[X_k] &= \int (x_k - \mu_k)^2 p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n. \end{aligned} \quad (4.61)$$

Nel caso di più variabili la covarianza può essere calcolata con la (4.20) per una qualunque coppia di variabili (x_i, x_k) :

$$\text{Cov}[X_i, X_k] = \int (x_i - \mu_i)(x_k - \mu_k) p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n. \quad (4.62)$$

Si hanno pertanto $n(n-1)/2$ covarianze diverse ed n varianze. Esse sono spesso scritte in una matrice simmetrica V , detta *matrice delle covarianze*:

$$V = \begin{bmatrix} \text{Var}[X_1] & \text{Cov}[X_1, X_2] & \dots & \text{Cov}[X_1, X_n] \\ \dots & \text{Var}[X_2] & \dots & \text{Cov}[X_2, X_n] \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \text{Var}[X_n] \end{bmatrix}, \quad (4.63)$$

dove gli elementi diagonali sono le n varianze e quelli non diagonali le $n(n-1)/2$ covarianze. La matrice è simmetrica perché ovviamente $\text{Cov}[X_i, X_k] = \text{Cov}[X_k, X_i]$; per questo motivo essa è stata scritta esplicitamente solo sopra la diagonale. In totale gli elementi diversi della matrice sono

$$\frac{n(n-1)}{2} + n = \frac{n(n+1)}{2}.$$

Sovente si introducono i coefficienti di correlazione $\rho[X_i, X_k]$ ottenuti generalizzando la (4.31) e la matrice delle covarianze viene espressa anche come:

$$V = \begin{bmatrix} \text{Var}[X_1] & \rho[X_1, X_2]\sigma[X_1]\sigma[X_2] & \dots & \rho[X_1, X_n]\sigma[X_1]\sigma[X_n] \\ \dots & \text{Var}[X_2] & \dots & \rho[X_2, X_n]\sigma[X_2]\sigma[X_n] \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \text{Var}[X_n] \end{bmatrix}. \quad (4.64)$$

Se le variabili sono incorrelate, i termini fuori dalla diagonale sono tutti nulli, ed i termini diagonali coincidono con le varianze delle singole variabili.

Invece della matrice delle covarianze talvolta si usa la matrice di correlazione:

$$C = \begin{bmatrix} 1 & \rho[X_1, X_2] & \dots & \rho[X_1, X_n] \\ \dots & 1 & \dots & \rho[X_2, X_n] \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & 1 \end{bmatrix}, \quad (4.65)$$

la quale è anch'essa simmetrica. Generalizzando la (4.24) in più dimensioni, la matrice delle covarianze si può esprimere come:

$$\langle (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\dagger \rangle = V, \quad (4.66)$$

dove $(\mathbf{X} - \boldsymbol{\mu})$ è un vettore colonna e il simbolo \dagger indica l'operazione di trasposizione. In questo modo il prodotto righe per colonne tra un vettore colonna ed una riga fornisce la matrice quadrata simmetrica (4.63).

Il calcolo della matrice delle covarianze e dei coefficienti di correlazione a partire da un insieme di dati può essere fatto con la nostra routine `Covardat`. Le istruzioni per l'uso si trovano nei commenti al codice.

Le matrici V e C compaiono spesso nel calcolo delle probabilità multidimensionale. Si può mostrare che esse godono della fondamentale proprietà di essere *definite positive*:

$$\mathbf{x}^\dagger V \mathbf{x} \geq 0, \quad (4.67)$$

dove l'uguaglianza vale solo se tutti gli n elementi del vettore \mathbf{x} sono nulli. Notate che nella forma quadratica (4.67) a sinistra compare un vettore riga \mathbf{x}^\dagger ed a destra un vettore colonna \mathbf{x} , in modo da ottenere un numero (non una matrice) che, nel caso V sia diagonale, è una somma di quadrati del tipo $\sum x_i^2 V_{ii}$.

Poiché $V^\dagger = V$, $(V^{-1})^\dagger = V^{-1}$, se $\mathbf{x} = V^{-1}\mathbf{y}$ si ottiene la relazione:

$$\mathbf{y}^\dagger V^{-1} V V^{-1} \mathbf{y} = \mathbf{y}^\dagger V^{-1} \mathbf{y} \geq 0 \quad (4.68)$$

la quale mostra che anche la matrice inversa è definita positiva. Un'altra utile proprietà è che, per ogni matrice definita positiva V , esiste una matrice H tale che:

$$H H^\dagger = V \implies H H^\dagger V^{-1} H = H \implies H^\dagger V^{-1} H = \mathbf{1}, \quad (4.69)$$

dove $\mathbf{1}$ è la matrice unità. Tutte queste proprietà sono dimostrate in dettaglio nel classico testo di Cramér [Cra51].

È possibile anche generalizzare le (4.11, 4.13) ed ottenere varie distribuzioni marginali e condizionate di una o più variabili integrando su quelle rimanenti. La generalizzazione è ovvia e non la riportiamo.

Quando le \mathbf{X} sono variabili gaussiane, esse hanno una funzione densità data dalla generalizzazione multidimensionale della gaussiana bivariata (4.40):

$$g(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} |V|^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\dagger V^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]; \quad |V| \equiv \det|V| \neq 0 \quad (4.70)$$

dove V è la matrice delle covarianze (4.63).

La densità (4.70) è detta *gaussiana multivariata* e può essere ricavata estendendo in più dimensioni gli argomenti con i quali si è dedotta la gaussiana bivariata. Ovviamente, è facile verificare che la gaussiana multivariata contiene quella bivariata come caso particolare. Infatti, se $\rho \neq \pm 1$:

$$V^{-1} = \left(\begin{vmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{vmatrix} \right)^{-1} = \frac{1}{\sigma_x^2 \sigma_y^2 (1 - \rho^2)} \begin{vmatrix} \sigma_y^2 & -\sigma_{xy} \\ -\sigma_{xy} & \sigma_x^2 \end{vmatrix},$$

da cui la (4.40).

Se H è una matrice che soddisfa la (4.69), attraverso la trasformazione

$$H\mathbf{Z} = \mathbf{X} - \boldsymbol{\mu} \quad (4.71)$$

la forma definita positiva che compare nell'esponente della gaussiana si trasforma in:

$$\gamma(\mathbf{X}) \equiv (\mathbf{X} - \boldsymbol{\mu})^\dagger V^{-1} (\mathbf{X} - \boldsymbol{\mu}) = \mathbf{Z}^\dagger H^\dagger V^{-1} H \mathbf{Z} = \mathbf{Z}^\dagger \mathbf{Z} = \sum_{i=1}^n Z_i^2, \quad (4.72)$$

dove si è utilizzata la terza delle (4.69). L'equazione (4.71) è pertanto l'equivalente della rotazione bidimensionale (4.44). Le nuove variabili \mathbf{Z} sono ancora gaussiane perché combinazioni lineari di variabili gaussiane (si veda l'esercizio 5.3 del prossimo capitolo); inoltre, è facile vedere che esse hanno media nulla e che, in base alle (4.66, 4.71) ed alla terza delle (4.69), sono anche variabili standard incorrelate:

$$\begin{aligned} \langle \mathbf{Z} \mathbf{Z}^\dagger \rangle &= H^{-1} \langle (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\dagger \rangle (H^\dagger)^{-1} \\ &= H^{-1} V (H^\dagger)^{-1} = (H^\dagger V^{-1} H)^{-1} = \mathbf{1}. \end{aligned} \quad (4.73)$$

Dal teorema 4.3 segue che le Z_i sono mutuamente indipendenti e che pertanto la (4.72) è una variabile $Q \sim \chi^2(n)$. La relazione (4.52) vale quindi anche per il caso n -dimensionale.

Se $|V| = 0$, tra le n variabili esistono una o più relazioni di dipendenza lineare; è allora necessario individuare un numero $r < n$ di variabili linearmente indipendenti e riformulare la (4.70). Pensiamo sia utile ricordarvi che due variabili sono *stocasticamente* indipendenti (o dipendenti) se vale (o non vale) la (4.12); la dipendenza matematica di tipo lineare implica invece l'esistenza di equazioni di vincolo ben definite tra le variabili. Due variabili linearmente dipendenti o vincolate corrispondono statisticamente ad una sola variabile. Se l'equazione è lineare, dal teorema 4.2 di Cauchy-Schwarz risulta $\rho = \pm 1$.

Pertanto, quando si ha a che fare con variabili aleatorie espresse da forme quadratiche definite positive del tipo

$$(\mathbf{X} - \boldsymbol{\mu})^\dagger W (\mathbf{X} - \boldsymbol{\mu}),$$

dove le \mathbf{X} sono variabili gaussiane e W è una matrice simmetrica, occorre sempre verificare se $|W| = 0$. In questo caso è necessario ridurre il numero di variabili determinando, se vi sono tra le variabili p equazioni lineari, $(n - p)$ nuove variabili linearmente indipendenti. La teoria dei sistemi lineari assicura che è sempre possibile determinare queste nuove variabili in modo da ottenere ancora una forma quadratica definita positiva di dimensione $n - p$, per la quale valgono ancora le (4.72, 4.73). È quindi possibile esprimere la forma quadratica come somma di $n - p$ variabili gaussiane standard indipendenti T_i :

$$(\mathbf{X} - \boldsymbol{\mu})^\dagger W (\mathbf{X} - \boldsymbol{\mu}) = \sum_{i=1}^{n-p} T_i^2. \quad (4.74)$$

Possiamo a questo punto generalizzare il teorema di Pearson 3.3 come segue:

Teorema 4.4 (Forme quadratiche). *Una variabile aleatoria $Q \sim \chi^2(\nu)$ è data dalla somma dei quadrati di ν variabili standard gaussiane stocasticamente indipendenti, oppure dalla forma quadratica definita positiva (4.74) di variabili gaussiane. I gradi di libertà sono dati in questo caso dal numero di variabili presenti meno il numero di eventuali relazioni lineari che esistono tra di esse.*

4.6 Insiemi di probabilità in più dimensioni

Consideriamo ora il calcolo dei livelli di probabilità in più dimensioni. L'analogo bidimensionale dell'integrale (3.94) è dato da:

$$P\{a \leq X \leq b, c \leq Y \leq d\} = \int_a^b \int_c^d p(x, y) \, dx \, dy. \quad (4.75)$$

Esso fornisce la probabilità che in una prova si ottengano le realizzazioni numeriche delle due variabili aleatorie entro dei limiti fissati. L'intervallo di probabilità si trasforma in questo caso in un rettangolo, come mostrato in Fig. 4.7a). Quando i limiti inferiori sono $a = -\infty, c = -\infty$, la (4.75) fornisce l'analogo bidimensionale della funzione di ripartizione (2.34).

La (4.75) può essere estesa in più dimensioni considerando una variabile $\mathbf{X} = (X_1, X_2, \dots, X_n)$ e una regione D nello spazio n -dimensionale. La probabilità che una prova fornisca un vettore compreso in D è data allora da:

$$P\{\mathbf{X} \in D\} = \int_D p(x_1, x_2, \dots, x_n) \, dx_1 \, dx_2 \dots \, dx_n. \quad (4.76)$$

Nel caso di variabili gaussiane indipendenti la (4.75) dà luogo al prodotto di due integrali di densità gaussiane monodimensionali. Quando gli intervalli di probabilità di X e Y sono simmetrici e centrati sulle rispettive medie, i livelli di probabilità si possono ottenere come prodotto delle funzioni (3.38) definite a pagina 83:

$$P\{|X - \mu_x| \leq a, |Y - \mu_y| \leq b\} = 2 \int_{\mu_x}^{\mu_x+a} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2} \frac{(x-\mu_x)^2}{\sigma_x^2}} dx \times \\ \times 2 \int_{\mu_y}^{\mu_y+b} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2} \frac{(y-\mu_y)^2}{\sigma_y^2}} dy = [2 E(a/\sigma_x)] \cdot [2 E(b/\sigma_y)] . \quad (4.77)$$

I livelli di probabilità corrispondenti alla legge 3σ monodimensionale si ottengono pertanto come *prodotto dei livelli di probabilità monodimensionali* (3.35). Ad esempio, la probabilità che entrambe le variabili siano comprese entro $\pm\sigma$, cioè entro una deviazione standard, è data da:

$$P\{|X - \mu_x| \leq \sigma_x, |Y - \mu_y| \leq \sigma_y\} = 0.683 \cdot 0.683 = 0.466 . \quad (4.78)$$

Consideriamo ora la gaussiana bivariata standardizzata per il caso di variabili incorrelate (4.35) ed operiamo la trasformazione in coordinate polari:

$$u = r \cos \theta , \quad v = r \sin \theta .$$

Poiché l'elemento di superficie $du dv$ si trasforma in $r dr d\theta$ (lo jacobiano della trasformazione è r), la (4.35) si trasforma nella funzione:

$$g(r, \theta; 0, 1) = \frac{1}{2\pi} r e^{-\frac{1}{2}r^2} , \quad 0 \leq 2\pi , \quad 0 \leq r < \infty , \quad (4.79)$$

la quale non è altro che la densità χ^2 con due gradi di libertà; essa mostra che, utilizzando le variabili standard, *l'ellisse di concentrazione di Fig. 4.7a) si trasforma in un cerchio col centro nell'origine* (Fig. 4.7b). Da qui è poi facile ricavare le probabilità di cadere entro l'ellisse di concentrazione avente gli assi principali pari a k volte le deviazioni standard delle due variabili:

$$P\{U^2 + V^2 \leq k^2\} = P\left\{\frac{(X - \mu_x)^2}{\sigma_x^2} + \frac{(Y - \mu_y)^2}{\sigma_y^2} \leq k^2\right\} = \\ = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \int_0^k r e^{-\frac{1}{2}r^2} dr d\theta = 1 - e^{-\frac{1}{2}k^2} , \quad (4.80)$$

da cui ottiene una sorta di legge 3σ nel piano:

$$P\left\{\frac{(X - \mu_x)^2}{\sigma_x^2} + \frac{(Y - \mu_y)^2}{\sigma_y^2} \leq k^2\right\} = 1 - e^{-\frac{k^2}{2}} = \begin{cases} 0.393 \text{ per } k = 1 \\ 0.865 \text{ per } k = 2 \\ 0.989 \text{ per } k = 3 . \end{cases} \quad (4.81)$$

Notate che l'intervallo di probabilità rettangolare $\pm\sigma$ della (4.78) ha una probabilità maggiore (0.466 contro 0.393) di quello della corrispondente ellisse

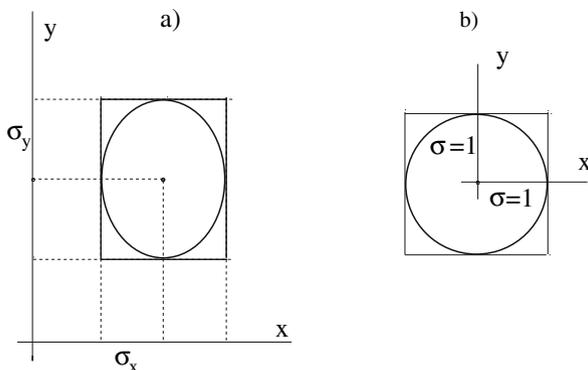


Fig. 4.7. Regione relativa alla probabilità $P\{|X - \mu_x| \leq \sigma_x, |Y - \mu_y| \leq \sigma_y\}$ ed ellisse di concentrazione corrispondente per variabili normali incorrelate (a); nel caso di variabili standard, il rettangolo e l'ellisse si trasformano rispettivamente in un quadrato e in un cerchio di semi-lato o raggio unitario centrati nell'origine (b)

(4.81), la quale per $k = 1$ ha i lati del rettangolo come assi principali. Questo fatto è mostrato anche in Fig. 4.7, dove si vede che il rettangolo o il quadrato di probabilità contiene l'ellisse o il cerchio di concentrazione.

Se, invece di due variabili standard (U, V) , consideriamo due variabili gaussiane (X, Y) di media nulla ed aventi la stessa varianza σ^2 e nell'integrale (4.80) operiamo la sostituzione $r \rightarrow r/\sigma$, otteniamo la funzione

$$\rho(r) = \frac{1}{2\pi} \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right), \quad x^2 + y^2 = r^2, \quad (4.82)$$

la quale, integrata in $dr d\theta$, dà la densità di probabilità nel piano per due variabili gaussiane indipendenti (X, Y) di uguale dispersione. Questa funzione è detta densità di Rayleigh, ed è l'analogo bidimensionale della densità di Maxwell (3.75), che è valida nello spazio.

Vi sarete accorti che gli intervalli di probabilità di una certa variabile sono diversi a seconda del numero globale di variabili che viene fatto variare e del tipo di dominio D che viene considerato nella (4.76). Fissando le idee al caso di due variabili (X, Y) , per un certo livello di probabilità fissato (ad esempio 68.3%), possiamo definire ben cinque intervalli di probabilità: un intervallo $(x \pm \sigma_x)$ relativo alla probabilità del 68.3% di trovare un valore di X per un qualunque valore di Y e l'analogo scambiando X con Y (e sono due), oppure l'intervallo $(\langle Y|x \rangle \pm \sqrt{\text{Var}[Y|x]})$ delle (4.54, 4.55) relativo alla probabilità del 68.3% di trovare un valore di Y per un valore fissato x_0 e l'analogo per X per un valore fissato y_0 (e sono quattro) ed infine possiamo definire una regione bidimensionale D in cui cade il 68.3% dei punti di (X, Y) . Le densità coinvolte nei cinque casi sono rispettivamente le densità marginali $p_X(x)$, $p_Y(y)$, le densità condizionate $p(y|x_0)$, $p(x|y_0)$, e la densità congiunta $p(x, y)$.

La regione del piano D dove effettuare le stime multidimensionali è quasi sempre l'ellisse di concentrazione:

$$Q = (\mathbf{X} - \boldsymbol{\mu})^\dagger V^{-1} (\mathbf{X} - \boldsymbol{\mu}) \rightarrow \sum_{i=1}^n \frac{(X_i - \mu_i)^2}{\sigma_i^2} \quad (4.83)$$

dove l'espressione a destra della freccia vale ovviamente in caso di variabili incorrelate. Questa regione possiede delle proprietà importanti che esaminiamo brevemente.

Nel caso di n variabili gaussiane, in base al teorema 4.4, la variabile Q ha distribuzione $\chi^2(n)$. Questo permette di utilizzare, nel calcolo dei livelli di probabilità associati all'ellisse di concentrazione multidimensionale (4.83), le tabelle D.3 e D.4 dei livelli di significatività della densità χ^2 riportate in Appendice D.

Dato che la densità marginale di ogni variabile è gaussiana, la variabile

$$Q_i = \frac{(X_i - \mu_i)^2}{\sigma_i^2} \quad (4.84)$$

segue la densità $\chi^2(1)$. I livelli di probabilità corrispondenti a $\{Q_i = 1, 4, 9\}$ non sono altro, come sappiamo, che i livelli di probabilità gaussiani del 68.3%, 95.4% e 99.7% corrispondenti ai valori $\{\sqrt{Q_i} = 1, 2, 3\}$ della variabile gaussiana standard non quadrata: $P\{Q_i \leq \chi^2\} = P\{T_i \leq \sqrt{\chi^2}\}$.

Le Tab. D.3 e D.4 sono due modi diversi di calcolare le probabilità relative alla distribuzione χ^2 : per ottenere i valori riportati nella Tab. D.4 per un determinato livello di probabilità α basta trovare il corrispondente livello $1 - \alpha$ in Tab. D.3 e moltiplicare il chi-quadrato ridotto $\chi_R^2(\nu)$ della tabella per i gradi di libertà ν del problema (si noti infatti che la Tab. D.4 riporta i valori di del χ^2 totale, non di quello ridotto).

La differenza tra i livelli di probabilità di ogni singola variabile e quelli congiunti è mostrata, per il caso bidimensionale, in Fig. 4.8, dove il valore $\chi^2 = 2.30$ è tratto dalla Tab. D.4 per $\nu = 2$. Guardando la figura dovete fare attenzione ad un fatto importante, che useremo spesso in seguito: *le proiezioni sugli assi della curva corrispondente a $\chi^2 = 1$ definiscono gli intervalli ad $1 - \sigma$ delle variabili, riferite alle rispettive densità marginali*. Questa proprietà, ovvia in una dimensione, si dimostra anche per due variabili ponendo uguale ad 1 la funzione χ^2 (4.52):

$$Q = \gamma(X, Y) = \frac{1}{1 - \rho^2} (u^2 - 2\rho uv + v^2) = 1 \quad (4.85)$$

e considerando l'intersezione di questa curva con la retta di regressione (4.54) $v = \rho u$. Si trova facilmente la soluzione, che è $u = \pm 1$, corrispondente ad un intervallo $(\mu_x \pm \sigma_x)$. Lo stesso vale ovviamente per la proiezione sull'asse y .

Analogamente al caso monodimensionale, spesso il modello gaussiano è ragionevolmente valido anche per una gran parte di fenomeni statistici in

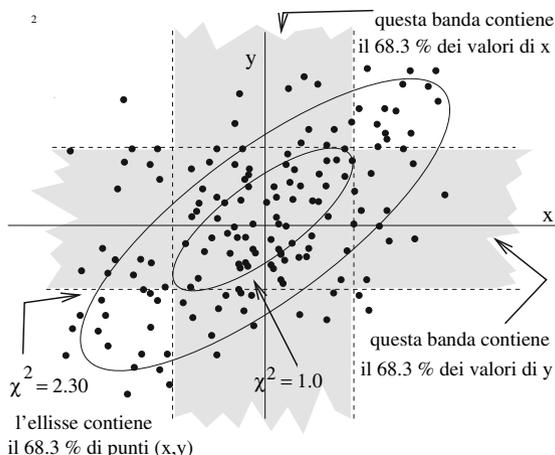


Fig. 4.8. Regioni di probabilità in una e due dimensioni. La curva corrispondente a $\chi^2 = 2.30$ contiene il 68.3% delle coppie di variabili. La proiezione sugli assi della curva $\chi^2 = 1$ fornisce gli intervalli di probabilità ad 1σ per ognuna delle variabili

più dimensioni. L'uso dell'ellisse di concentrazione e della densità χ^2 fornisce allora un mezzo molto potente per prevedere la probabilità di un risultato sperimentale costituito da una n -upla di valori (x_1, x_2, \dots, x_n) .

Se invece il problema non consente l'uso del modello gaussiano, è necessario ricorrere alla risoluzione degli integrali (4.76) definendo opportune regioni di probabilità, di solito ipercubiche o ellittiche, e tenendo conto con cura delle correlazioni tra le variabili. Spesso, per raggiungere questo difficile obiettivo, si ricorre ai metodi di simulazione Monte Carlo descritti nel cap. 7.

4.7 La distribuzione multinomiale

La densità binomiale (2.30) individua la distribuzione di probabilità della variabile $I = (I_1, I_2)$ definita come istogramma sperimentale di N eventi suddivisi in due canali con conteggi I_1 ed I_2 rispettivamente, quando le probabilità vere che una osservazione cada nei canali 1 e 2 sono rispettivamente $p \equiv p_1$ e $(1 - p) \equiv p_2$:

$$P\{I = n_1, n_2\} = b(x; N, p) = \frac{N!}{n_1!(N - n_1)!} p^{n_1} (1 - p)^{N - n_1} = \frac{N!}{n_1!n_2!} p_1^{n_1} p_2^{n_2}, \quad (4.86)$$

con $(p_1 + p_2) = 1$ e $(n_1 + n_2 = N)$. Questa equazione si presta ad una immediata generalizzazione al caso di un istogramma generico con un numero k di canali. La densità che si ottiene, detta *multinomiale*, rappresenta la probabilità che la variabile aleatoria $I = (I_1, I_2, \dots, I_k)$ fornisca un istogramma sperimentale

con un numero determinato di eventi n_i in ognuno dei canali, quando sono note le probabilità a priori p_i :

$$P\{I = n_1, n_2, \dots, n_k\} = b(\mathbf{n}; N, \mathbf{p}) = \frac{N!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k} \quad (4.87)$$

$$(p_1 + p_2 + \dots + p_k) = 1, \quad (n_1 + n_2 + \dots + n_k) = N. \quad (4.88)$$

Poiché ogni variabile, rispetto a tutte le altre, soddisfa i requisiti della densità binomiale, valgono le relazioni:

$$\langle I_i \rangle = N p_i, \quad \text{Var}[I_i] = N p_i (1 - p_i). \quad (4.89)$$

Le variabili (I_1, I_2, \dots, I_k) sono però dipendenti a causa della seconda delle (4.88). La loro covarianza può essere calcolata direttamente con la (4.62) oppure in modo molto più semplice utilizzando la legge di trasformazione della varianza. Il calcolo verrà svolto nel prossimo capitolo, nell'esercizio 5.11. Il risultato è la formula (5.85), che qui anticipiamo:

$$\text{Cov}[I_i, I_j] \equiv \sigma_{ij} = -N p_i p_j, \quad (i \neq j). \quad (4.90)$$

La covarianza è negativa, perché l'aumento del numero di eventi in un canale comporta la diminuzione del numero degli eventi negli altri canali.

Nel caso monodimensionale, la densità binomiale tende rapidamente alla densità gaussiana. La stessa proprietà vale anche per la densità multinomiale, il che è un fatto di grande importanza per studio degli istogrammi che faremo in statistica.

L'equazione (3.24), valida per $Np, N(1-p) > 10$, si può infatti anche scrivere come:

$$b(n; N, p) \propto \exp \left[-\frac{1}{2} \frac{(n - Np)^2}{Np(1-p)} \right] = \exp \left[-\frac{1}{2} \left(\frac{(n_1 - Np_1)^2}{Np_1} + \frac{(n_2 - Np_2)^2}{Np_2} \right) \right] \quad (4.91)$$

dove l'ultima uguaglianza, ottenuta ponendo $p \equiv p_1, (1-p) \equiv p_2, n \equiv n_1, (N-n) \equiv n_2$, suggerisce una forma simmetrica che può essere generalizzata al caso di n variabili (trovate la traccia della non facile dimostrazione in [Gne76]). Possiamo pertanto scrivere il risultato fondamentale:

$$b(\mathbf{n}; N, \mathbf{p}) \propto \exp \left[-\frac{1}{2} \sum_{i=1}^k \frac{(n_i - Np_i)^2}{Np_i} \right], \quad (Np_i > 10 \forall i), \quad (4.92)$$

dove la somma va estesa a tutte le k variabili che compaiono nelle somme (4.88). Abbiamo pertanto approssimato la densità multinomiale con una densità gaussiana multidimensionale, dove nell'esponentiale compare la somma dei quadrati delle variabili. Tuttavia, le variabili sono correlate attraverso la (4.88) e quindi i gradi di libertà sono in realtà solo $k-1$.

Se ora combiniamo questo risultato con il teorema 4.4, arriviamo all'importante conclusione che possiamo enunciare come:

Teorema 4.5 (Somma di Pearson). *In un istogramma di k canali, costruito a partire da un campione casuale di dimensione N e avente probabilità vera p_i in ogni canale, la variabile*

$$Q = \sum_{i=1}^k \frac{(I_i - Np_i)^2}{Np_i}, \quad (I_1 + I_2 + \cdots + I_k) = N,$$

dove I_i è il numero di eventi osservati in ogni canale, tende ad avere, per $N \rightarrow \infty$, densità χ^2 con $k - 1$ gradi di libertà.

Questo teorema, come vedremo, è la chiave di volta per l'impiego del test χ^2 in statistica e può già venire applicato con ottima approssimazione se $Np_i > 10$.

Notiamo anche che nella (4.93) non compare la somma dei quadrati di variabili standard, perché la varianza delle variabili I_i è data dalla (4.89) e le variabili sono correlate. Pertanto il teorema generalizza in modo non banale i risultati del teorema 3.3, che si riferivano a somme di quadrati di variabili gaussiane standard indipendenti.

Problemi

4.1. Se X e Y sono due variabili uniformi ed indipendenti negli intervalli $[0, a]$ e $[0, b]$ rispettivamente, trovare la densità congiunta $p(x, y)$.

4.2. Se X_i , $i = 1, 2, 3$ sono tre variabili gaussiane indipendenti di media μ_i e varianza σ_i^2 , trovare a) la striscia che contiene il 90% dei valori di X_1 , b) l'ellisse che contiene il 90% dei valori delle coppie (X_1, X_2) e c) l'ellissoide che contiene il 90% delle n -uple (X_1, X_2, X_3) .

4.3. Se X e Y sono variabili aleatorie indipendenti, vale l'uguaglianza $\langle Y|x \rangle = \langle Y \rangle$?

4.4. Calcolare approssimativamente per via grafica il coefficiente di correlazione dall'ellisse di concentrazione di Fig. 4.6 (in alto a destra).

4.5. Se $Z = aX + b$ e $U = cY + d$ con $ac \neq 0$, calcolare $\rho[Z, U]$.

4.6. In una popolazione la statura è una variabile gaussiana pari a $\langle X \rangle \pm \sigma[X] = 175 \pm 8$ cm per gli uomini e $\langle Y \rangle \pm \sigma[Y] = 165 \pm 6$ cm per le donne. Supponendo che non vi sia correlazione, trovare la percentuale di coppie con uomo e donna più alti di 180 e 170 cm rispettivamente.

4.7. Come risolvereste, in linea di principio, il problema precedente nel caso più realistico (le coppie tendono infatti ad avere statura omogenea) di un coefficiente di correlazione $\rho = 0.5$ tra la statura del marito e quella della moglie?

4.8. Nel lancio indipendente di due dadi si accetta il valore del secondo dado solo se il numero è dispari, altrimenti si assume come secondo valore quello del primo dado. Indicando con X e Y la coppia di valori ottenuti in ogni prova, trovare, per le due strategie, la densità di probabilità, le densità marginali di X e Y , la media e le deviazioni standard di X e Y e la loro covarianza.

Funzioni di variabili aleatorie

I libri di testo traboccano di trionfi dell'analisi lineare, i cui insuccessi sono sepolti così in profondità che le tombe non sono segnate e la loro esistenza passa inosservata.

Ian Stewart, *Dio gioca a dadi?*

5.1 Introduzione

Prima di passare alla statistica ci rimane ancora questo capitolo, in cui affrontiamo il problema seguente: se consideriamo più variabili aleatorie X, Y, \dots definite sullo stesso spazio di probabilità e le combiniamo in una funzione $Z = f(X, Y, \dots)$ del tipo (2.9), *otteniamo una nuova variabile aleatoria* Z ; se conosciamo la densità di probabilità congiunta $p_{XY\dots}(x, y, \dots)$ delle variabili aleatorie di partenza, quale sarà la densità $p_Z(z)$ della variabile Z ?

Per fissare le idee, consideriamo il caso di due variabili X e Y . Sono allora coinvolte due densità di probabilità p_{XY} e p_Z e una funzione $f(x, y)$, secondo lo schema seguente:

$$X, Y \sim p_{XY}(x, y) \implies Z = f(X, Y) \sim p_Z(z) .$$

Le densità p_{XY} è nota, $f(x, y)$ rappresenta un legame funzionale assegnato (somma, prodotto, esponenziale o altro) e il problema consiste nel determinare la densità p_Z della nuova variabile casuale Z . Questo schema, che può essere esteso in modo ovvio ad un qualunque numero di variabili, rappresenta l'essenza del problema che vogliamo risolvere. Le densità di probabilità verranno sempre indicate con la lettera $p(\dots)$, i legami funzionali con la lettera $f(\dots)$.

La variabile aleatoria Z resta definita in funzione dei risultati a e b delle prove che realizzano X e Y , secondo lo schema di Fig. 5.1:

$$Z = Z(a, b) = f(X(a), Y(b)) , \quad a \in A , \quad b \in B .$$

Nel caso di variabili discrete, la probabilità che Z appartenga ad un certo insieme numerico \mathbb{R}_Z è allora data, in base al teorema 1.1, dalla probabilità

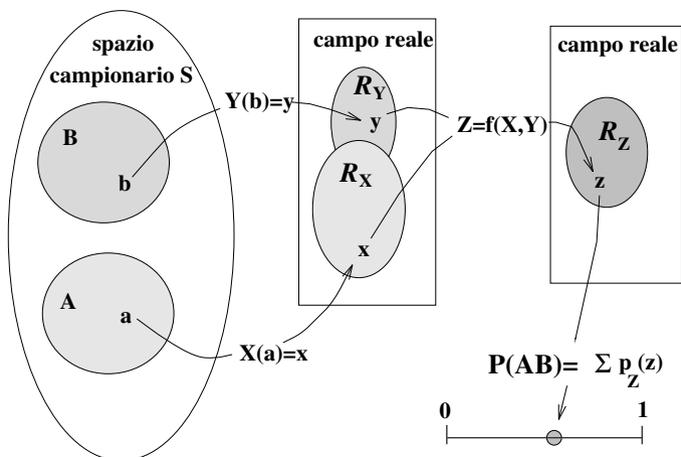


Fig. 5.1. Definizione della variabile aleatoria $Z = f(X, Y)$

di ottenere $\{X \in \mathbb{R}_X\}$ e $\{Y \in \mathbb{R}_Y\}$ (ovvero $a \in A$ e $b \in B$) sommata su tutti i casi che soddisfano la condizione $z \in \mathbb{R}_Z$:

$$P\{Z \in \mathbb{R}_Z\} = \sum_{[x,y:f(x,y) \in \mathbb{R}_Z]} P(AB) = \sum_{[x,y:f(x,y) \in \mathbb{R}_Z]} P\{X \in \mathbb{R}_X, Y \in \mathbb{R}_Y\} . \quad (5.1)$$

Se X e Y sono indipendenti, dal teorema 1.2 delle probabilità composte si ha anche:

$$P\{z \in \mathbb{R}_Z\} = \sum_{[x,y:f(x,y) \in \mathbb{R}_Z]} P\{X \in \mathbb{R}_X\} P\{Y \in \mathbb{R}_Y\} . \quad (5.2)$$

La Fig. 5.1 mostra schematicamente il significato della (5.1).

Consideriamo ora un semplice esempio: siano X e Y i punteggi che si ottengono lanciando due volte un dado ($1 \leq X, Y \leq 6$) e sia $Z = f(X, Y) = X + Y$ la somma dei due punteggi. Definiamo \mathbb{R}_Z l'insieme dei risultati minori di 5 ($Z = X + Y < 5$) e calcoliamo la probabilità $P\{Z \in \mathbb{R}_Z\}$. Se le due prove sono indipendenti, vale la (5.2) e la probabilità di ottenere una coppia (x, y) vale $1/6 \times 1/6 = 1/36$. La (5.1) prescrive di sommare le probabilità di tutte le coppie per le quali $Z < 5$: (1,1), (1,2), (1,3), (2,1), (2,2), (3,1). Si hanno 6 coppie e quindi $P\{Z < 5\} = 6/36 = 1/6$.

Nel seguito procederemo per passi successivi, considerando prima una variabile, poi due variabili, ed indicando infine come estendere il procedimento a n variabili. Mostriamo anche come sia possibile, limitandosi solo allo studio della trasformazione della media e della varianza attraverso la funzione f , arrivare ad una soluzione semplice e generale, anche se approssimata, del problema.

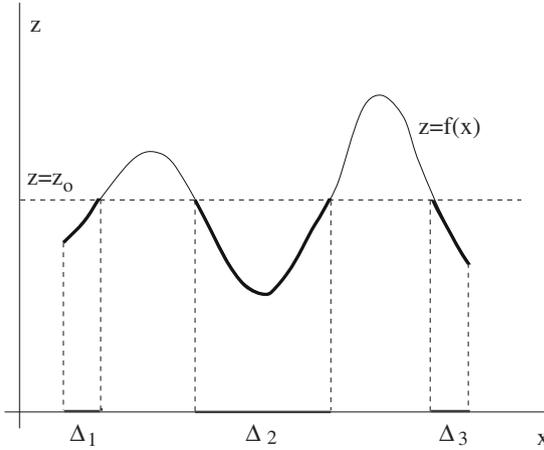


Fig. 5.2. La trasformazione $z = f(x)$

5.2 Funzione di una variabile aleatoria

Sia X una variabile casuale continua di densità $p_X(x)$ e sia Z una variabile casuale legata a X dalla relazione funzionale:

$$Z = f(X) . \quad (5.3)$$

Per determinare la densità $p_Z(z)$ basta applicare le note regole sul cambiamento di variabile in una funzione, *facendo però attenzione a come si trasformano gli intervalli di probabilità*. È quindi opportuno ricorrere agli integrali di probabilità definiti dalla funzione di ripartizione ed utilizzare le relazioni chiave (2.29, 2.36). Oltre a queste equazioni, ci serviremo del teorema di Leibnitz di derivazione di un integrale, la cui dimostrazione può essere trovata su molti testi di analisi matematica. Data una funzione $z = f(x)$, se $x = f^{-1}(z)$ esiste ed è derivabile in $[x_1, x_2]$, vale la relazione:

$$\begin{aligned} \frac{d}{dz} \int_{x_1=f^{-1}(z_1)}^{x_2=f^{-1}(z_2)} p(x) dx &= \left(\frac{df^{-1}}{dz} \right)_{z=z_2} p(x_2) - \left(\frac{df^{-1}}{dz} \right)_{z=z_1} p(x_1) \\ &= \frac{p(x_2)}{f'(x_2)} - \frac{p(x_1)}{f'(x_1)} , \end{aligned} \quad (5.4)$$

dove l'apice indica l'operazione di derivata.

Consideriamo ora una generica funzione $f(X)$ continua come in Fig. 5.2, e determiniamo la probabilità che Z sia minore di un determinato valore z_0 . In sostanza, dobbiamo trovare la probabilità che Z si trovi al di sotto della retta $z = z_0$ di Fig. 5.2.

Dalle (2.29, 5.1) e dalla figura otteniamo:

$$P\{Z \leq z_0\} \equiv F_Z(z) = \sum_i \int_{\Delta_i} p_X(x) dx \quad (5.5)$$

dove F_Z è la funzione cumulativa di Z e gli intervalli Δ_i sono indicati sempre in Fig. 5.2. Questi intervalli, tranne il primo e l'ultimo, hanno come estremi le radici reali x_1, x_2, \dots, x_n dell'equazione (5.3):

$$z_0 = f(x_1) = f(x_2) = \dots = f(x_n) . \quad (5.6)$$

L'equazione (2.36) ci assicura che, derivando la (5.5), otterremo la densità di probabilità cercata. Poiché dalla Fig. 5.2 si vede che l'estremo inferiore di ogni intervallo Δ_i non dipende da z (e quindi ha derivata nulla) oppure, quando dipende da z , ha derivata sempre negativa (la funzione decresce), applicando la formula di Leibnitz (5.4) alla (5.5) possiamo scrivere la densità $p_Z(z)$ come somma di termini tutti positivi, prendendo il valore assoluto delle derivate delle radici reali (5.6):

$$\frac{dF_Z(z)}{dz} = p_Z(z) = \frac{p_X(x_1)}{|f'(x_1)|} + \frac{p_X(x_2)}{|f'(x_2)|} + \dots + \frac{p_X(x_n)}{|f'(x_n)|} , \quad (5.7)$$

dove il membro a destra è una funzione di $z_0 \equiv z$ tramite l'inverso delle (5.6). Il risultato è sempre positivo, come si richiede per una densità di probabilità. Quando esiste una sola radice reale dell'equazione (5.3) e $p(x) \geq 0$, questa formula coincide con il solito metodo di sostituzione di variabile in un integrale. Questo metodo è già stato tacitamente applicato negli esercizi 3.10 e 3.11 sulle distribuzioni di Maxwell e di Boltzmann.

Applichiamo ora l'equazione fondamentale (5.7) a qualche altro caso significativo. Nel caso in cui l'equazione (5.3) sia data da:

$$Z = f(X) = aX + b \quad (5.8)$$

la (5.6) ammette la sola soluzione:

$$x_1 = \frac{z - b}{a} .$$

Poiché $f'(x) = a$, dalla (5.7) si ottiene:

$$p_Z(z) = \frac{1}{|a|} p_X\left(\frac{z - b}{a}\right) . \quad (5.9)$$

Se invece consideriamo il legame funzionale:

$$Z = f(X) = aX^2 , \quad (5.10)$$

la (5.6) ammette le soluzioni:

$$x_1 = -\sqrt{\frac{z}{a}} , \quad x_2 = +\sqrt{\frac{z}{a}} \quad (5.11)$$

che danno luogo alle derivate:

$$|f'(x_1)| = |f'(x_2)| = 2a\sqrt{\frac{z}{a}} = 2\sqrt{az},$$

le quali, inserite nella (5.7), permettono di determinare la densità cercata:

$$p_Z(z) = \frac{1}{2\sqrt{az}} \left[p_X \left(-\sqrt{\frac{z}{a}} \right) + p_X \left(\sqrt{\frac{z}{a}} \right) \right], \quad z \geq 0. \quad (5.12)$$

Se la densità $p_X(x)$ di partenza è la gaussiana (3.28) e si applica l'equazione (5.8), la (5.9) diventa:

$$p_Z(z) = \frac{1}{|a|\sigma\sqrt{2\pi}} \exp \left[-\frac{[z - (a\mu + b)]^2}{2a^2\sigma^2} \right] \quad (5.13)$$

la quale è una gaussiana di media e varianza date da

$$\mu_z = a\mu + b, \quad \sigma_z^2 = a^2\sigma^2. \quad (5.14)$$

Notiamo che, poiché la trasformazione (5.8) è lineare, le (5.14) si possono dedurre direttamente anche dalle (2.62, 2.63).

Nel caso di una trasformazione quadratica (5.10) di una variabile proveniente da una distribuzione gaussiana di media nulla, la (5.12) diventa:

$$p_Z(z) = \frac{1}{\sigma\sqrt{2\pi az}} \exp \left(-\frac{z}{2a\sigma^2} \right), \quad z > 0, \quad (5.15)$$

che è la densità gamma (3.57) con $k = 1/2$. La media e la varianza possono essere ottenute dalle (3.58), oppure integrando per parti le (2.53, 2.56), e sono date da:

$$\mu_z = a\sigma^2, \quad \sigma_z^2 = 2a^2\sigma^4. \quad (5.16)$$

5.3 Funzioni di più variabili aleatorie

Generalizziamo i risultati del paragrafo precedente al caso di funzioni di più variabili aleatorie.

Consideriamo per primo il caso semplice di una sola variabile Z funzione di n variabili aleatorie:

$$Z = f(X_1, X_2, \dots, X_n) \equiv f(\mathbf{X}). \quad (5.17)$$

L'analogo della funzione di ripartizione o cumulativa (5.5) è in questo caso dato dall'equazione:

$$P\{Z \leq z_0\} = F_Z(z) = \int \dots \int_{(\mathbf{X} \in D)} p_{\mathbf{X}}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n \quad (5.18)$$

dove $p_{\mathbf{X}}(x_1, x_2, \dots, x_n)$ è la densità di probabilità delle n variabili e D è l'insieme che comprende tutte le n -uple $\mathbf{X} = (x_1, x_2, \dots, x_n)$ tali che $P\{Z \leq z_0\}$, in accordo con la (5.1).

È bene notare che la densità di probabilità delle variabili \mathbf{X} compare solo come argomento dell'integrale, mentre il legame funzionale $Z = f(\mathbf{X})$ compare esclusivamente nella determinazione del dominio di integrazione D .

La derivazione della cumulativa (5.18) risolve in molti casi il problema della determinazione della densità p_z . Questo metodo è la generalizzazione di quello usato nel par. 3.9, dove abbiamo derivato la (3.61) per ottenere la densità χ^2 .

Per trattare il caso più generale di n variabili \mathbf{Z} funzioni di n variabili aleatorie \mathbf{X} di partenza, in alternativa all'uso della funzione cumulativa o di ripartizione, si può ricorrere ad una celebre formula basata sul teorema generale del cambiamento di variabile in una funzione integranda. Poiché questo teorema si trova dimostrato su molti testi di analisi matematica, qui ne riportiamo soltanto l'enunciato:

Teorema 5.1 (Cambiamento di variabile in funzioni densità). *Siano $\mathbf{X} \equiv (X_1, X_2, \dots, X_n)$ n variabili casuali con densità congiunta $p_{\mathbf{X}}(\mathbf{x})$ e siano $\mathbf{Z} \equiv (Z_1, Z_2, \dots, Z_n)$ n variabili legate alle \mathbf{X} dalle n relazioni funzionali*

$$\begin{aligned} Z_1 &= f_1(\mathbf{X}) \\ Z_2 &= f_2(\mathbf{X}) \\ &\dots\dots\dots \\ Z_n &= f_n(\mathbf{X}), \end{aligned} \tag{5.19}$$

le quali siano tutte invertibili e derivabili con derivata continua rispetto a tutti gli argomenti (vi sia cioè una corrispondenza biunivoca tra i due domini delle \mathbf{X} e delle \mathbf{Z} , per cui $X_1 = f_1^{-1}(\mathbf{Z})$, ecc..).

La densità di probabilità $p_{\mathbf{Z}}$ è data allora da:

$$\begin{aligned} p_{\mathbf{Z}}(z_1, z_2, \dots, z_n) &= p_{\mathbf{X}}(x_1, x_2, \dots, x_n) |J| \\ &= p_{\mathbf{X}}(f_1^{-1}(\mathbf{z}), f_2^{-1}(\mathbf{z}), \dots, f_n^{-1}(\mathbf{z})) |J|, \end{aligned} \tag{5.20}$$

dove $|J|$ è lo Jacobiano, definito come il valore assoluto del determinante:

$$|J| = \begin{vmatrix} \partial f_1^{-1}/\partial z_1 & \partial f_1^{-1}/\partial z_2 & \dots & \partial f_1^{-1}/\partial z_n \\ \partial f_2^{-1}/\partial z_1 & \partial f_2^{-1}/\partial z_2 & \dots & \partial f_2^{-1}/\partial z_n \\ \dots & \dots & \dots & \dots \\ \partial f_n^{-1}/\partial z_1 & \partial f_n^{-1}/\partial z_2 & \dots & \partial f_n^{-1}/\partial z_n \end{vmatrix}. \tag{5.21}$$

Ovviamente, la trasformazione è possibile se tutte le derivate sono continue e $|J| \neq 0$. Nel caso non vi sia un'unica trasformazione f_i ($i = 1, 2, \dots, n$) che sia

invertibile, occorre suddividere i domini delle \mathbf{X} e delle \mathbf{Z} in m sottoinsiemi tra loro disgiunti tra i quali esista una corrispondenza biunivoca e sommare la (5.20) su tali domini:

$$p_{\mathbf{Z}}(z_1, z_2, \dots, z_n) = \sum_{L=1}^m p_{\mathbf{X}}(f_{L1}^{-1}(\mathbf{z}), f_{L2}^{-1}(\mathbf{z}), \dots, f_{Ln}^{-1}(\mathbf{z})) |J_L| . \quad (5.22)$$

Il teorema può essere applicato anche al caso della trasformazione particolare (5.17). Consideriamo infatti una variabile Z funzione di 2 variabili:

$$Z = f(X_1, X_2) . \quad (5.23)$$

Per applicare il metodo del determinante jacobiano, definiamo $Z_1 \equiv Z$ e una variabile ausiliaria $Z_2 = X_2$. La (5.19) diventa:

$$Z_1 = f(X_1, X_2) , \quad Z_2 = X_2 . \quad (5.24)$$

La densità di $Z \equiv Z_1$ può essere allora trovata applicando la (5.20) ed integrando poi sulla variabile ausiliaria $Z_2 = X_2$. Dato che lo jacobiano vale:

$$|J| = \begin{vmatrix} \frac{\partial f_1^{-1}}{\partial z_1} & \frac{\partial f_1^{-1}}{\partial z_2} \\ 0 & 1 \end{vmatrix} = \left| \frac{\partial f_1^{-1}}{\partial z_1} \right| ,$$

dalla (5.20) si ottiene:

$$p_{\mathbf{Z}}(z_1, z_2) = p_{\mathbf{X}}(x_1, x_2) \left| \frac{\partial f_1^{-1}}{\partial z_1} \right| . \quad (5.25)$$

La densità della variabile Z si ottiene integrando sulla variabile ausiliaria Z_2 :

$$p_{Z_1}(z_1) = \int p_{\mathbf{Z}}(z_1, z_2) dz_2 . \quad (5.26)$$

Ricordando che $Z_1 \equiv Z$, $X_2 \equiv Z_2$ e che quindi

$$p_Z(z) \equiv p_{Z_1}(z_1) , \quad \frac{\partial f_1^{-1}}{\partial z_1} \equiv \frac{\partial f_1^{-1}}{\partial z} , \quad X_1 = f_1^{-1}(Z_1, X_2) = f_1^{-1}(Z, X_2) ,$$

possiamo scrivere la (5.26) come:

$$\begin{aligned} p_Z(z) &= \int p_{\mathbf{X}}(x_1, x_2) \left| \frac{\partial f_1^{-1}}{\partial z} \right| dx_2 \\ &= \int p_{\mathbf{X}}(f_1^{-1}(z, x_2), x_2) \left| \frac{\partial f_1^{-1}}{\partial z} \right| dx_2 , \end{aligned} \quad (5.27)$$

che rappresenta la densità di probabilità cercata.

Questa formula rappresenta una alternativa alla (5.18) per trovare la densità $p_Z(z)$ quando il legame funzionale è del tipo $Z = f(X_1, X_2)$.

Se le variabili X_1 e X_2 sono indipendenti, allora la densità $p_{\mathbf{X}}$ si fattorizza secondo la (4.6), e la (5.27) diventa:

$$p_Z(z) = \int p_{X_1}(f_1^{-1}(z, x_2)) p_{X_2}(x_2) \frac{\partial f_1^{-1}}{\partial z} dx_2 . \quad (5.28)$$

Quando la variabile Z è data dalla somma:

$$Z = X_1 + X_2 , \quad (5.29)$$

la funzione inversa f^{-1} e la sua derivata da inserire nella (5.27) sono date da:

$$X_1 = f_1^{-1}(Z, X_2) = Z - X_2 , \quad \frac{\partial f_1^{-1}}{\partial z} = 1 , \quad (5.30)$$

per cui si ottiene:

$$p_Z(z) = \int_{-\infty}^{+\infty} p_{\mathbf{X}}(z - x_2, x_2) dx_2 . \quad (5.31)$$

Se le due variabili X_1 e X_2 sono indipendenti, allora la densità di probabilità di partenza, che è del tipo (4.3), si fattorizza nella (4.6) e otteniamo l'integrale:

$$p_Z(z) = \int_{-\infty}^{+\infty} p_{X_1}(z - x_2) p_{X_2}(x_2) dx_2 , \quad (5.32)$$

che è detto *integrale di convoluzione*. Esso si incontra molte volte, sia in statistica sia in fisica sperimentale nell'analisi di misure di laboratorio (si veda il cap. 11).

Dopo tanta matematica, vi facciamo notare che quest'ultimo integrale ha una semplice spiegazione intuitiva: la probabilità di osservare un valore $Z = z$ è dato dalla probabilità di avere realizzato un valore x_2 , per la probabilità di avere un valore $x_1 = z - x_2$ che assicuri l'uguaglianza $z = x_1 + x_2$. Questa probabilità va sommata su tutti i possibili valori di X_2 . L'integrale di convoluzione appare quindi come una ulteriore applicazione delle leggi fondamentali (1.22, 1.23) per variabili continue.

Notiamo anche che la (5.31) può essere trovata anche tramite la funzione cumulativa (5.18). Infatti, in questo caso si ha:

$$\begin{aligned} P\{(X + Y \leq z)\} &= F_Z(z) = \int_{(X+Y \leq z)} p_{\mathbf{X}}(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{+\infty} dx_2 \int_{-\infty}^{z-x_2} p_{\mathbf{X}}(x_1, x_2) dx_1 , \end{aligned}$$

e, derivando rispetto a z tramite la (5.4), si riottiene la (5.31).

In SCILAB è presente la funzione `convol` che calcola la convoluzione di due funzioni date per punti. Copiando dal sito il nostro codice `Conguf` potete risolvere l'esercizio seguente e, con poche modifiche, anche gli altri esercizi ed in generale molti semplici problemi di convoluzione.

Esercizio 5.1

Trovare la densità di una variabile aleatoria

$$Z = X + Y$$

somma di due variabili indipendenti $X \sim N(\mu, \sigma^2)$ e $Y \sim U(a, b)$.

Risposta. Dalle (3.28, 3.79, 5.32) si ha immediatamente:

$$\begin{aligned} p_Z(z) &= \frac{1}{b-a} \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(z-y-\mu)^2}{2\sigma^2}\right] dy \\ &= \frac{1}{b-a} \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{[y-(z-\mu)]^2}{2\sigma^2}\right] dy . \end{aligned} \quad (5.33)$$

Questa densità non è altro che una gaussiana di media $(z-\mu)$ e varianza σ^2 integrata entro i limiti della densità uniforme di partenza.

Ricorrendo alla gaussiana cumulativa (3.43) possiamo scrivere:

$$p_Z(z) = \frac{1}{b-a} \left[\Phi\left(\frac{b-(z-\mu)}{\sigma}\right) - \Phi\left(\frac{a-(z-\mu)}{\sigma}\right) \right] . \quad (5.34)$$

Nelle misure di grandezze fisiche, spesso allo strumento viene associata una dispersione casuale di tipo uniforme, mentre alle operazioni di misura una dispersione di tipo gaussiano (vedi cap. 11). In questi casi la (5.34) fornisce la legge di dispersione totale della misura, ed è pertanto importante nello studio della propagazione degli errori, che verrà svolto nel par. 11.6. La Fig. 5.3 mostra che la forma di questa densità risultante è abbastanza simile ad una gaussiana.

Esercizio 5.2

Trovare la densità di una variabile aleatoria

$$Z = X + Y$$

somma di due variabili indipendenti $X, Y \sim U(0, 1)$.

Risposta. Poiché $0 \leq X, Y \leq 1$, ovviamente $0 \leq Z \leq 2$. Anche in questo caso, dalle (3.79, 5.32) si ha immediatamente:

$$p_Z(z) = \int u_Y(z-x) u_X(x) dx ,$$

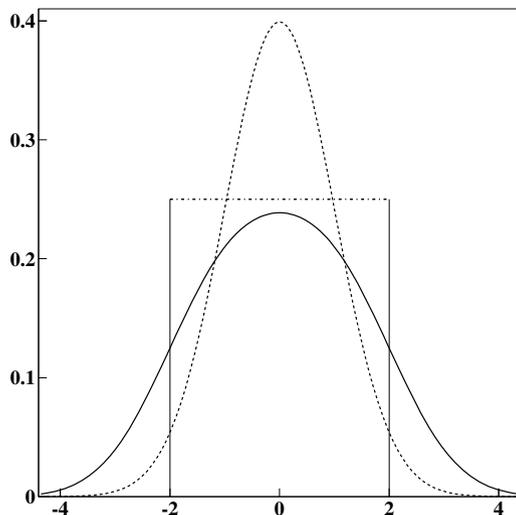


Fig. 5.3. Convoluzione di una gaussiana standard con $\mu = 0$ e $\sigma = 1$ (linea tratteggiata) con una densità uniforme limitata tra -2 e 2 (linea tratto-punto). La curva risultante è la linea piena

dove $u(x)$ rappresenta la densità uniforme:

$$u(x) = \begin{cases} 1 & \text{per } 0 \leq x \leq 1 \\ 0 & \text{altrimenti} \end{cases}$$

Gli argomenti della densità uniforme, che sono le variabili $(z - x)$ e (x) , devono quindi mantenersi limitati tra 0 e 1. L'integrale si scompone allora nei due termini:

$$p_Z(z) = \left[\int_0^z dx \right]_{0 \leq z \leq 1} + \left[\int_{z-1}^1 dx \right]_{1 < z \leq 2},$$

ed ha come risultato:

$$p_Z(z) = \begin{cases} z & \text{per } 0 \leq z \leq 1 \\ 2 - z & \text{per } 1 < z \leq 2 \\ 0 & \text{altrimenti} \end{cases}. \quad (5.35)$$

Questa densità è normalizzata, compresa tra 0 e 2, di forma triangolare e con massimo in $z = 1$. Anche questa funzione verrà discussa in dettaglio nello studio della propagazione degli errori di due misure affette da errori strumentali, che verrà svolto nel par. 11.6.

Esercizio 5.3

Trovare la densità della variabile

$$Z = X + Y ,$$

dove $X, Y \sim N(\mu, \sigma^2)$ sono due variabili gaussiane indipendenti.

Risposta. Anche in questo caso, dalle (3.28, 5.32) otteniamo immediatamente:

$$p_Z(z) = \frac{1}{2\pi\sigma_x\sigma_y} \int_{-\infty}^{+\infty} \exp \left[-\frac{(x - \mu_x)^2}{2\sigma_x^2} - \frac{(z - x - \mu_y)^2}{2\sigma_y^2} \right] dx .$$

L'integrale che compare in questa formula può essere risolto con i metodi indicati nell'esercizio 3.4. Esso è del tipo:

$$\int_{-\infty}^{+\infty} e^{-Ax^2+2Bx-C} dx = \sqrt{\frac{\pi}{A}} e^{-\frac{AC-B^2}{A}} ,$$

dove:

$$A = \frac{1}{2} \frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2 \sigma_y^2} , \quad B = \frac{\mu_x}{2\sigma_x^2} + \frac{z - \mu_y}{2\sigma_y^2} , \quad C = \frac{\mu_x^2}{2\sigma_x^2} + \frac{(z - \mu_y)^2}{2\sigma_y^2} . \quad (5.36)$$

Si ottiene pertanto la densità:

$$p_Z(z) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma_x^2 + \sigma_y^2}} \exp \left[-\frac{[z - (\mu_x + \mu_y)]^2}{2(\sigma_x^2 + \sigma_y^2)} \right] , \quad (5.37)$$

la quale è una gaussiana di media e deviazione standard date da

$$\mu_z = \mu_x + \mu_y , \quad \sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2} . \quad (5.38)$$

Poiché la trasformazione è lineare e le variabili sono indipendenti, le (5.38) sono in accordo con le (4.8, 4.19). L'esercizio ci dice però un fatto nuovo e molto importante: la composizione lineare di variabili gaussiane dà luogo ancora a densità gaussiane.

La (5.37) può essere facilmente dimostrata anche utilizzando la proprietà (B.4) delle funzioni generatrici dell'Appendice B.

Quando una densità conserva la sua forma funzionale per composizione lineare di più variabili, si dice che essa è stabile. È utile a questo punto ricordare il teorema Limite Centrale 3.1, secondo il quale la somma di N variabili aleatorie qualsiasi si distribuisce secondo la gaussiana per N abbastanza grande, in pratica per $N > 10$. Ebbene, se le variabili di partenza sono già gaussiane, questa condizione può essere rimossa e la proprietà vale per qualunque N . L'insieme di queste proprietà sta alla base del ruolo centrale che la densità gaussiana o normale assume sia nel calcolo delle probabilità sia in statistica.

Esercizio 5.4

Trovare la densità della variabile

$$Z = X + Y ,$$

dove X e Y sono due variabili intere poissoniane indipendenti rispettivamente di media μ_1 e μ_2 .

Risposta. La (5.32), per variabili intere $X, Y \geq 0$, va riscritta come:

$$p_Z(z) = \sum_{x=0}^z p_X(x) p_Y(z-x) , \quad (5.39)$$

dove la densità p_X, p_Y è la poissoniana (3.14). Si ha pertanto:

$$p_Z(z) = \sum_{x=0}^z \frac{\mu_1^x \mu_2^{z-x}}{x! (z-x)!} e^{-(\mu_1 + \mu_2)} .$$

Moltiplicando e dividendo per $z!$ e ricordando la formula del binomio di Newton:

$$(\mu_1 + \mu_2)^z = \sum_{k=0}^z \frac{z!}{k!(z-k)!} \mu_1^k \mu_2^{z-k} ,$$

si ottiene:

$$p_Z(z) = e^{-(\mu_1 + \mu_2)} \frac{1}{z!} \sum_{x=0}^z \frac{z!}{x!(z-x)!} \mu_1^x \mu_2^{z-x} = \frac{(\mu_1 + \mu_2)^z}{z!} e^{-(\mu_1 + \mu_2)} , \quad (5.40)$$

da cui risulta che la densità cercata è una poissoniana di media $(\mu_1 + \mu_2)$. Notiamo che, diversamente dal caso gaussiano, solo la somma, ma non la differenza di variabili poissoniane è poissoniana. Infatti, se $Z = Y - X$, è possibile avere $Z \leq 0$ e nella (5.40) compare il termine $(z+x)!$ invece del termine $(z-x)!$. È chiaro allora che la distribuzione della differenza è diversa da quella di Poisson.

Nei due esercizi che seguono determineremo le densità di Student e di Snedecor, di cui ci serviremo più avanti in statistica. Quindi, non saltate gli esercizi (leggetene almeno l'enunciato e la soluzione) e fate attenzione.

Esercizio 5.5

Trovare la densità di probabilità di una variabile Z formata dal rapporto tra una variabile gaussiana standard e la radice quadrata di una variabile Q_R di densità χ^2 ridotto con ν gradi di libertà. La variabile gaussiana e la variabile Q_R siano tra loro indipendenti.

Risposta. Denotiamo con X la variabile gaussiana standard di densità (3.42) e con Y la variabile χ^2 di densità (3.67). Dobbiamo allora trovare la densità della variabile

$$Z = \frac{X}{\sqrt{Y/\nu}} = \sqrt{\nu} \frac{X}{\sqrt{Y}}. \quad (5.41)$$

Poiché le variabili sono indipendenti, possiamo applicare la (5.28), dove

$$x = f^{-1}(z, y) = \frac{1}{\sqrt{\nu}} z \sqrt{y}, \quad \frac{\partial f^{-1}}{\partial z} = \frac{\sqrt{y}}{\sqrt{\nu}}.$$

Abbiamo allora, utilizzando il prodotto delle densità (3.42, 3.67):

$$p_Z(z) = \frac{1}{\sqrt{2\pi\nu} 2^{\nu/2} \Gamma(\frac{\nu}{2})} \int_0^\infty y^{\frac{\nu-1}{2}} \exp\left[-\frac{1}{2} y \left(\frac{z^2}{\nu} + 1\right)\right] dy.$$

Operiamo ora nell'integrale il cambio di variabile:

$$q = \frac{1}{2} y \left(\frac{z^2}{\nu} + 1\right), \quad y = \frac{2}{\left(\frac{z^2}{\nu} + 1\right)} q, \quad dy = \frac{2}{\left(\frac{z^2}{\nu} + 1\right)} dq,$$

ed otteniamo:

$$p_Z(z) = \frac{1}{\sqrt{2\pi\nu} 2^{\nu/2} \Gamma(\frac{\nu}{2})} \frac{2^{\nu/2}}{\sqrt{2}} \frac{1}{\left(\frac{z^2}{\nu} + 1\right)^{(\nu-1)/2}} \frac{2}{\left(\frac{z^2}{\nu} + 1\right)} \times \int_0^\infty q^{\frac{\nu-1}{2}} e^{-q} dq.$$

Ricordando l'espressione integrale (3.64) della funzione gamma, ed operando le necessarie semplificazioni algebriche, otteniamo infine:

$$p_Z(z) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{\pi}} \frac{1}{\sqrt{\nu}} \left(\frac{z^2}{\nu} + 1\right)^{-\frac{\nu+1}{2}}.$$

Questa è la famosa densità t di Student. Generalmente essa viene scritta come densità di una variabile t , ponendo anche $\sqrt{\pi} = \Gamma(1/2)$, come risulta dalle (3.65).

Ecco pertanto il risultato finale, relativo a una variabile t espressa come rapporto tra una variabile gaussiana standard e la radice quadrata di una variabile $\chi_R^2(\nu)$.

$$s_\nu(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{1}{2}) \Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\nu}} \left(\frac{t^2}{\nu} + 1\right)^{-\frac{\nu+1}{2}}. \quad (5.42)$$

I valori integrali della densità di Student sono riportati nella Tab. D.2 in Appendice D a pagina 522. Dalla tabella e dalla Fig. 5.4 potete verificare che, per un numero di gradi di libertà $> 20 - 30$, la densità coincide praticamente con una gaussiana.

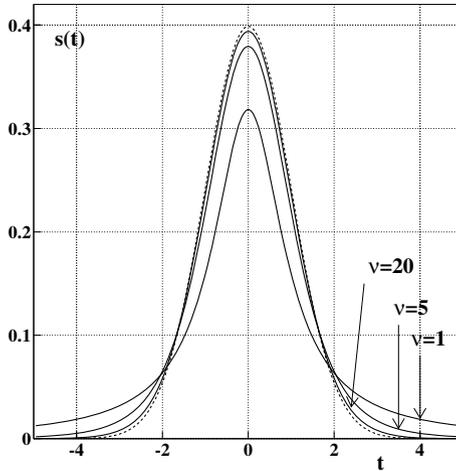


Fig. 5.4. Densità di Student per 1, 5 e 20 gradi di libertà. La linea tratteggiata è la gaussiana standard

I valori della media e della varianza della distribuzione si possono ricavare, come sempre, dalle (2.53, 2.56), ottenendo:

$$\mu = 0 \quad \sigma^2 = \frac{\nu}{\nu - 2}, \tag{5.43}$$

da cui risulta che la varianza è definita solo per $\nu > 2$. Per $\nu \leq 2$ la funzione $s_\nu(t)$ è un esempio, insolito ma non impossibile, di densità che non possiede varianza. In questo caso la parametrizzazione dell'intervallo di probabilità (3.94) in termini di deviazione standard non è più possibile, ed è necessario, per ricavare i livelli di probabilità richiesti da un problema, calcolare direttamente l'integrale della densità entro i limiti assegnati. Queste probabilità si possono anche ricavare dalla Tab. D.2.

In SCILAB esiste la funzione `cdfst` per il calcolo della densità di Student. Le sequenze di chiamata sono le seguenti:

```
[p, q]=cdfst(PQ, t, Df)
[t] =cdfst(T, Df, p, q)
[Df] =cdfst(Df, p, q, t)
```

Con la chiamata "PQ" viene calcolato l'integrale p (ed il suo complemento $q = 1-p$) tra $-\infty$ ed il valore t nel caso di Df gradi di libertà. Con la chiamata "T" viene calcolato il quantile t corrispondente alla probabilità p, mentre con la chiamata "Df" vengono trovati i gradi di libertà corrispondenti ai parametri indicati nella funzione.

Esercizio 5.6

Trovare la densità di probabilità di una variabile F formata dal rapporto tra due variabili indipendenti χ^2 ridotto con μ e ν gradi di libertà:

$$\mathbf{F} = \frac{Q_R(\mu)}{Q_R(\nu)}. \quad (5.44)$$

Poiché la tradizione vuole che si indichi il rapporto (5.44) sempre in lettere maiuscole, indicheremo con F i valori assunti dalla variabile di Snedecor \mathbf{F} .

Risposta. Per non appesantire la notazione, definiamo:

$$\mathbf{F} = \frac{Q_R(\mu)}{Q_R(\nu)} \equiv \frac{Y}{X} \equiv f(X, Y)$$

$$Y = f^{-1}(\mathbf{F}, X) = \mathbf{F}X, \quad \frac{\partial f^{-1}}{\partial \mathbf{F}} = X.$$

La densità χ^2 ridotto con ν gradi di libertà è data dall'integrando della (3.72):

$$p_\nu(x) = a_\nu x^{\frac{1}{2}(\nu-2)} e^{-\frac{1}{2}\nu x} dx,$$

dove:

$$a_\nu = \frac{\nu^{\frac{\nu}{2}}}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})}. \quad (5.45)$$

In questo caso la relazione base (5.28) diventa:

$$\begin{aligned} p_{\mu\nu}(F) &= \int_0^\infty p_\mu(f^{-1}(Fx)) p_\nu(x) \frac{\partial f^{-1}}{\partial F} dx \\ &= a_\mu a_\nu \int F^{\mu/2-1} x^{\mu/2-1} e^{-\mu Fx/2} x^{\nu/2-1} e^{-\nu x/2} x dx \\ &= a_\mu a_\nu F^{\mu/2-1} \int x^{\frac{1}{2}(\mu+\nu-2)} e^{-\frac{1}{2}(\mu F + \nu)x} dx. \end{aligned}$$

Eseguendo ora il cambio di variabile:

$$\omega = \frac{x}{2}(\mu F + \nu), \quad dx = \frac{2}{\mu F + \nu} d\omega,$$

otteniamo infine:

$$p_{\mu\nu}(F) = \frac{a_\mu a_\nu F^{\mu/2-1} 2^{\frac{1}{2}(\mu+\nu)}}{(\mu F + \nu)^{\frac{1}{2}(\mu+\nu)}} \int \omega^{\frac{1}{2}(\mu+\nu)-1} e^{-\omega} d\omega.$$

Ricordando la (5.45) e notando che l'integrale non è altro che la funzione $\Gamma[(\mu + \nu)/2]$ in accordo con la (3.65), otteniamo finalmente:

$$p_{\mu\nu}(F) = c_{\mu\nu} F^{\frac{1}{2}(\mu-2)} (\mu F + \nu)^{-\frac{1}{2}(\mu+\nu)}, \quad (5.46)$$

$$c_{\mu\nu} = \mu^{\frac{\mu}{2}} \nu^{\frac{\nu}{2}} \frac{\Gamma(\frac{\mu+\nu}{2})}{\Gamma(\frac{\mu}{2}) \Gamma(\frac{\nu}{2})},$$

che rappresenta la famosa densità F di Snedecor per la variabile \mathbf{F} . Essa è riportata in Fig. 5.5. La media e la varianza si calcolano come al solito dalle (2.53, 2.56), e sono date da:

$$\langle \mathbf{F} \rangle = \frac{\nu}{\nu - 2}, \quad \text{Var}[\mathbf{F}] = \frac{2\nu^2(\mu + \nu - 2)}{\mu(\nu - 2)^2(\nu - 4)}. \quad (5.47)$$

Queste due equazioni sono definite per $\nu > 2$ e $\nu > 4$, rispettivamente. Per i gradi di libertà $\mu, \nu \rightarrow \infty$ la densità F tende ad essere gaussiana; tuttavia, come si vede dalla figura, la convergenza verso questa forma è piuttosto lenta. I valori del rapporto $F \equiv F_u(\mu, \nu)$ corrispondenti ai percentili del 95% ($u = 0.95$) e del 99% ($u = 0.99$) e quindi ai livelli di significatività del 5% e dell'1% a destra della media,

$$u = \int_0^{F_u(\mu, \nu)} p_{\mu\nu}(F) \, dF, \quad (5.48)$$

sono molto usati in statistica nell'analisi della varianza. Essi sono riportati nelle tabelle D.5 e D.6 dell'Appendice D, alle pagine 525 e 526. I percentili ($u = 5\%$) e ($u = 1\%$), coincidenti coi livelli di significatività delle code a sinistra della media, in genere non vengono dati, perché tra i valori quantili sussiste la relazione:

$$F_u(\mu, \nu) = \frac{1}{F_{1-u}(\nu, \mu)}. \quad (5.49)$$

La (5.49) può essere dimostrata notando che, per definizione:

$$1 - u = \int_0^{F_{1-u}(\mu, \nu)} p_{\mu\nu}(F) \, dF = \int_{F_u(\mu, \nu)}^{\infty} p_{\mu\nu}(F) \, dF$$

e che, dato che la variabile F è un rapporto:

$$\text{se } F \text{ segue } p_{\mu, \nu}(F) \quad \text{allora} \quad \frac{1}{F} \text{ segue } p_{\nu, \mu}(F).$$

Si può dunque scrivere:

$$\begin{aligned} u &= P\{\mathbf{F} \leq F_u(\mu, \nu)\} = P\left\{\frac{1}{\mathbf{F}} \leq F_u(\nu, \mu)\right\} \\ &= P\left\{\frac{1}{\mathbf{F}} \geq F_{1-u}(\nu, \mu)\right\} = P\left\{\mathbf{F} \leq \frac{1}{F_{1-u}(\nu, \mu)}\right\}, \end{aligned}$$

da cui la (5.49).

In SCILAB esiste la funzione `cdf` per il calcolo della densità F . Le sequenze di chiamata sono le seguenti:

```
[p, q]=cdf(PQ, f, Dfn, Dfd)
[F]=cdf(F, Dfn, Dfd, p, q)
[Dfn]=cdf(Dfn, Dfd, p, q, f)
[Dfd]=cdf(Dfd, p, q, f, Dfn)
```

Con la chiamata "PQ" viene calcolato l'integrale p (ed il suo complemento $q = 1-p$) tra $-\infty$ ed il valore f nel caso di Dfn e Dfn gradi di libertà. Con

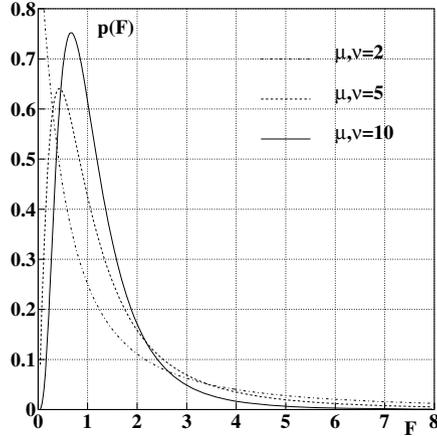


Fig. 5.5. Densità della variabile F di Snedecor data dal rapporto di due variabili χ^2 ridotto aventi rispettivamente μ e ν gradi di libertà

la chiamata “ F ” viene calcolato il quantile τ corrispondente alla probabilità p , mentre con le chiamate “ Dfn ” e “ Dfd ” vengono trovati i gradi di libertà corrispondenti ai parametri indicati nella funzione.

Sul problema delle funzioni di variabili aleatorie si possono scrivere dei libri (ed infatti sono stati scritti...). Se vi piace l’analisi e volete approfondire ulteriormente l’argomento, potete consultare il testo di Papoulis [Pap77].

5.4 La trasformazione della media e della varianza

Come avete visto, la determinazione della densità di probabilità di una variabile espressa come funzione di altre variabili aleatorie è un argomento piuttosto complesso. Negli esempi che abbiamo fin qui sviluppato, abbiamo già incontrato discrete complicazioni matematiche, pur essendoci limitati al caso più semplice di due sole variabili indipendenti.

Fortunatamente, come nel calcolo delle probabilità elementare, quasi sempre una soluzione approssimata ma soddisfacente del problema può essere ottenuta determinando, invece che le forme funzionali complete, solamente il valore centrale (media) e la dispersione (deviazione standard) delle funzioni densità coinvolte. Questo è quanto ora ci proponiamo di fare.

Cominciamo dalla situazione più semplice, quella di una variabile Z funzione di una sola variabile aleatoria X

$$Z = f(X) \tag{5.50}$$

di densità nota $p_X(x)$.

Supponendo che f sia invertibile, $X = f^{-1}(Z)$ e la media della variabile Z si ottiene dalla (2.53), dalla (5.7) e differenziando la (5.50):

$$\begin{aligned} \langle Z \rangle &= \int z p_Z(z) dz = \int f(x) p_X(x) \frac{dz}{|f'(x)|} \\ &= \int f(x) p_X(x) dx . \end{aligned} \quad (5.51)$$

Risulta pertanto che la media di Z è data dalla media di $f(X)$ sulla densità di partenza $p_X(x)$, in accordo con la (2.67), che è la definizione di valore atteso. Questo risultato, valido anche nel caso multidimensionale sotto le condizioni del teorema 5.1, consente di ottenere in modo corretto e abbastanza semplice il valore centrale della densità di Z .

In molti casi si ricorre però ad una formula approssimata, che si ottiene sviluppando al secondo ordine in serie di Taylor la funzione $f(x)$ nell'intorno del valor medio μ della variabile di partenza X :

$$f(x) \simeq f(\mu) + f'(\mu)(x - \mu) + \frac{1}{2} f''(\mu)(x - \mu)^2 . \quad (5.52)$$

Inserendo questo sviluppo nella (5.51), si verifica facilmente che il termine associato alla derivata prima si annulla e che vale pertanto la relazione approssimata:

$$\langle Z \rangle \simeq f(\mu) + \frac{1}{2} f''(\mu) \sigma^2 , \quad (5.53)$$

dove σ è la deviazione standard di X .

Questa importante relazione mostra che *la media della funzione f (cioè di Z) è pari alla funzione della media più un termine correttivo che dipende dalla concavità della funzione nell'intorno della media e dalla varianza di X .*

Se la relazione (5.50) è di tipo lineare,

$$Z = aX + b ,$$

allora la derivata seconda nella (5.53) si annulla e si ottiene:

$$\langle Z \rangle = f(\langle X \rangle) . \quad (5.54)$$

Nel caso non lineare, si può mostrare che $\langle Z \rangle \geq f(\langle X \rangle)$ se $f''(\langle X \rangle) > 0$, mentre $\langle Z \rangle \leq f(\langle X \rangle)$ se $f''(\langle X \rangle) < 0$.

Veniamo ora alla trasformazione della varianza. Analogamente alla (5.51) si può scrivere:

$$\text{Var}[Z] = \int [f(x) - \langle f(X) \rangle]^2 p_X(x) dx . \quad (5.55)$$

Ricordando il risultato approssimato (5.53) ed utilizzando per la funzione f lo sviluppo al secondo ordine (5.52), possiamo scrivere:

$$\begin{aligned}\text{Var}[Z] &\simeq \int \left[f(x) - f(\mu) - \frac{1}{2}f''(\mu)\sigma^2 \right]^2 p_X(x) dx \\ &= \int \left[f'(\mu)(x - \mu) + \frac{1}{2}f''(\mu)(x - \mu)^2 - \frac{1}{2}f''(\mu)\sigma^2 \right]^2 p_X(x) dx .\end{aligned}$$

Svolgendo il quadrato nell'integrando e ricordando la definizione (2.58) dei momenti di una distribuzione, dopo una rielaborazione un po' lunga ma facile si ottiene:

$$\text{Var}[Z] \simeq [f'(\mu)]^2 \sigma^2 + \frac{1}{4} [f''(\mu)]^2 (\Delta_4 - \sigma^4) + f'(\mu)f''(\mu) \Delta_3 , \quad (5.56)$$

dove Δ_i sono i momenti definiti nella (2.58) di pagina 59. Dato che questa è comunque una relazione approssimata, spesso è sufficiente conoscere solo l'ordine di grandezza dei momenti per poter effettuare una stima accettabile della varianza di Z . Se la densità di X è simmetrica intorno alla media, allora $\Delta_3 = 0$. Se, oltre ad essere simmetrica, la densità è gaussiana, allora, in base alla (3.33), $\Delta_4 = 3\sigma^4$ e la (5.56) diventa:

$$\text{Var}[Z] \simeq [f'(\mu)]^2 \sigma^2 + \frac{1}{2} [f''(\mu)]^2 \sigma^4 . \quad (5.57)$$

Se la densità $p_X(x)$ è simmetrica ma non gaussiana, in genere non si commette un grande errore utilizzando comunque la (5.57).

Quando la deviazione standard di $p_X(x)$ è piccola, cioè quando $\sigma^2 \gg \sigma^4$, è valida l'ulteriore approssimazione:

$$\text{Var}[Z] \simeq [f'(\mu)]^2 \sigma^2 , \quad (5.58)$$

che diventa esatta quando tra Z ed X esiste una relazione lineare. In quest'ultimo caso le (5.54, 5.58) coincidono con le (2.62, 2.63).

È utile anche verificare che quando $Z = aX^2$ e X segue la densità gaussiana, le (5.53, 5.57) forniscono il risultato corretto (5.16).

Trattiamo ora il caso più generale, costituito da una variabile Z funzione di n variabili X_i :

$$Z = f(X_1, X_2, \dots, X_n) \equiv f(\mathbf{X}) . \quad (5.59)$$

Questo caso rimane trattabile in modo semplice se si utilizza l'approssimazione lineare delle (5.54, 5.58). In altre parole, si suppone che *la media della funzione f coincida con la funzione delle medie e che la varianza di Z dipenda solo dalle derivate prime di f e dalle varianze delle distribuzioni delle \mathbf{X} .*

Per cominciare, consideriamo il caso semplice di due variabili,

$$Z = f(X_1, X_2)$$

e linearizziamo la funzione f nell'intorno delle medie μ_1, μ_2 delle due variabili di partenza X_1, X_2 :

$$\begin{aligned} z &\simeq f(\mu_1, \mu_2) + \frac{\partial f}{\partial x_1}(x_1 - \mu_1) + \frac{\partial f}{\partial x_2}(x_2 - \mu_2) \\ &\equiv f(\mu_1, \mu_2) + \frac{\partial f}{\partial x_1} \Delta x_1 + \frac{\partial f}{\partial x_2} \Delta x_2, \end{aligned} \quad (5.60)$$

dove le derivate sono calcolate in $x_1 = \mu_1, x_2 = \mu_2$.

La media di Z si ottiene considerando la (5.51) estesa a due variabili:

$$\langle Z \rangle = \int f(x_1, x_2) p_{\mathbf{X}}(x_1, x_2) dx_1 dx_2, \quad (5.61)$$

e sostituendo ad $f(x_1, x_2)$ lo sviluppo (5.60). Poiché i termini del tipo $(x_i - \mu_i) p_{\mathbf{X}}(x_1, x_2)$ si annullano integrando, il risultato finale è semplicemente, come sempre nell'approssimazione lineare, che la media della funzione coincide con la funzione delle medie:

$$\langle Z \rangle = f(\mu_1, \mu_2). \quad (5.62)$$

La generalizzazione al caso di n variabili è ovvia:

$$\langle Z \rangle = f(\mu_1, \mu_2, \dots, \mu_n). \quad (5.63)$$

La varianza di Z si trova invece considerando la generalizzazione della (5.55):

$$\text{Var}[Z] = \int [f(x_1, x_2) - f(\mu_1, \mu_2)]^2 p_{\mathbf{X}}(x_1, x_2) dx_1 dx_2, \quad (5.64)$$

e sostituendo ad $f(x_1, x_2) - f(\mu_1, \mu_2)$ lo sviluppo (5.60). Il risultato contiene ora delle novità interessanti:

$$\begin{aligned} \sigma_z^2 &\simeq \left[\frac{\partial f}{\partial x_1} \right]^2 \int (\Delta x_1)^2 p_{\mathbf{X}}(x_1, x_2) dx_1 dx_2 \\ &\quad + \left[\frac{\partial f}{\partial x_2} \right]^2 \int (\Delta x_2)^2 p_{\mathbf{X}}(x_1, x_2) dx_1 dx_2 \\ &\quad + 2 \left[\frac{\partial f}{\partial x_1} \frac{\partial f}{\partial x_2} \right] \int \Delta x_1 \Delta x_2 p_{\mathbf{X}}(x_1, x_2) dx_1 dx_2 \\ &= \left[\frac{\partial f}{\partial x_1} \right]^2 \sigma_1^2 + \left[\frac{\partial f}{\partial x_2} \right]^2 \sigma_2^2 + 2 \left[\frac{\partial f}{\partial x_1} \frac{\partial f}{\partial x_2} \right] \sigma_{12}. \end{aligned} \quad (5.65)$$

La varianza di z , nell'approssimazione lineare, risulta funzione delle varianze σ_i^2 delle singole variabili, della loro covarianza σ_{12} , e delle derivate calcolate nei punti $x_1 = \mu_1, x_2 = \mu_2$.

Questa legge generalizza la (4.19) di pagina 121. Se le due variabili sono indipendenti, esse hanno covarianza nulla e si ottiene:

$$\sigma_{12} = 0 \implies \sigma_z^2 = \left[\frac{\partial f}{\partial x_1} \right]^2 \sigma_1^2 + \left[\frac{\partial f}{\partial x_2} \right]^2 \sigma_2^2 . \quad (5.66)$$

Quando $Z = X_1 + X_2$ è data dalla somma di due variabili indipendenti, la densità risultante è data dall'integrale di convoluzione (5.32) e il calcolo esplicito di μ_z e σ_z^2 può risultare molto complicato, se lo sono le densità delle variabili di partenza. Tuttavia, in questo caso le (5.62) e (5.66) sono esatte e forniscono il risultato:

$$\mu_z = \mu_1 + \mu_2 , \quad \sigma_z^2 = \sigma_1^2 + \sigma_2^2 ,$$

che permette di *conoscere esattamente la media e la varianza di Z*, anche se la forma esplicita della densità corrispondente non è nota o è troppo complicata da calcolare.

Nel caso di trasformazioni lineari, è quindi sempre possibile, utilizzando i criteri dell'intervallo di probabilità e della legge 3σ illustrati nei parr. 3.6, 3.11, valutare in modo approssimato, ma estremamente semplice e in molti casi più che soddisfacente, la dispersione dei valori di z intorno al valor medio.

Le (5.62, 5.66) forniscono in genere buoni risultati (ma approssimati!) anche quando Z è il prodotto o il rapporto di due variabili indipendenti. In questo caso la (5.66) fornisce, sia per il prodotto sia per il rapporto, l'unico risultato (verificatelo come esercizio):

$$Z = X_1 X_2 , \quad Z = \frac{X_1}{X_2} , \quad Z = \frac{X_2}{X_1} \implies \frac{\text{Var}[Z]}{\langle Z \rangle^2} = \frac{\text{Var}[X_1]}{\langle X_1 \rangle^2} + \frac{\text{Var}[X_2]}{\langle X_2 \rangle^2} , \quad (5.67)$$

il quale mostra che la varianza relativa di Z (detta anche quadrato del coefficiente di variazione σ/μ) è la somma delle varianze relative delle variabili di partenza. Tuttavia, questo è un risultato approssimato, come mostrato nell'esercizio che segue.

Esercizio 5.7

Trovare la formula esatta della varianza per il prodotto XY di due variabili aleatorie indipendenti.

Risposta. *Se le due variabili sono indipendenti, sappiamo, dalla (4.9), che la media del prodotto è il prodotto delle medie. La varianza del prodotto è data allora da:*

$$\begin{aligned} \text{Var}[XY] &= \int (xy - \mu_x \mu_y)^2 p_X(x) p_Y(y) \, dx \, dy \\ &= \int (x^2 y^2 + \mu_x^2 \mu_y^2 - 2\mu_x \mu_y xy) p_X(x) p_Y(y) \, dx \, dy \\ &= \langle x^2 \rangle \langle y^2 \rangle + \mu_x^2 \mu_y^2 - 2\mu_x^2 \mu_y^2 = \langle x^2 \rangle \langle y^2 \rangle - \mu_x^2 \mu_y^2 . \end{aligned}$$

Tenendo presente la (2.66) possiamo scrivere:

$$\begin{aligned}\text{Var}[XY] &= (\sigma_x^2 + \mu_x^2)(\sigma_y^2 + \mu_y^2) - \mu_x^2 \mu_y^2 \\ &= \sigma_x^2 \sigma_y^2 + \mu_x^2 \sigma_y^2 + \mu_y^2 \sigma_x^2,\end{aligned}\quad (5.68)$$

che è la soluzione del problema.

Possiamo confrontare questa equazione con la (5.67) se ne dividiamo ambo i membri per il prodotto delle medie al quadrato:

$$\frac{\sigma_z^2}{\mu_z^2} = \frac{\sigma_x^2}{\mu_x^2} + \frac{\sigma_y^2}{\mu_y^2} + \frac{\sigma_x^2 \sigma_y^2}{\mu_x^2 \mu_y^2}.\quad (5.69)$$

Quest'ultima equazione mostra che la (5.67) è valida solo si verifica la condizione

$$\frac{\sigma_x^2}{\mu_x^2} + \frac{\sigma_y^2}{\mu_y^2} \gg \frac{\sigma_x^2 \sigma_y^2}{\mu_x^2 \mu_y^2},\quad (5.70)$$

il che avviene solo se le varianze relative sono piccole.

Torniamo ora alla (5.65) e notiamo che essa può essere espressa nella forma matriciale:

$$\sigma_z^2 \simeq \frac{\partial f / \partial x_1 \quad \partial f / \partial x_2}{\left| \begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{array} \right|} \left| \begin{array}{c} \partial f / \partial x_1 \\ \partial f / \partial x_2 \end{array} \right| \equiv T V T^\dagger,\quad (5.71)$$

dove V è la matrice simmetrica delle covarianze (4.63) scritta per il caso bidimensionale, T è la matrice delle derivate, detta anche gradiente o matrice del trasporto e \dagger indica l'operazione di trasposizione. Si può interpretare questa equazione dicendo che le varianze e covarianze delle variabili di partenza vengono trasformate o "trasportate" dalle matrici delle derivate, attraverso la funzione f , per ottenere la dispersione della variabile Z .

La (5.71) è immediatamente generalizzabile al caso n -dimensionale contemplato nella (5.59):

$$\begin{aligned}\sigma_z^2 \simeq \frac{\partial f / \partial x_1 \quad \partial f / \partial x_2 \quad \dots \quad \partial f / \partial x_n}{\left| \begin{array}{cccc} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{array} \right|} \left| \begin{array}{c} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ \dots \\ \partial f / \partial x_n \end{array} \right| \\ \equiv T V T^\dagger.\end{aligned}\quad (5.72)$$

Se tutte le variabili sono indipendenti, le covarianze sono nulle e la (5.72) fornisce direttamente la generalizzazione della (5.66):

$$\sigma_z^2 \simeq \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)_{x_i=\mu_i}^2 \sigma_i^2, \tag{5.73}$$

dove le derivate sono calcolate, come sempre, nei valori medi delle X_i .

Ricordiamo che le formule (5.60 - 5.73) sono esatte solo nel caso di trasformazioni lineari del tipo $Z = b + \sum a_i X_i$ con a e b coefficienti costanti. Esse però forniscono risultati abbastanza accurati anche nei casi non lineari, quando le densità coinvolte sono abbastanza simmetriche, il numero delle variabili di partenza è elevato e le varianze relative sono piccole.

Per i casi complicati fortunatamente esiste un metodo, basato sulle tecniche di simulazione, che permette di effettuare in modo molto semplice ed efficace il calcolo della varianza di una funzione di più variabili. Esso sarà descritto in dettaglio nel par. 7.9.

5.5 Medie e varianze per n variabili

Trattiamo ora il caso più generale, quello di m variabili Z_k funzioni di n variabili X_i :

$$\begin{aligned} Z_1 &= f_1(X_1, X_2, \dots, X_n) \\ Z_2 &= f_2(X_1, X_2, \dots, X_n) \\ \dots &= \dots\dots\dots \\ Z_m &= f_m(X_1, X_2, \dots, X_n) . \end{aligned} \tag{5.74}$$

Restando sempre nell'ambito dell'approssimazione lineare, le m medie delle variabili Z sono date ovviamente da:

$$\begin{aligned} \langle Z_1 \rangle &= f_1(\mu_1, \mu_2, \dots, \mu_n) \\ \langle Z_2 \rangle &= f_2(\mu_1, \mu_2, \dots, \mu_n) \\ \dots &= \dots\dots\dots \\ \langle Z_m \rangle &= f_m(\mu_1, \mu_2, \dots, \mu_n) . \end{aligned} \tag{5.75}$$

Per determinare le varianze (e le covarianze!) delle variabili Z_k , occorre generalizzare la (5.72). La procedura non presenta difficoltà concettuali: si deve partire dalla matrice delle covarianze (4.63) delle variabili \mathbf{X} , effettuare il prodotto righe per colonne con le matrici del trasporto T e T^\dagger ed ottenere la matrice delle covarianze delle variabili \mathbf{Z} :

$$\begin{aligned} V(\mathbf{Z}) &\simeq T V(\mathbf{X}) T^\dagger , \\ \text{Cov}[Z_i, Z_k] &\simeq \sum_{j=1}^n \sum_{l=1}^n T_{ij} \sigma_{jl} T_{lk}^\dagger , \quad (i, k = 1, 2, \dots, m) . \end{aligned} \tag{5.76}$$

Le matrici $V(\mathbf{X})$ e $V(\mathbf{Z})$ sono matrici quadrate e simmetriche di dimensione $n \times n$ e $m \times m$ rispettivamente, mentre la matrice T del trasporto è una matrice $m \times n$ data da:

$$T = \begin{pmatrix} \partial f_1 / \partial x_1 & \partial f_1 / \partial x_2 & \dots & \partial f_1 / \partial x_n \\ \partial f_2 / \partial x_1 & \partial f_2 / \partial x_2 & \dots & \partial f_2 / \partial x_n \\ \dots & \dots & \dots & \dots \\ \partial f_m / \partial x_1 & \partial f_m / \partial x_2 & \dots & \partial f_m / \partial x_n \end{pmatrix}, \quad (5.77)$$

mentre T^\dagger è la matrice $n \times m$ trasposta di T . Ricordiamo che, per la definizione 4.3, $\text{Cov}[X_i, X_i] \equiv \text{Var}[X_i] = \sigma_{ii}$, $\text{Cov}[X_i, X_j] = \sigma_{ij}$.

Se introduciamo la notazione compatta:

$$\frac{\partial f_i}{\partial x_k} \equiv (\partial_k f_i) \quad (5.78)$$

possiamo scrivere la (5.76) come:

$$\text{Cov}[Z_i, Z_k] \simeq \sum_{j,l=1}^n (\partial_j f_i) \sigma_{jl} (\partial_l f_k). \quad (5.79)$$

Questa equazione permette di calcolare in modo abbastanza semplice le dispersioni delle variabili Z e le loro eventuali covarianze, ovvero le loro eventuali correlazioni.

Ad esempio, se le variabili di partenza sono indipendenti, le covarianze sono nulle, cioè $\sigma_{ij} = 0$ per $i \neq j$ e la (5.79) diventa:

$$\text{Cov}[Z_i, Z_k] \equiv \sum_{j=1}^n (\partial_j f_i) \sigma_{jj} (\partial_j f_k). \quad (5.80)$$

Poiché spesso capita di effettuare calcoli di covarianze per funzioni di variabili aleatorie, vogliamo mostrarvi in dettaglio come fare. Tratteremo il caso di due variabili Z_1, Z_2 e X_1, X_2 , essendo ovvia la generalizzazione ai casi contenenti un numero maggiore di variabili.

Il problema consiste, essendo state ricavate dai dati le varianze e covarianze delle X_1, X_2 , nel determinare le varianze e covarianze delle variabili Z_1, Z_2 . In realtà, tutto è già contenuto nella (5.79). La covarianza $\text{Cov}[Z_1, Z_2]$ è data

infatti da:

$$\begin{aligned}
 \text{Cov}[Z_1, Z_2] &\simeq \sum_{j,l=1}^2 (\partial_j f_1) \sigma_{jl} (\partial_l f_2) \\
 &= (\partial_1 f_1) \sigma_1^2 (\partial_1 f_2) + \\
 &\quad (\partial_1 f_1) \sigma_{12} (\partial_2 f_2) + \\
 &\quad (\partial_2 f_1) \sigma_{21} (\partial_1 f_2) + \\
 &\quad (\partial_2 f_1) \sigma_2^2 (\partial_2 f_2) \\
 &= (\partial_1 f_1) (\partial_1 f_2) \sigma_1^2 + (\partial_2 f_1) (\partial_2 f_2) \sigma_2^2 + \\
 &\quad [(\partial_1 f_1) (\partial_2 f_2) + (\partial_2 f_1) (\partial_1 f_2)] \sigma_{12} ,
 \end{aligned} \tag{5.81}$$

dove nell'ultimo passaggio si è utilizzata l'uguaglianza $\sigma_{12} = \sigma_{21}$.

La notazione matriciale è comoda e compatta, ma, affinché non periate il senso di quello che si è fatto, vi ricordiamo che la (5.81) può essere dimostrata anche direttamente, ricorrendo alla definizione di covarianza (4.24) e sviluppando in serie al primo ordine le variabili z nell'intorno della loro media. Si ha infatti:

$$\begin{aligned}
 \text{Cov}[Z_1, Z_2] &= \langle (Z_1 - \langle Z_1 \rangle) (Z_2 - \langle Z_2 \rangle) \rangle \\
 &\simeq \langle [(\partial_1 f_1) \Delta X_1 + (\partial_2 f_1) \Delta X_2] [(\partial_1 f_2) \Delta X_1 + (\partial_2 f_2) \Delta X_2] \rangle .
 \end{aligned} \tag{5.82}$$

Se si prosegue il conto tenendo presente che

$$\begin{aligned}
 \langle (\Delta X_1)^2 \rangle &= \sigma_1^2 \\
 \langle (\Delta X_2)^2 \rangle &= \sigma_2^2 \\
 \langle (\Delta X_1) (\Delta X_2) \rangle &= \sigma_{12} = \sigma_{21}
 \end{aligned}$$

si ottiene esattamente la (5.81).

Esercizio 5.8

Due variabili Z_1 e Z_2 dipendono da due variabili gaussiane standard indipendenti X e Y secondo le funzioni:

$$\begin{aligned}
 Z_1 &= X + 3Y \\
 Z_2 &= 5X + Y .
 \end{aligned}$$

Trovare il coefficiente di correlazione lineare di Z_1 e Z_2 .

Risposta. Pur essendo X e Y indipendenti, la relazione funzionale crea una dipendenza tra Z_1 e Z_2 .

Definendo $X_1 \equiv X$ e $X_2 \equiv Y$ ed utilizzando la notazione (5.78), si trovano facilmente le derivate da utilizzare nei calcoli:

$$\begin{aligned}
 (\partial_1 f_1) &= 1 , & (\partial_2 f_1) &= 3 , \\
 (\partial_1 f_2) &= 5 , & (\partial_2 f_2) &= 1 .
 \end{aligned}$$

Per determinare il coefficiente di correlazione lineare, dato dalla (4.31), occorre prima determinare le varianze e la covarianza di Z_1 e Z_2 .

Le variabili di partenza sono indipendenti, quindi per determinare le varianze delle Z basta applicare la (5.80) e tenere presente che le variabili standard hanno varianza unitaria:

$$\text{Var}[Z_1] = (1)^2 \sigma_1^2 + (3)^2 \sigma_2^2 = 10 ,$$

$$\text{Var}[Z_2] = (5)^2 \sigma_1^2 + (1)^2 \sigma_2^2 = 26 .$$

Dato che le variabili X e Y sono standard e indipendenti, la covarianza tra le variabili Z si trova applicando la (5.81) con $\sigma_1^2 = \sigma_2^2 = 1$ e $\sigma_{12} = 0$:

$$\text{Cov}[Z_1, Z_2] = (5 \cdot 1) \sigma_1^2 + (3 \cdot 1) \sigma_2^2 + (1 \cdot 1 + 3 \cdot 5) \sigma_{12} = 5 + 3 = 8 .$$

Dalla (4.31) otteniamo infine:

$$\rho[Z_1, Z_2] = \frac{8}{\sqrt{10} \sqrt{26}} = 0.496 .$$

Esercizio 5.9

Di due variabili aleatorie X e Y sono note media, varianza e covarianza: $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}$. Si opera la trasformazione:

$$Z_1 = 5X + Y$$

$$Z_2 = X \cdot Y .$$

Trovare la covarianza $\text{Cov}[Z_1, Z_2]$ tra Z_1 e Z_2 .

Risposta. Siano μ_1 e μ_2 rispettivamente le medie di Z_1 e Z_2 . Dalla (5.82), sviluppando in serie le due nuove variabili in funzione delle vecchie otteniamo:

$$\begin{aligned} \text{Cov}[Z_1, Z_2] &= \langle (Z_1 - \mu_1)(Z_2 - \mu_2) \rangle \\ &= \left\langle \left(\frac{\partial Z_1}{\partial X} \Delta X + \frac{\partial Z_1}{\partial Y} \Delta Y \right) \left(\frac{\partial Z_2}{\partial X} \Delta X + \frac{\partial Z_2}{\partial Y} \Delta Y \right) \right\rangle \\ &= \frac{\partial Z_1}{\partial X} \frac{\partial Z_2}{\partial X} \langle (\Delta X)^2 \rangle + \frac{\partial Z_1}{\partial X} \frac{\partial Z_2}{\partial Y} \langle \Delta X \Delta Y \rangle \\ &\quad + \frac{\partial Z_1}{\partial Y} \frac{\partial Z_2}{\partial X} \langle \Delta X \Delta Y \rangle + \frac{\partial Z_1}{\partial Y} \frac{\partial Z_2}{\partial Y} \langle (\Delta Y)^2 \rangle \\ &= 5Y \text{Var}[X] + X \text{Var}[Y] + (5X + Y) \text{Cov}[X, Y] . \end{aligned} \quad (5.83)$$

Poiché lo sviluppo in serie si intende fatto nell'intorno dei valori medi, le variabili X e Y , che compaiono in questa equazione come frutto delle derivate, sono in realtà i valori μ_x e μ_y assunti come noti. Il problema è quindi completamente risolto.

Esercizio 5.10

Si supponga di considerare le coordinate misurate di un punto nel piano x - y come variabili aleatorie di deviazione standard di 0.2 cm in x e 0.4 cm in y . Le misure non sono correlate. Trovare la matrice delle covarianze in coordinate polari nel punto $(x, y) = (1, 1)$.

Risposta. Poiché le misure non sono correlate, la matrice delle covarianze delle variabili di partenza vale:

$$V_{xy} = \begin{vmatrix} 0.04 & 0 \\ 0 & 0.16 \end{vmatrix}.$$

La trasformazione in coordinate polari è data da:

$$r = \sqrt{x^2 + y^2}, \quad \varphi = \arctan \frac{y}{x}.$$

Ad essa corrisponde la matrice del trasporto:

$$T = \begin{vmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{vmatrix},$$

la quale, calcolata nel punto $(x, y) = (1, 1)$, cui corrisponde in coordinate polari il punto $(r, \varphi) = (\sqrt{2}, \pi/4)$, diventa:

$$T = \begin{vmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{2} & \frac{1}{2} \end{vmatrix}.$$

Il problema può essere ora risolto applicando la (5.76):

$$V_{r,\varphi} = \begin{vmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{2} & \frac{1}{2} \end{vmatrix} \begin{vmatrix} 0.04 & 0 \\ 0 & 0.16 \end{vmatrix} \begin{vmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{2} \\ \frac{1}{\sqrt{2}} & \frac{1}{2} \end{vmatrix} = \begin{vmatrix} 0.100 & 0.042 \\ 0.042 & 0.050 \end{vmatrix}.$$

La radice quadrata degli elementi diagonali della matrice $V_{r,\varphi}$ fornisce le deviazioni standard delle variabili r, φ . Pertanto, in corrispondenza della misura

$$x = 1 \pm 0.2 \text{ cm}, \quad y = 1 \pm 0.4 \text{ cm},$$

la trasformazione in coordinate polari fornisce i valori:

$$r = \sqrt{2} \pm \sqrt{0.100} = 1.41 \pm 0.32 \text{ cm}, \quad \varphi = \frac{\pi}{4} \pm \sqrt{0.050} = 0.78 \pm 0.22 \text{ rad}.$$

Gli elementi non diagonali della matrice mostrano anche che la trasformazione di variabili che si è operata ha introdotto una correlazione positiva tra r e φ .

Esercizio 5.11

Dimostrare la formula (4.90) sulla covarianza delle variabili (I_i, I_j) della distribuzione multinomiale.

Risposta. In una distribuzione multinomiale i contenuti $\{I_i = n_i\}$ dei canali sono correlati dalla relazione:

$$f(n_i) \equiv (n_1 + n_2 + \dots + n_k) = N . \quad (5.84)$$

Dato che N è una costante, $\text{Var}[N] = 0$. Applicando alla (5.84) la legge di trasformazione (5.79) otteniamo:

$$\begin{aligned} \text{Var}[N] &= \sum_{ij} (\partial_i f) \sigma_{ij} (\partial_j f) \\ &= \sum_{ij} \sigma_{ij} = \sum_i \sigma_i^2 + \sum_{i \neq j} \sigma_{ij} = 0 . \end{aligned}$$

Dall'ultimo passaggio di questa equazione e dalle (4.89) otteniamo:

$$\sum_{i \neq j} \sigma_{ij} = - \sum_i \sigma_i^2 = - \sum_i N p_i (1 - p_i) = - \sum_{i \neq j} N p_i p_j ,$$

dove si è tenuto conto che

$$(1 - p_i) = \sum_j p_j , \quad (i \neq j) .$$

Dalla precedente relazione otteniamo la (4.90):

$$\text{Cov}[I_i, I_j] \equiv \sigma_{ij} = -N p_i p_j . \quad (5.85)$$

Per finire, ricordiamo che in questo paragrafo abbiamo utilizzato solo l'approssimazione lineare e che per essa valgono tutte le limitazioni già discusse in precedenza.

Problemi

5.1. Se $X \sim U(0, 1)$, utilizzando la funzione cumulativa trovare la densità di $Z = -\ln X$.

5.2. Trovare la densità di probabilità di $Z = X^2$ dove la densità di X è $p_X(x) = 2(1-x)$, $0 \leq x \leq 1$.

5.3. Trovare la densità di probabilità $p_Z(z)$ della variabile $Z = X/Y$ conoscendo la densità congiunta $p_{XY}(x, y)$.

5.4. Le variabili X e Y sono indipendenti e hanno densità $p_X(x) = \exp[-x]$, $x \geq 0$ e $p_Y(y) = \exp[-y]$, $y \geq 0$ rispettivamente. Trovare la densità di $Z = X + Y$.

5.5. Trovare la densità di $X = \sum_{i=1}^n T_i$, dove T_i sono tempi aleatori indipendenti di legge esponenziale negativa.

5.6. Sapendo che le variabili X e Y sono indipendenti e hanno densità $p_X(x) = \exp[-x]$, $x \geq 0$ e $p_Y(y) = \exp[-y]$, $y \geq 0$ rispettivamente, calcolare la densità $p_{ZW}(z, w)$ delle variabili $Z = X/(X + Y)$, $W = X + Y$ e commentare il risultato.

5.7. Trovare la densità di $Z = XY$ dove X e Y sono due variabili uniformi $\sim U(0, 1)$ e calcolare $\langle Z \rangle$ e $\text{Var}[Z]$.

5.8. Due componenti T_1 e T_2 , entrambi con tempo medio di vita $1/\lambda$, funzionano in parallelo. Il secondo componente entra in funzione solo dopo il guasto del primo. Se il tempo di funzionamento dei due componenti segue la legge esponenziale, trovare la densità di probabilità del tempo T di funzionamento del sistema e la sua vita media.

5.9. Sono date due variabili aleatorie $Z_1 = 3X + 2Y$ e $Z_2 = XY$, dove X e Y sono variabili indipendenti. Trovare media, varianza, covarianza e coefficiente di correlazione lineare di Z_1 e Z_2 quando a) X e Y sono variabili standard e quando b) X e Y sono variabili di media e varianza unitaria.

5.10. Una ditta produce alberi di acciaio il cui diametro è una variabile gaussiana di parametri $\langle X \rangle = 5.450$ e $\sigma[X] = 0.020$ mm e cuscinetti il cui diametro interno è anch'esso una variabile gaussiana di parametri $\langle Y \rangle = 5.550$ e $\sigma[Y] = 0.020$ mm. L'albero va alloggiato entro il cuscinetto e l'accoppiamento è accettabile quando il gioco albero/cuscinetto è compreso tra 0.050 e 0.150 mm. Trovare la percentuale di assemblaggi scartati.

5.11. Lungo una strada passano in media, nell'unità di tempo, μ veicoli da A a B e λ veicoli nell'altro senso, da B ad A . Trovare la densità di probabilità del numero totale di veicoli N e la probabilità di osservare k veicoli da A a B su un totale di n .

Statistica di base

Dove Alinardo sembra dare informazioni preziose e Guglielmo rivela il suo metodo per arrivare a una verità probabile attraverso una serie di errori sicuri.

Umberto Eco, *Il nome della rosa*

6.1 Introduzione

Nei capitoli precedenti crediamo di avervi presentato tutti i risultati veramente importanti del calcolo delle probabilità.

Vediamo ora di entrare nel mondo affascinante della statistica iniziando con il porci la domanda fondamentale: cos'è la statistica e in cosa differisce dal calcolo delle probabilità? Una prima risposta la si può ottenere meditando attentamente sui due punti seguenti:

- *un problema di calcolo delle probabilità*: se ad una moneta attribuiamo una probabilità vera pari a $1/2$ di ottenere testa in un lancio, qual è la probabilità di ottenere in 1000 prove meno di 450 teste? Questo problema è stato risolto nell'esercizio 3.13 di pagina 109;
- *lo stesso problema in statistica*: se in 1000 lanci di una moneta si ottengono 450 teste, qual è la stima che si può dare della probabilità vera di ottenere testa, cioè quella che si otterrebbe in un numero infinito di lanci?

Come si vede, nell'approccio probabilistico si definisce un intervallo di probabilità *calcolato con i valori veri* dei parametri della distribuzione assunta come modello e si stima la probabilità di ottenere un certo *risultato sperimentale*.

Nell'approccio statistico, invece, partendo *dal valore sperimentale* si deve trovare un intervallo per localizzare il valore vero della probabilità. Ci accorgiamo a questo punto che per compiere questa stima manca un ingrediente essenziale: l'equivalente statistico della deviazione standard. Anticipiamo qui un risultato approssimato che verrà discusso nei prossimi paragrafi: spesso in statistica è lecito eseguire la stima delle deviazioni standard sostituendo alle quantità vere quelle misurate: $\sigma \simeq s$. La deviazione standard stimata s così definita viene spesso chiamata da fisici ed ingegneri (e in genere da

chi è abituato ad eseguire misure di laboratorio) come *errore statistico*. In sede internazionale [fSI93], il termine consigliato per i risultati provenienti da misure è quello di incertezza statistica (*statistical uncertainty*). Abbiamo quindi tre sinonimi: deviazione standard stimata (termine matematico), errore statistico (linguaggio dei fisici) ed incertezza statistica (termine consigliato in sede internazionale). Nel seguito noi useremo prevalentemente il termine errore statistico.

Utilizzando le formule dell'esercizio 3.13, otteniamo la Tab. 6.1, che fornisce gli intervalli di probabilità:

$$\mu \pm \sigma = 500.0 \pm 15.8 \simeq 500 \pm 16 = [484, 516] \quad (\text{calcolo delle probabilità})$$

$$x \pm s = 450.0 \pm 15.7 \simeq 450 \pm 16 = [434, 466] \quad (\text{statistica}).$$

Nonostante l'apparente analogia, questi due intervalli hanno un significato profondamente diverso: il primo, noti μ e σ , assegna una probabilità ad un insieme di valori di X , il secondo fornisce una previsione sul valore di μ .

Tabella 6.1. Differenza tra calcolo delle probabilità e statistica nel caso elementare di $x = 450$ successi in $n = 1000$ lanci di una moneta

<i>Calcolo delle probabilità</i>	<i>Statistica</i>
probabilità di	stima di
valori dello spettro	parametri (\hat{p})
probabilità vera: $p = 0.5$	frequenza: $f = x/n = 0.45$
valore atteso: $\langle X \rangle = 500$	valore misurato: $x = 450$
deviazione standard:	errore statistico o incertezza:
$\sigma[X] = \sqrt{np(1-p)} = 15.8$	$s = \sqrt{nf(1-f)} = 15.7$

L'esempio appena discusso si riferisce alla *stima* di un parametro vero incognito partendo dai dati. Come vedremo, questa operazione viene fatta attraverso gli *stimatori consistenti*, definiti nella (2.76) come variabili aleatorie $T_N(X)$ funzioni di un campione casuale di dimensione N che convergono in probabilità ad un valore determinato. Vi ricordiamo che la definizione 2.12 di campione casuale è a pagina 63.

L'altro campo di applicazione della statistica, la verifica delle ipotesi, cerca di rispondere a domande del tipo: se ripetendo due volte l'esperimento consistente in mille lanci di una moneta si ottengono 450 e 600 successi, quanto è probabile che nei due esperimenti si sia usata la stessa moneta? In questo capitolo affronteremo la verifica delle ipotesi nel caso più semplice, quello del rigetto o meno di una ipotesi iniziale. Più avanti, nel cap. 9, spiegheremo come ottimizzare la scelta tra più ipotesi alternative.

I concetti finora esemplificati col lancio della moneta possono essere definiti rigorosamente tramite lo spazio di probabilità (1.11), introducendo una

notazione leggermente diversa:

$$\mathcal{E}(\theta) \equiv (S, \mathcal{F}, P_\theta) \quad (6.1)$$

dove P_θ è una probabilità dipendente da un parametro θ , che fornisce una legge

$$P\{X \in A\} = \int_A p(x; \theta) dx \quad (6.2)$$

per il campione casuale (X_1, X_2, \dots, X_N) . Nel caso del lancio della moneta la probabilità è discreta e $p(x; \theta) = b(x; 1000, p)$, con $\theta = p$. È dunque del parametro θ che cercheremo una stima e sul parametro θ che vedremo come procedere alla verifica di ipotesi, formalizzando quanto esposto in via intuitiva negli esercizi 3.13-3.17.

Le sezioni seguenti fino alla 6.13 sono dedicate alla stima, le successive alla verifica di ipotesi. Per facilitarvi lo studio e la applicazione pratica di questi metodi abbiamo suddiviso l'esposizione per problemi (studio della probabilità, della media, della varianza, ecc.) e abbiamo intenzionalmente intitolato i paragrafi come “stima di ...” quando si tratta di un problema di stima dei parametri e come “verifica di ...” quando si discute di verifica di ipotesi.

6.2 Intervalli di confidenza

Nell'esempio introduttivo di Tab. 6.1 abbiamo definito, in modo intuitivo, un intervallo di ampiezza $2s$ per stimare la probabilità di ottenere testa come:

$$n \pm \sqrt{nf(1-f)} \simeq 450 \pm 16$$

sulla base del risultato di mille lanci.

Immaginiamo ora di ripetere lo stesso esperimento per ottenere un nuovo intervallo, che, prima di osservare il risultato dei lanci, è chiaramente un intervallo aleatorio che potremmo definire come

$$X \pm S .$$

In statistica, la probabilità che questo intervallo ha di contenere la vera probabilità p di osservare testa si chiama *livello di confidenza* dall'inglese *confidence level*, e si indica con $CL \equiv (1 - \alpha)$, mentre α è detto livello di significatività. È nostro interesse che questo intervallo abbia contemporaneamente un livello di confidenza elevato e un'ampiezza ridotta perché sia utile. Cerchiamo allora di definire un criterio per scegliere un intervallo con il CL desiderato seguendo l'interpretazione frequentista adottata nelle scienze sperimentali sulla base del famoso lavoro di Neyman [Ney37].

Consideriamo la densità $p(x; \theta)$ di forma funzionale nota, e supponiamo di voler determinare il valore del parametro θ fisso ed incognito. Se θ è un parametro di posizione come la media, per valori diversi di θ la densità scorrerà

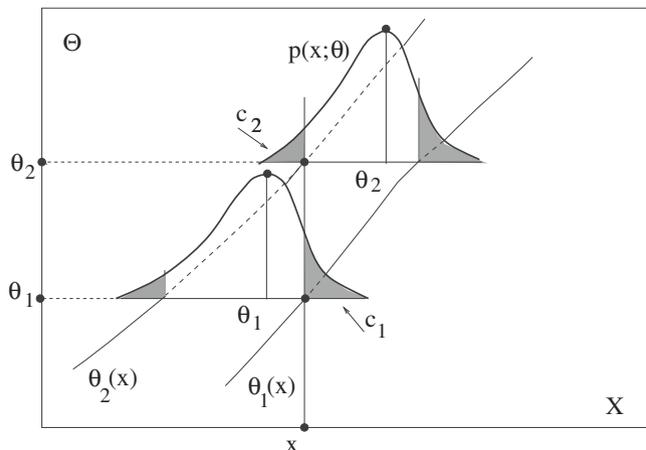


Fig. 6.1. Costruzione di un intervallo di confidenza di livello CL per il parametro di posizione θ . Le aree chiare non ombreggiate di $p(x; \theta)$ valgono tutte CL . Variando il valore di θ varia la posizione di $p(x; \theta)$ lungo l'asse dei valori misurati x . Introducendo l'asse verticale dei parametri Θ , si ottiene la situazione della figura, in cui si vede la fascia di confidenza di Neyman $[\theta_1(x), \theta_2(x)]$, la cui larghezza in x è l'intervallo di $p(x; \theta)$ che sottende un'area CL . Ottenendo in un esperimento il valore x e determinando le intersezioni della retta passante per x e parallela all'asse Θ con la fascia di confidenza, si determina l'intervallo $[\theta_1, \theta_2]$ che contiene il valore vero θ con un livello di confidenza $CL = 1 - \alpha$, dove $\alpha = c_1 + c_2$ è la somma delle due aree ombreggiate

lungo l'asse x come mostrato in Fig. 6.1. Le aree non ombreggiate delle densità di Fig. 6.1 corrispondono ai livelli di probabilità:

$$P\{x_1 \leq X \leq x_2; \theta\} = 1 - \alpha = \int_{x_1}^{x_2} p(x; \theta) dx, \quad (6.3)$$

dove α è la somma delle aree delle due code ombreggiate. L'unione di tutti gli intervalli $[x_1, x_2]$ di Fig. 6.1 dà luogo ad una regione, nel piano (x, θ) , detta *fascia di confidenza di livello $CL = 1 - \alpha$* . Questa fascia, se si guarda la Fig. 6.1 dall'alto, appare come nella Fig. 6.2. Le due curve che la delimitano sono funzioni monotone crescenti $\theta_1(x)$ e $\theta_2(x)$. Questo non è vero in generale ma spesso ci si può ricondurre a questa situazione riparametrizzando il problema (ad esempio, nel caso della distribuzione esponenziale negativa, utilizzando la media anziché la probabilità per unità di tempo come parametro).

Tenendo sempre presente le Figg. 6.1 e 6.2, supponiamo ora di aver misurato X ottenendo un valore x ; tracciamo da questo punto una parallela all'asse dei parametri, le cui intersezioni con la fascia di confidenza individuano l'intervallo $[\theta_1, \theta_2]$. Notiamo che questo intervallo è aleatorio, perché il valore x varia ad ogni esperimento. In generale dovremo quindi indicarlo con $[\Theta_1, \Theta_2]$.

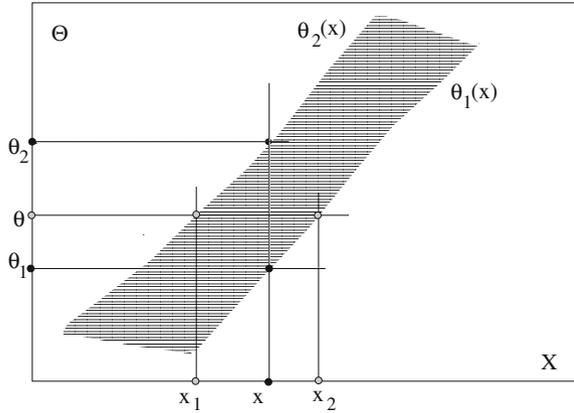


Fig. 6.2. Guardando dall'alto la Fig. 6.1 appare la fascia di confidenza di Neyman per un CL fissato, che permette la determinazione dell'intervallo di confidenza $\theta \in [\theta_1, \theta_2]$ a partire da un valore misurato x

Se il valore vero del parametro incognito è θ , dalle figure si vede che ad esso corrisponde sull'asse dei valori misurati l'intervallo $[x_1, x_2]$ e, per costruzione, si ha $P\{x_1 \leq X \leq x_2\} = CL$. Dato che quando $x = x_1$ sull'asse dei parametri $\theta = \theta_2$ e quando $x = x_2$ si ha invece $\theta = \theta_1$, per costruzione $x \in [x_1, x_2]$ se e solo se $\theta \in [\theta_1, \theta_2]$. Vale allora la relazione fondamentale:

$$P\{x_1 \leq X \leq x_2\} = P\{\theta_1 \leq \theta \leq \theta_2\} = CL, \quad (6.4)$$

che definisce l'intervallo di confidenza $[\theta_1, \theta_2]$ di livello di confidenza $CL = 1 - \alpha$. Quanto detto finora può essere riassunto nella:

Definizione 6.1 (Intervallo di confidenza). *Date due statistiche Θ_1 e Θ_2 (nel senso della def. 2.13 di pagina 64) con Θ_1 e Θ_2 variabili continue e $\Theta_1 \leq \Theta_2$ con probabilità 1, si dice che $I = [\Theta_1, \Theta_2]$ è un intervallo di confidenza per un parametro θ , di livello di confidenza $0 < CL < 1$, se, per ogni θ appartenente allo spazio dei parametri, la probabilità che I contenga θ vale CL :*

$$P\{\Theta_1 \leq \theta \leq \Theta_2\} = CL. \quad (6.5)$$

Se Θ_1 e Θ_2 sono variabili discrete, l'intervallo di confidenza è l'intervallo più piccolo che soddisfa la relazione:

$$P\{\Theta_1 \leq \theta \leq \Theta_2\} \geq CL. \quad (6.6)$$

Per evidenziare meglio questo concetto di intervallo che “ricopre” il parametro θ con una certa probabilità, il livello di confidenza viene anche chiamato “copertura”. In inglese viene usato il termine *coverage probability* [CB90] e la condizione (6.6) viene detta di *minimum overcoverage* (sovra-copertura).

Il fatto importante da notare è che nell'intervallo di confidenza (6.5) gli estremi *sono variabili aleatorie, mentre il parametro θ è fissato*; di conseguenza, il livello di confidenza va sempre riferito all'intervallo $I = [\Theta_1, \Theta_2]$ ed indica *la frazione di esperimenti che individuano correttamente il valore vero, in un insieme infinito di esperimenti ripetuti, ognuno dei quali trova un intervallo di confidenza diverso*. Ciò equivale ad affermare che ogni particolare intervallo $[\theta_1, \theta_2]$ viene ricavato con un metodo che dà il risultato corretto in una frazione *CL* degli esperimenti.

Questa interpretazione dei risultati statistici, che ha un chiaro significato operativo, molto vicino a quanto succede nella vita reale nelle misure di laboratorio e in genere negli esperimenti ripetuti, è detta *interpretazione frequentista* ed è quella diventata prevalente nelle scienze applicate grazie al lavoro pubblicato dallo statistico J. Neyman nel 1937 [Ney37].

In statistica è lecito chiamare intervallo di confidenza sia l'intervallo aleatorio $I = [\Theta_1, \Theta_2]$ sia le sue realizzazioni numeriche $[\theta_1, \theta_2]$. Talvolta si afferma che il *CL* è la probabilità che il valore vero θ sia contenuto in $[\theta_1, \theta_2]$ ma si deve sempre essere consci che questa dizione, nell'interpretazione frequentista, è scorretta perché θ non è una variabile aleatoria [Cou95].

In pratica, avendo ottenuto un valore x di una variabile X continua e conoscendo la densità $p(x; \theta)$ della (6.2), se θ è un parametro di posizione come la media, i valori $[\theta_1, \theta_2]$, relativi ad una realizzazione dell'intervallo aleatorio $[\Theta_1, \Theta_2]$ secondo il *CL* assegnato, possono essere trovati, come mostrato in Fig. 6.1, con le relazioni (fate attenzione alla posizione di θ_1 e θ_2):

$$\int_x^\infty p(z; \theta_1) dz = c_1, \quad \int_{-\infty}^x p(z; \theta_2) dz = c_2, \quad (6.7)$$

dove $CL = 1 - c_1 - c_2$. Nel caso X sia una variabile discreta, agli integrali vanno sostituite le somme sui corrispondenti valori dello spettro, come vedremo tra poco.

Se si sceglie un intervallo simmetrico, allora valgono le condizioni $c_1 = c_2 = (1 - CL)/2$. La scelta dell'intervallo simmetrico non è ovviamente l'unica ma è la più comune poiché, nel caso di densità di probabilità di forma regolare "a campana", questo è l'intervallo di ampiezza minima. A volte si vuole determinare, per un certo *CL*, solo il limite superiore di θ , cioè l'intervallo $(-\infty, \theta]$; in questo caso nelle (6.7) va posto $\theta = \theta_1 = \theta_2$ e $c_1 = CL$, $c_2 = 1 - CL$. Per il limite inferiore, cioè per l'intervallo $[\theta, +\infty)$, valgono invece le condizioni $\theta = \theta_1 = \theta_2$ e $c_1 = 1 - CL$, $c_2 = CL$.

Il tipo di intervallo da utilizzare è in genere determinato dalla natura del problema. È però di fondamentale importanza tenere presente che la scelta va fatta *prima di effettuare la misura*. Decidere ad esempio, nel caso di eventi rari, se dare un intervallo simmetrico o un limite superiore a seconda che la misura fornisca o meno dei risultati (tecnica detta di "flip-flop") porta ad un calcolo errato dei livelli da associare agli intervalli di confidenza [FC98].

Notiamo che le (6.7) valgono in generale, poiché quasi sempre un problema di stima richiede la individuazione di uno stimatore T_N , secondo la (2.76) di

pag. 67 (dove $\mu \equiv \theta$). La densità di T_N dipende in genere dal parametro da stimare e può essere indicata come $p(t; \theta)$, con θ parametro di posizione. Infatti, $\langle T_N \rangle = \int t p(t; \theta) dt = \tau(\theta)$, con $\tau(\theta) = \theta$, se lo stimatore è corretto (questi concetti verranno approfonditi più avanti, nel cap. 9). Se si riesce a trovare questa densità, si può effettuare la stima di θ attraverso le (6.7).

6.3 Determinazione degli intervalli di confidenza

Abbiamo fin qui visto due modi per determinare l'intervallo di confidenza: il semplice metodo grafico di Fig. 6.1, che richiede però la costruzione della fascia di confidenza di Neyman, oppure il calcolo degli integrali (6.7). Entrambi questi metodi sono in genere impegnativi dal punto di vista computazionale. Fortunatamente, se la forma della funzione densità ha qualche proprietà di invarianza rispetto ai parametri da stimare, si possono trovare metodi particolarmente semplici.

Consideriamo il caso in cui il parametro θ di Fig. 6.1 è la media di una curva normale, $\theta \equiv \mu$. Allora la forma di $p(x; \mu)$ è invariante per traslazioni e le funzioni $\theta_1(x)$ e $\theta_2(x)$ di Fig. 6.2 sono due rette parallele. Dalla Fig. 6.3 risulta la proprietà seguente:

$$\int_{\mu-t\sigma}^{\mu+t\sigma} p(x; \mu) dx = \int_{x-t\sigma}^{x+t\sigma} p(\mu; x) d\mu \tag{6.8}$$

che si può scrivere come:

$$P\{\mu-t\sigma \leq X \leq \mu+t\sigma\} = P\{-t\sigma \leq X-\mu \leq t\sigma\} = P\{X-t\sigma \leq \mu \leq X+t\sigma\} . \tag{6.9}$$

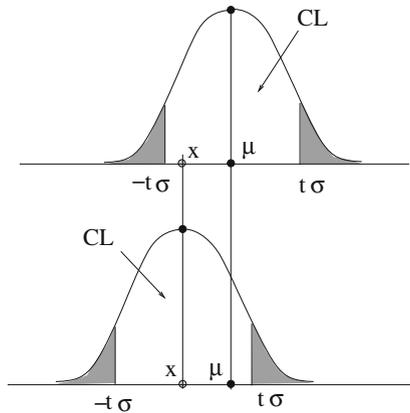


Fig. 6.3. Nel caso la forma della densità sia invariante per traslazione del parametro μ , il calcolo dell'intervallo di confidenza viene fatto in modo semplice con la formula (6.9)

In altre parole, *i livelli di probabilità dell'intervallo centrato su μ coincidono con i livelli di confidenza dell'intervallo aleatorio centrato su X* . Ad ogni esperimento il valore di X varia, ma la probabilità di copertura di μ è uguale a quella dell'intervallo di probabilità $\mu \pm t\sigma$ (si veda la Fig. 6.3). In termini di funzione di ripartizione $F(z; \theta)$ (per θ ed z scalari) questa proprietà può essere scritta come:

$$F(z; \theta) = 1 - F(\theta; z) . \quad (6.10)$$

Abbiamo perciò trovato la seguente regola pratica: nel caso gaussiano, o comunque quando vale la (6.10), basta centrarsi sul valore misurato ed assumere come livelli di confidenza i livelli di probabilità corrispondenti all'ampiezza dell'intervallo di stima centrato sulla media.

Ad esempio, nel caso di Tab. 6.1, dato che sappiamo che il numero x di successi in $n = 1000$ lanci di una moneta è una variabile gaussiana (perché $np, n(1-p) \gg 10$), possiamo stimare la probabilità vera come:

$$\begin{aligned} x \pm \sqrt{nf(1-f)} &= 450 \pm 16 \quad (CL = 68.3\%) \\ x \pm 2\sqrt{nf(1-f)} &= 450 \pm 32 \quad (CL = 95.4\%) \\ x \pm 3\sqrt{nf(1-f)} &= 450 \pm 48 \quad (CL = 99.7\%) . \end{aligned}$$

Come abbiamo appena visto, la variabile aleatoria $Q = (X - \mu)$, pur contenendo il parametro μ , ha distribuzione $N(0, \sigma^2)$ indipendente da μ . Le variabili aleatorie la cui distribuzione non dipende dal parametro che si vuole stimare sono dette quantità-pivot o quantità pivotali. Un altro esempio si ha quando θ è un parametro di scala, cioè $p(x; \theta) = h(x/\theta)/\theta$; in questo caso $Q = X/\theta$ è una quantità pivotale per θ .

Generalizzando la discussione, possiamo dire che se $Q(X, \theta)$ è una quantità pivotale, la probabilità $P\{Q \in A\}$ non dipende da θ per ogni $A \in \mathbb{R}$. Se questa distribuzione ha una densità nota (gaussiana standard, Student, χ^2 , ...), in corrispondenza di un CL assegnato si possono determinare facilmente i quantili q_1 e q_2 tali che

$$P\{q_1 \leq Q(X, \theta) \leq q_2\} = CL . \quad (6.11)$$

Se la relazione $q_1 \leq Q(X, \theta) \leq q_2$ può essere risolta rispetto a θ , possiamo scrivere, come nella (6.9):

$$P\{q_1 \leq Q(X, \theta) \leq q_2\} = P\{\theta_1(X) \leq \theta \leq \theta_2(X)\} = P\{\Theta_1 \leq \theta \leq \Theta_2\} , \quad (6.12)$$

ottenendo così la (6.5). Come si vede, se si trova una quantità pivotale risolvibile rispetto a θ secondo la (6.12), l'intervallo di confidenza può essere determinato *senza ricorrere agli integrali* (6.7).

6.4 Cenno all'approccio bayesiano

Le considerazioni fin qui svolte stanno alla base dell'approccio frequentista per la stima dei parametri.

Nell'approccio bayesiano, cui abbiamo accennato nel par. 1.3, il parametro da stimare *viene considerato come una variabile aleatoria* e l'intervallo di confidenza rappresenta la conoscenza ottenuta, dopo la misura, sul valore del parametro. Avendo ottenuto 450 teste in mille lanci di una moneta, in approssimazione normale ed attribuendo al valore atteso una probabilità a priori costante, risulta che dopo l'esperimento il valore atteso vale 450 ± 16 con un grado di credibilità (probabilità o *belief*) del 68%. In questo caso l'approccio bayesiano fornisce un risultato numerico uguale a quello frequentista, ma interpretato in altro modo. L'intervallo bayesiano dipende però dalle informazioni a priori. Se sapessimo che la moneta è fabbricata bene ($p = 0.5$) con una incertezza sulla bilanciatura di tipo gaussiano con, poniamo, $\sigma_p/p = 0.1\%$, potremmo sostituire alla distribuzione a priori uniforme (probabilità costante) quella gaussiana $N(p, \sigma_p^2)$, ottenendo un risultato diverso anche numericamente da quello frequentista.

Non approfondiremo ulteriormente questi aspetti, che sono trattati in dettaglio, dal punto di vista statistico, in [CB90] e [Pic96].

Ricordiamo infine che in fisica sono state recentemente proposte analisi bayesiane quando non è facile trovare quantità pivotali, come nel caso di piccoli campioni contenenti conteggi in presenza di fondo o di campioni provenienti da popolazioni gaussiane con vincoli fisici sulle variabili misurate (ad esempio, se X è una massa, vale la condizione *a priori* $\{X \geq 0\}$) [Cou95, D'A99]. Recentemente è stato però proposto, anche per questi casi, un approccio frequentista, che non necessita di assunzioni a priori sulla distribuzione dei parametri [FC98] e che sta incontrando il favore dei fisici sperimentali [JLPe00].

6.5 Alcune notazioni

Nel seguito sarà importante tenere presente le notazioni usate per le stime puntuali e le stime per intervalli di confidenza. La stima puntuale di una quantità vera si effettua attraverso i valori degli stimatori; ad esempio la media di un campione è una possibile stima della media vera. La notazione che useremo è: $m = \hat{\mu}$. Tratteremo in modo esteso le stime puntuali nei capp. 9 e 10.

In questo capitolo tratteremo invece in modo completo le stime per intervalli.

Nel caso gaussiano, un valore vero θ è contenuto con livello di confidenza $1 - \alpha = CL$ in un intervallo simmetrico centrato su x quando

$$\theta \in [x - t_{1-\alpha/2}s, x + t_{1-\alpha/2}s], \quad (6.13)$$

dove $s \simeq \sigma$ e $t_{\alpha/2} = -t_{1-\alpha/2}$ sono i quantili della gaussiana standard. È facile verificare che gli indici dei quantili si possono scrivere in termini del CL come:

$$\frac{\alpha}{2} = \frac{1 - CL}{2}, \quad 1 - \frac{\alpha}{2} = \frac{1 + CL}{2}. \quad (6.14)$$

In statistica spesso si scrive l'intervallo ad 1σ come:

$$\theta \in [x - s, x + s] \equiv x \pm s . \quad (6.15)$$

La prima notazione è preferita dai matematici, la seconda da fisici ed ingegneri, i quali quasi sempre sostituiscono il simbolo di appartenenza con quello di uguaglianza:

$$\theta \in x \pm s \xrightarrow{\text{fisici ed ingegneri}} \theta = x \pm s . \quad (6.16)$$

Se gli errori a destra e sinistra del valore centrale sono diversi, la notazione dei matematici ovviamente non cambia, mentre l'altra diventa:

$$\theta \in [x - s_1, x + s_2] \equiv x_{-s_1}^{+s_2} . \quad (6.17)$$

Solitamente viene considerato improprio assegnare ad s più di due cifre significative; se ad esempio la prima cifra significativa dell'errore s è il metro, non ha senso dare il risultato con la precisione del millimetro. Vale la seguente regola pratica, che qui riportiamo come:

Affermazione 6.2 (Cifre significative dell'errore statistico). *In statistica i risultati finali del tipo $x \pm s$ vanno presentati con l'errore statistico s dato a non più di due cifre significative e x arrotondato di conseguenza. Se non si segue questa regola, occorre spiegarne le ragioni.*

Notiamo che nei risultati intermedi, per ridurre gli errori di troncamento, è lecito usare un numero più elevato di cifre significative; tuttavia, nei *risultati finali*, si dovrebbe sempre seguire la regola 6.2. È pertanto scorretto scrivere:

$$35.923 \pm 1.407 , \quad 35.923 \pm 1.4 , \quad 35.9 \pm 1.407 ,$$

perché nel primo caso vi sono troppe cifre significative, nel secondo e nel terzo vi è discordanza tra le cifre significative del dato e dell'errore. È invece corretto scrivere:

$$35.9 \pm 1.4 \quad \text{oppure} \quad 36 \pm 1 .$$

6.6 Stima della probabilità da grandi campioni

Ci poniamo il problema seguente: se eseguiamo una prova di n tentativi bernoulliani e otteniamo x successi e quindi una frequenza $f = x/n$, come possiamo stimare la probabilità vera dell'evento in esame?

Dal calcolo delle probabilità sappiamo che effettuando n tentativi su un sistema che genera eventi con probabilità p costante si può ottenere un numero x di successi (spettro) compreso tra 0 ed n . Tuttavia, questi risultati non sono equiprobabili, ma si distribuiscono secondo la densità binomiale (2.30). Sappiamo anche che i valori capitano più di frequente intorno al valor medio np , con varianza $\sigma^2 = np(1-p)$, e che la forma della densità binomiale diventa gaussiana per $np, n(1-p) > 10$.

Con questo tipo di prova si osservano le realizzazioni numeriche $f = x/n$ della variabile aleatoria frequenza $F = X/n$, che, in base alle (2.63, 3.4, 3.6), ha media e varianza pari a

$$\langle F \rangle = \frac{\langle X \rangle}{n} = \frac{np}{n} = p, \quad (6.18)$$

$$\text{Var}[F] = \frac{\text{Var}[X]}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}. \quad (6.19)$$

Possiamo quindi definire la variabile standard (3.37)

$$T = \frac{F - p}{\sigma[F]}, \quad \text{che assume i valori} \quad t = \frac{f - p}{\sqrt{\frac{p(1-p)}{n}}}. \quad (6.20)$$

Se ora eseguiamo una prova, f è noto e p è incognito; *utilizzando l'approccio statistico*, possiamo allora determinare i valori di p per i quali il valore assunto dalla variabile standard risulta minore di un certo limite t assegnato:

$$\frac{|f - p|}{\sqrt{\frac{p(1-p)}{n}}} \leq |t|.$$

Eliminiamo i valori assoluti elevando al quadrato ambo i membri e risolviamo, *rispetto all'incognita* p , l'equazione di II grado risultante:

$$t^2 \geq \frac{(f-p)^2}{p(1-p)/n}, \quad (f-p)^2 \leq t^2 \frac{p(1-p)}{n}, \\ (t^2 + n)p^2 - (t^2 + 2fn)p + nf^2 \leq 0.$$

Poiché il termine in p^2 è sempre positivo, la disuguaglianza è soddisfatta per valori di p compresi nell'intervallo:

$$p \in \frac{(t^2 + 2fn) \pm \sqrt{t^4 + 4f^2n^2 + 4t^2nf - 4t^2nf^2 - 4n^2f^2}}{2(t^2 + n)},$$

da cui, in forma compatta, si ottiene la formula di Wilson:

$$p \in \frac{f + \frac{t^2}{2n}}{\frac{t^2}{n} + 1} \pm \frac{t \sqrt{\frac{t^2}{4n^2} + \frac{f(1-f)}{n}}}{\frac{t^2}{n} + 1}. \quad (6.21)$$

Il parametro t sta ad indicare un qualunque valore della variabile standard T ; porre $t = 1$ corrisponde a considerare una unità di errore statistico. Si noti che l'intervallo non è centrato sul valore misurato f , ma nel punto $(f + t^2/2n)/(t^2/n + 1)$, che è funzione di f e del numero dei tentativi effettuati. Questo effetto è dovuto alla notevole asimmetria della densità binomiale

per piccoli n , come si vede dalla Fig. 2.4. Per $n \gg 1$ vale l'approssimazione gaussiana, l'intervallo tende a centrarsi sul valore misurato e la (6.21), per $CL = 1 - \alpha$, diventa:

$$p \in f \pm t_{1-\alpha/2}s = f \pm t_{1-\alpha/2} \sqrt{\frac{f(1-f)}{n}}, \quad (6.22)$$

dove $|t_{\alpha/2}| = t_{1-\alpha/2}$ sono i valori quantili della gaussiana standard che definiscono gli estremi dell'intervallo che sottende un'area pari a $CL = 1 - \alpha$. Generalmente viene riportato l'intervallo ad 1σ con $t_{1-\alpha/2} = 1$:

$$p \in f \pm s = f \pm \sqrt{\frac{f(1-f)}{n}}. \quad (6.23)$$

che è detto intervallo di Wald. Moltiplicando per il numero di tentativi n , la (6.23) può essere facilmente espressa in funzione del numero di successi x . L'intervallo che si ottiene è allora relativo al numero atteso di successi, ovvero alla speranza matematica μ :

$$\mu \in x \pm \sqrt{x \left(1 - \frac{x}{n}\right)}. \quad (6.24)$$

Questa formula, che viene usata in pratica per $nf, n(1-f) > 20, 30$, è stata utilizzata in Tab. 6.1 nell'esempio introduttivo al capitolo. Essa è facile da ricordare, perché l'intervallo è centrato sul valore misurato, *la variabile T della (6.20) è pivotale perché $T \sim N(0, 1)$* , vale la (6.9), l'errore statistico ha la stessa espressione della deviazione standard della distribuzione binomiale (con la probabilità p sostituita dalla frequenza misurata f) e i livelli di confidenza CL sono gaussiani.

Gli intervalli (6.21) e (6.23) sono gli intervalli di confidenza rispettivamente per piccoli e grandi campionamenti.

Come abbiamo appena osservato, per grandi campionamenti la densità binomiale assume una forma gaussiana e l'intervallo di confidenza, come mostrato in Fig. 6.4a, è simmetrico. Il valore misurato f definisce in questo caso due code che valgono entrambe $(1 - CL)/2$, in modo tale da avere le due aree tra f e p_1 e tra f e p_2 uguali ognuna a $CL/2$.

Nel caso di piccoli campioni, come mostrato in Fig. 6.4b, le due code di area $(1 - CL)/2$ assumono forma diversa e l'intervallo di confidenza è asimmetrico rispetto alla frequenza osservata. La variabile T della (6.20) non è pivotale e non è più possibile assegnare in modo generale i livelli di confidenza all'intervallo (6.21).

Trattiamo per primo il caso dell'intervallo (6.23), relativo ai grandi campioni, che è il caso più facile. Se è valida la condizione $nf, n(1-f) \gg 1$, i livelli di confidenza associati all'intervallo (6.23) sono quelli della legge 3σ gaussiana (3.35), riportati nella Tab. D.1. È allora molto facile effettuare stime affidabili, come mostrato nell'esempio che segue.

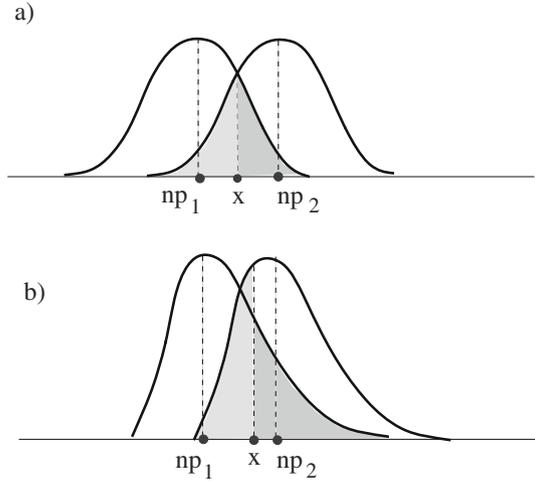


Fig. 6.4. Ad un livello di confidenza pari a CL , corrispondente nel caso simmetrico (a) alla somma di due aree relative agli intervalli (np_1, x) e (x, np_2) , corrispondono due code di area $(1 - CL)/2$ che sono uguali nel caso di grandi campionamenti (a), di forma diversa nel caso di piccoli campioni (b)

Esercizio 6.1

In sede di proiezione dei risultati elettorali, sono state esaminate 3000 schede, campionate casualmente dalla popolazione totale di schede in scrutinio. Il partito A ha ottenuto 600 voti. Dare la previsione finale (proiezione) del risultato.

Risposta. Dato che $n = 3000$, $f = 600/3000 = 0.20$ e $nf, n(1 - f) \gg 10$ possiamo utilizzare l'intervallo (6.23), con i livelli di confidenza gaussiani. Otteniamo pertanto:

$$p \in f \pm \sqrt{\frac{f(1-f)}{n}} = (20.0 \pm 0.7)\% \quad CL = 68.3\% ,$$

$$p \in f \pm 2\sqrt{\frac{f(1-f)}{n}} = (20.0 \pm 1.4)\% \quad CL = 95.4\% ,$$

$$p \in f \pm 3\sqrt{\frac{f(1-f)}{n}} = (20.0 \pm 2.1)\% \quad CL = 99.7\% .$$

Si noti la sorprendente precisione che si ottiene pur con un numero limitato di schede elettorali. A rigore, dato che la popolazione in questo caso è molto grande ma finita (milioni di elettori), a questi risultati andrebbe apportata, come mostreremo tra poco nel par. 6.13, una correzione che però qui è assolutamente trascurabile. La vera difficoltà, in questo tipo di previsioni, sta

nell'ottenere un campione veramente rappresentativo della popolazione totale. In generale si definisce come rappresentativo o casuale un campione nel quale un individuo di qualunque gruppo ha una probabilità di entrare proporzionale alla dimensione del gruppo rispetto alla popolazione totale (si veda anche la def. 6.4 più avanti). In campioni relativi ad un fenomeno naturale di tipo fisico, come quelli che si ottengono in laboratorio, se non si commettono errori sistematici durante le misure è la natura stessa che provvede a fornire un campione casuale. La situazione è invece molto diversa nel caso delle scienze sociali o biologiche, dove le tecniche di campionamento da una popolazione (sampling) sono tanto importanti e difficili da costituire un ramo speciale della statistica. Chi è interessato a questi metodi può consultare [Coc77].

La (6.22) permette anche di trovare la dimensione del campione necessaria per mantenere l'errore statistico al di sotto di un valore assoluto prefissato a priori. Si arriva facilmente a questo risultato se eleviamo al quadrato l'errore statistico presente nell'equazione e sostituiamo a f il valore vero p , ottenendo la varianza:

$$\sigma^2 = \frac{p(1-p)}{n} . \quad (6.25)$$

Per inciso, notiamo che questa equazione è identica alla (3.5), a meno della divisione per il fattore n^2 , dato che qui consideriamo la variabile $F = X/n$. Se ora uguagliamo a zero la derivata rispetto a p , otteniamo:

$$\frac{d}{dp} \left[\frac{p(1-p)}{n} \right] = \frac{1}{n} (1-2p) = 0 \implies p = \frac{1}{2} .$$

Sostituendo il valore di massimo $p = 1/2$ nella (6.25), otteniamo il limite superiore per la varianza vera, in funzione del numero di tentativi n :

$$\sigma_{max}^2 = \frac{1}{4n} . \quad (6.26)$$

La proprietà notevole di questa formula è che essa dà un limite superiore per la varianza *indipendentemente dal valore della probabilità vera p* . È pertanto possibile determinare una formula universale per il numero di tentativi necessari per mantenere l'intervallo di stima $\pm t_{1-\alpha/2} \sigma_{max}$ al di sotto di un certo valore prefissato, in corrispondenza di un certo livello di confidenza $CL = 1 - \alpha$. Essa si ottiene dalla (6.26):

$$n = \frac{t_{1-\alpha/2}^2}{4(t_{1-\alpha/2} \sigma_{max})^2} . \quad (6.27)$$

Esercizio 6.2

Trovare il numero di campionamenti necessario per avere un intervallo di stima assoluto minore del 4 per mille con un livello di confidenza del 99%.

Risposta. *Poiché il problema si riferisce ai grandi campionamenti, possiamo usare la Tab. D.1 da cui risulta che ad un CL del 99%, pari a una semi-area gaussiana di 0.495, corrisponde un valore $t_{1-0.005} = 2.57$. Inserendo questo valore nella (6.27) otteniamo immediatamente:*

$$n = \frac{(2.57)^2}{4 \cdot (0.002)^2} = 412\,806 .$$

6.7 Stima della probabilità da piccoli campioni

Veniamo ora alla seconda parte del problema relativo alla stima di una probabilità, cioè al caso dei piccoli campioni, in cui $x = nf < 10$.

In questo caso dovremmo usare l'intervallo (6.21), per il quale non sono però definiti livelli di confidenza universali, perché la forma della densità binomiale dipende fortemente, per piccoli campionamenti, sia dal numero di tentativi n sia dal valore di p . Risolviamo pertanto il problema con le (6.7) trascritte per una variabile discreta; se CL è il livello di confidenza richiesto, i valori p_1 e p_2 del corrispondente intervallo di confidenza sono determinati dagli estremi delle due code di Fig. 6.4b, che possono essere trovati con le equazioni dette di Clopper-Pearson:

$$\sum_{k=x}^n \binom{n}{k} p_1^k (1-p_1)^{n-k} = c_1 , \quad (6.28)$$

$$\sum_{k=0}^x \binom{n}{k} p_2^k (1-p_2)^{n-k} = c_2 . \quad (6.29)$$

La presenza in entrambe somme del valore x assicura il rispetto della condizione di sovra-copertura (6.6) per variabili discrete. Nel caso simmetrico $c_1 = c_2 = (1 - CL)/2 = \alpha/2$. La risoluzione di queste due equazioni rispetto a p_1 e p_2 permette la stima corretta della probabilità da piccoli campioni.

In alternativa a queste formule, si può usare la formula di Wilson (6.21) applicando alla frequenza $f = x/n$ la correzione di continuità, che in genere migliora la copertura degli intervalli di confidenza quando la variabile è discreta:

$$f_{\pm} = \frac{x \pm 0.5}{n} . \quad (6.30)$$

Assumendo l'approssimazione normale (dove $|t_{\alpha/2}| = t_{1-\alpha/2}$), si ottiene il seguente intervallo di stima:

$$p \in \frac{f_{\pm} + \frac{t_{\alpha/2}^2}{2n}}{\frac{t_{\alpha/2}^2}{n} + 1} \pm \frac{|t_{\alpha/2}| \sqrt{\frac{t_{\alpha/2}^2}{4n^2} + \frac{f_{\pm}(1-f_{\pm})}{n}}}{\frac{t_{\alpha/2}^2}{n} + 1} , \quad (6.31)$$

che in genere assicura una buona sovra-copertura ed è mediamente meno ampio dell'intervallo che si ottiene dalle (6.28, 6.29) [Rot10]. Le proprietà di copertura delle (6.28, 6.29, 6.31), in funzione della dimensione del campione e dei valori di probabilità, saranno esplorate più avanti nell'ambito delle tecniche di simulazione, nel par. 7.11 a pagina 295, la cui lettura è fortemente consigliata ai lettori interessati alla stima della probabilità.

Gli intervalli di stima appena definiti sono calcolati, mediante il codice `Stiprob` che riportiamo a pagina 195.

Durante l'inserimento dei dati, l'utente può scegliere, tramite una finestra di dialogo, se fare la stima a due code simmetriche oppure se procedere alla stima ad una coda a destra (ricerca dell'estremo inferiore) o a sinistra (estremo superiore). Nel primo caso `Stiprob` risolve le equazioni (6.28, 6.29) con una ricerca dicotomica che trova p_1 e p_2 dimezzando il passo s ad ogni ciclo, finché il valore $c_1 = c_2 = (1 - CL)/2$ calcolato s_1 o s_2 differisce da quello previsto SL_1 o SL_2 a meno della quantità `SMALL`, che determina l'accuratezza del procedimento. Nella ricerca dicotomica viene usata la routine `cdfbin(PQ,X,N,p,1-p)` che calcola i valori cumulativi (indicati con la sigla PQ) della distribuzione binomiale $b(X; N, p)$. Si noti che, nel caso a due code, viene calcolato e riportato in uscita anche l'intervallo approssimato della (6.31). Nei casi di stima a una coda SL_1 o SL_2 vengono posti pari a $(1 - CL)$ e si utilizza la routine `p=cdfbin(PrOmpr,CL,1-CL,x,N)`, che con il codice d'uso `PrOmpr` trova direttamente la probabilità p della binomiale $b(x; N, p)$ per la quale il valore misurato x è un quantile pari a CL .

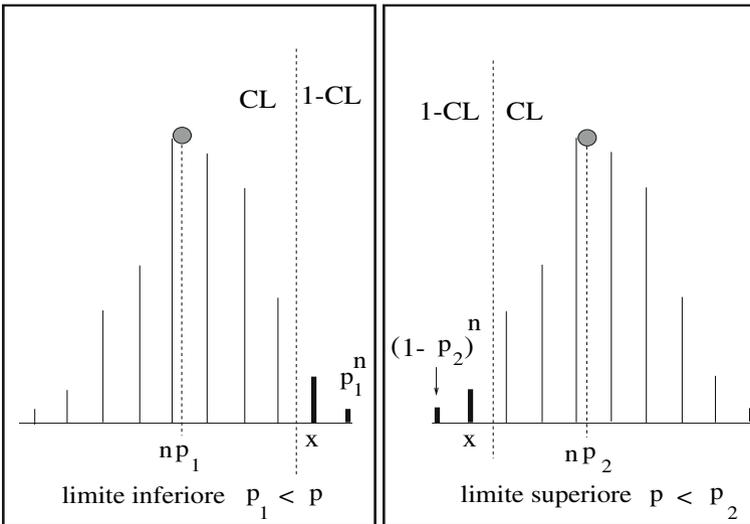


Fig. 6.5. Stima dei limiti inferiori e superiori di un probabilità con CL assegnato. Nei casi limite in cui $x = 0$ e $x = n$ si ha rispettivamente $(1 - p_2)^n = 1 - CL$ e $p_1^n = 1 - CL$

In uscita il codice stampa i valori p_1, p_2 e, per la stima a due code, il livello di confidenza calcolato $1-s_1-s_2$ che, a meno degli errori di arrotondamento numerico, coincide con il valore assegnato in ingresso.

```
// -----
//
// SCILAB script Stiprob: stima della probabilita' per piccoli
// campioni con il metodo di Neyman
// Versione interattiva.
//
// -----

SMALL=0.00001; LIMIT=0.005;
p1=0.5; p2=0.5; pp1=0.5; pp2=0.5;

st1="La stima a due code viene fatta";
st2="solo per valori p1<0.005 e (1-p2)<0.005.";
st3="Altrimenti si esegue la stima a una coda";
beep();
messagebox([st1;st2;st3]);

N = input("numero dei tentativi N:");
X = input("numero dei successi X:");
if X>N then warning("X deve essere <= N ! Stop"); abort; end;
CL = input("Livello di Confidenza CL (range [0,1]):");
beep();
YN=x_choose(["due code","lim. inf","lim. sup."],["Quale stima?"]);

// stima a due code

if YN==1 then
  SL1=0.5*(1-CL); SL2=SL1;

  //ricerca dicotomica di p1
  p1=0.5; s=p1; s1=0;
  while abs(s1-SL1)>SMALL & p1>SMALL,
    if X==0 then
      cum1=0;
    else
      cum1=cdfbin("PQ",X-1,N,p1,1-p1);
    end;
    s1=1 - cum1;
    s=s/2;
    if s1<SL1 then p1=p1+s;
    else p1=p1-s;
    end;
  end;
  p1=p1+0.5*s;
```

```

//ricerca dicotomica di p2
p2=0.5; s=p2; s2=0;
while abs(s2-SL2)>SMALL & (1-p2)>SMALL,
  if X==0 then
    cum2=0;
  else
    cum2=cdfbin("PQ",X,N,p2,1-p2);
  end;
  s2=cum2;
  s=s/2;
  if s2<SL2 then p2=p2-s;
  else p2=p2+s;
  end;
end;
p2 = p2 +0.5*s;
// valori approssimati con la formula di Wilson
P=0.5*(1.+CL);
ta = cdfnor("X",0.,1.,P,1.-P); // valore quantile gaussiano
ffp = max(0.,X/N - 0.5/N); // correzioni di continuita'
ffm = min(1.,X/N + 0.5/N);
psqm=sqrt(ta*ta/(4.*N*N) + ffm*(1.-ffm)/N);
psqp=sqrt(ta*ta/(4.*N*N) + ffp*(1.-ffp)/N);
pp1 = max(0.,(ffm + ta*ta/(2.*N) - ta*psqm)/(1. + ta*ta/N));
pp2 = min(1.,(ffp + ta*ta/(2.*N) + ta*psqp)/(1. + ta*ta/N));
end;

// stime a una coda

if (YN==3 | p1<LIMIT) then
  p2=cdfbin("PrOmpr",1-CL,CL,X,N);
elseif (YN==2 | 1-p2< LIMIT) then
  p1=cdfbin("PrOmpr",CL,1-CL,X-1,N);
end;

// mostra i risultati

write(%io(2)," -----");
if (YN==3 | p1<LIMIT) then
  write(%io(2)," eseguita stima a una coda ");
  if p1<LIMIT then
    write(%io(2)," a causa di p1<"+string(LIMIT)); end;
    write(%io(2)," Limite superiore CL= "+string(CL)+":");
    write(%io(2)," p2 = "+string(p2));
    write(%io(2)," media N*p2 = "+string(N*p2));
  elseif (YN==2 | 1-p2<LIMIT) then
    write(%io(2)," eseguita stima a una coda ");
    if 1-p2<LIMIT then
      write(%io(2)," a causa di (1-p2)<"+string(LIMIT)); end;
      write(%io(2)," Limite inferiore CL= "+string(CL)+":");
    end;
  end;
end;

```

```

write(%io(2)," p1 = "+string(p1));
write(%io(2)," media N*p1 = "+string(N*p1));
elseif YN==1 then
write(%io(2)," stima a due code con CL = "+string(CL));
write(%io(2)," CL ottenuto = "+string(1-s1-s2));
write(%io(2)," p1 = "+string(p1));
write(%io(2)," p2 = "+string(p2));
write(%io(2)," valori approssimati ");
write(%io(2)," p1 = "+string(pp1));
write(%io(2)," p2 = "+string(pp2));
end;
write(%io(2)," -----");

```

Nel caso $x = 0$ e nel caso $x = n$ le (6.28, 6.29), con $c_1 = c_2 = 1 - CL$, forniscono due importanti casi limite, mostrati nella Fig. 6.5:

$$x = n \implies p_1^n = 1 - CL, \quad (6.32)$$

$$x = 0 \implies (1 - p_2)^n = 1 - CL. \quad (6.33)$$

Da queste formule otteniamo, per un CL prefissato, il limite inferiore di una probabilità quando tutti i tentativi hanno avuto successo:

$$p_1 = \sqrt[n]{1 - CL} = e^{\frac{1}{n} \ln(1 - CL)} = 10^{\frac{1}{n} \log(1 - CL)} \quad (6.34)$$

e il limite superiore quando non si è registrato alcun successo:

$$p_2 = 1 - \sqrt[n]{1 - CL} = 1 - e^{\frac{1}{n} \ln(1 - CL)} = 1 - 10^{\frac{1}{n} \log(1 - CL)}, \quad (6.35)$$

dove l'uso dei logaritmi in base e o base 10 può essere utile nel caso di n molto grande. Nel codice `Stiprob` questi casi limite vengono trattati automaticamente se viene richiesta la stima ad una coda.

Nel caso della stima a due code, se si verifica la condizione $p_1 < \text{LIMIT}$ oppure $(1 - p_2) < \text{LIMIT}$, la stima a una coda viene forzata e, nei casi $x = 0$ e $x = n$, si eseguono le stime con le (6.34, 6.35). La costante `LIMIT` viene segnalata all'utente e può essere ovviamente modificata. Il valore che abbiamo assegnato è 0.05. In alternativa a `Stiprob`, potete utilizzare la Fig. 6.6, valida per $CL = 90\%$.

I limiti superiore e inferiore delle (6.34, 6.35) definiscono due intervalli ($0 \leq p \leq p_2$) e ($p_1 \leq p \leq 1$) la cui interpretazione frequentista è un po' sottile: su 100 esperimenti che stimano il limite inferiore p_1 con la (6.28) con $c_1 = 1 - CL$, mediamente una percentuale $100 \cdot CL$ di questi determinerà $p \in [p_1, 1]$ in modo corretto. Nel caso in cui tutti i tentativi abbiano avuto successo, dalla (6.34) risulta che, se $p < p_1$, si possono avere n successi in n tentativi, ma solo in una frazione di esperimenti minore di $1 - CL$. Un ragionamento simile vale per la stima del limite superiore (si veda anche, più avanti a pag. 202, la Fig. 6.7).

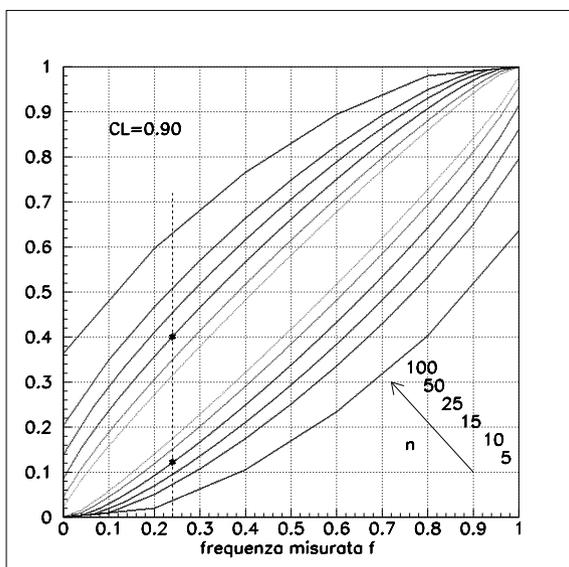


Fig. 6.6. Curve di Neyman corrispondenti ai limiti di probabilità per un $CL = 90\%$ in corrispondenza di una frequenza misurata. Andando dalla periferia verso il centro gli ellissoidi si riferiscono ad un numero di tentativi $n = 5, 15, 25, 50, 100$. Ad esempio, se in 25 tentativi si ottengono 6 successi, $f = 0.24$ e la probabilità vera è compresa entro i limiti $p_1 = 0.121$, $p_2 = 0.402$ come mostrato in figura

Quando n è grande e non si sono registrati successi, p_2 è piccola. Sviluppando al primo ordine l'esponenziale nella (6.35) intorno al punto iniziale $p_2 = 0$, si ottiene allora l'approssimazione:

$$p_2 \simeq -\frac{1}{n} \ln(1 - CL) , \quad (6.36)$$

che corrisponde all'equazione:

$$e^{-np_2} = (1 - CL) = \alpha , \quad (6.37)$$

la quale non è altro che è la probabilità poissoniana di non ottenere eventi quando la media è $\mu = np_2$.

Esercizio 6.3

In 20 prove bernoulliane indipendenti si sono ottenuti 5 eventi. Qual è la stima della probabilità dell'evento, con un CL del 90%?

Risposta. In questo caso abbiamo $x = 5$, $n = 20$, $f = 5/20 = 0.25$. Ci troviamo pertanto nel caso di un piccolo campionamento. Se introduciamo i dati $x = 5, n = 20, CL = 0.90$ nella routine **Stiprob** otteniamo i valori:

$$p_1 = 0.104 , \quad p_2 = 0.455 ,$$

che definiscono l'intervallo:

$$p \in [0.104, 0.455] \equiv 0.25^{+0.20}_{-0.15}, \quad CL = 90\%,$$

in accordo con la (6.17). Possiamo risolvere il problema anche in modo approssimato, applicando la (6.31) con un valore $t = 1.645$, dedotto dalla solita Tab. D.1 delle probabilità gaussiane, come valore intermedio delle aree 0.4495 e 0.4505. L'approssimazione consiste proprio in quest'ultima assunzione sui livelli gaussiani di t , e non nell'uso della (6.31), che è generale. Dalla (6.31) otteniamo i valori:

$$p \in [0.145, 0.405] = 0.28 \pm 0.13.$$

In modo ancora più approssimato, possiamo utilizzare la (6.22) sempre con lo stesso valore di t . Il risultato è:

$$p \in [0.09, 0.410] = 0.25 \pm 0.16.$$

Come si vede, i tre metodi forniscono risultati sensibilmente diversi. Questo fatto sarà approfondito in dettaglio con l'utilizzo di tecniche di simulazione, nel par. (7.11).

Esercizio 6.4

Da un'urna contenente 5 biglie bianche e nere in proporzione sconosciuta si eseguono 10 estrazioni (con reinserimento) e si estraggono 10 biglie nere. Trovare il limite inferiore del numero di biglie nere nell'urna per un $CL = 0.90$. Confrontare i risultati con quelli dell'esercizio 1.6 di pagina 32.

Risposta. Dalla (6.34) $p = (0.10)^{1/10} = 0.794$. Il limite inferiore del numero di biglie nere è $0.794 \cdot 5 = 3.97$, per cui possiamo affermare che l'urna contiene almeno 4 biglie nere con $CL = 0.90$. Un'urna con meno di 4 biglie nere può dare luogo a 10 estrazioni consecutive di 10 biglie nere, ma questo avviene in meno del 10% degli esperimenti.

Introducendo il concetto di "probabilità di sbagliare", che verrà utilizzato tra poco nella verifica delle ipotesi, possiamo anche dire che l'urna contiene almeno 4 biglie nere, con una probabilità di sbagliare del 10%.

È interessante confrontare questa soluzione frequentista del problema con quella bayesiana della Tab. 1.2 di pagina 33: la stima frequentista è indipendente da qualunque ipotesi a priori soggettiva sulla composizione iniziale dell'urna. Le ipotesi iniziali soggettive, come sappiamo, influenzano il risultato finale, come mostrano i risultati dell'esercizio 1.6 e del problema 1.12.

Le equazioni (6.34, 6.35) sono importanti in molti problemi di affidabilità, come mostrato negli esempi che seguono.

Esercizio 6.5

Un test di affidabilità sottopone una pompa di emergenza a 500 partenze a freddo. Se la pompa supera il test, qual è la probabilità che non parta in emergenza, con un livello di confidenza del 95%?

Risposta. Applicando la (6.35) si ottiene immediatamente:

$$p = 1 - \sqrt[500]{1 - CL} = 1 - \sqrt[500]{0.05} = 0.00597 ,$$

pari a circa lo 0.6%.

L'interpretazione di Neyman del risultato è la seguente: una pompa con probabilità di panne maggiore del 6 per mille può partire per 500 volte consecutive, ma questo accade in meno del 5% dei test. Si noti che questo risultato vale nello schema di prove indipendenti.

Esercizio 6.6

Quante prove consecutive senza guasti sono necessarie in un test per affermare, con una confidenza del 95%, che un dispositivo si guasterà in meno del 3% dei casi?

Risposta. Utilizzando i logaritmi decimali, dalla (6.35) si ottiene facilmente:

$$n = \frac{\log(1 - CL)}{\log(1 - p)} = \frac{\log(1 - 0.95)}{\log(1 - 0.03)} = 98.4 , \quad (6.38)$$

pari a circa 100 prove. Un dispositivo con probabilità di guasto $> 3\%$ può affrontare con successo 100 prove, ma questo succede in meno del 5% dei test.

6.8 Intervalli di stima per eventi poissoniani

La determinazione di intervalli di confidenza può essere estesa dal caso binomiale anche al caso poissoniano. Ad esempio, si può utilizzare in modo approssimato il codice `Stiprob` di pagina 195 con n molto grande (ad esempio $n = 1000$), oppure, in corrispondenza di un numero x di eventi osservato, stimare esattamente μ risolvendo le equazioni, analoghe alle (6.28, 6.29):

$$\sum_{k=x}^{\infty} \frac{\mu_1^k}{k!} \exp(-\mu_1) = c_1 , \quad \sum_{k=0}^x \frac{\mu_2^k}{k!} \exp(-\mu_2) = c_2 , \quad (6.39)$$

dove nella prima equazione la somma va estesa fino a valori tali da rendere trascurabile l'errore di troncamento. Nel caso simmetrico si ha come sempre $c_1 = c_2 = (1 - CL)/2$. Anche qui la presenza, in entrambe le somme, del valore

osservato x assicura la sovra-copertura dell'intervallo e il soddisfacimento della (6.6).

In approssimazione gaussiana, dato che per la poissoniana $\sigma^2 = \mu$, l'intervallo di confidenza per il valore atteso μ può essere trovato dalla quantità pivotale

$$\frac{|x - \mu|}{\sqrt{\mu}} \leq |t_{\alpha/2}|, \quad (6.40)$$

che è distribuita secondo la normale standard. Anche qui i valori $|t_{\alpha/2}| = t_{1-\alpha/2}$, come nella (6.31), rappresentano i valori quantili della gaussiana standard agli estremi dell'intervallo che sottende un'area pari a $CL = 1 - \alpha$. Introducendo, come nella (6.31), la correzione di continuità

$$x_{\pm} = x \pm 0.5 \text{ se } x \neq 0, \quad (6.41)$$

e risolvendo la (6.40) rispetto a μ si ottiene:

$$\mu \in x_{\pm} + \frac{t_{\alpha/2}^2}{2} \pm |t_{\alpha/2}| \sqrt{x_{\pm} + \frac{t_{\alpha/2}^2}{4}}. \quad (6.42)$$

L'esperienza numerica con dati simulati, che noi tratteremo più avanti nel par. 7.11 a pagina 295, mostra che questo intervallo ha ottime proprietà di sovra-copertura e può essere usato in alternativa alle (6.39) per $x > 10$ [Rot10]. Quando $x > 100$ la (6.42) può essere sostituita dall'intervallo usuale:

$$\mu \in x \pm |t_{\alpha/2}| \sqrt{x}. \quad (6.43)$$

La seconda delle (6.39), con $c_2 = 1 - CL$, permette anche di risolvere in modo generale e non approssimato il problema di trovare, per un CL assegnato, il limite superiore poissoniano al numero medio di eventi avendo ottenuto x eventi. Per $x = 0, 1, 2$, ponendo $\mu_2 \equiv \mu$, si ha ad esempio (si veda anche la Fig. 6.7):

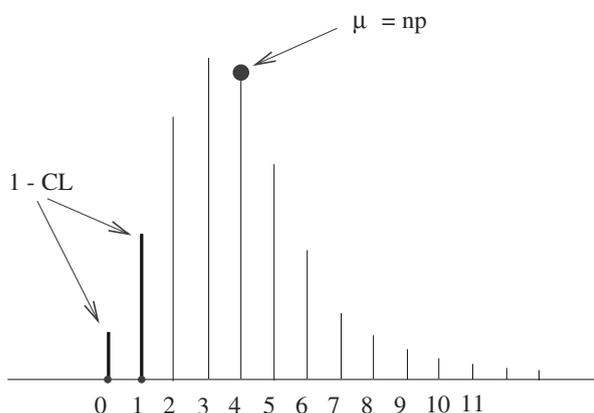
$$\begin{aligned} e^{-\mu} &= 1 - CL, \\ e^{-\mu} + \mu e^{-\mu} &= 1 - CL, \\ e^{-\mu} + \mu e^{-\mu} + \frac{\mu^2}{2} e^{-\mu} &= 1 - CL \end{aligned}$$

e così via. La Tab. 6.2 può evitarvi la fatica di risolvere queste equazioni. A titolo di esempio, essa mostra che, quando $\mu > 2.3$, mediamente *non si otterranno successi* in un numero di esperimenti $< 10\%$. Analogamente, se 4.74 è il limite superiore per $x = 1$ e $CL = 95\%$, significa che, quando $\mu > 4.74$, si possono ottenere i valori $x = 0, 1$ solo in un numero di esperimenti $< 5\%$, come mostrato anche in Fig. 6.7.

In alternativa alla tabella, si possono calcolare le probabilità cumulative poissoniane con la routine SCILAB `cdfpoi` invocata col modo d'uso `Xlam`.

Tabella 6.2. Limiti superiori μ_2 poissoniani al numero medio di eventi per x eventi osservati, per livelli di confidenza del 90 e 95%

x	90 %	95 %	x	90 %	95 %
0	2.30	3.00	6	10.53	11.84
1	3.89	4.74	7	11.77	13.15
2	5.32	6.30	8	13.00	14.44
3	6.68	7.75	9	14.21	15.71
4	7.99	9.15	10	15.41	16.96
5	9.27	10.51	11	16.61	18.21

**Fig. 6.7.** Rappresentazione grafica della seconda delle (6.39) per $x = 1$

Con la chiamata `cdfpoi(Xlam,CL,1-CL,x-1)` essa risolve la prima delle (6.39) fornendo il valore di μ_1 . Il limite superiore μ_2 della seconda delle (6.39) si ottiene con la chiamata `cdfpoi(Xlam,1-CL,CL,x)`. Ad esempio, potete verificare facilmente come, con la semplice istruzione `mu=cdfpoi(Xlam,.1,.9,0)`, si ottenga il limite superiore `mu=2.302585` per $x = 0$ e $CL = 90\%$, in accordo con la Tab. (6.2).

Esercizio 6.7

Si sono registrati 23 conteggi. Trovare il limite superiore con $CL = 0.95$.

Risposta. Dalla Tab. D.1 si vede che il valore quantile corrispondente alla coda di area $\alpha/2 = 0.500 - 0.475 = 0.025$ vale $|t_{\alpha/2}| = 1.96$. In approssimazione gaussiana, lecita perché $\mu > 10$, possiamo allora scrivere, in accordo con la Fig. 6.8:

$$\frac{\mu - 23}{\sqrt{\mu}} = 1.96 ,$$

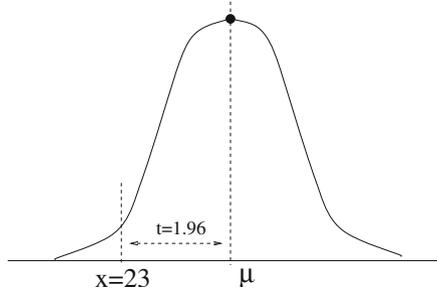


Fig. 6.8. Determinazione dell'estremo superiore μ in presenza di 23 conteggi, col metodo dell'esercizio 6.7

arrivando alla equazione di secondo grado in $\sqrt{\mu}$:

$$\mu - 1.96\sqrt{\mu} - 23 = 0 .$$

La soluzione positiva di questa equazione vale $\sqrt{\mu} = 5.874$, quindi il valore cercato è $\mu = 34.5$. Si noti che la soluzione approssimata

$$\mu = 23 + 1.96\sqrt{23} = 32.4$$

differisce sensibilmente dal risultato corretto in approssimazione gaussiana.

6.9 Stima della media da grandi campioni

A differenza della media vera μ , che è una quantità fissa, la media di un campione (o media campionaria) è una variabile aleatoria. Infatti se produciamo un campione casuale di dimensione N da una distribuzione, ne facciamo la media:

$$m = \frac{1}{N} \sum_{i=1}^N x_i \quad (6.44)$$

e ripetiamo il procedimento più volte, otterremo ogni volta un risultato diverso. La media campionaria va quindi vista come uno stimatore

$$M = \frac{1}{N} \sum_{i=1}^N X_i ,$$

del tipo (2.9, 2.70), secondo lo schema di pagina 64. Applicando l'operatore varianza alla variabile M otteniamo, sfruttando le solite proprietà (4.19, 5.73):

$$\text{Var}[M] = \text{Var} \left[\frac{1}{N} \sum_{i=1}^N X_i \right] = \frac{1}{N^2} \sum_{i=1}^N \text{Var}[X_i] .$$

Poiché le variabili X_i appartengono tutte allo stesso campione e sono indipendenti, $\text{Var}[X_i] = \sigma^2$, dove σ^2 è la varianza della distribuzione campionaria. Otteniamo dunque l'importante e semplice risultato:

$$\text{Var}[M] = \frac{1}{N^2} \sum_{i=1}^N \sigma^2 = \frac{1}{N^2} N \sigma^2 = \frac{\sigma^2}{N}. \quad (6.45)$$

Poiché σ^2 in genere non è nota, possiamo approssimarla con la varianza s^2 calcolata direttamente dal campione ed ottenere così l'intervallo ad 1σ

$$\mu \in m \pm \frac{\sigma}{\sqrt{N}} \simeq m \pm \frac{s}{\sqrt{N}} \quad (6.46)$$

centrato sul valor medio e di ampiezza pari all'errore statistico s/\sqrt{N} , che *contiene la media vera*. Con quale livello di confidenza? Il teorema Limite Centrale 3.1 afferma che la densità della media campionaria sarà gaussiana per $N \gg 1$, condizione che in pratica diventa $N > 10$. Nel caso gaussiano dalla (6.9) risulta allora che i livelli di confidenza cercati sono dati dai livelli di probabilità rispetto alla media della corrispondente densità gaussiana. Il problema è quindi completamente risolto, almeno per grandi campioni.

Spesso, per indicare esplicitamente che per N grande i livelli di confidenza sono gaussiani, la (6.46) viene messa nella forma:

$$P\{X_1 + X_2 + \dots + X_N \leq x\} \simeq \Phi\left(\frac{x - N\mu}{\sigma\sqrt{N}}\right), \quad (6.47)$$

dove le N variabili X_i hanno media μ e deviazione standard σ e Φ è la funzione cumulativa della gaussiana standard (3.43).

Vediamo ora per quali valori di N è lecito sostituire nella (6.46) la varianza (o la deviazione standard) vera con quella osservata.

Rappresentiamo il campione discusso nel paragrafo precedente, per la stima della probabilità, con un istogramma, assegnando il valore 0 agli insuccessi, ed il valore 1 ai successi. Nell'approssimazione $N - 1 \simeq N$, la media e la varianza di questo campione si ottengono dalle (2.52, 2.54):

$$\begin{aligned} m &= \sum_i x_i f_i = 0 \cdot \frac{N-x}{N} + 1 \cdot \frac{x}{N} = \frac{x}{N} \equiv f \\ s^2 &= \sum_i (x_i - m)^2 f_i = (-m)^2 \frac{N-x}{N} + (1-m)^2 \frac{x}{N} \\ &= f^2(1-f) + (1+f^2-2f)f = f(1-f), \end{aligned}$$

da cui si vede che, con la rappresentazione scelta, la media misurata del campione è uguale alla frequenza f e la varianza misurata è uguale a $f(1-f)$. Utilizzando ora la (6.46), otteniamo immediatamente:

$$\mu \equiv p = m \pm \frac{s}{\sqrt{N}} = f \pm \sqrt{\frac{f(1-f)}{N}}. \quad (6.48)$$

Per capire bene il significato di questo risultato, è bene ricordare che esso si basa sulla sostituzione della varianza vera con quella misurata e sull'approssimazione per grandi campioni $N \simeq N - 1$.

Nel caso della media, la (6.46) è una buona approssimazione per $N > 20, 30$, ed i livelli di confidenza sono gaussiani per $N > 10$. Dato che la condizione $N > 20, 30$ è quella prevalente, possiamo scrivere la formula che stima ad 1σ la media vera per grandi campioni come:

$$\mu \in m \pm \frac{s}{\sqrt{N}}, \quad (N > 20, 30, CL \text{ gaussiani}). \quad (6.49)$$

Tratteremo tra breve il caso dei piccoli campioni ($N < 20, 30$).

6.10 Stima della varianza da grandi campioni

In statistica si possono definire due tipi di varianze: l'una rispetto alla media vera, l'altra rispetto alla media campionaria:

$$S_\mu^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2, \quad S^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - M)^2, \quad (6.50)$$

le quali, per $N \rightarrow \infty$, tendono, nel senso della (2.76), alla varianza vera σ^2 . In genere si indica semplicemente con S^2 la varianza rispetto ad M , che è detta varianza campionaria. Solo se sarà strettamente necessario, nel seguito distingueremo le due varianze con la notazione della (6.50).

Per prima cosa, diamo ragione del fattore $N - 1$ che compare in S^2 della (6.50). Il punto della questione sta nella relazione algebrica:

$$\begin{aligned} \sum_i (X_i - \mu)^2 &= \sum_i (X_i - M + M - \mu)^2 = \sum_i [(X_i - M) + (M - \mu)]^2 \\ &= \sum_i (X_i - M)^2 + \sum_i (M - \mu)^2 + 2(M - \mu) \sum_i (X_i - M) \\ &= \sum_i (X_i - M)^2 + N(M - \mu)^2, \end{aligned} \quad (6.51)$$

dove nell'ultimo passaggio si è sfruttato il fatto che $\sum_i (X_i - M) = 0$. Ora osservate bene la (6.51): essa indica che la dispersione dei dati intorno alla media vera è pari alla dispersione dei dati intorno alla media campionaria più un termine che tiene conto della dispersione della media campionaria intorno alla media vera. Questa somma di fluttuazioni è proprio l'origine del termine $N - 1$. Infatti, invertendo la (6.51), abbiamo:

$$\sum_i (X_i - M)^2 = \sum_i (X_i - \mu)^2 - N(M - \mu)^2. \quad (6.52)$$

Applichiamo ora l'operatore di media a tutti i membri di questa relazione, seguendo la tecnica discussa nel par. 2.11. Dalle (2.61, 4.8, 6.45) otteniamo:

$$\begin{aligned} \left\langle \sum_i (X_i - M)^2 \right\rangle &= \sum_i \langle (X_i - M)^2 \rangle = \sum_i \langle (X_i - \mu)^2 \rangle - N \langle (M - \mu)^2 \rangle \\ &= \left(\sum_i \right) \sigma^2 - N \frac{\sigma^2}{N} = N\sigma^2 - \sigma^2 = (N - 1)\sigma^2 . \end{aligned} \quad (6.53)$$

Si ha in definitiva la giustificazione della seconda delle (6.50), poiché risulta che lo stimatore S^2 è non distorto in quanto gode della proprietà (2.78):

$$\langle S^2 \rangle = \left\langle \frac{1}{N-1} \sum_i (X_i - M)^2 \right\rangle = \sigma^2 . \quad (6.54)$$

La varianza campionaria (non corretta)

$$\frac{1}{N} \sum_i (X_i - M)^2 \quad (6.55)$$

è un esempio di *stimatore non corretto o distorto (biased)*. Infatti dalla (6.53) risulta:

$$\left\langle \frac{1}{N} \sum_{i=1}^N (X - M)^2 \right\rangle = \frac{N-1}{N} \sigma^2 . \quad (6.56)$$

Come abbiamo già accennato nel par. 2.11, uno stimatore è distorto quando la media degli stimatori T_N , calcolati da un campione di dimensione N , differisce dal limite per $N \rightarrow \infty$ (2.76) dello stimatore. Nel nostro caso, la media differisce dal parametro che si vuole stimare, che è la varianza, per un fattore $(N-1)/N = 1 - 1/N$. Il termine $1/N$ si chiama *fattore di distorsione (bias)*; nel nostro caso però esso si annulla al limite per $N = \infty$. Si dice allora che lo stimatore è *asintoticamente corretto*, poiché la media di stimatori ottenuti con N sufficientemente grande è molto vicina al limite in probabilità dello stimatore. Tutti questi concetti saranno poi approfonditi nel cap. 9.

L'equazione (6.54) si può anche interpretare richiamando il concetto di *gradi di libertà*, introdotto per la prima volta nel par. 3.9 a proposito della distribuzione χ^2 . Se si stima la media dai dati, la somma degli N quadrati degli scarti è in realtà formata da $N - 1$ termini, perché la media è vincolata ad essere proporzionale alla somma dei dati. La varianza di un campione è pertanto una grandezza univoca, a patto che *la media dei quadrati degli scarti venga fatta sui gradi di libertà della statistica*. In modo abbastanza generale, possiamo quindi dire:

Definizione 6.3 (Gradi di libertà di una statistica). *Il numero dei gradi di libertà ν di una statistica $T = t(X_1, X_2, \dots, X_N, \hat{\theta})$ dipende da un*

insieme noto di parametri $\hat{\theta}$, è dato dal numero N di variabili del campione meno il numero k di parametri $\hat{\theta}$ stimati dai dati:

$$\nu = N - k . \quad (6.57)$$

Cerchiamo ora di trovare ... la varianza della varianza. Questo non è un paradosso, perché la varianza campionaria, come la media, è una variabile aleatoria, che tende alla varianza vera per $N \rightarrow \infty$. Analogamente a quanto abbiamo fatto per la media, possiamo applicare l'operatore varianza alle quantità S_μ^2 e S^2 della (6.50) ed applicare la trasformazione (5.73). Il formalismo operatoriale, definito nelle (2.59, 2.66) e già applicato nella (6.53), semplifica di molto il calcolo, almeno nel caso di S_μ^2 . Abbiamo infatti:

$$\begin{aligned} \text{Var}[S_\mu^2] &= \frac{1}{N^2} \sum_i \text{Var}[(X_i - \mu)^2] = \frac{1}{N^2} \sum_i [\langle (X_i - \mu)^4 \rangle - \langle (X_i - \mu)^2 \rangle^2] \\ &= \frac{1}{N^2} [N\Delta_4 - N\sigma^4] = \frac{1}{N} (\Delta_4 - \sigma^4) , \end{aligned}$$

dove il momento di ordine 4 è stato indicato con la notazione della (2.58). Sostituendo ai valori veri Δ_4 e σ^2 i valori stimati dai dati s_μ^2 e

$$D_4 = \frac{1}{N} \sum_i (X_i - \mu)^4 ,$$

otteniamo la relazione:

$$\text{Var}[S_\mu^2] \simeq \frac{D_4 - s_\mu^4}{N} , \quad (6.58)$$

che permette di scrivere ad 1σ l'intervallo di confidenza della varianza come:

$$\sigma^2 \in s_\mu^2 \pm \sigma[S_\mu^2] = s_\mu^2 \pm \sqrt{\frac{D_4 - s_\mu^4}{N}} .$$

Un calcolo di questo genere può essere ripetuto per la varianza S^2 , ma in questo caso bisogna tenere conto anche della varianza della media campionaria M intorno alla media vera. Il risultato finale, abbastanza laborioso da ottenere, mostra che, rispetto alla (6.58), compaiono termini correttivi di ordine $1/N^2$ che pertanto, per grandi campioni, possono essere trascurati. La (6.58) può allora essere utilizzata anche in questo caso, sostituendo ad N il numero $(N - 1)$ di gradi di libertà. Nel seguito utilizzeremo la notazione:

$$\sigma^2 \in \begin{cases} s_\mu^2 \pm \sqrt{\frac{D_4 - s_\mu^4}{N}} & \text{media nota,} \\ s^2 \pm \sqrt{\frac{D_4 - s^4}{N - 1}} & \text{media ignota .} \end{cases} \quad (6.59)$$

Quali sono i livelli di confidenza da associare alla (6.59)? Si può mostrare che la varianza campionaria per variabili X qualsiasi tende alla densità gaussiana,

ma solo per campioni con $N > 100$. Se le variabili del campione sono gaussiane, allora, come vedremo meglio nel prossimo paragrafo, la distribuzione campionaria della varianza è riconducibile alla densità χ^2 , la quale converge alla densità gaussiana un po' più velocemente, all'incirca per $N > 30$. In generale la variabile aleatoria "varianza campionaria" tende a diventare una variabile gaussiana molto più lentamente della corrispondente media campionaria, che lo è già per $N > 10$. Questo fatto è del tutto naturale se si tiene conto che nella varianza la combinazione delle variabili campionarie è quadratica invece che lineare.

Per trovare l'intervallo di confidenza della deviazione standard, conoscendo l'intervallo $[s_1^2, s_2^2]$ della varianza per un CL assegnato, basta porre $\sigma \in [\sqrt{s_1^2}, \sqrt{s_2^2}]$. In modo approssimato, basta applicare la legge di trasformazione non lineare (5.56), dove $x = s^2$ e $z = s = \sqrt{x} = \sqrt{s^2}$. Ritenendo solo il primo termine, abbiamo:

$$\text{Var}[S] \simeq \frac{1}{4S^2} \text{Var}[S^2] = \frac{D_4 - s^4}{4(N-1)s^2} . \quad (6.60)$$

L'intervallo di confidenza ad 1σ per la deviazione standard è allora:

$$\sigma \in s \pm \sqrt{\frac{D_4 - s^4}{4(N-1)s^2}} , \quad (N > 100 , CL \text{ gaussiani}) . \quad (6.61)$$

Se le variabili che compongono il campione sono gaussiane, allora vale la relazione (3.33) $\Delta_4 = 3\sigma^4$, e le (6.59, 6.61) divengono rispettivamente:

$$\sigma^2 \in s^2 \pm \sigma^2 \sqrt{\frac{2}{N-1}} , \quad \sigma \in s \pm \frac{\sigma}{\sqrt{2(N-1)}} , \quad (6.62)$$

da cui, per $CL = 1 - \alpha$:

$$\frac{s^2}{1 + t_{1-\alpha/2}\sqrt{2/(N-1)}} \leq \sigma^2 \leq \frac{s^2}{1 - t_{1-\alpha/2}\sqrt{2/(N-1)}} , \quad (6.63)$$

$$\frac{s}{1 + t_{1-\alpha/2}\sqrt{1/[2(N-1)]}} \leq \sigma \leq \frac{s}{1 - t_{1-\alpha/2}\sqrt{1/[2(N-1)]}} , \quad (6.64)$$

dove $t_{1-\alpha/2}$ sono i quantili della densità gaussiana e dove a $N-1$ va sostituito N se le dispersioni sono calcolate rispetto alla media vera.

Le formule (6.49, 6.59, 6.61, 6.63, 6.64) risolvono il problema della stima di media, varianza e deviazione standard per grandi campioni. A volte, invece delle (6.63, 6.64), si usano un po' impropriamente le (6.62) dove a σ si sostituisce s nella formula dell'errore. Daremo tra breve alcuni esempi di stime statistiche, dopo che avremo trattato il problema della stima con piccoli campioni.

6.11 Stima di media e varianza da piccoli campioni

La stima di media e varianza per piccoli campioni ($N < 20, 30$ per la media e $N < 100$ per la varianza) porta ad intervalli e livelli di confidenza che dipendono dal tipo di popolazione del campione. Per questa ragione non è possibile, a differenza di quanto appena visto per i grandi campioni, arrivare in questo caso a formule di validità generale.

Tuttavia, applicando le nozioni di calcolo delle probabilità che abbiamo fin qui sviluppato, è possibile arrivare a una soluzione semplice e completa almeno per il caso più frequente ed importante, quello relativo ai campioni estratti da densità gaussiane.

La teoria dei campioni gaussiani si basa su due punti fondamentali. Il primo punto è che, come mostrato nell'esercizio 5.3, la media campionaria M , se è la somma di N variabili gaussiane, è anch'essa gaussiana per qualunque N . Allora, in base alla (6.45), anche la variabile

$$\frac{M - \mu}{\sigma} \sqrt{N}$$

è una variabile standard gaussiana per qualunque N , cioè una quantità pivotale. Il secondo punto può essere condensato in due importanti teoremi:

Teorema 6.1 (Indipendenza di M e S^2). *Se (X_1, X_2, \dots, X_N) è un campione casuale di dimensione N estratto da una popolazione avente densità normale $g(x; \mu, \sigma)$, M ed S^2 sono variabili aleatorie indipendenti.*

Dimostrazione. Dato che le X_i sono variabili gaussiane indipendenti, esse hanno una densità congiunta prodotto di N gaussiane monodimensionali. Introducendo le variabili standard $T_i = (X_i - \mu)/\sigma$, la densità congiunta diventa:

$$g(\mathbf{t}) = \frac{1}{(2\pi)^{N/2}} \exp \left[-\frac{1}{2} \sum_{i=1}^N t_i^2 \right]. \quad (6.65)$$

È facile verificare le relazioni:

$$\frac{1}{N} \sum_{i=1}^N T_i = \frac{M - \mu}{\sigma}, \quad \sum_{i=1}^N (T_i - M_t)^2 = (N - 1) \frac{S^2}{\sigma^2}, \quad (6.66)$$

dove $M = \sum_i X_i/N$, $M_t = \sum_i T_i/N$ sono le medie campionarie di X e T ed S^2 è data dalla (6.50).

Poiché la gaussiana standard $g(\mathbf{t})$ ha simmetria sferica, essa ha forma funzionale invariante per trasformazioni ortogonali del tipo

$$T_i = \sum_{j=1}^N a_{ij} Z_j, \quad \sum_{i=1}^N T_i^2 = \sum_{i=1}^N Z_i^2, \quad (6.67)$$

corrispondenti ad una rotazione degli assi. Le nuove variabili Z_i sono anch'esse gaussiane standard indipendenti $N(0, 1)$. Scegliamo ora una trasformazione ortogonale che soddisfa alle (6.67) e tale che:

$$\begin{aligned} Z_1 &= \frac{1}{\sqrt{N}} \sum_{i=1}^N T_i = \frac{M - \mu}{\sigma/\sqrt{N}}, \\ \sum_{i=2}^N Z_i^2 &= \sum_{i=1}^N T_i^2 - \frac{1}{N} \left(\sum_{i=1}^N T_i \right)^2 = \sum_{i=1}^N T_i^2 - \frac{2}{N} \left(\sum_{i=1}^N T_i \right)^2 + \frac{1}{N} \left(\sum_{i=1}^N T_i \right)^2 \\ &= \sum_{i=1}^N T_i^2 - \frac{2}{N} \sum_{i=1}^N \sum_{j=1}^N T_i T_j + \frac{1}{N^2} \sum_{i=1}^N \left(\sum_{j=1}^N T_j \right)^2 = \frac{(N-1)S^2}{\sigma^2}, \end{aligned}$$

dove si è tenuto conto delle (6.66) e del fatto che $\sum_{i=1}^N = N$. Risulta dunque che la (6.65) si può scrivere come:

$$g(\mathbf{t}) = \frac{1}{(2\pi)^{N/2}} \exp \left[-\frac{1}{2} \frac{(m - \mu)^2}{\sigma^2/N} \right] \exp \left[-\frac{1}{2} \frac{(N-1)s^2}{\sigma^2} \right].$$

Dal teorema di fattorizzazione 4.1 segue allora l'indipendenza di M e S^2 . \square

Teorema 6.2 (Varianza campionaria). *Se (X_1, X_2, \dots, X_N) è un campione casuale estratto da una popolazione avente densità normale $g(x; \mu, \sigma)$, la variabile*

$$Q_R = \frac{S^2}{\sigma^2} = \frac{1}{N-1} \sum_i \frac{(X_i - M)^2}{\sigma^2} \quad (6.68)$$

ha densità (3.72) χ^2 ridotto con $N-1$ gradi di libertà. Essa è pertanto una quantità pivotale rispetto a σ^2 .

Dimostrazione. Dal teorema precedente risulta che $(N-1)S^2/\sigma^2$ è somma dei quadrati di $(N-1)$ variabili standard indipendenti Z_i ($i = 2, 3, \dots, N$), da cui l'asserto.

Si noti che dividendo la (6.52) per σ^2 e scambiando i termini risulta:

$$\underbrace{(N-1) \frac{S^2}{\sigma^2}}_{N-1} + \underbrace{\frac{(M-\mu)^2}{\sigma^2/N}}_1 = \underbrace{\sum_i \frac{(X_i - \mu)^2}{\sigma^2}}_{N \text{ gradi di libertà}},$$

in accordo col teorema di additività 3.4 delle variabili χ^2 . \square

L'intervallo di confidenza per la media può allora essere trovato sfruttando i risultati dell'esercizio 5.5. Infatti la variabile (5.41), che in questo caso vale:

$$T = \frac{M - \mu}{\sigma} \sqrt{N} \frac{1}{\sqrt{Q_R}} = \frac{M - \mu}{\sigma} \sqrt{N} \frac{\sigma}{S} = \frac{M - \mu}{S} \sqrt{N}, \quad (6.69)$$

essendo il rapporto tra una variabile gaussiana standard e la radice quadrata di una variabile χ^2 ridotto, tra loro indipendenti, si distribuisce secondo la nota densità di Student con $N - 1$ gradi di libertà, i cui quantili sono tabulati in Tab. D.2 (notate che nella (6.69) compare N , ma che la variabile T ha densità con $N - 1$ gradi di libertà). La (6.69) mostra che l'uso della variabile di Student permette di eliminare dall'intervallo di confidenza la varianza vera (incognita) σ^2 e di definire una quantità pivotale per la stima di μ . Infatti, se $t_{1-\alpha/2} = -t_{\alpha/2}$ sono i valori quantili di T corrispondenti al livello di confidenza $CL = 1 - \alpha$ fissato come in Fig. 6.9, invertendo la (6.69) si ha:

$$m - t_{1-\alpha/2} \frac{s}{\sqrt{N}} \leq \mu \leq m + t_{1-\alpha/2} \frac{s}{\sqrt{N}} . \tag{6.70}$$

Come è facile vedere dalla Tab. D.2 e dalla Fig. 5.4, per $n > 20, 30$ la densità di Student in pratica coincide con la densità gaussiana e la (6.70) coincide con la (6.49).

Rimandiamo a tra poco gli esercizi sull'uso di queste formule ed occupiamoci prima della stima della varianza. In questo caso la (6.68) fornisce la quantità pivotale $S^2/\sigma^2 \sim \chi_R^2(N-1)$. Seguendo il procedimento indicato nelle (6.11, 6.12), definiamo prima l'intervallo di probabilità:

$$\chi_{R(\alpha)}^2 \leq \frac{S^2}{\sigma^2} \leq \chi_{R(1-\alpha/2)}^2 , \tag{6.71}$$

dove $\chi_{R(\alpha/2)}^2, \chi_{R(1-\alpha/2)}^2$ sono i valori quantili della variabile Q_R corrispondenti al livello di confidenza CL prefissato, come mostrato in Fig. 6.9. Quindi, dall'intervallo (6.71) otteniamo per inversione l'intervallo di stima della

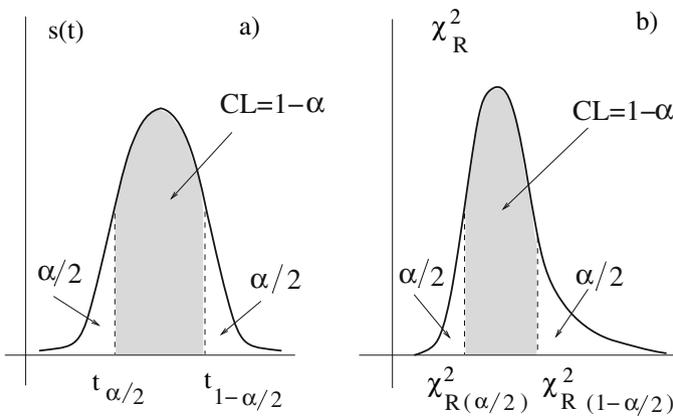


Fig. 6.9. Intervalli di confidenza associati ai valori quantili $t_{\alpha/2} < t_{1-\alpha/2}$ della variabile di Student a) e della variabile chi-quadrato ridotto b) nelle stime di media e varianza con piccoli campioni gaussiani

varianza in corrispondenza di un valore misurato s^2 :

$$\frac{s^2}{\chi_{R(1-\alpha/2)}^2} \leq \sigma^2 \leq \frac{s^2}{\chi_{R(\alpha/2)}^2}, \quad N - 1 \text{ gradi di libertà.} \quad (6.72)$$

I quantili α sono legati al livello CL dalle (6.14).

L'intervallo di confidenza per la deviazione standard è dato semplicemente da:

$$\frac{s}{\sqrt{\chi_{R(1-\alpha/2)}^2}} \leq \sigma \leq \frac{s}{\sqrt{\chi_{R(\alpha/2)}^2}}. \quad (6.73)$$

Per valori $N > 30$, la densità χ^2 può già considerarsi quasi gaussiana, quindi simmetrica, centrata sul valor medio $\langle Q_R \rangle = 1$ e con deviazione standard

$$\sigma_\nu \equiv \sigma [Q_R] = \sqrt{\frac{2}{N-1}},$$

in accordo con le (3.71). In questo caso si possono scrivere i valori quantili come:

$$\chi_{R(\alpha/2)}^2 = 1 - t_{1-\alpha/2} \sigma_\nu, \quad \chi_{R(1-\alpha/2)}^2 = 1 + t_{1-\alpha/2} \sigma_\nu,$$

e l'ampiezza dell'intervallo di confidenza con $CL = 1 - \alpha$ diventa:

$$\frac{s^2}{1 + t_{1-\alpha/2} \sigma_\nu} \leq \sigma^2 \leq \frac{s^2}{1 - t_{1-\alpha/2} \sigma_\nu} \quad (6.74)$$

che coincide con l'intervallo (6.63).

6.12 Come utilizzare la teoria della stima

Nei paragrafi precedenti abbiamo dedotto le formule fondamentali della stima di parametri. Esse sono quelle di uso corrente nelle analisi dei dati e vengono impiegate nei campi più diversi, dalla fisica, all'ingegneria e alla biologia. Il quadro d'insieme è riassunto nella Tab. 6.3.

La tabella mostra che, per quelli che in statistica vengono detti grandi campioni ($N > 30$ per la media ed $N > 100$ per la varianza), le formule ottenute risolvono il problema della stima in modo semplice e generale.

Risulta altresì evidente che il caso relativo alla stima della media e della dispersione per piccoli campioni non gaussiani resta ancora non ben definito. Tuttavia, questo è un caso che nella pratica capita assai raramente e per il quale non è possibile dare una soluzione generale, in quanto sia gli intervalli sia i livelli di confidenza dipendono dal tipo di distribuzione coinvolta nel problema. Le soluzioni, per alcuni casi di particolare interesse, si trovano nelle riviste specializzate o in libri un po' più tecnici del nostro. Se sfortunatamente vi capiterà di dover risolvere un problema di questo genere, potete consultare il testo [MGB91] ed i riferimenti lì riportati.

Tabella 6.3. Gli intervalli di confidenza ad 1σ e le distribuzioni da utilizzare per i livelli di confidenza (CL) nella stima di parametri. I campioni con $N > 30$ per la stima della media ed $N > 100$ per la stima della varianza sono detti in statistica *grandi campioni*

	<i>Variabili gaussiane</i>		<i>Variabili qualunque</i>	
	<i>Intervallo</i>	<i>CL</i>	<i>Intervallo</i>	<i>CL</i>
Probabilità				
$Nf < 10$	–	–	eq. (6.28, 6.29)	binomiale
$Nf > 10$				
$N(1 - f) > 10$	–	–	eq. (6.31)	Gauss
Frequenza				
$x < 10$	–	–	eq. (6.39)	Poisson
$x > 10$	–	–	eq. (6.40)	Gauss
Media				
$N < 30$	$m \pm \frac{s}{\sqrt{N}}$	Student	$\simeq m \pm \frac{s}{\sqrt{N}}$?
$N > 30$	$m \pm \frac{s}{\sqrt{N}}$	Gauss	$m \pm \frac{s}{\sqrt{N}}$	Gauss
Varianza				
$N < 100$	$\frac{s^2}{\chi_{R1}^2} \leq \sigma^2 \leq \frac{s^2}{\chi_{R2}^2}$	χ_R^2	$\simeq s^2 \pm \sqrt{\frac{D_4 - s^4}{N - 1}}$?
$N > 100$	eq. (6.63)	Gauss	$s^2 \pm \sqrt{\frac{D_4 - s^4}{N - 1}}$	Gauss
Dev. Std				
$N < 100$	$\sqrt{\frac{s^2}{\chi_{R1}^2}} \leq \sigma \leq \sqrt{\frac{s^2}{\chi_{R2}^2}}$	Maxwell	$\simeq s \pm \sqrt{\frac{D_4 - s^4}{4s^2(N - 1)}}$?
$N > 100$	eq. (6.64)	Gauss	$s \pm \sqrt{\frac{D_4 - s^4}{4s^2(N - 1)}}$	Gauss

Dalla Tab. 6.3 risulta che tutte le varianze degli stimatori frequenza, media, varianza, ecc. per grandi campioni sono della forma σ_0^2/N , dove σ_0 è una costante. Risulta pertanto pienamente verificata la condizione di Kolmogorov (2.75), sufficiente per la convergenza quasi certa di tutte queste stime verso i valori veri. Anche la convergenza debole risulta pertanto verificata, come si può vedere direttamente dalla disuguaglianza di Tchebychev. Infatti, la (3.93) può anche essere messa nelle forme:

$$P\{|X - \mu| \geq K \sigma\} \leq \frac{1}{K^2}, \quad P\{|X - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2},$$

dove, nell'ultima forma, si è posto $K = \epsilon/\sigma$. Se ora $T \equiv T_N$ è uno stimatore di Tab. 6.3, $\text{Var}[T_N] = \sigma_0^2/N$ ed otteniamo:

$$P\{|T_N - \mu| \geq \epsilon\} \leq \frac{\sigma_0^2}{N\epsilon^2}.$$

Per $N \rightarrow \infty$, la (2.72) è quindi verificata. La convergenza quasi certa della media campionaria verso la media vera con l'aumentare delle dimensione del campione è detta anche *legge dei grandi numeri*.

Copiando dal nostro sito la routine `Stimvs` e studiandone il codice, potete esplorare numericamente le formule per la stima di media e varianza nel caso gaussiano, per piccoli e grandi campioni.

Esercizio 6.8

Un campione gaussiano contenente N elementi ha media $m = 10$ e deviazione standard $s = 5$. Stimare, con un livello di confidenza del 95%, media, varianza e deviazione standard nei due casi $N = 10$ ed $N = 100$.

Risposta. Poiché il campione è gaussiano, possiamo utilizzare le (6.70, 6.72, 6.73), che richiedono la valutazione dei quantili $\alpha = 0.025$ e 0.975 delle distribuzioni di Student e χ^2 . Queste distribuzioni sono tabulate nell'Appendice D. Poiché la densità di Student è simmetrica, dalla Tab. D.2 risulta:

$$-t_{0.025} = t_{0.975} = \begin{cases} 2.26 \text{ per } N = 10 \text{ (9 gradi di libertà)} \\ 1.98 \text{ per } N = 100 \text{ (99 gradi di libertà)} \end{cases}.$$

Dalla (6.70) otteniamo allora la stima della media:

$$\begin{aligned} \mu &\in 10 \pm 2.26 \frac{5}{\sqrt{10}} = 10.0 \pm 3.6 \quad (N = 10, CL = 95\%) \\ \mu &\in 10 \pm 1.98 \frac{5}{\sqrt{100}} = 10.0 \pm 1.0 \quad (N = 100, CL = 95\%). \end{aligned} \quad (6.75)$$

Per la stima della dispersione, dobbiamo ricorrere ai quantili della densità χ_R^2 . Dalla Tab. D.3 otteniamo:

$$\chi_{R0.025}^2, \chi_{R0.975}^2 = \begin{cases} 0.30, 2.11 \text{ per } N = 10 \text{ (9 gradi di libertà)} \\ 0.74, 1.29 \text{ per } N = 100 \text{ (99 gradi di libertà)}. \end{cases}$$

Dalla (6.72) e dalla sua radice quadrata otteniamo allora la stima della varianza e della deviazione standard:

$$\begin{aligned} \sigma^2 &\in \left[\frac{25}{2.11}, \frac{25}{0.3} \right] = [11.9, 83.3] \quad (N = 10, CL = 95\%) \\ \sigma^2 &\in \left[\frac{25}{1.29}, \frac{25}{0.74} \right] = [19.3, 33.8] \quad (N = 100, CL = 95\%) \end{aligned} \quad (6.76)$$

$$\begin{aligned} \sigma &\in [\sqrt{11.9}, \sqrt{83.3}] = [3.4, 9.1] \quad (N = 10, CL = 95\%) \\ \sigma &\in [\sqrt{19.3}, \sqrt{33.8}] = [4.4, 5.9] \quad (N = 100, CL = 95\%). \end{aligned} \quad (6.77)$$

Come utile confronto, applichiamo le formule approssimate, che presuppongono livelli di confidenza gaussiani. Dall'esercizio 3.8, oppure direttamente da Tab. D.1, ricaviamo che in corrispondenza di $CL = 0.95$ si ha $t = 1.96$.

Per la stima della media possiamo allora utilizzare la (6.49):

$$\mu \in 10 \pm 1.96 \frac{5}{\sqrt{10}} = 10.0 \pm 3.1 \quad (N = 10) ,$$

$$\mu \in 10 \pm 1.96 \frac{5}{\sqrt{100}} = 10.0 \pm 1.0 \quad (N = 100) ,$$

la quale fornisce (entro gli arrotondamenti) un risultato identico alle (6.75) per $N = 100$, ed un risultato leggermente sottostimato per $N = 10$. Infatti, come si può vedere dalla Fig. 5.4, le code della densità di Student sottendono aree leggermente maggiori di quelle sottese dalla gaussiana standard.

Per stimare in modo approssimato i parametri di dispersione con $CL = 95\%$, utilizziamo le (6.63, 6.64) ancora con $t = 1.96$:

$$\sigma^2 \in \left[\frac{25}{1 + 1.96\sqrt{2/9}}, \frac{25}{1 - 1.96\sqrt{2/9}} \right] = [13.0, 328.7] \quad (N = 10) ,$$

$$\sigma^2 \in \left[\frac{25}{1 + 1.96\sqrt{2/99}}, \frac{25}{1 - 1.96\sqrt{2/99}} \right] = [19.5, 34.8] \quad (N = 100) ,$$

$$\sigma \in [3.6, 18.1] \quad (N = 10) ,$$

$$\sigma \in [4.4, 5.9] \quad (N = 100) .$$

Se confrontiamo questi risultati con i risultati corretti (6.76, 6.77), vediamo che le dispersioni approssimate sono accettabili solo per $N = 100$. Si possono usare anche le (6.62) con $s \simeq \sigma$ nei termini a destra, ottenendo gli intervalli $\sigma^2 \in [1.9, 47.1]$, $\sigma \in [2.7, 7.3]$ per $N = 10$ e $\sigma^2 \in [18.1, 31.9]$, $\sigma \in [4.3, 5.7]$ per $N = 100$. Come si vede, per grandi campioni si possono usare anche le (6.62). Potete anche risolvere il problema con la routine `Stimvs`.

Esercizio 6.9

L'analisi di un campione di 1000 resistenze elettriche del valore 1000Ω prodotte da una ditta ha mostrato che i valori si distribuiscono approssimativamente secondo una gaussiana di deviazione standard $s = 10 \Omega$ (in effetti i processi di produzione delle resistenze a impasto seguono abbastanza bene le condizioni del teorema Limite Centrale 3.1). Si decide di mantenere costante questo standard di produzione effettuando un controllo di qualità misurando periodicamente con un multimetro di grande accuratezza un campione di 5 resistenze. Definire i limiti statistici del controllo di qualità ad un livello di confidenza del 95%.

Risposta. Dobbiamo assumere il valore nominale delle resistenze come valore medio vero della produzione: $\mu = 1000 \Omega$.

La dispersione vera dei valori delle resistenze intorno alla media si può stimare dalla misura fatta sul campione di 1000 resistenze, applicando con $s \simeq \sigma$ la (6.62):

$$\sigma \in s \pm \frac{s}{\sqrt{2(N-1)}} = 10.0 \pm 0.2 ,$$

da cui si vede che il campione di 1000 resistenze consente di stimare la dispersione con una incertezza relativa del 2 %. Possiamo pertanto assumere $s = 10 \Omega$ come valore vero della deviazione standard.

Dato che si è osservato un campione di tipo gaussiano, dalla Tab. D.1 possiamo dire che entro $1000 \pm 1.96 \sigma \simeq 1000 \pm 20 \Omega$ cadrà il 95% dei valori. In sostanza solo 5 resistenze su cento devono stare al di fuori di

$$980 \Omega \leq R \leq 1020 \Omega . \quad (6.78)$$

Il problema è ora istituire dei controlli sulle resistenze prodotte per verificare che le condizioni iniziali si mantengano ragionevolmente costanti.

Campionando 5 resistenze, possiamo istituire un buon controllo di qualità attraverso la media del campione. Infatti, se vengono assunti come valori veri

$$\mu = 1000 \Omega , \quad \sigma \simeq s = 10 \Omega ,$$

in base alla (6.46), la media campionaria di 5 elementi sarà compresa nell'intervallo

$$\mu \pm \frac{\sigma}{\sqrt{5}} = 1000.0 \pm 4.5 \Omega \quad (6.79)$$

con livelli di confidenza gaussiani. In questo caso la distribuzione di Student non va usata, perché il campione è gaussiano ed abbiamo supposto di conoscere la deviazione standard vera. Se questa non fosse nota dalla misura delle 1000 resistenze, essa andrebbe dedotta dal campione ridotto delle 5 resistenze campionate; varrebbe ancora la (6.79), ma in questo caso avremmo al posto di σ la deviazione standard s delle 5 resistenze, e i livelli di confidenza seguirebbero la distribuzione di Student con 4 gradi di libertà, dato che $N < 10$.

Per un primo controllo di qualità si può allora utilizzare la (6.79), con un livello di confidenza del 95%, che dalla Tab. D.1 risulta associato ad un intervallo di circa $\pm 1.96 \sigma$. In questo caso la probabilità di sbagliare giudicando la produzione scadente è del 5%.

Un primo controllo di qualità segnalerà allora una possibile cattiva produzione quando la media del campione è al di fuori dell'intervallo

$$1000 \pm 1.96 \cdot \frac{10}{\sqrt{5}} \simeq (1000 \pm 9) \Omega ,$$

ovvero

$$991 \Omega \leq m(R) \leq 1009 \Omega \quad (\text{I controllo qualità, } CL = 95 \%) . \quad (6.80)$$

Il controllo globale può essere ulteriormente perfezionato verificando anche che la deviazione standard del campione non superi il valore iniziale di 10Ω . Infatti, invertendo la (6.73) e prendendone l'estremo superiore abbiamo:

$$s \leq \sqrt{\chi_{R\alpha}^2} \sigma, \quad (6.81)$$

dove $\sigma = 10 \Omega$ e $\chi_{R\alpha}^2$ è il valore di $Q_R(4)$, la variabile χ^2 ridotto con 4 gradi di libertà corrispondente al livello di confidenza prestabilito. Per un $CL = 95\%$, dalla Tab. D.3 otteniamo:

$$\chi_{R0.95}^2 = 2.37,$$

e quindi:

$$s \leq 10 \cdot \sqrt{2.37} \simeq 15 \Omega.$$

Il secondo controllo di qualità segnalerà allora una possibile cattiva produzione quando la dispersione del campione di 5 resistenze supera il limite di 15Ω :

$$s(R) \leq 15 \Omega \quad (\text{II controllo qualità, } CL = 95\%).$$

Se si richiede che i controlli di qualità I e II siano contemporaneamente soddisfatti, si assicura il mantenimento corretto sia del valore nominale sia della dispersione iniziale delle resistenze elettriche, entro il livello di confidenza arbitrariamente scelto.

Avendo richiesto un $CL = 0.95$, avremo un superamento di questi limiti mediamente 5 volte su cento. Un segnale fuori dalla fascia di confidenza, ma entro i limiti della statistica prevista, si chiama falso allarme. La situazione viene spesso sintetizzata graficamente nella carta di controllo della qualità, in cui si sceglie una zona di normalità, al di sopra o al di sotto della quale vi sono due fasce di allarme ed al di fuori di queste la zona proibita.

La Fig. 6.10 mostra una possibile carta di controllo per il valor medio delle resistenze, relativa al problema che stiamo discutendo. La zona di normalità corrisponde all'intervallo (6.80) di ampiezza $\pm 1.96 \sigma$; la zona di allarme va da 1.96 a 3.0σ , mentre la zona proibita è al di fuori di $\pm 3.0 \sigma$. Un valore nella

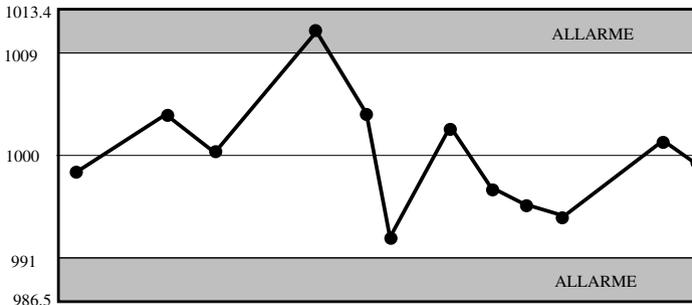


Fig. 6.10. Carta di controllo per il caso della produzione di resistenze elettriche discusso nell'esercizio 6.9

zona proibita può capitare in condizioni normali solo 3 volte su 1000 controlli (legge 3σ), un evento che può giustificare la sospensione della produzione e l'attivazione dei processi di manutenzione delle macchine (attenzione: questa è una decisione soggettiva che può variare da caso a caso).

Come esercizio, utilizzando la (6.81) potete fare una carta simile (carta S) anche per la dispersione. Nella zona di allarme mediamente dovremo avere 5 valori ogni cento controlli, vale a dire una probabilità a priori $\gamma = 0.05$. Il controllo di qualità può allora essere ulteriormente raffinato segnalando se capita un numero eccessivo di falsi allarmi, anche in assenza di segnali nella zona proibita. Se n è il numero di falsi allarmi avuti in N controlli si può applicare la (6.20), poiché qui assumiamo di conoscere la probabilità vera γ :

$$t_n = \frac{n - \gamma N}{\sqrt{N\gamma(1 - \gamma)}} = 4.6 \frac{n - 0.05N}{\sqrt{N}}, \quad (6.82)$$

ed arrestare la produzione, applicando ancora la legge 3σ , quando $t_n > 3.0$. Per una variabile gaussiana, questo valore corrisponde a una probabilità di sbagliare di $\simeq 1.5$ per mille. L'approssimazione gaussiana è lecita per $n > 10$, cui corrisponde $N\gamma > 10$, cioè $N > 200$ nella (6.82). Per $N = 200$, risulta che bisogna procedere alla manutenzione se $n \geq 19$.

6.13 Stime da una popolazione finita

Nelle stime fin qui eseguite abbiamo supposto che la popolazione fosse costituita da un insieme infinito di elementi. I risultati ottenuti valgono anche per popolazioni finite, a patto che, dopo ogni prova, l'elemento estratto venga reinserito nella popolazione (campionamento con reimmissione). È intuitivo però che vi siano delle modifiche da fare nel caso di campionamento senza reimmissione da una popolazione finita, poiché, se si esaurisse la popolazione, le quantità in esame diventerebbero certe e non sarebbero più stime statistiche. Iniziamo con la:

Definizione 6.4 (Campione casuale da una popolazione finita). *Un campione S di N elementi estratto da una popolazione finita di N_p elementi è detto casuale se è uno dei $N_p!/[N!(N_p - N)!]$ possibili insiemi ognuno dei quali ha la stessa probabilità di essere scelto.*

Se X è la variabile aleatoria che si vuole stimare attraverso il campione S , possiamo scrivere:

$$M = \frac{1}{N} \sum_{i=1}^N X_i = \frac{1}{N} \sum_{i=1}^{N_p} x_i I_i, \quad (6.83)$$

dove la seconda somma è fatta su tutti gli elementi N_p della popolazione e I_i è una variabile indicatrice (vedere eq. 2.8), definita come:

$$I_i = \begin{cases} 1 & \text{se } x_i \in S, \\ 0 & \text{altrimenti.} \end{cases}$$

I_i è una variabile binomiale a due valori e corrisponde pertanto al numero di successi che si possono ottenere in un solo tentativo, di probabilità N/N_p . Si ha pertanto:

$$\langle I_i \rangle = \frac{N}{N_p}, \quad \text{Var}[I_i] = \frac{N}{N_p} \left(1 - \frac{N}{N_p} \right), \quad \forall i. \quad (6.84)$$

Otteniamo allora:

$$\langle M \rangle = \frac{1}{N} \sum_{i=1}^{N_p} x_i \langle I_i \rangle = \frac{1}{N_p} \sum_{i=1}^{N_p} x_i = \langle X \rangle, \quad (6.85)$$

da cui si vede che *anche in questo caso la media campionaria è uno stimatore corretto della media vera.*

Troviamo ora la varianza della somma delle variabili campionate; ricordando la (5.65) possiamo scrivere:

$$\text{Var} \left[\sum_{i=1}^{N_p} x_i I_i \right] = \sum_{i=1}^{N_p} x_i^2 \text{Var}[I_i] + 2 \sum_i \sum_{j < i} x_i x_j \text{Cov}[I_i, I_j]. \quad (6.86)$$

Il calcolo della covarianza di (I_i, I_j) richiede la conoscenza della media $\langle I_i I_j \rangle$. Ricorrendo alle definizioni generali di pagina 59 e scrivendo solo i valori non nulli, abbiamo:

$$\langle I_i I_j \rangle = P\{x_j \in S | x_i \in S\} P\{x_i \in S\} = \frac{N-1}{N_p-1} \frac{N}{N_p}.$$

Dalla (4.25) di pagina 123 possiamo esprimere la covarianza come:

$$\text{Cov}[I_i, I_j] = \langle I_i I_j \rangle - \langle I_i \rangle \langle I_j \rangle = \frac{N}{N_p} \frac{N-1}{N_p-1} - \frac{N^2}{N_p^2} = -\frac{N(N_p-N)}{N_p^2(N_p-1)}.$$

Inserendo questo risultato nella (6.86) abbiamo:

$$\text{Var} \left[\sum_{i=1}^{N_p} x_i I_i \right] = \frac{N}{N_p} \frac{N_p-N}{N_p} \sum_{i=1}^{N_p} x_i^2 - 2 \frac{N(N_p-N)}{N_p^2(N_p-1)} \sum_i \sum_{j < i} x_i x_j. \quad (6.87)$$

Ricordando che $(\sum_i x_i)^2 = \sum_i x_i^2 + 2 \sum_i \sum_{j < i} x_i x_j$, dopo un po' di facile algebra otteniamo infine:

$$\text{Var} \left[\sum_{i=1}^{N_p} x_i I_i \right] = \frac{N(N_p-N)}{N_p-1} \left(\langle X^2 \rangle - \langle X \rangle^2 \right), \quad (6.88)$$

dove $\langle X^2 \rangle = \sum X_i^2/N_p$ e $\langle X \rangle^2 = (\sum X_i)^2/N_p^2$. La varianza della media campionaria vale allora:

$$\begin{aligned} \text{Var} \left[\frac{1}{N} \sum_{i=1}^{N_p} x_i I_i \right] &= \text{Var}[M] = \frac{N_p - N}{N(N_p - 1)} \left(\langle X^2 \rangle - \langle X \rangle^2 \right) \\ &= \frac{\text{Var}[X]}{N} \frac{N_p - N}{N_p - 1} \simeq \frac{\text{Var}[X]}{N} \left(1 - \frac{N}{N_p} \right). \end{aligned} \quad (6.89)$$

Questa equazione è il risultato fondamentale per la stima da popolazioni finite: dal confronto con le formule analoghe per popolazione infinita di Tab. 6.3, si vede che le varianze di medie, frequenze e proporzioni, calcolate da campioni estratti da popolazioni finite, vanno corrette col fattore $(N_p - N)/(N_p - 1) \simeq (1 - N/N_p)$. Lo stesso fattore va applicato se nella (6.89) alla varianza vera σ^2 si sostituisce quella stimata s^2 .

Ad esempio, nel caso dell'esercizio 6.1, considerando 30 milioni di elettori, la correzione sarebbe dell'ordine di $\sqrt{1 - 3/30\,000}$. Diverso è il caso dell'estrazione senza reimmissione da un'urna: se la frequenza delle biglie di un certo tipo fosse ad esempio $f = 15/30 = 0.50$ e l'urna contenesse 100 biglie, l'errore della frequenza passerebbe da $\sqrt{0.5(1 - 0.5)/30} = 0.09$ (popolazione infinita) al valore $\sqrt{0.5(1 - 0.5)/30} \sqrt{70/99} = 0.08$. La varianza si annullerebbe se si estraessero tutte le 100 biglie.

La varianza campionaria, a differenza della media, deve essere corretta per popolazioni finite. Infatti, dalle (6.53, 6.89) si ha:

$$\langle S^2 \rangle = \left\langle \frac{1}{N-1} \sum_i (X_i - M)^2 \right\rangle = \frac{1}{N-1} \left[N\sigma^2 - \frac{N_p - N}{N_p - 1} \sigma^2 \right] = \sigma^2 \frac{N_p}{N_p - 1}. \quad (6.90)$$

Risulta allora che lo stimatore non distorto per la varianza di una popolazione finita vale:

$$S^2 = \frac{N_p - 1}{N_p} \frac{1}{N-1} \sum_i (X_i - M)^2. \quad (6.91)$$

La correzione per la varianza $\text{Var}[S^2]$ è più complicata e può essere trovata in [KS73].

6.14 Verifica di una ipotesi

Abbiamo già accennato alla verifica delle ipotesi negli esercizi 3.13-3.17. Approfondiamo ora l'argomento considerando la verifica (accettazione o rifiuto) di una ipotesi, detta *ipotesi nulla* o, più semplicemente, ipotesi. L'argomento verrà poi completato più avanti, nel cap. 9, dove discuteremo i criteri per ottimizzare le scelte tra più ipotesi in alternativa.

Consideriamo dunque la densità di probabilità $p(x; \theta)$ della variabile X definita nella (6.2), dipendente da un parametro θ e supponiamo che, in base

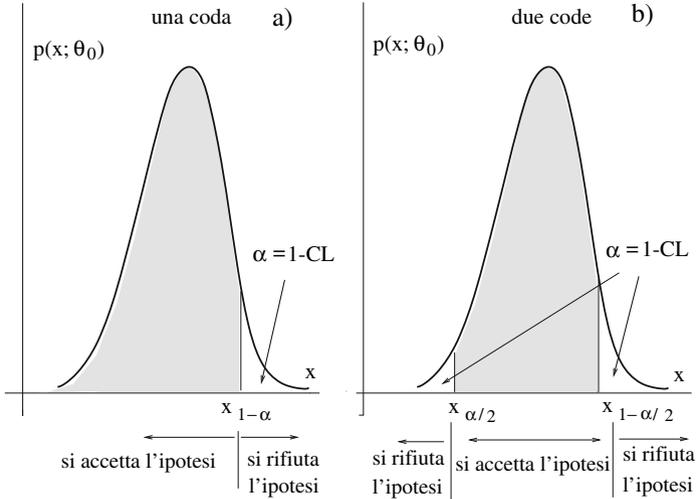


Fig. 6.11. Test a una coda a) o a due code b). L'uso dell'uno o dell'altro dipende dal tipo di problema, come evidenziato dagli esempi discussi nel testo. Il livello di confidenza CL è il valore dell'area ombreggiata definita dai valori quantili $x_{1-\alpha}$, $x_{\alpha/2}$ e $x_{1-\alpha/2}$. Il livello del test $\alpha = 1 - CL$ è il valore dell'area delle code al di fuori dell'intervallo di confidenza

ad argomenti a priori o a stime statistiche, venga assunta l'ipotesi $\theta = \theta_0$. Se l'esperimento fornisce un evento $\{X = x\}$, occorre decidere se accettare o meno il modello corrispondente all'ipotesi, in base al risultato ottenuto. Lo schema che si utilizza è in genere quello di Fig. 6.11: come si vede, l'intervallo di variabilità di X viene suddiviso in due regioni, una regione favorevole all'ipotesi ed una regione contraria, detta *regione critica*. I limiti di queste regioni sono date dai valori quantili $x_{\alpha/2}$, $x_{1-\alpha/2}$, corrispondenti ad un livello di significatività α stabilito a priori, detto anche *livello del test*. Si rifiuta l'ipotesi se il livello di significatività osservato $\alpha_0 < \alpha$, la si accetta in caso contrario.

Avendo osservato l'evento $\{X = x\}$, nell'ipotesi corrispondente a $p(x; \theta_0)$, di media μ , occorre quindi calcolare le probabilità:

$$P\{X < x\} = \alpha_0, \text{ test a una coda (a sinistra),} \tag{6.92}$$

$$P\{X > x\} = \alpha_0, \text{ test a una coda (a destra),} \tag{6.93}$$

$$P\{|X - \mu| > |x - \mu|\} = \alpha_0, \text{ test a due code.} \tag{6.94}$$

Se si scarta l'ipotesi, la probabilità di sbagliare, cioè di ottenere per puro caso valori "nelle code", per i quali le (6.92-6.94) danno $\alpha_0 < \alpha$ essendo vera l'ipotesi, non supera mai α . Questo tipo di errore è detto di I tipo (scartare una ipotesi giusta), per distinguerlo dall'altro possibile errore (accettare una ipotesi sbagliata), detto di II tipo, che verrà trattato più avanti. Il valore α viene anche detto dimensione dell'errore di I tipo.

Il livello di significatività osservato (*significance level*) definito nelle (6.92-6.94), che indicheremo spesso con SL , nella letteratura inglese viene detto anche *p-value*.

La verifica di una ipotesi può essere fatta anche con statistiche $T = t(X_1, X_2, \dots, X_N)$ che stimano il valore del parametro θ , cioè con degli stimatori. In questo caso alla densità $p(x; \theta_0)$ attribuita al campione si deve sostituire la densità della *distribuzione campionaria* dello stimatore e tutto procede come appena spiegato. Se lo stimatore non è distorto, in base alla (2.78), $\langle T_N(X) \rangle = \theta_0$. Stabilito allora un valore α per il livello del test, si rigetta l'ipotesi se, ottenuta una stima $\{T = t_0\}$, risulta:

$$P\{T < t_0\} = SL < \alpha, \text{ test a una coda (a sinistra)}, \quad (6.95)$$

$$P\{T > t_0\} = SL < \alpha, \text{ test a una coda (a destra)}, \quad (6.96)$$

$$P\{|T - \theta_0| > |t_0 - \theta_0|\} = SL < \alpha, \text{ test a due code.} \quad (6.97)$$

Il livello del test α viene scelto in modo arbitrario. In genere, come abbiamo indicato in precedenza, si scelgono valori piccoli, circa l'1-5%. Se l'evento ha un SL minore di α , esso si trova nelle code della densità di riferimento. In questo caso si giudica più probabile la falsità dell'ipotesi $\theta = \theta_0$ che non la sua validità. Il caso contrario viene interpretato come una *fluttuazione statistica*, lo scostamento viene attribuito al caso e l'ipotesi viene accettata.

Per finire, notiamo che in alcuni casi il confronto tra livello del test ed SL presenta ambiguità:

- Per una variabile aleatoria discreta, i livelli SL di due valori contigui possono “stare a cavallo” del valore deciso per il test. Ad esempio, nel test a una coda a destra, può succedere che:

$$P\{X > x_1\} = SL_1 < \alpha < SL_2 = P\{X > x_2\},$$

con $x_1 > x_2$. In questo caso si accetta l'ipotesi se $\{X = x_2\}$, la si respinge se $\{X = x_1\}$. In alternativa, se si dispone di un generatore di variabili uniformi $0 \leq U \leq 1$ (ad esempio, la routine `random` di un calcolatore), si può randomizzare il test definendo una probabilità p tale che:

$$SL_1 + p(SL_2 - SL_1) = \alpha \implies p = \frac{\alpha - SL_1}{SL_2 - SL_1}. \quad (6.98)$$

Se si verifica l'evento $\{X = x_1\}$, si *scarta* l'ipotesi se dalla routine `random` si ottiene $\{U = u \leq p\}$, la si *accetta* in caso contrario. In questo modo, per costruzione, il livello SL dell'evento coincide esattamente col livello del test α .

- Nel caso del test a due code, ad un livello α possono corrispondere limiti fiduciarî diversi, cioè intervalli di estremi diversi ma con lo stesso livello di confidenza $CL = 1 - \alpha$. Generalmente in questo caso si scelgono i due estremi, a destra e sinistra, che sottendono “code” aventi la stessa area $\alpha/2$. Questa convenzione è stata in effetti utilizzata nella (6.94).

6.15 Verifica di compatibilità tra due valori

Affrontiamo ora, per la prima volta, un semplice problema di verifica di ipotesi, la verifica di compatibilità tra due valori.

Supponiamo che siano stati eseguiti due esperimenti o campionamenti differenti e che si voglia verificare se i risultati x_1 e x_2 sono in accordo tra loro, ovvero se i dati provengono dalla stessa popolazione.

Se alla popolazione si attribuisce una densità nota con parametri fissati a priori, allora si può affrontare il problema col calcolo delle probabilità, verificando la compatibilità di ogni singolo dato con in valori assunti come veri. Esempi di questa tecnica sono stati dati negli esercizi 3.13-3.17. Tuttavia, spesso i parametri della la densità non sono noti ed è necessario procedere in ambito esclusivamente statistico.

Un modo elegante per affrontare il problema sta nel definire una nuova variabile aleatoria differenza

$$D = X_1 - X_2 ,$$

la quale *ha valore atteso nullo* nell'ipotesi che i due valori X_1 e X_2 provengano dalla stessa popolazione.

Poiché la varianza di una differenza è data, per variabili indipendenti, dalle (5.66, 5.67), possiamo scrivere:

$$\text{Var}[D] = \text{Var}[X_1] + \text{Var}[X_2] \simeq s_1^2 + s_2^2 ,$$

dove nell'ultimo passaggio si è utilizzata l'approssimazione per grandi campioni (si veda la Tab. 6.3) sostituendo ai valori veri incogniti quelli delle varianze misurate. Possiamo allora definire il valore standard:

$$|t_D| = \frac{|x_1 - x_2| - 0}{\sqrt{s_1^2 + s_2^2}} = \frac{|x_1 - x_2|}{\sqrt{s_1^2 + s_2^2}} \quad (6.99)$$

e calcolarne il livello di significatività secondo le densità di probabilità coinvolte nelle misure. Poiché la (6.99) presuppone l'approssimazione per grandi campioni, per il calcolo dei livelli di probabilità si usa quasi sempre la densità gaussiana, perché (come mostrato anche nell'esercizio 5.3) la differenza di due variabili gaussiane indipendenti è anch'essa gaussiana.

Generalmente la (6.99) viene usata per confrontare due frequenze oppure due medie campionarie.

Nel caso di due frequenze f_1 ed f_2 , indicando con N_1 ed N_2 (invece che con n) il numero di tentativi effettuati nelle due prove, dalla (6.23), valida per grandi campioni, si ottiene immediatamente:

$$t_f = |f_1 - f_2| \left[\frac{f_1(1 - f_1)}{N_1} + \frac{f_2(1 - f_2)}{N_2} \right]^{-1/2} , \quad (6.100)$$

con livelli di probabilità gaussiani.

Invece che in termini di probabilità, se $N_1 = N_2 = N$, il confronto può anche essere fatto in termini di numero di successi x_1 e x_2 , utilizzando sempre la (6.99) e la varianza definita nella (6.24):

$$t_x = |x_1 - x_2| \left[x_1 \left(1 - \frac{x_1}{N} \right) + x_2 \left(1 - \frac{x_2}{N} \right) \right]^{-1/2}. \quad (6.101)$$

Se $N_1 \neq N_2$, occorre rinormalizzare uno dei due valori (ad esempio $x_1 \rightarrow x_1 N_1 / N_2$) ricalcolandone la varianza, ma in questo caso è più conveniente il confronto in termini di frequenza, attraverso la (6.100).

Per il confronto di due medie m_1 ed m_2 , provenienti da due campioni diversi con varianze s_1^2 , s_2^2 e numero di elementi N_1 , N_2 rispettivamente, dalla (6.49) si ottiene il valore standard:

$$t_m = |m_1 - m_2| \left[\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \right]^{-1/2}, \quad (6.102)$$

al quale, se $N_1, N_2 > 20, 30$, vanno ovviamente applicati livelli di probabilità gaussiani.

Esercizio 6.10

Due gruppi di sperimentatori hanno fornito, nella misura di una stessa grandezza fisica, i valori:

$$x_1 = 12.3 \pm 0.5$$

$$x_2 = 13.5 \pm 0.8$$

ed affermano che entrambe le misure sono di tipo gaussiano. Verificare se le misure sono tra loro compatibili.

Risposta. Dalla (6.99) otteniamo:

$$t = \frac{|12.3 - 13.5|}{\sqrt{(0.5)^2 + (0.8)^2}} = \frac{1.20}{0.943} = 1.27.$$

Poiché si deve trovare lo scostamento assoluto tra le due misure, è necessario fare il test a due code. Dalla Tab. D.1 risulta:

$$P\{|T| \geq 1.27\} = 1 - 2 \cdot 0.3980 \simeq 0.20,$$

pari ad un livello di significatività del 20%. Poiché la probabilità di sbagliare rigettando l'ipotesi di compatibilità è troppo alta, le due misure si devono ritenere compatibili.

Notate bene che non si afferma che le misure sono compatibili, ma solo che esse non possono essere messe in dubbio sulla sola base della statistica. Come dovrete aver già capito da parecchio tempo, la distinzione può sembrare sottile, ma è fondamentale!

Esercizio 6.11

In un esperimento di percezione extrasensoriale (ESP) sono state predisposte 5 scatole, numerate da 1 a 5 e, nella scatola n. 3, è stato posto un oggetto-bersaglio. Duecento persone sono state invitate ad indovinare in quale scatola si trovasse il bersaglio e 62 di queste hanno indicato proprio la n. 3. In una prova di controllo con tutte le scatole vuote, facendo sempre credere alla platea che in una delle scatole si trovasse un bersaglio, le persone che hanno indicato la scatola n. 3 sono state 50. Stabilire se l'esperimento rivela o meno effetti ESP.

Risposta. Se utilizziamo il calcolo delle probabilità, possiamo procedere come nell'esercizio 3.17 ed assumere come ipotesi nulla che, se le persone tirano a indovinare, ognuna delle 5 scatole abbia una eguale probabilità di $1/5$ di essere scelta. Allora la distribuzione teorica assunta come vera è binomiale, di media e deviazione standard date da:

$$\mu = 40, \quad \sigma = \sqrt{40(1 - 40/200)} = 5.65.$$

Poiché i successi sono stati pari a 62, la variabile standard (3.37) di questo dato vale:

$$t = \frac{x - \mu}{\sigma} = \frac{62 - 40}{5.65} = 3.9$$

a cui corrisponde, come da Tab. D.1, un livello di significatività estremamente basso, $< 1 \cdot 10^{-4}$. L'ipotesi di presenza di effetti ESP sembra confermata dall'esperimento.

Tuttavia, la psicologia sperimentale afferma che il numero 3 è psicologicamente favorito (in genere lo sono tutti i numeri dispari rispetto a quelli pari). In altre parole, in una prova a vuoto con scatole numerate da 1 a 5, la maggior parte delle scelte dovrebbe localizzarsi sulla scatola n. 3, anche in assenza di bersaglio, semplicemente perché il 3 è un numero "più simpatico" degli altri. Occorre allora, in mancanza di un modello a priori, abbandonare il calcolo delle probabilità e affrontare il problema in ambito statistico, utilizzando solamente il numero di successi con bersaglio ($x_1 = 62$) e quello a scatola vuota ($x_2 = 50$). Gli errori statistici da associare a queste osservazioni si ottengono dalla (6.24):

$$s(50) = \sqrt{50 \left(1 - \frac{50}{200}\right)} = 6.12,$$

$$s(62) = \sqrt{62 \left(1 - \frac{62}{200}\right)} = 6.54.$$

Combinando adesso i risultati della prova a vuoto e di quella col bersaglio nella formula (6.101) abbiamo immediatamente:

$$|t| = \frac{|62 - 50|}{\sqrt{(6.12)^2 + (6.54)^2}} = 1.34.$$

A questo valore della variabile gaussiana standard corrisponde in Tab. D.1 un livello di significatività sulle due code di

$$P\{|T| > 1.34\} = 2 \cdot (0.5 - 0.4099) = 0.1802 \simeq 18\% .$$

L'analisi mostra pertanto che nel 18% dei casi, tirando ad indovinare ed essendo psicologicamente condizionati a favore del numero 3, si possono avere scostamenti di più di 12 punti (in più o in meno) tra la prova a vuoto e la prova col bersaglio. Il risultato è pertanto compatibile con il caso e non rivela affatto probabili effetti ESP.

Se i due campioni sono piccoli e la distribuzione campionaria è gaussiana, si può ancora fare uso della densità di Student, se si suppone che le due medie provengano da campioni aventi, oltre che la stessa media, anche la stessa dispersione: $\sigma_1 = \sigma_2 \equiv \sigma$. In questo caso la variabile standard della differenza assume il valore:

$$t_D = \frac{m_1 - m_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} = \frac{m_1 - m_2}{\sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} . \quad (6.103)$$

Utilizzando la (6.68), nella quale la varianza dei dati è calcolata rispetto alla media campionaria, definiamo il valore di χ^2 (non ridotto):

$$\chi^2 = \frac{(N_1 - 1)s_1^2}{\sigma_1^2} + \frac{(N_2 - 1)s_2^2}{\sigma_2^2} = \frac{1}{\sigma^2} [(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2] , \quad (6.104)$$

al quale, in base al teorema 3.4, vanno assegnati $N_1 + N_2 - 2$ gradi di libertà. In base ai risultati dell'esercizio 5.5, possiamo dire che il valore (5.41), che in questo caso diventa:

$$\begin{aligned} t_S &= \frac{m_1 - m_2}{\sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \frac{\sigma \sqrt{N_1 + N_2 - 2}}{\sqrt{[(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2]}} \\ &= \frac{m_1 - m_2}{s_{12} \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} , \quad s_{12} = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}} , \quad (6.105) \end{aligned}$$

si distribuisce secondo la densità di Student con $N_1 + N_2 - 2$ gradi di libertà. In questo modo il test sulla differenze delle medie può essere risolto in modo soddisfacente anche per campioni con almeno una decina di eventi, per i quali la media si può considerare una variabile gaussiana.

SCILAB fornisce la routine `nand2med(x1,x2,conf)` per il test tra le medie di due campioni `x1` e `x2`, in forma di vettori contenenti i dati grezzi. In uscita si ottengono due numeri: la differenza tra le medie e il valore

$s = t_{\text{conf}} s_{12} \sqrt{(1/N_1) + (1/N_2)}$ della (6.105), dove t è il quantile di Student per campioni con meno di 100 eventi, altrimenti è quello gaussiano. Per grandi campioni, l'errore vale $s \simeq t_{\text{conf}} \sqrt{(s_1^2/N_1) + (s_2^2/N_2)}$, in accordo con la (6.102). Si tenga presente che conf è il valore cumulativo ed è legato al livello CL del test a due code dalla relazione $\text{conf} = 0.5 + CL/2$. Se si vuole ottenere la stima ad 1σ dell'errore s sulla differenza delle medie si deve porre $\text{conf} = 0.5 + 0.683/2 = 0.8415$.

Occupiamoci infine della verifica di compatibilità tra due varianze. Essa, per campioni con meno di un centinaio di eventi, non può essere fatta utilizzando la densità gaussiana o di Student. Tuttavia, per campioni gaussiani esiste un test molto semplice ed utile, che fa uso della densità F di Snedecor trovata nell'esercizio 5.6.

Se s_N^2 ed s_M^2 sono due varianze ottenute da due campioni indipendenti rispettivamente con N ed M eventi e provenienti da popolazioni di densità gaussiane aventi la stessa varianza σ^2 , dalla (6.68) sappiamo che il valore

$$F_{N-1, M-1} = \frac{s_N^2/\sigma}{s_M^2/\sigma} = \frac{s_N^2}{s_M^2}, \quad (6.106)$$

essendo il rapporto tra due valori χ^2 ridotto indipendenti, si distribuirà secondo la densità F di Snedecor (5.46) con $(N-1, M-1)$ gradi di libertà. L'uso combinato della (6.106) e dei quantili della densità F delle tabelle D.5, D.6 viene detto analisi della varianza o F-test. In ambito SCILAB potete usare la nostra routine `Test2v`, che utilizza `cdf(F,P,Q,N,M)` per ottenere dalla distribuzione $p_{NM}(F)$ il valore cumulativo di F e `cdf(F,N,M,P,1-P)` per il calcolo di F a partire dal valore cumulativo P .

Esercizio 6.12

Due sperimentatori, che affermano di aver eseguito un campionamento dalla stessa popolazione avente densità gaussiana, hanno trovato rispettivamente varianze pari a

$$s_1^2 = 12.5, \quad s_2^2 = 6.4$$

con campioni di 20 e 40 eventi rispettivamente. Verificare se le due varianze campionarie sono compatibili con un unico valore vero σ^2 ad un livello del 2%.

Risposta. La variabile F è data da:

$$F = \frac{12.5}{6.4} = 1.95.$$

Se scegliamo di fare la verifica di compatibilità ad un livello di significatività del 2%, il rapporto sperimentale F deve essere minore del valore di F al 99% che si ricava da Tab. D.6:

$$F_{0.99}(19, 39) \simeq 2.4,$$

e maggiore del valore di F all'1 %, che si ricava con la relazione (5.49):

$$F_{0.01}(19, 39) = \frac{1}{F_{0.99}(39, 19)} \simeq \frac{1}{2.7} = 0.37 .$$

Poiché

$$0.37 < 1.95 < 2.4 ,$$

i due valori sono compatibili ad un livello del 2%.

Potete ripetere l'esercizio con **Test2v** variando il livello del test.

6.16 Stima della densità di una popolazione

Dopo l'analisi della media e della dispersione del campione, passiamo ora all'analisi della struttura globale (forma) del campione, con l'obiettivo di ottenere informazioni sulla densità della popolazione da cui è stato estratto.

La media e la varianza campionarie non sono le sole variabili aleatorie di interesse. Quasi sempre il campione viene presentato in forma di istogramma, suddiviso in un numero K di canali, di solito di ampiezza fissa Δx , in ognuno dei quali cade un numero di eventi n_i . Le quantità n_i danno la forma complessiva del campione e sono di cruciale importanza se si è interessati a studiare la struttura della densità della popolazione da cui il campione è stato estratto. *Queste quantità vanno considerate come variabili aleatorie I_i , perché variano da campionamento a campionamento.*

L'istogramma spesso viene normalizzato, nel qual caso, invece delle quantità n_i , vengono date le frequenze o probabilità misurate $f_i = n_i/N$, dove N è il numero totale di eventi presenti nel campione.

In Fig. 6.12 e nella (6.107) viene riportato, a titolo di esempio, un istogramma non normalizzato ottenuto simulando al calcolatore un campione di $N = 1000$ eventi estratto da una densità gaussiana con parametri $\mu = 70$, $\sigma = 10$. Nella (6.107) x_i e $n_i \equiv n(x_i)$ sono rispettivamente il punto medio e il contenuto del canale, il passo vale $\Delta x = 5$:

x	$n(x)$	x	$n(x)$
37.5	1	72.5	207
42.5	4	77.5	153
47.5	16	82.5	101
52.5	44	87.5	42
57.5	81	92.5	7
62.5	152	97.5	6
67.5	186		

(6.107)

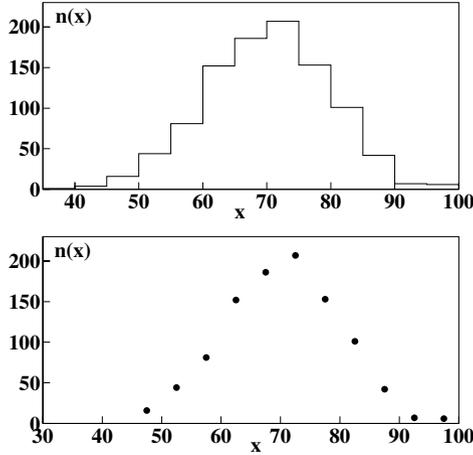


Fig. 6.12. Due rappresentazioni possibili di un istogramma ottenuto simulando al calcolatore l'estrazione di 1000 eventi da un popolazione di densità gaussiana di media $\mu = 70$ e deviazione standard $\sigma = 10$

Se $p(x)$ è la densità della popolazione del campione, definendo le variabili aleatorie I ed F associate agli eventi $\{I_i = n_i\}$ ed $\{F_i = n_i/N\}$, dalla (2.34) segue che:

$$\langle I_i \rangle = \mu_i = Np_i = N \int_{\Delta x} p(x) dx \simeq Np(x_0) \Delta x , \quad (6.108)$$

$$\langle F_i \rangle = p_i = \int_{\Delta x} p(x) dx \simeq p(x_0) \Delta x , \quad (6.109)$$

dove Δx è la larghezza del canale e x_0 un punto interno ad esso. Il termine più a destra nelle equazioni viene dal teorema della media integrale. Se l'ampiezza del canale è sufficientemente piccola e la densità è una funzione abbastanza regolare (funzione *smooth*), si può assumere che essa vari linearmente entro il canale; in questo caso, in base alla regola trapezoidale, x_0 è il punto medio del canale.

Nel caso di variabili aleatorie discrete, nelle (6.108, 6.109) all'integrale su Δx va sostituita la sommatoria sulle probabilità vere corrispondenti ai valori contenuti in Δx .

Nel par. 4.7 abbiamo visto, per un campione casuale di ampiezza N , che la probabilità globale di ottenere un particolare istogramma a partire da probabilità vere p_i , ($i = 1, 2, \dots, k$) date, tramite la (6.109), da una densità $p(x)$, segue la distribuzione multinomiale (4.87).

Abbiamo già notato, a proposito della (4.87), che il numero di eventi I_i caduti in un generico canale i -esimo (x_i, x_{i+1}) di ampiezza Δx segue la distribuzione binomiale. Infatti:

- se il fenomeno stocastico è stazionario nel tempo, la probabilità p_i di cadere nel canale resta costante;
- la probabilità di cadere nel canale non dipende dagli eventi che sono caduti o che cadranno negli altri canali.

Pertanto, possiamo dire che la variabile aleatoria I_i , *numero di eventi in un canale dell'istogramma*, è data dal verificarsi di eventi indipendenti determinati da una probabilità costante. Se il numero totale di eventi N è un parametro predeterminato, risulta allora che la probabilità che la variabile aleatoria I_i assuma il valore n_i sarà data dalla legge binomiale (2.30) con probabilità elementare p_i (si veda anche la (4.87)):

$$P\{I_i = n_i\} = b(n_i; N, p_i) = \frac{N!}{n_i!(N - n_i)!} p_i^{n_i} (1 - p_i)^{N - n_i}, \quad (6.110)$$

dove $(1 - p_i)$ indica la probabilità di cadere in un qualunque altro canale dell'istogramma diverso da quello considerato. La deviazione standard della binomiale considerata vale:

$$\sigma_i = \sqrt{N p_i (1 - p_i)}. \quad (6.111)$$

Questa quantità può essere stimata dai dati attraverso l'incertezza $s_i \equiv s(n_i)$. Se il canale ha più di 5, 10 eventi, poiché a maggior ragione $N(1 - p_i) = (N - n_i) > 5, 10$, si può utilizzare la (6.24), con $x \equiv n_i$ ed $n \equiv N$:

$$s_i = \sqrt{n_i \left(1 - \frac{n_i}{N}\right)}. \quad (6.112)$$

Se l'istogramma è normalizzato, la (6.112) va divisa per N , ottenendo così un risultato ben noto, l'errore statistico (6.23) sulla probabilità misurata $f_i = n_i/N$:

$$s\left(\frac{n_i}{N}\right) = \sqrt{\frac{n_i}{N^2} \left(1 - \frac{n_i}{N}\right)} = \sqrt{\frac{f_i(1 - f_i)}{N}}. \quad (6.113)$$

Le due formule appena mostrate sono di fondamentale importanza nell'analisi degli istogrammi e vengono dette *fluttuazioni casuali (o statistiche) dei canali*.

Se l'istogramma non è ottenuto con numero totale N di eventi costante, ma è raccolto controllando altri parametri, ad esempio un certo intervallo di tempo Δt , il numero N si trasforma da costante a variabile statistica \mathbf{N} poissoniana e le fluttuazioni del contenuto dei canali vanno calcolate in modo diverso.

Questo è un argomento statistico abbastanza sottile e vogliamo fare subito un esempio per chiarirvi le idee. Se esaminiamo più istogrammi, ognuno dei quali si riferisce al peso di 100 neonati, le fluttuazioni del numero dei neonati entro un certo intervallo di peso (percentile, come dicono i medici) obbediranno alle (6.112, 6.113). Se invece raccogliamo gli istogrammi del peso dei neonati mese per mese, alle fluttuazioni del numero dei neonati entro un certo

intervallo di peso si aggiungeranno quelle del numero totale \mathbf{N} dei neonati in un mese, che saranno poissoniane nell'intorno di un valore medio stabile (se si suppone, per semplificare, che le nascite siano stabili di mese in mese).

Procedendo nella discussione, osserviamo che, in base alla legge delle probabilità composte, in questo caso la probabilità di osservare N eventi, dei quali $\{I_i = n_i\}$ nel canale i -esimo, sarà data dal prodotto della probabilità poissoniana (3.14) di ottenere un totale di $\{\mathbf{N} = N\}$ eventi quando la media è λ e dalla probabilità binomiale (2.30) di avere n_i eventi nel canale in esame, su un totale di N , quando la probabilità vera è p_i :

$$\begin{aligned} P\{I_i = n_i, \mathbf{N} = N\} &= P\{I_i = n_i | \mathbf{N} = N\} P\{\mathbf{N} = N\} \\ &= p(n_i, N) = \frac{N!}{n_i!(N - n_i)!} p_i^{n_i} (1 - p_i)^{N - n_i} \frac{e^{-\lambda} \lambda^N}{N!}. \end{aligned}$$

Se ora definiamo $m_i = N - n_i$, ed utilizziamo le identità:

$$e^{-\lambda} = e^{-\lambda p_i} e^{-\lambda(1-p_i)}, \quad \lambda^N = \lambda^{N-n_i} \lambda^{n_i} = \lambda^{m_i} \lambda^{n_i},$$

possiamo esprimere questa probabilità nella forma:

$$p(n_i, m_i) = \frac{e^{-\lambda p_i} (\lambda p_i)^{n_i}}{n_i!} \frac{e^{-\lambda(1-p_i)} [\lambda(1-p_i)]^{m_i}}{m_i!},$$

la quale non è altro che il prodotto di due poissoniane, di media λp_i e $\lambda(1-p_i)$ rispettivamente. Da questo fatto, in base al teorema 4.1 di pagina 120, si deduce che le variabili I_i , numero di eventi nel canale i -esimo, ed $(\mathbf{N} - I_i)$, numero di eventi caduti negli altri canali dell'istogramma, *sono entrambe variabili poissoniane indipendenti*. Detto in altre parole, se \mathbf{N} è una variabile poissoniana ed I_i , per $\{\mathbf{N} = N\}$ fissato, una variabile binomiale, allora I_i ed $(\mathbf{N} - I_i)$ sono variabili poissoniane indipendenti.

Poiché sappiamo che la deviazione standard della poissoniana è pari alla radice quadrata della media, possiamo immediatamente modificare la (6.112) nella forma:

$$\sigma[I_i] \equiv \sigma_i = \sqrt{\lambda p_i} \simeq s(n_i) = \sqrt{n_i},$$

dove ai valori veri si sono sostituiti quelli misurati. L'errore statistico dei canali è dato allora da:

$$s(n_i) = \sqrt{n_i}, \quad s\left(\frac{n_i}{N}\right) = \frac{1}{N} \sqrt{n_i} = \sqrt{\frac{f_i}{N}}. \quad (6.114)$$

Riassumiamo ora in uno schema coerente i risultati ottenuti. La stima del numero di eventi vero μ_i (speranza matematica o valore atteso) nel canale i -esimo è dato dalla (6.108) ed il corrispondente intervallo di confidenza approssimato, di livello 68.3%, è dato da:

$$\mu_i \in n_i \pm \sqrt{n_i \left(1 - \frac{n_i}{N}\right)}, \quad (6.115)$$

per istogrammi raccolti con numero di eventi N determinato e da

$$\mu_i \in n_i \pm \sqrt{n_i}, \quad (6.116)$$

per istogrammi in cui il numero totale di eventi N è una variabile poissoniana. Per gli istogrammi normalizzati, le (6.115 e 6.116) si trasformano rispettivamente in:

$$p_i \in f_i \pm \sqrt{\frac{f_i(1-f_i)}{N}}, \quad f_i \equiv \frac{n_i}{N}, \quad (6.117)$$

$$p_i \in f_i \pm \sqrt{\frac{f_i}{N}}, \quad (6.118)$$

che rappresentano gli intervalli di stima per le quantità vere (6.109).

Queste formule sono valide per $n_i \geq 5, 10$, cioè per canali popolati da almeno una decina di eventi e in questo caso, come sappiamo, valgono i livelli di confidenza gaussiani.

Per canali con meno di 10 eventi, alla (6.117) va sostituita la (6.21) (con N al posto di n), e alla (6.115) la (6.21) moltiplicata per N . Le formule poissoniane (6.116, 6.118) restano invece invariate. In questi casi, tuttavia, i livelli di confidenza non sono gaussiani e vanno ricavati dalle distribuzioni binomiale e poissoniana, a seconda che N sia fisso o variabile.

Le conclusioni cui siamo pervenuti si prestano ad una soddisfacente rappresentazione intuitiva delle fluttuazioni del contenuto dei canali di un istogramma. Se consideriamo ad esempio un istogramma a due canali, nel caso di N costante il numero di eventi n_1 ed n_2 nei due canali è completamente correlato, dato che $n_1 + n_2 = N$. In realtà si ha a che fare con una sola variabile aleatoria e in questo caso gli errori statistici nei due canali sono uguali, come risulta evidente dalla (6.113), che è simmetrica in f e $(1-f)$, oppure dalla (6.112) dopo un minimo di algebra. In generale il fatto di avere N costante determina una correlazione tra i canali, data dalla covarianza (4.90), che statisticamente, per $n_i, n_j > 10$, può essere stimata con buona approssimazione come:

$$s(n_i, n_j) = -N f_i f_j. \quad (6.119)$$

Quando invece N è una variabile poissoniana, ogni canale dell'istogramma si comporta come un contatore poissoniano e indipendente di eventi e le fluttuazioni vanno come la radice quadrata del numero degli eventi.

La rappresentazione grafica dell'istogramma, per essere corretta, va completata con la rappresentazione degli errori statistici. Per convenzione questi errori, calcolati con le (6.112, 6.113) a seconda dei casi, si riportano in grafico come intervalli $\pm s_i$ centrati sui valori n_i . Questi intervalli, chiamati barre d'errore (*error bars*), definiscono una banda entro cui sono localizzati i valori veri μ_i o p_i delle (6.108, 6.109). Tuttavia, come sempre, bisogna ricordarsi che i livelli di confidenza, se $n_i > 5, 10$, seguono la legge 3σ . Pertanto, la banda totale entro cui sono contenuti con ragionevole certezza questi valori è

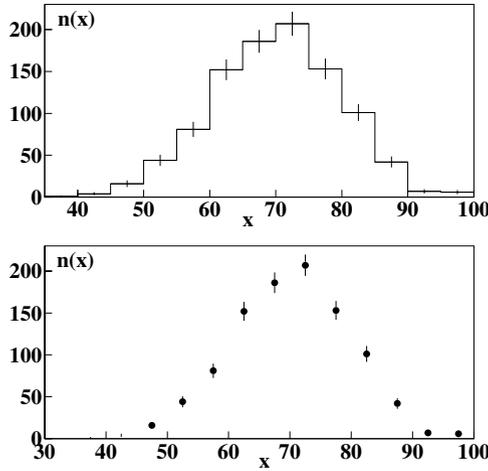


Fig. 6.13. Istogramma di Fig. 6.12 completato con le barre d'errore

in realtà larga 3 volte le barre d'errore riportate nei grafici. L'istogramma di Fig. 6.12, completato con le barre d'errore calcolate con la (6.112), è mostrato in Fig. 6.13.

Abbiamo dimostrato che le *fluttuazioni* negli eventi contenuti in un certo canale dell'istogramma obbediscono alle leggi di probabilità binomiale o poissoniana; questa è una legge del tutto generale, indipendente dalla densità $p(x)$ della popolazione da cui il campione proviene, che può essere qualunque. Come illustrato in Fig. 6.14, la densità determina invece la struttura globale del campione, che è data dai valori medi o centrali del contenuto dei canali.

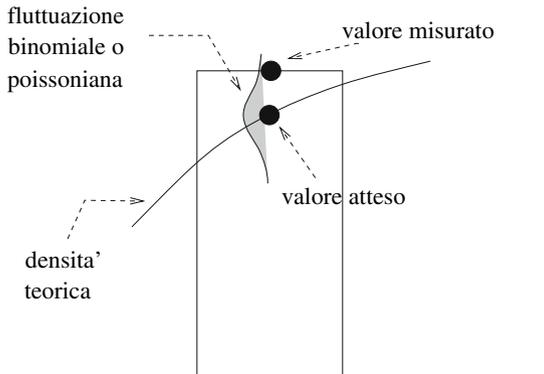


Fig. 6.14. Valore misurato, valore atteso in base ad un modello qualunque di densità e fluttuazione statistica del canale dell'istogramma. L'area ombreggiata va immaginata uscente ortogonalmente al foglio

Come verificare se la forma del campione sia o meno in accordo con un modello di densità scelto per la popolazione è l'argomento del prossimo paragrafo.

6.17 Verifica di compatibilità tra campione e popolazione

In questo paragrafo completiamo lo studio del campione (6.107). La media e la varianza calcolate dai dati valgono:

$$m = \frac{1}{1000} \sum_{i=1}^{10} x_i n_i = 70.09, \quad s^2 = \frac{1}{999} \sum_{i=1}^{10} (x_i - m)^2 n_i = 95.40.$$

Poiché $N = 1000$, gli intervalli di confidenza per media, varianza e deviazione standard si possono calcolare dalle (6.49, 6.62):

$$\begin{aligned} \mu &\in m \pm \frac{s}{\sqrt{N}} = 70.1 \pm 0.3, \\ \sigma^2 &\in s^2 \pm s^2 \sqrt{\frac{2}{N-1}} = 95.4 \pm 4.3, \\ \sigma &\in s \pm s \frac{1}{\sqrt{2(N-1)}} = 9.77 \pm 0.22 \end{aligned}$$

e poiché i livelli di confidenza da associare a questi intervalli sono gaussiani, si può affermare che i risultati sono compatibili con i valori veri $\mu = 70$ e $\sigma = 10$.

Verifichiamo ora se la popolazione del campione ha una densità gaussiana, assumendo che i valori della media e della deviazione standard siano quelli veri, $\mu = 70$, $\sigma = 10$. I valori veri o attesi del numero degli eventi in ogni canale sono allora dati dalla (6.108), dove $p(x)$ è una gaussiana, $N = 1000$ e $\Delta x = 5$:

$$\begin{aligned} \mu_i &= N p_i = N \int_{\Delta x} \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} \right] dx \\ &\simeq 1000 \frac{1}{10\sqrt{2\pi}} \exp \left[-\frac{(x_i - 70)^2}{200} \right] 5. \end{aligned} \quad (6.120)$$

È utile ora costruire la Tab. 6.4, dove nelle colonne da (1) a (7) sono riportati nell'ordine: il valore dello spettro nel punto medio del canale (1), il numero di eventi (2), il valore vero del numero di eventi calcolato con la (6.120) (3), l'errore statistico s_i della (6.112) (4), la deviazione standard

$$\sigma_i = \sqrt{\mu_i \left(1 - \frac{\mu_i}{N} \right)}$$

del numero vero di eventi (5), i valori della variabile standard

$$t_i = \frac{|n_i - \mu_i|}{s_i}$$

Tabella 6.4. Dati necessari per il confronto tra l'istogramma di Fig.6.13 e la densità gaussiana della (6.120)

1	2	3	4	5	6	7
x_i	n_i	μ_i	s_i	σ_i	t_i	t'_i
37.5	1	1	1.0	1.0	0.00	0.00
42.5	4	4.5	2.0	2.1	0.25	0.24
47.5	16	15.8	4.0	3.9	0.05	0.05
52.5	44	43.1	6.5	6.4	0.14	0.14
57.5	81	91.3	8.6	9.1	1.20	1.15
62.5	152	150.6	11.3	11.9	0.12	0.12
67.5	186	193.3	12.3	12.5	0.59	0.58
72.5	207	193.3	12.8	12.5	1.10	1.10
77.5	153	150.6	11.4	11.3	0.20	0.21
82.5	101	91.3	9.5	9.1	1.02	1.06
87.5	42	43.1	6.3	6.4	0.17	0.17
92.5	7	15.8	2.6	3.9	3.38	2.26
97.5	6	4.5	2.4	2.1	0.90	0.71

del numero misurato di eventi in ogni canale calcolata con l'errore statistico (6) e infine i valori

$$t'_i = \frac{|n_i - \mu_i|}{\sigma_i},$$

calcolati con la deviazione standard vera (7). In Fig. 6.15 è riportato l'istogramma di Fig. 6.13 (colonne 1,2 e 5 di Tab. 6.4) con l'aggiunta della linea continua, che rappresenta i valori veri (attesi) provenienti dalla gaussiana (6.120) e calcolati in colonna 3 di Tab. 6.4.

Una prima valutazione dell'accordo tra dati e modello può essere fatta "ad occhio": circa il 68 % dei punti sperimentali deve "toccare", entro una barra d'errore, il corrispondente valore vero. Nel nostro caso, dalla colonna 7 di Tab. 6.4 e dalla Fig. 6.15, risulta che su 13 punti si ottiene l'accordo col valore atteso per 9 punti (69 %) entro $\pm\sigma_i$, 12 punti (92 %) entro $\pm 2\sigma_i$, 13 valori (100 %) entro $\pm 3\sigma_i$. Come si vede, le fluttuazioni sembrano rispettare molto bene le percentuali della legge 3σ (3.35), il che significa che la densità assunta come modello fornisce una buona rappresentazione dei dati.

Notiamo che si arriva alle stesse conclusioni anche utilizzando gli errori statistici, cioè le variabili standard di colonna 6. L'unico canale anomalo è quello corrispondente al valore $x_i = 92.5$, che fornisce in questo caso un valore standard stimato di 3.38, a fronte di un valore corretto di 2.26. Il disaccordo è originato dal basso contenuto del canale, che contiene solo 7 eventi. Per questo motivo, se si usano gli errori statistici, in genere si accettano discrepanze maggiori di 3σ nei canali con meno di 10 eventi.

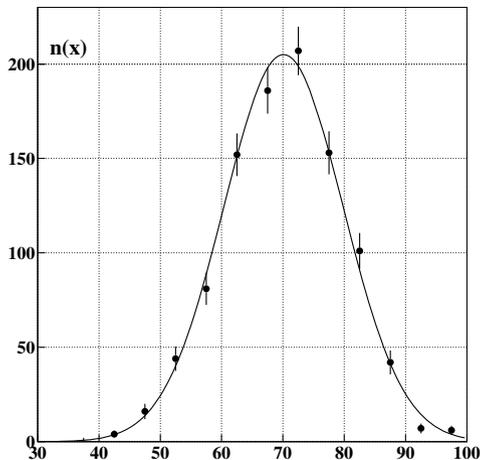


Fig. 6.15. Confronto tra l'istogramma di Fig. 6.13 ed i valori attesi sulla base della densità gaussiana di media 70 e deviazione standard 10

Oltre a questo test qualitativo, in statistica è possibile fare test quantitativamente precisi, il più importante dei quali, per il confronto tra densità della popolazione e istogramma, è il test χ^2 (*chi-square test*).

Per eseguire questo test occorre conoscere (o assumere come ipotesi nulla) le probabilità vere p_i di ogni canale ed eseguire la somma su tutti i K canali:

$$\chi^2 = \sum_{i=1}^K \frac{(n_i - \mu_i)^2}{\mu_i} = \sum_{i=1}^K \frac{(n_i - Np_i)^2}{Np_i}. \quad (6.121)$$

Se il numero totale N di eventi è variabile e si somma sui canali con più di 5, 10 eventi, la (6.121) risulta essere approssimativamente la somma del quadrato di K variabili gaussiane standard indipendenti (dato che per una distribuzione di Poisson $\sigma_i^2 = \mu_i$) e pertanto, in base al teorema di Pearson 3.3, essa si distribuisce approssimativamente come la densità χ^2 (3.66) con K gradi di libertà. I valori integrali della corrispondente densità χ^2 ridotto (3.72) sono riportati in Tab. D.3.

Se invece il numero totale N di eventi è costante, le variabili della somma (6.121) sono correlate, ma il teorema 4.5 della somma di Pearson ci assicura ancora che il risultato è una variabile χ^2 , ma questa volta con $(K - 1)$ gradi di libertà. Notiamo che, se N è costante, è sbagliato scrivere:

$$\chi^2 = \sum_{i=1}^K \frac{(n_i - Np_i)^2}{Np_i(1 - p_i)} \quad (\text{falso!}),$$

perché la somma è sì composta dal quadrato di variabili standard, ma queste sono correlate.

In conclusione, occorre sempre sommare il quadrato delle differenze tra le frequenze osservate e quelle vere e dividere per le frequenze vere, avendo cura di ricordarsi che i gradi di libertà sono pari al numero dei canali se N è una variabile poissoniana, mentre vanno diminuiti di una unità se N è costante. Tutto questo deriva dal teorema 4.5 di Pearson applicato all'analisi statistica degli istogrammi.

Eseguiamo ora il test χ^2 sui dati dell'istogramma (6.107). Nel nostro caso, poiché il primo canale contiene solo un evento (il contenuto non è una variabile gaussiana), occorre sommare tra loro i primi due canali e poi sommare gli altri undici canali rimanenti. Il valore χ^2 ridotto che si ottiene dai dati di Tab. 6.4 è allora:

$$\chi_R^2(11) = \frac{1}{11} \left[\frac{(n_1 + n_2 - \mu_1 - \mu_2)^2}{\mu_1 + \mu_2} + \sum_{i=3}^{13} \frac{(n_i - \mu_i)^2}{\mu_i} \right] = \frac{8.987}{11} = 0.82 . \quad (6.122)$$

Poiché N è costante, il numero dei gradi di libertà è uguale al numero dei punti della somma meno uno, cioè 11.

Col calcolo del χ^2 si raggiunge, per così dire, il massimo della sintesi, perché i risultati di Tab. 6.4 e di Fig. 6.15, relativi al confronto tra dati e modello, vengono condensati in un solo numero, nel nostro caso 0.82. Ovviamente, se rifacessimo un altro esperimento, otterremmo un numero diverso, perché il χ_R^2 della (6.122) è il valore assunto da una variabile aleatoria che ha la densità di Tab. D.3.

Procediamo ora al test χ^2 vero e proprio. Cerchiamo in Tab. D.3, nella riga corrispondente a 11 gradi di libertà, l'area corrispondente al valore di 0.82: troviamo un valore di circa il 60 %. In termini formali, la situazione corrisponde all'equazione:

$$P\{Q_R \geq 0.82\} \simeq 0.60 ,$$

dove Q_R è una variabile che segue la densità χ^2 ridotto con 11 gradi di libertà.

Definendo il livello di significatività come:

$$P\{Q_R \leq 0.82\} = 1 - P\{Q_R > 0.82\} \simeq 1 - 0.60 = 0.40 ,$$

se il modello è valido, sono possibili valori di χ^2 ridotto minori di quello trovato in almeno il 40 % delle prove. Poiché un livello di significatività del 40% rende altamente probabile l'errore di I tipo, l'ipotesi va accettata. In conclusione, abbiamo generato al calcolatore un campione di 1000 eventi da una distribuzione gaussiana con $\mu = 70$ e $\sigma = 10$. Successivamente, abbiamo eseguito l'analisi del campione ed effettuato i test statistici rispetto alla densità vera. Questi test hanno mostrato buon accordo tra dati e modello.

Come vedete, la logica del test χ^2 è esattamente uguale a quella utilizzata con la Tab. D.1 per le variabili gaussiane. Ciò che cambia è solo il tipo di variabile e la funzione di test. Non occorre quindi ridiscutere il livello di significatività e l'errore di I tipo in ulteriore dettaglio.

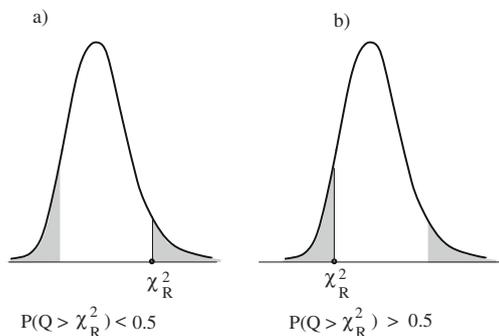


Fig. 6.16. Livello di significatività osservato nel caso di test a due code (aree ombreggiate) quando il valore trovato χ_R^2 corrisponde a) ad una probabilità $P\{Q_R > \chi_R^2\} < 0.5$ oppure b) a $P\{Q_R > \chi_{R\alpha}^2\} > 0.5$. Le zone ombreggiate hanno aree uguali.

L'unico punto su cui fare attenzione, come abbiamo già spiegato, è che la densità χ^2 , a differenza della gaussiana, non è simmetrica. Si rende pertanto necessario tabulare i valori integrali in tutto il campo di esistenza della funzione, invece che in una sola delle due metà, e questo fatto porta a probabilità cumulative estratte dalla tabella che possono superare il 50%.

Il test χ^2 può essere fatto ad una o due code. Nel test ad una coda il livello del test riguarda in genere la coda a destra ed il livello di significatività osservato SL (o p -value) vale

$$SL = P\{Q_R \geq \chi_R^2(\nu)\}. \quad (6.123)$$

Se il test è a due code, come in Fig. 6.16, allora

$$SL = 2P\{Q_R(\nu) > \chi_R^2(\nu)\} \text{ se } P\{Q_R(\nu) > \chi_R^2(\nu)\} < 0.5 \quad (6.124)$$

$$SL = 2P\{Q_R(\nu) < \chi_R^2(\nu)\} \text{ se } P\{Q_R(\nu) > \chi_R^2(\nu)\} > 0.5.$$

Occorre fare ancora alcune ulteriori considerazioni quando i valori di χ^2 sono troppo alti o troppo bassi: in ognuno di questi due casi test indica che *le fluttuazioni dei dati intorno ai valori assunti come veri non sono di tipo statistico*. Quando la (6.123) fornisce valori troppo piccoli:

$$P\{Q_R > \chi_R^2(\nu)\} < 0.01,$$

il valore di χ^2 ridotto trovato si situa nella coda a destra e risulta troppo alto. In questo caso è altamente probabile che la densità della popolazione assunta come modello non sia quella vera. Può anche essere che gli errori assegnati ai canali siano stati calcolati male e risultino sottostimati. Se si è sicuri di aver calcolato bene gli errori, il risultato del test è il rifiuto dell'ipotesi.

Molto più raramente capita che il valore di χ^2 sia troppo piccolo:

$$P\{Q_R > \chi_R^2(\nu)\} > 0.99 .$$

Questo caso si verifica quando le probabilità a priori p_i sono calcolate da una popolazione che, avendo un numero eccessivo di parametri, tende ad *interpolare* i dati, oppure quando gli errori sono stati erroneamente sovrastimati. Approfondiremo meglio questi concetti nel cap. 10.

Spesso il χ^2 di un istogramma viene calcolato dividendo per le frequenze misurate invece che per quelle vere:

$$\chi^2 = \sum_{i=1}^K \frac{(n_i - Np_i)^2}{n_i} . \quad (6.125)$$

In questo caso il denominatore di ogni addendo è approssimato ma è *indipendente da ogni modello*. Se si opera nell'ambito di un modello, è più corretto e consistente dividere per le frequenze vere. Tuttavia, se nella verifica di ipotesi si considerano solo i canali con più di 5-10 eventi, l'uso delle frequenze misurate porta quasi sempre a risultati equivalenti a quelli ottenuti con quelle vere. Nei test con procedure di minimizzazione, che vedremo nei capitoli 9 e 10, si utilizzano spesso al denominatore le frequenze misurate, perché, essendo indipendenti dai modelli, esse restano costanti durante la procedura di adattamento del modello ai dati, semplificando di molto gli algoritmi di calcolo.

La nostra routine `Chisq(n,mu,nu)`, assegnati in ingresso la matrice dei dati `n`, quella dei valori attesi `mu` ed i gradi di libertà `nu`, calcola il χ^2 con la (6.121) e fornisce in uscita un vettore a tre posizioni: χ^2 , χ^2/ν , *SL*.

Esercizio 6.13

Fare l'analisi statistica completa dell'esperimento delle 10 monete di Tab. 2.2 e Fig. 2.1, assumendo come modello che i lanci siano indipendenti e che tutte le monete abbiano una probabilità a priori testa/croce pari a 1/2.

Risposta. I dati sono riportati ancora per comodità nella nuova Tab. 6.5, dove sono state anche calcolate le quantità di base che ci serviranno nella analisi. Per non fare confusione, occorre per prima cosa definire i parametri:

- numero di lanci per prova: $n = 10$;
- numero di prove: $N = 100$;
- numero totale di lanci: $N \cdot n = M = 1000$.

Il primo test che possiamo fare è quello sul numero totale di successi, cioè sul numero totale di teste. Dalle colonne 1 e 2 della tabella otteniamo:

$$x = 2 \cdot 5 + 3 \cdot 13 + \dots = 521 \quad \text{successi} .$$

Tabella 6.5. Istogramma relativo all'esperimento costituito da $N = 100$ prove, ognuna consistente nel lancio di 10 monete (vedi anche Tab. 2.2). Le colonne contengono: il numero di teste possibili (1), il numero di volte (successi) in cui si è ottenuto il numero di teste riportato in prima colonna (2), le speranze matematiche, o numero vero di eventi, date dal numero di prove per la probabilità binomiale della (2.30) con $n = 10$, $p = 1/2$ (3), l'errore statistico dell'istogramma calcolato col numero vero di eventi (4) $\sigma_i = \sqrt{\mu_i \left(1 - \frac{\mu_i}{N}\right)}$, e con quello misurato (5) $s_i = \sqrt{n_i \left(1 - \frac{n_i}{N}\right)}$, i valori del χ^2 per ogni punto, ottenuti utilizzando nel calcolo la probabilità vera (6) e la frequenza misurata (7)

<i>Spettro</i> (<i>n. teste</i>)	<i>Suc-</i> <i>cessi</i>	<i>Bino-</i> <i>miale</i>	<i>Dev. std.</i> "vera"	<i>Dev. std.</i> <i>stimata</i>	χ^2 "vero"	χ^2 <i>stimato</i>
x_i	n_i	μ_i	σ_i	s_i	$\frac{(n_i - \mu_i)^2}{\mu_i}$	$\frac{(n_i - \mu_i)^2}{n_i}$
1	2	3	4	5	6	7
0	0	0.0	0.0	0.0	0.00	0.00
1	0	1.0	1.0	0.0	1.00	—
2	5	4.4	2.0	2.2	0.08	0.07
3	13	11.7	3.2	3.4	0.14	0.13
4	12	20.5	4.0	3.2	3.52	6.02
5	25	24.6	4.3	4.3	0.01	0.01
6	24	20.5	4.0	4.3	0.60	0.51
7	14	11.7	3.2	3.5	0.45	0.38
8	6	4.4	2.0	2.4	0.58	0.43
9	1	1.0	1.0	1.0	0.00	0.00
10	0	0.0	0.0	0.0	0.00	0.00

Poiché la speranza matematica è $Mp = 500$ e la dispersione vera, calcolata con la (3.5), vale $\sigma = \sqrt{500(1 - 1/2)} = 15.8$, otteniamo il valore standard:

$$t = \frac{x - Mp}{\sqrt{Mp(1 - p)}} = \frac{521 - 500}{15.8} = 1.33 ,$$

al quale corrisponde in Tab. D.1 un livello di significatività

$$P\{|T| \geq 1.33\} = 2 \cdot (0.5000 - .4082) = 0.1836 \simeq 18.4\% ,$$

il quale permette di affermare che, ripetendo varie volte mille lanci indipendenti con monete non truccate, in circa il 18% dei casi si avranno scostamenti (in più o in meno) da 500 maggiori di 21.

Nell'esercizio 2.6 abbiamo calcolato la media e la varianza dell'istogramma di Tab. 6.5:

$$m = 5.21 , \quad s^2 = 2.48 , \quad s = 1.57 ,$$

ed i corrispondenti valori veri della binomiale con $n = 10$ e $p = 1/2$:

$$\frac{Mp}{N} = np = \mu = 5.00, \quad \sigma^2 = 2.50, \quad \sigma = 1.58.$$

Il test sulla media, eseguito con la (6.46), fornisce allora:

$$t = \frac{m - \mu}{\sigma/\sqrt{N}} = \frac{5.21 - 5.00}{0.158} \simeq \frac{5.21 - 5.00}{0.157} = 1.34.$$

Il test sulla media risulta uguale al test sul numero totale di lanci, perché vale l'identità

$$\frac{Mp - x}{\sqrt{Mp(1-p)}} = \frac{Nnp - Nx/N}{\sqrt{\frac{N^2}{N}np(1-p)}} = \frac{np - x/N}{\sqrt{np(1-p)}/\sqrt{N}} = \frac{\mu - m}{\sigma/\sqrt{N}}.$$

Per il test sulla varianza possiamo utilizzare la formula per grandi campioni (6.62) e calcolare il valore standard:

$$t = \frac{2.50 - 2.48}{2.50\sqrt{\frac{2}{N}}} = \frac{2.50 - 2.48}{0.35} = 0.06,$$

cui corrisponde in Tab. D.1 un livello di significatività

$$P\{|T| \geq 0.06\} = 0.94 = 94\%.$$

Veniamo ora al test χ^2 , che è il test finale complessivo sulla forma del campione. Raggruppando i primi e gli ultimi 3 canali, in modo da avere sempre un numero di eventi per canale ≥ 5 , otteniamo:

$$\chi_R^2(6) = \frac{1}{6} \left[\frac{(5 - 5.4)^2}{5.4} + \frac{(13 - 11.7)^2}{11.7} + \dots + \frac{(7 - 5.4)^2}{5.4} \right] = \frac{5.18}{6} = 0.86.$$

Il numero dei gradi di libertà è uguale a 6, poiché il numero totale di eventi è fissato e nella somma compaiono 7 termini. A questo valore di χ_R^2 corrisponde in Tab. D.3 un livello di significatività

$$P\{Q_R > 0.86\} \simeq 0.52$$

del 50% circa: in pratica abbiamo ottenuto un evento $\{Q_R = \chi_R^2\}$ nell'intorno del valore più probabile.

Se nel calcolo del χ^2 avessimo diviso per le frequenze misurate come nella (6.125), avremmo ottenuto $\chi_R^2 = 7.42/6 = 1.24$ ed un livello di significatività di circa il 25%. I due risultati, per quanto analoghi dal punto di vista statistico, differiscono sensibilmente a causa della dimensione del campione ($N = 100$) piuttosto ridotta per il test χ^2 .

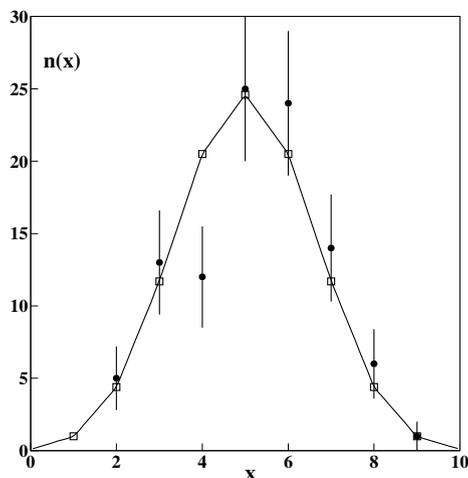


Fig. 6.17. Dati sperimentali con barre d'errore e previsioni della densità binomiale (quadrati vuoti) per 100 prove consistenti nel lancio di 10 monete. Per facilitare il confronto, i punti discreti della densità binomiale sono stati uniti da tratti di retta

In tutti i test eseguiti si sono ottenuti livelli di significatività molto elevati, il che dimostra un buon accordo statistico tra i dati e il modello che assume lanci indipendenti di monete non truccate.

I dati sperimentali con i relativi errori statistici e i valori teorici dati dalla distribuzione binomiale sono riportati anche in Fig. 6.17.

Abbiamo finalmente analizzato in modo completo ed esauriente l'esperimento delle 10 monete, la cui presenza ... incombeva minacciosa da tempo. Vi facciamo notare che non si tratta di un esperimento tanto facile da interpretare, se ci sono volute 242 pagine di testo per completarne l'analisi.

6.18 Verifica di ipotesi con test non parametrici

Abbiamo visto come applicare il test χ^2 nel confronto tra un istogramma e una densità modello dipendente da parametri. Vediamo ora come modificare questa procedura quando si confrontano tra loro due o più campioni, senza assumere per la loro popolazione una specifica funzione di densità. Questi test vengono detti *non parametrici*.

Notiamo prima di tutto che, se il dato sperimentale consiste in una sola frequenza corrispondente ad un numero di successi > 10 , l'uso del test χ^2 è equivalente all'uso del test gaussiano a due code su una variabile standard.

Infatti, le variabili

$$T = \frac{X - \mu}{\sigma}, \quad Q(1) = \frac{(X - \mu)^2}{\sigma^2},$$

seguono l'una la densità gaussiana, l'altra quella χ^2 con un grado di libertà. Potete facilmente controllare come, assegnando a caso un valore di T ed eseguendo il test a due code con la Tab. D.1 ed eseguendo il test ad una coda per la variabile $Q(1)$ con la Tab. D.3, si ottengano risultati identici.

Nel caso di una coppia di variabili gaussiane, la verifica di compatibilità può essere effettuata con la densità gaussiana o di Student, secondo il metodo della differenza del par. 6.15. In alternativa, se l'esperimento misura la frequenza di realizzazione di un evento in due campionamenti indipendenti, spesso viene usata l'analisi delle *tabelle di contingenza* con il test χ^2 . Questo metodo consiste nel costruire una tabella contenente il numero di successi n_a ed n_b ottenuti con due serie di tentativi N_a ed N_b :

	<i>Successi</i>	<i>Insuccessi</i>	<i>Totale</i>
campione A	n_a	$N_a - n_a$	N_a
campione B	n_b	$N_b - n_b$	N_b
totale	$n_a + n_b$	$N_a + N_b - n_a - n_b$	$N_a + N_b = N$

Supponendo che i due campioni provengano da un medesimo processo stocastico con probabilità vera p , si costruisce la tavola di contingenza attesa:

	<i>Successi</i>	<i>Insuccessi</i>	<i>Totale</i>
campione A	pN_a	$(1 - p)N_a$	N_a
campione B	pN_b	$(1 - p)N_b$	N_b
totale	$p(N_a + N_b)$	$(1 - p)(N_a + N_b)$	$N_a + N_b = N$

Ogni riga della tabella di contingenza sperimentale costituisce un istogramma a due canali e la tabella attesa associata fornisce il corrispondente valore atteso del numero di successi e insuccessi.

In base ai teoremi di Pearson 4.4, 4.5 ed al teorema 3.4 di additività del χ^2 la quantità

$$\chi^2 = \frac{(n_a - pN_a)^2}{pN_a} + \frac{[N_a - n_a - (1 - p)N_a]^2}{(1 - p)N_a} + \frac{(n_b - pN_b)^2}{pN_b} + \frac{[N_b - n_b - (1 - p)N_b]^2}{(1 - p)N_b}, \quad (6.126)$$

si può considerare come il valore assunto da una variabile χ^2 con 2 gradi di libertà.

Se p è incognita e l'ipotesi nulla afferma solo che essa è la medesima per i campioni A e B , si può effettuare una stima puntuale dai dati. Se i successi e gli insuccessi osservati sono tutti > 10 , allora si pone

$$p \simeq \hat{p} = \frac{n_a + n_b}{N_a + N_b} = \frac{n_a + n_b}{N} . \quad (6.127)$$

Poiché questa assunzione introduce una ulteriore relazione di dipendenza tra i dati, in base alla definizione 6.3 la (6.126) rappresenta i valori presi da una variabile χ^2 con un solo grado di libertà. Notiamo che il metodo è approssimato, perché la probabilità vera è stata stimata dalla frequenza osservata (6.127). La tendenza verso la densità di χ^2 della variabile (6.126) si ha ovviamente per $N_a, N_b \rightarrow \infty$. Tuttavia il metodo viene accettato e fornisce buoni risultati quando vale la condizione $N > 40$, $n_a, n_b > 10$, la quale assicura che il numero di successi sia gaussiano e che la stima della probabilità vera sia affidabile. Vi ricordiamo che questo punto è stato ampiamente discusso nel par. 6.6.

Esercizio 6.14

Per provare la validità di un vaccino, si sono studiati due gruppi di cavie, uno vaccinato, l'altro non vaccinato. I risultati sono riportati nella seguente tavola di contingenza:

	<i>Ammalate</i>	<i>Sane</i>	<i>Totale</i>
cavie vaccinate	10	41	51
cavie non vaccinate	18	29	47
totale	28	70	98

Dire se i risultati sono significativi per il test a una coda a destra ad un livello del 5%.

Risposta. Se assumiamo come ipotesi nulla H_0 che il vaccino non sia efficace, allora le differenze tra i due gruppi di cavie sono solo casuali. Sotto questa ipotesi la miglior stima della probabilità di contrarre la malattia sarà data dalla frequenza (6.127):

$$\hat{p} = \frac{28}{98} = 0.286 \simeq 29\% .$$

Analogamente, la probabilità di non contrarre la malattia sarà pari a 0.714. Possiamo allora costruire la tavola di contingenza attesa (cioè dei valori attesi sotto H_0 : $0.286 \cdot 51 = 14.6$, ecc.):

	<i>Ammalate</i>	<i>Sane</i>	<i>Totale</i>
cavie vaccinate	14.6	36.4	51
cavie non vaccinate	13.4	33.6	47
totale	28.0	70.0	98

Dalla formula (6.126) abbiamo:

$$\chi^2 = \frac{(10 - 14.6)^2}{14.6} + \frac{(18 - 13.4)^2}{13.4} + \frac{(41 - 36.4)^2}{36.4} + \frac{(29 - 33.6)^2}{33.6} = 4.24 .$$

Dalla Tab. D.3 si vede che $SL = 5\%$ corrisponde, con un grado di libertà, ad un valore $\chi^2(1) = 3.84$, mentre il valore trovato, $\chi^2(1) = 4.24$, corrisponde a $SL \simeq 4\%$. Concludiamo che il vaccino supera il test di efficacia al livello del 5%, perché la probabilità di sbagliare, scartando H_0 ed affermando che il vaccino è efficace, è solo del 4%.

Dato che questo problema richiede l'uso di una tavola di contingenza a due entrate, possiamo usare in alternativa il metodo del par. 6.15 e la formula (6.100). Dalla tabella di contingenza sperimentale si ha in questo caso:

$$t_f = \left| \frac{10}{51} - \frac{18}{47} \right| \left[\frac{\frac{10}{51}(1 - \frac{10}{51})}{51} + \frac{\frac{18}{47}(1 - \frac{18}{47})}{47} \right]^{-1/2} = 2.074 . \quad (6.128)$$

Notate che qui non si utilizza la tabella di contingenza attesa. Il valore $t_f^2 = 4.30$ corrisponde ad un χ^2 con un grado di libertà in accordo col valore 4.24 trovato in precedenza con la tavola di contingenza attesa. Avremmo trovato due valori identici se nella (6.128) avessimo usato al denominatore, invece che le frequenze sperimentali, quelle della tabella di contingenza attesa. Dalla Tab. D.1 si ottiene, per $t = 2.07$, un valore di 0.4808, pari ad un livello di significatività, per un test a due code, di:

$$SL = 2 \cdot (0.5000 - 0.4808) = 0.0384 \simeq 3.8\% ,$$

in accordo col test χ^2 effettuato in precedenza.

Va notato che entrambi i metodi usati sono approssimati: il metodo che utilizza la tabella di contingenza attesa assume che le probabilità vere coincidano con le frequenze calcolate dai dati supponendo che il vaccino sia inefficace, mentre il metodo della differenza utilizza, al posto degli errori veri, gli errori statistici calcolati dalle frequenze sperimentali.

Mostriamo ora come il test del χ^2 possa essere esteso a tavole di contingenza di qualunque dimensione. In generale, possiamo considerare la tabella di contingenza seguente:

	Canale 1	Canale 2	...	Canale c	Totale
campione 1	n_{11}	n_{12}	...	n_{1c}	$\sum_j n_{1j}$
campione 2	n_{21}	n_{22}	...	n_{2c}	$\sum_j n_{2j}$
...
campione r	n_{r1}	n_{r2}	...	n_{rc}	$\sum_j n_{rj}$
totale	$\sum_i n_{i1}$	$\sum_i n_{i2}$...	$\sum_i n_{ic}$	$\sum_{ij} n_{ij} = N$

la quale è composta da r righe (i campioni in forma di istogramma) e c colonne (i canali degli istogrammi). Nel caso di un solo istogramma, il test si riduce al caso discusso nel par. 6.17.

Come al solito, supponiamo di voler verificare se i campioni sono omogenei, se cioè provengono tutti dalla stessa popolazione. Se questa ipotesi nulla è vera, allora ad ogni colonna della tabella potremo associare una probabilità vera (incognita) p_j , che moltiplicata per il numero totale di elementi $\sum_j n_{ij} = N_i$ di ogni campione (riga), fornirà, in ogni cella, il valore atteso $p_j N_i$ del numero di eventi. Tenendo presente anche in questo caso i teoremi di Pearson 4.4, 4.5 ed il teorema 3.4 di additività del χ^2 , possiamo dire che la quantità

$$\chi^2 = \sum_{ij} \frac{(n_{ij} - p_j \sum_k n_{ik})^2}{p_j \sum_k n_{ik}} = \sum_{ij} \frac{(n_{ij} - N_i p_j)^2}{N_i p_j} \quad (6.129)$$

segue una distribuzione di χ^2 .

Il valore incognito della probabilità si può stimare dai dati generalizzando la (6.127) e sommando per colonne:

$$\hat{p}_j = \frac{\sum_{i=1}^r n_{ij}}{N} \quad (6.130)$$

dove N è il gran totale della tabella.

Veniamo ora al calcolo dei gradi di libertà. Dal teorema di Pearson 4.5 sappiamo che ogni riga contribuisce con $(c - 1)$ gradi di libertà. Tuttavia questo numero, che vale $r(c - 1)$, va diminuito di un valore pari al numero di relazioni (6.130) con le quali abbiamo stimato le probabilità vere. Queste ultime non sono pari a al numero di colonne c , ma a $(c - 1)$, perché la probabilità dell'ultima colonna resta determinata dalla relazione di chiusura:

$$\hat{p}_c = 1 - \sum_{j=1}^{c-1} \hat{p}_j .$$

Abbiamo pertanto un numero di gradi di libertà che vale:

$$\nu = r(c - 1) - (c - 1) = (r - 1)(c - 1) . \quad (6.131)$$

In conclusione, per la verifica di una ipotesi H_0 , con un livello di significatività $SL = \alpha$ prefissato (in genere $\alpha = 0.01-0.05$), su uno o più istogrammi raccolti in una tabella di contingenza con $(i = 1, 2, \dots, r)$ righe e $(j = 1, 2, \dots, c)$ colonne, si procede come segue: si calcola il χ^2 ridotto

$$\chi_R^2(\nu) = \frac{1}{\nu} \sum_{ij} \frac{(n_{ij} - N_i \hat{p}_j)^2}{N_i \hat{p}_j} = \frac{1}{\nu} \left(\sum_{ij} \frac{n_{ij}^2}{N_i \hat{p}_j} - N \right) , \quad (6.132)$$

$$\nu = (r - 1)(c - 1) , \quad (6.133)$$

dove N_i sono i totali di riga, N è il gran totale della tabella e \hat{p}_j è data dalla (6.130); si rifiuta H_0 ad un livello di significatività α se dalla Tab. D.3 risulta

$$P\{Q_R(\nu) \geq \chi_R^2(\nu)\} < \alpha. \quad (6.134)$$

Questo test è il pilastro della analisi statistica non parametrica delle frequenze.

Si noti che, nel caso non parametrico, non si rigetta l'ipotesi quando il χ^2 è piccolo ed i valori di probabilità sono alti, poiché non va preso in considerazione il caso in cui il modello a priori di densità contenga un numero troppo elevato di parametri. Il test è quindi sempre *a una coda*, quella di destra. Tuttavia, è bene notare che un valore di χ^2 troppo piccolo, corrispondente a valori di probabilità > 0.99 nella (6.134), indica una coincidenza sospetta tra la tabella di contingenza sperimentale e quella attesa, dovuta in genere a *fluttuazioni non casuali* o a correlazioni nei dati. In questi casi occorre cautela nell'accettare l'ipotesi nulla, specialmente se si è in presenza di un numero elevato di gradi di libertà. Vi raccomandiamo, a questo proposito, il problema 6.11.

Potremmo evitare di usare il test χ^2 e ricorrere sempre al test gaussiano delle differenze come nel caso delle tabelle di contingenza (2×2)? La risposta è negativa e vogliamo spiegarvene il motivo.

Nel caso di tavole di contingenza con più di due dimensioni potremmo pensare di fare una somma di differenze di frequenze divise per il relativo errore, come nella (6.100). Se togliamo il valore assoluto della differenza, anche la somma di queste variabili aleatorie sarà gaussiana e potremmo eseguire il test di significatività con la Tab. D.1. Tuttavia, una somma di grandi fluttuazioni (positive e negative) potrebbe fornire un buon valore della variabile standard anche nel caso di una ipotesi grossolanamente sbagliata. Se, per evitare questo inconveniente, mantenessimo il valore assoluto nella somma delle differenze, allora non avremmo più una variabile finale gaussiana e non potremmo pertanto eseguire il test.

La variabile χ^2 , invece, essendo una somma di quadrati di fluttuazioni, combina sempre positivamente tutte le fluttuazioni dei dati rispetto al valore vero ed è pertanto più affidabile. In linguaggio tecnico, possiamo dire che il test χ^2 , rispetto a quello gaussiano delle differenze, minimizza l'errore di II tipo (accettare una ipotesi sbagliata) e che pertanto, in base alla terminologia che introdurremo nel cap. 9, è un test più potente.

In ambito SCILAB è possibile scrivere una routine estremamente sintetica, che abbiamo chiamato `Tabcont(tabel)`, per l'analisi di tabelle di contingenza `tabel` di qualsiasi dimensione:

```
// funzione SCILAB per il calcolo del chi quadrato,
// del chi quadrato ridotto e del livello di significativita' SL
// a partire da una tabella di contingenza tabel
// Output: [cont] = [chi2 ; chi2 ridotto; livello SL]
```

```

function [cont]=Tabcont(tabel)

// rilevazione dimensioni tabella
[r c]= size(tabel);

// gran totale
GT = sum(tabel);

// probabilita' vere stimate (matrice riga somma delle colonne)
P = sum(tabel,'r')./GT;

// totali per riga (matrice colonna somma delle righe)
T = sum(tabel,'c');

// creazione della tabella di contingenza attesa
tabatt = T*P;

// calcolo del chi quadrato
cont = Chisq(tabel,tabatt,(r-1)*(c-1));

endfunction

```

La routine sfrutta la funzione SCILAB `sum` per sommare righe e colonne della tabella. P è la matrice delle probabilità (6.130), T quella dei totali di riga N_i della (6.132). Il prodotto righe per colonne $T*P$ è la tabella di contingenza attesa $N_i\hat{p}_j$. Il vettore di uscita `cont`, che contiene i valori di χ^2 , χ^2/ν , SL , viene ottenuto con la routine `Chisq` descritta a pag. 239, che potete copiare dal nostro sito.

Esercizio 6.15

In una ditta di cinghie dentate di gomma tre operatori X , Y e Z eseguono un test visivo sulla bontà del prodotto. Essi possono accettare il pezzo oppure scartarlo per un difetto di tipo A o B . Un controllo ha fornito la tabella seguente:

	Tipo A	Tipo B	Buono	Totale
operatore X	10	54	26	90
operatore Y	20	50	50	120
operatore Z	30	35	35	100
totale	60	139	111	310

Stabilire se i criteri di controllo dei tre operatori sono omogenei ad un livello dell'1%.

Risposta. Utilizzando la (6.130) otteniamo le probabilità

$$\hat{p}_A = \frac{60}{310} = 0.193, \quad \hat{p}_B = \frac{139}{310} = 0.448, \quad \hat{p}_{buono} = \frac{111}{310} = 0.358.$$

Queste probabilità permettono di ottenere la tabella di contingenza attesa:

	Tipo A	Tipo B	Buono	Totale
operatore X	17.5	40.4	32.2	90
operatore Y	23.2	53.8	43.0	120
operatore Z	19.3	44.8	35.8	100
totale	60.0	139.0	111.0	310

Se rifate i conti, potrete trovare qualche piccola differenza rispetto ai valori sopra riportati a causa degli arrotondamenti.

Il valore del χ^2 si ottiene dalla (6.132):

$$\chi^2 = \frac{10^2}{17.5} + \frac{54^2}{40.4} + \cdots + \frac{50^2}{53.8} + \cdots + \frac{35^2}{35.8} - 310 = 18.88 .$$

In base alla (6.133), i gradi di libertà sono:

$$\nu = (3 - 1)(3 - 1) = 4 .$$

Il χ^2 ridotto (6.132) vale pertanto:

$$\chi_R^2(4) = \frac{18.88}{4} = 4.72 .$$

Dalla Tab. D.3 risulta che a questo valore corrisponde un livello di significatività $P\{Q_R(4) \geq \chi_R^2(4)\}$ molto piccolo, poiché al valore $\chi_R^2(4) = 3.71$ corrisponde già un livello di significatività del 5 per mille; il valore trovato di 4.73 corrisponde dunque ad un livello ancora minore. La routine **Tabcont** fornisce infatti $SL = 8.3 \cdot 10^{-4}$. Possiamo concludere che gli operatori seguono criteri diversi, poiché l'ipotesi di omogeneità, ad un livello dell'1%, va rifiutata.

6.19 Stima della correlazione

Quando si calcola da un insieme finito di dati (x_i, y_i) il coefficiente di correlazione campionario si ottiene generalmente un valore diverso da zero, anche quando le variabili sono incorrelate. Il problema è quindi, avendo ottenuto dai dati un coefficiente di correlazione r , verificare se esso è compatibile o meno con un valore nullo (verifica di ipotesi), oppure stimare l'intervallo di confidenza entro il quale è localizzato il coefficiente di correlazione vero ρ (stima di un parametro).

La stima campionaria del coefficiente di correlazione (4.31) per un insieme finito di N elementi richiede prima la definizione della covarianza campionaria $S(X, Y) \equiv S_{XY}$.

Senza perdere di generalità possiamo considerare una coppia di variabili centrate (a media vera nulla), per le quali la covarianza vera vale;

$$\text{Cov}[X, Y] = \langle (X - \mu_x)(Y - \mu_y) \rangle = \langle XY \rangle . \quad (6.135)$$

Per individuare eventuali fattori di distorsione (*bias*) è necessario, analogamente a quanto fatto per la varianza nella (6.53), trovare il valor medio vero della somma campionaria:

$$\left\langle \sum_{i=1}^N (X_i - M_X)(Y_i - M_Y) \right\rangle ,$$

che in generale avrà $\{M_X = m_x \neq 0\}$, $\{M_Y = m_y \neq 0\}$ anche quando i valori veri delle medie sono nulli. Applicando le proprietà di linearità dell'operatore media e notando che

$$\sum_i M_X Y_i = M_X \sum_i Y_i = \frac{1}{N} \sum_j X_j \sum_i Y_i$$

e così via, abbiamo:

$$\begin{aligned} \left\langle \sum_{i=1}^N (X_i - M_X)(Y_i - M_Y) \right\rangle &= \left\langle \sum_i (X_i Y_i - M_X Y_i - M_Y X_i + M_X M_Y) \right\rangle \\ &= \left\langle \sum_i X_i Y_i \right\rangle - \frac{1}{N} \left\langle \sum_i X_i \sum_j Y_j \right\rangle \\ &= N \langle XY \rangle - \frac{1}{N} \left\langle \sum_i X_i \sum_j Y_j \right\rangle . \end{aligned} \quad (6.136)$$

L'ultimo termine di questa equazione si può scrivere come:

$$\left\langle \sum_i X_i \sum_j Y_j \right\rangle = \left\langle \sum_i X_i Y_i \right\rangle + \sum_{i \neq j} \langle X_i Y_j \rangle .$$

Ora, X_i ed Y_j sono indipendenti, perché sono valori provenienti da eventi diversi campionati indipendentemente (ricordiamo che la correlazione esiste nella coppia (X_i, Y_i) , relativa alla stessa prova!). Dalla (4.9) si può allora scrivere

$$\langle X_i Y_j \rangle = \langle X \rangle \langle Y \rangle = 0 ,$$

poiché per ipotesi le medie vere delle due variabili sono nulle. Dato che $\langle \sum XY \rangle = N \langle XY \rangle$, la (6.136) si risolve nell'equazione:

$$\left\langle \sum_{i=1}^N (X_i - M_X)(Y_i - M_Y) \right\rangle = \left\langle \sum_i X_i Y_i \right\rangle - \frac{1}{N} \left\langle \sum_i X_i Y_i \right\rangle = (N-1) \langle XY \rangle , \quad (6.137)$$

che, ricordando la (6.135), implica:

$$\text{Cov}[X, Y] = \left\langle \frac{1}{N-1} \sum_{i=1}^N (X_i - M_X)(Y_i - M_Y) \right\rangle. \quad (6.138)$$

In definitiva, la covarianza campionaria *non distorta* vale:

$$s(x, y) = \frac{1}{N-1} \sum_{i=1}^N (x_i - m_x)(y_i - m_y). \quad (6.139)$$

Questo risultato mostra che tanto per la varianza che per la covarianza, se le stime sono fatte rispetto alle medie campionarie, il fattore N al denominatore va sostituito con $N - 1$. Il coefficiente di correlazione lineare campionario assume allora la forma compatta:

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_i (x_i - m_x)(y_i - m_y)}{\sqrt{\sum_i (x_i - m_x)^2 \sum_j (y_j - m_y)^2}} = \frac{\sum_i (x_i y_i) / N - m_x m_y}{s_x s_y}, \quad (6.140)$$

dove nell'ultimo passaggio è stata utilizzata la (4.25). A questo valore corrisponde la variabile "coefficiente di correlazione campionario R ".

Per il calcolo di $s(x, y)$ o di r da un insieme di dati grezzi, potete usare la nostra routine `Covardat(x, flag)`, dove \mathbf{x} è la matrice $r \times c$ contenente c valori di r variabili diverse e il risultato è la (6.139) quando `flag=0` e la (6.140) quando `flag=1`.

La densità di probabilità di R è stata trovata nel 1915 da R.A. Fisher. Questa distribuzione (dedotta anche in [Cra51]), nel caso di N variabili gaussiane con coefficiente di correlazione vero $\rho = 0$, ha densità di probabilità data da:

$$c(r) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{1}{2}) \Gamma(\frac{\nu}{2})} (1-r^2)^{(\nu-2)/2}, \quad (6.141)$$

dove $\nu = N - 2$. Anche la stima del coefficiente di correlazione vero a partire dai dati campionari è un problema difficile, la cui brillante soluzione è dovuta sempre a R.A. Fisher, che nel 1921 dimostrò il teorema:

Teorema 6.3 (Variabile Z di Fisher). *Se R è un coefficiente di correlazione campionario ottenuto da N coppie di variabili normali, la variabile*

$$Z = \frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) \quad (6.142)$$

possiede, per $N \rightarrow \infty$, distribuzione normale con media e varianza date da:

$$\langle Z \rangle = \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right), \quad \text{Var}[Z] = \frac{1}{N-3}, \quad (6.143)$$

essendo ρ il coefficiente di correlazione vero.

Dimostrazione. La dimostrazione completa del teorema è piuttosto complicata e generalmente i testi rimandano all'articolo originale di Fisher oppure al suo celebre testo [Fis41].

Tuttavia, possiamo dare una parziale dimostrazione del teorema considerando, come in [Gal84], il caso in cui $\rho = 0$ ed applicando alla densità (6.141) la trasformazione inversa della (6.142):

$$R = \frac{e^{2Z} - 1}{e^{2Z} + 1} = \tanh Z ,$$

dove \tanh rappresenta la tangente iperbolica. Dato che:

$$1 - \tanh^2 x = \frac{1}{\cosh^2 x} , \quad \frac{d \tanh x}{dx} = \frac{1}{\cosh^2 x} ,$$

indicando con A tutti i coefficienti costanti della (6.141) e tenendo presente che $\nu = N - 2$, dalla uguaglianza $c(r) dr \equiv f(z) dz$, si ottiene:

$$f(z) = A (1 - \tanh^2 z)^{(N-4)/2} \frac{1}{\cosh^2 z} = A \frac{1}{\cosh^{N-2} z} .$$

Per il coseno iperbolico vale lo sviluppo in serie:

$$\cosh z = \frac{1}{2}(e^z + e^{-z}) = 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \dots \simeq e^{z^2/2} ,$$

dove l'ultima approssimazione vale per z dell'ordine dell'unità. A causa della presenza del logaritmo, i valori di z si mantengono abbastanza limitati, tranne nel caso estremo $r \simeq \pm 1$. Possiamo allora, con buona approssimazione, arrestare questo sviluppo in serie al secondo ordine e scrivere:

$$f(z) \simeq A e^{-(N-2)z^2/2} .$$

Risulta dunque che la variabile Z è approssimativamente normale, con media nulla e varianza $\text{Var}[Z] = 1/(N - 2)$. Il risultato corretto è tuttavia rappresentato dalle (6.143). \square

In realtà il teorema è molto potente, perché in pratica vale per $N > 10$ e dà buoni risultati anche con variabili non gaussiane.

Per mostrarvi le prodigiose proprietà della trasformazione di Fisher, abbiamo riportato in Fig. 6.18 l'istogramma di 50 000 coefficienti di correlazione campionari r ottenuti con 20 coppie di variabili uniformi $(X, X + V)$ dell'esercizio 4.2 e il relativo istogramma della variabile Z .

Nella stima della correlazione da un insieme di dati $(\{R = r\}, \{Z = z\})$ si esegue la trasformazione (6.142), si applica la teoria della stima per variabili gaussiane ed infine si antitrasforma applicando l'esponenziale ad ambo i membri della (6.142):

$$e^{2z} = \frac{1+r}{1-r} \implies r = \frac{e^{2z} - 1}{e^{2z} + 1} . \quad (6.144)$$

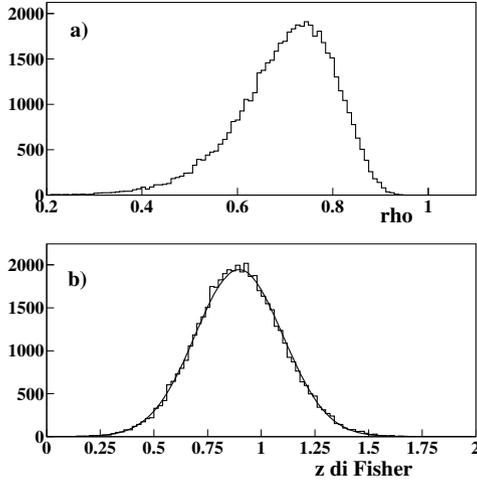


Fig. 6.18. Istogramma di 50 000 coefficienti di correlazione tra 20 coppie di variabili uniformi $(X, X+V)$ a); corrispondente istogramma della variabile Z di Fisher e della gaussiana (curva piena) che meglio si adatta ai dati b)

Una formula più semplice per l'intervallo di stima può essere ottenuta trovando l'errore su r applicando alla (6.144) la legge di propagazione degli errori e la (6.143):

$$s_r = \frac{d}{dz} \left(\frac{e^{2z} - 1}{e^{2z} + 1} \right) \sigma_z = \frac{4 e^{2z}}{(e^{2z} + 1)^2} \frac{1}{\sqrt{N-3}}$$

e sostituendo poi il valore di z della (6.142):

$$s_r = \frac{1-r^2}{\sqrt{N-3}} \quad \text{da cui} \quad \rho \in r \pm \frac{1-r^2}{\sqrt{N-3}} \quad CL \simeq 68\%. \quad (6.145)$$

La procedura è chiarita negli esempi che seguono.

Esercizio 6.16

Valutare la compatibilità tra valori veri e simulati nell'esercizio 4.2 di pagina 127.

Risposta. L'esercizio si riferisce a 10 000 coppie di dati simulati, per i quali si sono ottenuti i valori:

$$r_1 = 1.308 \cdot 10^{-2}, \quad r_2 = 0.7050, \quad r_3 = -0.7131,$$

a fronte di valori veri dati da:

$$\rho_1 = 0, \quad \rho_2 = 0.7071, \quad \rho_3 = -0.7071.$$

Trasformiamo i valori sperimentali con la (6.142) e i valori veri con la prima delle (6.143):

$$z_1 = 1.308 \cdot 10^{-2}, \quad z_2 = 0.8772, \quad z_3 = -0.8935,$$

$$\mu_1 = 0, \quad \mu_2 = 0.8814, \quad \mu_3 = -0.8814.$$

I valori z_i sono relativi ad una variabile gaussiana di media μ_i e deviazione standard:

$$\sigma = \sqrt{\frac{1}{N-3}} = \sqrt{\frac{1}{9997}} = 0.0100.$$

Possiamo allora costruire le tre variabili standard:

$$t_1 = \frac{|z_1 - \mu_1|}{\sigma} = 1.38, \quad t_2 = \frac{|z_2 - \mu_2|}{\sigma} = 0.42, \quad t_3 = \frac{|z_3 - \mu_3|}{\sigma} = 1.21,$$

alle quali, come si vede dalla Tab. D.1, corrispondono elevati livelli di significatività. I dati si scostano infatti di 1.38, 0.42 o 1.21 deviazioni standard dalla loro media, indicando un buon accordo tra simulazione e valori attesi.

Esercizio 6.17

Determinare dai dati di Tab. 6.6 il coefficiente di correlazione altezza/torace.

Risposta. La tabella, riprodotta anche in Fig. 6.19, rappresenta un istogramma bidimensionale. Dalla sua struttura si deduce facilmente l'ampiezza delle classi: ad esempio, i soldati con perimetro toracico tra 78 e 82 cm (valore centrale 80) e altezza tra 160 e 164 cm (valore centrale 162) sono in numero di 110 e così via.

Se t_i ed s_i sono i valori spettrali del torace e della altezza, gli istogrammi marginali $n(t_i)$ e $n(s_i)$ hanno forma di tipo gaussiano, come si può vedere facilmente dai grafici (fatelo per esercizio). Le medie e le deviazioni standard degli istogrammi marginali sono una stima delle corrispondenti quantità vere delle densità marginali delle altezze e dei perimetri del torace. Esse si possono calcolare tramite le (2.40, 2.41). Con ovvia notazione, si ottiene:

$$m_t = \frac{1}{1665} (72 \cdot 3 + 30 \cdot 76 + \dots) = 85.71,$$

$$s_t = \left[\frac{(72 - 85.71)^2 \cdot 3 + (76 - 85.71)^2 \cdot 30 + \dots}{1664} \right]^{1/2} = 4.46,$$

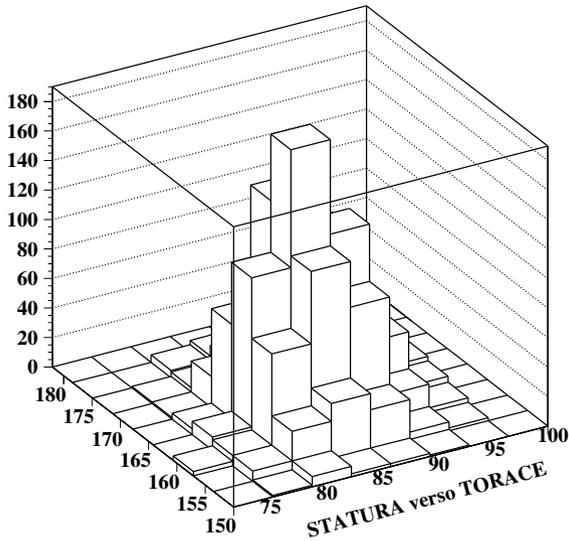
$$m_s = \frac{1}{1665} (150 \cdot 13 + 154 \cdot 109 + \dots) = 163.71,$$

$$s_s = \left[\frac{(150 - 163.71)^2 \cdot 13 + (154 - 163.71)^2 \cdot 109 + \dots}{1664} \right]^{1/2} = 5.79.$$

Le due dispersioni s_t ed s_s sono state calcolate dividendo per $(1665-1)$, poiché si sono utilizzate le medie empiriche.

Tabella 6.6. Statura e perimetro toracico (in cm) di 1665 soldati italiani della prima guerra mondiale. I dati sono riportati da [Bol62]

		<i>Perimetro toracico</i>								
		72	76	80	84	88	92	96	100	Totali
	150		1	7	2	2	1			13
S	154		7	27	39	28	6	2		109
t	158	3	7	69	118	87	21	5		310
a	162		9	110	190	126	46	9	2	492
t	166		4	68	145	114	58	12	5	406
u	170		1	22	46	69	46	12	2	198
r	174		1	4	15	35	22	6		83
a	178			5	8	11	16			40
	182				2	10	2			14
Totali		3	30	312	565	482	218	46	9	1665

**Fig. 6.19.** Istogramma bidimensionale dei dati di Tab. 6.6

La stima ad 1σ delle medie della statura e del torace è data dalla (6.46):

$$\begin{aligned}\mu_t &\in 85.71 \pm \frac{4.46}{\sqrt{1665}} \simeq 85.7 \pm 0.1 \\ \mu_s &\in 163.71 \pm \frac{5.79}{\sqrt{1665}} \simeq 163.7 \pm 0.1 .\end{aligned}$$

Veniamo ora allo studio della correlazione. I dati dovrebbero essere correlati, perché l'esperienza mostra che uomini bassi con torace enorme, o viceversa, sono molto rari. Calcoliamo pertanto la covarianza campionaria (4.23):

$$\begin{aligned}s_{st} &= \frac{1}{1664} [(72 - 85.71)(158 - 163.71) \cdot 3 + \\ &\quad (76 - 85.71)(150 - 163.71) \cdot 1 + \\ &\quad (76 - 85.71)(154 - 163.71) \cdot 7 + \dots] = 6.876 .\end{aligned}$$

Il coefficiente di correlazione campionario (6.140) vale allora:

$$r_{st} = \frac{s_{st}}{s_s s_t} = \frac{6.876}{5.79 \cdot 4.46} = 0.266 .$$

Vi ricordiamo che, per il calcolo di s_{st} o di r_{st} potete usare le routine di SCILAB `covar` e `correl`.

Stimiamo ora, con un $CL = 95\%$, il coefficiente di correlazione altezza/torace dei soldati. Per prima cosa, calcoliamo la variabile Z corrispondente alla correlazione campionaria $r = 0.266$:

$$z = \frac{1}{2} \ln \frac{1 + 0.266}{1 - 0.266} = 0.272 .$$

Questa è una variabile approssimativamente gaussiana, di deviazione standard data dalla seconda delle (6.143). Poiché i dati si riferiscono a 1665 soldati, abbiamo:

$$\sigma = \sqrt{\frac{1}{N - 3}} = \sqrt{\frac{1}{1662}} = 0.024 .$$

Per una variabile gaussiana, i limiti di confidenza al 95% sono dati da una variabile standard $t = 1.96$:

$$\mu_z \in 0.272 \pm 1.96 \cdot 0.024 = 0.272 \pm 0.047 = [0.225, 0.319] .$$

L'intervallo che abbiamo trovato contiene con $CL = 95\%$ il valore μ_z . L'intervallo di confidenza per il coefficiente di correlazione vero ρ si trova infine inserendo nella seconda delle (6.144) i valori (0.225, 0.319):

$$\rho \in [0.221, 0.309] = 0.266_{-0.045}^{+0.043} , \quad CL = 95\% .$$

Come si vede, dai dati risulta in modo certo la correlazione altezza/torace. Lo stesso risultato può essere ottenuto in questo caso anche con la formula approssimata (6.145).

Poniamoci ora la domanda: un soldato alto sui 170 cm, che torace deve avere per essere considerato nella norma? Se assumiamo l'istogramma della tabella come campione di riferimento, possiamo rispondere alla domanda stimando dai dati, con le (4.54, 4.55), la media (sulla retta di regressione) e la deviazione standard del torace t condizionate dalla statura s :

$$m(t|s) = m_t + r_{st} \frac{s_t}{s_s} (s - m_s) = 85.71 + 0.204 (170 - 163.71) \simeq 86.9 \text{ cm}$$

$$s(t|s) = [s_t^2 (1 - r_{st}^2)]^{1/2} = [19.89(1 - 0.071)]^{1/2} \simeq 4.3 \text{ cm.}$$

Sempre nell'ipotesi di un modello gaussiano, possiamo dire che il perimetro toracico del soldato, per essere nella norma, deve essere compreso, con legge 3σ , nell'intervallo 86.9 ± 4.3 cm.

Problemi

6.1. Da un'urna contenente 400 biglie rosse e nere (in proporzione sconosciuta) si estraggono 30 biglie, di cui 5 rosse. Trovare, ad un livello di confidenza del 95%, il numero iniziale di biglie rosse R contenute nell'urna nel caso di a) estrazione con reinserimento utilizzando l'approssimazione per grandi campioni, b) estrazione con reinserimento utilizzando il codice `Stiprob`, c) estrazione senza reinserimento utilizzando l'approssimazione per grandi campioni.

6.2. In un esperimento si sono ottenuti 20 conteggi. Trovare il limite superiore del valore atteso dei conteggi, con $CL = 90\%$. Utilizzare l'approssimazione gaussiana della distribuzione di Poisson.

6.3. Lo scarto quadratico medio del peso di una popolazione adulta è $\sigma = 12$ Kg. Trovare la media $\langle S \rangle$ e la deviazione standard $\sigma[S]$ per un campione di $N = 200$ individui.

6.4. Eseguendo un numero incognito N di prove, ognuna con probabilità nota a priori $p = 0.2$, si ottengono $n = 215$ successi. Stimare N con $CL=90\%$.

6.5. Se X_i ($i = 1, 2, \dots, N$) è un campione casuale estratto da una popolazione normale, dimostrare che la covarianza tra la media M e lo scarto campionario ($X_i - M$) è nulla.

6.6. Vengono sommate 25 variabili gaussiane aventi tutte la stessa media μ ignota e $\sigma = 1$, ottenendo il valore di 245. Si trovi a) l'intervallo standard di stima di μ ($CL=68\%$) e b) il limite superiore di μ con $CL=95\%$.

6.7. Due macchine producono alberi di acciaio. Le medie di due campioni di 10 alberi hanno fornito i diametri: $\mu_1 \in 5.36 \pm 0.05$ e $\mu_2 \in 5.21 \pm 0.05$ mm. Applicare il test di Student per verificare l'omogeneità dei pezzi prodotti dalle due macchine.

6.8. Ad un gruppo di 70 ammalati è stato somministrato un farmaco, ad un altro gruppo di 58 ammalati del semplice zucchero (placebo). Giudicare se il farmaco è efficace in base alla tabella di contingenza ottenuta:

	<i>Medicina</i>	<i>Placebo</i>	<i>Totale</i>
guariti	40	28	68
ammalati	30	30	60
totale	70	58	128

6.9. Il conteggio X del numero di autobus arrivati ad un casello autostradale in 5 minuti, ripetuto per $N = 100$ prove, ha dato il seguente istogramma, con lo spettro discreto di X diviso in 11 canali (da 0 a 10):

n. di autobus x_i	0	1	2	3	4	5	6	7	8	9	10
n. di prove n_i	1	5	6	19	20	17	15	8	8	1	0

Dire, con un test a due code, se i dati sono compatibili con un processo poissoniano.

6.10. Secondo un modello genetico un certo tipo di pianta dovrebbe fornire in rapporto tra piselli verdi e gialli di 3:1. In un campione di 500 piselli se ne sono trovati 356 verdi. Dire se il modello può essere accettato ad un livello del 5%.

6.11. Una serie di 600 lanci con due dadi diversi ha dato luogo alla tabella di contingenza relativa alle 6 facce:

	1	2	3	4	5	6	<i>Totale</i>
DADO 1	101	105	103	95	90	106	600
DADO 2	99	105	98	103	96	99	600
TOTALE	200	210	201	198	186	205	1200

Dire se a) le due serie di lanci sono tra loro compatibili e se b) i dadi e i lanci sono regolari.

6.12. Un campione di 40 coppie di dati gaussiani ha coefficiente di correlazione $r = 0.15$. Fare la stima per intervallo di ρ con $CL \simeq 68\%$.

6.13. Dopo l'introduzione del nuovo codice della strada le morti per incidente nel fine settimana sono diminuite da 60 a 33. Qual'è la probabilità di sbagliare affermando che il calo è merito del nuovo codice?

6.14. Un conteggio poissoniano ha fatto registrare il valore $x = 167$ in 10 secondi. Si sa che la media dei conteggi X è funzione dell'efficienza di conteggio ϵ sul segnale e del fondo b secondo la relazione $\mu = \epsilon\nu + b$, ovvero:

$$X \sim \text{Poisson}(\epsilon\nu + b) .$$

L'efficienza ϵ è stata determinata come $\epsilon \pm \sigma_\epsilon = 0.90 \pm 0.10$, mentre il fondo è stimato in base ad un valore di $b = 530$ conteggi in 100 secondi. Stimare il valore della frequenza della sorgente ν con errore.

6.15. Un insieme di contatori di raggi cosmici ha rilevato in uno stesso intervallo di tempo i seguenti conteggi:

Contatore	1	2	3	4	5	6
Conteggi	29	19	18	25	17	10

Valutare se il flusso sui contatori è omogeneo al livello del test del 5%.

6.16. Due sorgenti A e B forniscono il seguente risultato:

$$A = 240 \text{ conteggi in } 10 \text{ secondi}$$

$$B = 670 \text{ conteggi in } 10 \text{ secondi .}$$

Uno sperimentatore afferma di aver ottenuto, in una nuova misura con le sorgenti A e B insieme, il risultato

$$C = 10500 \text{ conteggi in } 100 \text{ secondi.}$$

Valutare la correttezza di questa affermazione.

6.17. I seguenti 47 intertempi si riferiscono agli intervalli di funzionamento osservati tra guasti consecutivi di un dispositivo elettronico.

intervallo	$0 - 120h$	$120 - 240h$	$240 - 360h$	$360 - 480h$
frequenza	22	12	7	6

Dire se i dati sono in accordo con l'ipotesi di una legge esponenziale con $\lambda = 0.005 \text{ h}^{-1}$ al livello dell'1%.

6.18. L'osservazione dei tempi di arrivo di eventi nel corso di un esperimento di conteggio, durato un tempo prefissato, ha dato il seguente risultato.

intervallo	$[0 - 1] \text{ s}$	$[1 - 2] \text{ s}$	$[2 - 4] \text{ s}$	$s > 4 \text{ s}$
prove	370	268	219	151

Verificare se i dati sono un accordo con una frequenza pari a 0.5 s^{-1} .

6.19. Quattro misure di una grandezza fisica forniscono i valori

$$1.12 \quad 1.13 \quad 1.10 \quad 1.09 .$$

Se esse provengono da popolazioni normali con eguale errore di misura, dato dalla varianza $\sigma^2 = 4 \cdot 10^{-4}$, verificare se le misure sono compatibili.

6.20. La legge pone un limite alla concentrazione di un certo inquinante nell'aria a 55 parti per milione (ppm). Una serie di 10 misure ripetute ha fornito un valore medio di 58 ppm. Sapendo che l'incertezza di una singola misura (deviazione standard) è ± 10 ppm, dire se il dato supera i limiti consentiti ad un livello del test del 5%.

6.21. Un esperimento fornisce il seguente risultato:

13 valori < -1 , 25 valori in $[-1, 0]$, 44 valori in $(0, 1]$, 16 valori > 1 .

Giudicare, attraverso il livello di significatività, se i dati sono compatibili con la distribuzione normale standard $N(0, 1)$.

6.22. Calcolare il numero N di lampadine necessario per essere sicuri al 95% di avere una illuminazione di 1000 ore, sapendo che ogni lampadina ha una probabilità di vita di tipo esponenziale negativo con media $\tau = 100$ ore.

Assumere di usare una lampadina finché non si brucia e poi di accendere la successiva, fino ad arrivare a 1000 ore.

Assumere anche la distribuzione gaussiana per la variabile somma dei tempi di vita delle lampadine.

6.23. In sei esemplari di un cavo per alta tensione lungo 100 metri sono stati rilevati rispettivamente

18, 14, 10, 10, 21, 17

punti con difetti di isolamento.

Valutare se questi dati possono essere considerati in accordo con uno standard di meno di 15 difetti ogni 100 metri.

6.24. Il fondo atteso di un esperimento di conteggio (misurato accuratamente in fase di calibrazione) è di 10 conteggi/s. In una prova (fondo più eventuale segnale) si registrano in un secondo 25 conteggi. Utilizzando l'approssimazione gaussiana, trovare il limite superiore dei conteggi per il solo segnale con $CL = 95\%$.

Il metodo Monte Carlo

Le proprietà davvero uniche del cervello umano mi sembrano caratterizzate proprio dal possente sviluppo e dall'impiego intensivo della funzione di simulazione. E ciò al livello più profondo delle funzioni conoscitive, quello su cui si basa il linguaggio e che questo, senza dubbio, esprime solo in parte.

Jacques Monod, *Il caso e la necessità*

7.1 Introduzione

Con il termine di “metodo Monte Carlo” oppure “metodo MC”, vengono in generale denominate tutte quelle tecniche che fanno uso di variabili aleatorie artificiali (cioè generate al calcolatore) per la risoluzione di problemi matematici.

Indubbiamente questo non è un modo molto efficiente per trovare la soluzione di un problema, in quanto la procedura di campionamento simulato porta ad un risultato che è sempre affetto dall'errore statistico. Nella pratica però ci si trova spesso di fronte a situazioni in cui è troppo difficile, se non addirittura impossibile, utilizzare i tradizionali procedimenti numerici o analitici ed in tutti questi casi il metodo Monte Carlo diventa l'unica alternativa disponibile.

L'applicazione di questo metodo non è ristretta solo ai problemi di natura statistica, come potrebbe forse sembrare dall'utilizzo delle distribuzioni di probabilità, ma include tutti quei casi in cui si riesce a trovare un collegamento tra il problema in esame ed il comportamento di un certo sistema aleatorio: il valore di un integrale definito, che non è certamente una grandezza casuale, può, ad esempio, essere calcolato anche usando dei numeri casuali.

Le basi teoriche del metodo Monte Carlo (o metodo di simulazione) sono conosciute da molto tempo ed il primo esempio dell'impiego di numeri casuali per la risoluzione di integrali definiti si trova addirittura in un libro (“*Essai d'arithmétique morale*”) scritto nel 1777 da Georges-Louis Leclerc, conte di Buffon, matematico e naturalista francese, in cui è delineato un metodo per il calcolo approssimato del valore di π (si veda il problema 7.13).

Per oltre un secolo e mezzo esso fu però usato solo sporadicamente e soprattutto per fini didattici. La sua prima applicazione sistematica si è avuta solo nella prima metà degli anni '40 a Los Alamos, ad opera dell'équipe di scienziati, guidata da Enrico Fermi, che sviluppò il progetto delle prime bombe atomiche.

In questo stesso periodo nasce anche il termine “Monte Carlo”, che fa ovviamente riferimento alla città del Principato di Monaco famosa per il suo casinò e, più precisamente, alla roulette, uno dei mezzi meccanici più semplici per generare delle variabili aleatorie. La paternità del nome va in particolare attribuita ai matematici J. Von Neumann e S. Ulam, che lo adottarono come nome in codice delle loro ricerche segrete, condotte utilizzando numeri casuali, sui processi di diffusione ed assorbimento dei neutroni nei materiali fissili (per maggiori notizie storiche, potete consultare [HH64]).

Dopo gli anni '50 si è avuto un decisivo impulso nell'impiego del metodo con l'utilizzo dei calcolatori, grazie ai quali esso è passato in pochi anni dal ruolo di curiosità matematica a quello di strumento indispensabile per la ricerca scientifica. Questo è avvenuto non solo perché i calcolatori permettono la rapida esecuzione dei lunghi calcoli che sono spesso necessari per ottenere un risultato significativo, ma anche, come vedremo in seguito, per la facilità con cui essi possono generare dei numeri casuali. Attualmente si hanno applicazioni nei più svariati campi di ricerca, dalla fisica nucleare alla chimica, dalla meccanica statistica e quantistica all'economia.

In questo capitolo daremo una descrizione dei principi fondamentali del metodo e dei dettagli tecnici più importanti necessari per realizzare codici Monte Carlo per la risoluzione di problemi di tipo statistico. Altre applicazioni significative verranno illustrate più avanti, nel cap. 8.

7.2 Cos'è il metodo Monte Carlo?

Come già abbiamo accennato, in tutte le librerie di sistema dei calcolatori esistono da molto tempo routine di servizio che generano valori di variabili che possiamo considerare aleatorie, con densità uniforme in $[0, 1]$. La presenza di queste routine è dovuta al fatto che, in base al teorema 3.5 ed a quanto spiegheremo dettagliatamente tra poco, attraverso di esse è possibile generare variabili aleatorie aventi una qualsiasi altra densità di probabilità, discreta o continua.

La realizzazione di questo tipo di routine, che apparentemente può sembrare semplice, costituisce in realtà un problema matematico complesso, come si può vedere consultando i riferimenti [BS91, Cha75, Com91, FLJW92, Jam90, MNZ90, PFTW92]. La libreria di SCILAB mette a disposizione la routine `rand`, particolarmente semplice, e la routine `grand`, più sofisticata, che utilizzeremo nel seguito. La chiamata a `grand` è del tipo:

```
x = grand(m,n,tipo,p1,p2);
```

dove x è una matrice di numeri aleatori di dimensione $m \times n$, “tipo” è il tipo di distribuzione da cui sono estratti i numeri e $p1, p2, \dots$ sono i parametri della distribuzione. Ad esempio, per estrarre un numero casuale uniforme nell'intervallo $[0, 1)$ basta scrivere:

```
x = grand(1,1,unf,0,1);
```

e per generare un vettore di 2000 numeri e farne l'istogramma evidenziandone i parametri statistici bastano le due istruzioni:

```
x = grand(1,2000,unf,0,1);
clf();Histplote(20,x,errors=1,stat=1)
```

in cui abbiamo fatto uso della nostra routine `Histplote`. Le numerose possibilità di `grand` sono descritte in dettaglio nella guida in linea di SCILAB.

Le routine che generano numeri casuali vanno inizializzate con uno o più numeri interi detti semi (*seeds*). Se i semi non vengono cambiati, ad ogni nuova prova del programma i risultati (pur casuali), saranno sempre gli stessi. In altre parole, con gli stessi semi si genererà *sempre lo stesso campione*. Nel caso di `grand` l'istruzione da collocare all'inizio del programma per avere sempre una certa sequenza casuale è la seguente:

```
grand(setsd,123432);
```

dove il seme 123432 è arbitrario e va cambiato ogni volta che si desidera ottenere sequenze diverse. Omettendo questa istruzione, la routine `grand` utilizza il seme corrente di SCILAB che varia ad ogni chiamata, garantendo la variabilità statistica del risultato.

Dobbiamo a questo punto mettervi subito in guardia sulla affidabilità delle routine di attualmente in circolazione, diverse delle quali non superano molti dei test casuali standard. Per procedere a simulazioni corrette è allora necessario utilizzare routine ben documentate e di provata affidabilità, come quelle presenti in SCILAB e nelle librerie scientifiche più collaudate.

Da ora in poi:

- con il simbolo ξ (con o senza indici) denoteremo sempre una variabile aleatoria uniforme nell'intervallo $[0, 1]$;
- con ξ (con o senza indici) denoteremo i valori assunti dalla variabile $\xi \sim U(0, 1)$ in una o più prove specifiche. In pratica questi sono i valori numerici forniti in uscita dalla routine `grand`.

Per simulare l'esperimento più semplice, il lancio della moneta, basta allora dividere a metà l'intervallo unitario e, per ogni numero generato ξ , definire l'evento come “testa” quando $0 \leq \xi \leq 0.5$ e come “croce” quando $0.5 < \xi \leq 1$ (naturalmente anche la convenzione inversa andrebbe ugualmente bene). Le istruzioni SCILAB sono le seguenti:

```

// Routine MCmoneta per la simulazione del lancio di una moneta
// Lo spettro e' nel vettore punti (11 posizioni)
// le frequenze simulate sono nel vettore monete (11 posizioni)

punti = [0 1 2 3 4 5 6 7 8 9 10];
monete = [0 0 0 0 0 0 0 0 0 0 0];
// scommentare se si desidera sempre lo stesso campione
// grand("setsd",137778); // cambiare 137778 per una frequenza
// diversa
for k=1:100,
    xi = grand(1,10,"unf",0,1); // lancio delle 10 monete
    teste=0;
    for j=1:10,
        if(xi(j) < 0.5) then teste=teste+1, end; // conta le teste
    end;
    monete(teste+1) = monete(teste+1) + 1; // frequenza teste
end;
clf(); Histfreque(monete,punti,stat=1,errors=1); // display
end;

```

È facile riconoscere che la routine opera secondo i seguenti passi:

- genera 10 numeri casuali $\xi_1, \xi_2, \dots, \xi_{10}$;
- conta il numero x di volte in cui $\xi_i \leq 0.5$ ($i = 1, 2, \dots, 10$);
- ripete i due passi precedenti per $N = 100$ volte, mettendo in istogramma i valori di x via via trovati;

Alla fine del ciclo, se denotiamo con n_i il contenuto di un generico canale dell'istogramma, il rapporto n_i/N ci dà, per ogni canale, la stima della probabilità cercata. L'istogramma che si ottiene è riportato in Fig. 7.1.

Occorre a questo punto notare quanto segue:

- mediante i passi 1) e 2), con cui simuliamo il lancio di 10 monete, non facciamo altro che generare una variabile aleatoria dalla distribuzione binomiale (2.30) e, per N tendente all'infinito, l'istogramma di frequenza che otteniamo tende quindi ad essere proprio $b(x; 10, 0.5)$. Potete facilmente verificare questo fatto simulando un numero molto grande di lanci, come $N \geq 20\,000$;
- poiché, per ovvie ragioni, possiamo ripetere questo algoritmo solo per un numero finito di volte, la simulazione ha come risultato delle frequenze affette da un certo errore, che è poi lo stesso che otterremmo se cercassimo di trovare x eseguendo dei veri lanci. Se denotiamo con f_i il rapporto n_i/N , la migliore stima che siamo in grado di dare per $b(x_i; 10, 0.5)$ risulta:

$$b(x_i; 10, 0.5) \simeq f_i \pm \sqrt{\frac{f_i(1-f_i)}{N}}. \quad (7.1)$$

Naturalmente la precisione della nostra stima aumenta all'aumentare di N ma il miglioramento che otteniamo non è molto elevato: quadruplicando il

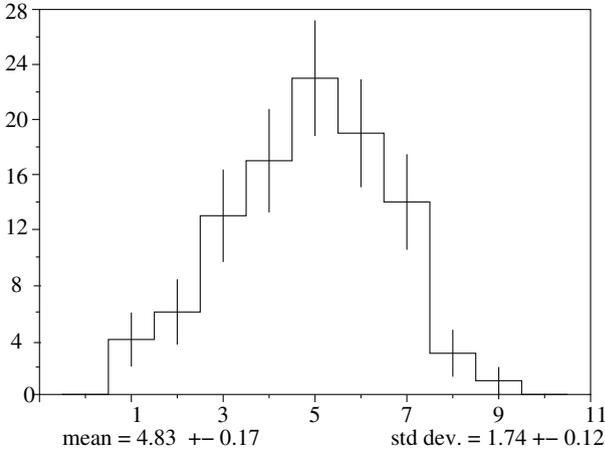


Fig. 7.1. Numero di teste ottenuto in 100 lanci simulati di 10 monete. Si confronti il risultato con quello dell'esperimento reale di Fig. 6.17 a pagina 242

numero di simulazioni, per esempio, dimezziamo solamente l'errore e, come dimostreremo tra poco, tale andamento non è legato a questo particolare esempio ma è una caratteristica generale di tutti i calcoli MC.

Nel prossimo paragrafo daremo una giustificazione teorica del metodo MC, riscrivendo in maniera più generale e matematicamente corretta le indicazioni, piuttosto qualitative, che abbiamo finora sviluppato.

7.3 Fondamenti matematici

Consideriamo una variabile T associata ad un qualsiasi fenomeno aleatorio come funzione di k variabili casuali (X_1, X_2, \dots, X_k) :

$$T = f(X_1, X_2, \dots, X_k) . \quad (7.2)$$

Come abbiamo spiegato nei capitoli precedenti, l'intero processo è in ultima analisi caratterizzato dai valori medi e dai parametri di dispersione di T ; questi ultimi sono a loro volta esprimibili, in base alla (2.66), come differenza di valori medi. Il fenomeno in esame può quindi sempre essere studiato sulla base di valori attesi (2.67), tramite la risoluzione di uno o più integrali (o somme) del tipo:

$$I = \langle T \rangle = \int_D f(x_1, x_2, \dots, x_k) p(x_1, x_2, \dots, x_k) dx_1 \dots dx_k , \quad (7.3)$$

dove $p(x_1, x_2, \dots, x_k)$ è la densità di probabilità normalizzata delle variabili (X_1, X_2, \dots, X_k) , definite in $D \in \mathbb{R}^k$,

$$\int_D p(x_1, x_2, \dots, x_k) dx_1 \dots dx_k = 1 . \quad (7.4)$$

Da un punto di vista strettamente formale, tutti i calcoli MC che sono simulazioni di processi stocastici, in cui si generano casualmente degli eventi secondo certe distribuzioni di probabilità, sono equivalenti all'approssimazione del valore di un integrale definito o di una somma.

Se simuliamo N prove della variabile aleatoria T utilizzando diversi insiemi indipendenti di numeri casuali *estratti dalla densità* p , per le proprietà della media di un campione (vedere parr. 2.11 e 6.9), sappiamo che la quantità:

$$T_N = \frac{1}{N} \sum_{i=1}^N f(x_{1i}, x_{2i}, \dots, x_{ki}) , \quad (7.5)$$

è una stima consistente e corretta di I .

Se supponiamo che la distribuzione della variabile aleatoria T , come è quasi sempre verificato in pratica, abbia una varianza finita σ_T^2 , applicando la disuguaglianza di Tchebycheff (3.92) e la formula (6.45) della varianza della media, otteniamo la relazione:

$$P \left\{ |T_N - I| \leq \frac{\sigma_T}{\sqrt{N}} \epsilon \right\} \geq 1 - \frac{1}{\epsilon^2} \quad \epsilon \geq 1 . \quad (7.6)$$

Un'altra proprietà molto utile (sempre assumendo σ_T^2 finita) è quella fornita dal teorema Limite Centrale 3.1, secondo il quale la densità di probabilità di T_N tende "asintoticamente" ad essere una gaussiana con media I e deviazione standard σ_T/\sqrt{N} . Come abbiamo già notato, il requisito di asintoticità viene in realtà soddisfatto per N abbastanza basso ($N \geq 10$) per cui possiamo quasi sempre scrivere che:

$$P \left\{ |T_N - I| \leq \frac{3\sigma_T}{\sqrt{N}} \right\} \simeq 0.997 . \quad (7.7)$$

Le (7.6, 7.7) mostrano che i valori simulati convergono in probabilità verso la quantità da stimare I come σ_T/\sqrt{N} . Risulta pertanto che (salvo i casi in cui, come accenneremo più avanti, si riesce a "manipolare" T riducendone la varianza) l'unico modo per aumentare la precisione della stima T_N è quello di aumentare il numero N di eventi simulati. Questa lenta convergenza degli stimatori MC può richiedere un notevole tempo di calcolo nella simulazione di sistemi complessi.

7.4 Generazione di variabili aleatorie discrete

Se la funzione di densità della variabile discreta che si vuole generare è nota, allora, tenendo presente che:

- è possibile costruire, con la (2.28), la funzione cumulativa;
- valgono le (3.87, 3.89) di pagina 102;
- si dispone della routine `grand`,

è immediato rendersi conto che il metodo più efficiente, per simulare variabili discrete, consiste nell'estrarre un numero casuale $0 \leq \xi \leq 1$, considerarlo come variabile cumulativa e determinare il valore quantile corrispondente al valore cumulativo estratto.

L'uso della cumulativa è il metodo di base per tutte le simulazioni MC.

Consideriamo un segmento di lunghezza unitaria, suddividiamolo poi in k intervalli ed assegniamo ad ognuno di essi una lunghezza p_i pari alla probabilità che si verifichi il corrispondente evento $\{X = x_i\}$ (Fig. 7.2). Poiché la probabilità che un numero casuale uniforme $0 \leq \xi \leq 1$ cada in un particolare intervallo è esattamente pari alla lunghezza dell'intervallo:

$$P\{0 \leq \xi \leq p_1\} = p_1, \tag{7.8}$$

$$P\{p_1 < \xi \leq p_1 + p_2\} = p_2, \tag{7.9}$$

.....,

$$P\{p_1 + p_2 + \dots + p_{k-1} < \xi \leq 1\} = p_k, \tag{7.10}$$

si dovrà considerare estratto il valore x_j corrispondente all'estremo superiore $p_1 + p_2 + \dots + p_j$ dell'intervallo entro cui è contenuto il numero ξ generato con la routine `grand`. Tale procedimento può essere così schematizzato:

Algoritmo 7.1 (Generazione di variabili discrete). Per generare una variabile casuale discreta X , che può assumere un insieme finito di valori spettrali x_1, x_2, \dots, x_k con probabilità p_1, p_2, \dots, p_k , occorre:

- costruire la funzione cumulativa F_j :

$$F_j = \sum_{i=1}^j p_i \quad (j = 1, 2, \dots, k); \tag{7.11}$$

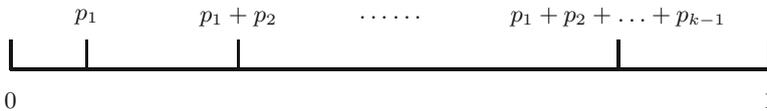


Fig. 7.2. Generazione di variabili casuali discrete. L'intervallo $[0, 1]$ viene suddiviso in k segmenti di lunghezza p_1, p_2, \dots, p_k ed il sottointervallo entro cui cade un numero casuale ξ individua l'evento generato

- generare un numero $0 \leq \xi \leq 1$;
- trovare l'indice j ($1 \leq j \leq k$) per cui è verificata la disuguaglianza:

$$F_{j-1} < \xi \leq F_j \quad (7.12)$$

(quando $j = 1$, definiamo $F_{j-1} = 0$);

- porre $\{X = x_j\}$.

Esercizio 7.1

Scrivere un codice per simulare il lancio di una coppia di dadi (non truccati).

Risposta. Dobbiamo, innanzitutto, trovare la probabilità puntuale e quella cumulativa F associate ad ogni punteggio. Se con S denotiamo la somma dei punti che si ottiene in un singolo lancio, queste probabilità sono rappresentate dalla Tab. 7.1.

Si possono usare due differenti metodi per determinare in quale intervallo del vettore F si collocano i numeri casuali ξ :

- ricerca sequenziale: si esaminano in serie, uno dopo l'altro, tutti i sottointervalli iniziando dal primo (o dall'ultimo);
- ricerca binaria (o dicotomica): si inizia l'esame dal sottointervallo corrispondente all'indice centrale del vettore; se quello cercato è maggiore (minore) si elimina la prima (seconda) metà del vettore F , ci si riposiziona al centro della metà rimasta e si ripetono le operazioni precedenti fino a trovare l'indice j che soddisfa la relazione (7.12).

Tabella 7.1. Distribuzione di probabilità p_i ($i = 1, \dots, 11$) e funzione cumulativa F_i ($F_i = \sum_{j=1}^i p_j$) del punteggio S_i che si ottiene lanciando una coppia di dadi

i	S_i	p_i	$F_i = \sum_{j=1}^i p_j$
1	2	1/36	1/36
2	3	2/36	3/36
3	4	3/36	6/36
4	5	4/36	10/36
5	6	5/36	15/36
6	7	6/36	21/36
7	8	5/36	26/36
8	9	4/36	30/36
9	10	3/36	33/36
10	11	2/36	35/36
11	12	1/36	1

È facile dimostrare che in quest'ultimo procedimento il numero di confronti necessari per individuare questo indice è pari al minimo intero m che verifica la disuguaglianza $2^m \geq k$ (se $k = 1000$, ad esempio, $m = 10$). Per questo motivo esso è senz'altro più vantaggioso, in termini di tempo di calcolo, rispetto a quello sequenziale, che necessita di un numero medio di confronti pari a $k/2$.

La routine SCILAB per la ricerca dicotomica `dsearch(x,val)` data una cumulativa discreta `val` trova l'indice corrispondente al numero dell'intervallo in cui cade un valore `x`. Le numerose possibilità di questa routine sono descritte nella guida in linea. Altri algoritmi di ricerca dicotomica sono descritti in [KR89, PFTW92].

Il codice SCILAB `Dadi` per la simulazione del lancio di due dadi è allora il seguente:

```
// Routine Dadi per la simulazione del lancio di due dadi
// Lo spettro e' nel vettore punti
// le frequenze simulate sono nel vettore lanci

cum = [0. 1/36 3/36 6/36 10/36 15/36 21/36 26/36 30/36 33/36
                                             35/36 1.];

punti = [2 3 4 5 6 7 8 9 10 11 12];
lanci = [0 0 0 0 0 0 0 0 0 0 0];

//scommentare se si vuole sempre lo stesso campione
//grand("setsd",678543); // cambiare 678543 per una frequenza
diversa

for k=1:20000,
    xi = grand(1,1,"unf",0,1);
    ind = dsearch(xi,cum); // ricerca dicotomica nella
                           cumulativa
    lanci(ind)=lanci(ind)+1; // frequenza simulata
end;
clf(); Histfreque(lanci,punti,stat=1,errors=1); // display
end;
```

La nostra routine `Histfreque` produce il risultato di Fig. 7.3, ottenuta simulando 20 000 lanci. Dall'istogramma, attraverso la relazione (7.7), possiamo stimare il valor medio $\langle S \rangle$ del punteggio che si ottiene lanciando una coppia di dadi come:

$$\langle S \rangle = 7.00 \pm 0.03, \quad CL \simeq 99.7\% .$$

Esercizio 7.2

Simulare il lancio della moneta col metodo della cumulativa.

Risposta. Consideriamo l'ultima colonna della Tab. 2.3 di pagina 52, che contiene proprio i valori cumulativi (7.11) relativi al numero di successi nel lancio

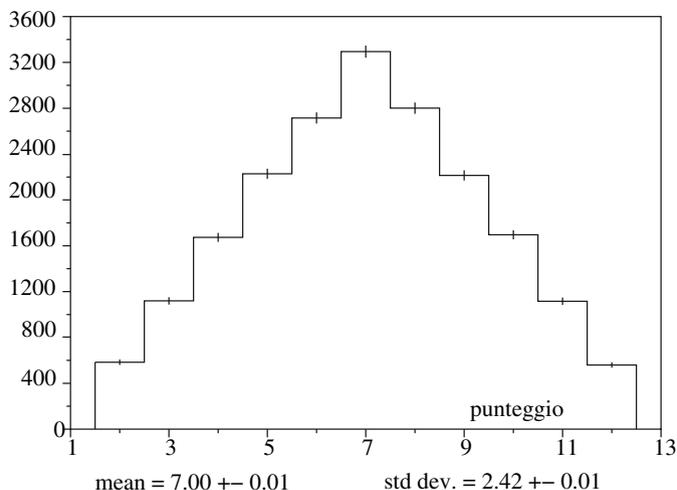


Fig. 7.3. Istogramma di 20 000 lanci simulati di una coppia di dadi

di 10 monete. Se generiamo una variabile ξ il segmento in cui essa è compreso individua il numero di teste ottenute nella prova; ad esempio, il numero $\xi = 0.35$ cade all'interno del quinto segmento:

$$p_0 + p_1 + p_2 + p_3 < \xi < p_0 + p_1 + p_2 + p_3 + p_4 \quad (7.13)$$

e corrisponde perciò al valore $x = x_5 = 4$ teste.

Vi lasciamo come esercizio scrivere il facile codice di simulazione, che è molto più efficiente di quello visto nel par. 7.2.

7.5 Generazione di variabili aleatorie continue

Nel caso delle variabili continue basta sfruttare ancora il teorema 3.5 per arrivare immediatamente allo stesso tipo di procedimento che abbiamo usato per le variabili discrete.

Consideriamo una generica densità di probabilità $p(x)$, avente cumulativa $F(x)$ e definita su un intervallo qualsiasi $[a, b]$. Se generiamo a caso $\{\xi = \xi\}$ e calcoliamo il corrispondente valore:

$$x = F^{-1}(\xi) , \quad (7.14)$$

otteniamo, in base alla relazione (3.87), una generazione casuale di X secondo la densità $p(x)$.

Per generare variabili aleatorie di densità qualsiasi basta quindi risolvere un integrale ed invertire la funzione risultante mediante un algoritmo che possiamo così enunciare:

Algoritmo 7.2 (Trasformazione inversa). Per generare una variabile casuale continua X , di densità $p(x)$ e definita nell'intervallo $[a, b]$, occorre:

- generare un numero $0 \leq \xi \leq 1$;
- risolvere, rispetto a x , l'equazione:

$$\int_a^x p(t) dt = \xi . \quad (7.15)$$

Iniziate a prendere confidenza con questo procedimento, di fondamentale importanza per il metodo MC, seguendo attentamente gli esercizi che seguono. Vi consigliamo anche di riguardare l'esercizio 3.12 di pag. 103.

Esercizio 7.3

Generare una variabile X uniformemente distribuita nell'intervallo $[a, b]$.

Risposta. In questo caso dalla (3.79) si ha:

$$p(x) = \frac{1}{(b-a)} \quad (7.16)$$

e l'equazione (7.15) diventa:

$$\int_a^x \frac{dx}{(b-a)} = \xi , \quad (7.17)$$

da cui il risultato:

$$x = a + \xi(b-a) . \quad (7.18)$$

Esercizio 7.4

Generare casualmente, in un cerchio di raggio R , punti distribuiti in maniera uniforme con densità ρ (punti/cm²) costante.

Risposta. È conveniente individuare un punto P qualsiasi all'interno di un cerchio mediante la coppia di coordinate polari r e φ (con $0 \leq r \leq R$ e $0 \leq \varphi \leq 2\pi$).

Per trovare le densità di probabilità $p(\varphi)$ e $q(r)$ di queste variabili, osserviamo che $p(\varphi) d\varphi$ è data dal rapporto tra il numero di punti contenuto nel settore infinitesimo tratteggiato in Fig. 7.4a) ed il numero totale di punti contenuti nella superficie del cerchio:

$$p(\varphi) d\varphi = \frac{\rho R^2 d\varphi/2}{\rho\pi R^2} = \frac{d\varphi}{2\pi} . \quad (7.19)$$

Analogamente, per $q(r)$ otteniamo (vedere Fig. 7.4b):

$$q(r) dr = \frac{\rho 2\pi r dr}{\rho\pi R^2} = \frac{2r}{R^2} dr . \quad (7.20)$$

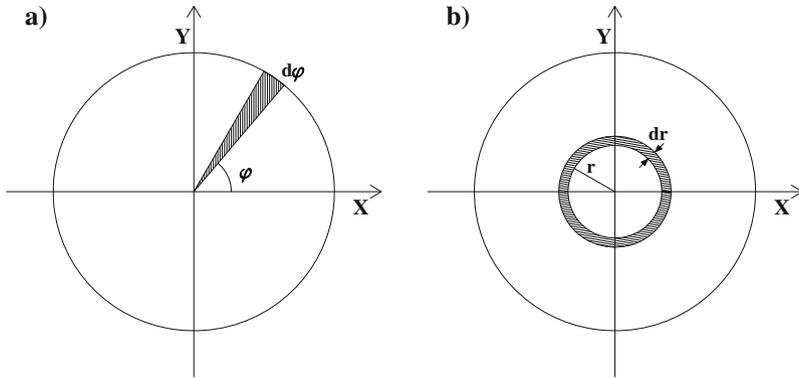


Fig. 7.4. a) Nel caso di punti uniformemente distribuiti nel cerchio, $p(\varphi) d\varphi$ è pari al rapporto tra l'area infinitesima tratteggiata ed il numero totale di punti sulla superficie circolare. b) Come nella figura precedente, ma per $q(r) dr$

Le corrispondenti cumulative risultano allora:

$$\xi_1 = P(\varphi) = \int_0^\varphi p(\varphi) d\varphi = \frac{\varphi}{2\pi} , \quad (7.21)$$

$$\xi_2 = Q(r) = \int_0^r q(r) dr = \frac{r^2}{R^2} . \quad (7.22)$$

Applicando di nuovo l'algoritmo 7.2 otteniamo:

$$\begin{cases} \varphi = 2\pi\xi_1 \\ r = R\sqrt{\xi_2} . \end{cases} \quad (7.23)$$

dove ξ_1 e ξ_2 sono due numeri casuali indipendenti.

Esercizio 7.5

Generare in modo isotropo (con densità costante) punti distribuiti su una superficie sferica di raggio R .

Risposta. La posizione di un generico punto P posto su una superficie sferica (che supponiamo, senza perdere in generalità, essere al centro di un sistema di assi cartesiani) viene di solito definita (come in Fig. 7.5a) mediante la terna di coordinate (R, φ, ϑ) con:

$$\begin{cases} 0 \leq \varphi \leq 2\pi & (\varphi = \text{angolo azimutale}) \\ 0 \leq \vartheta \leq \pi & (\vartheta = \text{angolo polare}) . \end{cases} \quad (7.24)$$

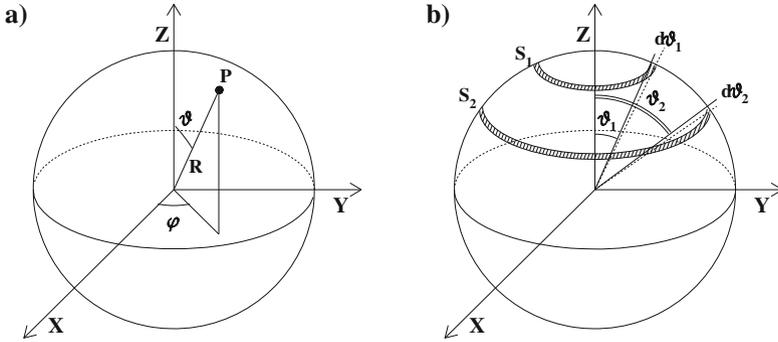


Fig. 7.5. a) Il generico punto P sulla superficie sferica viene individuato mediante la terna di coordinate sferiche (R, ϑ, φ) . b) Per ottenere la stessa densità di punti casuali sulle superfici S_1 ed S_2 , non si deve generare uniformemente secondo ϑ , ma secondo $(1 - \cos \vartheta)$ oppure $\cos \vartheta$

Le formule che permettono di passare nel sistema di coordinate ortogonali XYZ sono:

$$\begin{cases} x = R \operatorname{sen} \vartheta \cos \varphi \\ y = R \operatorname{sen} \vartheta \operatorname{sen} \varphi \\ z = R \cos \vartheta . \end{cases} \quad (7.25)$$

Poiché R rimane costante lungo la superficie della sfera, dobbiamo determinare solo le funzioni di generazione di ϑ e φ . Consideriamo una superficie infinitesima $d\Omega$ attorno a P ; la sua area è:

$$d\Omega = \operatorname{sen} \vartheta \, d\vartheta \, d\varphi . \quad (7.26)$$

Se ora denotiamo con n_{tot} e dn il numero totale dei punti sulla sfera e su $d\Omega$, per la condizione di isotropia dobbiamo porre che la densità dei punti sulla superficie sferica sia costante:

$$\frac{n_{tot}}{4\pi} = \frac{dn}{d\Omega} . \quad (7.27)$$

La probabilità $p(\Omega) \, d\Omega$ per un punto qualsiasi sulla superficie sferica di trovarsi entro $d\Omega$ risulta quindi:

$$p(\Omega) \, d\Omega = \frac{dn}{n_{tot}} = \frac{d\Omega}{4\pi} = \frac{\operatorname{sen} \vartheta \, d\vartheta \, d\varphi}{4\pi} \quad (7.28)$$

e le probabilità $p(\varphi) \, d\varphi$ e $q(\vartheta) \, d\vartheta$ che un punto si trovi nell'intervallo $[\varphi, \varphi + d\varphi]$ qualunque sia ϑ e nell'intervallo $[\vartheta, \vartheta + d\vartheta]$ qualunque sia φ , sono date

dalle rispettive densità marginali, definite nella (4.11):

$$p(\varphi) d\varphi = \frac{1}{4\pi} d\varphi \int_0^\pi \sin \vartheta d\vartheta = \frac{1}{2\pi} d\varphi, \quad (7.29)$$

$$q(\vartheta) d\vartheta = \frac{1}{4\pi} \sin \vartheta d\vartheta \int_0^{2\pi} d\varphi = \frac{\sin \vartheta}{2} d\vartheta. \quad (7.30)$$

Le cumulative corrispondenti a queste densità sono:

$$\xi_1 = P(\varphi) = \frac{\varphi}{2\pi}, \quad (7.31)$$

$$\xi_2 = Q(\vartheta) = \frac{1 - \cos \vartheta}{2} \quad (7.32)$$

e le formule per la generazione casuale di φ e ϑ risultano infine:

$$\begin{cases} \varphi = 2\pi\xi_1 \\ \vartheta = \arccos(1 - 2\xi_2) \end{cases} \quad (7.33)$$

Ricordatevi quindi la seguente considerazione non proprio intuitiva: isotropia sulla superficie sferica significa uniformità in φ ma non in ϑ . Dalle equazioni precedenti ricaviamo infatti che la condizione di isotropia è soddisfatta quando $\xi = (1 - \cos \vartheta)$ (oppure $\xi = \cos \vartheta$), e non ϑ , è una variabile uniformemente distribuita. Questa proprietà può essere compresa qualitativamente da un punto di vista geometrico se consideriamo le due superfici tratteggiate S_1 ed S_2 di Fig. 7.5b), che sono rispettivamente comprese tra $[\vartheta_1, \vartheta_1 + d\vartheta]$ e $[\vartheta_2, \vartheta_2 + d\vartheta]$. Generando in maniera uniforme in ϑ , otterremmo all'incirca lo stesso numero di punti sulle due superfici ma, essendo l'area di S_2 molto maggiore di quella di S_1 , la densità dei punti risultante sarebbe molto maggiore ai poli che all'equatore. La funzione $(1 - \cos \vartheta)$ è così interpretabile come una funzione "di peso" da assegnare ad ogni valore di ϑ perché valga la relazione (7.27).

Giudicando solo in base agli esercizi precedenti, l'algoritmo 7.2 sembrerebbe risolvere tutti i nostri problemi di generazione casuale. In realtà la situazione non è così semplice, perché la funzione $p(x)$ da integrare può essere nota solo numericamente, oppure l'integrale che compare nel termine di sinistra della (7.15) può non essere risolubile o dare come risultato una funzione non invertibile analiticamente.

In tutti questi casi esiste un'ampia varietà di procedimenti alternativi a cui fare ricorso; nei prossimi paragrafi ne presenteremo alcuni tra quelli di più comune impiego.

7.6 Metodo del rigetto

Dopo aver costruito un rettangolo (delimitato dai vertici A, B, C, D in Fig. 7.6), di base $(b - a)$ ed altezza h , che racchiude una densità di probabilità $p(x)$ (ovviamente h deve essere maggiore o uguale al massimo p_{max}

assunto da $p(x)$ in $[a, b]$, scegliamo in modo casuale al suo interno dei punti uniformemente distribuiti le cui generiche coordinate (x_i, y_i) risultano, per la (7.18):

$$\begin{cases} x_i = a + \xi_1(b - a) \\ y_i = \xi_2 h \end{cases} \quad (7.34)$$

con $0 \leq \xi_1, \xi_2 \leq 1$.

Consideriamo ora la probabilità che la variabile uniforme $a \leq X \leq b$ cada entro l'intervallo infinitesimo $[x_i, x_i + dx]$:

$$P\{x_i < X < x_i + dx\} = P\{x_i < a + \xi_1(b - a) < x_i + dx\} = \frac{dx}{(b - a)} \quad (7.35)$$

e la probabilità, condizionata da x_i , che una variabile uniforme $0 \leq Y \leq h$ sia minore di $p(x_i)$:

$$\begin{aligned} P\{Y \leq p(x)|x_i < X < x_i + dx\} &\simeq P\{Y \leq p(x_i)\} \\ &= P\{\xi_2 \leq p(x_i)/h\} = \frac{p(x_i)}{h}. \end{aligned} \quad (7.36)$$

In base al teorema 1.20 delle probabilità composte, la probabilità che si verifichino entrambe le condizioni precedenti vale:

$$\begin{aligned} P\{x_i < X < x_i + dx, Y \leq p(x_i)\} \\ &= P\{Y \leq p(x)|x_i < X < x_i + dx\} \cdot P\{x_i < X < x_i + dx\} \\ &\simeq \frac{p(x_i) dx}{h(b - a)} = \varepsilon p(x_i) dx. \end{aligned} \quad (7.37)$$

Essa coincide, a parte il fattore costante $\varepsilon = 1/[(h(b - a))]$, con la probabilità che X sia compresa in $(x_i, x_i + dx)$.

Queste semplici considerazioni sono alla base della tecnica del rigetto, che applica per tentativi la (7.37) e che possiamo così enunciare:

Algoritmo 7.3 (Rigetto elementare). *Per generare una variabile casuale continua X , di densità di probabilità qualsiasi $p(x)$ definita nell'intervallo finito $[a, b]$ ed avente valore massimo p_{max} , occorre:*

- estrarre un punto $x \in [a, b]$: $x = a + \xi_1(b - a)$;
- calcolare $p(x)$;
- estrarre un punto y compreso tra 0 ed h ($h \geq p_{max}$): $y = \xi_2 h$;
- se $y \leq p(x)$ si accetta x , altrimenti lo si scarta e si riprende il procedimento dall'inizio.

È chiaro che per applicare questo procedimento basta conoscere l'espressione analitica di $p(x)$, mentre non è richiesta alcuna informazione sulla sua cumulativa; tale vantaggio viene però pagato in rendimento poiché sono necessari almeno due numeri casuali uniformi per generare una variabile di densità $p(x)$.

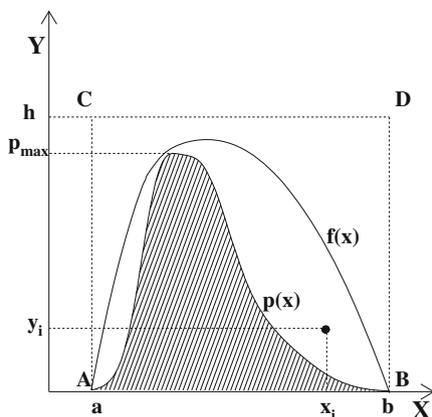


Fig. 7.6. Generazione di variabili casuali mediante il metodo del rigetto. La densità di probabilità $p(x)$ viene delimitata entro un rettangolo (oppure entro una funzione $f(x)$) ed un punto generato uniformemente all'interno di esso viene "accettato" se si trova nella regione definita da $p(x)$ e dall'asse delle ascisse (zona tratteggiata)

La costante ε , uguale alla frazione dei punti accettati rispetto alla totalità di quelli generati, rappresenta l'efficienza di generazione o, in maniera equivalente, l'inverso del numero medio di tentativi necessari per accettare un punto. Essa, come è semplice dedurre anche in base a elementari considerazioni geometriche, dipende dal rapporto tra l'area di $p(x)$ e quella del rettangolo $ABCD$ di Fig. 7.6.

Per ottimizzare il metodo, occorrerebbe generare i punti da accettare o scartare non più all'interno di un rettangolo ma entro una curva $f(x)$ che contenga $p(x)$ e che ne segua abbastanza bene l'andamento, in modo tale da massimizzare il rapporto tra le rispettive aree (vedere Fig. 7.6). Questa estensione rende il metodo del rigetto valido anche per funzioni definite in un intervallo non limitato in quanto basta trovare una funzione $f(x)$ che sia definita nello stesso intervallo. Per rendere agevole (e soprattutto rapida) l'operazione di campionamento occorre che $f(x)$ abbia una cumulativa $F(x)$ invertibile analiticamente. In tal caso, mediante le equazioni:

$$\begin{cases} x_i = F^{-1}(\xi_1) \\ y_i = \xi_2 f(x_i) \end{cases}, \quad (7.38)$$

si genera uniformemente un punto all'interno di $f(x)$ che, analogamente a quanto abbiamo appena visto, deve essere poi accettato se $y_i \leq p(x_i)$.

Per dimostrare dal punto di vista formale questo nuovo procedimento, basta riscrivere le relazioni (7.35) e (7.36), che ora diventano rispettivamente:

$$P\{x_i < X < x_i + dx\} \simeq \frac{f(x_i) dx}{\int_a^b f(x) dx} \quad (7.39)$$

e:

$$P\{Y \leq p(x) | x_i < X < x_i + dx\} \simeq P\{\xi_2 f(x_i) \leq p(x_i)\} = \frac{p(x_i)}{f(x_i)}. \quad (7.40)$$

Inserendo queste due ultime relazioni nella (7.37), otteniamo il risultato:

$$P\{x_i < X < x_i + dx, Y \leq p(x_i)\} \simeq \frac{p(x_i) dx}{\int_a^b f(x) dx} \propto p(x_i) dx. \quad (7.41)$$

L'efficienza ε del metodo viene ottenuta integrando la relazione precedente su tutti i valori di x :

$$\varepsilon = \frac{\int_a^b p(x) dx}{\int_a^b f(x) dx} = \frac{1}{\int_a^b f(x) dx}. \quad (7.42)$$

In questo caso, la costante ε , che nella (7.37) rappresentava l'inverso dell'area $h(b-a)$ del rettangolo ABCD, è l'inverso dell'integrale di $f(x)$, cioè dell'area sottesa da questa funzione con l'asse delle ascisse.

L'algoritmo precedente viene così generalizzato come:

Algoritmo 7.4 (Rigetto ottimizzato). *Per generare una variabile aleatoria X , avente una densità di probabilità qualsiasi $p(x)$ definita nell'intervallo (limitato o illimitato) $[a, b]$ e delimitata entro una funzione $f(x)$ avente una cumulativa $F(x)$ invertibile analiticamente, occorre:*

- estrarre un punto $x \in [a, b]$ distribuito secondo $f(x)$: $x = F^{-1}(\xi_1)$;
- calcolare $p(x)$;
- estrarre dalla densità uniforme un punto y compreso tra 0 ed $f(x)$: $y = \xi_2 f(x)$;
- se $y \leq p(x)$ si accetta x , altrimenti lo si scarta e si riprende il procedimento dall'inizio.

È possibile formulare anche una terza versione del metodo del rigetto (ideata dal matematico americano J. Von Neumann negli anni '50) quando la funzione $p(x)$ da cui si intende generare la variabile X è fattorizzabile come il prodotto di due funzioni:

$$p(x) = g(x)h(x), \quad (7.43)$$

dove $h(x)$ abbia una cumulativa $H(x)$ invertibile analiticamente e $g(x)$ sia limitata nell'intervallo $[a, b]$.

Possiamo pertanto scrivere:

$$H(x) = \frac{\int_a^x h(x) dx}{\int_a^b h(x) dx} \quad (7.44)$$

e:

$$0 \leq g(x) \leq G. \quad (7.45)$$

Utilizziamo come al solito due numeri uniformi indipendenti ξ_1 e ξ_2 e definiamo le condizioni:

$$\begin{cases} x_i = H^{-1}(\xi_1) \\ \xi_2 \leq \frac{g(x_i)}{G} \end{cases} \quad (7.46)$$

Le relazioni (7.39) e (7.40) risultano ora:

$$P\{x_i \leq X < x_i + dx\} \simeq \frac{h(x_i) dx}{\int_a^b h(x) dx}, \quad (7.47)$$

$$P\{\xi_2 \leq g(x)/G | x_i \leq X < x_i + dx\} = P\{\xi_2 \leq g(x_i)/G\} = g(x_i)/G \quad (7.48)$$

e quindi:

$$P\{x_i \leq X < x_i + dx, \xi_2 \leq g(x)/G\} \simeq \frac{1}{G \int_a^b h(x) dx} h(x_i) g(x_i) dx. \quad (7.49)$$

L'efficienza ε viene ottenuta, anche in questo caso, integrando su tutti i valori di x :

$$\varepsilon = \frac{\int_a^b h(x) g(x) dx}{G \int_a^b h(x) dx} \leq \frac{G \int_a^b h(x) dx}{G \int_a^b h(x) dx} = 1. \quad (7.50)$$

Se $p(x)$ è già normalizzata, allora $\int_a^b h(x) g(x) dx = 1$ e l'efficienza è semplicemente:

$$\varepsilon = \frac{1}{G \int_a^b h(x) dx} < 1. \quad (7.51)$$

La (7.43) può essere quindi riscritta come:

$$p(x) = G \cdot \int_a^b h(x) dx \cdot \frac{h(x)}{\int_a^b h(x) dx} \frac{g(x)}{G} \equiv Ch^*(x)g^*(x), \quad (7.52)$$

dove $h^*(x)$ è una densità normalizzata e $0 \leq g^*(x) \leq 1$. La costante C , se $p(x)$ è normalizzata, è l'inverso dell'efficienza e deve pertanto soddisfare alla condizione $C \geq 1$. L'algoritmo diventa allora:

Algoritmo 7.5 (Rigetto pesante). Per generare i valori di una variabile casuale X , avente una densità di probabilità qualsiasi $p(x)$ definita nell'intervallo (finito o infinito) $[a, b]$ e fattorizzabile come $p(x) = Cg(x)h(x)$, dove $C \geq 1$, $0 \leq g(x) \leq 1$ e dove $h(x)$ è una densità di probabilità avente una cumulativa $H(x)$ invertibile analiticamente, occorre:

- generare $0 \leq \xi_1 \leq 1$;
- generare un punto $x \in [a, b]$ distribuito secondo $h(x)$: $x = H^{-1}(\xi_1)$;
- calcolare $g(x)$;
- generare $0 \leq \xi_2 \leq 1$;
- se $\xi_2 \leq g(x)$ si accetta x , altrimenti lo si scarta e si riprende il procedimento dall'inizio.

Questo algoritmo può essere facilmente ricordato se si considera $h(x)$ come densità di base degli eventi e $g(x)$ come funzione “peso”: generato un punto $x_i = H^{-1}(\xi_1)$, esso sarà tanto più importante (“pesante”) quanto più $g(x_i)$ è vicino ad uno. Questa condizione viene tenuta presente nella seconda generazione, quando si accetta l’evento solo se $\xi_2 \leq g(x_i)$.

Notiamo infine che se nella (7.52) la funzione peso vale $g(x) = p(x)/h$ e $h(x) = 1/(b - a)$ otteniamo l’algoritmo 7.3, mentre se la funzione peso vale $g(x) = p(x)/h(x)$ abbiamo l’algoritmo 7.4.

Esercizio 7.6

Generare una variabile distribuita nell’intervallo $[0, \pi/2]$ con densità di probabilità:

$$p(x) = x \sin x \, dx . \quad (7.53)$$

Risposta. In questo caso non possiamo applicare il metodo di inversione della cumulativa (algoritmo 7.2) in quanto l’equazione:

$$\int_0^x x \sin x = \sin x - x \cos x = \xi \quad (7.54)$$

non è invertibile analiticamente per x . Applichiamo perciò la tecnica del rigetto utilizzando i tre procedimenti che abbiamo appena derivato.

- *Algoritmo 7.3.*

Delimitiamo $p(x)$ entro il quadrato di base $[0, \pi/2]$ ed altezza $\pi/2$, come in Fig. 7.7. Con riferimento alle equazione (7.34), abbiamo ora che $a = 0$; $b = \pi/2$; $h = \pi/2$. La variabile richiesta viene perciò simulata, con un’efficienza di estrazione di circa il 40% ($\varepsilon = 4/\pi^2$), mediante le seguenti istruzioni SCILAB:

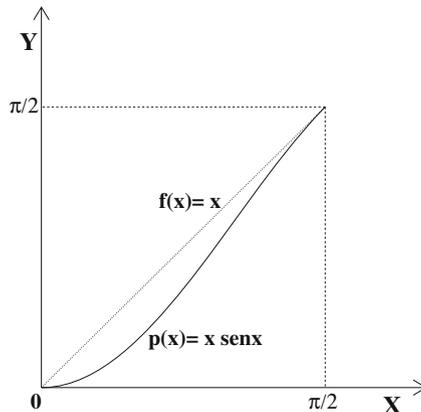


Fig. 7.7. Andamento delle funzioni $f(x) = x$ e $p(x) = x \sin(x)$ nell’intervallo $[0, \pi/2]$

```

pig2= 1.570796;
n = 20000;
xv = zeros(1,n);
for k=1:n;
    csi2=1; px=0.;
    while (csi2>px),
        x = pig2 * sqrt(grand(1,1,"unf",0,1));
        px= sin(x);
        csi2= grand(1,1,"unf",0,1);
        xv(k) = x;
    end;
end;
clf(); Histplote(20,xv,stat=1,errors=1); // display

```

• *Algoritmo 7.4.*

Se generiamo dei punti uniformemente distribuiti all'interno della curva $f(x) = x$ (come in Fig. 7.7), raddoppiamo l'efficienza di generazione rispetto all'algoritmo precedente in quanto il rapporto tra le aree di $p(x)$ ed $f(x)$ risulta:

$$\varepsilon = \int_0^{\pi/2} p(x) dx \Big/ \int_0^{\pi/2} f(x) dx = \left(\frac{8}{\pi^2}\right) \simeq 81\% . \quad (7.55)$$

Per compiere tale operazione, occorre prima normalizzare $f(x)$, calcolandone l'area nell'intervallo $[0, \pi/2]$:

$$\int_0^{\pi/2} x dx = \frac{\pi^2}{8} \quad (7.56)$$

e poi risolvere l'equazione:

$$\int_0^x \left(\frac{8}{\pi^2}\right)x dx = \xi_1 , \quad (7.57)$$

che dà come risultato:

$$x = \left(\frac{\pi}{2}\right)\sqrt{\xi_1} . \quad (7.58)$$

In questo modo abbiamo generato l'ascissa x_i del generico punto casuale, mentre la sua ordinata viene ricavata generando uniformemente tra 0 ed $f(x) = x$, condizione che fornisce l'equazione $y = x\xi_2$. Il ciclo di generazione casuale diventa così :

```

while (csi>px),
    x = pig2 * sqrt(grand(1,1,"unf",0,1));
    px= x*sin(x);
    csi= x * grand(1,1,"unf",0,1);
    xv(k) = x;
end;

```

- *Algoritmo 7.5.*
Sfruttando la (7.56), è immediato fattorizzare $p(x)$ come:

$$p(x) = \left(\frac{\pi^2}{8}\right) \left(\frac{8}{\pi^2}\right) x \operatorname{sen} x \quad (7.59)$$

ed applicare questo procedimento ponendo $C = \pi^2/8$; $g(x) = \operatorname{sen} x$ e $h(x) = (8/\pi^2)x$. Il ciclo da eseguire diventa:

```
while (csi>px),
  x = pig2 * sqrt(grand(1,1,"unf",0,1));
  px= sin(x);
  csi= grand(1,1,"unf",0,1);
  xv(k) = x;
end;
```

7.7 Metodo di ricerca lineare

Quando la cumulativa $F(x)$ non è esprimibile analiticamente, possiamo sempre calcolare numericamente l'integrale

$$F(x) = \int_a^x p(x) dx \quad (7.60)$$

per N valori diversi x_1, x_2, \dots, x_N (con $x_1 = a$ ed $x_N = b$) e costruire così per punti tale funzione (vedere Fig. 7.8):

$$\begin{aligned} F(x_1) &= 0 \\ F(x_2) &= F(x_1) + \int_{x_1}^{x_2} p(x) dx, \\ F(x_3) &= F(x_2) + \int_{x_2}^{x_3} p(x) dx, \\ &\dots\dots\dots \\ F(x_N) &= F(x_{N-1}) + \int_{x_{N-1}}^{x_N} p(x) dx = 1. \end{aligned} \quad (7.61)$$

In questo modo ci siamo riportati al caso già discusso nel par. 7.4, relativo alla generazione di una variabile discreta. Infatti, attraverso la relazione:

$$F_{j-1} = \sum_{i=1}^{j-1} p(x_i) < \xi_i \leq \sum_{i=1}^j p(x_i) = F_j, \quad (1 \leq j \leq N) \quad (7.62)$$

siamo subito in grado di determinare il particolare sottointervallo $[x_{j-1}, x_j]$ in cui è compresa la variabile aleatoria che dobbiamo generare.

Se supponiamo che $F(x)$ vari linearmente entro ogni sottointervallo, interpolando linearmente tra i due estremi x_{j-1} ed x_j del sottointervallo prescelto si ottiene:

$$x = x_{j-1} + \frac{\xi - F_{j-1}}{(F_j - F_{j-1})} (x_j - x_{j-1}). \quad (7.63)$$

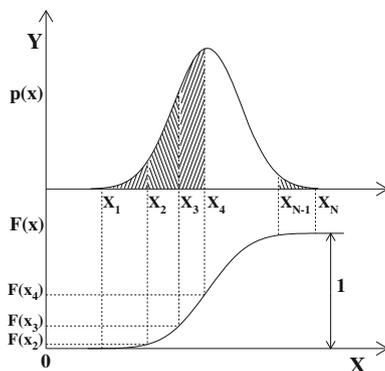


Fig. 7.8. Generazione di variabili casuali mediante il metodo di ricerca lineare. Si costruisce per punti la funzione cumulativa $F(x)$ per riportarsi in un caso analogo a quello della generazione di una variabile discreta

Dato che la cumulativa $F(x)$ ha, per costruzione, un andamento monotono crescente ed è pertanto una funzione regolare, l'approssimazione (7.63) dà in generale un risultato molto buono. Tutte queste considerazioni sono riassunte nel seguente procedimento:

Algoritmo 7.6 (Ricerca lineare). Per generare una variabile casuale X , avente una densità di probabilità qualsiasi $p(x)$ definita nell'intervallo limitato o illimitato $[a, b]$ occorre:

- calcolare numericamente la cumulativa per N punti x_1, x_2, \dots, x_N ($x_1 = a$; $x_N = b$):

$$F_j = \int_a^{x_j} p(x) dx \quad 1 \leq j \leq N ; \quad (7.64)$$

- generare un numero $0 \leq \xi \leq 1$;
- determinare l'indice j tale che: $F_{j-1} < \xi \leq F_j$;
- calcolare x mediante la relazione (7.63).

7.8 Metodi particolari di generazione casuale

In alcuni casi nessuno degli algoritmi finora discussi riesce a generare in modo semplice o sufficientemente rapido valori di variabili aleatorie.

Fortunatamente, esistono ormai da molto tempo diversi algoritmi consolidati per risolvere "ad hoc" e in modo efficiente molti problemi particolari di generazione casuale.

Nel seguito vi mostreremo, come esempio, solo alcuni dei metodi utilizzati per generare variabili da densità gaussiane e poissoniane mentre, per avere

una rassegna completa di tutti (o quasi) i differenti algoritmi di generazione, vi consigliamo di consultare le referenze [Fis96], [Knu81], [Mor84] e [Rub81].

Ricordiamo pure che la funzione `grand` di SCILAB permette di generare variabili aleatorie distribuite secondo le densità più comuni.

a) Generazione di variabili gaussiane.

Come abbiamo visto nel par. 3.6, per ottenere una densità $g(x)$, avente μ e σ qualsiasi, basta ottenere un valore standard

$$t = \frac{x - \mu}{\sigma} \quad (7.65)$$

dalla densità gaussiana standard (avente $\mu = 0$ e $\sigma = 1$):

$$g(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \quad (-\infty \leq t \leq +\infty) \quad (7.66)$$

ed effettuare la trasformazione inversa della (7.65):

$$x = t\sigma + \mu . \quad (7.67)$$

A questo punto non siamo però in grado di esprimere analiticamente la cumulativa di $g(t)$ e, di conseguenza, utilizzare l'algoritmo 7.2.

Dato l'intervallo infinito di variazione di t , non è utilizzabile nemmeno l'algoritmo 7.3, mentre sono invece applicabili gli algoritmi 7.4 (vedere esercizio 7.5), 7.5 e 7.6 (vedere, ad esempio, [TC93]). Ad essi vengono tuttavia di solito preferiti altri metodi, più semplici od efficienti. Uno di questi, basato sul teorema Limite Centrale, utilizza la somma di variabili aleatorie ed è descritto nel problema 7.7. Potete anche vedere le routine `Simgauss1` e `Gauss1` sul nostro sito.

Il procedimento che viene più frequentemente usato per la generazione gaussiana è quello ideato alla fine degli anni '50 dai matematici americani G.E.P. Box e M.E. Muller, i quali hanno mostrato che, contrariamente quanto si potrebbe supporre intuitivamente, è più facile generare simultaneamente non una ma due variabili gaussiane indipendenti. Consideriamo infatti la gaussiana standard bidimensionale in coordinate polari (r, φ) dell'equazione (4.79):

$$g(r, \varphi) dr d\varphi = \frac{r}{2\pi} e^{-r^2/2} dr d\varphi = p(r)q(\varphi) dr d\varphi \quad (7.68)$$

e calcoliamo le funzioni cumulative del modulo e dell'angolo del vettore polare:

$$\xi_1 = P(r) = \int_0^r p(r) dr = 1 - e^{(-r^2/2)} , \quad (7.69)$$

$$\xi_2 = Q(\varphi) = \int_0^\varphi q(\varphi) d\varphi = \frac{\varphi}{2\pi} . \quad (7.70)$$

A questo punto è facile immaginare il motivo della scelta compiuta: sia $p(r)$ che $q(\varphi)$ hanno delle cumulative che sono invertibili analiticamente. Applicando l'algoritmo 7.2 otteniamo facilmente:

$$\begin{cases} r = \sqrt{-2 \log \xi_1} \\ \varphi = 2\pi\xi_2 . \end{cases} \quad (7.71)$$

Passando alle coordinate cartesiane (z_1, z_2) :

$$\begin{cases} z_1 = r \cos \varphi = \sqrt{-2 \log \xi_1} \cos(2\pi\xi_2) \\ z_2 = r \sin \varphi = \sqrt{-2 \log \xi_1} \sin(2\pi\xi_2) , \end{cases} \quad (7.72)$$

otteniamo i valori di una coppia di variabili gaussiane standard indipendenti partendo da due numeri casuali ξ_1 e ξ_2 .

Per rendere più veloce l'algoritmo si può utilizzare un ingegnoso espediente, descritto in [Knu81]: se generiamo casualmente un punto di coordinate cartesiane (v_1, v_2) all'interno del cerchio unitario centrato sull'origine, la somma $s = v_1^2 + v_2^2$ è un numero casuale uniformemente compreso tra 0 ed 1 (vi lasciamo come esercizio la semplice dimostrazione di questa affermazione) che possiamo usare al posto di ξ_1 , mentre l'angolo definito da questo punto e dall'asse delle ascisse rappresenta l'angolo casuale $2\pi\xi_2$. In tal modo si evita il calcolo diretto delle funzioni trigonometriche poiché il coseno ed il seno che compaiono nelle equazione (7.72) vengono calcolati attraverso i rapporti v_1/\sqrt{s} e v_2/\sqrt{s} .

Per ottenere le coordinate (v_1, v_2) si ha lo svantaggio di dover utilizzare la tecnica del rigetto, poiché occorre prima generare due numeri v_1, v_2 uniformemente distribuiti nell'intervallo $[-1, 1]$ e poi accettare solo le coppie che soddisfano la condizione: $v_1^2 + v_2^2 \leq 1$. Tuttavia l'efficienza di questa operazione ($\simeq 78\%$, pari al rapporto tra cerchio unitario ed il quadrato ad esso circoscritto) è abbastanza elevata e quest'ultimo procedimento è risultato, sul calcolatore da noi utilizzato, all'incirca un 20% più veloce del metodo "classico" descritto dalle equazioni (7.72). L'algoritmo di generazione gaussiana può essere quindi così schematizzato:

Algoritmo 7.7 (Generazione gaussiana). *Per generare due variabili gaussiane normalizzate indipendenti Z_1, Z_2 occorre:*

- generare due numeri casuali indipendenti $0 \leq \xi_1, \xi_2 \leq 1$;
- definire $v_1 = 2\xi_1 - 1$; $v_2 = 2\xi_2 - 1$ e calcolare la somma $s = v_1^2 + v_2^2$;
- se $s > 1$ ripetere il procedimento dall'inizio;
- se $s \leq 1$ generare gli eventi $\{Z_1 = z_1\}, \{Z_2 = z_2\}$ come:

$$\begin{cases} z_1 = v_1 \sqrt{\frac{-2 \log s}{s}} \\ z_2 = v_2 \sqrt{\frac{-2 \log s}{s}} . \end{cases} \quad (7.73)$$

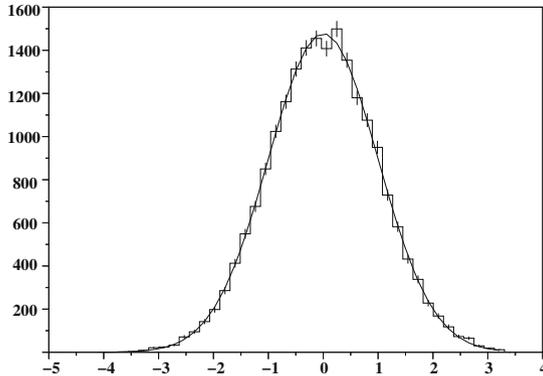


Fig. 7.9. Istogramma di un campione di 20 000 numeri casuali generati con la routine Gauss2. La curva continua rappresenta la densità gaussiana standard

Questo algoritmo è realizzato nella routine Gauss2, qui riportata, che dà come risultato l'istogramma di Fig. 7.9.

```
// routine Gauss2: generazione di due
// variabili gaussiane col metodo di Box & Muller
// iset e g2 sono variabili globali
// se iset = 1 si utilizza la variabile globale
// g2 precedentemente calcolata

global iset g2;
iset=0;

function x = Gauss2(mu, sigma)
s = 2;
if(iset==0) then
while (s>1),
v1 = 2. * grand(1,1,"unf",0,1) - 1;
v2 = 2. * grand(1,1,"unf",0,1) - 1;
s = v1*v1 + v2*v2;
end
ls = sqrt(-2.*log(s)/s);
g1 = v1 * ls;
g2 = v2 * ls;
iset=1;
else
iset=0;
g1=g2;
end;
x = mu + sigma*g1
endfunction;
```

```
// esempio di chiamata con grafico dei risultati

n=20000; s=40; // si generano 20000 numeri in 40 canali
g = zeros(1,n);
for k=1:n, g(k) = Gauss2(0,1); end;
dx= (max(g)-min(g))/s;
x= [min(g)+0.5*dx:(max(g)-min(g))/s:max(g)-0.5*dx];
clf();
plot(x,n*dx*exp(-x^2/2)/sqrt(2*3.14159265)); // curva gaussiana
Histplote(s,g,stat=1,errors=1); // istogramma simulato
end;
```

b) Generazione di variabili poissoniane.

Ogni processo stocastico in cui vengono generate, con probabilità λ costante nel tempo, variabili discrete e indipendenti è caratterizzato da due proprietà fondamentali:

- l'intervallo di tempo tra due eventi successivi è una variabile aleatoria di densità esponenziale;
- il numero di eventi che cadono entro un certo intervallo di tempo di ampiezza Δt prefissata è una variabile poissoniana di media $\mu = \lambda \Delta t$.

Nell'esercizio 3.12 abbiamo visto che il tempo τ tra due eventi stocastici successivi viene simulato attraverso l'equazione

$$\tau = -\frac{1}{\mu} \log \xi . \quad (7.74)$$

Per ricavare il numero x di eventi che cadono entro $\Delta t = 1$ basta quindi generare una sequenza $\tau_0, \tau_1, \dots, \tau_x, \tau_{x+1}$ di intervalli temporali (con $\tau_0 = 0$) fino a quando è verificata la disuguaglianza:

$$\sum_{i=0}^x \tau_i \leq 1 < \sum_{i=0}^{x+1} \tau_i \quad x = 0, 1, \dots \quad (7.75)$$

Tenendo conto della (7.74) possiamo riscrivere questa disuguaglianza come:

$$-\sum_{i=0}^x \log \xi_i \leq \mu < -\sum_{i=0}^{x+1} \log \xi_i \quad x = 0, 1, \dots \quad (7.76)$$

oppure, eliminando i logaritmi:

$$\prod_{i=0}^x \xi_i \geq e^{-\mu} > \prod_{i=0}^{x+1} \xi_i \quad x = 0, 1, \dots \quad (7.77)$$

In conclusione abbiamo dedotto il seguente:

Algoritmo 7.8 (Generazione poissoniana). *Per generare una variabile poissoniana X di media μ occorre:*

- definire $k = 0$; $s = 1$;
- generare un numero casuale $0 \leq \xi_k \leq 1$;
- porre $s = s \xi_k$;
- se $s < e^{-\mu}$ porre $x = k$; altrimenti porre $k = k + 1$ e ripetere il procedimento dal secondo passo.

È molto facile dedurre che il tempo di calcolo di questo algoritmo aumenta proporzionalmente a μ ; per tale motivo, quando μ è abbastanza grande ($\mu \geq 30$) conviene senz'altro sfruttare il fatto che, come rilevato nel par. 3.3, la poissoniana viene molto ben approssimata da una gaussiana avente media e varianza μ . Data una variabile gaussiana normalizzata Z , possiamo ottenere un'altra variabile gaussiana G con le caratteristiche sopra indicate attraverso la relazione:

$$G = \mu + Z\sqrt{\mu} . \quad (7.78)$$

Per ottenere i valori di X , che è una variabile intera ≥ 0 , dobbiamo però arrotondare G , variabile reale compresa nell'intervallo $(-\infty, +\infty)$, al più vicino intero non negativo:

$$X = \max\left\{0, [G + 0.5]\right\} = \max\left\{0, [\mu + Z\sqrt{\mu} + 0.5]\right\} . \quad (7.79)$$

La routine `Poiss`, dato come ingresso $\mu = \text{mu}$, genera una variabile poissoniana sfruttando, per $\mu < 30$, l'algoritmo 7.8 e, per $\mu \geq 30$, la relazione (7.79) e la subroutine `Gauss2` che abbiamo appena presentato:

```
// routine Poiss: generazione del numero di eventi
// osservati x da una poissoniana di media mu

function x = Poiss(mu)
  if(mu<30) then
    s=1; x=0;
    while (s>exp(-mu)), s = s*grand(1,1,"unf",0,1); x=x+1; end;
    x= x-1;
    if(x<0) then x=0; end;
  else // approssimazione gaussiana
    x = int(Gauss2(mu,sqrt(mu)) + 0.5);
    if(x<0) then x=0; end;
  end;
endfunction;
```

Un utile esercizio da fare con la routine `Poiss`, è la verifica della Tab. 6.2 del par. 6.7. Ponendo $\mu = 2.3$ e generando campioni sempre più grandi, osserverete che il numero di volte in cui vi sono zero eventi estratti tenderà al 10%, che ponendo $\mu = 4.74$ il numero di successi $x \leq 1$ tenderà al 5% e così via. Per questa ed altre prove potete copiare dal nostro sito la routine `Simpois`, che elabora istogrammi poissoniani.

Per simulare variabili aleatorie aventi distribuzioni qui non considerate, basta ricorrere alle definizioni ed ai teoremi del calcolo delle probabilità; ad esempio, per simulare una variabile $Q(\nu) \sim \chi^2(\nu)$ basta ricordare il teorema 3.3 e sommare i quadrati di ν variabili gaussiane standard gaussiane. Seguendo le definizioni standard si ottengono senz'altro algoritmi corretti, anche se non sempre particolarmente rapidi ed efficienti.

7.9 Studio Monte Carlo di distribuzioni

Una delle applicazioni più interessanti e spettacolari del metodo MC è probabilmente la determinazione di distribuzioni statistiche comunque complicate. Come abbiamo visto nel cap. 5, lo studio delle distribuzioni di variabili aleatorie funzioni di altre variabili aleatorie richiede tecniche matematiche alquanto laboriose, tanto che la soluzione analitica, nella maggior parte dei casi, può essere alla portata solo di abili matematici o non essere nemmeno ottenibile in pratica.

Il metodo Monte Carlo ha rivoluzionato questo ramo della statistica applicata, in quanto consente di ottenere, con procedure elementari, la soluzione richiesta, anche se nella forma approssimata di istogrammi e non nella forma analitica corretta. Dall'istogramma simulato si possono poi estrarre i parametri che servono (come media, varianza, area delle code) con errore statistico trascurabile rispetto al problema che si sta trattando.

La procedura è schematizzata in Fig. 7.10: si generano le variabili aleatorie (X_1, X_2, X_3, \dots) secondo le rispettive densità e si calcola, ad ogni estrazione, la variabile $Z = f(X_1, X_2, X_3, \dots)$; alla fine del ciclo si presenta il risultato come istogramma della variabile Z . Il passo di istogrammazione può essere variato a piacere e reso congruo col problema in esame. Dall'istogramma simulato si calcolano infine le quantità statistiche fondamentali che caratterizzano la distribuzione. Si può anche eseguire il *best fit* dell'istogramma (con i metodi che vi illustreremo nel par. 9.7) utilizzando funzioni empiriche come esponenziali, polinomi, somme di gaussiane, ottenendo così una forma analitica, indipendente dal passo dell'istogramma, *che interpola la soluzione vera* con il grado di accuratezza desiderato. Se si dispone di un calcolatore, il metodo è talmente semplice da poter essere illustrato in modo completo con un facile esempio.

Si supponga di voler trovare la distribuzione della variabile aleatoria

$$Z = \frac{\text{sen } X}{\text{sen } Y}, \quad (7.80)$$

dove X e Y sono angoli gaussiani di media e deviazione standard pari a:

$$\begin{aligned} \theta_x &= 20^\circ, & \sigma_x &= 3^\circ, \\ \theta_y &= 13^\circ, & \sigma_y &= 3^\circ. \end{aligned} \quad (7.81)$$

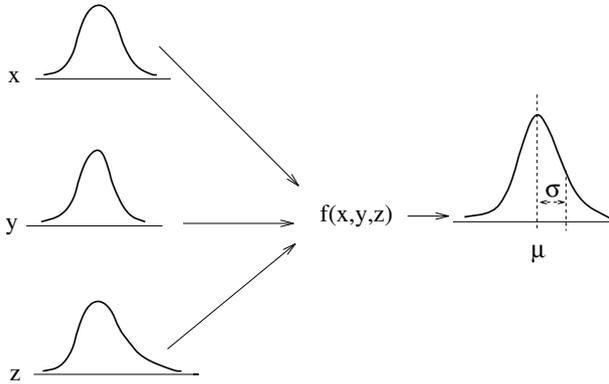


Fig. 7.10. Determinazione di una distribuzione con il metodo MC

Utilizzando le formule approssimate (5.63, 5.66) del par. 5.4, possiamo stimare media e varianza della distribuzione incognita come:

$$\langle Z \rangle = \frac{\text{sen } \theta_x}{\text{sen } \theta_y} = 1.52 , \quad (7.82)$$

$$\sigma[Z] = \left[\left(\frac{\cos \theta_x}{\text{sen } \theta_y} \right)^2 \left(\frac{\pi}{180} \right)^2 \sigma_x^2 + \left(\frac{\text{sen } \theta_x \cos \theta_y}{\text{sen}^2 \theta_y} \right)^2 \left(\frac{\pi}{180} \right)^2 \sigma_y^2 \right]^{1/2} = 0.41 , \quad (7.83)$$

dove la varianza è stata trasformata in radianti, dato che z è un numero reale espresso nel sistema decimale.

Come illustrato nel par. 5.4, queste due formule sono approssimate: la (7.82) vale solo se la relazione tra Z e (X, Y) è lineare, mentre la (7.83) richiede sia la validità della dipendenza lineare sia la presenza di errori percentuali piccoli.

Poiché queste condizioni sono violate in modo drastico dalle (7.80, 7.81), in questo caso non siamo affatto certi né della bontà delle (7.82, 7.83) né del tipo di distribuzione della variabile Z . Anzi, le nostre conoscenze di calcolo delle probabilità ci autorizzano a supporre una distribuzione non gaussiana per Z . Vediamo come la simulazione sia in grado di risolvere agevolmente tutti questi problemi.

Utilizzando la routine `grand` abbiamo generato, con la routine `Rifraz` qui riportata, 20 000 variabili gaussiane (7.81) e ne abbiamo fatto il rapporto dei seni come nella (7.80): il risultato è l'istogramma di Fig. 7.11a).

```
// routine Rifraz: simulazione della misura
// di un indice di rifrazione
```

```
n=20000;
g = zeros(1,n);
```

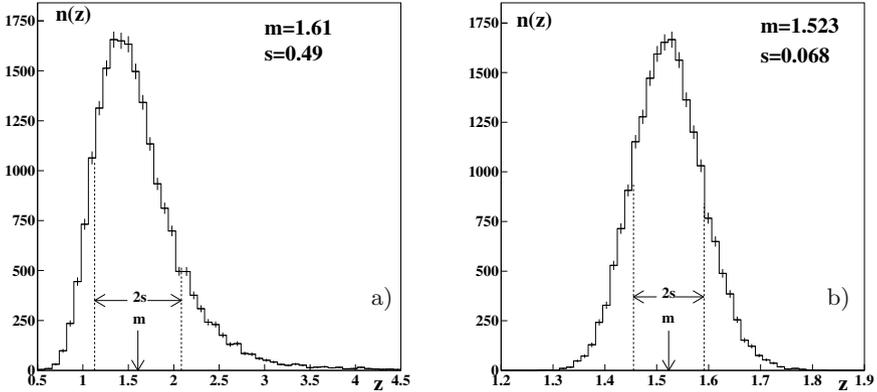


Fig. 7.11. Istogramma di 20 000 variabili Z della (7.80), dove X e Y sono variabili gaussiane date a) dalle (7.81); b) dalla (7.84)

```
rad=57.296; // conversione gradi radianti
for k=1:n,
    a1=20/rad; a2=3/rad; a3=13/rad; t=100; // valori in radianti
    while (t>4.5), // esclusione dei valori piu' alti (trascurabili)
        t = sin(grand(1,1,'nor',a1,a2))/sin(grand(1,1,'nor',a3,a2));
    end;
    g(k) = t;
end;
clf(); Histplote(40,g,stat=1,errors=1); // istogramma simulato
end;
```

Come si vede, i parametri μ e σ , stimati dall'istogramma:

$$\langle Z \rangle \simeq 1.61, \quad \sigma[Z] \simeq 0.49,$$

hanno valori abbastanza differenti dalle predizioni delle (7.82, 7.83). L'errore statistico su queste due quantità, stimato dalle formule fondamentali di Tab. 6.3 di pagina 213, risulta dell'ordine di 0.002-0.003 e può essere pertanto trascurato nelle considerazioni che seguono. Esso comunque può essere reso piccolo a piacere aumentando il numero di variabili simulate.

La simulazione rivela anche che *la forma della densità si discosta in modo sensibile dalla curva normale o gaussiana*. Dall'istogramma risulta che gli eventi nell'intervallo

$$\mu \pm \sigma \simeq m \pm s = 1.61 \pm 0.49 = [1.12, 2.10]$$

sono pari a 14 895, corrispondenti a una percentuale del 75%.

Come si vede, a parte la non conoscenza della forma analitica esatta della densità istogrammata in Fig. 7.11a), tutte le altre informazioni sono facilmente

accessibili con la simulazione. È anche interessante diminuire la dispersione di X e Y e vedere quello che succede. L'intuizione ci dice che dovremmo eliminare la coda lunga della densità, che è dovuta alla non linearità della funzione seno. Operiamo allora come in precedenza, sostituendo però alle (7.81) le condizioni:

$$\begin{aligned} \theta_x &= 20^\circ, & \sigma_x &= 0.5^\circ, \\ \theta_y &= 13^\circ. & \sigma_y &= 0.5^\circ. \end{aligned} \quad (7.84)$$

La (7.82) rimane invariata, mentre la (7.83) predice in questo caso il valore:

$$\sigma[Z] \simeq 0.068 .$$

Il risultato della simulazione è mostrato in Fig. 7.11b). Ora la densità è diventata “quasi” normale, con media e deviazione standard praticamente coincidenti con le previsioni delle (7.82, 7.83). Nell'intervallo $m \pm s$ si trovano 13 723 eventi, pari al 68.6%, un valore in pieno accordo con la legge 3σ .

7.10 Determinazione degli intervalli di confidenza

Le tecniche MC hanno un ruolo fondamentale anche nella determinazione degli intervalli di confidenza, perché consentono di risolvere in modo approssimato gli integrali (6.7) di pagina 184 anche quando è complicato o impossibile trovare la forma funzionale della densità $p_Z(z; \theta)$, dove $Z = f(X_1, X_2, \dots)$ è una variabile aleatoria funzione di altre variabili primarie di distribuzione nota. Il parametro (o i parametri) θ sono in genere relativi alle distribuzioni delle variabili X_1, X_2, \dots

Supponendo di dover effettuare la stima di un parametro $\theta \in \Theta$, il metodo segue la definizione 6.1: se z_{mis} è il valore misurato (spesso uno stimatore di θ ricavato da un campione di dimensione n), per ogni valore di θ si deve ottenere, con i metodi descritti nel paragrafo precedente, un istogramma rappresentativo della densità $p_Z(z; \theta)$. Le condizioni espresse dagli integrali (6.7) vengono approssimate contando, per ogni istogramma generato, la frazione f (frequenza) di eventi per cui $z \leq z_{\text{mis}}$. Gli estremi dell'intervallo di confidenza $[\theta_1, \theta_2]$ si trovano quando

$$1 - f \simeq \int_{z_{\text{mis}}}^{\infty} p(z; \theta_1) dz = c_1, \quad f \simeq \int_{-\infty}^{z_{\text{mis}}} p(z; \theta_2) dz = c_2. \quad (7.85)$$

Se il parametro è continuo, si considera una griglia discreta di valori di passo $\Delta\theta$ tale da consentire la precisione desiderata.

Nella (7.85) l'uguaglianza è stata sostituita col simbolo \simeq , che va inteso come *circa uguale entro l'errore statistico*. Se vogliamo fissare la precisione attraverso la deviazione standard, dalla (6.18) segue che dobbiamo simulare un numero n di osservazioni tale che gli errori

$$\sqrt{\frac{c_1(1-c_1)}{n}}, \quad \sqrt{\frac{c_2(1-c_2)}{n}} \quad (7.86)$$

assumano il valore desiderato. Supponiamo ad esempio di voler determinare l'intervallo standard simmetrico con $CL = 0.683$; in questo caso, come sappiamo, $(1 - CL)/2 = c_1 = c_2 = 0.158$. Se generiamo un istogramma con 100 000 eventi, nelle code da ricercare ve ne saranno circa 15 800 e dalla (7.86) risulta che l'errore statistico da associare ad f vale circa 0.001.

Quando la densità $p_Z(z, \theta)$ dipende da un insieme di parametri $\theta \in \Theta$, si costruisce una griglia multidimensionale per tutti i parametri e si registrano gli insiemi di valori per i quali, come nelle (7.85), vengono soddisfatti i livelli corrispondenti agli insiemi di confidenza scelti.

Se sono coinvolte densità congiunte $p_Z(\mathbf{z}; \theta)$, le regioni di confidenza ed i relativi livelli possono essere molto complicati e variano a seconda del problema in esame [JLPe00]. Il caso più semplice si ha quando le variabili sono gaussiane ed i parametri θ sono i valori medi veri incogniti. Se $z_{i \text{ mis}}$ e $z_{i \text{ sim}}$ sono rispettivamente i valori misurati e simulati e θ_i i valori di griglia dei parametri, si includono nell'insieme di confidenza di livello CL i punti θ della griglia multidimensionale per i quali sia approssimativamente CL la frequenza relativa f con la quale si è verificata la condizione

$$\chi^2 = \sum_i \frac{(z_{i \text{ sim}} - \theta_i)^2}{\sigma_i^2} \leq \sum_i \frac{(z_{i \text{ mis}} - \theta_i)^2}{\sigma_i^2} = \chi_0^2. \quad (7.87)$$

Spesso il CL richiesto corrisponde alla regione $\Delta\chi^2 \leq 1$, i cui estremi definiscono, come mostrato nella Fig. 4.8 di pagina 146, le deviazioni standard di ogni variabile. Questa tecnica si chiama, come vedremo più avanti nel cap. 11, propagazione degli errori.

Il metodo basato sulle (7.85-7.87), detto *metodo della griglia*, è del tutto generale, ma risulta spesso laborioso. In alcuni casi è però possibile, se la densità di Z è invariante rispetto ai parametri θ , procedere in modo molto più facile e diretto.

Ci rifacciamo qui ai risultati del par. 6.2 di pagina 181. Se vale la (6.10), la seconda delle (7.85) diviene

$$c_2 = \int_{-\infty}^{z_{\text{mis}}} p(z; \theta_2) dz = F(z_{\text{mis}}; \theta_2) = 1 - F(\theta_2; z_{\text{mis}}),$$

equazione che, quando F è invertibile, viene immediatamente risolta rispetto a θ_2 ponendo

$$\theta_2 = F^{-1}(1 - c_2; z_{\text{mis}}). \quad (7.88)$$

Analogamente, si ottiene per θ_1 :

$$\theta_1 = F^{-1}(c_1; z_{\text{mis}}). \quad (7.89)$$

In altre parole, si tratta di trovare i quantili di ordine $1 - c_2$ e c_1 della distribuzione F con parametro z_{mis} . Quando non è possibile procedere all'inversione analitica di F , θ_1 e θ_2 sono facilmente ricavabili simulando un istogramma di questa distribuzione.

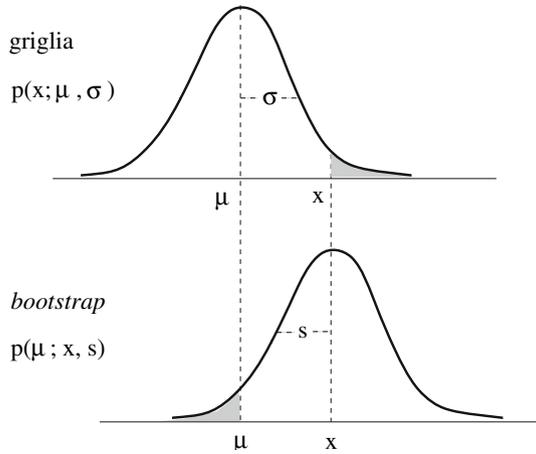


Fig. 7.12. Ricerca dell'intervallo di confidenza, quando vale la (6.10), nel caso di simmetria ed invarianza di forma della densità rispetto ai parametri considerati (in questo caso μ). Col metodo della griglia (curva in alto), si trova il valore di μ per il quale l'area ombreggiata ha il valore consistente col CL assegnato (in genere $1 - CL$ o $(1 - CL)/2$). Col metodo *bootstrap* (curva in basso), si ottiene un istogramma dalla densità che ha come parametro il valore misurato x e si trova l'intervallo di confidenza di μ come se fosse un intervallo di probabilità. L'area delle code ombreggiate è uguale

Nel caso gaussiano valgono le (6.9, 6.10) di pagina 185 e quindi, se Z ha media θ (ad esempio è lo stimatore della media campionaria) la (6.10) diventa

$$F(z_{\text{mis}} - \theta) = 1 - F(\theta - z_{\text{mis}})$$

dove F rappresenta la funzione di ripartizione della gaussiana di media nulla (la varianza non interessa), simmetrica rispetto allo zero. Se $CL = 1 - \alpha$ e $c_1 = c_2 = \alpha/2$, la (7.88) e la (7.89) danno

$$\theta_1 - z_{\text{mis}} = F^{-1}(\alpha/2), \quad \theta_2 - z_{\text{mis}} = F^{-1}(1 - \alpha/2)$$

cioè

$$\theta_1 = z_{\text{mis}} + t_{\alpha/2}, \quad \theta_2 = z_{\text{mis}} + t_{1-\alpha/2} .$$

(Si guardi anche la Fig. 7.12, con $\mu = \theta_1$ e $x = z_{\text{mis}}$.) In sostanza, per risolvere il problema con il Monte Carlo, ci si centra sul valore misurato e si trova la localizzazione di θ_2 dalla coda della densità, simulando un campione dalla densità gaussiana $g(z; z_{\text{mis}})$ e valutando θ_2 attraverso l'istogramma. Si procede in modo simile per θ_1 .

Questa tecnica [Buc84, DH99, Rip86], che in pratica sostituisce ai parametri veri quelli stimati, fa parte di una classe di algoritmi Monte Carlo detti di *bootstrap*¹, di cui parleremo più diffusamente nel prossimo paragrafo

¹ In italiano si dovrebbe dire, ma è poco usato, *algoritmi a laccio*.

Applichiamo ora questi principi generali al caso concreto della (7.80). Se le (7.81) rappresentano una misura, alle quantità vere θ_x e θ_y vanno sostituiti i valori sperimentali t_x e t_y . Supporremo che la quantità $z \equiv n_{\text{mis}} = (\text{sen } t_x / \text{sen } t_y) = 1.52$ sia la misura dell'indice di rifrazione n ottenuto con un goniometro ottico misurando le direzioni dei raggi luminosi con errori di $\pm 3^\circ$ oppure di $\pm 0.5^\circ$. Il risultato, come avviene in genere per le misure di laboratorio, va dato ad un livello di confidenza del 68.3% centrato sul valore misurato $n_{\text{mis}} = 1.52$.

Consideriamo, per cominciare, il caso di errori grandi ($\sigma = 3^\circ$). Per applicare il metodo della griglia, essendo gli angoli misurati delle variabili gaussiane indipendenti, occorre calcolare la distribuzione di $\text{sen } x / \text{sen } y$ dove X e Y sono variabili gaussiane di media θ_x e θ_y e deviazione standard pari all'errore di misura di 3° , e determinare i due indici di rifrazione per i quali il valore misurato $n = 1.52$ è rispettivamente il quantile di valore α e $1 - \alpha$ con $\alpha = 0.1585$. Il calcolo va eseguito per tutti i possibili valori di θ_x e θ_y . Il risultato, ottenuto con la routine **Gridrif** e mostrato anche in Fig. 7.13, fornisce l'intervallo:

$$n \in [1.10, 1.95] = 1.52_{-0.43}^{+0.42}, \quad CL \simeq 68.3\% . \quad (7.90)$$

Applichiamo ora il metodo *bootstrap* ed utilizziamo le (7.88, 7.89) sostituendo ai valori θ_x e θ_y delle (7.81) i valori t_x e t_y degli angoli misurati ed accumulando un istogramma di $Z = \text{sen } X / \text{sen } Y$ generando X e Y secondo le densità normali $g(x; t_x, \sigma_x)$ e $g(y; t_y, \sigma_y)$. Questo istogramma non è altro che quello di Fig. 7.11a), il quale va ora considerato come proveniente da una popolazione di densità $p(n; t_x, t_y)$. Con 100 000 eventi simulati, si sono ottenuti con la routine **Bootrif** 15 850 eventi negli intervalli $(-\infty, 1.17]$ e $[2.03, +\infty)$, cioè

$$n \in [1.17, 2.03] = 1.52_{-0.35}^{+0.51}, \quad CL \simeq 68.3\% , \quad (7.91)$$

che è un risultato differente rispetto al risultato corretto della (7.90). La ragione della discrepanza è la forte asimmetria della distribuzione, come si vede dalla Fig. 7.13.

Nel caso rappresentato in Fig.7.11 b) e relativo ad un errore di $\pm 0.5^\circ$, la densità delle misure ha una forma all'incirca invariante entro i valori degli angoli considerati e l'intervallo di confidenza diventa simmetrico rispetto al valore misurato. In questo caso, entrambi i metodi discussi precedentemente, *bootstrap* e griglia, danno lo stesso risultato:

$$n \in [1.45, 1.59] = 1.52 \pm 0.07 , \quad CL \simeq 68.3\% ,$$

dove l'errore degli estremi, che è di qualche per mille, non è riportato. La simulazione mostra che, per ottenere una misura affidabile, è opportuno avere errori di misura dell'ordine del mezzo grado. Questa è peraltro l'accuratezza tipica dei goniometri ottici ordinari.

Nel par. 11.6 applicheremo le tecniche di simulazione al caso estremamente importante della propagazione degli errori di misura.

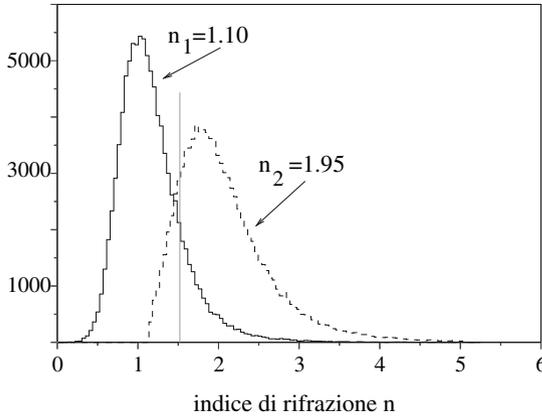


Fig. 7.13. Campioni simulati dalla densità $p(n; \theta_x, \theta_y)$ per i valori θ_x e θ_y corrispondenti agli estremi dell'intervallo (7.90). La linea è il valore misurato

7.11 Simulazione degli esperimenti di conteggio

Nei parr. 6.7 e 6.8 abbiamo introdotto gli intervalli di confidenza per la stima di probabilità e frequenze. Qui vogliamo esplorare, tramite simulazione, le proprietà di copertura statistica di questi intervalli.

Nel caso della stima della probabilità, avendo ottenuto x successi in n tentativi e avendo quindi osservato una frequenza $f = x/n$, l'intervallo con livello di confidenza CL può essere trovato con le formule (6.22, 6.28, 6.29, 6.31), che per comodità riportiamo:

- la formula generale per il caso binomiale, detta di Clopper-Pearson, è la stima frequentista $p \in [p_1, p_2]$, dove $CL = 1 - c_1 - c_2$ e p_1 e p_2 sono le soluzioni delle equazioni:

$$\sum_{k=x}^n \binom{n}{k} p_1^k (1 - p_1)^{n-k} = c_1, \tag{6.28}$$

$$\sum_{k=0}^x \binom{n}{k} p_2^k (1 - p_2)^{n-k} = c_2; \tag{6.29}$$

- l'approssimazione gaussiana con correzione di continuità, dove $f_{\pm} = (x \pm 0.5)/n$ fornisce l'intervallo (detto di Wilson):

$$p \in \frac{f_{\pm} + \frac{t_{\alpha/2}^2}{2n}}{\frac{t_{\alpha/2}^2}{n} + 1} \pm \frac{|t_{\alpha/2}| \sqrt{\frac{t_{\alpha/2}^2}{4n^2} + \frac{f_{\pm}(1 - f_{\pm})}{n}}}{\frac{t_{\alpha/2}^2}{n} + 1}; \tag{6.31}$$

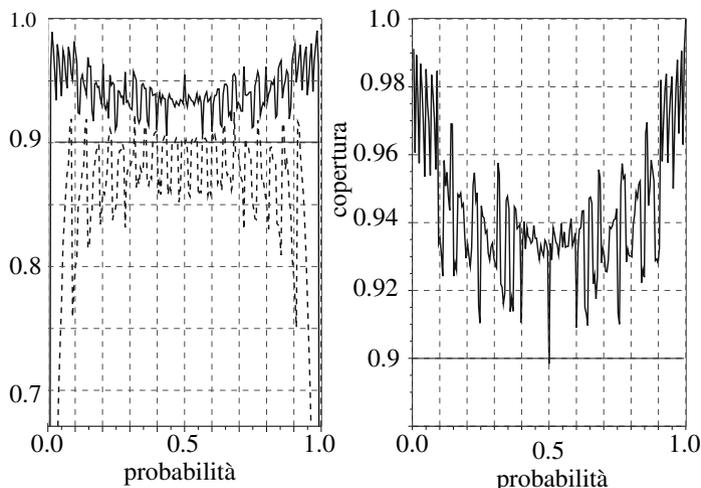


Fig. 7.14. Curve di copertura, nel caso della distribuzione binomiale con $n = 30$ e $CL = 0.90$, per gli intervalli di confidenza frequentisti (6.28, 6.29) (curva nel riquadro a destra), per l'intervallo della (6.31) (curva piena nel riquadro a sinistra) e per l'intervallo approssimato della (6.22) (curva tratteggiata a sinistra). Le formule frequentiste e la (6.31) forniscono coperture equivalenti

- l'approssimazione per grandi campioni porta all'intervallo usuale (detto di Wald) riportato in tutti i manuali di statistica:

$$p \in f \pm |t_{\alpha/2}| \sqrt{\frac{f(1-f)}{n}}. \quad (6.22)$$

Assegnato un numero di tentativi n ed una probabilità p , il codice di simulazione `Binocov`, riportato a pagina 297, estrae a caso un valore x di successi con la routine `grand`, calcola l'intervallo di confidenza con le formule appena considerate e conta quante volte si ha un successo, cioè quando il valore vero p è contenuto nell'intervallo. La curva di copertura si ottiene ripetendo la procedura 10 000 volte per ogni valore di p e riportando in grafico, con la routine `plot`, la frazione di inclusioni. I valori p_1 e p_2 delle formule frequentiste (6.28, 6.29) sono ottenuti per inversione della cumulativa della distribuzione binomiale dalla routine SCILAB `cdfbin`. I metodi numerici che permettono questa inversione sono descritti anche in [Gro02, PFTW92]. Per i casi limite $x = 0$ e $x = N$ vengono usate le formule (6.34, 6.35).

Le curve di copertura sono riportate in Fig. 7.14 per una distribuzione binomiale con $n = 30$ per intervalli con $CL = 0.90$, cui corrisponde un quantile gaussiano $|t_{\alpha/2}| = t_{1-\alpha/2} = 1.645$, dove $\alpha = 1 - CL$. La struttura delle curve appare irregolare a causa del valore discreto della variabile considerata, il numero x di successi. Il risultato, abbastanza sorprendente, mostra come la formula approssimata (6.22) (curva tratteggiata nel riquadro a sinistra della

Fig. (7.14)) fornisca una copertura assolutamente insoddisfacente, quasi sempre ben al di sotto del livello di confidenza assegnato. Contrariamente alla pratica usuale, questa formula andrebbe quindi usata solo per campioni abbastanza grandi, almeno con $n > 300$ [Rot10]. L'intervallo frequentista delle (6.28, 6.29), nel riquadro a destra, nonostante il comportamento irregolare della curva, fornisce una sovra-copertura corretta, con valori sempre sopra il livello di confidenza scelto. Un altro risultato interessante è che l'intervallo semplificato (6.31), che può essere calcolato agevolmente, fornisce una copertura equivalente a quella delle formule corrette frequentiste, che richiedono però l'uso di software statistico o programmi *ad hoc* per trovare i valori p_1, p_2 .

In conclusione, la simulazione ha mostrato che l'uso della (6.31) dovrebbe essere ben più diffuso di quanto non sia ora, perché *questa formula, a differenza della (6.22), fornisce intervalli affidabili già per $n > 10$* [BCD01, Rot10], come si può verificare direttamente con la routine `Binocov`.

```
// -----
//
// codice SCILAB Binocov: simulazione delle proprieta' di
// copertura di intervalli di confidenza per il caso bimomiale
//
// Routine SCILAB utilizzate:
// grand(1,1,'def')
//
// plot per il display dei risultati
// -----

// valori iniziali
Mn=200;
Nrep = 10000;
// scommentare se si vuole la stessa sequenza
// grand('setsd',5876756);           // seme dei numeri casuali

// azzeramenti e definizione dei vettori
prob=zeros(1,Mn);
y = zeros(1,Mn);
c = zeros(1,Mn);
w = zeros(1,Mn);
alo=0.5*(1.-CL);
aup=0.5*(1.-CL);
//      input da terminale
CL = input("livello di confidenza (es. 0.95):");
if(CL>1.) then
    CL=CL/100.;
end;
N = input("numero di prove:");
write(%io(2),"    CL = "+string(CL));
write(%io(2),"    N = "+string(N));
```

```

// determinazione del quantile t
tq = 0.5 + 0.5*CL;
t = cdfnor('X',0.,1.,tq,1-tq);
write(%io(2)," t = "+string(t));

// simulazione della copertura
pp=0.;
dp=1./(Mn-2);
for n=1:Mn,
    pp=pp+dp;
    if(pp>1) then pp=1.; end;
    prob(n)=pp;
    if modulo(n,20)==0 then write(%io(2)," prob = "+string(pp));
    end;
Ncount=0;
Nwilson=0;
Nclopper=0;
for k=1:Nrep, // ciclo di simulazione della copertura
    x = grand(1.,1.,'bin',N,pp);
    f=x/N;
    // formula di Wald
    if x==0 then
        p1=0.;
        p2=1.-(alo)**(1./N);
    elseif x==N then
        p1=(aup)**(1./N);
        p2=1.;
    else
        p1 = max(0.,f - t*sqrt(f*(1-f)/N));
        p2 = min(1.,f + t*sqrt(f*(1-f)/N));
    end;
    if((p1<=pp)&(pp<=p2)) then
        Ncount=Ncount+1;
    end;
    // formula di Wilson con correzione di continuita'
    fp=f+0.5/N;
    fm=f-0.5/N;
    t2n=t*t/N;
    w1 =
        max(0.,(fm+0.5*t2n-t*sqrt((0.25*t2n+fm*(1.-fm))/N))/(1.+t2n));
    w2 =
        min(1.,(fp+0.5*t2n+t*sqrt((0.25*t2n+fp*(1.-fp))/N))/(1.+t2n));
    if((w1<=pp)&(pp<=w2)) then
        Nwilson=Nwilson+1;
    end;
    // formula di Clopper Pearson
    if x==0 then
        c1=0.;

```

```

    c2=1.- (alo)**(1./N);
elseif x==N then
    c1= (aup)**(1./N);
    c2=1.;
else
    c1 = cdfbin('PrOmpr',1.-alo,alo,x-1,N);
    c2 = cdfbin('PrOmpr',aup,1.-aup,x,N);
end;
if((c1<=pp)&(pp<=c2)) then
    Ncloppe=Ncloppe+1;
end;
end;
y(n) = Ncount/Nrep;
w(n) = Nwilson/Nrep;
c(n) = Ncloppe/Nrep;
end;
// display dei risultati
clf();
subplot(1,2,1);
xgrid(2);
plot(prob,y,'--k',prob,w,'-k');
subplot(1,2,2);
xgrid(2);
plot(prob,c,'-k');

// -----

```

Nel caso della stima di una frequenza di conteggio μ a partire da un conteggio misurato x , le formule da utilizzare sono le (6.39, 6.42, 6.43), che anche qui riportiamo per comodità:

- la formula generale per il caso poissoniano è la stima frequentista dell'intervallo $\mu \in [\mu_1, \mu_2]$, dove i valori μ_1, μ_2 sono le soluzioni delle equazioni:

$$\sum_{k=x}^{\infty} \frac{\mu_1^k}{k!} \exp(-\mu_1) = c_1, \quad \sum_{k=0}^x \frac{\mu_2^k}{k!} \exp(-\mu_2) = c_2; \quad (6.39)$$

- in approssimazione gaussiana con correzione di continuità ($x_{\pm} = x \pm 0.5$) si ottiene:

$$\mu \in x_{\pm} + \frac{t_{\alpha/2}^2}{2} \pm |t_{\alpha/2}| \sqrt{x_{\pm} + \frac{t_{\alpha/2}^2}{4}}; \quad (6.42)$$

- quando x è grande, la (6.42) può essere sostituita dall'intervallo usuale:

$$\mu \in x \pm |t_{\alpha/2}| \sqrt{x}. \quad (6.43)$$

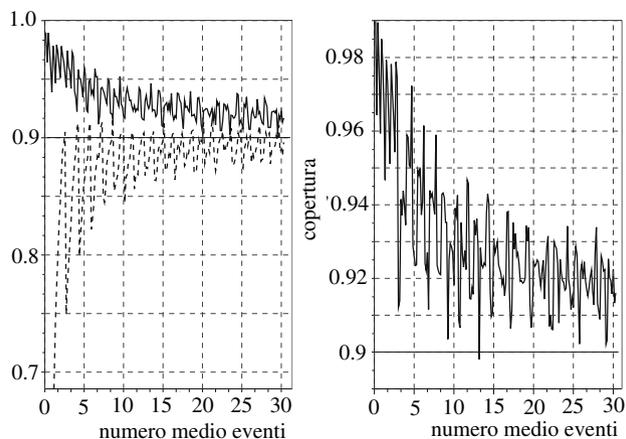


Fig. 7.15. Curve di copertura, nel caso della distribuzione di Poisson con $CL = 0.90$, per gli intervalli di confidenza frequentisti (6.39, 6.29) (curva nel riquadro a destra), per l'intervallo della (6.42) (curva piena nel riquadro a sinistra) e per l'intervallo approssimato della (6.43) (curva tratteggiata a sinistra). Le formule frequentiste e la (6.42) forniscono coperture equivalenti

Le curve di copertura di questi tre intervalli, per $\mu \leq 30$, sono riportate in Fig. 7.15. I grafici sono stati ottenuti con la nostra routine `Poisscov`, che qui non riportiamo perché essa ha una struttura del tutto simile a `Binocov`, che è stata appena descritta in dettaglio.

Come nel caso binomiale, anche qui vediamo come la copertura data dalla formula approssimata (6.43) sia assolutamente insoddisfacente. Le formule frequentiste e quella in approssimazione gaussiana con correzione di continuità sono in pratica equivalenti e danno un buona sovra-copertura (curve piene in Fig. 7.15). Anche qui, come nel caso binomiale, la simulazione mostra la superiorità dell'intervallo gaussiano con correzione di continuità (6.42) su quello approssimato (6.43), che andrebbe usato solo quando la media supera qualche decina [Rot10].

7.12 *Bootstrap* non parametrico

Nel paragrafi precedenti abbiamo visto come ricavare numericamente le proprietà di una variabile aleatoria attraverso campioni casuali ottenuti con modelli parametrici di densità di probabilità. Spesso, se non ci sono ulteriori informazioni, ai parametri del modello (medie, varianze o altro) si assegnano i valori misurati invece di quelli veri.

Questo metodo, detto *bootstrap parametrico*, è in realtà ancora più generale e si può spesso applicare anche quando *non si ha alcuna informazione sulla densità di probabilità della variabile aleatoria da esaminare* [DH99, DE83,

Efr79, Efr82, ET93, PFTW92]. Riprendiamo, ad esempio, l'esercizio 6.17 nel quale abbiamo determinato un intervallo di confidenza per il coefficiente di correlazione vero delle coppie dati di altezza/torace riportati nella Tab. 6.6, supponendo che la densità di probabilità bivariata $p(x, y)$ delle coppie di dati fosse data dalla funzione gaussiana bidimensionale.

Cerchiamo ora di determinare un intervallo di confidenza per il coefficiente di correlazione torace/altezza senza supporre una densità gaussiana per queste due variabili e non utilizzando pertanto la trasformazione di Fisher.

Con il metodo *bootstrap*, si assume che il campione ottenuto rappresenti la vera densità di probabilità dei dati da cui poi vengono estratti dei campioni casuali per la stima dei parametri da determinare. Un numero pari a $n = 1665$ coppie correlate (s_i, t_i) di valori torace/altezza viene quindi estratto (con reimmissione) dal campione sperimentale fino ad ottenere un nuovo campione "virtuale" di 1665 elementi. Alcuni dei dati originali così saranno presenti più di una volta all'interno di questo campione simulato, mentre alcuni altri non saranno considerati. Da esso, con le stesse operazioni dell'esercizio 6.17 ricaviamo tramite la (6.140) un valore r_{st}^* .

Ripetendo per un numero sufficientemente grande di volte questa operazione otteniamo un campione abbastanza numeroso di coefficienti r_{st}^* da cui ricavare l'intervallo di confidenza per il coefficiente di correlazione.

L'istogramma ottenuto con 10000 diversi valori simulati di r_{st}^* con la nostra routine `Bootcor` è mostrato nella Fig. 7.16. Da esso abbiamo ricavato l'intervallo di confidenza

$$r_{st}^* \in [0.221, 0.309] = 0.266_{-0.045}^{+0.043} \quad (CL \simeq 95.4\%)$$

che è esattamente coincidente con quello trovato nell'esercizio 6.17!

L'essenza del metodo si può sintetizzare come segue: *attraverso l'estrazione con reimmissione, generare da un unico campione reale un insieme di campioni fittizi di pari dimensione e determinare da questo insieme l'intervallo di confidenza per il parametro desiderato.*

La generazione di campioni fittizi o artificiali a partire dal campione reale (e aventi la stessa dimensione n) corrisponde a sostituire la densità di probabilità incognita $p(\mathbf{x})$ con una distribuzione di probabilità discreta $p_n^*(\mathbf{x})$ ad n componenti, la quale assegna ad ogni valore osservato la stessa probabilità $1/n$. Dal campione fittizio $(\mathbf{x}_1^*, \dots, \mathbf{x}_n^*)$, generato da questa distribuzione, si ricava un valore t^* ; questa operazione, ripetuta per r volte, fornisce un campione di valori $t_1^*, t_2^*, \dots, t_r^*$ da cui si stimano le proprietà della statistica $T = t(\mathbf{X}_1, \dots, \mathbf{X}_n)$ che si vuole utilizzare per stimare il parametro θ .

Consideriamo ora la funzione cumulativa $F_n^*(\mathbf{x})$ di $p_n^*(\mathbf{x})$ che assume i valori $(0, 1/n, 2/n, \dots, n/n)$ e che, in maniera formale, viene scritta come:

$$F_n^*(\mathbf{x}) = \frac{N(\mathbf{x}_i \leq \mathbf{x})}{n} \quad (7.92)$$

dove con il simbolo $N(A)$ abbiamo denotato quante volte si verifica l'evento A . Dalla (7.92) è facile convincersi che F^* segue una distribuzione binomiale di probabilità $p = F(\mathbf{x})$, dove F è la cumulativa della densità vera ma incognita

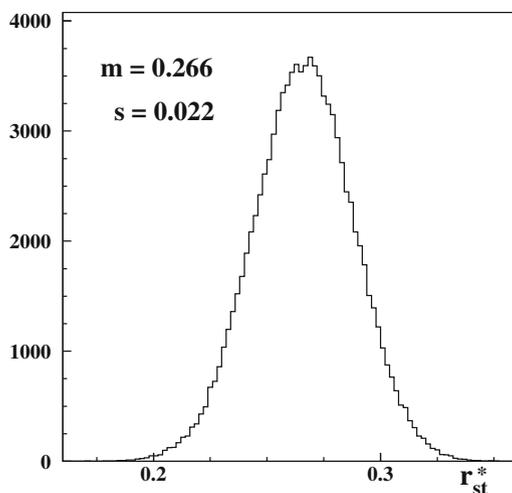


Fig. 7.16. Istogramma di bootstrap ottenuto con 10 000 campioni per r_{st}^*

$p(\mathbf{x})$. Da questa osservazione segue direttamente che:

$$E[F_n^*(\mathbf{x})] = nF(\mathbf{x}) = np$$

$$\text{Var}[F_n^*(\mathbf{x})] = \frac{1}{n}\{F(\mathbf{x})[1 - F(\mathbf{x})]\} = \frac{p(1-p)}{n}.$$

Dal teorema Limite Centrale e dalla legge dei grandi numeri è facile ricavare che, per n grande, $F_n^*(\mathbf{x})$ tende ad avere distribuzione gaussiana e ad approssimare sempre meglio la vera cumulativa $F(\mathbf{x})$. Da questa semplice proprietà dipende la bontà del metodo *bootstrap*, che a prima vista può apparire quasi miracolosa!²

Poiché nel *bootstrap* si sostituisce, come abbiamo visto, $F_n^*(\mathbf{x})$ a $F(\mathbf{x})$, gli errori di questo tipo di approssimazione sono dovuti alla differenza $(\hat{t} - \theta)$ tra il valore vero θ del parametro e la sua stima statistica corretta \hat{t} ed alla differenza $(\hat{t} - \hat{t}^*)$, dove \hat{t}^* è ricavato dal campione simulato di *bootstrap*.

Quest'ultima differenza è dovuta sia alle dimensioni finite del campione sperimentale e di quelli di *bootstrap* (rispettivamente n ed r) sia al particolare stimatore T utilizzato. Ad esempio, se T non è una funzione lineare, non è di solito vero che asintoticamente $\hat{t}^* = \hat{t}$ e questo porta ad un errore sistematico (*bias*) nella stima *bootstrap*.

Mentre è possibile ridurre l'errore statistico aumentando il numero r dei campioni *bootstrap* (mentre n è di solito fissato), più difficile risulta stimare quanto è importante l'errore sistematico per il particolare caso considerato. Occorrerebbe quindi trovare stimatori che siano il più possibile pivotali (che

² In effetti, a causa della sua "oltraggiosa" semplicità, il metodo, dopo i famosi lavori di Efron [Efr79, Efr82], ha impiegato un certo tempo ad affermarsi tra gli statistici.

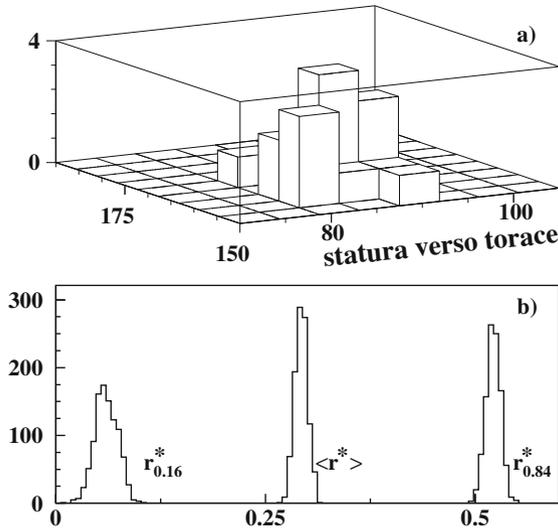


Fig. 7.17. a) Istogramma bidimensionale statura/torace per un sottogruppo di 20 soldati estratto dalla Tab. 6.6. b) istogramma con media e limiti dell'intervallo di confidenza al 68% dei campioni di *bootstrap*

abbiano cioè la stessa distribuzione sia rispetto a $F_n^*(\mathbf{x})$ che a $F(\mathbf{x})$, ma non sempre questo è fattibile. In [DH99] trovate un elenco dei vari metodi che vengono applicati con successo in questi casi.

In generale, tuttavia, il metodo *bootstrap* funziona abbastanza bene per diverse classi di stimatori anche quando non si hanno valori molto elevati sia di r che di n , come capita spesso nei casi reali.

Supponiamo, ad esempio, di conoscere statura e perimetro toracico solo per un gruppo casuale di $n = 20$ soldati estratto a caso dal campione completo di Tab. 6.6. Da questo campione, ottenuto con la routine `Bootcor20` e riportato nella parte a) di Fig. 7.17, abbiamo ricavato l'intervallo di confidenza di Fisher per il coefficiente di correlazione vero ρ ripetendo il procedimento dell'esercizio 6.17:

$$\rho \in [0.06, 0.507] = 0.307_{-0.247}^{+0.200}, \quad CL = 68\%. \quad (7.93)$$

Data la ridotta numerosità del campione, questa stima è necessariamente più rozza di quella ricavata in precedenza dato che, per la (6.143), l'ampiezza degli intervalli di confidenza varia proporzionalmente a $\propto 1/\sqrt{n-3}$.

Supponiamo ora di non conoscere la teoria di Fisher, che ha portato alla (7.93). Possiamo procedere comunque applicando il metodo *bootstrap*. Utilizzando ancora la routine `Bootcor20`, si possono generare 1000 replicazioni del campione, ognuna con 20 elementi, e calcolare l'intervallo di stima di r^* . Se utilizzate la routine, noterete che i due intervalli, quello di Fisher e quello ottenuto con i mille campioni *bootstrap*, sono molto simili.

Per verificare la stabilità del metodo, abbiamo ripetuto 1000 volte questa procedura ricavando il valor medio $\langle r^* \rangle$ e gli estremi di minimo ($r_{0.16}^*$) e massimo ($r_{0.84}^*$) dell'intervallo di confidenza con $CL = 68\%$ ³. Abbiamo riportato i diversi valori via via ottenuti nella parte b) di Fig. 7.17.

Nonostante il basso valore $n = 20$ che abbiamo scelto, i valori medi delle quantità in istogramma differiscono solo di qualche % dai corrispondenti valori della (7.93), approssimazione questa senz'altro accettabile per quasi tutti i problemi concreti. Inoltre, già con $r = 1000$, la dispersione di $\langle r^* \rangle$, $r_{0.16}^*$ e $r_{0.84}^*$ è abbastanza contenuta; la maggior parte dei campioni *bootstrap* fornisce valori per l'intervallo di confidenza uguali, entro il 5%, a quelli della (7.93); solo una piccola percentuale dei campioni si discosta (anche se di poco) di oltre il 10%.

Generalizzando, possiamo dire che, tutte le volte che non sono note le proprietà statistiche di una certa quantità che si vuole stimare da un campione, *l'intervallo di confidenza può essere stimato col metodo bootstrap*.

Va tuttavia tenuto presente che non si può sapere in anticipo quali siano i “cattivi” campioni *bootstrap*, quelli cioè che portano ad un grande errore sistematico, e questa rimane in definitiva la limitazione principale del procedimento. Questi problemi di applicabilità sono discussi in dettaglio in [DH99]. Noi suggeriamo, per verificare caso per caso se sia applicabile il metodo *bootstrap*, di collaudare la procedura su dati simulati, verificando quanto i risultati che si ottengono siano compatibili coi valori veri, che in questo caso sono noti.

Problemi

7.1. Simulare il comportamento di un giocatore che punta sui numeri del lotto che ritardano da più di 90 settimane su una certa ruota. Questa strategia aumenta la probabilità di vincita?

7.2. Risolvere con un codice di simulazione il problema di Monty 1.1 di pag. 33.

7.3. Risolvere con un codice di simulazione il problema dell'incontro 1.7 di pag. 34.

7.4. Scrivere un codice di simulazione che calcoli la probabilità che la distanza tra due punti estratti a caso in modo uniforme all'interno di un cerchio sia minore del raggio.

7.5. Generare, mediante l'algoritmo 7.3, variabili casuali x distribuite secondo la densità di probabilità semi-gaussiana:

$$p(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2} \quad \forall x \geq 0$$

utilizzando la funzione $f(x) = ke^{-x}$. Trovare il valore di k che dà la massima efficienza di generazione. Con il valore di x così ottenuto, generare una variabile $Z \sim N(0, 1)$.

³ Per ridurre i tempi di calcolo, qui abbiamo utilizzato un codice FORTRAN.

7.6. Determinare, con un codice di simulazione, i livelli di confidenza dell'equazione (6.21) di pag. 189 nel caso $n = 5$, $p = 0.25$ e $t = 1, 2, 3$.

7.7. Utilizzando il teorema Limite Centrale trovare un algoritmo di generazione di variabili gaussiane standard.

7.8. Generare una coppia di variabili normali standard X e Y aventi coefficiente di correlazione lineare ρ .

7.9. Come si può utilizzare l'algoritmo del rigetto per estrarre a caso uniformemente dei punti all'interno di un cerchio di raggio R ?

7.10. Generare l'istogramma della variabile $Z = Y/X$ dove X e Y sono variabili gaussiane standard.

7.11. In una popolazione la statura è una variabile gaussiana pari a $\langle X \rangle \pm \sigma[X] = 175 \pm 8$ cm per gli uomini e $\langle Y \rangle \pm \sigma[Y] = 165 \pm 6$ cm per le donne. Sapendo che tra la statura del marito e della moglie esiste una correlazione positiva $\rho = 0.5$ (le coppie tendono ad avere statura omogenea), trovare, mediante un codice di simulazione, la percentuale di coppie con uomo e donna più alti di 180 e 170 cm rispettivamente (si vedano anche i problemi 4.6, 4.7 di pagina 148).

7.12. Risolvere con un codice di simulazione il problema 5.9 di pagina 177.

7.13. Un piano infinito è diviso da linee parallele a distanza unitaria. Un ago di lunghezza unitaria è gettato in modo casuale sul piano. Si può mostrare che la probabilità che l'ago cada a cavallo di una linea è $p = 2/\pi$ (problema di Buffon [Gne76]). Stimare il valore di π con un codice Monte Carlo.

7.14. Nel problema di asimmetria destra-sinistra (o alto-basso) gli eventi possono realizzarsi "a sinistra" con probabilità P oppure "a destra" con probabilità $1 - P$. Simulare 5000 esperimenti destra-sinistra, e contare il numero di volte in cui, su $N = 50$ prove, si hanno n_s eventi a sinistra ed $n_d = N - n_s$ eventi a destra. Fare l'istogramma della asimmetria $A = (n_s - n_d)/N$, e confrontare la deviazione standard dei dati con quella vera discussa nel problema 11.8 di pagina 491.

7.15. Verificare, con la routine `Binocov`, avendo fissato $CL = 95\%$, a partire da quale valore di n la formula approssimata (6.22) garantisce una differenza tra copertura e $CL < 5\%$.

7.16. Verificare, con la routine `Poisscov`, avendo fissato $CL = 95\%$, a partire da quale valore di μ la formula approssimata (6.43) fornisce una differenza tra copertura e $CL < 2\%$.

7.17. Le quotazioni medie (in \$) delle azioni (a) e delle obbligazioni (o) della borsa di New York negli anni 1950-1959 sono riportati nella tabella (dati da [Spi76]):

	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959
a	35.2	39.9	41.9	43.2	40.1	53.3	54.1	49.1	40.7	55.2
o	102.4	100.9	97.4	97.8	98.3	100.1	97.1	91.6	94.85	94.65

Una teoria economica prevede una correlazione negativa tra le quotazioni delle azioni e delle obbligazioni. Utilizzare il metodo *bootstrap* non parametrico per verificare il modello ad un livello del 5%.

Applicazioni del metodo Monte Carlo

PRIMO POSTULATO DI THUMB

È meglio risolvere un problema con forte approssimazione e sapere la verità più o meno al 10 per cento, che esigere una soluzione esatta e non sapere la verità per niente

Arthur Bloch, *Il secondo libro di Murphy*

8.1 Introduzione

I campi di applicazione del metodo Monte Carlo (nel seguito abbreviato spesso come MC) sono praticamente illimitati, ed è veramente difficile affrontare in modo organico un problema di questo genere entro un ambito ristretto come quello di un capitolo. Tuttavia, abbiamo pensato che fosse utile presentarvi qui almeno la risoluzione di alcuni problemi, partendo dall'impostazione generale fino all'estremo dettaglio del codice di simulazione. In questo modo potrete acquisire una completa padronanza della tecnica, così da essere in grado di applicare e adattare con successo il metodo ai vostri problemi specifici.

Gli esempi trattati in questo capitolo sono volutamente scelti in modo arbitrario e disomogeneo: la diffusione di particelle nei materiali (tipico problema di fisica sperimentale, chimica o ingegneria), il calcolo del numero ottimale di addetti ad un impianto (una istruttiva applicazione alla gestione industriale), alcune applicazioni dell'algoritmo di Metropolis (studio dei sistemi aventi un gran numero di componenti identici, che sono studiati in economia, ingegneria, fisica e chimica), ed infine il calcolo del valore di un integrale definito (qui siamo soprattutto nella fisica teorica e nella matematica).

8.2 Studio dei fenomeni di diffusione

La propagazione di un insieme di particelle in un dato materiale, neutroni in un reattore nucleare, elettroni in un metallo o in un semiconduttore, fotoni nell'atmosfera, viene di solito denominata con il termine di *diffusione*.

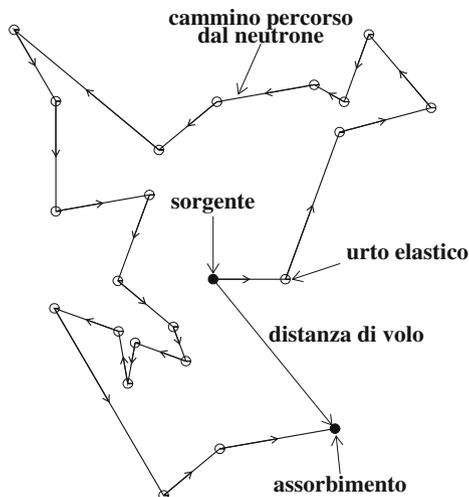


Fig. 8.1. Rappresentazione schematica della traiettoria percorsa da un neutrone termico durante il processo di diffusione

In questo tipo di processi ciascuna particella, partendo da uno stato iniziale, percorre durante un certo intervallo di tempo una traiettoria predeterminata fino ad un istante casuale in cui essa subisce un *urto*, cioè un'interazione con il mezzo attraversato (con la possibile produzione, in alcuni casi, di particelle secondarie), per effetto del quale cambia, sempre in maniera aleatoria, direzione e modulo della velocità e prosegue il proprio cammino fino alla successiva collisione. Lo schema generale è quello di Fig. 8.1.

In linea di principio, questo sistema può essere descritto mediante l'equazione di Boltzmann (o del trasporto), la cui trattazione è riportata in molti testi avanzati di fisica, chimica o ingegneria, come [Lam66].

Sfortunatamente, come ha giustamente commentato J.Lamarsh [Lam66], *“è molto più facile ricavare l'equazione di Boltzmann che risolverla”*, poiché essa ha una complessa forma integro-differenziale, dove compaiono parametri che, per la complicata configurazione dei sistemi reali, sono estremamente variabili in funzione sia dello spazio sia dell'energia delle particelle.

Il metodo MC resta così quasi sempre l'unico approccio disponibile per risolvere questi problemi e, nonostante la grande varietà dei processi di diffusione, è anche possibile delineare uno schema di simulazione “a blocchi” di validità quasi universale.

Consideriamo, come esempio significativo, il caso della diffusione di neutroni all'interno di un materiale. Qui, analogamente a quanto avviene per i raggi γ e per le altre particelle elettricamente neutre, il calcolo delle traiettorie si presenta molto semplice poiché i neutroni seguono, tra due collisioni successive con i nuclei atomici del mezzo attraversato, percorsi in linea retta.

Consideriamo, come in Fig. 8.1, una sorgente puntiforme che emette in modo isotropo neutroni di energia cinetica fissa e pari a 0.0038 elettron-volts¹ (essi fanno parte della categoria dei neutroni “termici”, così chiamati perché in equilibrio termico con la materia alla temperatura ambiente di 27° C), situata al centro di una sfera omogenea di raggio infinito, costituita dal materiale diffusore.

Il procedimento MC che vi presentiamo può essere suddiviso in vari passi, che sono comuni a tutti i calcoli diffusivi (vedere anche [RvN63]):

a) *Geometria, materiali ed interazioni.*

Poiché il mezzo in cui si propagano i neutroni è infinito ed omogeneo, il nostro sistema risulta molto semplificato e non è quindi necessario introdurre alcun dato geometrico.

Abbiamo invece bisogno di conoscere i parametri che definiscono le interazioni dei neutroni con i nuclei atomici. Esse (al pari di quanto avviene per tutte le altre particelle) vengono descritte mediante una grandezza, denominata “sezione d’urto macroscopica totale” (o, più in breve, Σ_T), che dà la probabilità, per unità di lunghezza percorsa, di avere un qualsiasi tipo di reazione.

Σ_T è una proprietà intrinseca di ogni materiale, rimane costante in mezzi omogenei, dipende dalla velocità ma non dal cammino precedente della particella considerata. In caso di una sorgente non monoenergetica, occorre perciò conoscere con sufficiente precisione l’andamento di Σ_T nell’intervallo di energia considerato. Supponiamo, come è ragionevole fare per i neutroni termici, che le interazioni possibili siano di due tipi: assorbimento oppure diffusione elastica (cambio di direzione senza perdita di energia); con questa ipotesi la sezione d’urto macroscopica totale viene ovviamente decomposta come:

$$\Sigma_T = \Sigma_a + \Sigma_{el} , \quad (8.1)$$

dove Σ_a e Σ_{el} sono le sezioni d’urto macroscopiche dei due processi considerati. Dato che i neutroni che consideriamo sono monoenergetici, abbiamo bisogno di soli due numeri, che leggeremo come parametri iniziali, per definire le loro interazioni.

Altri dati di cui avremo bisogno (e che leggeremo sempre all’inizio del programma) sono il numero di massa del nucleo diffusore (necessario per il calcolo delle traiettorie) e la velocità dei neutroni, in quanto siamo interessati anche a sapere il tempo necessario per il loro assorbimento.

b) *Cinematica.*

Il punto di emissione è prefissato e, per comodità, lo porremo al centro del sistema di coordinate, mentre invece la direzione di volo dei neutroni

¹ L’elettron-volt, o più brevemente eV, è un’unità di lavoro molto usata nella fisica atomica e nucleare ed è, per definizione, uguale all’energia cinetica acquisita da un elettrone quando passa tra due punti fra cui vi è una differenza di potenziale di 1 Volt. È facile mostrare che vale la relazione: $1\text{eV} = 1.60210^{-12}\text{erg}$.

uscenti dalla sorgente viene determinata, in base alla (7.33), generando a caso gli angoli φ e ϑ attraverso le relazioni:

$$\begin{cases} \varphi = 2\pi\xi_1 \\ \vartheta = \arccos(1 - 2\xi_2) . \end{cases} \quad (8.2)$$

Si calcolano poi i coseni direttori:

$$\begin{cases} \alpha = \text{sen } \vartheta \cos \varphi \\ \beta = \text{sen } \vartheta \text{sen } \varphi \\ \gamma = \cos \vartheta \end{cases} \quad (8.3)$$

che, con la conoscenza del punto di partenza, permettono di definire univocamente i parametri che descrivono la traiettoria iniziale di ogni particella.

c) *Tracciamento.*

In base allo schema di Fig. 8.1, poiché tra due collisioni successive il neutrone non subisce altre interazioni, un passo del tracciamento coincide in questo caso con la distanza di volo. Essa può venire calcolata notando che, in base alla definizione stessa di sezione d'urto macroscopica, l'interazione del neutrone in uno qualunque dei modi permessi è un processo stocastico avente le stesse caratteristiche di quelli che abbiamo trattato nel par. 3.8. Il neutrone, infatti, durante il suo percorso nel mezzo attraversato, subisce urti (cioè "genera" eventi) non correlati, discreti e con probabilità Σ_T costante; per questo, analogamente alla (3.48), la probabilità di avere una distanza x tra un urto e l'altro risulta:

$$p(x) dx = \Sigma_T e^{-x\Sigma_T} dx . \quad (8.4)$$

Questa distribuzione (la cui media $1/\Sigma_T$ rappresenta il libero cammino medio tra 2 urti successivi), attraverso la relazione (3.91), che in questo caso diventa:

$$d = -\frac{1}{\Sigma_T} \ln \xi_3 , \quad (8.5)$$

permette di associare ad ogni neutrone una distanza d , pari alla lunghezza del "passo" della simulazione, tra il punto di partenza e quello dove avviene la collisione successiva.

Ora, conoscendo gli angoli di emissione e la distanza percorsa, siamo in grado di calcolare le coordinate del punto in cui avviene la reazione del neutrone con un nucleo del materiale diffusore. A questo punto dobbiamo valutare gli effetti di questa interazione sulla traiettoria della particella, sapendo che la probabilità che essa sia assorbita è data da (Σ_a/Σ_T) , mentre quella di essere diffusa è (Σ_{el}/Σ_T) . La scelta tra queste due alternative viene compiuta estraendo un nuovo numero casuale ξ_4 : se $0 \leq \xi_4 \leq (\Sigma_a/\Sigma_T)$ il neutrone è assorbito e l'evento generato ha termine; se invece $(\Sigma_a/\Sigma_T) < \xi_4 \leq 1$ si ha una diffusione elastica ed occorre generare una nuova direzione di volo, come mostrato in Fig. 8.1.

Per i neutroni termici il processo di urto, rispetto alla direzione incidente, è isotropo nel sistema del centro di massa neutrone-nucleo²; si devono perciò generare in tale sistema gli angoli azimutale (φ_{cm}) e polare (ϑ_{cm}) di emissione:

$$\begin{cases} \cos \vartheta_{cm} = 1 - 2\xi_5 \\ \varphi_{cm} = 2\pi\xi_6 \end{cases} \quad (8.6)$$

ed effettuare la trasformazione nel sistema del laboratorio, che risulta essere [Lam66]:

$$\begin{cases} \varphi = \varphi_{cm} \\ \cos \vartheta = \frac{1 + A \cos \vartheta_{cm}}{\sqrt{A^2 + 2A \cos \vartheta_{cm} + 1}} \end{cases} \quad (8.7)$$

dove con A si è indicato il numero di massa dei nuclei bersaglio.

I coseni direttori della nuova direzione di volo del neutrone (α' , β' , γ') vengono quindi calcolati mediante le formule:

$$\begin{cases} \alpha' = \mu\alpha + a(\alpha\gamma \operatorname{sen} \varphi + \beta \cos \varphi) \\ \beta' = \mu\beta + a(\beta\gamma \operatorname{sen} \varphi - \alpha \cos \varphi) \\ \gamma' = \mu\gamma - a(1 - \gamma^2) \operatorname{sen} \varphi \end{cases} \quad (8.8)$$

dove:

$$a = \sqrt{\frac{1 - \mu^2}{1 - \gamma^2}} \quad ; \quad \mu = \cos \vartheta \quad ; \quad |\gamma| \neq 1. \quad (8.9)$$

Se $|\gamma| = 1$ si ha invece:

$$\begin{cases} \alpha' = \gamma b \cos \varphi \\ \beta' = b \operatorname{sen} \varphi \\ \gamma' = \gamma \mu \end{cases} \quad (8.10)$$

² Consideriamo un insieme di N particelle di massa m_1, m_2, \dots, m_N aventi rispettivamente una velocità, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N$ rispetto ad un sistema di riferimento predeterminato. Il punto definito mediante l'equazione:

$$\mathbf{r}_{cm} = \frac{\sum_{i=1}^N m_i \cdot \mathbf{r}_i}{\sum_{i=1}^N m_i}$$

(dove \mathbf{r}_i è la distanza dell' i -esima particella dal centro del sistema di coordinate) è detto centro di massa dell'insieme. In modo analogo si può definire la sua velocità come:

$$\mathbf{v}_{cm} = \frac{\sum_{i=1}^N m_i \cdot \mathbf{v}_i}{\sum_{i=1}^N m_i}.$$

Un sistema di riferimento in cui $\mathbf{v}_{cm} = 0$ viene chiamato sistema del centro di massa dell'insieme di particelle. Esso è spesso usato nella fisica poiché permette di trattare in maniera semplice, come avviene nel nostro caso, molte reazioni nucleari.

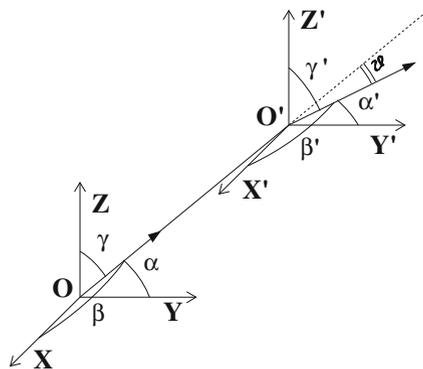


Fig. 8.2. Le terne di coseni direttori (α, β, γ) e $(\alpha', \beta', \gamma')$ rappresentano rispettivamente la direzione di volo del neutrone prima e dopo l'urto che avviene nel punto O' mentre con ϑ è indicato l'angolo polare di diffusione del neutrone rispetto alla direzione incidente

con $b = \sqrt{1 - \mu^2}$. Nella Fig. 8.2 è rappresentata la geometria del processo di urto nel sistema del laboratorio. Per la dimostrazione di queste formule potete consultare le nostre pagine web [RPP].

d) *Memorizzazione dell'evento.*

Nel caso di diffusione elastica, prima di ritornare al punto precedente per continuare a seguire il cammino della particella considerata, si aggiornano le quantità più rilevanti della traiettoria:

- la distanza percorsa totale d_i tra la sorgente ed il punto in cui avviene l'ultima collisione (che supponiamo essere l' i -esima) considerata;
- le proiezioni di d_i sugli assi cartesiani:

$$x_i = x_{i-1} + d_i \alpha \quad (8.11)$$

$$y_i = y_{i-1} + d_i \beta \quad (8.12)$$

$$z_i = z_{i-1} + d_i \gamma \quad (8.13)$$

(ovviamente $x_0 = y_0 = z_0 = 0$);

- il tempo di volo d_i/v (v è il modulo della velocità dei neutroni considerati; nel nostro caso $v = 2.2 \cdot 10^5$ cm/s);
- il numero delle collisioni subite.

Quando invece il neutrone è assorbito, dopo un numero k di urti, si calcola la quantità:

$$r = \sqrt{x_k^2 + y_k^2 + z_k^2} \quad (8.14)$$

che rappresenta la distanza tra la sorgente ed il punto finale di arresto (vedere la Fig. 8.2).

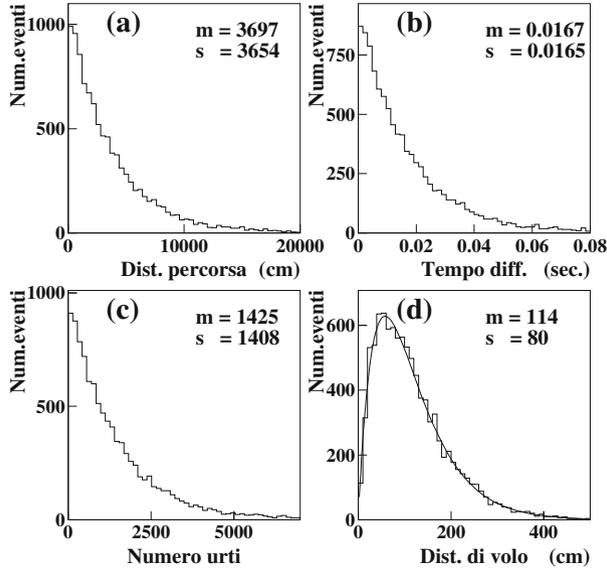


Fig. 8.3. Istogrammi finali ottenuti nella simulazione della diffusione di 10 000 neutroni nel carbonio

Il procedimento ha termine quando è stato generato il numero prefissato di particelle. Nel nostro sito web [RPP] è riportato il codice `Neutroni` che esegue questa simulazione.

I risultati forniti dal programma generando 10 000 eventi, nel caso in cui il nucleo diffusore è il carbonio ($A = 12$, $\Sigma_{el} = 0.3851 \text{ cm}^{-1}$, $\Sigma_a = 2.728 \cdot 10^{-4} \text{ cm}^{-1}$), sono riportati negli istogrammi di Fig. 8.3. Dalla analisi di questi dati, oltre a ricavare, dai valori medi degli istogrammi, i principali parametri che caratterizzano la diffusione dei neutroni nel carbonio, possiamo ottenere diverse altre utili indicazioni su questo processo.

Prendiamo in esame l'istogramma (a) di Fig. 8.3 (distanza a zig-zag percorsa dai neutroni): l'errore sulla media m dell'istogramma è $s/\sqrt{10\,000} \simeq 36$ (s è la deviazione standard dell'istogramma). Entro l'errore statistico, c'è coincidenza tra media e deviazione standard e l'istogramma proviene perciò da una densità di probabilità esponenziale negativa (vedere par. 3.3). È facile capire questa caratteristica se osserviamo che l'unica quantità che influenza il cammino percorso è Σ_a ; richiamando le stesse considerazioni che ci hanno portato alla (8.4), la densità di probabilità iniziale deve quindi essere:

$$p(x) = \Sigma_a e^{-x\Sigma_a} \quad (8.15)$$

ed infatti, sempre entro gli errori statistici, è verificata la relazione:

$$\mu_H = \sigma_H = \frac{1}{\Sigma_a}. \quad (8.16)$$

Simili considerazioni valgono anche per l'istogramma (b), che rappresenta il tempo intercorso tra emissione ed assorbimento. Tale grandezza non è altro che il rapporto tra la distanza percorsa x e la velocità v dei neutroni, per cui questa distribuzione ha le stesse caratteristiche della precedente. Per ricavare esplicitamente formula della densità $p(t)$ (con $t = x/v$) associata a questo istogramma basta perciò applicare la legge di trasformazione (5.7):

$$p(t) = p(x(t)) \frac{dx}{dt} = v \Sigma_a e^{-tv \Sigma_a}. \quad (8.17)$$

Passiamo ora a considerare la distribuzione del numero di urti subiti da ogni particella prima di essere assorbita (istogramma (c)). A prima vista, tenendo conto che ogni interazione può essere considerata un evento raro, uniforme e stazionario, ci si potrebbe aspettare, per tale istogramma, una distribuzione poissoniana ma, anche in questo caso, c'è coincidenza tra media e deviazione standard, caratteristica indicazione di una distribuzione esponenziale negativa. Contrariamente a quanto avviene, per esempio, nel lancio di una moneta, in cui la probabilità di avere testa o croce non è influenzata dai lanci precedenti, l'alternativa diffusione elastica-assorbimento ha effetto sulle interazioni successive poiché, se si verifica il secondo processo, ha termine il cammino del neutrone e viene così esclusa la possibilità di avere altri urti. In base a queste considerazioni è facile convincersi che risulta poissoniano solo il numero di urti subito da ogni particella *per unità di lunghezza percorsa*: questa grandezza viene influenzata solo dalla diffusione elastica ed in tale processo ogni collisione è sicuramente indipendente da quella precedente.

La distribuzione del numero di urti segue invece la legge geometrica (3.7) di pagina 73 con probabilità di successo, cioè di assorbimento, $p = \Sigma_a / \Sigma_T$. Come abbiamo notato nel par. 3.8, questa distribuzione, quando $p \ll 1$ (come nel nostro caso), è praticamente indistinguibile dall'esponenziale con parametro p (3.50); per questo motivo, sempre entro gli errori statistici, media e varianza dell'istogramma sono pari a Σ_T / Σ_a .

Per interpretare l'istogramma (d), notiamo che la distanza di volo (8.14) è il modulo di un vettore le cui componenti x_k , y_k , z_k sono realizzazioni di una somma di variabili casuali indipendenti e provenienti da popolazioni con varianza finita:

$$x_k = \sum_{i=1}^{N_{urti}} d_i \alpha_i; \quad y_k = \sum_{i=1}^{N_{urti}} d_i \beta_i; \quad z_k = \sum_{i=1}^{N_{urti}} d_i \gamma_i. \quad (8.18)$$

Per il teorema Limite Centrale già tante volte richiamato, se x_k , y_k e z_k fossero variabili gaussiane, per il teorema 3.3 di Pearson, la densità di probabilità $p(r)$ della distanza di volo r sarebbe una distribuzione maxwelliana. In questo caso specifico non è però verificata una delle condizioni di applicazione del teorema Limite Centrale: è vero che $d_i \alpha_i$, $d_i \beta_i$ e $d_i \gamma_i$ sono indipendenti e provengono da popolazioni con varianza finita, ma nelle sommatorie (8.18) il numero N_{urti} di collisioni per evento non è fisso ma è anch'esso una variabile casuale. Variabili di questo tipo vengono denominate *somme aleatorie*;

per derivarne la distribuzioni di probabilità sono necessari metodi matematici abbastanza sofisticati, che sono sviluppati in alcuni testi come [Pap77] e che noi tralasciamo, scrivendo solo il risultato finale [Lam66]:

$$p(r) = \frac{r}{L^2} e^{-r/L}, \quad L = \sqrt{\frac{\Sigma_{el}}{3\Sigma_a\Sigma_t^2}}, \quad (8.19)$$

in cui il parametro caratteristico L (nel nostro caso $L = 71.4$ cm) viene denominato, con ovvio significato, *lunghezza di diffusione*. È facile verificare che la media e la deviazione standard della distribuzione risultano rispettivamente $\mu = 2L = 112.6$ e $\sigma = \sqrt{2}L = 79.4$, valori in accordo, entro gli errori statistici, con quelli che abbiamo ricavato dall'istogramma simulato (d). In questo grafico abbiamo anche riportato la funzione di *best fit* spiegata nel cap. 9 calcolata prendendo la formula (8.19) come modello della popolazione statistica dell'istogramma. L'accordo migliore con la distribuzione simulata è stato ottenuto per $L = 56.8 \pm 0.4$ ed il valore del χ^2 risultante dal programma di *fit* è stato $\chi^2 = 1.0$, indicazione che il modello assunto interpreta correttamente la forma dell'istogramma.

8.3 Simulazione dei processi stocastici

Con il metodo MC diventa molto conveniente anche lo studio dell'evoluzione temporale di un qualsiasi processo stocastico. Come esempio, prenderemo in esame un classico argomento di ricerca operativa³: lo studio dei fenomeni di attesa.

Le code di persone davanti ad uno sportello sono il caso più noto e familiare dei sistemi che stiamo per prendere in esame: da esse la teoria dei fenomeni di attesa ha preso il nome e gran parte della terminologia, ma le classi di processi che si possono così analizzare sono assai numerose e molte di esse riguardano da vicino anche diversi altri momenti della nostra vita quotidiana. La regolazione del traffico automobilistico urbano o di una stazione ferroviaria, la programmazione dei voli di una compagnia aerea o del numero di casse in un supermercato, il funzionamento di un magazzino o del sistema di manutenzione di uno stabilimento, sono solo alcuni dei problemi che possono essere interpretati e risolti mediante la teoria delle code, la cui conoscenza è perciò indispensabile quando ci si deve occupare delle metodologie legate alla gestione industriale, economica o sociale.

La struttura di un fenomeno di attesa comprende, in maniera schematica, un certo numero di "stazioni di servizio" o "canali" (che possono essere impiegati oppure operai addetti ad una certa mansione, linee di comunicazione,

³ La ricerca operativa è quel settore delle scienze economiche che si occupa dell'analisi dei sistemi produttivi al fine di fornire soluzioni adatte per una organizzazione ottimale.

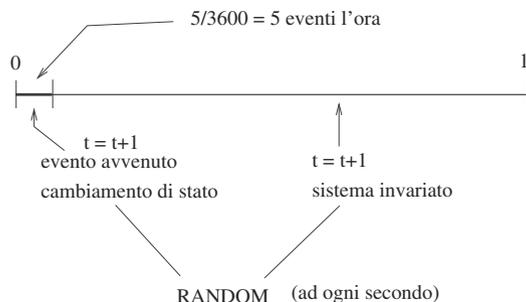


Fig. 8.4. Simulazione sincrona di un evento di probabilità pari a 5 eventi/ora

posti in un parcheggio, ecc.) che svolgono il lavoro richiesto dai “clienti” (persone in attesa di un servizio, macchine da riparare, merci da spedire, ecc.). Il sistema servente, quando arriva un nuovo cliente, può trovarsi in due diverse configurazioni:

- vi è almeno una stazione libera: la domanda del cliente viene subito eseguita e questa operazione occupa una o più stazioni per un certo periodo di tempo;
- tutte le stazioni sono impegnate: il cliente in attesa di essere “servito” viene messo in una “coda”, le cui caratteristiche variano di molto al variare del sistema considerato.

Lo scopo di questo studio è ovviamente quello di valutare qual è il rendimento complessivo del processo tenendo conto dei costi legati sia ai tempi di attesa dei clienti sia al numero o ai periodi di inattività delle stazioni serventi (per trattazione più completa, si può consultare [CS61]).

La parte più difficile per la simulazione di questo tipo di processi consiste nel definire in modo corretto un “orologio” che riproduca la sequenza temporale continua degli eventi. Tale problema può essere risolto in due modi diversi, che nella letteratura vengono denominati *simulazione sincrona* e *simulazione asincrona* [BFS87].

Nella simulazione sincrona il problema dell’orologio viene risolto in modo estremamente semplice: si definisce una unità di tempo (il secondo, il minuto, l’ora . . .) arbitraria purché piccola rispetto ai tempi caratteristici del sistema e ad ogni ciclo di scansione si fa avanzare l’orologio di una unità. La simulazione è così *discreta*, ma si può anche arrivare alla simulazione di un processo continuo quanto più l’unità di avanzamento temporale in un ciclo è piccola rispetto ai tempi di transizione di stato del sistema.

Conoscendo allora le probabilità temporali $0 < \lambda_i \ll 1$ (nell’unità di misura scelta) di ogni possibile evento o cambiamento di stato del sistema, si esegue un’ estrazione casuale di un insieme di variabili aleatorie uniformi $0 \leq \xi_i \leq 1$ e si fa accadere l’evento se $\xi_i < \lambda_i$, altrimenti si lascia invariato il sistema. Per ogni intervallo di tempo prefissato Δt si calcolano quindi le medie tem-

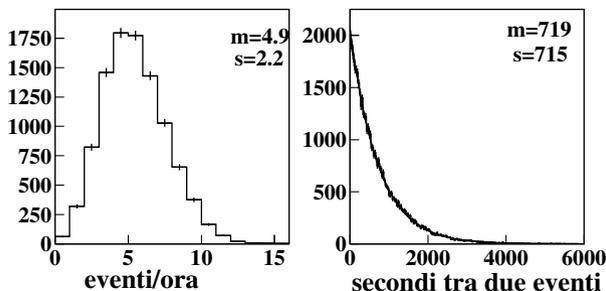


Fig. 8.5. Istogrammi del numero medio di eventi all'ora e dei tempi tra due eventi successivi generati con la simulazione sincrona di Fig. 8.4. La linea continua riportata nell'istogramma dei tempi è il *fit* dei dati con la legge esponenziale negativa

porali e le deviazioni standard delle grandezze caratteristiche del sistema e se ne fanno eventualmente grafici ed istogrammi alla fine della simulazione. La Fig. 8.4 mostra come fare per simulare il guasto di una macchina che in media fa registrare 5 guasti all'ora, scegliendo come unità di tempo il secondo.

Poniamoci ora la domanda fondamentale: la simulazione sincrona è in accordo con la legge generale dei fenomeni stocastici indipendenti, che prevede un numero di eventi poissoniano ed eventi contigui separati da intervalli di tempo esponenziali? La risposta è sì purché valga la condizione $\lambda_i \ll 1$. Infatti, in questo caso la probabilità di un numero n di istanti di tempo elementari tra eventi contigui è data dalla densità geometrica (3.50), che, come ripetutamente notato, quando $p \ll 1$, diventa un'ottima approssimazione della densità esponenziale (3.50) con parametro p .

A conforto di questa asserzione, riportiamo in Fig. 8.5 gli istogrammi relativi alla simulazione, per 10000 ore, del numero di richieste di interventi all'ora e del numero di secondi tra due eventi contigui per il caso illustrato in Fig. 8.4, dove $\lambda = 5/3600 \text{ s}^{-1}$: come potete vedere, la soluzione è praticamente indistinguibile dal caso continuo sia per il numero di eventi all'ora (poissoniana) sia per gli intertempi (esponenziale). Notate che, pur con un tempo medio tra due eventi di circa $3600/5 = 720$ secondi (12 minuti), si possono avere attese lunghe quasi due ore! Questo è uno dei motivi delle grandi fluttuazioni nelle medie temporali che spesso si hanno nei fenomeni stocastici, che la simulazione è in grado di riprodurre accuratamente.

Veniamo ora all'altro tipo di simulazione discreta, quella asincrona [Bun86, Ros96]. In questo caso si sfrutta il fatto che il sistema cambia il proprio stato solo in corrispondenza di eventi ben precisi (ad esempio, arrivo di un nuovo cliente o fine del servizio di una stazione) rimanendo, in caso contrario, sostanzialmente invariato. Così il tempo dell'orologio simulato non avanza in modo costante ma per intervalli variabili, ottenuti tramite la (3.12), che scandiscono l'arrivo di un nuovo evento ed in corrispondenza dei quali vengono aggiornati tutti gli indicatori che descrivono lo stato del sistema da studiare (supponia-

mo di nuovo per semplicità che ogni cambiamento avvenga con durata nulla). L'istante in cui avviene un nuovo evento è ricavato facendo riferimento ad una "lista" in cui devono essere riportati tutti i possibili tipi di eventi che possono avvenire nel sistema e l'istante in cui essi accadranno.

In genere la simulazione sincrona dà luogo a codici di calcolo molto più semplici di quelli che utilizzano la simulazione asincrona, poiché quest'ultima deve garantire l'ordinamento corretto dei tempi di arrivo, il che risulta spesso difficile, specialmente se il sistema è complicato. Tuttavia, la scelta della simulazione asincrona risulta a volte assolutamente preferibile, in quanto i tempi richiesti dalla simulazione sincrona, che esegue un ciclo fisso anche nei casi (e sono la maggioranza!) in cui il sistema rimane invariato, possono risultare inaccettabili.

I diagrammi di flusso tipici della simulazione sincrona e asincrona sono riportati in Fig. 8.6.

L'obiettivo della simulazione di questo tipo di sistemi è in genere la determinazione del valore di alcune grandezze caratteristiche *mediate* entro intervalli di tempo prefissati. Per fissare le idee, potremmo essere interessati al numero X di clienti in attesa ad una stazione di servizio o in fila alla cassa di un supermercato, mediato entro un'ora (assunta come unità di tempo). Per ricavare queste grandezze mediate è necessario registrare i tempi t_i delle variazioni (ad esempio quando si aggiunge un nuovo cliente alla coda) e calcolare ad ogni variazione la grandezza

$$x_i(t_i - t_{i-1}) \equiv x_i \Delta t_i, \quad (8.20)$$

dove x_i è il valore della variabile (discreta o continua) *prima* che avvenga la variazione a t_i . Se osserviamo il fenomeno per un tempo lungo t ore (ad esempio un giorno), possiamo definire la grandezza mediata come:

$$m_t = \frac{\sum_i x_i \Delta t_i}{t}. \quad (8.21)$$

Se dividiamo il tempo in intervalli $\Delta t_i \equiv \delta t$ abbastanza piccoli, in modo da contenere non più di un cambiamento di stato, e tutti uguali tra loro (come nella simulazione sincrona) e assegniamo ad ogni δt il valore x_i dell'ultimo cambiamento di stato, questa formula non è altro che la normale media di un campione di $n = t/\delta t$ eventi.

La media e la varianza di X si possono ottenere mediante un campionamento simulato su N cicli di durata pari a t (ad esempio per molti giorni, una settimana o un mese), applicando le formule standard della statistica:

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^N (m_t)_i, \quad \langle x^2 \rangle = \frac{1}{N} \sum_{i=1}^N (m_t)_i^2, \quad (8.22)$$

$$s^2(x) = \frac{N}{N-1} \left[\langle x^2 \rangle - \langle x \rangle^2 \right]. \quad (8.23)$$

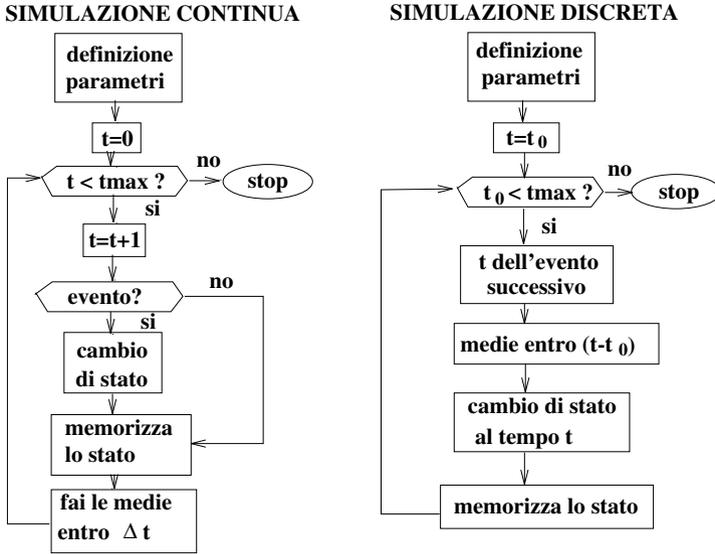


Fig. 8.6. Diagrammi di flusso per la simulazione sincrona e asincrona dei fenomeni di attesa

In questi casi la stima della varianza di X va sempre fatta secondo la (8.23), accumulando via via le medie parziali, dato che il valore finale della media sarà noto solo alla fine della simulazione.

In questo tipo di problemi occorre tener conto che le fluttuazioni sul risultato finale dipendono, oltre che dall'errore statistico, anche dalla variabilità nel tempo dei parametri che descrivono lo stato del processo da studiare. Quasi sempre, tuttavia, i sistemi che interessano raggiungono, dopo un periodo più o meno lungo dall'inizio del loro funzionamento, uno stato stazionario a partire dal quale le quantità (8.22, 8.23) non hanno più variazioni apprezzabili. Determinare quando viene raggiunto lo stato stazionario non è un problema risolvibile in generale per via analitica poiché vengono coinvolte le particolari caratteristiche di ogni singolo processo come la intensità del traffico dei clienti o la forma delle densità dei tempi di arrivo e di servizio. Per tale motivo, di solito si stampano i risultati ad intervalli regolari di tempo simulato in modo tale da poter osservare empiricamente le variazioni e la convergenza al valore di equilibrio delle variabili di stato del sistema in esame.

Anche la valutazione dell'errore statistico risulta abbastanza complicata poiché tutti gli eventi sono correlati tra di loro (il tempo di attesa di un generico cliente dipende, per esempio, da quello dei clienti immediatamente precedenti) e l'uso delle formule relative ad osservazioni indipendenti porterebbe così a dare una valutazione sbagliata di questa grandezza.

Un risultato generale della statistica, che comunque resta sempre valido, è quello relativo all'errore della media $\langle X \rangle$ che, nel caso di regime stazionario,

diminuirà come la radice quadrata del numero dei cicli di campionamento, i quali ovviamente sono proporzionali al tempo totale simulato T_{sim} :

$$\sigma_{MC} \propto \frac{1}{\sqrt{T_{sim}}} . \quad (8.24)$$

Uno degli algoritmi più utilizzati per il calcolo dell'errore statistico è il metodo delle *batch means* (medie a lotti) che tratteremo tra poco nel par. 8.6. In alternativa, possono essere applicati due semplici procedimenti:

- una volta determinato l'intervallo di tempo simulato entro cui si vuole studiare un certo processo, si ripete l'intero programma, con diversi insiemi di numeri casuali, un numero di volte (almeno 15 o 20) abbastanza grande da poter applicare il teorema Limite Centrale per il calcolo dell'errore sulla media di differenti risultati;
- se si vuole invece trovare l'intervallo minimo di tempo simulato necessario per ottenere una precisione prefissata sui risultati finali, occorre effettuare un breve test preliminare per ricavare l'errore sul risultato in corrispondenza di un piccolo valore di T_{sim} . Sfruttando la legge di proporzionalità (8.24) sull'errore statistico, si ricava quindi facilmente il valore cercato.

8.4 Il numero di addetti ad un impianto: simulazione sincrona

Per esemplificare e chiarire meglio le considerazioni precedenti risolviamo, mediante la simulazione sincrona, il problema di determinare il numero ottimale di addetti da adibire alla sorveglianza ed alla manutenzione di un certo numero di macchine utensili.

Supponiamo che le caratteristiche principali del sistema da studiare, dedotte in base alle osservazioni compiute, risultino essere le seguenti:

- nell'impianto vi sono 10 macchine, per ognuna di esse si hanno in media 3 richieste di intervento in un'ora sia per la riparazione di eventuali guasti sia durante la normale fase operativa di lavoro;
- la durata di ogni intervento (che richiede l'attività di una sola persona), qualunque sia l'addetto che la compie e la macchina che lo richiede, segue anch'essa una legge esponenziale negativa con una media di 2 interventi condotti a termine in un'ora da ciascun addetto;
- il costo orario legato all'inattività di un addetto (c_1) è valutato in 70 euro, mentre quello legato all'inattività di una macchina (c_2) è di 20 euro.

Per prima cosa occorre procedere ad una descrizione sintetica (ma completa!) del sistema. Ricordiamo che mediamente si hanno 3 richieste di intervento per ora per macchina e che ogni addetto ripara mediamente due macchine all'ora. Obiettivo della ottimizzazione è minimizzare il costo dell'impianto, sapendo che una macchina in attesa di intervento costa 20 euro/ora e un operatore

Tabella 8.1. Descrizione logica utilizzata per la simulazione del sistema “numero di addetti ad un impianto”. La variabile di stato di una macchina può assumere tre valori: macchina in funzione (0), in riparazione (1), in attesa di riparazione (2)

<i>Stato macchina</i>	<i>Variazione di stato</i>	<i>Addetti inattivi</i>	<i>Variazione addetti in.</i>	<i>Stato macchina</i>
funzionante	funziona	invariati	–	0
	richiede intervento	se > 0	– 1	1
		se = 0		–
in riparazione	continua	invariati	–	1
	fine intervento	aumenta	+ 1	0
in attesa	–	se > 0	– 1	1
	–	se = 0	–	2

inattivo costa 70 euro/ora. La simulazione dovrà quindi determinare il numero medio di operatori inattivi $\langle N_{oi} \rangle$ e di macchine in attesa $\langle N_{mi} \rangle$ e minimizzare il costo orario medio

$$C = 70 \cdot \langle N_{oi} \rangle + 20 \cdot \langle N_{mi} \rangle . \quad (8.25)$$

La descrizione sintetica del sistema è riportata in Tab. 8.1.

Il secondo passo da fare consiste nella scrittura di un codice di simulazione come quello riportato qui sotto. L'unità di misura scelta è il secondo e quindi le probabilità di richiesta di intervento e di fine riparazione sono rispettivamente $3/3600$ e $2/3600$. Il programma, ad ogni secondo, esamina lo stato delle macchine (inizialmente poste tutte funzionanti, $Ms[k]=0$). Se la macchina funziona, richiede un intervento se la variabile casuale uniforme $\xi \leq 3/3600$; se invece la macchina è servita da un meccanico, avviene una fine servizio se $\xi \leq 2/3600$. Questa scelta casuale è operata dalla routine `evento(λ)`. Infine, se la macchina è in attesa di intervento e vi sono meccanici liberi essa viene servita ($Ms[k]=1$), altrimenti resta in attesa ($Ms[k] = 2$). Questa è la fase di cambiamento di stato, indicata nella seconda colonna di Tab. 8.1 ed anche nello schema a blocchi di Fig. 8.6.

```

/ codice SCILAB Optsinc: simulazione sincrona per la
// ottimizzazione del numero di addetti ad un impianto
//
// Routine SCILAB utilizzate:
// grand(1,1,'def') (chiamata di default)
// per estrarre un valore dalla distr. uniforme U(0,1)
// plot2d, Histplote (utente) per il display dei risultati
//

```

```

// valori iniziali e azzeramenti
Mn=10; // 10 macchine
H = 12000; // 12000 ore di simulazione
grand('setsd',5876756); // seme dei numeri casuali
On=0; Ol=0; med=0; minh=0; oinh=0; // azzeramenti
enne=0; med2=0; olib=0; olib2=0;
memac=0; simac=0; meope=0; sigad=0;

// Ms(Mn) = 0, 1, 2 e' il vettore di stato macchina
// azzerata Ms e gli indicatori orari Oih e Mih
Ms=zeros(1,Mn); Oih=zeros(1,H); Mih=zeros(1,H); Ore=zeros(1,H);
Oih1=zeros(1,H); Mih1=zeros(1,H);

// input da terminale
On = input("numero degli operatori:");

Hour = 60; // Unita' di misura: minuto. 1 ora = 60 minuti
Mp = 3/Hour; // 3 interventi/ora
Op = 2/Hour; // 2 fine interventi/ora

Ol=On; // Ol = numero istantaneo di operatori inattivi
Oi=0; // Oi = numero cumulato di operatori inattivi
Mi=0; // Mi = contano le macchine in attesa di intervento

// simula per H ore in passi di 3600/Hour secondi

for n=1:Hour*H,
    for k=1:Mn,
        select Ms(k),
        case 0 then
            if(grand(1,1,'def')<Mp) then
                if(Ol>0) then Ms(k)=1; Ol=Ol-1;
                else Ms(k)=2;
            end;
        end;
        case 1 then
            if(grand(1,1,'def')<Op) then
                Ms(k)=0; Ol=Ol+1;
            end;
        case 2 then
            if(Ol~=0) then,
                Ms(k)=1; Ol=Ol-1;
            end;
        end;
    end;
    for k=1:Mn, if(Ms(k)==2) then Mi=Mi+1; end; end;
    Oi = Oi + Ol; ;

```

```

if(modulo(n,Hour*24)==0 & n~=0) then
    enne=enne+1;
    Ore(enne) = 24*enne;
    minh= Mi/(Hour*24);
    med = med + minh;
    med2 = minh*minh + med2;
    oinh = Oi/(Hour*24);
    olib = olib + oinh;
    olib2 = oinh*oinh + olib2;
    memac = med/enne;
    simac = sqrt(med2/enne - memac*memac);
    meope = olib/enne;
    sigad = sqrt(olib2/enne - meope*meope);
//      plot delle curve
    Oih1(enne) = oinh;      // operatori inattivi/ora
    Mih1(enne) = minh;     // macchine in attesa/ora
    Oih(enne) = meope;     // media operatori inattivi/ora
    Mih(enne) = memac;     // media macchine in attesa/ora
    x=zeros(1,enne);
    y = zeros(1,enne); z = zeros(1,enne);
    y1= zeros(1,enne); z1= zeros(1,enne);
    for j=1:enne, x(j)=Ore(j); y(j) =Oih(j); z(j) =Mih(j);
        y1(j)=Oih1(j); z1(j)=Mih1(j); end;
    if(enne>3) then
        clf();
        subplot(2,2,1)
        plot2d(x,y); xtitle('addetti inattivi');
        subplot(2,2,2)
        plot2d(x,z); xtitle('macchine inattive');
        subplot(2,2,3)
        Histplote(30,y1,stat=1);
        subplot(2,2,4)
        Histplote(30,z1,stat=1);
    end;
    Oi=0 ; Mi = 0;      // azzera gli indicatori giornalieri
end;
end;

```

Ad ogni secondo viene sommato il numero di operatori inattivi O_i e quello delle macchine in attesa di intervento M_i , che sono le variabili che si vogliono studiare. Ogni 24 ore queste grandezze sono mediate secondo la (8.21) e le corrispondenti medie e varianze sono calcolate progressivamente con le (8.22, 8.23). La Fig. 8.7 mostra l'andamento di O_i e M_i per un periodo di osservazione di 500 giorni.

Le medie e le varianze finali di queste quantità sono riportate in Tab. 8.2 con la valutazione del costo globale tramite la (8.25). Dall'esame della figura e della tabella si vede che le fluttuazioni intorno ai valori medi sono notevoli. In effetti, l'impianto non è molto efficiente, perché avendo 10 macchine con

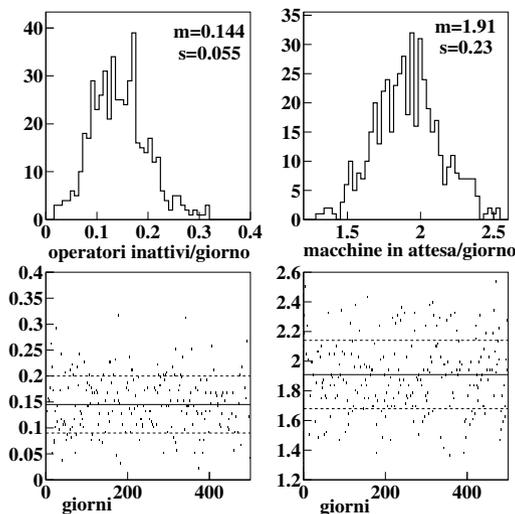


Fig. 8.7. Le figure in basso mostrano l'andamento giornaliero del numero di macchine in attesa di riparazione e del numero di addetti inattivi per ora per il problema "numero ottimale di addetti ad un impianto" nel caso di 5 addetti; la proiezione sulle ordinate dei valori giornalieri è riportata negli istogrammi in alto. La media m e i valori $m \pm s$ degli istogrammi in alto sono le linee continue e tratteggiate dei diagrammi in basso

Tabella 8.2. Variazione del costo globale dello stabilimento al variare del numero degli operatori utilizzati. Gli errori indicano la deviazione standard dei dati, non l'errore sulla media

<i>Numero addetti</i>	<i>Ore di inattività addetti/ora</i> ($\langle N_{oi} \rangle \pm s_{oi}$)	<i>Ore di attesa macchine/ora</i> ($\langle N_{mi} \rangle \pm s_{mi}$)	<i>Costo Globale orario</i>
3	$1.5 \cdot 10^{-3} \pm 3 \cdot 10^{-3}$	5.00 ± 0.21	100, 10
4	$2.21 \cdot 10^{-2} \pm 1.84 \cdot 10^{-2}$	3.38 ± 0.26	69, 10
5	0.148 ± 0.055	1.91 ± 0.23	48, 60
6	0.53 ± 0.12	0.87 ± 0.15	52, 40

una media di 3 richieste di intervento l'ora ed un tempo medio di servizio di mezz'ora, si avrà sempre un gran numero di macchine ferme (circa 5 o 6) anche con 10 addetti, uno per macchina. Data la legge esponenziale assunta per i tempi di richiesta sia di intervento sia di servizio, che sono pure tra loro confrontabili, è naturale attendersi ampie fluttuazioni nelle grandezze

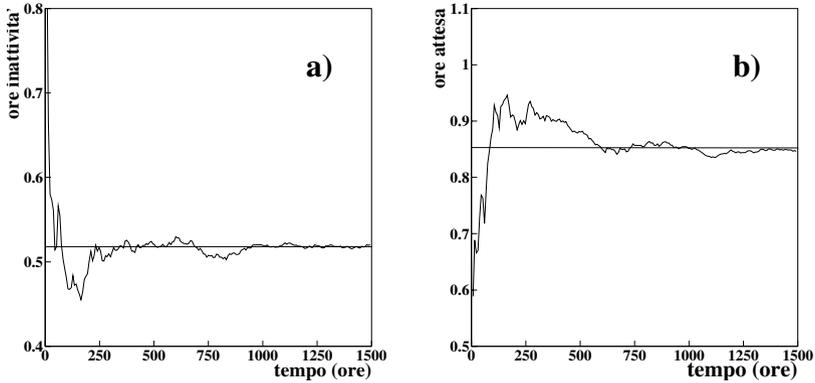


Fig. 8.8. Variazione, in funzione del tempo della simulazione, del numero medio di addetti inattivi (a) e della frazione di macchine in attesa di riparazione (b) per ogni ora di funzionamento dell'impianto e nel caso di 6 addetti. La linea continua rappresenta il valore asintotico di tali quantità calcolato analiticamente come in [DMM⁺75]

giornaliere osservate. Essendo però il tipo di macchina un dato assegnato a priori, la simulazione risolve comunque il problema di minimizzazione dei costi, mostrando che il numero ottimale di addetti è 5, come mostrato in Tab. 8.2.

Nella Fig. 8.8 abbiamo invece riportato, nel caso di 6 operai, l'andamento al variare del tempo di simulazione del numero medio di ore di addetti inattivi e della frazione di macchine in attesa di riparazione per ogni ora di funzionamento dell'impianto. Le fluttuazioni presenti all'inizio della simulazione sono dovute sia alle particolari condizioni iniziali assunte per il sistema, lontane da quelle stazionarie, sia alla quantità insufficiente di dati. Dai due riquadri in basso della Fig. 8.7 si può comunque dedurre che la condizione di stazionarietà viene raggiunta abbastanza rapidamente.

Va anche notato che le medie e deviazioni standard di Tab. 8.2 sono ricavate da una osservazione di 500 giorni. Considerando come esempio il caso ottimale di 5 operatori, dalle formule statistiche della Tab. 6.3 di pag. 213, in accordo anche con la (8.24), possiamo stimare gli errori su queste quantità come:

$$\sigma[\langle N_{mi} \rangle] \simeq \frac{1.91}{\sqrt{500}} = 0.08, \quad \sigma[s(N_{mi})] \simeq \frac{0.23}{\sqrt{1000}} = 0.007$$

$$\sigma[\langle N_{oi} \rangle] \simeq \frac{0.148}{\sqrt{500}} = 0.007, \quad \sigma[s(N_{oi})] \simeq \frac{0.12}{\sqrt{1000}} = 0.004.$$

8.5 Il numero di addetti ad un impianto: simulazione asincrona

La simulazione sincrona ha lo svantaggio di compiere numerosi cicli inutili, nei quali il sistema non cambia di stato. Per rendere il programma del paragrafo precedente più veloce abbiamo cambiato l'unità di tempo dal secondo al minuto, ottenendo risultati differenti di qualche per cento da quelli di Tab. 8.2. In questo schema non è inoltre possibile usare distribuzioni dei tempi di richieste di intervento e di fine servizio diversi dall'esponenziale. Un metodo più generale e veloce per descrivere il sistema è quello di utilizzare la simulazione asincrona.

In questo caso il tempo non scorre ad intervalli regolari ma a salti, scandito dall'accadere degli eventi. Nel caso del problema trattato in precedenza, gli eventi sono di due tipi, richiesta di intervento e fine intervento. Si è reso quindi necessario, nel codice di simulazione `Optasinc`, che troverete nelle nostre pagine web [RPP], introdurre i vettori `Tmac[k]` e `Top[k]` che contengono rispettivamente, per ogni macchina, i tempi di richiesta di intervento e di fine servizio, determinati dalla legge esponenziale attraverso la routine `texp`. In fase di inizializzazione, vengono estratti una serie di tempi di richieste di intervento mentre i tempi di intervento vengono "congelati" con un tempo "infinito" `BIG`. L'orologio della simulazione avanza ricercando il minimo dei tempi contenuti in questi due vettori, utilizzando la routine `mini`. Nel caso una macchina sia in attesa di riparazione o un operatore sia inattivo bisogna avere l'avvertenza di "fermare" il tempo di richiesta o di fine intervento introducendo nei vettori `Tmac` e `Top` il tempo `BIG`.

Supponiamo ora di aver già considerato un certo numero di eventi; quello successivo sarà o una nuova richiesta di intervento oppure la fine del servizio di un addetto. Se l'evento è una richiesta di intervento della macchina `kmac`, si aggiorna il tempo simulato, si impegna un operatore (se disponibile), si "congela" la macchina (`Tmac[kmac]=BIG`) e si simula il tempo di fine intervento (`Top[kmac] = t + texp(Op)`); se non vi sono operatori liberi, si pone la macchina nello stato 2 (`Ms[kmac]=2`). Se invece l'evento è la fine dell'intervento dell'operatore sulla macchina `kop`, si genera un tempo di nuova richiesta di intervento `Tmac[kop] = t + texp(Mp)`, si congela il tempo di riparazione della macchina (`Top[kop] = BIG`) e si aggiornano i vettori e gli indici di stato, come commentato nel programma. Se vi sono altre macchine in coda, il meccanico che ha appena terminato il proprio lavoro si occupa subito di una macchina in attesa (la numero `n`), con un intervento che avrà termine all'istante `Top[n]= t + texp(Op)`.

Il numero di macchine in attesa `Mi` e il numero di operatori liberi `O1` vengono moltiplicati per il tempo tra due eventi (`t-tprec`), ottenendo le nuove variabili `oin` e `min`. Qui va notato un punto importante: questo intervallo di tempo moltiplica `O1` e `Mi` *calcolati al tempo tprec*, in quanto tra `tprec` e `t` la situazione rimane quella del cambiamento di stato avvenuto in `tprec`. Questo

passo del codice è equivalente a quello della simulazione sincrona in cui, ad ogni secondo, si sommano le stesse quantità, anche se rimanevano invariate. Le medie giornaliere della (8.21) vengono poi ottenute dividendo oin e min per l'intervallo $(t - \text{tprint})$ che, a meno di fluttuazioni di qualche decina di minuti dovute alla selezione discreta dei tempi, vale circa 24 ore. Le medie progressive e le relative varianze vengono poi calcolate come nel caso della simulazione sincrona. La simulazione ha termine quando il tempo t raggiunge un valore prefissato TIME .

Come si vede, la logica e la struttura del codice risultano più complicate che nel caso della simulazione sincrona; tuttavia, il tempo di esecuzione è risultato circa 15 volte più breve, perché in questo caso si calcola la transizione di stato del sistema solo quando avviene effettivamente un evento. Il programma ha ovviamente fornito gli stessi risultati della simulazione sincrona, che abbiamo già riportato in Tab. 8.2 e Fig. 8.7.

Per programmi semplici come questi il guadagno di tempo non è importante, ma per modelli complessi, in cui il tempo di esecuzione dei programmi può essere anche di qualche ora, un guadagno di questa entità risulta determinante.

Potete ora riprendere dalle pagine web [RPP] il programma appena descritto e complicarne la struttura cercando di descrivere un modello più realistico sperimentando direttamente come sia possibile studiare sistemi complessi in modo tutto sommato semplice. Ad esempio, si possono introdurre tipi diversi di richieste di servizio (guasto, manutenzione, ...) con tempi diversi di intervento, considerare la disponibilità dei pezzi di ricambio, oppure assegnare ad ogni addetto tempi di riparazione personalizzati e così via. L'inserimento di questi dettagli in un modello analitico rende rapidamente intrattabile il problema, mentre, con un programma di simulazione, è possibile aggiungere in modo modulare nuovi pezzi di codice senza complicare eccessivamente sia la struttura sia la gestione del modello.

8.6 Algoritmo di Metropolis

L'algoritmo di Metropolis è un metodo sofisticato per riuscire a produrre un campione da distribuzioni per le quali non sia facilmente utilizzabile una delle tecniche spiegate nel cap. 7.

Le distribuzioni alle quali meglio si applica sono quelle specificate da densità scritte come

$$p(\mathbf{x}) = \frac{h(\mathbf{x})}{Z},$$

dove \mathbf{x} è un vettore aleatorio d -dimensionale. A causa della condizione di normalizzazione, si deve avere $Z = \sum_{\mathbf{x}} h(\mathbf{x})$ nel caso discreto e $Z = \int h(\mathbf{x}) d\mathbf{x}$ in quello continuo. La costante di normalizzazione Z è indispensabile per ricavare qualsiasi quantità legata a $p(\mathbf{x})$ (come media, varianza e percentili), tuttavia in alcuni casi il calcolo può rivelarsi in pratica impossibile. Questo avviene, ad esempio, nell'ambito della fisica o della chimica quando si studiano

sistemi composti da un gran numero d di identici componenti elementari come le molecole in un gas o in un solido cristallino (tipicamente $\sim 10^{23}$, numero che rappresenta grosso modo il numero di atomi contenuti in un cm^3 di materia). Supponiamo che ciascun componente elementare possa assumere k stati. Se $g(\mathbf{x})$ esprime un parametro macroscopico del sistema (come temperatura, pressione, momento magnetico), il calcolo della sua media, $\sum_{\mathbf{x}} g(\mathbf{x})p(\mathbf{x})$, richiederebbe di valutare $h(\mathbf{x})$ per ognuna delle k^d possibili configurazioni del sistema, un numero che può essere enorme rispetto alle nostre risorse di calcolo.

L'algoritmo di Metropolis è molto potente per risolvere questi problemi, in quanto non è necessario conoscere Z , né valutare $h(\mathbf{x})$ per tutti i valori di \mathbf{x} . L'idea è quella di generare una sequenza di numeri la cui distribuzione di frequenza asintotica tenda a $p(\mathbf{x})$.

Immaginiamo quindi un sistema in cui gli eventi “nascono” assumendo un valore \mathbf{x} in accordo con la densità $p(\mathbf{x})$ e poi possano “migrare” ad un altro valore \mathbf{y} secondo una probabilità di transizione *arbitraria* $t(\mathbf{x} \rightarrow \mathbf{y})$. I sistemi in cui queste probabilità dipendono solo dal valore di partenza e di arrivo sono detti *catene di Markov*, e sono di importanza fondamentale nello studio di molti processi stocastici [RC99]. Una condizione sufficiente affinché la catena converga ad una popolazione distribuita come $p(\mathbf{x})$ è che essa si stabilizzi in una situazione di equilibrio in cui ogni transizione avviene con probabilità uguale a quella inversa. Una delle condizioni sotto cui questo si verifica è data dall'equazione detta del bilancio dettagliato:

$$p(\mathbf{x}) t(\mathbf{x} \rightarrow \mathbf{y}) = p(\mathbf{y}) t(\mathbf{y} \rightarrow \mathbf{x}) , \quad (8.26)$$

in cui il termine di destra (sinistra) esprime la probabilità che il sistema evolva da \mathbf{x} a \mathbf{y} (da \mathbf{y} a \mathbf{x}). La funzione arbitraria $t(\mathbf{x} \rightarrow \mathbf{y})$ viene in genere scritta come

$$t(\mathbf{x} \rightarrow \mathbf{y}) = q(\mathbf{x}, \mathbf{y}) \alpha(\mathbf{x}, \mathbf{y}) . \quad (8.27)$$

Alla distribuzione ausiliaria $q(\mathbf{x}, \mathbf{y})$ si richiede che, per ogni fissato valore \mathbf{x} appartenente allo spettro di \mathbf{X} , essa sia una distribuzione di probabilità sullo spettro di \mathbf{X} . In altre parole, per ogni \mathbf{x} fissato, $q(\mathbf{x}, \mathbf{y}) \geq 0$ per ogni valore di \mathbf{y} e $\sum_{\mathbf{y}} q(\mathbf{x}, \mathbf{y}) = 1$ nel caso discreto e $\int q(\mathbf{x}, \mathbf{y}) d\mathbf{y} = 1$ nel caso continuo. Un'altra condizione necessaria è che sia possibile generare velocemente dalla distribuzione $q(\mathbf{x}, \cdot)$.

La probabilità $\alpha(\mathbf{x}, \mathbf{y})$ di accettare il valore proposto dalla distribuzione ausiliaria è invece definita dall'algoritmo di Metropolis in modo tale da “guidare” l'evoluzione del sistema verso stati sempre più probabili (in cui, ad esempio, sia minima l'energia totale o la temperatura o la pressione). In questo modo si arriva in una situazione stazionaria in cui i parametri macroscopici del sistema non variano con il tempo ed in cui viene soddisfatta anche la (8.26).

L'algoritmo prevede N passi; se $\mathbf{x}^{(i)}$ è il valore generato al passo i , per ottenere il valore successivo si procede come segue:

Algoritmo 8.1 (Algoritmo di Metropolis).

- 1) Generare un valore \mathbf{y} dalla distribuzione ausiliaria $q(\mathbf{x}^{(i)}, \cdot)$.
- 2) Generare un valore ξ dalla distribuzione uniforme su $(0, 1)$.
- 3) Calcolare la probabilità di accettazione

$$\alpha(\mathbf{x}^{(i)}, \mathbf{y}) = \min \left\{ 1, \frac{h(\mathbf{y}) q(\mathbf{y}, \mathbf{x}^{(i)})}{h(\mathbf{x}^{(i)}) q(\mathbf{x}^{(i)}, \mathbf{y})} \right\} \quad (8.28)$$

dove il rapporto che vi compare si dice rapporto di accettazione.

- 4) Se $\xi \leq \alpha(\mathbf{x}^{(i)}, \mathbf{y})$, allora porre $\mathbf{x}^{(i+1)} = \mathbf{y}$, altrimenti porre $\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)}$.
Se $i < N$ tornare al passo 1.

È facile mostrare che i valori ottenuti con la (8.28) seguono una densità che soddisfa al bilancio dettagliato. Infatti, poiché dalle (8.26, 8.27) segue:

$$p(\mathbf{x}) q(\mathbf{x}, \mathbf{y}) \alpha(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}) q(\mathbf{y}, \mathbf{x}) \alpha(\mathbf{y}, \mathbf{x}),$$

tenendo conto che dalla (8.28), se $p(\mathbf{y}) q(\mathbf{y}, \mathbf{x}) / [p(\mathbf{x}) q(\mathbf{x}, \mathbf{y})] < 1$, segue asintoticamente $\alpha(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}) q(\mathbf{y}, \mathbf{x}) / [p(\mathbf{x}) q(\mathbf{x}, \mathbf{y})]$ e $\alpha(\mathbf{y}, \mathbf{x}) = 1$, si ottiene l'identità

$$p(\mathbf{x}) q(\mathbf{x}, \mathbf{y}) \frac{p(\mathbf{y}) q(\mathbf{y}, \mathbf{x})}{p(\mathbf{x}) q(\mathbf{x}, \mathbf{y})} = p(\mathbf{y}) q(\mathbf{y}, \mathbf{x}).$$

Fissato dunque un valore iniziale $\mathbf{x}^{(0)}$, e ripetendo N volte i passi 1-4 si ottiene un campione $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)})$ senza bisogno di conoscere Z , in quanto il passo 3 dipende dal rapporto $p(\mathbf{y})/p(\mathbf{x}^{(i)}) = h(\mathbf{y})/h(\mathbf{x}^{(i)})$.

Ma che genere di campione abbiamo ottenuto? Evidentemente non un campione casuale, perché le variabili generate sono dipendenti, essendo la generazione di $\mathbf{x}^{(i+1)}$ dipendente da $\mathbf{x}^{(i)}$. Inoltre il valore iniziale $\mathbf{x}^{(0)}$ ha poco a che fare con $p(\cdot)$ perché viene di solito scelto in modo arbitrario. Tuttavia, sotto semplici ipotesi, si può mostrare [RC99] la validità del seguente:

Teorema 8.1. *Se $q(\mathbf{x}, \mathbf{y}) > 0$ per ogni \mathbf{x} e per ogni \mathbf{y} appartenenti allo spettro di \mathbf{X} , allora la proprietà*

$$\lim_{N \rightarrow \infty} \frac{\sum_{i=1}^N g(\mathbf{x}^{(i)})}{N} = \langle g(\mathbf{X}) \rangle = \begin{cases} \sum_{\mathbf{x}} g(\mathbf{x}) p(\mathbf{x}), & \text{caso discreto;} \\ \int g(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}, & \text{caso continuo} \end{cases} \quad (8.29)$$

è valida per ogni punto iniziale $\mathbf{x}^{(0)}$. La validità della (8.29) si mantiene anche se la condizione $q(\mathbf{x}, \mathbf{y}) > 0$ non è soddisfatta per tutte le coppie (\mathbf{x}, \mathbf{y}) , a condizione che, per ogni insieme A dello spettro con $p(A) > 0$, $q(\mathbf{x}, \mathbf{y})$ sia tale che A è raggiungibile con probabilità positiva a partire da ogni \mathbf{x} .

Nella sua formulazione più semplice, l'algoritmo viene utilizzato con $q(\mathbf{x}, \mathbf{y}) = q(\mathbf{y}, \mathbf{x})$, nel qual caso α dipende solo dal rapporto $p(\mathbf{y})/p(\mathbf{x})$. Spesso si estrae \mathbf{X} dalla distribuzione uniforme, entro l'intervallo di esistenza di $p(\mathbf{x})$. Potete a questo punto fare qualche test con la nostra routine **Metrop**, e risolvere il problema 8.7.

L'esecuzione dell'algoritmo di Metropolis ci fornisce dunque lo stimatore

$$\frac{\sum_{i=1}^N g(\mathbf{X}^{(i)})}{N} \equiv M$$

per $\langle g(\mathbf{X}) \rangle$. La varianza di α non è $\sum_{i=1}^N \text{Var}[g(\mathbf{X}^{(i)})]/N^2$, in quanto le variabili aleatorie simulate non sono indipendenti. Un modo per allentare la dipendenza è suddividere la successione simulata $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)})$ in k blocchi consecutivi di b elementi ciascuno (con b e k tali che $kb = N$)

$$(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(b)}, \mathbf{x}^{(b+1)}, \dots, \mathbf{x}^{(2b)}, \dots, \mathbf{x}^{((k-1)b+1)}, \dots, \mathbf{x}^{(kb)})$$

e calcolare la media campionaria di g in ciascun blocco:

$$(\langle g(\mathbf{x}) \rangle^{(1)}, \dots, \langle g(\mathbf{x}) \rangle^{(k)}) .$$

All'aumentare della dimensione dei blocchi, i blocchi non consecutivi sono sempre più distanti (in termini di iterazioni) e perciò sempre meno correlati. Si può mostrare che anche la correlazione tra blocchi consecutivi tende ad annullarsi al crescere di b , perciò ci si avvicina a una situazione di incorrelazione ed è naturale utilizzare la stima della varianza campionaria della successione delle medie dei blocchi per stimare l'errore di $\langle g(\mathbf{x}) \rangle^{(i)}$. Questo errore è associato a una media campionaria di b termini, pertanto, per ottenere l'errore associato a $m \equiv \langle g(\mathbf{x}) \rangle$, che è una media di $N = kb$ termini, dobbiamo dividere ancora per k , trovando infine la *stima delle batch means* (con un *CL* del 95.4%):

$$\langle g(\mathbf{X}) \rangle \in \langle g(\mathbf{x}) \rangle \pm 2 \sqrt{\frac{1}{k(k-1)} \sum_{i=1}^k [\langle g(\mathbf{x}) \rangle^{(i)} - \langle g(\mathbf{x}) \rangle]^2} . \quad (8.30)$$

Questa espressione sottintende la validità del Teorema Centrale Limite per la distribuzione di α ed è dimostrabile, sotto le ipotesi del teorema 8.1, per gli algoritmi di Metropolis aperiodici, cioè quando non esiste una partizione dello spazio degli stati i cui insiemi vengono visitati secondo una sequenza sempre uguale durante la simulazione.

Per calcolare la stima delle *batch means*, una volta fissato il numero di iterazioni N , bisogna scegliere b . Questo è un problema molto complesso e a tutt'oggi non è possibile dare una regola pratica di uso immediato.

8.7 Il modello di Ising

Vediamo ora un'applicazione dell'algoritmo di Metropolis al noto modello di Ising, utilizzato dal fisico tedesco per spiegare alcuni comportamenti osservati nella magnetizzazione dei materiali.⁴

⁴ Questo modello ha importanti applicazioni anche in ambiti completamente diversi dalla fisica, perché descrive bene la evoluzione di sistemi in cui vi sono cambia-

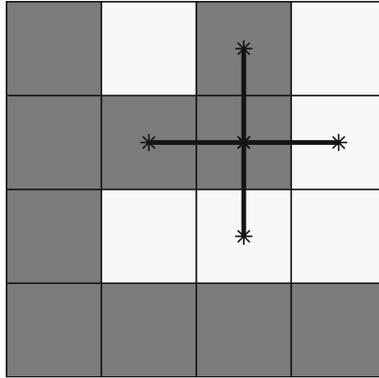


Fig. 8.9. Un reticolo di 4×4 atomi di un solido cristallino con i relativi momenti magnetici di spin (-1 colore scuro e 1 colore chiaro) e l'interazione tra primi vicini, rappresentata dalla croce

L'immagine binaria formata da 4×4 pixel⁵ di Fig. 8.9, rappresenta una porzione bidimensionale di un solido cristallino, dove su ogni pixel risiede un atomo del solido con il suo campo magnetico di spin: al colore scuro associamo spin -1, mentre al colore chiaro associamo spin +1. Quando la temperatura lo consente, l'interazione esistente tra gli spin nucleari a livello microscopico determina il comportamento del materiale a livello macroscopico, conferendogli proprietà ferromagnetiche o antiferromagnetiche. In Fig. 8.9 è stato evidenziato il più semplice modello di interazione microscopica, nel quale l'atomo al centro della croce ivi tracciata interagisce solo con i vicini più prossimi in senso orizzontale e verticale. Il modello di Ising, con interazione diretta solo tra primi vicini, definisce l'energia di un materiale ferromagnetico con $n \times n$ atomi nel modo seguente:

$$H(\mathbf{x}) = -\beta \sum_{i,j;i \sim j} x_i x_j, \quad \beta > 0,$$

dove $\mathbf{x} = (x_1, \dots, x_{n^2})$, x_i indica lo spin dell'atomo indicizzato da i e la somma è estesa ai soli primi vicini ($i \sim j$).

In assenza di un campo magnetico esterno, a ciascuna configurazione viene associata una probabilità che dipende dall'energia del sistema e dalla

menti di stato a seguito di interazioni di contatto tra gli individui. Ad esempio, esso è utilizzato per studiare l'impatto sociale di nuove idee e la dinamica delle opinioni in società complesse [KH96] e per prevedere il comportamento dei mercati finanziari [Voi03].

⁵ Il termine pixel nasce dalla contrazione delle parole inglesi *picture element* ed indica la più piccola unità omogenea che compone una immagine artificiale.

temperatura T :

$$p(\mathbf{x}) = \frac{\exp\left\{-\frac{H(\mathbf{x})}{T}\right\}}{\sum_{\mathbf{x}} \exp\left\{-\frac{H(\mathbf{x})}{T}\right\}} \equiv \frac{h(\mathbf{x})}{Z}.$$

La formula evidenzia che le configurazioni a bassa energia, cioè quelle con atomi vicini con lo stesso spin, hanno probabilità più elevata.

L. Onsager nel 1944 ha risolto il problema della trattazione analitica esatta del modello di Ising in due dimensioni, come ad esempio quello del calcolo del numero atteso di atomi con spin 1 a una data temperatura, utile per determinare il momento magnetico totale del materiale. Tuttavia il più realistico modello in tre dimensioni non è stato risolto e bisogna ricorrere alla simulazione, che per semplicità illustriamo nel caso bidimensionale.

Le configurazioni ammissibili di tutti gli spin nel caso bidimensionale sono 2^{n^2} , quindi non è possibile, se non per n molto ridotto, calcolare tutte le $p(\mathbf{x})$ e procedere alla simulazione diretta come nella sezione 7.4. Ricorriamo pertanto all'algoritmo di Metropolis nel quale scegliamo una distribuzione ausiliaria che ad ogni passo scelga un atomo a caso e che ne proponga il cambio di spin. È abbastanza facile convincersi che la formula per questa distribuzione è la seguente:

$$q(\mathbf{x}, \mathbf{y}) = \frac{1}{n^2}$$

se \mathbf{x} e \mathbf{y} differiscono per lo spin di un solo atomo, mentre $q(\mathbf{x}, \mathbf{y}) = 0$ in tutti gli altri casi. Per calcolare facilmente il rapporto di accettazione, possiamo osservare che l'energia del sistema è esprimibile così:

$$H(\mathbf{x}) = \beta(n^-(\mathbf{x}) - n^+(\mathbf{x}))$$

dove n^+ ed n^- indicano rispettivamente il numero di coppie di atomi primi vicini con segno concorde e discorde. Allora, essendo $q(\mathbf{x}, \mathbf{y}) = q(\mathbf{y}, \mathbf{x})$, il rapporto di accettazione è semplicemente

$$\exp\{\beta[(n^+(\mathbf{y}) - n^+(\mathbf{x})) - (n^-(\mathbf{y}) - n^-(\mathbf{x}))]/T\}.$$

Ora, se la distribuzione ausiliaria ha scelto l'atomo i , le differenze all'esponente dipendono esclusivamente dai segni dei primi vicini di quell'atomo. Indicando rispettivamente con $n^+(x_i)$ e con $n^-(x_i)$ il numero di primi vicini con lo stesso segno di x_i e con segno opposto, e tenendo conto che $n^-(x_i) = 4 - n^+(x_i)$, il rapporto di accettazione diviene semplicemente

$$\exp\{2\beta[n^+(y_i) - n^+(x_i)]/T\} = \exp\{2\beta[4 - 2n^+(x_i)]/T\}.$$

Fissata dunque una configurazione iniziale qualunque (per esempio scegliendo a caso lo spin di ciascun atomo), per procedere con l'algoritmo basta controllare i segni dell'atomo interessato dal cambiamento e dei suoi vicini. Se $g(\mathbf{x}^{(i)})$ è il numero di atomi con spin 1 al passo i dell'algoritmo, il termine successivo

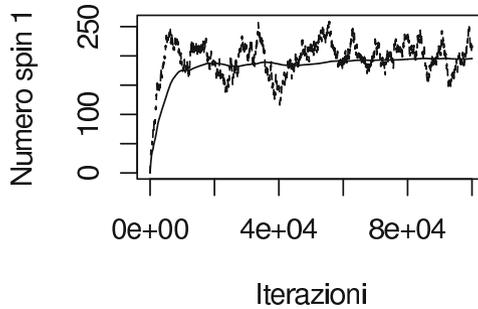


Fig. 8.10. Tracciato della simulazione di Metropolis per il modello di Ising: sono rappresentati il numero di spin 1 (linea tratteggiata) e la media campionaria di tale numero al crescere delle iterazioni (linea continua)

$g(\mathbf{x}^{(i+1)})$ per il calcolo della media (8.29) è invariato se l'atomo prescelto non cambia spin o si ottiene facilmente aggiungendo 1 o -1 in caso contrario.

Dal punto di vista algoritmico, l'indice dell'atomo da cambiare si ottiene generando una variabile uniforme ξ_1 in $(0, 1)$ e prendendo il più piccolo intero i che supera $\xi_1 n^2$. Dato i , se una seconda variabile uniforme ξ_2 risulta minore della probabilità di accettazione

$$\alpha(\mathbf{x}, \mathbf{y}) = \min \{1, \exp\{2\beta[4 - 2n^+(x_i)]/T\}\}$$

si procede al cambio di spin. Il programma `Ising` esegue questo algoritmo.

Abbiamo eseguito una simulazione del modello di Ising con $N = 100\,000$ iterazioni, $\beta = 0.3$ e $T = 1$, partendo dalla configurazione di tutti spin -1 in un reticolo $n \times n$ con $n = 20$. Supponiamo che ci interessi il numero atteso di atomi con spin 1, quindi $g(\mathbf{x})$ sarà questo numero nella configurazione \mathbf{x} . In Fig. 8.10 abbiamo riportato in linea tratteggiata il tracciato della successione $\{g(\mathbf{x}^{(i)})\}$. In linea continua compare il tracciato della media campionaria $\langle g(\mathbf{x}) \rangle$ calcolata con un numero crescente di iterazioni.

Possiamo notare alcune cose:

- 1) Dopo qualche migliaio di iterazioni il tracciato $\{g(\mathbf{x}^{(i)})\}$ si stabilizza e inizia ad oscillare attorno al suo presumibile valore atteso.
- 2) Il tracciato della media campionaria tende invece a convergere a un valore costante, che secondo il teorema 8.1 è $\langle g(\mathbf{X}) \rangle$.
- 3) Il fatto di essere partiti da una zona di evidente bassa probabilità (in quanto viene rapidamente abbandonata dall'algoritmo), induce una distorsione nella stima della media, poiché durante una simulazione di lunghezza finita la configurazione di tutti spin -1 non verrà mai raggiunta spontaneamente. L'averla inclusa significa averle assegnato un peso superiore a quello che le compete. Possiamo allora scartare un numero iniziale di iterazioni (per esempio le prime 10 000) per raggiungere una zona di alta probabilità e iniziare ad accumulare dati da questo punto per il calcolo di $\langle g(\mathbf{x}) \rangle$.

- 4) La quantità di dati da accumulare può essere decisa attraverso il tracciato della media campionaria, fermandosi quando quest'ultimo presenta oscillazioni di ampiezza inferiore a una certa soglia (che è soggettiva e dipende dalla precisione che desideriamo raggiungere).
- 5) La dimensione b dei gruppi per il metodo delle *batch means* va incrementata sino a quando la autocorrelazione della serie $\{g(\mathbf{x})^{(i)}\}_{i \geq 1}$ non diventi trascurabile. (Per la rappresentazione grafica della autocorrelazione si veda un libro sulle serie storiche, come [BD02]). Questo può a sua volta richiedere di aumentare il numero di iterazioni per non avere un numero di gruppi troppo piccolo. Il nostro esempio possiede proprio queste caratteristiche.

Nel caso della Fig. 8.10, scartando le prime 10 000 iterazioni e tenendo le successive 90 000 la stima intervallare (8.30), con $b = 4500$ (cioè 20 gruppi), è

$$197.5 \pm 2 \times 4.1 = (189.3, 205.7) .$$

Concludiamo con un'importante avvertenza. Potete provare, usando il codice `Ising` a fare nuove simulazioni con temperature T inferiori a 1. Troverete che quando la temperatura è tale che $\beta/T > 0.44$ il tracciato di $g(\mathbf{x})$ andrà abbastanza rapidamente verso una delle due mode della sua distribuzione (cioè $g(\mathbf{x}) = 0$ o $g(\mathbf{x}) = 400$), senza riuscire a passare dall'una all'altra. In situazioni di questo tipo, sebbene il teorema 8.1 continui a valere, non è praticamente possibile effettuare il numero di iterazioni necessarie per visitare le zone di maggior probabilità dello spazio degli stati, e quindi la media campionaria del tracciato non è affatto una buona stima di quella vera.

8.8 Calcolo di integrali definiti

Il calcolo numerico del valore di un integrale definito, oltre ad essere una delle applicazioni più note e diffuse del metodo MC, è anche un tipico esempio di uso delle tecniche di simulazione per problemi che a prima vista sembrerebbero non ammettere un approccio statistico. Come vedremo tra poco, con il metodo MC è conveniente risolvere integrali multidimensionali, dove gli altri metodi numerici presentano diversi problemi di applicazione. Nel seguito, tuttavia, per snellire e semplificare la notazione, considereremo solo le funzioni di una sola variabile, tenendo presente che tutta la trattazione può venire molto facilmente estesa al caso di funzioni multidimensionali.

Vi sono due diversi approcci fondamentali che possono essere utilizzati per calcolare l'integrale definito:

$$I = \int_a^b f(x) dx \tag{8.31}$$

utilizzando numeri casuali.

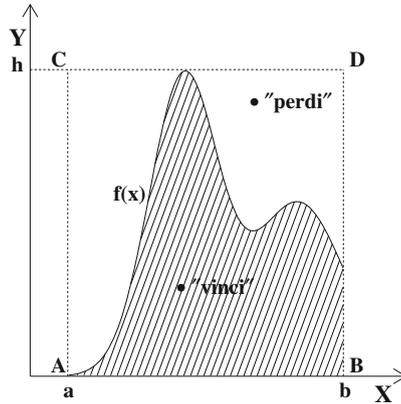


Fig. 8.11. Rappresentazione grafica del metodo “vinci o perdi”: per valutare l’integrale (8.31) si conta quante volte è “colpita” la zona tratteggiata generando a caso punti distribuiti uniformemente nel rettangolo ABCD di base $(b - a)$ e altezza $h \geq$ al valore massimo assunto da $f(x)$

Il primo metodo, detto “vinci o perdi” (in inglese *hit or miss*), si basa sull’interpretazione geometrica del valore di un integrale definito come misura dell’area sottesa da $f(x)$ e dall’asse delle ascisse nell’intervallo di integrazione $[a, b]$ (zona tratteggiata di Fig. 8.11).

Sfruttando da un diverso punto di vista le proprietà che stanno alla base del metodo del rigetto, possiamo infatti determinare I moltiplicando l’area di un rettangolo che racchiude $f(x)$ per la probabilità che un punto P qualsiasi all’interno del rettangolo si trovi anche all’interno dell’area sottesa da $f(x)$. Ricordando la relazione (7.37) abbiamo:

$$p = \frac{\text{area tratteggiata}}{\text{area rettangolo ABCD}} = \frac{I}{h \cdot (b - a)} \quad (8.32)$$

da cui:

$$I = pA, \quad \text{con } A = h \cdot (b - a). \quad (8.33)$$

Se generiamo all’interno del rettangolo ABCD N punti casuali uniformemente distribuiti $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$, e contiamo il numero N_S di “successi”, vale a dire il numero di volte in cui $y_i \leq f(x_i)$, attraverso il rapporto N_S/N otteniamo una valutazione approssimata di p (vedere Fig. 8.11) e la stima dell’integrale (8.31) diventa:

$$I = pA \simeq p_N A = A \frac{N_S}{N} \equiv I_N^{VP}. \quad (8.34)$$

Questo valore va considerato come la realizzazione di nuova variabile statistica I_N^{VP} . Per ricavare l’errore ad essa associato, basta osservare che N_S segue una

distribuzione binomiale di media $Np = NI/A$ e varianza $Np(1-p)$; otteniamo quindi:

$$\text{Var}[\mathbf{I}_N^{VP}] = \frac{A^2}{N^2} \text{Var}[\mathbf{N}_S] = \frac{I(A-I)}{N} \simeq \frac{I_N^{VP}(A-I_N^{VP})}{N}. \quad (8.35)$$

Il secondo metodo di integrazione MC, detto metodo della media o metodo grezzo (in inglese *crude Monte Carlo*), considera invece x come una variabile aleatoria uniformemente distribuita nell'intervallo di integrazione $[a, b]$. Sfruttando la definizione (2.67) di valore atteso di una funzione di variabile aleatoria, possiamo scrivere l'identità:

$$I = (b-a) \int_a^b \frac{1}{(b-a)} f(x) dx = \langle f(x) \rangle (b-a), \quad (8.36)$$

con $\langle f(x) \rangle$ pari alla media dei valori che la funzione da integrare assume in $[a, b]$.

Uno dei modi per valutare approssimativamente $\langle f(x) \rangle$ è quello di calcolare la media di N valori $f(x_1), f(x_2), \dots, f(x_N)$, con x_1, x_2, \dots, x_N scelti casualmente ed in maniera uniforme entro $[a, b]$:

$$I \simeq \frac{(b-a)}{N} \sum_{i=1}^N f(x_i) \equiv I_N^M. \quad (8.37)$$

La varianza della variabile statistica \mathbf{I}_N^M viene facilmente ricavata sfruttando le proprietà dell'operatore varianza introdotte nel par. 2.9:

$$\text{Var}[\mathbf{I}_N^M] \equiv \sigma_{I_N^M}^2 = \frac{(b-a)^2}{N^2} \text{Var} \left[\sum_{i=1}^N f(X_i) \right] = \frac{(b-a)^2}{N} \text{Var}[f(X_i)]. \quad (8.38)$$

Poiché $f(X_i)$ ora è vista come una funzione della variabile casuale X_i , tenendo presente la (8.36) e la (2.48) di pagina 58 otteniamo:

$$\begin{aligned} \text{Var}[\mathbf{I}_N^M] &= \frac{1}{N} \left[(b-a) \int_a^b f^2(x) dx - \left(\int_a^b f(x) dx \right)^2 \right] \\ &\simeq \frac{(b-a)^2}{N(N-1)} \left[\sum_{i=1}^N f^2(x_i) - \frac{1}{N} \left(\sum_{i=1}^N f(x_i) \right)^2 \right]. \end{aligned} \quad (8.39)$$

Il termine all'interno delle parentesi quadrate è una misura di quanto $f(x)$ si discosta dal suo valor medio nella regione di integrazione, per cui $\text{Var}[\mathbf{I}_N^M]$ dipende fortemente dalla larghezza del codominio di $f(x)$.

Esercizio 8.1

Calcolare con il metodo MC l'integrale

$$I = \int_0^{\frac{\pi}{2}} \sqrt{\text{sen } x} dx. \quad (8.40)$$

Risposta. Nonostante l'apparente semplicità, questo integrale non è risolvibile analiticamente. La soluzione, ricavabile per via numerica, vale:

$$I = \int_0^{\frac{\pi}{2}} \sqrt{\sin x} \, dx = 1.19814 \dots \quad (8.41)$$

Applichiamo i due metodi di integrazione MC presentati in precedenza utilizzando la routine `MCinteg`, reperibile nelle nostre pagine web [RPP].

a) Metodo della media secondo le equazioni (8.37, 8.39).

Generando 1000 punti casuali abbiamo ottenuto il risultato $I = 1.199 \pm 1.2 \cdot 10^{-2}$, riportato anche nella prima riga della Tab. 8.3.

Come abbiamo già notato in precedenza, se si vuole ottenere una soluzione molto precisa (per esempio entro qualche per mille) occorre generare un grande numero di punti, data la bassa velocità di convergenza del risultato MC ad I .

b) Metodo “vinci o perdi” (eqq. 8.34, 8.35).

In questo caso abbiamo $h = 1$, $(b - a) = \pi/2$ e generando sempre 1000 punti, abbiamo ottenuto $I = 1.202 \pm 2.1 \cdot 10^{-2}$, risultato con un errore quasi doppio rispetto a quello precedente.

Nella terminologia del metodo MC, si definisce come efficienza η di un programma di simulazione la quantità:

$$\eta = 1/(t_N \text{var}[S_N]) \quad (8.42)$$

in cui S_N è la stima fornita dal programma e t_N il tempo necessario per calcolarla. Nel nostro esempio, dato che il t_N dei due algoritmi utilizzati è all'incirca uguale, con il metodo “vinci o perdi” abbiamo quindi diminuito di circa 4 volte l'efficienza del programma di integrazione numerica.

Tabella 8.3. Confronto tra le stime dell'integrale (8.40) ottenute generando 1 000 punti casuali col codice `MCinteg` ed utilizzando i diversi algoritmi MC spiegati nel testo

Algoritmo	I_N	σ_{I_N}
Metodo della media	1.199	$1.2 \cdot 10^{-2}$
Metodo “vinci o perdi”	1.202	$2.1 \cdot 10^{-2}$
Campion. importanza	1.1987	$2.5 \cdot 10^{-3}$
Camp. stratificato		
a) proporzionale	1.1981	$1.0 \cdot 10^{-3}$
b) ottimale	1.19785	$6.8 \cdot 10^{-4}$

Tale differenza dipende ovviamente dal particolare integrale considerato, ma è facile dimostrare (si veda, ad esempio, [Jam80]) che con l'algoritmo "vinci o perdi" si ottiene sempre un risultato meno preciso rispetto a quello della media. Questo fatto può essere compreso intuitivamente osservando che, con il metodo "vinci o perdi" per ogni punto generato x_i , invece di aggiungere il corrispondente valore $f(x_i)$, viene sommato il valore 1 con probabilità $f(x_i)$. Si sostituisce così una stima a quello che è il valore esatto e, per questo, si introduce un errore supplementare.

8.9 Campionamento ad importanza

Il metodo della media viene denominato "grezzo", in contrapposizione a tecniche più raffinate, attraverso le quali si riesce a migliorare, a parità di N , la precisione del risultato, in accordo con la (7.7), riducendo la varianza σ_T associata alla distribuzione della variabile casuale simulata T_i .

Nel caso dell'integrazione, questa operazione viene relativamente facilitata in quanto σ_T coincide con la deviazione standard della funzione da integrare (vedere eq. 8.38) ed ogni sua manipolazione è in genere abbastanza semplice.

Ad esempio, commentando la (8.39) abbiamo notato che quanto più grande è la variazione di $f(x)$ nella regione di integrazione, tanto più elevato sarà anche l'errore sul risultato mentre, al contrario, esso diventa più preciso quando i valori generati di $f(x)$ non sono troppo dissimili tra loro. Si può quindi rendere più precisa la stima di MC di I quando si riesce a trovare una funzione $g(x)$ integrabile tale che $g(x) \simeq f(x)$, in modo tale da riscrivere I come:

$$I = \int_a^b f(x) dx = \int_a^b \frac{f(x)}{g(x)} g(x) dx = \int_a^b \frac{f(x)}{g(x)} dG(x), \quad (8.43)$$

con

$$G(x) = \int g(x) dx. \quad (8.44)$$

Se supponiamo che $g(x)$ sia anche normalizzata entro $[a, b]$, è facile concludere che l'equazione (8.43) non rappresenta altro che la media della funzione di variabile casuale $f(X)/g(X)$, dove X è una variabile aleatoria di densità $g(x)$, in accordo con la (2.67). Possiamo quindi scrivere I come

$$I = \left\langle \frac{f(X)}{g(X)} \right\rangle. \quad (8.45)$$

Denotando con X_1, X_2, \dots, X_N una campione generato casualmente secondo la densità di probabilità $g(x)$ otteniamo:

$$I \simeq \frac{1}{N} \sum_{i=1}^N \frac{f(X_i)}{g(X_i)} = I_N^{CI}, \quad X \sim g(x). \quad (8.46)$$

Invece di generare uniformemente x per integrare $f(x)$, si genera una variabile aleatoria distribuita come $g(x)$ per integrare $f(x)/g(x)$, dando così più peso a quelle parti più “importanti” di $f(x)$, da cui il nome di campionamento per importanza, dato a questo algoritmo.

La varianza finale dipende ora dal rapporto $f(x)/g(x)$ e, richiamando la (8.39), essa risulta:

$$\text{var}[I_N^{CI}] = \frac{1}{N} \left[\int_a^b \frac{f^2(x)}{g^2(x)} dG(x) - I^2 \right] = \frac{1}{N} \left[\int_a^b \frac{f^2(x)}{g(x)} dx - I^2 \right] \quad (8.47)$$

oppure, in maniera approssimata, come nella (8.39)

$$\text{var}[I_N^{CI}] \simeq \frac{1}{N(N-1)} \left\{ \sum_{i=1}^N \frac{f^2(x_i)}{g^2(x_i)} - \frac{1}{N} \left[\sum_{i=1}^N \frac{f(x_i)}{g(x_i)} \right]^2 \right\}. \quad (8.48)$$

Ricordiamo che in queste formule x è campionato dalla densità $g(x)$.

Dalla (8.47) si nota immediatamente come la varianza diventi nulla se $g(x) = f(x)/I$. Sfortunatamente, nei problemi reali, non potremo però mai utilizzare tale equazione poichè occorre conoscere I , che è esattamente il problema da cui eravamo partiti. Tuttavia la (8.47) dimostra quantitativamente come $g(x)$ debba essere scelta il più possibile uguale ad $f(x)$, per fare in modo che il rapporto tra le due funzioni vari in maniera molto limitata nella regione di integrazione e per massimizzare così il guadagno in precisione raggiungibile.

8.10 Campionamento stratificato

L'idea che sta alla base del campionamento stratificato (*stratified sampling*), tecnica molto nota ed usata in statistica, è simile a quella che abbiamo appena descritto: si concentra un maggior numero di punti in quelle zone che sono più importanti dal punto di vista del calcolo, ma con la differenza che, invece di cambiare la funzione da integrare, si suddivide la regione di integrazione in vari sottointervalli, generando poi punti in maniera uniforme ma con densità differenti a seconda del sottointervallo considerato.

Dividiamo l'intervallo $[a, b]$ in k segmenti definiti dai punti $a = \alpha_0 < \alpha_1 < \alpha_2 < \dots < \alpha_k = b$ e denotiamo con $\Delta_j = (\alpha_j - \alpha_{j-1})$ ed N_j la larghezza ed il numero di punti da generare nel generico j -esimo sottointervallo. Per le note proprietà di somma dell'integrale, possiamo inoltre scrivere:

$$I = \int_a^b f(x) dx = \sum_{j=1}^k \int_{\alpha_{j-1}}^{\alpha_j} f(x) dx = \sum_{j=1}^k I_j. \quad (8.49)$$

Nel campionamento stratificato, ciascuno degli integrali I_j che compaiono nel termine di destra dell'equazione precedente viene approssimato con il metodo

della media:

$$I_j = \int_{\alpha_{j-1}}^{\alpha_j} f(x) dx \simeq \frac{\Delta_j}{N_j} \sum_{i=1}^{N_j} f(\alpha_{j-1} + \Delta_j \xi_{ij}) = \frac{\Delta_j}{N_j} \sum_{i=1}^{N_j} f(x_{ij}) \quad (8.50)$$

(x_{ij} rappresenta l' i -esimo punto generato dalla distribuzione uniforme nel j -esimo sottointervallo) e la stima di I diventa così :

$$I = \sum_{j=1}^k \Delta_j \langle f_j(X) \rangle \simeq \sum_{j=1}^k \sum_{i=1}^{N_j} \frac{\Delta_j}{N_j} f(x_{ij}) = I_N^{CS} \quad (8.51)$$

in cui con il simbolo $\langle f_j(X) \rangle$ abbiamo rappresentato il valor medio di $f(X)$ nel j -esimo sottointervallo.

La varianza globale, che non è altro che la somma delle varianze dei singoli I_j presi separatamente, risulta in base alla (8.39):

$$\text{var}[I_N^{CS}] = \sum_{j=1}^k \frac{\Delta_j^2}{N_j} \sigma_j^2 = \sum_{j=1}^k \frac{1}{N_j} \left\{ \Delta_j \int_{\alpha_{j-1}}^{\alpha_j} f^2(x) dx - \left[\int_{\alpha_{j-1}}^{\alpha_j} f(x) dx \right]^2 \right\} \quad (8.52)$$

(σ_j^2 rappresenta la varianza di $f(x)$ nel j -esimo sottointervallo), oppure, in maniera approssimata:

$$\text{var}[I_N^{CS}] \simeq \sum_{j=1}^k \frac{\Delta_j^2}{N_j(N_j - 1)} \left\{ \sum_{i=1}^{N_j} f^2(x_{ij}) - \frac{1}{N_j} \left[\sum_{i=1}^{N_j} f(x_{ij}) \right]^2 \right\}. \quad (8.53)$$

L'errore sul risultato finale dipende ora, oltre che dall'andamento di $f(x)$, sia dal modo in cui si suddivide il dominio di integrazione sia da come si distribuiscono i punti all'interno di ogni sottointervallo.

Una volta fissati i vari sottointervalli Δ_j , dalla (8.52) con un procedimento non molto complicato ma noioso, si può dedurre che la migliore scelta per N_j si ha quando [Coc77]:

$$N_j = \frac{N \Delta_j \sigma_j}{\sum_{j=1}^k \Delta_j \sigma_j}. \quad (8.54)$$

Questa equazione indica, come era logico attendersi, di concentrare la generazione casuale nei sottointervalli più larghi ed in quelli in cui $f(x)$ presenta le maggiori variazioni. Anche in questo caso però tale risultato non è direttamente applicabile in quanto i valori σ_j non sono ovviamente noti a priori. Per utilizzare la formula precedente, si effettua di solito un breve test preliminare per ottenere una stima abbastanza corretta di σ_j da cui ricavare poi un valore opportuno per N_j . Nel fare questo occorre operare un ragionevole compromesso tra il maggior tempo di calcolo necessario e l'aumento in precisione che si riesce ad ottenere sul risultato finale.

Quando è troppo lungo o complicato utilizzare questo procedimento, è possibile dimostare (vedere di nuovo [Coc77]) che il miglior modo di procedere è quello di generare un numero di punti proporzionale alla lunghezza di ciascun sottointervallo:

$$N_j = N \frac{\Delta_j}{(b-a)} . \quad (8.55)$$

Questa proprietà può essere compresa intuitivamente se osserviamo che con il campionamento stratificato proporzionale si migliora, rispetto al metodo della media, l'uniformità di generazione dei punti casuali riducendo così le fluttuazioni statistiche dovute a loro addensamenti in zone particolari dell'intervallo di integrazione.

La lunghezza dei sottointervalli può invece essere ottimizzata solo quando i punti N_j vengono scelti in base all'equazione (8.54) e il dominio di integrazione è contemporaneamente suddiviso in un gran numero di strati (vedere sempre [Coc77]), condizioni che non sempre sono in pratica soddisfatte. In caso contrario non ci sono regole precise di comportamento; di solito, per semplificare i programmi, i sottointervalli Δ_j vengono presi tutti della stessa lunghezza:

$$\Delta_j = \frac{(b-a)}{k} \quad \forall j = 1, \dots, k . \quad (8.56)$$

Esercizio 8.2

Calcolare l'integrale 8.40 utilizzando il campionamento per importanza e quello stratificato.

Risposta.

a) *Campionamento per importanza.*

Dato che in un intorno dello zero della funzione $\sin x$ si ha:

$$\sin x \simeq x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots \quad (8.57)$$

possiamo tentare di approssimare la funzione da integrare ($\sqrt{\sin x}$) con $h(x) = \sqrt{x}$ e di scegliere quindi questa forma per $g(x)$, come mostrato in Fig. 8.12. Abbiamo quindi la densità normalizzata

$$g(x) = \frac{3}{\pi} \sqrt{\frac{2}{\pi}} \sqrt{x} , \quad (8.58)$$

cui corrisponde la funzione cumulativa:

$$G(x) = \int_0^x g(x) dx = \left(\frac{2}{\pi} x \right)^{3/2} , \quad (8.59)$$

e l'equazione (7.14) di pagina 270 dà come risultato:

$$x = \frac{\pi}{2} (\xi)^{2/3} . \quad (8.60)$$

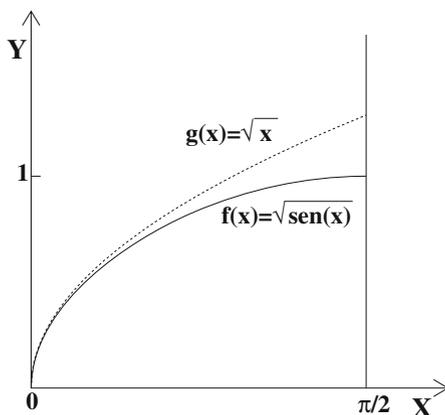


Fig. 8.12. Confronto tra le funzioni $f(x) = \sqrt{\text{sen } x}$ e $h(x) = \sqrt{x}$ nell'intervallo $[0, \pi/2]$

Le variabili x generate con questa formula sono utilizzate per il calcolo dell'integrale secondo la (8.46). Il calcolo è eseguito dalla nostra routine **Mcinteg**.

Dal risultato ottenuto con $N = 1000$, riportato nella terza riga della Tab. 8.3, si nota chiaramente un miglioramento di quasi un ordine di grandezza della precisione rispetto al metodo della media. Dato che, pure in questo caso, il tempo di calcolo è all'incirca uguale per i due algoritmi, il corrispondente guadagno in efficienza risulta essere di un fattore quasi 100! Per applicare questo metodo, $g(x)$ deve avere una forma abbastanza semplice (è necessario conoscere analiticamente la sua primitiva), pur dovendo riprodurre abbastanza bene l'andamento, che può essere molto complicato, di $f(x)$. Nel caso multidimensionale, per minimizzare il tempo di calcolo, è decisamente preferibile che $G(x_1, x_2, \dots, x_n)$ sia una funzione separabile:

$$G(x_1, x_2, \dots, x_n) = g(x_1) \cdot g(x_2) \cdot \dots \cdot g(x_n) .$$

Le classi di funzioni che soddisfano a queste condizioni sono poche: essenzialmente, le trigonometriche, le esponenziali, i polinomi di ordine non molto elevato e qualche loro combinazione.

b) **Campionamento stratificato.**

Effettuiamo questo calcolo utilizzando la tecnica di campionamento stratificato uniforme, prendendo sottointervalli uguali e generando in ognuno di essi lo stesso numero di punti.

Suddividendo il dominio di integrazione in 20 sottointervalli e generando 1000 punti abbiamo ottenuto il valore riportato nella quarta riga della Tab. 8.37. Rispetto al metodo della media, notiamo un miglioramento nella deviazione standard di un fattore $\simeq 12$, il che, tenendo conto

di circa un 30% di tempo di calcolo in più, si traduce pur sempre in un aumento di efficienza di oltre 2 ordini di grandezza.

Con 10 000 punti e 100 strati abbiamo ottenuto un risultato notevolmente accurato:

$$I = 1.198154 \pm 0.000023 .$$

A parità di punti e di strati, la precisione può essere ulteriormente incrementata se il numero di punti in ogni sottointervallo viene determinato con il metodo ottimale (equazione (8.54)), come è mostrato nell'ultima riga della Tab. 8.37. Anche questo algoritmo è presente nella routine

`Mcinteg`.

Oltre a quelli che vi abbiamo appena presentato, vi sono diversi altri metodi con cui si riesce ad ottenere una riduzione della varianza; chi fosse interessato può consultare [Kah56], [KW86], [Rub81], [Rip86]).

8.11 Integrali multidimensionali

Abbiamo mostrato come attraverso l'uso delle tecniche di riduzione della varianza, si riesce a migliorare, anche in maniera notevole, la precisione sulla stima MC del valore di un integrale definito. Però, nonostante questi progressi, per funzioni ad una dimensione questo metodo risulta sempre meno efficiente degli altri procedimenti di approssimazione numerica la cui velocità di convergenza va come N^{-k} con $k \geq 2$

La situazione però cambia quando passiamo al caso multidimensionale, poiché l'errore del metodo MC è indipendente dalla dimensionalità d dell'integrale mentre la precisione delle altre tecniche numeriche varia come $N^{-k/d}$ con $k \geq 2$ (un procedimento che richiede N punti in una dimensione, ne richiede, a parità di precisione, N^2 in due dimensioni, N^3 in tre dimensioni e così via). Inoltre tutti gli altri metodi assumono un andamento polinomiale della funzione da integrare per cui presentano problemi di applicazione in caso di discontinuità. Per tutte queste ragioni, il metodo MC inizia ad essere conveniente per integrali a 5 dimensioni, diventando l'unico in concreto utilizzabile a 10 o più dimensioni.

Nel caso in cui dobbiate affrontare per la prima volta il problema di calcolare l'integrale multidimensionale:

$$I = \int_{\Omega} f(x_1, \dots, x_d) dx_1 \dots dx_d \quad (8.61)$$

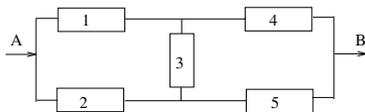
mediante il metodo MC, vi consigliamo innanzitutto di utilizzare dei programmi già pronti piuttosto che iniziare a scrivere un vostro codice personale. Sul mercato sono disponibili diversi codici affidabili che sfruttano le tecniche di riduzione della varianza precedentemente introdotte (vedere, ad esempio, [PFTW92]). In questi codici, per semplificare e rendere più affidabile

l'algoritmo di calcolo, l'integrazione viene effettuata su un dominio avente limiti di integrazione indipendenti (un ipercubo o un iperrettangolo) ottenuto effettuando un opportuno cambiamento di coordinate. Se la trasformazione di coordinate risulta troppo complessa, è preferibile considerare geometricamente l'integrale (8.61) come il volume di un solido W nello spazio a $d + 1$ dimensioni (x_1, \dots, x_d, y) ed applicare il metodo "vinci o perdi" che abbiamo introdotto nel par 8.8.

Problemi

8.1. Spesso nei controlli di qualità si usa lo scarto massimo: si sceglie un parametro caratteristico X , lo si considera come variabile aleatoria normale e si fa la differenza tra i valori massimi e minimi di X riscontrati in un lotto di controllo di n pezzi. Sapendo che lo scarto quadratico medio della produzione di X è $\sigma = 0.5$, scrivere un codice di simulazione per trovare il valore $\Delta = x_{\max} - x_{\min}$ al di sopra del quale scartare la produzione con $CL = 99\%$ per un lotto di controllo di $n = 100$ pezzi.

8.2. Una macchina si compone di cinque pezzi, secondo lo schema di flusso in figura.



Il sistema funziona se il "flusso di funzionamento" passa da A a B , cioè se funzionano almeno i pezzi $(1,3,5)$, $(2,3,4)$, $(1,4)$ e $(2,5)$. Sapendo che ogni pezzo può guastarsi con tempi dati dalla legge esponenziale negativa con valori medi rispettivamente di $\langle t_1 \rangle, \langle t_2 \rangle, \langle t_4 \rangle, \langle t_5 \rangle = 3$ giorni e $\langle t_3 \rangle = 5$ giorni, determinare con una simulazione il tempo di funzionamento medio dell'apparecchio.

8.3. Modificare la routine di integrazione `Mcinteg` (utilizzata negli esercizi 8.1 e 8.2) adattandola al calcolo dell'integrale della gaussiana standard

$$(1/\sqrt{2\pi}) \int_0^x \exp(-t^2/2) dt$$

e confrontare i dati con quelli di Tab. D.1 per vari valori di x . Escludere il campionamento ad importanza.

8.4. Utilizzando tutti i metodi di risoluzione discussi nel testo, calcolare con il metodo MC, eventualmente modificando la routine `Mcinteg`, l'integrale

$$I = \int_0^1 \log(1+x) dx .$$

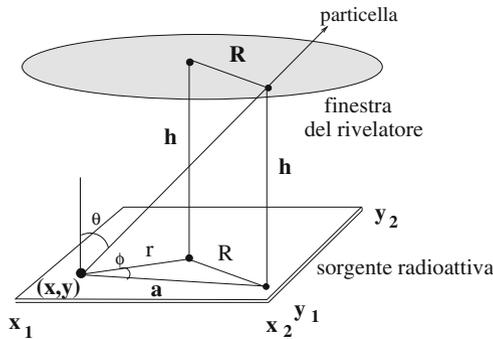
8.5. Calcolare con il metodo MC l'integrale: $\int_{-1}^1 \int_{-1}^1 \int_{-1}^1 (x_1^2 + x_2^2 + x_3^2) dx_1 dx_2 dx_3 .$

8.6. Calcolare, col metodo vinci o perdi, l'area dell'ellisse

$$\frac{x^2}{2} + y^2 = 1 .$$

8.7. Costruite un algoritmo di Metropolis per stimare la media e la varianza di una distribuzione gaussiana standard, usando una simulazione di lunghezza N . Usate come distribuzione ausiliaria una $U(-a, a)$. Estraiete poi un campione di dimensione N da una gaussiana standard con la routine `grand` di SCILAB. Confrontate le traiettorie degli stimatori al crescere di N con $a = 2$. Ripetete gli esperimenti cambiando il valore di a : la velocità di convergenza delle traiettorie degli stimatori di media e varianza è influenzata da a ? È ragionevole usare $a > 3$?

8.8. Calcolare l'efficienza geometrica del rivelatore in figura, avente una finestra di ingresso di raggio $R_d = 3$ cm, considerando una sorgente piana di lati $(x_2 - x_1) = 4$ cm e $(y_2 - y_1) = 6$ cm, posta ad una distanza $h = 5$ cm dal rivelatore. La sorgente è radioattiva ed emette particelle in modo isotropo secondo gli angoli (θ, ϕ) da punti (x, y) distribuiti in modo uniforme. Utilizzare la geometria suggerita in figura.



8.9. La distribuzione di von Mises ha densità:

$$p(x) = \frac{1}{2\pi I_0(c)} e^{c \cos(x)}, \quad -\pi \leq x \leq \pi ,$$

dove $I_0(c)$ è la funzione di Bessel modificata del primo tipo e di ordine zero. Utilizzare l'algoritmo di Metropolis per generare un campione da questa distribuzione senza calcolare $I_0(c)$.

8.10. Generare diversi campioni dal modello di Ising con $\beta = 0.3$ e $T = 1$, accrescendo di volta in volta la dimensione $n \times n$ del reticolo. Sulla base del tracciato del numero di atomi con spin 1 e tenendo fisso a 100 000 il numero N di iterazioni, qual è a vostro parere il massimo valore di n per il quale l'algoritmo può dirsi "funzionante"?

Inferenza statistica e verosimiglianza

Parlando della conseguenza più probabile, bisogna tener presente che, in realtà, la probabilità di passaggio in stati con entropia maggiore è talmente più grande rispetto alla probabilità di una sua diminuzione più o meno percepibile, che questa ultima non può praticamente essere osservata in natura.

L.D. Landau, E.M. Lifchitz e L.P. Pitaevskij,
Fisica statistica

9.1 Introduzione

Nel cap. 6 vi abbiamo presentato la teoria della stima e la verifica delle ipotesi nell'ambito della statistica elementare: da un insieme di dati si effettua la stima di una quantità (media, probabilità o coefficiente di correlazione) con il relativo errore e il relativo intervallo di confidenza. Successivamente, se necessario, si verifica, attraverso il test χ^2 o altri test, la compatibilità di questa stima con un modello, denominato generalmente ipotesi nulla H_0 .

In questo schema, che potremmo definire “statico”, il modello da verificare è dato a priori e non è modificato dalle informazioni provenienti dal campione raccolto per effettuare la stima.

Si può però adottare un approccio più “dinamico” ed efficiente, che consiste nel determinare, in base ai dati raccolti, *il modello di popolazione più verosimile*, nel modificare conseguentemente gli intervalli di stima (che, come sappiamo, possono dipendere dal modello di popolazione assunta) e nell'eseguire quindi il test di compatibilità tra dati e modello. Lo schema è quello di Fig. 9.1, dove la frecce in grassetto indicano le differenze rispetto al modello statico, considerato finora. I metodi che consentono di determinare da un insieme di dati il modello di popolazione più verosimile migliorando la stima dei parametri si chiamano metodi di adattamento ottimale, oppure, secondo la denominazione inglese, *metodi di best fit*. Nel seguito useremo liberamente questo termine inglese, che è di uso corrente.

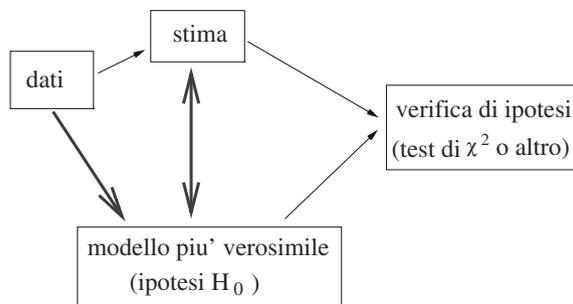


Fig. 9.1. Stima di parametri e verifica di ipotesi. Le frecce in grassetto indicano le fasi di ottimizzazione del modello

L'ottimizzazione del modello sulla base dei dati avviene in pratica quasi sempre assumendo nota la forma della funzione densità (binomiale, gaussiana, poissoniana, uniforme o altro) e considerando i parametri caratteristici della distribuzione quali la media, la varianza o i limiti di esistenza, *come parametri liberi*.

Nel corso del capitolo useremo la notazione seguente: se \mathbf{X} è un vettore di variabili aleatorie, ognuna di esse può dare luogo ad un campione, secondo la def. 2.12 di pagina 63. Diremo quindi che

$$\mathbf{x} = (x_1, x_2, \dots, x_m) \quad (9.1)$$

sono i valori osservati di $\mathbf{X} = (X_1, X_2, \dots, X_m)$ in una prova e che

$$\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{mi})$$

sono i valori assunti da \mathbf{X} nell' i -esima prova. Introduciamo inoltre una nuova notazione, che distingue tra la variabile \mathbf{X} ed un campione di dimensione n :

$$\underline{\mathbf{X}} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) . \quad (9.2)$$

I valori assunti dal campione dopo l'esperimento (dato da n prove) saranno:

$$\underline{\mathbf{x}} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) . \quad (9.3)$$

Considereremo dunque uno spazio di probabilità $(S, \mathcal{F}, P_\theta)$ dipendente da uno o più parametri, secondo la (6.1) di pagina 181. La densità da ottimizzare è pertanto del tipo $p(\mathbf{x}; \theta)$, dove $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ è un parametro p -dimensionale. Ottimizzare la densità significa allora determinare i valori dei parametri θ che, *fissata a priori una forma funzionale*, meglio si adattano al campione di dati raccolto.

Consideriamo per cominciare una sola variabile aleatoria X e siano $\underline{x} = (x_1, x_2, \dots, x_n)$ i valori osservati in n prove indipendenti. Se $p(x; \theta)$ è la densità di X (dipendente da un parametro θ), applicando la legge delle probabilità

composte al caso di n prove indipendenti effettuate sulla stessa variabile ed il teorema 4.1 di pagina 120, possiamo associare all'osservazione la densità di probabilità:

$$L(\boldsymbol{\theta}; \underline{\mathbf{x}}) = p(x_1; \boldsymbol{\theta})p(x_2; \boldsymbol{\theta}) \cdots p(x_n; \boldsymbol{\theta}) \equiv \prod_{i=1}^n p(x_i; \boldsymbol{\theta}) . \quad (9.4)$$

Questo prodotto, visto come funzione di $\boldsymbol{\theta}$, viene detto *funzione di verosimiglianza (likelihood function)* e, per ogni $\boldsymbol{\theta}$ fissato, rappresenta la probabilità, a meno dei fattori differenziali sui quali eventualmente integrare, di ottenere i valori $\underline{\mathbf{x}}$.

Nel caso di m variabili, la funzione di verosimiglianza si generalizza in modo ovvio tramite le (9.1-9.3):

$$\begin{aligned} L(\boldsymbol{\theta}; \underline{\mathbf{x}}) &= p(x_{11}, x_{21}, \dots, x_{m1}; \boldsymbol{\theta}) p(x_{12}, x_{22}, \dots, x_{m2}; \boldsymbol{\theta}) \dots \\ &\times p(x_{1n}, x_{2n}, \dots, x_{mn}; \boldsymbol{\theta}) = \prod_{i=1}^n p(\mathbf{x}_i; \boldsymbol{\theta}) , \end{aligned} \quad (9.5)$$

dove la produttoria si intende estesa a tutti gli n valori sperimentali ottenuti per ognuna delle m variabili \mathbf{X} . La definizione generale della funzione di verosimiglianza comprende anche il caso di prove non indipendenti, tuttavia non viene riportata, visto che non prenderemo in considerazione questa eventualità.

Come vedremo, le proprietà matematiche della funzione di verosimiglianza che interessano in statistica sono le stesse del suo logaritmo. È allora possibile eliminare la produttoria nelle (9.4, 9.5) definendo la nuova funzione

$$\mathcal{L} = -\ln(L(\boldsymbol{\theta}; \underline{\mathbf{x}})) = -\sum_{i=1}^n \ln(p(\mathbf{x}_i; \boldsymbol{\theta})) , \quad (9.6)$$

dove va notato il segno $-$ davanti al logaritmo. Se si adotta questa convenzione, ad un massimo di L corrisponde un minimo di \mathcal{L} . Nel seguito, per semplificare i calcoli, utilizzeremo spesso, invece della funzione di verosimiglianza L , il suo logaritmo negativo \mathcal{L} .

9.2 Il metodo della massima verosimiglianza

Il metodo della massima verosimiglianza per la stima di $\boldsymbol{\theta}$, che nel seguito verrà abbreviato con la notazione inglese ML (da *Maximum Likelihood*), fu introdotto da R.A. Fisher nel 1912. Esso può essere enunciato nel modo seguente:

Definizione 9.1 (Metodo della massima verosimiglianza (ML)). *Dato un insieme di valori osservati $\underline{\mathbf{x}} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ da un campione casuale*

$\underline{\mathbf{X}} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ di densità ipotizzata $p(\mathbf{x}; \boldsymbol{\theta})$, dove $\boldsymbol{\theta}$ è un parametro che varia in un insieme Θ , la stima di massima verosimiglianza $\hat{\boldsymbol{\theta}}$ di $\boldsymbol{\theta}$ è il punto di massimo (se esiste) della funzione di verosimiglianza (9.5). Sinteticamente:

$$\max_{\boldsymbol{\theta} \in \Theta} [L(\boldsymbol{\theta}; \underline{\mathbf{x}})] = \max_{\boldsymbol{\theta} \in \Theta} \left[\prod_{i=1}^n p(\mathbf{x}_i; \boldsymbol{\theta}) \right] = L(\hat{\boldsymbol{\theta}}; \underline{\mathbf{x}}). \quad (9.7)$$

In alternativa, il principio richiede di minimizzare la funzione logaritmica \mathcal{L} della (9.6). In questo caso, il punto di estremo della funzione si ottiene facilmente risolvendo rispetto a $\boldsymbol{\theta}$ le equazioni di verosimiglianza:

$$\frac{\partial \mathcal{L}}{\partial \theta_k} = \sum_{i=1}^n \left[\frac{1}{p(\mathbf{x}_i; \boldsymbol{\theta})} \frac{\partial p(\mathbf{x}_i; \boldsymbol{\theta})}{\partial \theta_k} \right] = 0, \quad (k = 1, 2, \dots, p). \quad (9.8)$$

Affinché il principio risulti chiaro, vanno dette subito tre cose:

- *prima della prova*, la funzione di verosimiglianza $L(\boldsymbol{\theta}; \underline{\mathbf{X}})$ è proporzionale alla densità di probabilità di $(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$. Si ha in generale proporzionalità e non coincidenza perché nella massimizzazione della funzione di verosimiglianza vengono talvolta omessi fattori costanti ininfluenti sul punto di massimo $\hat{\boldsymbol{\theta}}$;
- nella massimizzazione della verosimiglianza *dopo la prova*, si utilizzano i valori osservati $\underline{\mathbf{x}}$ delle variabili $\underline{\mathbf{X}}$. Le $\underline{\mathbf{x}}$ sono pertanto parametri fissati e non più variabili;
- dopo la prova la funzione di verosimiglianza dipende allora solo dai valori $\boldsymbol{\theta}$ che rappresentano le variabili rispetto alle quali massimizzare (o minimizzare). Il punto di massimo (o di minimo) $\hat{\boldsymbol{\theta}}$ rappresenta la stima ML del parametro che si può effettuare dai dati.

Vediamo ora come funziona il principio.

Esercizio 9.1

In n tentativi si sono ottenuti x successi. Trovare la stima ML della probabilità p .

Risposta. Assegniamo alla popolazione la densità binomiale (2.30), che per comodità riscriviamo:

$$b(n, p; x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}.$$

Poiché in questo caso abbiamo un solo valore, la funzione di verosimiglianza coincide con la binomiale, che va massimizzata rispetto a p , essendo noti e fissati dalla prova effettuata i valori di x ed n . Poiché il termine contenente i fattoriali non contiene p , esso può essere trascurato nella fase di massimizzazione. Minimizziamo per semplicità la verosimiglianza logaritmica (9.6), che in questo caso diventa:

$$\mathcal{L} = -x \ln(p) - (n-x) \ln(1-p).$$

Il minimo rispetto a p si trova annullando la derivata:

$$\frac{d\mathcal{L}}{dp} = -\frac{x}{p} + \frac{n-x}{1-p} = 0 ,$$

da cui risulta, utilizzando la notazione della (9.7):

$$\hat{p} = \frac{x}{n} = f . \quad (9.9)$$

È anche facile vedere, dalla derivata seconda, che questo è il valore di minimo della funzione.

Il risultato mostra che la stima ML della probabilità coincide con la frequenza osservata. All'inizio del libro abbiamo postulato l'esistenza della probabilità e la convergenza ad essa della frequenza empirica, come indicato nella (1.3). Alternativamente, avremmo potuto assumere come punto di partenza la stima ML e da questa dedurre la (1.3). In effetti, la (9.9) è un caso particolare di questo principio: gli stimatori fondamentali della statistica possono essere dedotti assumendo come punto di partenza il metodo ML. Nel seguito vedremo presto altri esempi di questo fatto.

I due esempi che seguono vi mostreranno una funzione di verosimiglianza prodotto di densità.

Esercizio 9.2

Due esperimenti hanno fornito rispettivamente x_1 successi su n_1 tentativi e x_2 successi su n_2 tentativi. Trovare la stima ML di p .

Risposta. Operando come nell'esercizio precedente, possiamo scrivere, a meno di fattori costanti inessenziali, la funzione ML come il prodotto di due binomiali aventi la stessa probabilità p :

$$L = p^{x_1} p^{x_2} (1-p)^{n_1-x_1} (1-p)^{n_2-x_2} .$$

Operando con i logaritmi otteniamo:

$$\mathcal{L} = -(x_1 + x_2) \ln(p) - (n_1 - x_1 + n_2 - x_2) \ln(1-p) ,$$

da cui:

$$\frac{d\mathcal{L}}{dp} = -\frac{x_1 + x_2}{p} + \frac{(n_1 + n_2) - x_1 - x_2}{1-p} = 0 \implies \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} .$$

La stima ML non è altro che la somma dei successi sul numero totale di tentativi.

Esercizio 9.3

Dati gli n valori x_i di una variabile gaussiana monodimensionale, trovare la stima ML di media e varianza.

Risposta. Dato che abbiamo a che fare con n misure, la funzione di verosimiglianza è il prodotto di n gaussiane aventi tutte gli stessi parametri μ e σ :

$$L(\mu, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2}.$$

La verosimiglianza logaritmica (9.6) è pertanto:

$$\mathcal{L}(\mu, \sigma) = +\frac{n}{2} \ln(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2,$$

la quale fornisce, tramite l'annullamento delle derivate, le stime richieste:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \implies \hat{\mu} = \sum_{i=1}^n \frac{x_i}{n} \equiv m \\ \frac{\partial \mathcal{L}}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0 \implies \hat{\sigma}^2 = \sum_{i=1}^n \frac{(x_i - m)^2}{n}, \end{aligned}$$

dove nella stima della varianza si è dovuta sostituire alla media μ la soluzione del sistema $\hat{\mu} = m$.

È interessante notare che la stima ML della media coincide con la media m usata in statistica, mentre la stima ML di σ^2 fornisce lo stimatore (6.55) che è corretto solo asintoticamente.

Quando la funzione di verosimiglianza è derivabile rispetto alle componenti di θ , il metodo ML si riduce ad un problema di analisi differenziale. Tuttavia, le componenti di θ possono essere discrete oppure, pur essendo continue, possono non ammettere derivate continue. Bisogna allora ricorrere a metodi alle differenze finite o a considerazioni di estremo che dipendono dal tipo di problema. L'esercizio seguente ne è un esempio.

Esercizio 9.4

Il campionamento di una variabile estratta da un popolazione di densità uniforme (3.79):

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

ha dato gli n valori

$$\underline{x} = x_1 < x_2 \cdots < x_{n-1} < x_n.$$

Trovare la stima ML degli estremi a e b .

Risposta. *Avendo ordinato i valori in ordine crescente, deve valere la condizione:*

$$a \leq x_1, \quad b \geq x_n.$$

La funzione di verosimiglianza (9.5) ha l'espressione:

$$L(a, b; \underline{x}) = \frac{1}{(b-a)^n} \quad a \leq x_1, \quad b \geq x_n.$$

In questo caso, pur avendo a che fare con parametri continui, il massimo della funzione non può essere trovato per differenziazione. Dato che la funzione di verosimiglianza in questo caso è massima quando è minimo il denominatore, la stima ML di a e b coincide rispettivamente col più piccolo e col più grande dei valori osservati:

$$\hat{a} = x_1, \quad \hat{b} = x_n.$$

È importante notare che la parte rimanente dei valori osservati:

$$x_2, x_3, \dots, x_{n-1}$$

non fornisce alcuna informazione alla stima di a e b . Dato che gli estremi x_1 e x_n contengono tutta l'informazione per la stima, essi formano una statistica sufficiente per (a, b) .

Il concetto di statistica sufficiente sarà sviluppato in maggiore dettaglio nel prossimo paragrafo.

Abbiamo mostrato in cosa consiste il metodo ML e quali risultati fornisce. Riteniamo però utile suggerirvi un argomento intuitivo che spiega, come risulta dagli esercizi precedenti, *perché il metodo funziona*, cioè come mai esso fornisce delle stime ragionevoli dei parametri. Osservate la Fig. 9.2, dove il processo di massimizzazione della funzione di verosimiglianza viene schematizzato come uno scorrimento della densità sull'asse delle ascisse (variazione dei parametri legati alla media) e come un suo allargamento o restringimento (variazione dei parametri di dispersione). I valori osservati sono sull'asse delle ascisse, rappresentati dai punti in grassetto; essi si addenseranno in una o più regioni, con qualche valore disperso nelle altre zone. La funzione di verosimiglianza è data dal *prodotto delle ordinate dei dati osservati calcolate attraverso la funzione di densità*. Come si vede dalla figura, il massimo di verosimiglianza si ottiene quando la densità è "aggiustata al meglio" sui dati. A questo massimo corrisponde la stima ML dei parametri.

È ora arrivato il momento di compiere un piccolo sforzo di astrazione e studiare i risultati fondamentali della teoria degli stimatori. Dopo questo passaggio, che faremo nei prossimi due paragrafi, la teoria della stima vi apparirà più chiara e probabilmente anche più interessante.

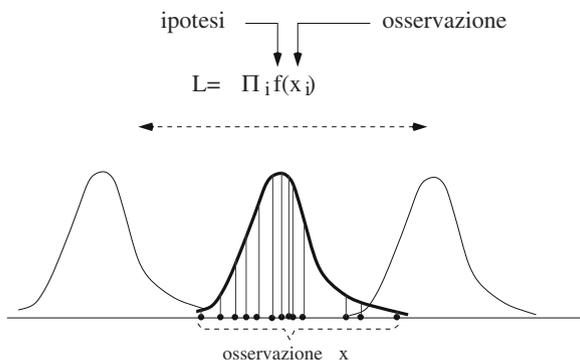


Fig. 9.2. Giustificazione intuitiva del metodo della massima verosimiglianza: la densità migliore è quella in grassetto, rispetto alla quale il prodotto delle ordinate è massimo

9.3 Proprietà degli stimatori

In questa sezione, le definizioni sono date per θ scalare, ma la 9.2, la 9.3 e la 9.4 valgono senza modifiche anche per parametri multidimensionali.

Per cominciare, torniamo alla definizione di stimatore, cui avevamo accennato brevemente nel par. 2.11.

Definizione 9.2 (Stimatore). Dato un campione \underline{X} di dimensione n da una variabile aleatoria m -dimensionale \mathbf{X} con densità $p(\mathbf{X}; \theta)$, $\theta \in \Theta$, si dice stimatore del parametro θ una statistica (secondo la definizione di pag. 64):

$$T_n(\underline{X}) \equiv t_n(\underline{X}) \quad (9.10)$$

con valori in Θ .

Lo stimatore è quindi definito come una funzione $T : S \rightarrow \Theta$ che porta dallo spazio campionario S allo spazio Θ dei parametri, utilizzata per stimare θ . Come sappiamo, la media M e la varianza S^2 di un campione sono stimatori dei parametri μ e σ^2 . Le convenzioni che useremo per stimatori e statistiche sono le seguenti: T_n (oppure T) è una variabile aleatoria data dalla (9.10), $t_n(\cdot)$ è la forma funzionale in esame, t_n (oppure t) è un valore assunto da T_n (o T) dopo una prova.

Uno stimatore ragionevole dovrebbe avvicinarsi sempre più al vero valore del parametro con l'aumentare delle osservazioni. Una proprietà desiderabile in questo senso è la consistenza:

Definizione 9.3 (Stimatore consistente). Uno stimatore T_n del parametro θ è detto consistente se converge in probabilità verso θ secondo la (2.72):

$$\lim_{n \rightarrow \infty} P\{|T_n - \theta| < \epsilon\} = 1, \quad \forall \epsilon > 0. \quad (9.11)$$

Come si vede, la consistenza dello stimatore richiede solo la convergenza in probabilità, anche se, per i casi considerati in questo testo, vale anche la convergenza quasi certa della (2.73).

L'insieme delle soluzioni delle equazioni di verosimiglianza (se si ha un massimo), che abbiamo indicato come $\hat{\theta}$, costituisce una *famiglia di stimatori* del parametro θ .

Utilizzando il valore atteso di uno stimatore (si veda il par. 2.11 per questo importante concetto), dalla disuguaglianza di Tchebychev (3.92) è facile mostrare che una condizione sufficiente perché valga la (9.11) è

$$\lim_{n \rightarrow \infty} \langle T_n \rangle = \theta \quad (9.12)$$

$$\lim_{n \rightarrow \infty} \text{Var}[T_n] = \lim_{n \rightarrow \infty} \left(\langle T_n^2 \rangle - \langle T_n \rangle^2 \right) = 0 . \quad (9.13)$$

Se n rimane finito, è lecito attendersi che anche la media vera della distribuzione degli stimatori T_n sia pari a θ . Tuttavia, non sempre questa condizione si verifica: un esempio che abbiamo già discusso più volte è quello della varianza campionaria non corretta (6.55). È quindi necessario introdurre la:

Definizione 9.4 (Stimatore corretto o non distorto (*unbiased*)). *Uno stimatore T_n di un parametro θ è corretto quando*

$$\langle T_n \rangle = \theta, \quad \forall n . \quad (9.14)$$

In caso contrario lo stimatore è detto distorto (biased) e si ha

$$\langle T_n \rangle = \theta + b_n , \quad (9.15)$$

dove b_n viene detto errore sistematico (bias).

Qualora non valga la (9.14) ma la (9.12) resti valida, allora

$$\lim_{n \rightarrow \infty} b_n = 0$$

e lo stimatore si dice asintoticamente corretto.

La varianza (6.55) è uno stimatore consistente, distorto e asintoticamente corretto.

Dopo la consistenza e la correttezza, la terza proprietà importante di uno stimatore è l'efficienza.

Definizione 9.5 (Stimatore più efficiente). *Dati due stimatori corretti T_n e Q_n di uno stesso parametro θ , si dice che T_n è più efficiente di Q_n se vale la relazione:*

$$\text{Var}[T_n] < \text{Var}[Q_n], \quad \forall \theta \in \Theta . \quad (9.16)$$

Chiaramente, dovendo scegliere tra due stimatori, a parità di altre condizioni quello più efficiente è preferibile, perché permette di ottenere intervalli di confidenza più piccoli nella stima di θ .

Un'importante proprietà che riguarda le statistiche è la sufficienza, introdotta da R.A. Fisher nel 1925. Riportiamo la formulazione più semplice di questa proprietà, che fa uso della funzione di verosimiglianza:

Definizione 9.6 (Statistica sufficiente). *La statistica S_n si dice sufficiente se la funzione di verosimiglianza (9.5) può essere fattorizzata nel prodotto di due funzioni h ed g tali che:*

$$L(\theta; \underline{\mathbf{x}}) = g(s_n(\underline{\mathbf{x}}), \theta) h(\underline{\mathbf{x}}), \quad (9.17)$$

dove la funzione $h(\underline{\mathbf{x}})$ non dipende da θ .

Nel caso di un parametro multidimensionale, la funzione g si scrive come $g(s_n(\underline{\mathbf{x}}), q_n(\underline{\mathbf{x}}), \dots, \theta)$ e si dice che le statistiche S_n, Q_n, \dots sono congiuntamente sufficienti per θ .

Su alcuni testi, la 9.17 è riportata come *teorema di fattorizzazione*. Infatti, la definizione di sufficienza viene talvolta espressa in termini molto generali e successivamente la (9.17) viene dimostrata essere una condizione necessaria e sufficiente per la validità della sufficienza. In questo testo noi adottiamo il teorema di fattorizzazione come *definizione* di sufficienza.

In pratica, la statistica sufficiente contiene tutte le informazioni sul parametro che si vuole stimare. Infatti, derivando la (9.17) rispetto a θ per ottenere il massimo di verosimiglianza, la funzione $h(\underline{\mathbf{x}})$ gioca il ruolo di una semplice costante, ed è pertanto ininfluenza nella stima di θ . Risulta altresì chiaro che, se S è una statistica sufficiente, una statistica $Q = f(S)$, dove f è una funzione invertibile di S , è anch'essa sufficiente, poiché la funzione di verosimiglianza può essere scritta nella forma

$$L(\theta; \underline{\mathbf{x}}) = g(f^{-1}(q), \theta) h(\underline{\mathbf{x}}). \quad (9.18)$$

Un buon esempio di statistica sufficiente, discusso nell'esercizio 9.4, è dato dai valori estremi di un campione estratto dalla densità uniforme.

Un esempio di statistiche congiuntamente sufficienti è dato invece dagli stimatori di media e varianza per campionamenti gaussiani:

$$T_n = \sum_i X_i^2, \quad Q_n = \sum_i X_i.$$

Infatti, dall'esercizio 9.3 si ha:

$$\begin{aligned} L(\mu, \sigma; \underline{\mathbf{x}}) &= \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2} \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2} (\sum_i x_i^2 - 2\mu \sum_i x_i + \mu^2 \sum_i 1)}, \end{aligned}$$

da cui:

$$L(\mu, \sigma; \underline{x}) = \frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}(t_n(\underline{x}) - 2\mu q_n(\underline{x}) + n\mu^2)} h \equiv g(t_n(\underline{x}), q_n(\underline{x}), \mu, \sigma) h ,$$

con h che in questo caso è una costante. Questa trattazione formale corrisponde al fatto pratico, ormai a voi ben noto, che per stimare media e varianza di un campione basta calcolare la media dei quadrati e il quadrato della media.

9.4 Teoremi sugli stimatori

In questo paragrafo abbiamo raccolto tutti i risultati importanti della teoria degli stimatori ML ed il teorema sul limite inferiore di Cramér-Rao, valido anche per altri stimatori. Considereremo per semplicità popolazioni con densità di probabilità $p(\mathbf{x}; \theta)$ dipendente da un parametro scalare θ . Tuttavia le formule restano valide anche nel caso di un vettore di parametri, se si sostituisce a θ il vettore $\boldsymbol{\theta}$ e alla derivata parziale $\partial/\partial\theta$ quella $\partial/\partial\theta_k$ rispetto a un qualunque elemento k -esimo del vettore $\boldsymbol{\theta}$. La generalizzazione a più parametri sarà discussa solo nei casi in cui essa non è del tutto ovvia.

Il primo teorema collega gli stimatori ML alle statistiche sufficienti, mostrando che la statistica sufficiente è quella che condensa nel modo migliore le informazioni su θ contenute nei dati.

Teorema 9.1 (Statistiche sufficienti). *Se $S_n = s_n(\underline{\mathbf{X}})$ è una statistica sufficiente per θ , lo stimatore ML $\hat{\theta}$, se esiste, è sempre funzione di S_n .*

Dimostrazione. Dalla (9.17) risulta che $L(\theta; \underline{\mathbf{x}})$ ha il massimo $\hat{\theta}$ nel punto in cui $g(s_n(\underline{\mathbf{x}}), \theta)$ ha il suo massimo e questo punto evidentemente può dipendere da $\underline{\mathbf{x}}$ solo attraverso s_n . Il teorema si estende facilmente ad un insieme di statistiche congiuntamente sufficienti. \square

Teorema 9.2 (Riparametrizzazioni). *Se $\eta \in H$ è un parametro funzione di un altro parametro θ attraverso una funzione biunivoca $g: \Theta \rightarrow H$*

$$\eta = g(\theta) , \tag{9.19}$$

la stima ML di η è data da:

$$\hat{\eta} = g(\hat{\theta}) , \tag{9.20}$$

dove $\hat{\theta}$ è la stima ML di θ .

Dimostrazione. Dimostriamo innanzi tutto che η e θ hanno la stessa funzione di verosimiglianza. Infatti, dato che g è invertibile abbiamo:

$$L_\theta(\theta; \underline{\mathbf{x}}) = L_\theta(g^{-1}(\eta); \underline{\mathbf{x}}) \equiv L_\eta(\eta; \underline{\mathbf{x}}) , \tag{9.21}$$

dove si è evidenziato il fatto che l'uguaglianza si realizza attraverso due forme funzionali diverse, L_θ ed L_η . Ad esempio, se $\eta = \ln \theta$ e θ è la media di una gaussiana, si ha:

$$L_\eta \propto \exp \left[-\frac{1}{2} \frac{(x - e^\eta)^2}{\sigma^2} \right].$$

Notiamo pure che questa trasformazione non dà luogo alle complicazioni viste nel caso delle variabili aleatorie, dove occorre anche cambiare gli elementi differenziali attraverso i determinanti jacobiani, perché ciò che si trasforma sono i *parametri*, non le *variabili*.

Poiché, per definizione:

$$L_\theta(\hat{\theta}; \underline{\mathbf{x}}) \geq L_\theta(\theta; \underline{\mathbf{x}}) \quad \text{per ogni } \theta,$$

e Θ è in corrispondenza biunivoca con H tramite g , applicando la (9.21) ad entrambi i membri della disuguaglianza, si ha:

$$L_\eta(g(\hat{\theta}); \underline{\mathbf{x}}) \geq L_\eta(\eta; \underline{\mathbf{x}}) \quad \text{per ogni } \eta,$$

e quindi

$$\hat{\eta} = g(\hat{\theta}). \quad \square$$

Questo teorema è molto utile: ad esempio, tutte le quantità funzioni della media campionaria possono essere considerate come risultato di stime ML.

Supporremo ora che valgano le seguenti proprietà, dette di regolarità:

- θ è un punto interno di Θ ;
- la funzione $p(x; \theta)$ è una densità di probabilità per ogni $\theta \in \Theta$;
- l'insieme $\{x : p(x; \theta) > 0\}$, (il *supporto* della densità) non dipende da θ ;
- se $\theta_1 \neq \theta_2$, allora esiste almeno un insieme $B \in \mathcal{F}$ per il quale $\int_B p(x; \theta_1) dx \neq \int_B p(x; \theta_2) dx$;
- $p(x; \theta)$ è derivabile per ogni $\theta \in \Theta$;
- le operazioni di integrazione o somma in x e di derivazione rispetto a θ possono essere scambiate due volte.

Trattando in termini di verosimiglianza, per mantenere le idee chiare occorre sempre chiedersi se il ragionamento è svolto considerando $L(\theta; \underline{\mathbf{X}})$ come *funzione aleatoria* di $\underline{\mathbf{X}}$ essendo θ fisso o come *funzione* del parametro θ per una osservazione particolare $\underline{\mathbf{x}}$.

Se utilizzerete questo accorgimento, facendo attenzione alla notazione maiuscola o minuscola, la lettura di ciò che segue vi risulterà più facile. Ad esempio, tutte le medie del tipo

$$\left\langle \left(\frac{\partial}{\partial \theta} \ln L \right)^2 \right\rangle,$$

si riferiscono alle funzioni delle variabili aleatorie $\ln L(\boldsymbol{\theta}; \mathbf{X})$ con un certo $\boldsymbol{\theta}$ fissato, mentre il metodo ML si applica *all'insieme di valori osservati* \underline{x} e considera la verosimiglianza funzione di $\boldsymbol{\theta}$.

Il teorema più importante sulla varianza di uno stimatore è quello formulato dagli statistici Cramér e Rao negli anni 1944-45, che ne stabilisce un limite inferiore. Il teorema non è difficile, se si tengono presenti alcune relazioni fondamentali.

La prima relazione utile si basa sul fatto che, se valgono le condizioni di regolarità, la densità di probabilità è normalizzata indipendentemente da θ :

$$\int p(\mathbf{x}; \theta) \, d\mathbf{x} = 1 ,$$

e quindi:

$$\int \frac{\partial p(\mathbf{x}; \theta)}{\partial \theta} \, d\mathbf{x} = \frac{\partial}{\partial \theta} \int p(\mathbf{x}; \theta) \, d\mathbf{x} = 0 . \quad (9.22)$$

Da questa relazione si può poi dimostrare che:

$$\left\langle \frac{\partial}{\partial \theta} \ln p(\mathbf{X}; \theta) \right\rangle = 0 . \quad (9.23)$$

Infatti:

$$\begin{aligned} \int \frac{\partial p(\mathbf{x}; \theta)}{\partial \theta} \, d\mathbf{x} &= \int \frac{1}{p(\mathbf{x}; \theta)} \frac{\partial p(\mathbf{x}; \theta)}{\partial \theta} p(\mathbf{x}; \theta) \, d\mathbf{x} \\ &= \int \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} p(\mathbf{x}; \theta) \, d\mathbf{x} \equiv \left\langle \frac{\partial}{\partial \theta} \ln p(\mathbf{X}; \theta) \right\rangle = 0 . \end{aligned}$$

Ovviamente vale anche la relazione:

$$\begin{aligned} \frac{\partial}{\partial \theta} \int \frac{\partial p(\mathbf{x}; \theta)}{\partial \theta} \, d\mathbf{x} &= \int \frac{\partial^2 p(\mathbf{x}; \theta)}{\partial \theta^2} \, d\mathbf{x} \\ &= \int \frac{1}{p(\mathbf{x}; \theta)} \frac{\partial^2 p(\mathbf{x}; \theta)}{\partial \theta^2} p(\mathbf{x}; \theta) \, d\mathbf{x} = \left\langle \frac{1}{p} \frac{\partial^2 p}{\partial \theta^2} \right\rangle = 0 . \end{aligned} \quad (9.24)$$

La (9.23) mostra che la derivata rispetto a θ del logaritmo di una densità, detta *score function*, è sempre una funzione di \mathbf{X} a media nulla. La varianza della “score function”, che viene chiamata *informazione attesa di Fisher* per $\boldsymbol{\theta}$ contenuta nella distribuzione di \mathbf{X} , e viene generalmente indicata con $I(\theta)$, in base alla (2.66) di pagina 61, è data da:

$$\begin{aligned} \text{Var} \left[\frac{\partial}{\partial \theta} \ln p(\mathbf{X}; \theta) \right] &= \left\langle \left(\frac{\partial}{\partial \theta} \ln p(\mathbf{X}; \theta) - \left\langle \frac{\partial}{\partial \theta} \ln p(\mathbf{X}; \theta) \right\rangle \right)^2 \right\rangle \\ &= \left\langle \left(\frac{\partial}{\partial \theta} \ln p(\mathbf{X}; \theta) \right)^2 \right\rangle \equiv I(\theta) . \end{aligned} \quad (9.25)$$

Si noti che vale la relazione notevole:

$$I(\theta) = \left\langle \left(\frac{\partial}{\partial \theta} \ln p(\mathbf{X}; \theta) \right)^2 \right\rangle = - \left\langle \frac{\partial^2}{\partial \theta^2} \ln p(\mathbf{X}; \theta) \right\rangle. \quad (9.26)$$

Infatti:

$$\left\langle \frac{\partial^2}{\partial \theta^2} \ln p \right\rangle = \left\langle \frac{\partial}{\partial \theta} \frac{\partial \ln p}{\partial \theta} \right\rangle = \left\langle -\frac{1}{p^2} \left(\frac{\partial p}{\partial \theta} \right)^2 + \frac{1}{p} \frac{\partial^2 p}{\partial \theta^2} \right\rangle = - \left\langle \left(\frac{\partial \ln p}{\partial \theta} \right)^2 \right\rangle$$

dove si è usata la (9.24).

Va ben tenuto presente che le fondamentali relazioni (9.22 - 9.26), *sono valide per qualunque densità* che soddisfa le condizioni di regolarità indicate a pag. 358. Vi consigliamo di fare qualche esercizio e verificarle per alcune delle densità note, come la binomiale e la gaussiana.

Fatto pure importante, queste relazioni valgono anche per la funzione di verosimiglianza, che, come sappiamo, a parte un fattore costante, è la densità di probabilità di un campione di dimensione n di una o più variabili aleatorie. Per esempio, l'informazione di Fisher su θ contenuta nella verosimiglianza è legata a quella contenuta in $p(\mathbf{x}_i; \theta)$ dalla cruciale relazione, analoga alla (9.25):

$$\left\langle \left(\frac{\partial}{\partial \theta} \ln L \right)^2 \right\rangle = \left\langle \left(\frac{\partial}{\partial \theta} \sum_i \ln p(\mathbf{X}_i; \theta) \right)^2 \right\rangle = n \left\langle \left(\frac{\partial}{\partial \theta} \ln p \right)^2 \right\rangle = nI(\theta), \quad (9.27)$$

dove si sono utilizzate le (5.80, 9.5, 9.25) e si è tenuto conto che il campione consiste di n realizzazioni indipendenti della variabile \mathbf{X} .

Un'ultima relazione utile riguarda un qualunque stimatore (corretto o distorto) T_n di θ :

$$\left\langle T_n \frac{\partial \ln L(\theta; \underline{\mathbf{X}})}{\partial \theta} \right\rangle = - \left\langle T_n \frac{\partial \mathcal{L}(\theta; \underline{\mathbf{X}})}{\partial \theta} \right\rangle = \frac{\partial \tau(\theta)}{\partial \theta}, \quad (9.28)$$

dove $\tau(\theta) = \langle T_n \rangle$ e $\tau(\theta)$ si suppone derivabile rispetto a θ . Osservate che $\tau(\theta) = \theta$ se lo stimatore è corretto e che l'ultimo membro della (9.28) in questo caso vale 1. Dall'equivalenza tra funzione di densità del campione e funzione di verosimiglianza e dalle condizioni di regolarità, si ha:

$$\begin{aligned} \frac{\partial \tau(\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta} \langle T_n \rangle \\ &= \int T_n \frac{\partial L(\theta; \underline{\mathbf{x}})}{\partial \theta} d\underline{\mathbf{x}} \\ &= \int T_n \frac{1}{L(\theta; \underline{\mathbf{x}})} \frac{\partial L(\theta; \underline{\mathbf{x}})}{\partial \theta} L(\theta; \underline{\mathbf{x}}) d\underline{\mathbf{x}} = \int T_n \frac{\partial \ln L(\theta; \underline{\mathbf{x}})}{\partial \theta} L(\theta; \underline{\mathbf{x}}) d\underline{\mathbf{x}} \\ &= \left\langle T_n \frac{\partial}{\partial \theta} \ln L(\theta; \underline{\mathbf{X}}) \right\rangle. \end{aligned}$$

Siamo ora in grado di dimostrare il:

Teorema 9.3 (Limite di Cramér-Rao). *Sia $p(\mathbf{x}; \theta)$ la densità di probabilità di \mathbf{X} e sia T_n uno stimatore di θ di varianza finita basato sul campione $\underline{\mathbf{X}}$. Se $\tau(\theta) = \langle T_n \rangle$ è derivabile e l'informazione $I(\theta)$ (9.25) della densità p resta finita per ogni θ , la varianza dello stimatore non può mai essere minore di $\tau'(\theta)^2 / \{nI(\theta)\}$:*

$$\text{Var}[T_n] \geq \frac{\tau'(\theta)^2}{n \left\langle \left(\frac{\partial}{\partial \theta} \ln p(\mathbf{X}; \theta) \right)^2 \right\rangle} = \frac{\tau'(\theta)^2}{nI(\theta)}. \quad (9.29)$$

Dimostrazione. Dalle (9.23, 9.28) risulta:

$$\left\langle T_n \frac{\partial}{\partial \theta} \ln L \right\rangle = \left\langle (T_n - \theta) \frac{\partial}{\partial \theta} \ln L \right\rangle = \tau'(\theta). \quad (9.30)$$

Elevando al quadrato questa espressione ed applicando la disuguaglianza (4.29) di Cauchy-Schwarz, si ottiene:

$$\langle (T_n - \theta)^2 \rangle \left\langle \left(\frac{\partial}{\partial \theta} \ln L \right)^2 \right\rangle \geq \left\langle (T_n - \theta) \left(\frac{\partial}{\partial \theta} \ln L \right) \right\rangle^2 = \tau'(\theta)^2. \quad (9.31)$$

Dalla (9.27) si ottiene allora:

$$\langle (T_n - \theta)^2 \rangle \equiv \text{Var}[T_n] \geq \frac{\tau'(\theta)^2}{nI(\theta)}. \quad \square$$

Se T_n è uno stimatore corretto di θ , il limite inferiore di Cramér-Rao diventa $1/\{nI(\theta)\}$. Il teorema di Cramér-Rao, stabilendo un limite inferiore alla varianza di uno stimatore, permette di valutarne l'efficienza in modo quantitativamente preciso. Infatti lo stimatore corretto ideale, in base alla definizione 9.5, sarà quello a varianza minima, ovvero lo stimatore per il quale vale la relazione:

$$\text{Var}[T_n] = \left\langle \left(\frac{\partial \ln L}{\partial \theta} \right)^2 \right\rangle^{-1} = - \left\langle \frac{\partial^2 \ln L}{\partial \theta^2} \right\rangle^{-1} = \frac{1}{nI(\theta)}. \quad (9.32)$$

Questo stimatore, tra quelli corretti, viene considerato come il più efficiente o come il *miglior stimatore*. È anche ovvio definire l'efficienza di un generico stimatore corretto come:

$$\varepsilon(T) = \left[\text{Var}[T_n] nI(\theta) \right]^{-1}. \quad (9.33)$$

Per il miglior stimatore corretto vale la condizione $\varepsilon(T) = 1$.

Uno stimatore che non sia il più efficiente viene anche detto *non ammissibile*.

Se l'informazione è grande, allora la varianza dello stimatore più efficiente, ovvero *l'intervallo di confidenza ottimale* di θ , risulta piccolo. Questo spiega la denominazione di *informazione* data ad I .

Nel caso di un parametro p -dimensionale θ , stimato da $T_n(\underline{X})$ in modo corretto, la generalizzazione della (9.32) fornisce la matrice di covarianza di T_n , il cui elemento (i, j) è dato da:

$$(NI_{ij})^{-1} = \text{Cov}[T_i, T_j] = \langle (T_i - \theta_i)(T_j - \theta_j) \rangle = - \left\langle \frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right\rangle^{-1}. \quad (9.34)$$

dove T_i , $i = 1, \dots, p$, è la i -esima componente di T_n .

Esercizio 9.5

Trovare l'informazione sulla probabilità contenuta nella distribuzione binomiale e quella sulla media nel caso delle distribuzioni di Poisson e di Gauss. Che considerazioni vi suggeriscono i risultati ottenuti?

Risposta. *Indicando con $b(x; p)$, $p(x; \mu)$ e $g(x; \mu, \sigma)$ rispettivamente le densità binomiale, poissoniana e gaussiana, dalle (2.30, 3.14, 3.28) si ottiene facilmente:*

$$\begin{aligned} \ln b(X; p) &= \ln n! - \ln(n - X)! - \ln X! + X \ln p + (n - X) \ln(1 - p), \\ \ln p(X; \mu) &= X \ln \mu - \ln X! - \mu, \\ \ln g(X; \mu, \sigma) &= \ln \left(\frac{1}{\sqrt{2\pi}\sigma} \right) - \frac{1}{2} \left(\frac{X - \mu}{\sigma} \right)^2. \end{aligned}$$

Notate che queste funzioni sono variabili aleatorie, perché funzioni di X (indicato con lettera maiuscola). Derivando rispetto ai parametri di interesse si ottiene:

$$\begin{aligned} \frac{\partial}{\partial p} \ln b(X; p) &= \frac{X}{p} - \frac{n - X}{1 - p} = \frac{X - np}{p(1 - p)}, \\ \frac{\partial}{\partial \mu} \ln p(X; \mu) &= \frac{X}{\mu} - 1 = \frac{X - \mu}{\mu}, \\ \frac{\partial}{\partial \mu} \ln g(X; \mu, \sigma) &= -\frac{X - \mu}{\sigma} \left(-\frac{1}{\sigma} \right) = \frac{X - \mu}{\sigma^2}. \end{aligned} \quad (9.35)$$

Poiché al numeratore di tutte queste derivate compare la differenza $X - \mu$, esse hanno media nulla, in accordo con la (9.23).

L'informazione può ora essere calcolata applicando la (9.26) alle (9.35), utilizzando il quadrato della derivata prima o la derivata seconda, come è più conveniente:

$$\begin{aligned} I(p) &= \frac{1}{p^2(1 - p)^2} \langle (X - np)^2 \rangle = \frac{np(1 - p)}{p^2(1 - p)^2} = \frac{n}{p(1 - p)}, \\ I(\mu) &= \frac{1}{\mu^2} \langle (X - \mu)^2 \rangle = \frac{\sigma^2}{\mu^2} = \frac{1}{\mu} = \frac{1}{\sigma^2}, \\ I(\mu) &= \frac{1}{\sigma^4} \langle (X - \mu)^2 \rangle = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}, \end{aligned} \quad (9.36)$$

dove si è usata la solita relazione $\sigma^2 = \langle (X - \mu)^2 \rangle$ e l'espressione esplicita della varianza per le tre distribuzioni considerate (si veda ad esempio la Tab. 3.1 di pagina 108).

Quali considerazioni suggeriscono ora questi risultati?

Per prima cosa, vediamo che in genere l'informazione è proporzionale all'inverso della varianza: una densità "stretta", con valori poco dispersi intorno alla media, avrà informazione elevata, come è intuitivo. Inoltre, se introduciamo lo stimatore "frequenza"

$$T_1 \equiv F = \frac{X}{n} ,$$

che stima la probabilità sulla base del numero di successi X e lo stimatore "media campionaria"

$$T_n \equiv M = \sum_i \frac{X_i}{n} ,$$

che stima la media μ sulla base di n valori X , vediamo che per questi stimatori il limite di Cramér-Rao coincide con l'errore statistico ricavato per questi stimatori nel cap. 6:

$$\begin{aligned} \text{Var}[F] &= \frac{1}{I(p)} = \frac{p(1-p)}{n} , \\ \text{Var}[M] &= \frac{1}{nI(\mu)} = \frac{\sigma^2}{n} . \end{aligned}$$

Questo errore viene valutato, per grandi campioni, con l'approssimazione $p \simeq f$ e $\sigma^2 \simeq s^2$, la quale dà luogo agli intervalli di stima (6.23, 6.46), ormai a voi ben noti.

Ne deduciamo che la frequenza è lo stimatore migliore della probabilità che compare come parametro nella densità binomiale e che la media campionaria è lo stimatore migliore della media delle distribuzioni poissoniana e gaussiana.

Veniamo ora al teorema cardine, quello che assegna al metodo ML il ruolo fondamentale nella stima dei parametri.

Teorema 9.4 (Stimatore più efficiente). *Se T_n è uno stimatore non distorto di $\tau(\theta)$ a varianza minima (il miglior stimatore), esso coincide con lo stimatore di massima verosimiglianza, se questo esiste:*

$$T_n = \tau(\hat{\theta}) .$$

Dimostrazione. Per il miglior stimatore vale il limite di Cramér-Rao, che è dato dal limite inferiore della (9.29):

$$\langle [T_n - \tau(\theta)]^2 \rangle \frac{nI(\theta)}{[\tau'(\theta)]^2} = 1 . \quad (9.37)$$

Questa relazione è soddisfatta se e solo se

$$\frac{\partial \ln L}{\partial \theta} = \frac{nI(\theta)}{\tau'(\theta)} [T_n - \tau(\theta)] . \quad (9.38)$$

Infatti, elevando al quadrato la (9.38), passando al valore atteso e ricorrendo alla (9.27) si ottiene facilmente la (9.37).

Viceversa, se vale la (9.37), vale anche la disuguaglianza di Cauchy-Schwarz (9.31) come uguaglianza stretta, per cui $\partial \ln L / \partial \theta = c [T_n - \tau(\theta)]$ per qualche costante c . Tenendo conto anche della (9.30) possiamo scrivere:

$$\frac{nI(\theta)}{\tau'(\theta)} = \frac{1}{\tau'(\theta)} \left\langle \left(\frac{\partial \ln L}{\partial \theta} \right)^2 \right\rangle = c \left\langle [T_n - \tau(\theta)] \frac{1}{\tau'(\theta)} \frac{\partial \ln L}{\partial \theta} \right\rangle = c ,$$

da cui $c = nI(\theta) / \tau'(\theta)$ e la (9.38).

Se ora fissiamo le \mathbf{x} e consideriamo θ variabile, la stima ML si ottiene annullando la (9.38):

$$\frac{\partial \ln L}{\partial \theta} = \frac{nI(\hat{\theta})}{[\tau'(\hat{\theta})]^2} [T_n - \tau(\hat{\theta})] = 0 , \quad \text{da cui} \quad T_n = \tau(\hat{\theta}) . \quad \square$$

Passiamo ora alle proprietà asintotiche degli stimatori ML, la prima delle quali è la consistenza. Mostriamo che questa sussiste con un'argomentazione euristica; una dimostrazione rigorosa può essere trovata in [Cra51] o anche in [Azz01, EDJ+71].

Le (9.23, 9.26) mostrano che, per la legge dei grandi numeri (si veda a pagina 214),

$$\frac{1}{n} \frac{\partial \ln L}{\partial \theta} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \ln p(\mathbf{x}_i; \theta)}{\partial \theta}$$

converge in probabilità (e anche quasi certamente) a

$$\left\langle \frac{\partial \ln p(\mathbf{X}; \theta)}{\partial \theta} \right\rangle = 0$$

e che

$$\frac{1}{n} \frac{\partial^2 \ln L}{\partial \theta^2} = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ln p(\mathbf{x}_i; \theta)}{\partial \theta^2}$$

converge in probabilità (e anche quasi certamente) a

$$\left\langle \frac{\partial^2 \ln p(\mathbf{X}; \theta)}{\partial \theta^2} \right\rangle = -I(\theta) < 0$$

se θ è il vero valore del parametro. Questo indica che, al crescere di n , la derivata prima della funzione di verosimiglianza nel punto θ tende in probabilità a zero e che la derivata seconda è negativa in probabilità. Siamo

quindi indotti a pensare che la distanza tra θ e il punto di massimo della log-verosimiglianza $\hat{\theta}$ tenda in probabilità a zero, il che corrisponde appunto alla consistenza dello stimatore ML.

Consideriamo ora una serie di esperimenti, in ognuno dei quali si ottiene un valore \mathbf{x} e si massimizza la funzione di verosimiglianza rispetto a θ . L'insieme delle stime $\hat{\theta}_i$ così ottenute forma un campione della variabile aleatoria $\hat{\theta}$. Se eseguiamo una serie infinita di esperimenti, otterremo la distribuzione di $\hat{\theta}$. Avrò una densità gaussiana? Il teorema di normalità asintotica, assicura che, se la dimensione n del campione è sufficientemente grande, la risposta è affermativa. Della dimostrazione diamo solo un accenno, evitando in particolare di trattare con precisione i termini trascurabili (secondo la convergenza in probabilità) negli sviluppi in serie di Taylor.

Sviluppiamo la derivata del logaritmo di L nell'intorno del valore vero θ , considerando per semplicità una sola variabile aleatoria X :

$$\left. \frac{\partial \ln L}{\partial \theta} \right|_{\hat{\theta}} = \left. \frac{\partial \ln L}{\partial \theta} \right|_{\theta} + \left. \frac{\partial^2 \ln L}{\partial \theta^2} \right|_{\theta} (\hat{\theta} - \theta) + \frac{1}{2} \left. \frac{\partial^3 \ln L}{\partial \theta^3} \right|_{\theta} (\hat{\theta} - \theta)^2 + \dots \quad (9.39)$$

Se n è la dimensione del campione, possiamo scrivere:

$$\begin{aligned} \frac{1}{n} \left. \frac{\partial \ln L}{\partial \theta} \right|_{\hat{\theta}} &= \frac{1}{n} \sum_{i=1}^n \left. \frac{\partial \ln p(x_i; \theta)}{\partial \theta} \right|_{\theta} + \frac{1}{n} \sum_{i=1}^n \left. \frac{\partial^2 \ln p(x_i; \theta)}{\partial \theta^2} \right|_{\theta} (\hat{\theta} - \theta) \\ &\quad + \frac{1}{2n} \sum_{i=1}^n \left. \frac{\partial^3 \ln p(x_i; \theta)}{\partial \theta^3} \right|_{\theta} (\hat{\theta} - \theta)^2 + \dots \end{aligned} \quad (9.40)$$

Poiché vale l'ipotesi di consistenza dello stimatore ML, per n sufficientemente grande $(\hat{\theta} - \theta)$ diventerà piccolo e nella (9.40), se i valori medi delle derivate restano limitati (condizione di regolarità), si possono trascurare i termini di ordine superiore al primo. Inoltre, considerando che per $\theta = \hat{\theta}$ la derivata prima di $\ln L$ si annulla, la (9.40) diventa:

$$\frac{1}{n} \sum_{i=1}^n \left. \frac{\partial \ln p(x_i; \theta)}{\partial \theta} \right|_{\theta} \simeq -\frac{1}{n} \sum_{i=1}^n \left. \frac{\partial^2 \ln p(x_i; \theta)}{\partial \theta^2} \right|_{\theta} (\hat{\theta} - \theta) .$$

La precedente relazione può essere riscritta come segue:

$$\frac{\sum_{i=1}^n \left. \frac{\partial \ln p(x_i; \theta)}{\partial \theta} \right|_{\theta}}{\sqrt{nI(\theta)}} \bigg/ \frac{-\frac{1}{n} \sum_{i=1}^n \left. \frac{\partial^2 \ln p(x_i; \theta)}{\partial \theta^2} \right|_{\theta}}{I(\theta)} \simeq \sqrt{nI(\theta)} (\hat{\theta} - \theta) .$$

Per il teorema Limite Centrale, il numeratore al primo membro converge in distribuzione ad una gaussiana standard, mentre, per la legge dei grandi numeri e la (9.26), il denominatore converge quasi certamente a 1. Si può poi mostrare che da ciò discende la convergenza in distribuzione ad una gaussiana per l'intero rapporto al primo membro, e quindi ciò vale anche per il secondo

membro, che consideriamo identico a meno di termini trascurabili. Scriveremo dunque che, per n grande, si ha approssimativamente

$$(\hat{\theta} - \theta) \sim N\left(0, \frac{1}{nI(\theta)}\right). \quad (9.41)$$

Possiamo enunciare finalmente l'importante:

Teorema 9.5 (Normalità asintotica). *Se valgono le condizioni di regolarità enunciate a pag. 358, gli stimatori ML sono asintoticamente normali con valore atteso pari al vero valore del parametro e hanno efficienza asintotica pari a 1.*

Dimostrazione. La (9.41) mostra che lo stimatore $\hat{\theta}$, per n grande, si distribuisce in modo normale, con media θ e varianza data dal limite di Cramér-Rao (9.29). \square

In pratica, i valori di n per cui vale l'approssimazione gaussiana della distribuzione di $\hat{\theta}$ dipendono dal tipo di densità campionaria $p(x; \theta)$ e dal tipo di stimatore. Nel caso della media campionaria la normalità viene raggiunta, come sappiamo, molto rapidamente ($n > 10$). Per altri stimatori, come la varianza campionaria, la convergenza verso la normale è molto più lenta.

Gli argomenti appena visti mostrano che il metodo ML fornisce stimatori consistenti, asintoticamente corretti (con un fattore di distorsione $O(1/n)$, come nell'esercizio 9.3), asintoticamente normali con varianza data dal limite di Cramér-Rao. Stimatori di questo tipo sono detti BAN (*Best Asimptotically Normal*).

Queste proprietà fanno della massima verosimiglianza il metodo più usato in statistica per la stima puntuale dei parametri, in tutti quei casi in cui è nota a priori la densità $p(x; \theta)$.

9.5 Intervalli di confidenza

La stima puntuale dei parametri attraverso la massimizzazione della verosimiglianza fornisce anche gli elementi per effettuare la stima per intervalli. Infatti, come abbiamo visto:

- a) dal teorema 9.5, $(\hat{\theta} - \theta)$ è asintoticamente normale con media nulla e varianza $1/[nI(\theta)]$;
- b) nel dimostrare il teorema 9.5, si è visto che $(\partial \ln L / \partial \theta)$ ha media nulla, varianza $nI(\theta)$ e distribuzione approssimativamente normale per n grande.

Questi due metodi danno in pratica gli stessi risultati. Se si utilizza l'informazione stimata $I(\hat{\theta})$ invece di quella attesa (che corrisponde a porre, in statistica elementare, $s \simeq \sigma$), si trovano in genere gli intervalli di confidenza già visti nel cap. 6. La distorsione dell'intervallo introdotta da questa approssimazione è

studiata in dettaglio in [EDJ⁺71] ed è dell'ordine di $1/n$. Potete approfondire un po' questi aspetti risolvendo il problema 9.8.

Vediamo ora in dettaglio un terzo metodo per la determinazione degli intervalli di confidenza, che è di fondamentale importanza nei casi multidimensionali. Vi anticipiamo la conclusione, che è:

- c) la variabile $2[\ln L(\hat{\theta}) - \ln L(\theta)]$ si distribuisce asintoticamente come $\chi^2(p)$, dove p è la dimensione di θ .

Per semplicità, consideriamo un'approssimazione asintotica di $2[\ln L(\hat{\theta}) - \ln L(\theta)]$ nel caso monodimensionale, sviluppando il logaritmo negativo di $L(\theta)$ attorno a $\hat{\theta}$, quando θ è il vero valore del parametro, fino al secondo ordine. Evitando ancora di trattare dettagliatamente i termini trascurabili secondo la convergenza in probabilità, si ha:

$$\begin{aligned} \mathcal{L}(\theta) &\simeq \mathcal{L}(\hat{\theta}) + \mathcal{L}'(\hat{\theta})(\theta - \hat{\theta}) + \frac{1}{2} \mathcal{L}''(\hat{\theta})(\theta - \hat{\theta})^2 \\ &\simeq \mathcal{L}(\hat{\theta}) + \frac{n}{2} \frac{\mathcal{L}''(\theta)}{n} (\theta - \hat{\theta})^2 \simeq \mathcal{L}(\hat{\theta}) - \frac{nI(\theta)}{2} (\theta - \hat{\theta})^2 . \end{aligned}$$

Il termine d'errore nella prima riga, che è $o(\theta - \hat{\theta})^2$, viene trascurato grazie alla consistenza di $\hat{\theta}$, la quale giustifica anche la sostituzione di $\hat{\theta}$ con θ nell'argomento di \mathcal{L}'' ; per l'ultimo passaggio si è invece utilizzata la legge dei grandi numeri. Possiamo quindi scrivere

$$2[\ln L(\hat{\theta}) - \ln L(\theta)] \simeq nI(\theta) (\hat{\theta} - \theta)^2 . \tag{9.42}$$

Operiamo ora una riparametrizzazione tramite una funzione invertibile $\eta(\theta)$ tale che si abbia $nI_\eta(\eta) = 1$; la funzione $\eta(\cdot)$ che possiede questo requisito è una qualunque primitiva di $\sqrt{nI(\theta)}$. La (9.42), riformulata in termini di η diventa:

$$\ln L_\eta(\hat{\eta}) - \ln L_\eta(\eta) \simeq \frac{1}{2} (\hat{\eta} - \eta)^2 . \tag{9.43}$$

In virtù del teorema 9.5 e della riparametrizzazione effettuata, sappiamo che, approssimativamente, $\hat{\eta} \sim N(\eta, 1)$, perciò, gli intervalli di confidenza per η di semiampiezza pari a una o due deviazioni standard sono rispettivamente $\hat{\eta} \pm 1$ ed $\hat{\eta} \pm 2$, con un livello di confidenza (approssimato) del 68.3% e del 95.4%. Grazie alla (9.43), gli estremi di intervalli con lo stesso livello di confidenza si trovano risolvendo rispetto a η le equazioni

$$\begin{aligned} \ln L_\eta(\hat{\eta}) - \ln L_\eta(\eta) &= 0.5 \\ \ln L_\eta(\hat{\eta}) - \ln L_\eta(\eta) &= 2 . \end{aligned}$$

Il medesimo risultato si ricava osservando che $(\hat{\eta} - \eta)^2 \sim \chi^2(1)$, poiché $(\hat{\eta} - \eta)$ ha distribuzione gaussiana standard. Allora, se $1 - \alpha = CL$ è il livello di confidenza, gli estremi dell'intervallo corrispondente saranno dati da quei valori η tali che:

$$2(\ln L_\eta(\hat{\eta}) - \ln L_\eta(\eta)) = \chi_\alpha^2(1) , \tag{9.44}$$

dove $\chi_\alpha^2(1)$ è il quantile di ordine α della distribuzione χ^2 con un grado di libertà. Se $\alpha = 0.683$, $\chi_\alpha^2(1) = 1.00$, il che ci riporta a risolvere rispetto a η l'equazione $[\ln L_\eta(\hat{\eta}) - \ln L_\eta(\eta)] = 0.5$.

Abbiamo dunque trovato una riparametrizzazione che produce intervalli simmetrici attorno a $\hat{\eta}$ con livelli di confidenza propri della distribuzione gaussiana.

Ora però viene il bello del ragionamento: non è necessario effettuare la riparametrizzazione, *basta sapere che esiste*. Infatti, grazie al teorema 9.2, sappiamo che *numericamente* le verosimiglianze L_θ ed L_η sono uguali; basta allora utilizzare la verosimiglianza originale L_θ e trovare i valori θ_1 e θ_2 per i quali

$$2\Delta[\ln L] \equiv 2 \left(\ln L(\hat{\theta}) - \ln L(\theta) \right) = \chi_\alpha^2(1) , \quad (9.45)$$

ed ottenere così l'intervallo di stima di θ

$$\theta_1 \leq \theta \leq \theta_2 , \quad CL = 1 - \alpha . \quad (9.46)$$

Per determinare ad esempio gli intervalli con $CL = 68.3\%$ e $CL = 95.4\%$, si trovano gli estremi che danno $2\Delta[\ln L] = 1$ oppure $2\Delta[\ln L] = 4$. Per una applicazione di questo metodo, si veda il problema 9.8.

È quindi possibile, se la funzione di verosimiglianza è sufficientemente regolare, trovare gli intervalli ed i relativi livelli di confidenza. I livelli di confidenza sono gaussiani ma *non corrispondono ad intervalli gaussiani*, poiché in generale risultano asimmetrici rispetto alla stima puntuale $\hat{\theta}$. Gli intervalli che si avvicinano maggiormente a quelli gaussiani sono quelli del parametro trasformato η della (9.19), della cui esistenza sappiamo e di cui ci siamo serviti per determinare la variazione della verosimiglianza in funzione dei livelli di confidenza. La situazione è schematizzata in Fig. 9.3.

La consistenza di $\hat{\theta}$ assicura comunque che, aumentando n , anche l'intervallo di stima (9.46) tende ad essere simmetrico rispetto a θ .

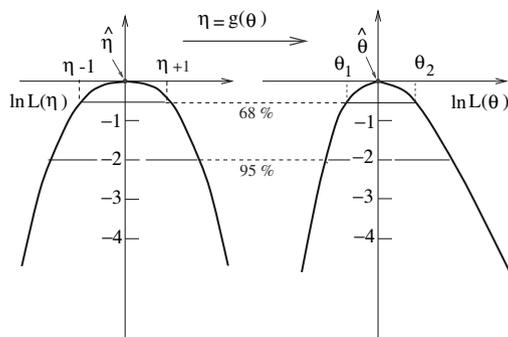


Fig. 9.3. Determinazione degli intervalli di confidenza nella stima ML di un parametro

Nel caso di un parametro p -dimensionale θ , la frontiera dell'insieme di confidenza si trova utilizzando la condizione equivalente alla (9.45):

$$2\Delta[\ln L] \equiv 2 \left(\ln L(\hat{\theta}) - \ln L(\theta) \right) = \chi_{\alpha}^2(p), \quad (9.47)$$

dove χ_{α}^2 è il quantile corrispondente al livello di confidenza CL assegnato e la distribuzione asintotica $2\Delta[\ln L]$ è $\chi^2(p)$. I valori χ_{α}^2 , in funzione dei gradi di libertà, si leggono in Tab. D.4 di pagina 524. Dalla tabella si vede ad esempio che, con due gradi di libertà, le regioni racchiuse dai contorni $2\Delta[\ln L] \simeq 2.4$ ed $2\Delta[\ln L] \simeq 6.0$, corrispondenti rispettivamente a $CL \simeq 68\%$ e $CL \simeq 95\%$, sono equivalenti ai valori di $2\Delta[\ln L] \simeq 1$ e $2\Delta[\ln L] \simeq 4$ per il caso monodimensionale.

La (9.47) richiede la esplorazione numerica delle ipersuperfici χ^2 ed è in genere di difficile applicazione. In pratica, quasi sempre si determinano p intervalli di confidenza monodimensionali, ciascuno di livello α , sempre per via numerica, facendo variare, secondo una griglia di valori, un parametro θ_i alla volta e massimizzando la verosimiglianza rispetto agli altri parametri per ogni valore di θ_i . L'intervallo $(\theta_{i1}, \theta_{i2})$ si ottiene risolvendo un'equazione analoga alla (9.45) rispetto a θ_i :

$$2[\ln L(\hat{\theta}) - \ln L(\theta_i, \hat{\theta}(\theta_i))] = \chi_{\alpha}^2(1) \quad (9.48)$$

dove $L(\theta_i, \hat{\theta}(\theta_i))$ è la verosimiglianza massimizzata rispetto a tutte le altre componenti di θ avendo fissato θ_i e $L(\hat{\theta})$ è il valore di *best fit* ottenuto massimizzando con tutti i parametri liberi. La procedura è giustificata considerando la seguente eguaglianza:

$$\ln L(\hat{\theta}) - \ln L(\theta) = [\ln L(\hat{\theta}) - \ln L(\theta_i, \hat{\theta}(\theta_i))] + [\ln L(\theta_i, \hat{\theta}(\theta_i)) - \ln L(\theta)].$$

Il primo membro ha distribuzione asintotica $\chi^2(p)$, mentre il secondo addendo del secondo membro ha distribuzione asintotica $\chi^2(p-1)$, essendo di fatto $2\Delta[\ln L]$ quando θ_i è noto. Questo suggerisce, per analogia col teorema 3.4 di additività del χ^2 discusso nell'Appendice B, che il primo termine al secondo membro abbia distribuzione asintotica $\chi^2(1)$.

Gli errori trovati con la (9.47) e la (9.48) hanno un significato diverso, come mostrato a suo tempo nella Fig. 4.8 di pagina 146, la cui corrispondente interpretazione statistica, nel caso di due parametri, è mostrata in Fig. 9.4. Il contorno della regione più scura corrisponde alla (9.47) per una variazione di una unità del χ^2 : in base alla Tab. D.4, con due gradi di libertà, questo contorno ha un livello di confidenza di quasi il 40% (il valore preciso è 39.3%, come si vede dalla (4.81) di pagina 143). La (9.48) corrisponde invece al livello di confidenza gaussiano in una dimensione per $\hat{\theta}_2$, dato dalla regione chiara della figura.

Gli errori forniti dai programmi di minimizzazione, se non vengono richiesti i contorni delle regioni χ^2 , sono calcolati con la (9.48) con $\chi_{\alpha}^2(p) = 1$ e si riferiscono a un $CL = 0.68$ per ogni singolo parametro, indipendentemente dagli altri.

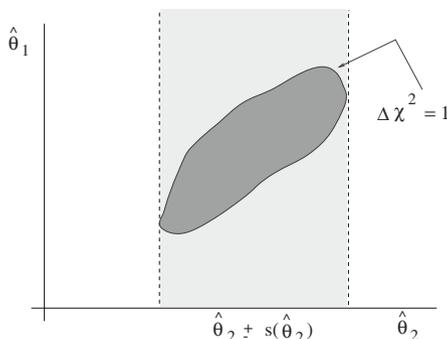


Fig. 9.4. Forme assunte dalle regioni di confidenza risultanti dalla (9.47) (regione più scura) e dalla (9.45) (fascia più chiara). La regione scura va associata all'intervallo aleatorio che contiene il valore vero della coppia $(\hat{\theta}_1, \hat{\theta}_2)$ con una probabilità del 39.3%; la fascia chiara, ottenuta proiettando la regione scura sull'asse delle ascisse, è la realizzazione dell'intervallo aleatorio che ha una probabilità del 68.3% di contenere il valore vero di $\hat{\theta}_2$ (o di $\hat{\theta}_1$, se si proietta sull'altro asse)

9.6 Il metodo dei minimi quadrati e la massima verosimiglianza

Consideriamo l'osservazione di n variabili gaussiane indipendenti, provenienti da n distribuzioni gaussiane differenti. La funzione di verosimiglianza è in questo caso:

$$L(\boldsymbol{\theta}; \underline{x}) = \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}\sigma_i(\boldsymbol{\theta})} \exp\left(-\frac{1}{2} \frac{(x_i - \mu_i(\boldsymbol{\theta}))^2}{\sigma_i^2(\boldsymbol{\theta})}\right) \right], \quad (9.49)$$

dove in generale i parametri $\boldsymbol{\mu}$ e $\boldsymbol{\sigma}$ possono dipendere a loro volta da un parametro multidimensionale $\boldsymbol{\theta}$. Si noti che qui è stata generalizzata la (9.4), dove le n misure provenivano dalla stessa popolazione. Nella (9.49), infatti, le n popolazioni sono diverse, pur avendo tutte la medesima densità gaussiana. Poiché anche in questo caso la verosimiglianza rappresenta la densità di probabilità del campione, ad essa si può senz'altro applicare tutta la teoria svolta fino a questo punto.

Il metodo ML permette allora la determinazione dei parametri $\boldsymbol{\sigma}$ e $\boldsymbol{\mu}$ attraverso la massimizzazione della (9.49) o la minimizzazione del suo logaritmo negativo:

$$\mathcal{L} \equiv -\ln L(\boldsymbol{\theta}; \underline{x}) = -\sum_{i=1}^n \ln\left(\frac{1}{\sqrt{2\pi}\sigma_i(\boldsymbol{\theta})}\right) + \frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu_i(\boldsymbol{\theta}))^2}{\sigma_i^2(\boldsymbol{\theta})}. \quad (9.50)$$

Poiché

$$-\sum_{i=1}^n \ln\left(\frac{1}{\sqrt{2\pi}\sigma_i}\right) = \frac{1}{2} \sum_{i=1}^n \ln(2\pi) + \frac{1}{2} \sum_{i=1}^n \ln(\sigma_i^2)$$

e il primo termine è inessenziale perché costante, trascurando anche il fattore moltiplicativo comune $1/2$, la (9.50) equivale alla ricerca del minimo della funzione:

$$\mathcal{L} \equiv -\ln L(\boldsymbol{\theta}; \underline{\mathbf{x}}) = \sum_{i=1}^n \ln(\sigma_i^2(\boldsymbol{\theta})) + \sum_{i=1}^n \frac{(x_i - \mu_i(\boldsymbol{\theta}))^2}{\sigma_i^2(\boldsymbol{\theta})}. \quad (9.51)$$

Se supponiamo che le deviazioni standard σ_i siano note, oppure approssimabili con le s_i stimate in precedenti esperimenti, il primo termine della (9.50) è costante, ed il metodo ML si riduce alla ricerca del minimo della funzione χ^2 :

$$\chi^2 = \sum_{i=1}^n \frac{(x_i - \mu_i(\boldsymbol{\theta}))^2}{s_i^2}, \quad (9.52)$$

attraverso l'annullamento delle derivate:

$$\frac{\partial \chi^2(\mathbf{x}, \boldsymbol{\mu}(\hat{\boldsymbol{\theta}}_j))}{\partial \theta_j} = 0, \quad (j = 1, 2, \dots, k). \quad (9.53)$$

La (9.53) rappresenta il metodo dei minimi quadrati, detto anche metodo LS (dall'inglese *Least Squares*), che sarà discusso in dettaglio nel cap. 10. Qui il principio appare come una conseguenza del metodo ML *quando i dati provengono da popolazioni aventi densità gaussiane di varianza nota e media incognita da determinare*.

La caratteristica importante del metodo LS è che esso, per venire impiegato, richiede soltanto la conoscenza del valore atteso e della varianza delle variabili osservate. Inoltre, dopo il processo di minimizzazione, è comunque possibile calcolare il χ^2 utilizzando i dati di minimo (*best fit*) $\boldsymbol{\mu}(\hat{\boldsymbol{\theta}})$:

$$\hat{\chi}^2 = \sum_{i=1}^n \frac{(x_i - \mu_i(\hat{\boldsymbol{\theta}}))^2}{s_i^2}, \quad (9.54)$$

e procedere, in alcuni casi, al test χ^2 , *in modo simile a quanto già discusso nei parr. 6.17, 6.18*.

Sebbene le quantità $\hat{\theta}_i$ siano delle stime e non i valori veri, sotto certe ipotesi, cui accenneremo nel cap. 10, quando la dimensione n del campione è grande la variabile (9.54) tende effettivamente alla densità χ^2 con $(n - p)$ gradi di libertà, dove p è sempre la dimensione di $\boldsymbol{\theta}$ [SW89].

È possibile in questo modo verificare se le forme funzionali assunte per le medie μ_i , sono compatibili con i dati. È importante notare che, mentre la parte matematica di minimizzazione o massimizzazione può essere sempre eseguita, la verifica delle ipotesi tramite il test del χ^2 è significativa solo se le variabili coinvolte sono gaussiane.

Il metodo LS, ricavato ancora una volta come conseguenza del metodo ML, trova una importante applicazione nello studio degli istogrammi, come vi mostriamo subito nel prossimo paragrafo.

9.7 Adattamento di densità (*best fit*) ad istogrammi

Qui riprendiamo e completiamo gli argomenti che abbiamo già trattato nei parr. 6.16 e 6.17, ovvero la stima della densità della popolazione attraverso l'istogramma e la verifica di compatibilità tra campione e modello di popolazione.

In questo caso la variabile aleatoria definita nella (9.51) è il numero I_i di eventi caduti nel canale i -esimo di ampiezza Δ_i dell'istogramma (che viene osservato come $I_i = n_i$), mentre μ_i è il numero atteso o teorico di eventi, che in base alla (6.109) è dato da:

$$\mu_i(\boldsymbol{\theta}) = N \int_{\Delta_i} p(x; \boldsymbol{\theta}) dx \simeq N p(x_{0i}; \boldsymbol{\theta}) \Delta_i \equiv N p_i(\boldsymbol{\theta}), \quad (9.55)$$

dove la densità ipotizzata $p_i(x_{0i}, \boldsymbol{\theta})$ è calcolata nel punto centrale x_{0i} del canale e N è la dimensione del campione. Qui la variabile x della (9.55), a differenza del paragrafo precedente, rappresenta i valori dello spettro di X , cioè la variabile lungo l'ascissa dell'istogramma.

La funzione di verosimiglianza è proporzionale alla probabilità multinomiale (4.87) di avere, dati k canali, n_i eventi nel canale i -esimo. Omettendo i fattori indipendenti da $\boldsymbol{\theta}$ abbiamo:

$$L(\boldsymbol{\theta}; \underline{n}) = \prod_{i=1}^k [p_i(\boldsymbol{\theta})]^{n_i}, \quad (9.56)$$

$$\mathcal{L} = -\ln L(\boldsymbol{\theta}; \underline{n}) = -\sum_{i=1}^k n_i \ln [p_i(\boldsymbol{\theta})]. \quad (9.57)$$

Per effettuare la stima ML del parametro p -dimensionale $\boldsymbol{\theta}$, possiamo massimizzare la (9.56) oppure minimizzare la (9.57). In genere, poiché si usano codici di minimizzazione, considereremo solo la minimizzazione delle verosimiglianze logaritmiche.

È interessante vedere che la minimizzazione della (9.57) contiene in sé il metodo dei minimi quadrati che è quello usato più comunemente. Deriviamo infatti la (9.57) rispetto alla componente j -esima di $\boldsymbol{\theta}$, θ_j . Cambiando segno, si ottiene

$$\sum_{i=1}^k \frac{n_i}{p_i(\boldsymbol{\theta})} \frac{\partial p_i(\boldsymbol{\theta})}{\partial \theta_j} = \sum_{i=1}^k \frac{n_i - N p_i(\boldsymbol{\theta})}{p_i(\boldsymbol{\theta})} \frac{\partial p_i(\boldsymbol{\theta})}{\partial \theta_j}$$

dove l'uguaglianza deriva dal vincolo $\sum_i p_i(\boldsymbol{\theta}) = 1$, che comporta l'annullamento della somma delle derivate parziali. È facile accorgersi che il secondo membro dell'uguaglianza coincide, a meno di una costante moltiplicativa, con la derivata parziale del termine:

$$\chi^2 = \sum_i \frac{(n_i - N p_i(\boldsymbol{\theta}))^2}{N p_i(\boldsymbol{\theta})}, \quad (9.58)$$

quando il denominatore viene considerato costante.

I parametri di *best fit* $\hat{\theta}$ possono essere dunque trovati minimizzando la (9.58) rispetto al numeratore. Si riconosce facilmente che questa procedura non è altro che una applicazione del metodo LS dei minimi quadrati rappresentato dalla (9.54). Poiché in questa approssimazione il denominatore del χ^2 deve essere considerato costante, nel processo di minimizzazione a volte si utilizza il χ^2 *modificato*:

$$\chi^2 = \sum_i \frac{(n_i - Np_i(\theta))^2}{n_i}, \quad (9.59)$$

dove gli errori statistici sono stati stimati con la approssimazione $\sigma_i^2 = Np_i \simeq n_i$. Abbiamo quindi a disposizione, per la stima di θ , la (9.57) e la (9.59). I due metodi danno risultati equivalenti, se il campione è grande.

Tuttavia, la (9.57) è la più generale, perché vale sempre, anche per campioni piccoli. La (9.59), invece, vale a rigore se tutti i canali hanno più di una decina di eventi. In pratica spesso si è meno restrittivi, e si considera grande un campione che ha più di 5 eventi per canale. Se ciò non si verifica, i canali con pochi eventi si possono raggruppare in un solo canale con quelli adiacenti prima di effettuare la minimizzazione del χ^2 .

Una volta effettuata la minimizzazione, si può procedere al *test del* χ^2 calcolato con la (9.58).

Il numero di gradi di libertà è pari a $(\nu - p)$, dove ν vale k oppure $k - 1$ a seconda che la dimensione del campione N sia rispettivamente variabile o costante.

Con questa procedura si realizza in pieno lo schema di Fig. 9.1: si formula un modello di popolazione di densità $p(\mathbf{x}; \theta)$ e dai dati si determina la densità più verosimile, data dal valore $\hat{\theta}$ del parametro di *best fit* con relativo errore; si esegue poi il test χ^2 per verificare l'ipotesi sulla forma funzionale scelta per la densità.

Vediamo ora come vanno in pratica le cose riprendendo alcuni casi già trattati nell'ambito della statistica di base.

Esercizio 9.6

Esequire il best fit dell'istogramma (6.107) ottenuto artificialmente al calcolatore simulando l'estrazione di 1000 dati da una gaussiana di parametri veri $\mu = 70$, $\sigma = 10$:

x_i	n_i	x_i	n_i
37.5	1	72.5	207
42.5	4	77.5	153
47.5	16	82.5	101
52.5	44	87.5	42
57.5	81	92.5	7
62.5	152	97.5	6
67.5	186		

Risposta. Supponendo che l'istogramma provenga da una gaussiana di parametri incogniti (come succederebbe per dati reali e non simulati), eseguiamo il best fit dei dati con una gaussiana. La (9.55) diventa allora:

$$\begin{aligned} N p_i(\boldsymbol{\theta}) &\equiv N p_i(\mu, \sigma) = N p_i(x_i; \mu, \sigma) \Delta_i \\ &= 1000 \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2}\right] \cdot 5, \end{aligned} \quad (9.60)$$

poiché $N = 1000$ e $\Delta_i = 5$.

Per eseguire il best fit abbiamo usato la nostra routine `Gaussfit`, che utilizza la funzione di minimizzazione `Nlinfit`, di cui troverete, nel nostro sito, la descrizione del codice e dei metodi statistici utilizzati nella stima e nella determinazione degli errori sui parametri.

Il programma esegue la minimizzazione di una funzione data dall'utente, supponendo che essa rappresenti una funzione χ^2 dipendente da p parametri. Se si utilizza la verosimiglianza della (9.57), si minimizza la funzione $-2 \ln L$. Alla fine della minimizzazione, si calcola il χ^2 finale (9.58) e l'utente può eseguire il test χ^2 .

Poiché per la funzione χ^2 (sia in fase di minimizzazione che in fase di test) si richiede un contenuto di eventi per canale superiore a 5, nell'applicazione delle formule (9.59, 9.58) si calcola il χ^2 raggruppando i primi due canali. Ad esempio, la (9.58) diventa ($n_1 = 1, n_2 = 4$):

$$\chi^2 = \frac{(n_1 + n_2 - N p_1(\mu, \sigma) - N p_2(\mu, \sigma))^2}{N p_1(\mu, \sigma) + N p_2(\mu, \sigma)} + \frac{(n_3 - N p_3(\mu, \sigma))^2}{N p_3(\mu, \sigma)} + \dots$$

I risultati sono riportati nella Tab. (9.61), che contiene nell'ordine: la formula usata nella minimizzazione, il valore dei parametri di best fit con errore, il valore di $\hat{\chi}^2$, ottenuto dalla (9.58), utilizzato nel test ed il livello di significatività SL osservato (p -value) (6.123) per il test ad una coda, ottenuto dalla Tab. D.3 con $(\nu - k)$ gradi di libertà. Poiché N è costante e si sono raggruppati i primi due canali, $\nu = (13 - 1 - 1) = 11$; dato che si è minimizzato rispetto a due parametri, $k = 2$ e quindi $(\nu - k) = 9$.

Equazione	$\hat{\mu}$	$\hat{\sigma}$	χ^2	SL
(9.57)	70.09 ± 0.31	9.73 ± 0.22	$8.67 \simeq 48\%$	(9.61)
(9.59)	69.97 ± 0.31	9.59 ± 0.22	$10.1 \simeq 35\%$	

La tabella mostra che i due metodi usati hanno portato a risultati equivalenti, anche se il campione utilizzato non era particolarmente numeroso. La stima con la formula approssimata (9.59) sembra tuttavia sottostimare leggermente il valore di σ .

Si può notare anche un buon accordo dei dati col modello vero utilizzato nella simulazione ($\mu = 70, \sigma = 10$) e una differenza trascurabile tra i valori di best fit e quelli calcolati direttamente dai dati nel par. 6.16.

Riprendiamo ora una nostra vecchia conoscenza, l'esperimento delle 10 monete, che avevamo trattato per l'ultima volta nell'esercizio 6.13 a pagina 239.

Esercizio 9.7

Determinare la probabilità p di ottenere testa con il lancio di una moneta eseguendo il best fit, con una densità binomiale, dell'istogramma dell'esperimento delle 10 monete di Tab. 2.2 (che per comodità riportiamo):

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	0	0	5	13	12	25	24	14	6	1	0

Risposta. In questo caso la popolazione di riferimento ha densità binomiale (2.30) e la (9.55) si scrive:

$$\begin{aligned}
 N p_i(\theta) &\equiv N p_i(p) = N b(x_i; 10, p) \\
 &= N \frac{10!}{x_i!(10 - x_i)!} p^{x_i} (1 - p)^{10 - x_i}, \tag{9.62}
 \end{aligned}$$

dove $N = 100$ è il numero delle prove e manca l'intervallo Δ_i perché qui siamo in presenza di una variabile discreta. La probabilità p è il parametro incognito da determinare attraverso il best fit.

La procedura da utilizzare è la stessa dell'esercizio precedente. La sola differenza è che, nel calcolo del χ^2 , si devono raggruppare i primi e gli ultimi tre canali ($n_1 = 0, n_2 = 1, n_3 = 5$) e ($n_9 = 6, n_{10} = 1, n_{11} = 0$) per avere un numero di eventi > 5 . Il numeratore della funzione χ^2 diventa allora:

$$\begin{aligned}
 &[0 + 0 + 5 - Nb(10, p; 0) - Nb(10, p; 1) - Nb(10, p; 2)]^2 \\
 &+ [13 - Nb(10, p; 3)]^2 + \dots
 \end{aligned}$$

I canali dell'istogramma sono pertanto pari a 7 ed i gradi di libertà della suddivisione sono $\nu = 7 - 1 = 6$ perché il numero totale N di prove è fissato. Un ulteriore grado di libertà viene perso perché la probabilità p viene determinata dai dati ($k = 1$). I gradi di libertà effettivi sono pertanto $(\nu - k) = 5$.

I risultati ottenuti con il nostro codice `Monetafit` sono riportati in Tab. 9.63, che è analoga a quella descritta nell'esercizio precedente.

Equazione	\hat{p}	$\hat{\chi}^2$	SL
(9.57)	0.522 ± 0.015	$4.34 \simeq 50\%$	
(9.59)	0.528 ± 0.017	$4.84 \simeq 45\%$	

Anche in questo caso, i due metodi forniscono risultati simili. Tuttavia, nel caso di piccoli campioni come questo vi consigliamo, come regola generale, di usare la (9.57).

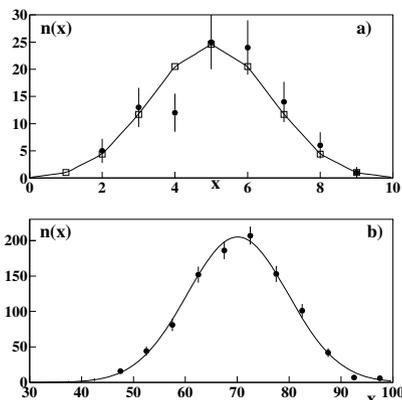


Fig. 9.5. Dati sperimentali con barra d'errore e curve di *best fit* a) per una densità binomiale (esercizio 9.7) e b) gaussiana (esercizio 9.6). Per facilitare il confronto, i punti discreti (quadrati vuoti) della binomiale di Fig. a) sono stati uniti da tratti di retta

L'ultima volta che ci siamo occupati delle 10 monete nell'ambito della statistica elementare, nell'esercizio 6.13, abbiamo stimato la probabilità direttamente dai dati, notando che l'esperimento forniva complessivamente 521 teste su 1000 lanci. Dalla (6.23) avevamo ottenuto:

$$p \in 0.521 \pm \sqrt{\frac{0.521(1 - 0.521)}{1000}} = 0.521 \pm 0.016 .$$

In questo caso l'ottimizzazione del parametro non ha portato a risultati diversi da quelli della stima diretta della probabilità p dai dati.

I dati sperimentali e le curve di *best fit* ottenute negli ultimi due esercizi sono rappresentati in Fig. 9.5.

9.8 La media pesata

Un'importante applicazione del metodo dei minimi quadrati consiste nel trovare la media di variabili gaussiane aventi tutte la stessa media vera μ ma varianza diversa.

Questo è un caso comune nell'attività di laboratorio, dove spesso è necessario combinare tra loro i risultati di misure di una stessa grandezza (stessa media vera) eseguite con apparati diversi (diverso errore di misura).

Come nel paragrafo precedente, deriviamo il risultato a partire dalla stima di massima verosimiglianza sotto l'ipotesi di osservazioni gaussiane, per

poi verificare che questo è ottenibile più generalmente come conseguenza del metodo dei minimi quadrati.

Per farvi capire meglio il problema, vi proponiamo prima un quesito interessante. Supponete di avere un campione di n osservazioni indipendenti e con la stessa media μ e deviazione standard σ ; raggruppateli ora in due campioni rispettivamente di k e $n - k$ osservazioni. Dato che per le medie vere vale la relazione $\mu = (\mu_1 + \mu_2)/2$, intuitivamente siamo portati a pensare che la media delle n osservazioni dovrebbe essere equivalente alla metà della somma delle due medie parziali, $m = (m_1 + m_2)/2$.

Questa è però una conclusione sbagliata. Infatti, come è facile verificare, si ha:

$$\frac{1}{n} \sum_{i=1}^n x_i \neq \frac{1}{2} \left[\frac{1}{k} \sum_{i=1}^k x_i + \frac{1}{n-k} \sum_{i=k+1}^n x_i \right].$$

L'uguaglianza è valida solo se $k = n - k$, cioè se i due sottocampioni hanno lo stesso numero di eventi. La spiegazione di questo paradosso è sottile e concettualmente importante: è sbagliato mediare le due medie parziali, perché, se i due sottocampioni sono diversi, esse hanno varianze σ^2/k e $\sigma^2/(n-k)$ differenti. Possiamo pensare allora di "pesare" le due medie in modo da avere comunque il risultato corretto. Se definiamo come peso l'inverso della varianza, allora $p_1 = (\sigma^2/k)^{-1}$ e $p_2 = [\sigma^2/(n-k)]^{-1}$ e si avrà

$$m = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{p_1 + p_2} \left[p_1 \frac{1}{k} \sum_{i=1}^k x_i + p_2 \frac{1}{n-k} \sum_{i=k+1}^n x_i \right].$$

Questa è la soluzione giusta, come vi mostreremo ora in modo generale, derivando la *formula della media pesata*.

Specializziamo la (9.50) al caso di n osservazioni gaussiane indipendenti suddivise in k sottogruppi di dimensione n_i ciascuno ($\sum_{i=1}^k n_i = n$), aventi tutte media μ ma varianza σ_i^2 secondo il sottogruppo. Il logaritmo negativo della funzione di verosimiglianza è

$$\mathcal{L} \equiv -\ln L(\mu, \boldsymbol{\sigma}; \underline{x}) = -\sum_{i=1}^k n_i \ln \left(\frac{1}{\sqrt{2\pi}\sigma_i} \right) + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{(x_{ij} - \mu)^2}{\sigma_i^2} \quad (9.64)$$

dove x_{ij} è l'osservazione j -esima nel sottogruppo i -esimo. Considerando le σ_i note, il primo termine di questa espressione non dipende da alcun parametro e può pertanto essere trascurato nella ricerca del minimo di \mathcal{L} , che diviene una funzione dipendente dal solo parametro μ . La condizione di minimo, indicando con m_i la media del sottogruppo i -esimo, è data allora da:

$$\frac{d\mathcal{L}}{d\mu} = \frac{1}{2} \frac{d\chi^2}{d\mu} = -\sum_i \sum_j \frac{x_{ij} - \mu}{\sigma_i^2} = -\sum_i \frac{n_i m_i}{\sigma_i^2} + \mu \sum_i \frac{n_i}{\sigma_i^2} = 0, \quad (9.65)$$

ovvero:

$$\hat{\mu} \equiv m = \frac{\sum_i (n_i m_i) / \sigma_i^2}{\sum_i n_i / \sigma_i^2}, \quad (9.66)$$

che è la famosa formula della media pesata. Con questa formula si trova il “baricentro” dei dati, pesando ognuno di essi con un fattore pari a:

$$p_i = \frac{n_i}{\sigma_i^2}. \quad (9.67)$$

Se i dati sono tutti estratti dalla stessa popolazione, allora $\sigma_i = \sigma$, e la (9.66) si trasforma nella solita formula della media campionaria:

$$\hat{\mu} \equiv m = \frac{(1/\sigma^2) \sum_i n_i m_i}{(1/\sigma^2) \sum_i n_i} = \sum_i \sum_j \frac{x_{ij}}{n}.$$

Per trovare l'errore statistico della media pesata basta applicare alla (9.66), come nel caso della media campionaria, la legge di trasformazione (5.73) per variabili indipendenti:

$$\text{Var} \left[\frac{\sum_i n_i M_i / \sigma_i^2}{\sum_i n_i / \sigma_i^2} \right] = \left(\frac{1}{\sum_i n_i / \sigma_i^2} \right)^2 \sum_i \frac{n_i^2 \text{Var}[M_i]}{\sigma_i^4}.$$

Poiché $\text{Var}[M_i] = \sigma_i^2 / n_i$, si ottiene:

$$\sigma_{\hat{\mu}}^2 = \left(\frac{1}{\sum_i n_i / \sigma_i^2} \right)^2 \sum_i \frac{n_i}{\sigma_i^2} = \frac{1}{\sum_i n_i / \sigma_i^2}. \quad (9.68)$$

La formula completa ad una deviazione standard della media pesata è data allora da:

$$\mu \in \frac{\sum_{i=1}^k m_i p_i}{\sum_{i=1}^k p_i} \pm \sqrt{\frac{1}{k \sum_{i=1}^k p_i}}, \quad p_i = \frac{n_i}{\sigma_i^2}. \quad (9.69)$$

Se tutti i dati provengono dalla stessa popolazione, allora $\sigma_i = \sigma$ e la formula si trasforma nella (6.46). I livelli di confidenza da assegnare all'intervallo di errore sono gaussiani (cioè seguono la legge 3σ), perché si riferiscono a variabili gaussiane combinate linearmente. Se le variabili non sono gaussiane, anche in questo caso il teorema Limite Centrale assicura che la normalità verrà raggiunta per n maggiore di una decina.

In pratica spesso le varianze non sono note e vengono stimate dai dati sperimentali ponendo $\sigma_i \simeq s_i$. È possibile mostrare che anche in questo caso i livelli di confidenza attribuibili all'intervallo (9.69) sono gaussiani quando n è elevato. La dimostrazione fa leva sulla consistenza delle s_i come stimatori delle σ_i , sul teorema Limite Centrale per la convergenza in distribuzione delle M_i e sull'indipendenza tra le osservazioni. Un'ipotesi necessaria alla dimostrazione è che il peso di ciascun sottogruppo di osservazioni non diventi trascurabile rispetto agli altri, il che si esprime richiedendo che n_i/n tenda a una costante per $n \rightarrow \infty$.

Possiamo ora mettere a frutto la nostra conoscenza sulla teoria degli stimatori chiedendoci se la media pesata è uno stimatore efficiente. Dato che essa è uno stimatore ML, sulla base del teorema 9.4 deduciamo che o essa è lo stimatore più efficiente oppure non esiste lo stimatore ottimale di Cramér-Rao per le somme pesate di dati. Applicando la (9.27) alla (9.65) è immediato vedere però che la media pesata è lo stimatore più efficiente, perché soddisfa al limite di Cramér-Rao del teorema 9.3:

$$\sum_i n_i I_i(\mu) = \left\langle -\frac{d^2 \ln L}{d\mu^2} \right\rangle = \left\langle -\frac{d}{d\mu} \sum_i \sum_j \frac{(x_{ij} - \mu)}{\sigma_i^2} \right\rangle = \sum_i p_i, \quad (9.70)$$

che è proprio l'inverso della varianza della (9.69). (Qui $I_i(\mu)$ indica l'informazione di Fisher per μ contenuta nella distribuzione di X_{ij}).

Esercizio 9.8

Utilizzando la routine `random` di un calcolatore, sono stati estratti dalla densità uniforme 20 numeri $0 \leq x_i \leq 1$. Essi sono riportati nella tabella seguente:

0.198	0.530	0.005	0.147
0.898	0.445	0.573	0.943
0.127	0.870	0.859	0.608
0.605	0.729	0.160	0.555
0.202	0.313	0.782	0.112

Eeguire la media dell'intero campione, la media pesata delle medie parziali dei primi 15 e degli ultimi 5 dati (le prime 3 colonne e l'ultima colonna della tabella) e confrontare i risultati.

Risposta. La media e deviazione standard dei tre campioni richiesti sono date, con ovvia notazione, da:

$$\begin{aligned} m_{20} &= 0.485, & s_{20} &= 0.302 \\ m_{15} &= 0.489, & s_{15} &= 0.298 \\ m_5 &= 0.473, & s_5 &= 0.347. \end{aligned}$$

Questi dati, in base alle (3.82), sono i valori assunti da una variabile uniforme di media vera $\mu = 0.5$ e deviazione standard vera $\sigma = 1/\sqrt{12} = 0.289$.

Applicando la (6.46) ai tre campioni otteniamo:

$$\begin{aligned} m_{20} &\in 0.485 \pm \frac{0.302}{\sqrt{20}} = 0.485 \pm 0.068 \\ m_{15} &\in 0.489 \pm \frac{0.298}{\sqrt{15}} = 0.489 \pm 0.077 \\ m_5 &\in 0.473 \pm \frac{0.347}{\sqrt{5}} = 0.47 \pm 0.15. \end{aligned}$$

La media pesata delle due medie parziali, che sono indipendenti poiché provengono da campioni che non hanno dati in comune, è data dalla (9.69):

$$\mu \in \frac{0.489 \cdot 168.7 + 0.473 \cdot 41.6}{168.7 + 41.6} \pm \frac{1}{\sqrt{168.7 + 41.6}} = 0.486 \pm 0.069 ,$$

dove $p_{15} = 15/s_{15}^2 = 168.7$ e $p_5 = 5/s_5^2 = 41.6$ sono i pesi. Come si vede, il risultato è praticamente identico a quello della grande media m_{20} .

Se non sapessimo la formula della media pesata, avremmo potuto applicare uno stimatore diverso, e precisamente quello dato dalla media aritmetica delle due medie parziali col relativo errore ottenuto dalla legge (5.67, 5.73):

$$\begin{aligned} \mu &\in \frac{m_{15} + m_5}{2} \pm \frac{1}{2} \sqrt{\frac{s_{15}^2}{15} + \frac{s_5^2}{5}} \\ &= \frac{0.489 + 0.473}{2} \pm 0.5 \sqrt{0.077^2 + 0.155^2} = 0.481 \pm 0.086 . \end{aligned}$$

È accettabile questo stimatore?

Se indichiamo con M_p ed M_s la media pesata e la media aritmetica possiamo verificare che questi stimatori sono non distorti. Infatti, applicando la (9.14) otteniamo:

$$\begin{aligned} \langle M_p \rangle &= \left\langle \frac{\sum_i M_i p_i}{\sum_i p_i} \right\rangle = \frac{\sum_i \langle M_i \rangle p_i}{\sum_i p_i} = \left\langle \mu \frac{\sum_i p_i}{\sum_i p_i} \right\rangle = \mu \\ \langle M_s \rangle &= \left\langle \frac{\sum_i M_i}{n} \right\rangle = \sum_i \frac{\langle M_i \rangle}{n} = \mu \frac{n}{n} = \mu . \end{aligned}$$

La differenza cruciale sta nel fatto che M_p , essendo uno stimatore ML, è il più efficiente. Si vede infatti che l'errore statistico di M_s è maggiore di circa il 20% di quello di M_p . Lo stimatore M_s è pertanto non ammissibile. In parole povere, anche M_s stima un intervallo che contiene la media vera (in senso frequentista), ma la larghezza di questo intervallo è maggiore di quella di M_p .

9.9 Verifica delle ipotesi

In questo e nei prossimi paragrafi completiamo l'importante argomento della verifica (o test) delle ipotesi, che abbiamo introdotto per la prima volta a livello un po' intuitivo negli esercizi 3.13-3.17 (pagine 109-112) e nel par. 6.14 a pagina 220, senza specificare esattamente le alternative rispetto alla quale l'ipotesi nulla veniva verificata.

Avendo definito la funzione di verosimiglianza, siamo ora in grado di affrontare l'argomento con maggior precisione, specificando una ipotesi nulla e una ipotesi alternativa e introducendo un criterio di ottimalità per la scelta

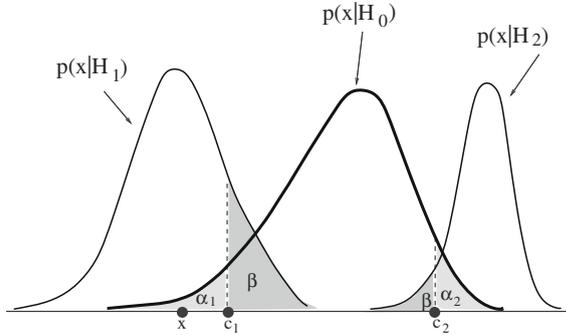


Fig. 9.6. Rappresentazione grafica delle quantità coinvolte nel test tra due ipotesi

tra le due ipotesi. L'argomento è però molto più vasto, perché si riescono a trattare casi in cui sia l'ipotesi nulla sia quella alternativa sono in realtà insiemi di ipotesi. Noi qui ci limiteremo a dare le idee di base, che però consentono già di definire una serie di metodi che sono applicabili in modo semplice e diretto a molti casi concreti.

Supporremo d'ora in poi che la funzione di verosimiglianza $L(\theta; x)$ non sia semplicemente proporzionale, ma sia coincidente con la funzione di densità. Se l'osservazione consiste di una sola misura, allora $L(\theta; x) = p(x; \theta)$, dove p è la densità della variabile aleatoria X . Di solito l'ipotesi viene verificata controllando se il valore di uno stimatore di θ (spesso funzione di una statistica sufficiente per θ) appartiene ad un "insieme critico". La funzione di verosimiglianza considerata sarà allora la densità di probabilità della distribuzione campionaria dello stimatore.

Sia data una osservazione x e si formulino due ipotesi H_0 (la principale, detta ipotesi nulla) ed H_1 , che rappresenta una possibile alternativa ad H_0 . Se le due ipotesi hanno come unico effetto un valore diverso dei parametri nella funzione di densità, in corrispondenza a queste due ipotesi si possono costruire, con notazione ovvia, due funzioni di verosimiglianza: $L(\theta_0; x)$ relativamente ad $H_0 : \theta = \theta_0$ e $L(\theta_1; x)$ relativamente ad $H_1 : \theta = \theta_1$. Per non appesantire il formalismo, considereremo qui solo il caso monodimensionale, ma tutte le conclusioni che trarremo valgono anche per una osservazione x e/o un insieme di parametri θ .

Nella verifica di più ipotesi, la situazione è quella di Fig. 9.6 ed il linguaggio quello delle Tab. 9.1 e 9.2, dove compaiono, oltre ai termini a voi già noti, alcuni termini nuovi.

Se $p(x|H_0)$ è la densità dello stimatore corrispondente all'ipotesi nulla, si accetta questa ipotesi se $c_1 \leq x \leq c_2$, la si rifiuta se x è nella regione critica, definita da $x < c_1$, $x > c_2$. L'area sottesa dalla regione critica è il livello di significatività SL , corrispondente alla probabilità di sbagliare rifiutando l'ipotesi H_0 quando è vera (errore di I tipo). Nel caso del test a due code il

Tabella 9.1. La terminologia dei test statistici

<i>Nome</i>	<i>Significato</i>
ipotesi nulla H_0	modello di riferimento
ipotesi alternativa H_1	modello alternativo
errore del I tipo α	scartare H_0 quando è vera
errore del II tipo β	accettare H_0 quando è vera H_1
livello di significatività SL	probabilità dell'errore del I tipo
area β	probabilità dell'errore del II tipo
livello del test α	valore SL fissato a priori
regione critica o di rifiuto	intervallo ($x < c_1$ o $x > c_2$) di Fig. 9.6
potenza del test $\pi = 1 - \beta$	scartare H_0 quando è vera H_1
test più potente	quello che, fissato α , ha $1 - \beta$ massimo
test a una coda	una sola coda a sinistra o a destra
test a due code	due code a destra e sinistra

livello di significatività è la somma delle due aree delle code indicate con α_1 e α_2 in Fig. 9.6. Se invece H_0 è sbagliata e l'ipotesi giusta è data dalla densità $p(x|H_1)$, l'area della coda, indicata come β sempre in Fig. 9.6, corrisponde alla probabilità di rifiutare l'ipotesi giusta, perché in questo caso l'ipotesi nulla H_0 viene accettata (errore di II tipo). L'area $\pi = 1 - \beta$ viene detta potenza del test, e corrisponde alla probabilità di scartare H_0 se è vera H_1 . La densità $p(x|H_2)$ di Fig. 9.6 indica la situazione simmetrica ad H_1 quando il massimo della densità relativa alla ipotesi alternativa si trova a destra del massimo dell'ipotesi nulla.

Se si fissa a priori un livello di significatività α , detto livello del test, relativo a due valori di coda c_1 e c_2 , si accetta l'ipotesi nulla H_0 quando $x \in [c_1, c_2]$ con probabilità pari a:

$$P\{c_1 \leq X \leq c_2 | H_0\} \equiv P\{X \in A | H_0\} = \int_{c_1}^{c_2} L(\theta_0; x) dx = 1 - \alpha, \quad (9.71)$$

dove A indica l'intervallo di accettazione $[c_1, c_2]$ nel caso monodimensionale e un sottoinsieme del supporto della densità di \mathbf{X} nel caso multidimensionale.

La potenza del test è la probabilità di scartare H_0 essendo vera H_1 , cioè la probabilità di ottenere valori nella regione critica quando è vera H_1 :

$$1 - P\{X \in A | H_1\} = \int_{-\infty}^{c_1} L(\theta_1; x) dx + \int_{c_2}^{+\infty} L(\theta_1; x) dx = 1 - \beta, \quad (9.72)$$

dove β è la probabilità dell'errore del II tipo, ovvero accettare H_0 quando è vera H_1 :

$$P\{X \in A | H_1\} = \int_{c_1}^{c_2} L(\theta_1; x) dx = \beta. \quad (9.73)$$

Tabella 9.2. Verifica tra due ipotesi: terminologia e corrispondenti livelli di probabilità

Ipotesi vera	Decisione	
	H_0	H_1
H_0	Decisione corretta $1 - \alpha$	Errore di I specie α
H_1	Errore di II specie β	Decisione corretta $1 - \beta$

Per un test ideale, in cui le densità corrispondenti ad H_0 ed H_1 hanno supporto disgiunto, $\beta = 0$ e la potenza $1 - \beta = 1$ è massima.

Queste sono le definizioni legate alla verifica delle ipotesi nel caso più generale di un test a due code, essendo ovvia la modifica necessaria per il test ad una coda, dove la regione di rifiuto ha forma $(-\infty, c)$ oppure $(c, +\infty)$.

9.10 Test di potenza massima

Quando si effettuano test statistici, in genere si fissa a priori il livello di significatività α richiesto dal problema, si costruisce uno stimatore $T = T_n(\underline{X})$ (la media, la varianza, il chi-quadrato, ecc.), si trova la distribuzione campionaria dello stimatore e si determinano da questa i valori quantili t_α o $t_{1-\alpha}$ dello stimatore corrispondenti alla probabilità α . Cambiando lo stimatore, per uno stesso valore di α , e fissata una ipotesi alternativa H_1 , la potenza del test risulta dipendere dallo stimatore prescelto. Poniamoci allora la fondamentale domanda: data una osservazione di una variabile X , quale è il test di potenza massima, cioè quello che ha la migliore regione critica?

La questione può avere in alcuni casi notevole interesse pratico, perché i test statistici spesso hanno un costo economico e/o gestionale elevato (si pensi ai test di controllo di qualità in ambito industriale) ed è quindi importante scegliere, dato un campione di dimensione fissata, il test più potente, per il quale l'errore di II tipo risulta minimo e di conseguenza il criterio di decisione più affidabile.

Una risposta alla domanda è data dal teorema di Neyman-Pearson:

Teorema 9.6 (di Neyman-Pearson). *Si consideri, sui parametri θ di una funzione di verosimiglianza, l'ipotesi nulla H_0 e l'ipotesi alternativa H_1 :*

$$H_0 : \theta = \theta_0 , \quad H_1 : \theta = \theta_1 .$$

Il rapporto di verosimiglianza:

$$R(X) = \frac{L(\theta_0; X)}{L(\theta_1; X)} , \quad (9.74)$$

assume valori grandi se è vera H_0 e valori piccoli se è vera H_1 . Il test di potenza massima tra tutti quelli di livello α è allora dato da:

$$\text{si rigetta } H_0 \text{ se } \left\{ R(X) = \frac{L(\theta_0; X)}{L(\theta_1; X)} \leq r_\alpha \right\}, \quad (9.75)$$

dove r_α è il valore di $R(X)$ di Fig. 9.7, tale che $P\{R(X) \leq r_\alpha | H_0\} = \alpha$.

Dimostrazione. Chiamiamo Ω (di livello α) la regione di rifiuto del test NP e Ω' (di livello α') una regione di rifiuto di H_0 attraverso un qualsiasi altro test diverso da quello NP. Il teorema vale se si dimostra che, a parità di livello, la probabilità dell'errore di II tipo è minima per il test NP:

$$\alpha = \alpha' \implies \beta' > \beta. \quad (9.76)$$

Per ipotesi si ha :

$$\alpha = \int_{\Omega} L(\theta_0; x) dx = \alpha' = \int_{\Omega'} L(\theta_0; x) dx.$$

Se $\beta' \neq \beta$ possiamo scrivere (si veda la Fig. 9.6):

$$\begin{aligned} \beta' - \beta &\equiv \Delta\beta = 1 - \int_{\Omega'} L(\theta_1; x) dx - \left[1 - \int_{\Omega} L(\theta_1; x) dx \right] \\ &= \int_{\Omega} L(\theta_1; x) dx - \int_{\Omega'} L(\theta_1; x) dx. \end{aligned} \quad (9.77)$$

Se $x \in \Omega$, allora la (9.75) vale sempre:

$$\frac{L(\theta_0; x)}{L(\theta_1; x)} \leq r_\alpha \implies L(\theta_1; x) \geq \frac{1}{r_\alpha} L(\theta_0; x),$$

mentre, se $x \in (\Omega' - \Omega)$:

$$\frac{L(\theta_0; x)}{L(\theta_1; x)} > r_\alpha \implies L(\theta_1; x) < \frac{1}{r_\alpha} L(\theta_0; x), \quad x \notin \Omega.$$

Sostituendo queste due ultime relazioni nella (9.77) e considerando che gli integrali su $\Omega \cap \Omega'$ coincidono, otteniamo la (9.76):

$$\Delta\beta > \frac{1}{r_\alpha} \left[\int_{\Omega} L(\theta_0; x) dx - \int_{\Omega'} L(\theta_0; x) dx \right] = \frac{1}{r_\alpha} [\alpha - \alpha'] = 0,$$

dimostrando così il teorema. \square

Applichiamo il teorema alla verifica dell'ipotesi di due medie diverse $H_0 : \mu = \mu_0$ contro $H_1 : \mu = \mu_1$ con campioni estratti da popolazioni di densità gaussiane di uguale varianza σ^2 . In questo caso si ottiene facilmente:

$$\frac{L_0}{L_1} = \exp \left\{ \frac{1}{2\sigma^2} \left[\sum (x_i - \mu_1)^2 - \sum (x_i - \mu_0)^2 \right] \right\} \leq r_\alpha,$$

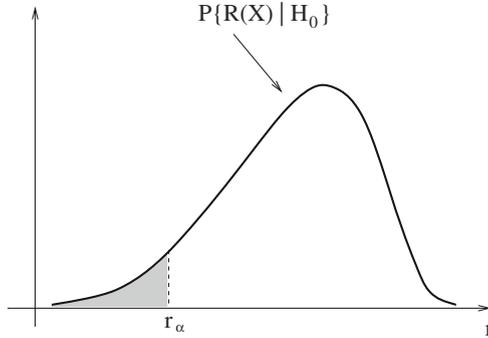


Fig. 9.7. Utilizzando il rapporto R di verosimiglianza, l'ipotesi nulla viene rifiutata se il valore del rapporto risulta minore del limite r_α , che determina il livello di significatività dato dall'area ombreggiata

e quindi, passando ai logaritmi:

$$2(\mu_0 - \mu_1) \sum_i x_i \leq 2\sigma^2 \ln r_\alpha + n(\mu_0^2 - \mu_1^2) . \tag{9.78}$$

Dividendo per $2n(\mu_0 - \mu_1)$, se $(\mu_0 - \mu_1) > 0$ si ottiene:

$$m_n = \frac{1}{n} \sum_i x_i \leq \frac{\sigma^2}{n(\mu_0 - \mu_1)} \ln r_\alpha + \frac{\mu_0 + \mu_1}{2} = m_0 . \tag{9.79}$$

In questo caso il test di Neyman-Pearson col calcolo esplicito di r_α è equivalente al test sulla media campionaria: si scarta H_0 se $\{M_n < m_0\}$. Il valore m_0 è determinato dalla condizione sull'errore di I tipo sotto H_0 :

$$\alpha = P\{M_n < m_0\} = P\left\{ \frac{M_n - \mu_0}{\sigma/\sqrt{n}} < \frac{m_0 - \mu_0}{\sigma/\sqrt{n}} \right\} ,$$

che è soddisfatta quando m_0 verifica l'eguaglianza

$$\frac{m_0 - \mu_0}{\sigma/\sqrt{n}} = t_\alpha \iff m_0 = \mu_0 + \frac{\sigma}{\sqrt{n}} t_\alpha$$

con t_α pari al quantile di ordine α della gaussiana standard.

Se $\mu_0 < \mu_1$, occorre cambiare segno alla disuguaglianza (9.78) e si scarta H_0 se $\{M_n > m_0 = \mu_0 + (\sigma/\sqrt{n}) t_{1-\alpha}\}$.

È facile vedere che il test sulla media campionaria è il più potente anche nel caso di test su medie di distribuzioni esponenziali e poissoniane [Gal84].

Esercizio 9.9

Si assume come ipotesi nulla H_0 che una variabile X sia gaussiana di media $\mu_0 = 10$ e deviazione standard $\sigma = 10$. Trovare la dimensione del campione e

la regione critica ottimale per accettare H_0 con $\alpha = 0.05$ e potenza $1 - \beta = 0.95$ rispetto all'ipotesi alternativa H_1 che la gaussiana sia di parametri $\mu_1 = 20$ e stessa $\sigma = 10$.

Risposta. Dalla (9.79) sappiamo che il test sulla media è quello di potenza massima, che ottimizza la regione critica. Se $\alpha = 0.05$ e $\beta = 0.05$, dalla Tab. D.1 otteniamo:

$$\frac{m_0 - \mu_0}{\sigma/\sqrt{n}} = t_{1-\alpha} = 1.645, \quad \frac{m_0 - \mu_1}{\sigma/\sqrt{n}} = t_\beta = -1.645.$$

Il valore m_0 è l'estremo della regione critica. Risolvendo nelle incognite m_0 ed n otteniamo:

$$n = \frac{\sigma^2(t_{1-\alpha} - t_\beta)^2}{(\mu_1 - \mu_0)^2} = 10.8, \quad m_0 = \frac{\mu_1 t_{1-\alpha} - \mu_0 t_\beta}{t_{1-\alpha} - t_\beta} = 15.$$

Il test richiesto deve quindi campionare $n = 11$ valori di X , calcolare la media campionaria M , accettare H_0 se $\{M \leq 15\}$, accettare H_1 se $\{M > 15\}$.

Esercizio 9.10

Una variabile S di distribuzione binomiale con p piccola ed n grande si può approssimare con una variabile di distribuzione normale $N(\mu, \sigma^2) = N(np, np)$. Si assume come ipotesi nulla H_0 che $p = 0.01$. Trovare la dimensione del campione e la regione critica ottimali per accettare H_0 con $\alpha = 0.05$ e potenza $1 - \beta = 0.80$ rispetto all'ipotesi alternativa H_1 che $p = 0.02$.

Risposta. Ripetendo il procedimento dell'esercizio precedente abbiamo:

$$\frac{s - 0.01n}{\sqrt{0.01n}} = t_{0.95} = 1.645, \quad \frac{s - 0.02n}{\sqrt{0.02n}} = t_{0.20} = -0.842,$$

dove i valori 1.645 e 0.842 si ricavano interpolando da Tab. D.1. Sottraendo la prima equazione dalla seconda si trova facilmente n :

$$0.01n = \sqrt{n}(0.1191 + 0.1645), \quad \text{da cui } n = 804$$

e quindi $s = 13$. Occorre un campione di dimensione 800 circa; se $s \leq 13$ si accetta H_0 , altrimenti si sceglie H_1 .

9.11 Funzioni test

Il test di Neyman-Pearson richiede la conoscenza della densità $p(r)$ del rapporto di verosimiglianza R , il che, pur essendo sempre possibile in linea di principio con le tecniche sviluppate nel cap. 5 sulle funzioni di variabili aleatorie, in pratica si rivela quasi sempre un'impresa piuttosto difficile e laboriosa.

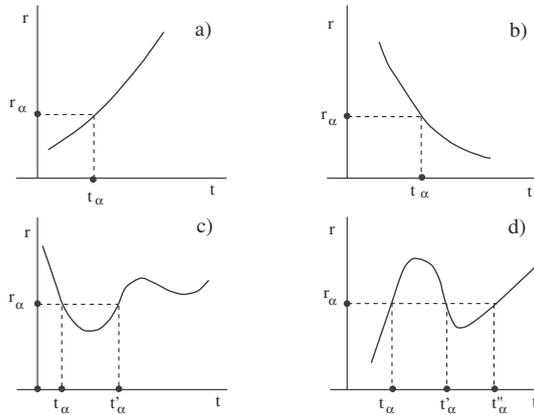


Fig. 9.8. Quando il rapporto di verosimiglianza è esprimibile come funzione di un'altra statistica, all'intervallo $R \leq r_\alpha$ relativo ad R può corrispondere un intervallo $T \leq t_\alpha$ a), $T \geq t_\alpha$ b), $t_\alpha \leq T \leq t'_\alpha$ c), $T \leq t_\alpha, t'_\alpha \leq T \leq t''_\alpha$ d), che dipende dal tipo di funzione $R = \psi(T)$

Fortunatamente, come avete appena visto nella (9.79), il problema si semplifica drasticamente se R può essere espresso nella forma

$$R = \psi(T) , \tag{9.80}$$

dove T è una statistica di distribuzione nota. Infatti, un test attraverso l'operatore T determina i limiti t_α corrispondenti al livello di significatività prescelto; tramite la (9.80) si può quindi determinare il limite r_α ed applicare il teorema 9.6. Tuttavia, questa operazione non deve necessariamente essere fatta in pratica: basta sapere che, se vale la (9.80), il test $R(X)$ può essere sostituito da un test $T(X)$. Anche questo test gode della proprietà di avere potenza massima.

Attraverso la (9.80) il test ad una coda della (9.75) $R(X) \leq r_\alpha$ può trasformarsi in un test ad una coda a destra o a sinistra, in un test a due code, oppure in un test con una regione critica più complicata a seconda del legame funzionale tra R e T , come mostrato in Fig. 9.8. Ad esempio, dalla (9.79) si vede che R è funzione della media campionaria. Ne segue che lo stimatore della media $T = \sum X_i/n$, è la funzione test della media più potente per campioni estratti da popolazioni aventi densità gaussiane, cioè quello che massimizza $1 - \beta$ una volta scelto il livello del test α .

Esercizio 9.11

Una ditta vince una gara d'appalto per la fornitura di un componente elettronico affermando che la percentuale di pezzi difettosi all'origine è minore dell'1%. Trovare il limite al di sotto del quale deve restare il numero di pezzi difettosi riscontrati in un lotto di controllo di 1000 pezzi perché si possa affermare, ad

un livello α del 5%, che la dichiarazione della ditta è corretta. Trovare inoltre la potenza di questo test rispetto alle ipotesi alternative di una probabilità di difetto del 2% e del 3% e dire se il test è di potenza massima.

Risposta. Se l'affidabilità è dell'1%, il numero di scarti su un lotto di 1000 pezzi è una variabile binomiale di valore atteso (media vera) pari a 10. Dato che la probabilità è piccola (1%) ed il valore della media è grande, è lecito utilizzare l'approssimazione gaussiana della distribuzione di Poisson. Siamo ricondotti al caso dell'esercizio 9.10, con la differenza che qui n è fisso e β incognito. Poiché il problema richiede un livello di significatività del 5%, dalla Tab. D.1 e dalle (3.43, 3.44) vediamo che il test a una coda

$$P\{T \geq t_{1-\alpha}\} = 1 - \Phi(t_{1-\alpha}) = 0.05$$

è soddisfatto per un valore della variabile standard $T = t_{1-\alpha} \simeq 1.645$. Se S è il numero degli scarti, dalla (3.37) abbiamo:

$$\frac{S - \mu}{\sigma} = \frac{S - 10}{\sqrt{10}} \geq 1.645, \quad \text{da cui: } S \geq 15.2.$$

Dato che la variabile considerata è discreta, non esiste un valore critico che coincide col valore SL assegnato. Potremmo a questo punto randomizzare il test ponendo nella (6.98) $\alpha_L = 0.05$ e

$$SL_2 = P\{S \geq 15\} = 1 - \Phi\left(\frac{15 - 10}{\sqrt{10}}\right) = 0.057,$$

$$SL_1 = P\{S \geq 16\} = 1 - \Phi\left(\frac{16 - 10}{\sqrt{10}}\right) = 0.029,$$

dove si è utilizzata la Tab. D.1. Come mostrato nel problema 9.15, il test randomizzato accetta il lotto 1 volta su 4 quando $S = 16$. Qui preferiamo procedere in modo più semplice (anche se approssimato), adottando come regola di decisione la accettazione del lotto se $S \leq 15$ ed il rifiuto se si trovano più di 15 pezzi difettosi.

La potenza del test è data dalla probabilità $1 - \beta$ di scartare H_0 (affidabilità dell'1%) quando è vera l'ipotesi alternativa H_1 . Ricordiamo che anche qui, come al solito, β è l'area di Fig. 9.6, che indica la probabilità dell'errore di II tipo. Considerando come H_1 un'affidabilità del 2%, dalla (3.43) e dalla Tab. D.1 otteniamo:

$$1 - \beta = 1 - P\{S < 16; H_1\} = P\{S \geq 16; H_1\}$$

$$= 1 - \Phi\left(\frac{16 - 20}{\sqrt{20}}\right) \simeq 0.81. \quad (9.81)$$

La potenza del test è di circa l'80% e la probabilità di accettare H_0 essendo vera H_1 è di circa il 20%.

Ripetendo il conto considerando H_1 corrispondente ad un'affidabilità del 3% otteniamo:

$$1 - \beta = 1 - \Phi \left(\frac{16 - 30}{\sqrt{30}} \right) \simeq 0.995, \quad (9.82)$$

da cui si vede che le ipotesi $H_0 : 1\%$ e $H_1 : 3\%$ danno luogo a distribuzioni con supporto praticamente disgiunto. In pratica, l'errore di II tipo viene commesso con probabilità apprezzabile per l'ipotesi alternativa di una affidabilità dell'ordine del 2%.

Questo test è di potenza massima, in base alla (9.79). Verifichiamolo anche considerando come funzione di verosimiglianza la distribuzione di Poisson, in approssimazione alla distribuzione binomiale, (3.14) e come ipotesi alternativa H_1 una affidabilità del 2%, troviamo il rapporto di verosimiglianza (9.74):

$$R(S) = \frac{10^S e^{-10}}{20^S e^{-20}} = e^{10} \left(\frac{1}{2} \right)^S \equiv \psi(S), \quad (9.83)$$

da cui si vede che vale la (9.80). Poiché il fattore esponenziale è costante, il legame tra R ed S è di tipo monotono decrescente, come in Fig. 9.8 b). La funzione inversa $S = \psi^{-1}(R)$ vale:

$$S = \frac{10 - \ln R}{\ln 2}. \quad (9.84)$$

Al livello del test prescelto corrisponde un intervallo $S \geq 15.5$ (valore intermedio tra i valori limite discreti trovati su S) ed un intervallo su R all'incirca pari a

$$R \simeq r_\alpha \leq e^{10}(0.5)^{15.5} = 0.47,$$

per il rigetto dell'ipotesi nulla.

Il test in esame è pertanto quello di potenza massima tra tutti i test possibili al livello del 5%.

Fino a questo momento abbiamo considerato ipotesi alternative del tipo $H_1 : \theta = \theta_1$, dette *puntuali o semplici*. Consideriamo ora il caso di ipotesi alternative, dette *composte*, che considerano un insieme di valori, come $H_1 : \theta > \theta_1$. In questi casi occorre esplorare la potenza del test per tutto l'insieme dei valori dei parametri dell'ipotesi alternativa, in modo da trovare il *test uniformemente più potente*. Non affrontiamo l'argomento, alquanto complesso, ma ci limitiamo ad osservare che il test del rapporto di verosimiglianza di solito si comporta bene da questo punto di vista. Nei casi (in verità non molti) in cui la potenza è calcolabile esplicitamente, si può comunque determinare per quali ipotesi alternative un test scelto risulta sufficientemente potente esaminando la *funzione di potenza*, definita, in funzione dei parametri della ipotesi alternativa, come:

$$\pi(\theta) = 1 - \beta(\theta) = 1 - P\{X \in A; H_1\}, \quad (9.85)$$

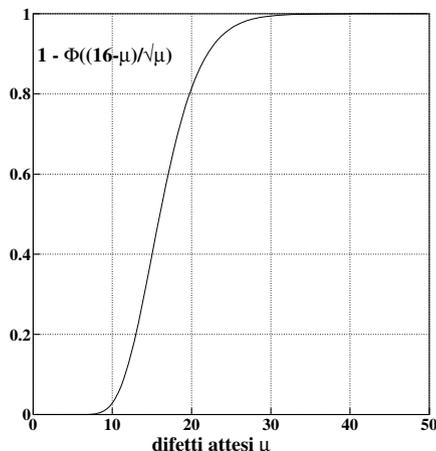


Fig. 9.9. Funzione di potenza nel caso dell'esercizio 9.11

dove A è la regione di accettazione di H_0 , cioè l'insieme complementare della regione critica.

Esercizio 9.12

Trovare la funzione di potenza nel caso dell'esercizio 9.11.

Risposta. Il criterio di decisione ottenuto consisteva nel rifiuto dell'ipotesi H_0 per un numero di difetti $S \geq 16$ su un lotto di 1000. Utilizzando l'approssimazione di Gauss (9.81) della densità di Poisson, la funzione di potenza (9.85) risulta:

$$\pi(\mu) = 1 - \Phi\left(\frac{16 - \mu}{\sqrt{\mu}}\right), \quad (9.86)$$

dove μ è il valore atteso del numero di pezzi difettosi. La funzione, riportata in Fig. 9.9, mostra che il supporto della distribuzione campionaria sotto H_0 (1% di difetti) diventa disgiunto da quello sotto H_1 quando si richiede in alternativa una percentuale di difetti superiore al $\geq 3\%$. Alla stessa conclusione eravamo già arrivati, in modo più intuitivo, già nell'esercizio 9.11.

9.12 Test sequenziali

La difficoltà nel trovare la distribuzione della variabile R di Neyman-Pearson può essere superata, oltre che con la trasformazione (9.80), anche in un altro modo, molto elegante ma approssimato.

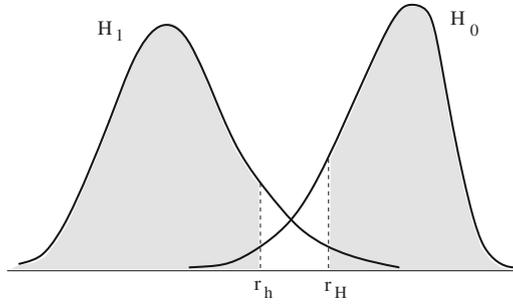


Fig. 9.10. In un test tra due ipotesi tramite il rapporto di verosimiglianza r , si accetta H_0 se $R > r_H$; si accetta H_1 se $R < r_h$; si rimane indecisi se $r_h < R < r_H$

Siano date due ipotesi H_0 ed H_1 e le corrispondenti funzioni di verosimiglianza $L(\theta_0; x) \equiv L_0(x)$ e $L(\theta_1; x) \equiv L_1(x)$. Il rapporto:

$$R = \frac{L_0(X)}{L_1(X)} \equiv \frac{L_0}{L_1} \tag{9.87}$$

tende ad assumere valori grandi se è vera H_0 , valori piccoli se è vera H_1 . Definiamo allora il test delle due ipotesi come segue:

$$\begin{aligned} \text{si accetta } H_0 \text{ se } (L_0/L_1) > r_H &\implies L_0 > r_H L_1 \\ \text{si accetta } H_1 \text{ se } (L_0/L_1) < r_h &\implies L_0 < r_h L_1 \\ \text{non si decide se } r_h < (L_0/L_1) < r_H . \end{aligned} \tag{9.88}$$

Le aree corrispondenti alle probabilità dell'errore di I e II tipo α e β sono mostrate in Fig. 9.10.

Per una osservazione data di dimensione n , lo spettro della variabile R viene quindi suddiviso in tre regioni disgiunte, R_0 , R_1 ed R_c , corrispondenti rispettivamente alle decisioni di accettare H_0 , accettare H_1 e di non prendere nessuna decisione. L'intervallo R_0 corrisponde allora alla condizione $R > r_H$ mentre R_1 è l'intervallo $R < r_h$. Dall'esame della Fig. 9.10 e dalle (9.88) risulta chiaro che le probabilità di prendere decisioni corrette sono date da:

$$1 - \alpha = \int_{R_0} L_0(x) dx \geq r_H \int_{R_0} L_1(x) dx = \beta r_H , \tag{9.89}$$

$$1 - \beta = \int_{R_1} L_1(x) dx \geq \frac{1}{r_h} \int_{R_1} L_0(x) dx = \frac{1}{r_h} \alpha , \tag{9.90}$$

dove, come al solito, α e β sono le probabilità assegnate di compiere errori di I e di II specie. Da queste disuguaglianze è immediato ricavare il limite inferiore di r_h e quello superiore di r_H :

$$\frac{\alpha}{1 - \beta} \leq r_h , \quad r_H \leq \frac{1 - \alpha}{\beta} . \tag{9.91}$$

In conclusione, possiamo enunciare la (9.88) come segue: in un test tra due ipotesi alternative H_0 ed H_1 , fissati a priori i livelli α e β , se

$$R \leq \frac{\alpha}{1-\beta} \leq r_h \quad \text{si accetta } H_1, \quad (9.92)$$

$$\frac{\alpha}{1-\beta} < R < \frac{1-\alpha}{\beta} \quad \text{non si prende alcuna decisione,} \quad (9.93)$$

$$R \geq \frac{1-\alpha}{\beta} \geq r_H \quad \text{si accetta } H_0. \quad (9.94)$$

Questo test è approssimato, ma *prescinde dalla conoscenza della distribuzione di probabilità del rapporto di verosimiglianza* (9.74).

Quali errori di primo e secondo tipo α' e β' sono effettivamente associati al test (9.92-9.94)? A questo quesito si può rispondere facilmente, anche se in modo parziale, poiché si dimostra che *si ha sempre* $\alpha' + \beta' \leq \alpha + \beta$. Infatti, dalle (9.89-9.91) si ha facilmente:

$$1 - \alpha' \geq \beta' r_H \geq \beta' \frac{1-\alpha}{\beta},$$

$$1 - \beta' \geq \frac{\alpha'}{r_h} \geq \alpha' \frac{1-\beta}{\alpha},$$

da cui:

$$\frac{\alpha'}{1-\beta'} \leq \frac{\alpha}{1-\beta}, \quad \frac{1-\alpha'}{\beta'} \geq \frac{1-\alpha}{\beta},$$

e quindi:

$$\alpha'(1-\beta) + \beta'(1-\alpha) \leq \alpha(1-\beta') + \beta(1-\alpha') \implies \alpha' + \beta' \leq \alpha + \beta. \quad (9.95)$$

Esercizio 9.13

Eseguire il test approssimato del rapporto di verosimiglianza sui dati dell'esercizio 9.11, assumendo un errore di I specie del 5% ed un errore di II specie del 20%.

Risposta. Nell'esercizio 9.11 avevamo trovato, studiando la funzione di probabilità della variabile somma, collegata alla variabile R di Neyman-Pearson dalla (9.84), che l'ipotesi H_0 di affidabilità all'1% andava scartata, ad un livello di significatività α del 5%, per un numero di pezzi difettosi maggiore di 15 su un lotto di 1000 pezzi. Avevamo anche trovato una probabilità β di circa il 20% per l'errore di II specie rispetto all'ipotesi H_1 di una affidabilità del 2%.

Anche in questo caso si chiede $\alpha = 0.05$ e $\beta = 0.20$. Dalle (9.92-9.94) otteniamo:

$$\begin{array}{ll} \text{si accetta } H_0 \text{ se} & R \geq 4.75 \\ \text{non si decide se} & 0.0625 < R < 4.75 \\ \text{si accetta } H_1 \text{ se} & R \leq 0.0625. \end{array}$$

Dalla (9.84) risulta che ai valori $r_H = 4.75$ ed $r_h = 0.0625$ corrispondono rispettivamente i valori:

$$s_h = 18.4, \quad s_H = 12.2.$$

L'ipotesi nulla di una affidabilità minore dell'1% (H_0) va quindi mantenuta se il numero dei pezzi difettosi resta al di sotto di 12, mentre se questo valore supera 18 va accettata l'ipotesi H_1 , relativa ad una affidabilità maggiore del 2%. Valori tra 12 e 18 si riferiscono ad una affidabilità stimata tra l'1 e il 2% e rappresentano pertanto una zona di incertezza nella quale è bene non prendere alcuna decisione.

Il fatto che i limiti nelle (9.92-9.94) non dipendano dalla dimensione n del campione suggerisce una applicazione del metodo del rapporto di verosimiglianza da effettuarsi ripetutamente nel corso di un esperimento, man mano che n aumenta, finché ci si trova nella zona di indecisione. Più precisamente, si continua a campionare finché

$$r_h < R_n < r_H, \quad n \geq 1 \quad (9.96)$$

e ci si ferma la prima volta che la (9.96) non è verificata. I limiti dell'intervallo di indecisione sono da specificare in funzione del livello del test e dell'errore di II tipo. Si noti che sia la dimensione del campione N sia il valore finale del rapporto di verosimiglianza

$$R_N = \frac{L_0(\theta_0; X_N)}{L_1(\theta_1; X_N)}, \quad (9.97)$$

al momento in cui si esce dalla zona di indecisione, sono ora variabili aleatorie.

Questi test sono detti *sequenziali*.

Se supponiamo di aver determinato r_h ed r_H in modo che le probabilità di errore di I e II tipo siano α e β , anche in questo caso valgono le (9.91), perciò un test sequenziale di livello approssimato, che è quello che si è in grado di condurre in pratica, avrà come zona di indecisione

$$\frac{\alpha}{1-\beta} < R_n < \frac{1-\alpha}{\beta}, \quad n \geq 1. \quad (9.98)$$

Anche in questo caso per i livelli effettivi del test vale la (9.95): $\alpha' + \beta' \leq \alpha + \beta$.

Il test sequenziale è dunque approssimato ma, a parità di potenza richiesta, richiede in media, rispetto ai test con dimensione n fissa, una dimensione minore, consentendo un risparmio medio notevole nelle operazioni di campionamento, anche del 50%. Intuitivamente, questo accade perché spesso si hanno sequenze di casi favorevoli (o sfavorevoli) che fanno assumere rapidamente ad R_N valori al di fuori dell'intervallo di indecisione (9.98) e consentono quindi di scegliere subito l'ipotesi H_0 (od H_1). Sotto condizioni molto generali, è possibile dare una dimostrazione di questa proprietà molto utile dei test sequenziali [MGB91, Pic96], che noi omettiamo. Vi presenteremo invece, nel prossimo esercizio, un approccio diverso che fa uso di tecniche di simulazione.

Esercizio 9.14

Applicare il test sequenziale per $\alpha = 0.05$ e $\beta = 0.20$, con N variabile, al caso dell'esercizio 9.11, assumendo che i pezzi difettosi siano l'1% (ipotesi H_0) o il 2% (ipotesi alternativa H_1). Stimare, con metodi di simulazione, le distribuzioni dei difetti rilevati S_N e dei pezzi campionati N .

Risposta. Se N è variabile, possiamo scrivere la (9.83) come:

$$R_N = \left(\frac{1}{2}\right)^{S_N} e^{N/100},$$

dove S_N è il numero di pezzi difettosi. Passando ai logaritmi otteniamo:

$$\ln R_N = -S_N \ln 2 + \frac{N}{100} \simeq -\frac{S_N}{1.44} + \frac{N}{100}$$

e la (9.98) diventa:

$$1.44 \ln \left(\frac{\alpha}{1-\beta} \right) - \frac{1.44}{100} N \leq -S_N \leq -\frac{1.44}{100} N + 1.44 \ln \left(\frac{1-\alpha}{\beta} \right).$$

Cambiando i segni ed inserendo i valori $\alpha = 0.05$ e $\beta = 0.20$ l'intervallo di indecisione, entro il quale continuare a campionare, diventa:

$$-2.25 + 0.0144 N < S_N < 3.99 + 0.0144 N. \quad (9.99)$$

Per valori di S_N a sinistra di questo intervallo si accetta H_0 , per valori a destra si accetta H_1 .

Nella (9.99) compaiono le due variabili aleatorie discrete S_N ed N , la cui distribuzione è molto difficile da determinare con metodi analitici. Possiamo però aiutarci con una simulazione, procedendo come segue: si considera un numero $0 \leq \xi \leq 1$ fornito dalla routine `rndm` e si incrementa N di uno; se $\xi \leq 0.01$ (ipotesi H_0) si incrementa di uno anche S_N ; se S_N ed N soddisfano la (9.99) si procede ad un ulteriore passo, se $S_N \geq 3.99 + 0.0144 N$ si sceglie H_1 , se $S_N \leq -2.25 + 0.0144 N$ si accetta H_0 . Lo stesso test può essere simulato sotto l'ipotesi alternativa H_1 , incrementando S_N se $\xi \leq 0.02$. Il risultato, ottenuto ripetendo il test al calcolatore per 100 000 volte, col nostro codice `Sequen` è riportato in Fig. 9.11. La simulazione mostra, in Fig. 9.11a) e b), che il numero medio di pezzi campionati è 451 ± 1 sotto H_0 e 570 ± 1 sotto H_1 , circa la metà di quelli necessari per il test a campione fisso, che, come mostrato nell'esercizio 9.10, è di circa 800, a parità dei livelli $\alpha = 0.05$ e $\beta = 0.20$. Si vede inoltre che N ha una distribuzione con una coda di tipo esponenziale a destra, con un piccolo (ma non trascurabile) numero di test in cui si raggiungono valori superiori a quello del test a N fisso. La densità di S_N sotto H_0 , riportata in Fig. 9.11 c), è di tipo esponenziale, con un valor medio di circa 5 pezzi. Sopra gli istogrammi di Fig. a) e b) è riportata anche una stima dei valori dei livelli $1 - \alpha'$ e $1 - \beta'$ "sperimentali",

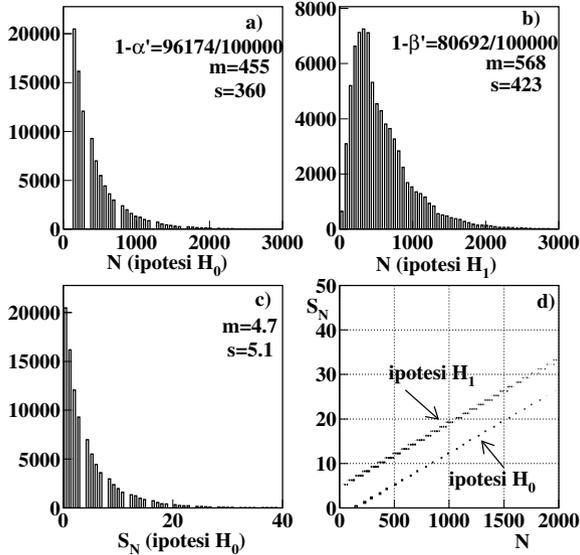


Fig. 9.11. Risultati della simulazione di 100 000 test sequenziali dell'esercizio 9.14. Istogramma del numero N di pezzi da campionare se la probabilità di difetto è dell'1% (ipotesi H_0) a); lo stesso se la probabilità di difetto è del 2% (ipotesi H_1) b); istogramma del numero degli scarti S_N sotto H_0 c); punti nel piano (N, S_N) corrispondenti all'accettazione delle due ipotesi d)

cioè il numero di test in cui si è accettata H_0 quando è vera (Fig. a)) e H_1 quando è vera (Fig. 9.11 b). Dai dati si ricavano i valori $\alpha' \in (3.8 \pm 0.3)\%$ e $\beta' \in (19.3 \pm 0.3)\%$ (verificate l'errore statistico come esercizio). Essi non coincidono con i valori fissati a priori perchè il test coinvolge variabili discrete. La somma $\alpha' + \beta' \in (23.1 \pm 0.4)\%$ è minore del valore $\alpha + \beta = 25\%$, in accordo con la (9.95). In Fig. 9.11 d) sono riportati, nel piano (N, S_N) i valori ottenuti nei 100 000 test, sotto le due ipotesi. I punti si dispongono appena al di sopra o al di sotto delle due rette limite della (9.99). Questa curva permette la determinazione per via grafica dell'intervallo di indecisione; ad esempio, quando $N = 500$, se $S_N < 5$ si sceglie H_0 , se $S_N > 11$ si sceglie H_1 . Va anche notato come in ogni istogramma compaia evidente la struttura discreta del problema.

Problemi

9.1. Un'urna contiene biglie bianche e nere in proporzione di 2 : 1 o di 1 : 2 (non si sa a favore di quale colore). Se in un esperimento si fanno 4 estrazioni con reimmissione, trovare la stima ML della proporzione in funzione dei risultati possibili di biglie nere estratte ($x = 0, 1, 2, 3, 4$).

9.2. Utilizzando il metodo del problema precedente, trovare la stima ML della probabilità p in funzione dei successi x ottenuti in $n = 3$ tentativi, considerando i 9 valori possibili: $p = 0.1, 0.2, 0.3, \dots, 0.9$.

9.3. Se $\hat{\mu}$ e $\hat{\sigma}$ sono stime ML di μ e σ , trovare la stima ML del quantile $F(x_\alpha) = \alpha$.

9.4. Avendo osservato n tempi t_i da una popolazione avente densità esponenziale $\lambda \exp[-\lambda t]$, fare la stima ML di λ .

9.5. Un metodo per stimare il numero N di elementi di una popolazione finita consiste nell'estrarre un campione casuale di n elementi, marchiarli e reimmetterli nella popolazione. Si estrae poi un secondo campione di n elementi e si determina il numero x di elementi marchiatati. Se in un esperimento si fissa $n = 150$ e si trova $x = 37$, fare la stima ML di N , a) in modo esatto e b) nell'approssimazione $N \gg n$. (*Suggerimento:* utilizzare la legge ipergeometrica (1.32) di pagina 25.)

Che metodi si potrebbero usare per determinare in modo approssimato la stima per intervallo di N , cioè $N \in \hat{N} \pm \sigma[\hat{N}]$?

9.6. Se X è una variabile di Bernoulli di densità $b(x) = p^x(1-p)^{1-x}$, verificare se le statistiche $S = X_1 + X_2$ e $P = X_1 X_2$ sono sufficienti.

9.7. Si ottiene da un campione di dimensione n la somma $w = \sum_i x_i^2$. Sapendo che la variabile ha distribuzione normale $N(0, \sigma^2)$, con media nulla nota e varianza da determinare, applicate i metodi a) e b) del par. 9.5 per determinare l'intervallo di confidenza di σ^2 .

9.8. Campionando da una popolazione normale $N(\mu = 0, \sigma^2)$ si sono ottenuti i valori: $x_i = 1.499, 5.087, 0.983, 2.289, 1.045, -1.886$. Trovare la stima ML puntuale e per intervallo della varianza σ^2 , per $CL = 95.4\%$. Fate ora la stima con la (6.72) di pagina 212. Questo risultato è più corretto del precedente?

9.9. Da 100 valori simulati di $X \sim N(0, \sigma^2)$ si è ottenuto $w = \sum_i x_i^2 = 1044$. Stimate la varianza con i metodi dei due problemi precedenti.

9.10. Due campioni dei diametri (in cm) di cuscinetti a sfere prodotti dalla stessa macchina in due settimane differenti hanno numero di eventi, media e deviazione standard rispettivamente pari a $n_1 = 100$, $m_1 = 2.08$, $s_1 = 0.16$ e $n_2 = 200$, $m_2 = 2.05$, $s_2 = 0.15$. Fare la stima ML e per intervallo del diametro medio.

9.11. Un monitor di radioattività rileva o meno la presenza di particelle nucleari senza contarne il numero. Esaminando 50 campioni casuali omogenei della stessa quantità di sostanza nella stessa unità di tempo, 45 sono risultati positivi al monitor. Fare la stima ML del numero medio di particelle emesse dalla quantità di sostanza nell'unità di tempo.

9.12. La registrazione di $N = 1000$ tempi di arrivo ha fornito l'istogramma:

t	2	4	6	8	10	12	14	16	18	20
$n(t)$	472	276	132	51	36	12	11	7	1	2

da cui risulta che tra 0 e 2 secondi si sono registrati 472 intertempi, tra 2 e 4 secondi 276 intertempi e così via. Eseguire il *best fit* utilizzando la densità esponenziale $e(t) = \lambda \exp(-\lambda t)$ determinando la stima ML di λ . Verificare la bontà dell'ipotesi col test χ^2 .

9.13. Secondo l'ipotesi nulla H_0 , una variabile binaria X assume i valori 0 e 1 con probabilità $1 - \varepsilon$ ed ε rispettivamente: $P\{X = 0, 1; H_0\} = (1 - \varepsilon), \varepsilon$. Sia $0 < \varepsilon \ll 1$ una quantità piccola e positiva. Si consideri l'ipotesi alternativa $P\{X = 0, 1; H_1\} = \varepsilon, (1 - \varepsilon)$. Si calcoli il livello di significatività e la potenza del test seguente: si accetta H_0 se in una prova $\{X = 0\}$, si accetta H_1 se $\{X = 1\}$.

9.14. Nella situazione del problema precedente, si consideri il risultato di due prove indipendenti ed il test seguente: si accetta H_0 se $\{X_1 = 0, X_2 = 0\}$; si sceglie H_1 se $\{X_1 = 1, X_2 = 1\}$; si lancia una moneta per scegliere a caso tra H_0 ed H_1 quando $\{X_1 + X_2 = 1\}$. Trovare il livello di significatività e la potenza di questo test e confrontare il risultato con quello del problema precedente.

9.15. Come si esegue il test randomizzato nel caso dell'esercizio 9.11?

9.16. Una ditta di automobili utilizza delle sospensioni che resistono in media 100 ore ad un test estremo di fatica (ipotesi H_0). Un fornitore propone un tipo di sospensione che a suo dire resiste in media 110 ore (ipotesi alternativa H_1). Sapendo che la distribuzione dei tempi è esponenziale negativa, progettare un test sul nuovo tipo di sospensione. a) Tenere la dimensione del campione fissa, utilizzare l'approssimazione normale per la media, e fare in modo che la probabilità di sbagliarsi se le nuove sospensioni sono uguali alle vecchie sia l'1% e la probabilità di accettare sospensioni buone sia il 95%. b) Risolvere ora il problema con un test sequenziale e valutare, con un codice di simulazione, le distribuzioni della media dei tempi e del numero di pezzi necessari al test. Conviene utilizzare il campione a dimensione fissa o il test sequenziale? Quale strategia adattereste?

9.17. Trovare la funzione di potenza nel caso del problema 9.16. (*Suggerimento:* confrontare la (3.57) di pagina 93 con la (3.67) di pagina 97.)

9.18. Si vuole verificare l'ipotesi nulla che la probabilità di avere testa nel lancio di una certa moneta sia $p = 0.5$ contro l'ipotesi alternativa $p = 0.3$. Qual è il numero n di lanci necessario per decidere tra queste due ipotesi, supponendo un livello del 10% e una potenza del 90%, cioè $\alpha = 0.10$ e $\beta = 0.10$? Si utilizzi l'approssimazione normale.

9.19. Con gli stessi valori di α e β del problema precedente, trovare il numero di successi x in funzione del numero dei lanci n che permetta di scegliere tra le due ipotesi con un test sequenziale. Determinare, lanciando una moneta o con un codice di simulazione, il numero medio di lanci n necessario per decidere tra le due ipotesi. Si confronti il risultato con quello del problema precedente.

9.20. Se X ha densità $p(x) = (1 + \theta)x^\theta$, $0 \leq x \leq 1$, si trovi, con un campione di $n = 100$ eventi, la miglior regione critica per il test di $\theta = \theta_0 = 1$ contro $\theta = \theta_1 = 2$ ad un livello $\alpha = 0.05$. Determinare la potenza del test e la funzione di potenza.

Minimi quadrati

È infine necessario, dopo che tutte le condizioni del problema sono espresse in modo opportuno, determinare i coefficienti in modo da rendere gli errori il più possibile piccoli. A tal fine, il metodo che mi è parso più semplice e più generale consiste nel minimizzare la somma dei quadrati degli errori.

Adrian Legendre, Nuovi metodi per la determinazione delle orbite delle comete

10.1 Introduzione

Nel par. 9.6 abbiamo mostrato che il metodo dei minimi quadrati (metodo LS, *Least Squares*) discende dal principio di massima verosimiglianza (metodo ML, *Maximum Likelihood*) quando le variabili sono gaussiane.

Storicamente le cose sono andate però in modo diverso. Infatti, mentre la massima verosimiglianza fu introdotta da Fisher agli inizi del 1900, il metodo dei minimi quadrati fu applicato per la prima volta dal matematico francese Legendre fin dal 1803, come indicato nell'epigrafe di questo capitolo. Successivamente Laplace, nel suo famoso trattato "Théorie Analytique des Probabilités" del 1812, dimostrò che il metodo LS produce stime non distorte anche nel caso di variabili non gaussiane. Il passo decisivo per la collocazione corretta del metodo LS nell'ambito della statistica fu poi compiuto da Gauss nel 1821, con la dimostrazione che gli stimatori LS, quando i valori attesi delle variabili aleatorie sono funzioni lineari dei parametri da determinare, sono stimatori non distorti ed efficienti (a varianza minima). Questo fondamentale teorema fu esteso e meglio formalizzato da Markov nel 1912, ed è oggi noto come teorema di Gauss-Markov.

In base a questi risultati possiamo dire che, quando non sono note a priori le densità delle popolazioni campionarie, in alternativa al metodo della massima verosimiglianza può essere usato il metodo dei minimi quadrati in tutti quei casi in cui si possono esprimere i valori attesi come combinazioni lineari dei parametri. Il principio LS non è solo quindi una conseguenza, ma anche

un complemento al metodo ML, relativamente alla classe degli stimatori non distorti a varianza minima. Se i valori attesi sono una combinazione non lineare dei parametri, il metodo LS è ancora utilizzabile, ma a prezzo della perdita delle proprietà di non distorsione e minima varianza. Accenneremo brevemente, nel corso di questo capitolo, anche al caso non lineare.

Il metodo LS, come illustrato nel par. 9.6, consiste nel trovare il minimo di una espressione di tipo χ^2 , come la (9.52). Generalizzando, possiamo definire il metodo come:

Definizione 10.1 (Metodo dei Minimi Quadrati (LS)). *Siano y_1, y_2, \dots, y_n i valori osservati una variabile aleatoria Y (detta “variabile risposta”) tali che $\langle Y_i \rangle = \mu_i(x_i, \boldsymbol{\theta}) \equiv \mu_i(\boldsymbol{\theta})$. Nella funzione $\mu_i(x_i, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$ è un vettore di parametri di dimensione $p + 1$ e x_1, \dots, x_n sono gli n valori osservati di una variabile X detta “predittore”. Le varianze $\text{Var}[Y_i] = \sigma_i^2$ sono note. La stima $\hat{\boldsymbol{\theta}}$ dei minimi quadrati di $\boldsymbol{\theta}$ è quella che rende minima la quantità:*

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{[y_i - \mu_i(\boldsymbol{\theta})]^2}{\sigma_i^2}. \quad (10.1)$$

Il predittore può essere anche un vettore \mathbf{X} . In tal caso scriveremo $\mu_i(\mathbf{x}_i, \boldsymbol{\theta})$, in corrispondenza dell' i -esimo valore osservato di \mathbf{X} .

Dalla definizione si comprende che le funzioni $\mu_i(\boldsymbol{\theta})$, al variare di $\boldsymbol{\theta}$, costituiscono una classe di modelli che mettono in relazione la variabile (aleatoria o deterministica) X con $\langle Y \rangle$; in virtù di tale relazione, che permette di valutare $\langle Y_i \rangle$ senza aver osservato y_i , la X viene chiamata predittore, mentre la Y viene chiamata risposta.

La minimizzazione della (10.1) rispetto a $\boldsymbol{\theta}$ ha lo scopo di identificare il miglior modello in questa classe. Notiamo che, a differenza del metodo ML, qui non è necessario conoscere la distribuzione delle variabili per stimare $\boldsymbol{\theta}$, ma solo il loro valore atteso e la varianza. Un'altra caratteristica del metodo è che le varianze σ_i^2 non dipendono da $\boldsymbol{\theta}$.

Dopo aver stimato $\boldsymbol{\theta}$, se non si è sicuri del modello, è opportuno eseguire un test di adattamento ai dati. Se le varianze σ_i^2 sono note, è lecito utilizzare il test χ^2 di adattamento, in modo simile a quanto fatto nei paragrafi 6.17 e 6.18. Per evitare equivoci, ricordiamo che:

- il procedimento di minimizzazione è puramente matematico e non statistico, e consiste nel trovare il minimo della quantità “tipo χ^2 ” (10.1);
- la quantità

$$(\chi^2)_{min} \equiv \chi^2(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \frac{[y_i - \mu_i(\hat{\boldsymbol{\theta}})]^2}{\sigma_i^2} \quad (10.2)$$

ha distribuzione $\chi^2(n - p - 1)$ se le Y_i sono gaussiane indipendenti con valore atteso $\mu_i(\boldsymbol{\theta})$, a sua volta funzione lineare di $\boldsymbol{\theta}$.

Con questa precisazione, pensiamo che indicare sempre come χ^2 la quantità da minimizzare, anche se essa può *non* avere distribuzione $\chi^2(n-p-1)$, non porti a confusione¹.

Se il test di adattamento non rifiuta l'ipotesi nulla o se si è già sicuri del modello adottato, può essere utile dare degli intervalli di confidenza per θ oppure per $\mu(\theta)$. Vedremo come effettuare tale operazione nel caso lineare gaussiano. Gli intervalli di confidenza per $\mu(\theta)$ servono a prevedere $\langle Y_i \rangle$ in corrispondenza di valori di x_i per i quali non sono stati fatti esperimenti.

Nel par. 9.7 abbiamo presentato un esempio del metodo LS per l'adattamento di funzioni di densità ad istogrammi, come conseguenza del principio della massima verosimiglianza. Tuttavia l'applicazione più importante, che è l'oggetto di questo capitolo, riguarda la ricerca di forme funzionali, contenute nelle funzioni $\mu_i(\theta)$. L'obiettivo che ci proponiamo è quello di darvi un metodo (anche se non necessariamente il migliore in tutti i casi) per la stima e la verifica dell'adattamento utilizzabile in ogni situazione e di mettervi in grado, per ogni tipo di problema, di scegliere l'algoritmo giusto e di utilizzare in modo appropriato i codici di minimizzazione, interpretandone correttamente i risultati dal punto di vista statistico.

Non descriveremo in dettaglio gli algoritmi di questi codici, perché questo è un aspetto essenzialmente di calcolo numerico, più che di statistica. Per maggiori dettagli, si possono consultare i testi [BR92], [PFTW92], oppure direttamente i manuali d'uso dei codici di minimizzazione, come [Jam92]. Nelle nostre pagine web [RPP] troverete anche le routine `Linfit` e `Nlinfit` per eseguire minimi quadrati lineari e non lineari ed ulteriori informazioni di carattere tecnico-numerico.

Nel seguito del capitolo useremo qualche volta, oltre ai termini italiani corretti, anche i termini inglesi *fit* e *best fit* che fanno parte ormai del gergo comune a molti ricercatori. Per evitare equivoci, riportiamo qui sotto la corrispondenza tra i termini utilizzati:

adattamento	=	<i>fit</i>
miglior adattamento	=	<i>best fit</i>
stime LS dei parametri $\hat{\theta}$	=	parametri di <i>best fit</i> $\hat{\theta}$
curva di regressione o curva dei minimi quadrati	=	curva di <i>best fit</i>

Cominciamo ora col discutere due dei tipi di dipendenza tra variabili che si possono presentare e sui quali ci concentreremo per la maggior parte di questo capitolo.

¹ I gradi di libertà ora sono $n-p-1$ e non $n-p$ perché nella def. 10.1 la dimensione di θ è $p+1$.

10.2 Predittori osservati senza errore

In statistica un caso molto frequente è dato da una variabile aleatoria Y espressa come:

$$Y = f(x) + Z, \quad f(x) \equiv f(x, \boldsymbol{\theta}), \quad (10.3)$$

dove $f(x)$ è una funzione nota (a meno del parametro $\boldsymbol{\theta}$) del predittore non aleatorio x . Osservate che si può sempre porre $\langle Y \rangle = f(x)$ e $\langle Z \rangle = 0$, perché se fosse $\langle Z \rangle = z_0$ basterebbe ridefinire f come $f(x) + z_0$ e si ricadrebbe ancora nella (10.3). Questo accorgimento permette di rappresentare Y come somma di una componente deterministica (il suo valore atteso), che coglie la parte essenziale della relazione col predittore x , e di una parte aleatoria, che rappresenta le fluttuazioni attorno alla media.

Nel caso di un insieme di valori x_i fissati e osservabili senza errore, un campionamento della Y ripetuto più volte per ciascun valore x_i fornisce un grafico come quello di Fig. 10.1 a), con le x_i fisse e le Y_i che fluttuano intorno ai loro valori medi; se si effettua una sola prova e la deviazione standard delle Z_i è nota, il grafico si disegna come in Fig. 10.1 b), con le *barre d'errore*, calcolate a partire dalle deviazioni standard σ_i delle Z_i , centrate sui valori misurati y_i .

Il compito del metodo dei minimi quadrati è determinare la *forma funzionale* $f(x)$ che lega $\langle Y \rangle$ alla variabile deterministica x . In altre parole, occorre determinare, per ogni x , la media $f(x, \hat{\boldsymbol{\theta}})$ come la curva *rispetto alla quale le fluttuazioni di Y sono casuali*, dove “casuali” significa che la dipendenza tra $\langle Y \rangle$ e x è interamente compresa in f , e che quanto rimane è un puro termine

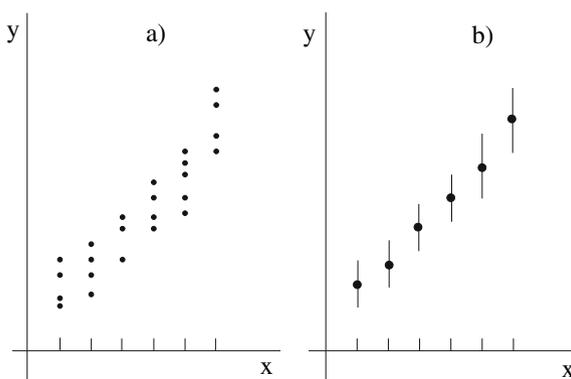


Fig. 10.1. Campionamento di una variabile aleatoria y funzione di una variabile non aleatoria x ; campionamenti ripetuti a) e rappresentazione convenzionale nel caso di un solo campionamento b)

d'errore. Se nella (10.1) $\mu_i(\boldsymbol{\theta}) = f(x_i, \boldsymbol{\theta})$, il χ^2 è dato da:

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_i^2}, \quad (10.4)$$

dove al denominatore compare la varianza di Y che, in base alla (10.3), è data da:

$$\sigma_i^2 \equiv \text{Var}[Y_i] = \text{Var}[Z_i].$$

Se la relazione è lineare, si ha ad esempio $f(x, \boldsymbol{\theta}) = \theta_1 + \theta_2 x$.

La ricerca di una densità a partire da un istogramma, ottenuta minimizzando la (9.59), è riconducibile al caso del predittore non aleatorio. Infatti, la funzione $f(x_i, \boldsymbol{\theta}) = Np_i(\boldsymbol{\theta})$ coincide con il valore atteso di $Y_i = I_i$, le variabili x sono i valori predeterminati dello spettro nel caso di spettro discreto, o il punto medio dello spettro entro il canale predeterminato Δx nel caso di spettro continuo e non vi è errore su x . L'errore è solo su Y ed è quello statistico, dato dalle fluttuazioni nel numero di osservazioni cadute durante il campionamento nei diversi canali dell'istogramma.

Un'altra situazione molto frequente viene dalle misure di grandezze fisiche, una misurata con errore trascurabile e sostanzialmente sotto il controllo dello sperimentatore (la x), l'altra determinata con meno precisione (la Y). In questo caso la funzione $y = f(x, \boldsymbol{\theta})$ rappresenta la legge fisica che lega le due variabili.

Veniamo ora al secondo caso di dipendenza tra variabili:

$$Y = f(X) + Z, \quad f(X) \equiv f(X, \boldsymbol{\theta}), \quad (10.5)$$

dove X e Z sono variabili aleatorie e $\langle Z \rangle = 0$. Poiché a differenza della (10.3), X è ora una variabile aleatoria, tra X ed Y esiste una *correlazione statistica* in generale non nulla. In un campionamento ripetuto con n prove indipendenti, noi osserviamo altrettanti valori del vettore aleatorio (X_i, Y_i) , che indichiamo con le coppie (x_i, y_i) , $i = 1, \dots, n$. Ponendole su un grafico, si ottiene una nuvola di punti come quella di Fig. 10.2 a). La densità congiunta di (X, Y) , la cui forma è quella della (4.3), è rappresentabile come in Fig. 10.2 b), indicando con sfumature di colore diverso i valori degli integrali della densità nelle varie regioni del piano.

Come nella (10.3), assumiamo che l'argomento di f nella relazione funzionale che definisce Y venga osservato esattamente. Se siamo interessati non tanto al meccanismo aleatorio che genera X quanto alla relazione tra X e Y , possiamo ricondurci alla (10.4) fissando un valore x e utilizzando la distribuzione di Y condizionata a $\{X = x\}$ per il calcolo del χ^2 :

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\text{Var}[Y_i|x_i]}, \quad (10.6)$$

dove abbiamo sostituito le σ_i^2 con le varianze condizionate $\text{Var}[Y_i|x_i]$ e il valore atteso delle Y_i con i rispettivi valori attesi condizionati $f(x_i, \boldsymbol{\theta}) = \langle Y_i|x_i \rangle$. Dunque, avendo fissato x_1, \dots, x_n , otteniamo $\hat{\boldsymbol{\theta}}$ minimizzando la (10.6).

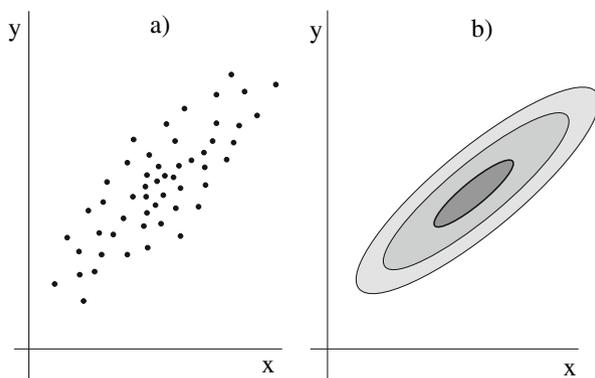


Fig. 10.2. Campionamento di coppie di variabili aleatorie correlate a) e rappresentazione della densità congiunta corrispondente b)

Per come è stato costruito il modello, $\langle Y|x \rangle$ è dato da $f(x)$ qualunque sia la distribuzione di (X, Z) . Al contrario, quest'ultima influenza $\text{Var}[Y|x]$, essendo

$$\text{Var}[Y|x] = \text{Var}[f(X)|x] + \text{Var}[Z|x] = \text{Var}[Z|x].$$

Se X e Z sono indipendenti, allora

$$\text{Var}[Y|x] = \text{Var}[Z|x] = \text{Var}[Z], \quad \text{e} \quad \text{Var}[Y] = \text{Var}[f(X)] + \text{Var}[Z] \quad (10.7)$$

e il denominatore della (10.6) diventa costante, mentre se X e Z non sono indipendenti non si può garantire che $\text{Var}[Z|x] = \text{Var}[Z]$.

Il legame della (10.5) è molto comune ed il caso di Tab. 6.6 è uno di questi. Infatti, il perimetro toracico di un certo soldato sarà correlato non al valore medio della statura, ma *alla statura di quello stesso soldato*. Possiamo anche pensare ad uno studio finanziario che cerchi di correlare l'andamento dell'indice di borsa di Milano con quello della borsa di New York: l'indice di Milano sarà correlato al *valore corrente* della borsa di New York; a questa correlazione si aggiungeranno altre variabili casuali che conterranno le altre fonti di variazione dell'indice di Milano.

Le formule (10.3) e (10.5) sono formalmente le stesse, tuttavia, il significato della funzione f è diverso: nel primo caso x è una variabile non aleatoria e $f(x)$ rappresenta una relazione funzionale analitica tra x e $\langle Y \rangle$, mentre nel secondo caso $f(X)$ induce la correlazione statistica tra di esse. In questo senso, facendo riferimento alla (10.5), possiamo chiamare $f(X)$ *funzione di correlazione* tra X e Y .

Per chiarire meglio il concetto, quando X e Z sono indipendenti, $\text{Var}[Z] = \text{Var}[Z|x]$, e dalla (10.7) abbiamo:

$$\sigma_y^2 \equiv \text{Var}[Y] = \text{Var}[f(X)] + \text{Var}[Z|x]. \quad (10.8)$$

Siccome $\text{Var}[Y|x] = \text{Var}[Z|x]$, la (10.8) può essere riscritta come

$$\text{Var}[Y|x] = \sigma_y^2 \left(1 - \frac{\text{Var}[f(X)]}{\sigma_y^2} \right). \quad (10.9)$$

Confrontiamo questa equazione con la (4.55) di pag. 135, che rappresenta la varianza di Y condizionatamente a $X = x$ nel caso in cui (X, Y) sia gaussiana bivariata e che per comodità riportiamo:

$$\text{Var}[Y|x] = \sigma_y^2(1 - \rho^2).$$

Per analogia tra questa espressione e la (10.9), siamo portati a definire una forma più generale del coefficiente di correlazione tra X e Y :

$$\rho = \pm \frac{\sigma[f(X)]}{\sigma[Y]}, \quad (10.10)$$

dove il segno è quello della covarianza tra X e Y , $\text{Cov}[X, Y]$. Si noti che dalla (10.7) e dalla (10.10) segue che $\rho = \pm 1$ quando $\text{Var}[Z] = 0$ e che $\rho = 0$ quando $f(X)$ è costante. È anche facile mostrare (si veda il problema 10.1) che, se $f(X) = a + bX$, il coefficiente di correlazione è pari alla covarianza tra X e Y diviso il prodotto tra le deviazioni standard:

$$\rho = \pm \frac{\sigma[f(X)]}{\sigma[Y]} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}. \quad (10.11)$$

Riassumendo, se si impone che due variabili abbiano distribuzione congiuntamente gaussiana, allora nel cap. 4 abbiamo mostrato che *la funzione di correlazione tra di esse non può che essere di tipo lineare*. Vale infatti l'equazione (4.54) della retta di regressione per $\langle Y|x \rangle$:

$$f(x) = \langle Y|x \rangle = \mu_y + \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x).$$

Questa situazione qui appare come un caso particolare della (10.5) per il quale la funzione di correlazione $f(X)$ può essere qualunque, ed il coefficiente di correlazione viene definito in generale secondo la (10.10).

10.3 Errori sui predittori

Un caso particolare ma importante che riguarda lo studio di una relazione funzionale tra $\langle Y \rangle$ ed $\langle X \rangle$ si ha quando

$$X = x_0 + X_R, \quad Y = f(x_0) + Z, \quad f(x_0) \equiv f(x_0, \boldsymbol{\theta}), \quad (10.12)$$

dove X_R ed Z sono indipendenti e con valore atteso nullo. La relazione che determina Y è identica alla (10.3), solo che x_0 è osservata *a meno di fluttuazioni statistiche*, che è un modo per dire che il predittore x_0 è ignoto ed è la variabile aleatoria X ad essere osservata.

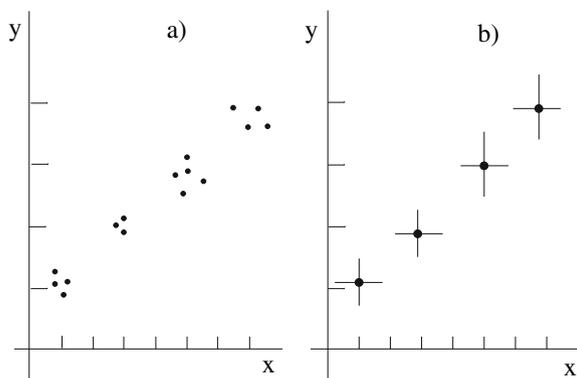


Fig. 10.3. Campionamento di una variabile aleatoria Y funzione del valor medio di un'altra variabile aleatoria X ; campionamenti ripetuti a) e rappresentazione convenzionale nel caso di un solo campionamento b)

Un campionamento ripetuto di (X, Y) , in corrispondenza di n valori veri $\mathbf{x}_0^\dagger = (x_{01}, \dots, x_{0n})$, dà luogo ad un grafico come quello di Fig. 10.3 a), con le x_i e le y_i che fluttuano intorno ai loro valori medi x_{0i} e $f(x_{0i})$; se si effettua una sola prova e le deviazioni standard delle fluttuazioni X_{Ri} ed Z_i sono note, il grafico si disegna come in Fig. 10.3 b), con le deviazioni standard di X_{Ri} e Z_i disegnate come barre d'errore sui valori misurati x_i e y_i .

Questo caso capita spesso nelle misure di laboratorio, quando si misurano con errori rilevanti due grandezze legate da una legge fisica di tipo deterministico. Anche qui il compito dei minimi quadrati è quello di trovare la legge fisica come curva di regressione o *best fit*, depurando il risultato dalle incertezze introdotte dalla misura.

Dalla definizione 10.1 risulta che il χ^2 da minimizzare è in questo caso:

$$\chi^2(\boldsymbol{\theta}, \mathbf{x}_0) = \sum_i \frac{(x_i - x_{0i})^2}{\sigma_{x_i}^2} + \sum_i \frac{[y_i - f(x_{0i}, \boldsymbol{\theta})]^2}{\sigma_{y_i}^2}. \quad (10.13)$$

Infatti, le variabili osservate sono le coppie (x_i, y_i) , mentre i parametri liberi sono ora i p parametri $\boldsymbol{\theta}$ e gli n parametri x_{0i} , che rappresentano i valori medi delle X_i . Il punto cruciale è che, in accordo con la (10.12), f è funzione delle variabili medie x_0 , non di quelle misurate. I parametri liberi da stimare sono $n + p + 1$, cioè la somma delle dimensioni di \mathbf{x}_0 e di $\boldsymbol{\theta}$.

La minimizzazione della (10.13) non presenta difficoltà se si dispone di un programma di minimizzazione *non lineare* del χ^2 , come la nostra routine `Nlinfit`. Tuttavia, il numero dei valori sperimentali raddoppiato e il numero di parametri liberi ($n + p + 1$) possono causare non pochi problemi se si ha a che fare con grandi campioni. Approssimiamo perciò il punto di minimo della (10.13) ponendo $x_{0i} = x_i$ nella prima sommatoria, che così è sempre zero, e cerchiamo di calcolare gli errori σ_{y_i} sviluppando $f(X)$ attorno ad x_0 in serie

di Taylor arrestata al primo ordine:

$$f(X) = f(x_0) + (X - x_0)f'(x_0) + o(X - x_0)^2. \quad (10.14)$$

Se $\text{Var}[X_R]$ è sufficientemente piccola da poter trascurare i termini di ordine superiore al primo (anche se aleatori), possiamo applicare l'operatore varianza alla (10.14) per ottenere

$$\text{Var}[f(X)] \simeq f'^2(x_0)\sigma_x^2.$$

Siccome $f'^2(x_0)$ dipende dal valore ignoto x_0 , lo sostituiamo con il valore osservabile x , per ottenere la cosiddetta *varianza effettiva* σ_E^2 :

$$\text{Var}[Y - f(X)] = \text{Var}[Y] + \text{Var}[f(X)] \simeq \sigma_y^2 + f'^2(x)\sigma_x^2 \equiv \sigma_E^2. \quad (10.15)$$

dove abbiamo anche usato l'ipotesi di indipendenza tra X_R e Z .

Usando perciò i termini $(y_i - f(x_i, \boldsymbol{\theta}))$ al numeratore del χ^2 e la varianza effettiva al denominatore, si ottiene il semplice risultato

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_{Ei}^2} = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_{yi}^2 + f'^2(x_i, \boldsymbol{\theta})\sigma_{xi}^2}, \quad (10.16)$$

dove compaiono solo i valori misurati (x_i, y_i) e le varianze delle loro distribuzioni di provenienza attraverso la varianza effettiva. Nella (10.16) i parametri liberi tornano di nuovo ad essere solo i p parametri $\boldsymbol{\theta}$.

Se i valori di $f'^2(x_i, \boldsymbol{\theta})$ sono ignoti, la minimizzazione della (10.16) non ricade più nella definizione 10.1 e, anche se f è una funzione lineare nei parametri $\boldsymbol{\theta}$, la minimizzazione del χ^2 dà luogo a equazioni non lineari a causa della derivata di f al denominatore. Tuttavia, se si dispone solo di programmi di minimizzazione lineare o si vuole ridurre il tempo di calcolo, si può utilizzare un metodo iterativo, ponendo all'inizio $\sigma_{Ei}^2 = \sigma_{yi}^2$ e nel ciclo k -esimo $\sigma_{Ei}^2 = \sigma_{yi}^2 + f'^2(x_i, \boldsymbol{\theta}^{(k-1)})\sigma_{xi}^2$, utilizzando le stime $\boldsymbol{\theta}^{(k-1)}$ del ciclo precedente. In questo modo il denominatore torna ad essere indipendente dai parametri in corso di minimizzazione nel ciclo corrente e si ripristina la linearità [Ore82]. Si può mostrare però che questa procedura produce stimatori non consistenti (anche se più vicini al valore vero del parametro rispetto ai metodi che ignorano gli errori su X) [Lyb84], per cui, se si dispone di un codice di minimizzazione non lineare, conviene senz'altro usare la (10.16), che fornisce risultati praticamente coincidenti con la formula generale (10.13).

La varianza effettiva si presta ad una intuitiva interpretazione geometrica, mostrata in Fig. 10.4: il termine $f'^2\sigma_x^2$, che viene aggiunto alla varianza iniziale σ_y^2 , non è altro che l'effetto delle fluttuazioni in x riportato, attraverso la tangente dell'angolo α (cioè la derivata $f'(x)$), sull'asse delle y .

Se le variabili X_i e Y_i avessero covarianza $\text{Cov}[X_i, Y_i]$ non nulla, allora, tramite la (5.65) si può ulteriormente generalizzare la (10.16) come segue:

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_{yi}^2 + f'^2(x_i, \boldsymbol{\theta})\sigma_{xi}^2 - 2f'(x_i, \boldsymbol{\theta})\sigma_{x_i y_i}}. \quad (10.17)$$

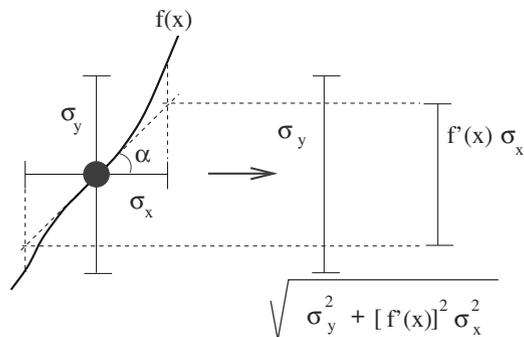


Fig. 10.4. Interpretazione geometrica della varianza effettiva

10.4 Rette dei minimi quadrati

In questo paragrafo esaminiamo un caso particolarmente semplice della relazione (10.3), alla quale riconduciamo anche la (10.5), considerando X e Z indipendenti e condizionandoci ai valori osservati di X . Perciò scriveremo sempre la x in minuscolo.

Quando la funzione che lega $\langle Y|x \rangle$ a x è una retta, cioè $f(x, \theta) = a + bx$, con $\theta = (a, b)$, la relazione tra x e Y diviene

$$Y = a + bx + Z, \quad (10.18)$$

dove assumiamo anche che $\text{Var}[Y] = \text{Var}[Z] = \text{Var}[Z|x] = \sigma_z^2$ indipendentemente dal valore di x . La varianza σ_z^2 raramente è nota. Avendo osservato il campione casuale $(x_1, y_1), \dots, (x_n, y_n)$, il χ^2 da minimizzare si può scrivere come:

$$\chi^2(a, b) = \frac{1}{\sigma_z^2} \sum_i (y_i - a - bx_i)^2. \quad (10.19)$$

La minimizzazione della (10.19) non dipende da σ_z^2 e richiede l'annullamento delle derivate parziali:

$$\begin{aligned} \frac{\partial \chi^2}{\partial a} &= -2 \left[\sum_i y_i - n a - b \sum_i x_i \right] = 0, \\ \frac{\partial \chi^2}{\partial b} &= -2 \left[\sum_i x_i y_i - a \sum_i x_i - b \sum_i x_i^2 \right] = 0. \end{aligned} \quad (10.20)$$

Per semplificare la notazione, poniamo

$$S_x = \sum_{i=1}^n x_i, \quad S_y = \sum_{i=1}^n y_i, \quad S_{xx} = \sum_{i=1}^n x_i^2, \quad S_{xy} = \sum_{i=1}^n x_i y_i. \quad (10.21)$$

Usando questa convenzione, le (10.20) danno allora luogo al sistema di equazioni:

$$a n + b S_x = S_y \quad (10.22)$$

$$a S_x + b S_{xx} = S_{xy} \quad (10.23)$$

ovvero, in forma matriciale

$$\begin{pmatrix} n & S_x \\ S_x & S_{xx} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix} \quad (10.24)$$

dove i parametri a e b rappresentano le incognite le medie campionarie sono coefficienti numerici noti. Con i metodi dell'analisi matriciale si ottiene subito la stima dei minimi quadrati dei parametri:

$$\hat{a} = \frac{1}{D} [S_{xx} S_y - S_x S_{xy}] \quad (10.25)$$

$$\hat{b} = \frac{1}{D} [n S_{xy} - S_x S_y] \quad (10.26)$$

dove

$$D = n S_{xx} - S_x^2 \quad (10.27)$$

rappresenta il determinante del sistema (10.24). I parametri ottenuti sono soprasssegnati dal simbolo “^” per indicare che si tratta di *valori stimati dai dati*, non dei valori veri a e b .

Utilizzando la relazione (4.23), dopo un po' di algebra, la (10.26) può essere scritta come:

$$\hat{b} = \frac{S_{xy}}{S_x^2} = \frac{S_{xy}}{S_x S_y} \frac{S_y}{S_x} = r \frac{S_y}{S_x} \quad (10.28)$$

dove r è la stima sperimentale (6.140) del coefficiente di correlazione lineare ρ della (10.11). Dalla (10.22) ricaviamo $\hat{a} = \langle y \rangle - \hat{b} \langle x \rangle$. Con i coefficienti così riespressi, la retta dei minimi quadrati stimata, $\hat{y}(x) = \hat{a} + \hat{b}x$, risulta essere una stima della retta di regressione (4.54), che abbiamo riportato alla fine del paragrafo precedente:

$$\hat{y}(x) = \langle y \rangle + r \frac{S_y}{S_x} (x - \langle x \rangle) \quad (10.29)$$

Dunque concludiamo che *la retta di regressione rappresenta il luogo dei punti della media della distribuzione condizionata per variabili gaussiane oppure la retta dei minimi quadrati per variabili anche non gaussiane.*

L'errore sulle stime dei parametri, che verrà indicato come al solito col simbolo s , si ottiene applicando la legge (5.65) sulla varianza di funzioni di variabili aleatorie, la quale permette di valutare l'effetto di propagazione delle fluttuazioni delle variabili Y (cioè della parte casuale) sui parametri \hat{a} e \hat{b} attraverso le (10.25, 10.26):

$$\sigma_f^2 = \sigma_z^2 \sum_i \left(\frac{\partial f}{\partial y_i} \right)^2 \quad (10.30)$$

La covarianza $\text{Cov}[Y_i, Y_j]$ non compare perché si suppone che le osservazioni Y_i siano tra loro indipendenti.

Tenendo presente che

$$\frac{\partial S_y}{\partial y_i} = 1 \quad \text{e} \quad \frac{\partial S_{xy}}{\partial y_i} = x_i ,$$

si ottiene:

$$\frac{\partial \hat{a}}{\partial y_i} = \frac{1}{D} [S_{xx} - S_x x_i] , \quad \frac{\partial \hat{b}}{\partial y_i} = \frac{1}{D} [n x_i - S_x] . \quad (10.31)$$

Applicando la (10.30) e tenendo presente la (10.27) è ora facile vedere che:

$$\text{Var}[\hat{a}] = \frac{\sigma_z^2}{D^2} \sum_i ([S_{xx} - S_x x_i]^2) = \sigma_z^2 \frac{S_{xx}}{D} , \quad (10.32)$$

$$\text{Var}[\hat{b}] = \frac{\sigma_z^2}{D^2} \sum_i ([n x_i - S_x]^2) = \sigma_z^2 \frac{n}{D} . \quad (10.33)$$

Queste sono espressioni esatte delle varianze degli stimatori, perché \hat{a} e \hat{b} dipendono linearmente dalle y_i . Ci domandiamo se questi errori si comportano come $1/\sqrt{n}$, cioè come l'errore della stima della media di un campione casuale. Indicata con \bar{s}_x^2 la varianza campionaria di (x_1, \dots, x_n) con denominatore n anziché $n-1$:

$$\bar{s}_x^2 = \frac{S_{xx}}{n} - \frac{S_x^2}{n^2} , \quad (10.34)$$

con qualche passaggio algebrico si trova che è effettivamente così:

$$\begin{aligned} \text{Var}[\hat{a}] &= \frac{\sigma_z^2}{n} \left(1 + \frac{\langle x \rangle^2}{\bar{s}_x^2} \right) \\ \text{Var}[\hat{b}] &= \frac{\sigma_z^2}{n} \frac{1}{\bar{s}_x^2} . \end{aligned}$$

Se (x_1, \dots, x_n) si comporta come un campione casuale, \bar{s}_x^2 e $\langle x \rangle^2$ convergono quasi certamente a valori costanti e dunque gli errori convergono a zero con velocità $1/\sqrt{n}$. Abbiamo già avuto modo di notare più volte come questo sia un comportamento comune a moltissimi stimatori statistici.

Per quanto riguarda il calcolo degli errori, questo non è possibile se σ_z^2 è incognita. Per stimare σ_z^2 , riscriviamo la (10.18): $Z = Y - (a + bx)$. Dunque, essendo $\sigma_z^2 = \text{Var}[Z]$, una sua stima è ricavabile dalla varianza campionaria dei residui dalla retta dei minimi quadrati,

$$\hat{z}_i = y_i - (\hat{a} + \hat{b}x_i) = y_i - \hat{y}_i . \quad (10.35)$$

La media dei residui è zero: infatti abbiamo $\langle \hat{z} \rangle = \langle y \rangle - \hat{a} - \hat{b} \langle x \rangle$. Ma la (10.29) mostra che la retta dei minimi quadrati passa per il punto $(\langle x \rangle, \langle y \rangle)$, dunque

$\langle y \rangle = \hat{a} + \hat{b} \langle x \rangle$ e $\langle \hat{z} \rangle = 0$. Allora la varianza campionaria dei residui è

$$s_z^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{z}_i^2 = \frac{1}{n-2} \sum_{i=1}^n [y_i - \hat{a} - \hat{b}x_i]^2 \quad (10.36)$$

dove il denominatore $(n-2)$ fa sì che lo stimatore della varianza sia non distorto. Le stime della varianza di \hat{a} e \hat{b} sono perciò

$$s^2(\hat{a}) = \frac{s_z^2}{n} \left(1 + \frac{\langle x \rangle^2}{\bar{s}_x^2} \right) \quad (10.37)$$

$$s^2(\hat{b}) = \frac{s_z^2}{n} \frac{1}{\bar{s}_x^2}. \quad (10.38)$$

Ora abbiamo gli elementi per calcolare degli intervalli di confidenza per i parametri della retta di regressione, *quando le Y_i sono assunte gaussiane indipendenti con media $(a + bx_i)$ e varianza σ_z^2* . Partiamo dal fatto che se sapessimo che $b = 0$, il problema dei minimi quadrati coinciderebbe con quello della stima della media del campione (y_1, \dots, y_n) . Questo è stato risolto nel par. 6.11, mostrando che M (cioè \hat{a}) e S^2/σ^2 sono indipendenti e hanno rispettivamente distribuzione gaussiana e χ^2 ridotto con $n-1$ gradi di libertà. Perciò $\sqrt{n}(M - \mu)/S$ aveva distribuzione t di Student con $n-1$ gradi di libertà. Un simile risultato continua a valere quando dobbiamo stimare anche b : infatti sia \hat{a} sia \hat{b} sono combinazioni lineari di variabili gaussiane indipendenti, mentre $s^2(\hat{a})/\sigma^2(\hat{a})$ e $s^2(\hat{b})/\sigma^2(\hat{b})$ hanno distribuzione χ^2 ridotto con $n-2$ gradi di libertà. Allora, si dimostra che $(\hat{a} - a)/s(\hat{a})$ e $(\hat{b} - b)/s(\hat{b})$ hanno entrambe distribuzione t di Student con $n-2$ gradi di libertà e dunque, indicando con $t_{1-\alpha/2}$ i quantili di questa distribuzione, gli intervalli di confidenza per a e b con $CL = 1 - \alpha$ sono:

$$\hat{a} \pm t_{1-\alpha/2} s(\hat{a}), \quad \hat{b} \pm t_{1-\alpha/2} s(\hat{b}). \quad (10.39)$$

L'unico modo non tedioso di dimostrare per bene questi risultati è quello che fa uso del teorema di Cochran e delle proprietà dei proiettori ortogonali su spazi lineari. Rimandiamo gli interessati a [Bal98].

Una volta stimato il modello, questo può essere utilizzato per valutare $\langle Y|x \rangle$ in corrispondenza di un punto x (anche non incluso tra i punti sperimentali) (x_1, \dots, x_n) , oppure per prevedere il valore che osserveremmo per Y in corrispondenza di x .

Per prevedere $\langle Y|x \rangle$ è naturale utilizzare la stima $\hat{y}(x) = \hat{a} + \hat{b}(x)$. L'errore ad essa associato dipende dalla covarianza tra \hat{a} e \hat{b} . Infatti applicando la (5.65) e le (10.32, 10.33) otteniamo la stima:

$$s^2(\hat{y}(x)) = \left(\frac{\partial \hat{y}}{\partial \hat{a}} \right)^2 s^2(\hat{a}) + \left(\frac{\partial \hat{y}}{\partial \hat{b}} \right)^2 s^2(\hat{b}) + 2 \frac{\partial \hat{y}}{\partial \hat{a}} \frac{\partial \hat{y}}{\partial \hat{b}} s(\hat{a}, \hat{b}) \quad (10.40)$$

dove compare appunto la stima della covarianza $s(\hat{a}, \hat{b})$. Con la solita notazione $\Delta X = X - \langle X \rangle$, applicando la (5.82) e sviluppando al primo ordine \hat{a} e \hat{b} nelle variabili y_i , possiamo calcolare la covarianza $\text{Cov}[\hat{a}, \hat{b}]$:

$$\text{Cov}[\hat{a}, \hat{b}] = \left\langle \Delta \hat{a} \Delta \hat{b} \right\rangle = \sum_i \frac{\partial \hat{a}}{\partial y_i} \frac{\partial \hat{b}}{\partial y_i} \langle (\Delta Y_i)^2 \rangle = \sum_i \frac{\partial \hat{a}}{\partial y_i} \frac{\partial \hat{b}}{\partial y_i} \text{Var}[Y_i] ,$$

dove alle derivate non si applica il valore atteso perché sono costanti nei valori misurati y_i . Con l'aiuto delle (10.31) e con la notazione (10.21) si ottiene in definitiva:

$$\begin{aligned} \text{Cov}[\hat{a}, \hat{b}] &= \frac{\sigma_z^2}{D^2} \sum_i (S_{xx} - S_x x_i) (n x_i - S_x) \\ &= \frac{\sigma_z^2}{D^2} [n S_{xx} S_x - n S_{xx} S_x - n S_x S_{xx} + S_x^2 S_x] \\ &= -\sigma_z^2 S_x \frac{[n S_{xx} - S_x^2]}{D^2} = -\sigma_z^2 \frac{S_x}{D} \end{aligned} \quad (10.41)$$

da cui, con la (10.34), la stima

$$s(\hat{a}, \hat{b}) = -s_z^2 \frac{S_x}{D} = -\frac{s_z^2}{n} \frac{\langle x \rangle}{\bar{s}_x^2} .$$

Sostituendo questa stima e le (10.37, 10.38) nella (10.40) troviamo infine:

$$s^2(\hat{y}) = \frac{s_z^2}{n} \left[1 + \frac{(x - \langle x \rangle)^2}{\bar{s}_x^2} \right] . \quad (10.42)$$

Poiché $\hat{y}(x)$ è combinazione lineare di gaussiane indipendenti e $s^2(\hat{y})$ è un multiplo di s^2 , è di nuovo possibile dimostrare che

$$\frac{\hat{y}(x) - (a + bx)}{s(\hat{y}(x))}$$

è distribuita come una variabile di Student con $n - 2$ gradi di libertà e quindi un intervallo di confidenza con un CL = $1 - \alpha$ per $y(x) = a + bx$ è dato da

$$y(x) \in \hat{y}(x) \pm t_{1-\alpha/2} s(\hat{y}(x)) . \quad (10.43)$$

La (10.42) è piuttosto istruttiva, perché segnala il pericolo di fare estrapolazioni: l'errore di previsione infatti cresce all'incirca come $(x - \langle x \rangle)$ man mano che x si allontana dal baricentro dei punti sperimentali, rappresentato da $\langle x \rangle$. E questo assumendo che al di fuori dello spettro sperimentale la relazione tra x e $\langle Y|x \rangle$ continui a essere lineare, cosa che non possiamo affermare o smentire basandoci solo sul campione osservato. Se non lo è, la (10.42) non è neppure più una stima appropriata della varianza. Di questa osservazione si trova conferma disegnando gli intervalli (10.43) al variare di x , ed utilizzando

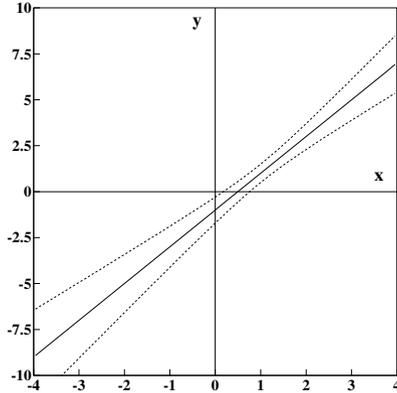


Fig. 10.5. Fascia di confidenza $\hat{y} \pm t_{1-\alpha/2} s(\hat{y})$ della retta dei minimi quadrati. $s(\hat{y})$ è stimato con la (10.42) per $\langle x \rangle = \langle y \rangle = 1$, $s_x = 1$ e $s/n = 0.5$

come ordinata $\hat{y}(x)$. Il risultato è una fascia attorno alla retta dei minimi quadrati come mostrato in Fig. 10.5. Ricordate però che questa *non* è un insieme di confidenza simultaneo con $CL = 1 - \alpha$, ma solo l'unione di tanti intervalli univariati.

Una nuova osservazione Y per un dato valore $X = x$ ha sempre valore atteso $a + bx$; useremo perciò $\hat{y}(x)$ come stimatore di Y . Tuttavia, nella previsione, si deve tener conto anche della varianza della parte casuale di Y , oltre a quella di $\hat{y}(x)$. Pertanto, l'errore associato alla stima è questa volta dato da

$$\text{Var} [\hat{y}(x) + Z|x] = \text{Var} [\hat{y}(x)] + \text{Var} [Z|x] = \text{Var} [\hat{y}(x)] + \sigma_z^2 \quad (10.44)$$

dove abbiamo potuto sommare le varianze perché la nuova parte casuale di Y è indipendente dalle fluttuazioni da cui dipendono gli stimatori \hat{a} e \hat{b} .

Sostituendo nella (10.44) lo stimatore ottenuto nella (10.42) e inserendo s_z^2 al posto di σ_z^2 abbiamo che l'errore è

$$s^2(\hat{a} + \hat{b}x + z) = s_z^2 + s^2(\hat{y}) = s_z^2 \left[1 + \frac{1}{n} \left(1 + \frac{(x - \langle x \rangle)^2}{\bar{s}_x^2} \right) \right]. \quad (10.45)$$

In definitiva, l'intervallo di stima di Y in corrispondenza di un x fissato e per un certo livello di confidenza $1 - \alpha$, diventa:

$$Y(x) \in \hat{a} + \hat{b}x \pm t_{1-\alpha/2} s_z \sqrt{1 + \frac{1}{n} \left(1 + \frac{(x - \langle x \rangle)^2}{\bar{s}_x^2} \right)}, \quad (10.46)$$

dove, come per la stima di $a + bx$, $t_{1-\alpha/2}$ è il quantile della distribuzione di Student con $n - 2$ gradi di libertà.

Esercizio 10.1

Riprendere l'esercizio 6.17 di pagina 254 e ricalcolare l'intervallo di stima per il perimetro toracico di un soldato alto 170 cm.

Risposta. La soluzione trovata a suo tempo non teneva conto dell'incertezza dovuta al fatto di aver usato medie, varianze e coefficiente di correlazione campionari invece di quelli veri. Se ora ricordiamo il risultato dell'esercizio 6.17,

$$t \in m(t|s) \pm s(t|s) = 86.9 \pm 4.3 \text{ cm}$$

e facciamo uso della (10.46), possiamo tenere conto di questo fatto ed eseguire la stima corretta:

$$\begin{aligned} t &\in 86.9 \pm 4.3 \sqrt{1 + \frac{1}{1665} \left(1 + \frac{(170 - 163.7)^2}{5.79^2}\right)} \\ &= 86.9 \pm 4.3 \cdot 1.0006 \simeq 86.9 \pm 4.3 . \end{aligned}$$

Il risultato, anche dopo la correzione, resta praticamente invariato. I livelli di confidenza si calcolano con la densità di Student con $(1665 - 2)$ gradi di libertà e si possono considerare a tutti gli effetti gaussiani. Pertanto, il 68% dei soldati alti 170 cm ha il perimetro toracico compreso tra 82.6 e 91.2 cm. L'intervallo di confidenza (10.42) che contiene la media vera (in senso frequentista) con $CL \simeq 68\%$ per $x = 170$ cm, vale invece:

$$86.9 \pm 0.1 \text{ cm} .$$

10.5 Minimi quadrati lineari: caso generale

In questo paragrafo generalizziamo al caso multidimensionale la discussione fatta nel par. 10.4.

Un problema più generale di minimizzazione lineare si ha quando abbiamo p predittori raccolti in un vettore \mathbf{x} e l'equazione corrispondente alla (10.18) è espressa da

$$Y = \mu(\mathbf{x}, \boldsymbol{\theta}) + Z = f(\mathbf{x}, \boldsymbol{\theta}) + Z = \theta_0 + \sum_{j=1}^p \theta_j x_j + Z , \quad (10.47)$$

dove x_j è l'elemento j -esimo di \mathbf{x} . Come nel par. 10.4, poniamo $\langle Z|\mathbf{x} \rangle = 0$ e $\text{Var}[Z|\mathbf{x}] = \sigma_z^2$. Perciò

$$\langle Y|\mathbf{x} \rangle = \theta_0 + \sum_{j=1}^p \theta_j x_j , \quad \text{Var}[Y|\mathbf{x}] = \sigma_z^2 .$$

Le componenti di \mathbf{x} possono essere variabili differenti o funzioni della stessa variabile, o un insieme delle due cose, secondo la convenienza. Per esempio potremmo avere che la risposta dipende da un solo predittore tramite un polinomio di grado p

$$f(\mathbf{x}, \boldsymbol{\theta}) = \theta_0 + \sum_{j=1}^p \theta_j x^j, \tag{10.48}$$

con $\mathbf{x}^\dagger = (x, x^2, \dots, x^p)$, oppure, più in generale:

$$f(\mathbf{x}, \boldsymbol{\theta}) = \theta_0 + \sum_{j=1}^p \theta_j f_j(x), \tag{10.49}$$

con $\mathbf{x}^\dagger = (f_1(x), f_2(x), \dots, f_p(x))$. Se $p = 1$ nella (10.48), ricadiamo nel caso particolare della retta dei minimi quadrati. Il problema di partenza è ancora quello di stimare $\boldsymbol{\theta}$ a partire da un campione casuale $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$, con $p + 1 < n$.

Avendo a che fare con un problema multidimensionale sarà conveniente passare alla notazione matriciale. Per facilitarvi la lettura, vi ricordiamo brevemente alcune regole del calcolo matriciale con due generiche matrici A e B . Indicando con \dagger la trasposizione, ovvero lo scambio delle righe con le colonne, valgono le proprietà

$$(AB)^\dagger = B^\dagger A^\dagger, \quad (A + B)^\dagger = A^\dagger + B^\dagger \quad (A^\dagger)^\dagger = A \tag{10.50}$$

se le dimensioni delle matrici sono compatibili.

Se A è una matrice quadrata, la matrice inversa è quella matrice A^{-1} tale che $AA^{-1} = I$, dove I è la matrice identità. Se esiste la matrice inversa, si dice che A ha rango pari al numero delle sue righe, o anche che ha “rango pieno”. Se B è una matrice quadrata a sua volta invertibile, valgono inoltre le proprietà:

$$(AB)^{-1} = B^{-1}A^{-1}, \quad A^{-1}A = I, \quad (A^{-1})^{-1} = A. \tag{10.51}$$

Se A è una matrice con m righe e k colonne, con $k \leq m$ e di rango k , la matrice $k \times k$

$$A^\dagger A = (A^\dagger A)^\dagger \tag{10.52}$$

è una matrice simmetrica e definita positiva. Cioè, dato un qualunque vettore \mathbf{c} non nullo e di dimensioni compatibili: $\mathbf{c}^\dagger A^\dagger A \mathbf{c} > 0$.

Oltre a ricordare queste formule, vale la pena di rileggere il par. 4.5.

Torniamo ora al nostro problema. Indicando con x_{ij} il valore del j -esimo elemento di \mathbf{x}_i , definiamo la matrice dei predittori:

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}. \tag{10.53}$$

Indichiamo poi con \mathbf{y} il vettore colonna delle risposte (y_1, \dots, y_n) e con $\boldsymbol{\theta}$ il vettore colonna dei parametri $(\theta_0, \theta_1, \dots, \theta_p)$. La colonna di uni nella matrice X serve a tener conto della parte costante del modello, che è bene includere sempre, a meno che non si sia certi che $\theta_0 = 0$. In questo modo possiamo esprimere la (10.47) per tutti i punti sperimentali con una sola equazione matriciale:

$$\mathbf{Y} = X \boldsymbol{\theta} + \mathbf{Z} \quad (10.54)$$

dove \mathbf{Z} è il vettore colonna (Z_1, \dots, Z_n) delle fluttuazioni casuali a media nulla, incorrelate, con varianze marginali costanti: $\text{Var}[Z_i] = \sigma_z^2$.

Il χ^2 da minimizzare per stimare $\boldsymbol{\theta}$, che ricaviamo dalla (10.1), è

$$\chi^2(\boldsymbol{\theta}) = \frac{1}{\sigma_z^2} \sum_{i=1}^n [y_i - \theta_0 - \sum_{j=1}^p \theta_j x_{ij}]^2, \quad (10.55)$$

e il suo punto di minimo dipende solo dal numeratore, che possiamo riscrivere in forma matriciale

$$\sigma_z^2 \chi^2 = (X\boldsymbol{\theta} - \mathbf{y})^\dagger (X\boldsymbol{\theta} - \mathbf{y}). \quad (10.56)$$

È facile vedere che la condizione di minimo della (10.55)

$$\frac{\partial \chi^2}{\partial \theta_k} = 0, \quad k = 0, \dots, p$$

implica le cosiddette *equazioni normali*:

$$\sum_{i=1}^n \left[\left(y_i - \theta_0 - \sum_{j=1}^p \theta_j x_{ij} \right) x_{ik} \right] = 0, \quad k = 0, 1, \dots, p, \quad (10.57)$$

avendo l'accortezza di porre $x_{i0} = 1$ per l'equazione con $k = 0$. La condizione di minimo espressa dalle equazioni normali (10.57) assume ora la forma compatta:

$$X^\dagger \mathbf{y} = (X^\dagger X) \boldsymbol{\theta} \quad \text{oppure} \quad \boldsymbol{\beta} = \boldsymbol{\alpha} \boldsymbol{\theta}, \quad (10.58)$$

dove abbiamo introdotto le matrici $\boldsymbol{\beta} = X^\dagger \mathbf{y}$ e $\boldsymbol{\alpha} = X^\dagger X$, secondo una notazione abbastanza diffusa [BR92, PFTW92]. Questa equazione, risolta nelle incognite $\boldsymbol{\theta}$, invertendo $X^\dagger X$, che assumiamo di rango $p + 1$, fornisce le stime LS dei parametri e rappresenta finalmente la soluzione del problema dei minimi quadrati lineari con più predittori (o dei minimi quadrati generali o della regressione multipla):

$$\hat{\boldsymbol{\theta}} = (X^\dagger X)^{-1} X^\dagger \mathbf{y} \equiv \boldsymbol{\alpha}^{-1} \boldsymbol{\beta}. \quad (10.59)$$

Potete vedere queste equazioni fondamentali codificate in SCILAB con la nostra routine `Linfit`.

Nel caso $\langle Y|x \rangle = a + bx$ è facile vedere che le equazioni normali (10.58) corrispondono proprio alle (10.24).

Veniamo ora allo studio degli errori sulle stime ottenute, usando il formalismo matriciale e tenendo presente che la (10.59) non è che un caso particolare della (5.74). Infatti possiamo scrivere la (10.59) come

$$\hat{\theta}_j = f_j(y_1, \dots, y_n), \quad j = 0, \dots, p,$$

dove ogni funzione f_j è una combinazione lineare degli elementi di \mathbf{y} , e applicare quindi la trasformazione delle varianze (5.76). In questa trasformazione, l'elemento di posizione ji della matrice del trasporto, dato da $\partial f_j / \partial y_i$, non è altro che l'elemento di posto ji della matrice $(X^\dagger X)^{-1} X^\dagger$. Applicando questa legge alla (10.59) si ottiene dunque la matrice delle covarianze di $\hat{\boldsymbol{\theta}}$:

$$\begin{aligned} V(\hat{\boldsymbol{\theta}}) &= (X^\dagger X)^{-1} X^\dagger V(\mathbf{Y}) X [(X^\dagger X)^{-1}]^\dagger \\ &= (X^\dagger X)^{-1} X^\dagger \sigma_z^2 I X [(X^\dagger X)^{-1}]^\dagger \\ &= \sigma_z^2 [(X^\dagger X)^{-1}]^\dagger = \sigma_z^2 (X^\dagger X)^{-1}, \end{aligned} \tag{10.60}$$

dove si sono utilizzate le (10.50, 10.51), la simmetria di $(X^\dagger X)$ e il fatto che $V(\mathbf{Y}) = \sigma_z^2 I$, cioè la matrice diagonale con elementi diagonali tutti pari a σ_z^2 .

Questa importante relazione mostra che *l'inverso della matrice $\boldsymbol{\alpha} = X^\dagger X$ contiene tutte le informazioni sugli errori nelle stime dei parametri*. Infatti in questa matrice, detta *matrice degli errori*, è una matrice quadrata di dimensione $(p+1) \times (p+1)$, data dal numero di parametri liberi del *fit*, ed è simmetrica e definita positiva. Gli elementi diagonali sono le varianze delle stime LS dei parametri $\hat{\boldsymbol{\theta}}$, gli elementi non diagonali rappresentano le covarianze tra tutte le coppie di stimatori $(\hat{\theta}_i, \hat{\theta}_j)$.

Un'utile verifica di questa affermazione si può fare applicando la (10.60) al caso bidimensionale (retta dei minimi quadrati); si ottiene infatti:

$$V(\hat{a}, \hat{b}) = \frac{\sigma_z^2}{nS_{xx} - S_x^2} \begin{pmatrix} S_{xx} & -S_x \\ -S_x & n \end{pmatrix},$$

in accordo con le (10.24, 10.32, 10.33, 10.41).

La valutazione di $\langle Y | \mathbf{x} \rangle$ in corrispondenza di un punto \mathbf{x} , come nel par. 10.4, avviene tramite la trasformazione lineare $\hat{y}(\mathbf{x}) = \mathbf{x}^\dagger \hat{\boldsymbol{\theta}}$, dove possiamo pensare a \mathbf{x}^\dagger come a una riga della matrice X o a un nuovo insieme di valori dei predittori. Applicando ancora la (5.74) e la (5.76) a questo nuovo caso particolare di trasformazione di variabili, si ottiene immediatamente la varianza di $\hat{y}(\mathbf{x})$:

$$\text{Var}[\hat{y}(\mathbf{x})] = \mathbf{x}^\dagger V(\hat{\boldsymbol{\theta}}) \mathbf{x}, \tag{10.61}$$

che rappresenta la generalizzazione della (10.40).

La stima corretta dell'errore σ_z , analogamente alla (10.36) si ricava da:

$$s_z^2 = \frac{(X \hat{\boldsymbol{\theta}} - \mathbf{y})^\dagger (X \hat{\boldsymbol{\theta}} - \mathbf{y})}{n - p - 1}. \tag{10.62}$$

Osservate che i gradi di libertà al denominatore sono sempre dati dal numero n di valori osservati meno il numero di parametri stimati.

Nel caso in cui le Y_i siano gaussiane indipendenti possiamo ricavare gli intervalli di confidenza per ogni θ_j analoghi a quelli che compaiono nelle (10.39):

$$\theta_j \in \hat{\theta}_j \pm t_{1-\alpha/2} s(\hat{\theta}_j), \quad (10.63)$$

dove $s(\hat{\theta}_j)$ è ricavabile dalla (10.60), sostituendo s_z a σ_z :

$$s(\hat{\theta}_j) = s_z \sqrt{(X^\dagger X)_{jj}^{-1}}. \quad (10.64)$$

Questa formula è utilizzata nella nostra routine `Linfit`.

L'intervallo di confidenza di livello $1-\alpha$ per $y(\mathbf{x}) = \langle Y | \mathbf{x} \rangle$ assume la forma (analoga alla (10.43))

$$y(\mathbf{x}) \in \hat{y}(\mathbf{x}) \pm t_{1-\alpha/2} s_z \sqrt{\mathbf{x}^\dagger (X^\dagger X)^{-1} \mathbf{x}} \quad (10.65)$$

dove $t_{1-\alpha/2}$ è il percentile della variabile di Student con $n-p-1$ gradi di libertà. Per quanto riguarda l'intervallo di confidenza per Y (simile alla (10.46)) in corrispondenza di un dato valore \mathbf{x} , esso è identico a quello della (10.65) dopo aver sommato 1 al termine sotto radice quadrata:

$$Y(\mathbf{x}) \in \hat{y}(\mathbf{x}) \pm t_{1-\alpha/2} s_z \sqrt{1 + \mathbf{x}^\dagger (X^\dagger X)^{-1} \mathbf{x}}. \quad (10.66)$$

Spesso la procedura di *fit* è resa complicata e di difficile interpretazione dalle correlazioni tra le stime LS dei parametri, i cui valori cambiano da un *fit* all'altro se decidiamo di aggiungere altri predittori a quelli usati nel primo *fit*, aumentando p nella (10.47). Per ottenere stime incorrelate si deve diagonalizzare la matrice $(X^\dagger X)^{-1}$ ricorrendo a sofisticate tecniche di calcolo matriciale oppure ortogonalizzando la matrice X in modo che soddisfi la condizione

$$\sum_i x_{ik} x_{il} = 0 \quad \text{se } k \neq l. \quad (10.67)$$

In questo modo $X^\dagger X$ è diagonale e anche la matrice degli errori, che ne è l'inversa, risulta diagonale, le covarianze sono tutte nulle e i parametri risultano tra loro incorrelati. Anche se importanti nella pratica, non discuteremo qui i metodi di diagonalizzazione, che sono piuttosto laboriosi e che si possono trovare sui testi dedicati alle tecniche di calcolo numerico, come [PFTW92].

Esercizio 10.2

Ricavare le equazioni normali nel caso di una relazione quadratica

$$Y = f(X; a, b, c) = a + bX + cX^2.$$

Risposta. Con la notazione (10.21) la (10.58) diventa:

$$\begin{pmatrix} S_1 & S_x & S_{x^2} \\ S_x & S_{x^2} & S_{x^3} \\ S_{x^2} & S_{x^3} & S_{x^4} \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \\ S_{x^2y} \end{pmatrix}.$$

Il confronto di queste equazioni con le (10.24) vi suggerisce immediatamente la generalizzazione delle equazioni normali ad un polinomio di qualunque grado.

10.6 Minimi quadrati lineari pesati

Nei paragrafi 10.4 e 10.5 abbiamo considerato le diverse osservazioni della variabile risposta come incorrelate e con varianza costante. Una prima deviazione da questa ipotesi è che le varianze non siano costanti, cioè che

$$V \equiv V(\mathbf{Y}) = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}, \quad (10.68)$$

ovvero $\text{Var}[Y_i] = \sigma_i^2$ per ogni i e $\text{Cov}[Y_i, Y_j] = 0$, per $i \neq j$. Più in generale, anche le covarianze potrebbero essere non nulle, con una matrice V non diagonale.

Se tutti gli elementi di V fossero ignoti, si porrebbe un difficile problema di stima. Per il caso della (10.68), assumiamo di conoscere i rapporti σ_i/σ_j , cioè quanto la risposta della prova i è più (o meno) variabile rispetto a quella della prova j . Allora potremo scrivere la matrice delle covarianze come $V = \sigma_z^2 W^{-1}$, dove

$$W^{-1} = \begin{pmatrix} \frac{1}{w_1} & 0 & \dots & 0 \\ 0 & \frac{1}{w_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{w_n} \end{pmatrix}. \quad (10.69)$$

I pesi (w_1, \dots, w_n) sono tutti noti: in tal modo $\sigma_i^2 = \sigma_z^2/w_i$ e anche i rapporti σ_i/σ_j sono tutti noti e pari a w_j/w_i .

Se vale la (10.69), allora il χ^2 da minimizzare, che scriviamo direttamente per il problema della regressione multipla con $\mathbf{x}_i^\dagger = (1, x_{i1}, \dots, x_{ip})$, è

$$\begin{aligned} \chi^2(\boldsymbol{\theta}) &= \sum_{i=1}^n \frac{[y_i - \mathbf{x}_i^\dagger \boldsymbol{\theta}]^2}{\sigma_i^2} = \frac{1}{\sigma_z^2} \sum_{i=1}^n \left[\sqrt{w_i} y_i - \sqrt{w_i} \mathbf{x}_i^\dagger \boldsymbol{\theta} \right]^2 \\ &= \frac{1}{\sigma_z^2} \sum_{i=1}^n \left[\tilde{y}_i - \tilde{\mathbf{x}}_i^\dagger \boldsymbol{\theta} \right]^2, \end{aligned} \quad (10.70)$$

dove abbiamo posto $\tilde{y}_i = \sqrt{w_i} y_i$ e $\tilde{\mathbf{x}}_i = \sqrt{w_i} \mathbf{x}_i$. Passando alla notazione matriciale, indichiamo con $W^{\frac{1}{2}}$ la matrice diagonale con le radici quadrate dei pesi sulla diagonale e poniamo $\tilde{\mathbf{y}} = W^{\frac{1}{2}} \mathbf{y}$ e $\tilde{X} = W^{\frac{1}{2}} X$. Il modello lineare che stiamo considerando nelle variabili trasformate è perciò

$$\tilde{\mathbf{Y}} = \tilde{X} \boldsymbol{\theta} + \mathbf{Z} \quad (10.71)$$

dove \mathbf{Z} è lo stesso del modello (10.54). Risulta immediatamente che minimizzare il χ^2 della (10.70) equivale a minimizzare

$$\chi^2 = (\tilde{X}\boldsymbol{\theta} - \tilde{\mathbf{y}})^\dagger (\tilde{X}\boldsymbol{\theta} - \tilde{\mathbf{y}}) . \quad (10.72)$$

Questo χ^2 è formalmente identico a quello della (10.56), perciò abbiamo, tenendo conto che $W^{\frac{1}{2}}W^{\frac{1}{2}} = W$, della simmetria di W e delle (10.50):

$$\hat{\boldsymbol{\theta}} = (\tilde{X}^\dagger \tilde{X})^{-1} \tilde{X}^\dagger \tilde{\mathbf{y}} = (X^\dagger W X)^{-1} X^\dagger W \mathbf{y} \equiv \boldsymbol{\alpha}^{-1} \boldsymbol{\beta} , \quad (10.73)$$

dove le matrici definite nella (10.58) ora diventano $\boldsymbol{\alpha} = (X^\dagger W X)$ e $\boldsymbol{\beta} = X^\dagger W \mathbf{y}$.

Per quanto riguarda gli errori delle stime, si ha subito, dalla (10.60),

$$V(\hat{\boldsymbol{\theta}}) = \sigma_z^2 (\tilde{X}^\dagger \tilde{X})^{-1} = \sigma_z^2 (X^\dagger W X)^{-1} , \quad (10.74)$$

mentre la stima di σ_z^2 , guardando la (10.62), risulta

$$s_z^2 = \frac{(\tilde{X} \hat{\boldsymbol{\theta}} - \tilde{\mathbf{y}})^\dagger (\tilde{X} \hat{\boldsymbol{\theta}} - \tilde{\mathbf{y}})}{n - p - 1} = \frac{(X \hat{\boldsymbol{\theta}} - \mathbf{y})^\dagger W (X \hat{\boldsymbol{\theta}} - \mathbf{y})}{n - p - 1} . \quad (10.75)$$

Se le osservazioni Y_i della risposta sono gaussiane indipendenti, per trovare gli intervalli di confidenza per le θ_j , basterà dunque applicare la (10.63) con l'errore della (10.64) sostituito da $s_z \sqrt{(X^\dagger W X)_{jj}^{-1}} = s_z \sqrt{\boldsymbol{\alpha}_{jj}^{-1}}$. Nella routine **Linfit** troverete codificate le equazioni (10.70-10.75) per i minimi quadrati pesati.

La "previsione" di $\tilde{\mathbf{y}}(\mathbf{x})$ in corrispondenza di un valore \mathbf{x} non è probabilmente di molto interesse, dunque non conviene applicare direttamente la (10.65) a $\tilde{\mathbf{y}}$. Basta osservare che continua a valere $\hat{\mathbf{y}}(\mathbf{x}) = \mathbf{x}^\dagger \hat{\boldsymbol{\theta}}$ per la variabile risposta non trasformata e dunque la stima della varianza di $\hat{\mathbf{y}}(\mathbf{x})$ è $s_z^2 \mathbf{x}^\dagger (X^\dagger W X)^{-1} \mathbf{x}$, come nella (10.61). L'intervallo di confidenza di livello $1 - \alpha$ è perciò come quello della (10.65):

$$y(\mathbf{x}) \in \hat{\mathbf{y}}(\mathbf{x}) \pm t_{1-\alpha/2} s_z \sqrt{\mathbf{x}^\dagger (X^\dagger W X)^{-1} \mathbf{x}} . \quad (10.76)$$

L'intervallo di previsione per $Y(\mathbf{x})$ si ricava da questo aggiungendo alla stima della varianza di $\hat{\mathbf{y}}(\mathbf{x})$ quella della parte casuale del modello per Y in corrispondenza di \mathbf{x} :

$$Y(\mathbf{x}) = \mathbf{x}^\dagger \boldsymbol{\theta} + \frac{z}{\sqrt{w}}$$

cioè $\text{Var}[z/\sqrt{w}] = s_z^2/w$. Qui w è o un nuovo peso, se \mathbf{x} non è un punto sperimentale, o uno dei vecchi pesi in caso contrario. Risulta infine, come per la (10.66),

$$Y(\mathbf{x}) \in \hat{\mathbf{y}}(\mathbf{x}) \pm t_{1-\alpha/2} s_z \sqrt{\frac{1}{w} + \mathbf{x}^\dagger (X^\dagger W X)^{-1} \mathbf{x}} . \quad (10.77)$$

Quanto detto fin qui è valido se W è una matrice diagonale. Ma osserviamo la trasformazione $\tilde{\mathbf{y}} = W^{\frac{1}{2}}\mathbf{y}$ assieme alla ovvia fattorizzazione di V , cioè $V = \sigma_z^2 W^{-1} = \sigma_z^2 W^{-\frac{1}{2}} W^{-\frac{1}{2}}$. Se V non è diagonale e poniamo $V = \sigma_z^2 W^{-1}$, con W matrice nota, riusciamo ad ottenere una fattorizzazione analoga applicando la (4.69): $W^{-1} = H H^\dagger$. Qui H gioca lo stesso ruolo di $W^{-\frac{1}{2}}$. Rileggendo il par. 4.5, si comprende che la trasformazione $\tilde{\mathbf{Y}} = H^{-1}\mathbf{Y}$ rende $\tilde{\mathbf{Y}}$ un vettore gaussiano con matrice delle covarianze pari a $\sigma_z^2 I$ e vettore delle medie $\tilde{\mathbf{X}}\boldsymbol{\theta} = H^{-1}\mathbf{X}\boldsymbol{\theta}$, riconducendoci alle ipotesi utilizzate nel paragrafo precedente per ricavare tutti gli intervalli di confidenza. Dunque i risultati dalla (10.72) alla (10.76) valgono per una qualunque matrice delle covarianze $\sigma_z^2 W^{-1}$ senza modifiche. Anche la (10.77) continua a valere inserendo al posto di $1/w$ l'elemento di posto i sulla diagonale di W^{-1} se si prende in considerazione un punto sperimentale \mathbf{x}_i , oppure un nuovo coefficiente in caso contrario.

10.7 Fit con errori sistematici

Quando si effettua una procedura di *best fit* capita abbastanza frequentemente di avere a disposizione dati sperimentali che sono affetti da errori sia statistici che sistematici.

Nel prossimo capitolo faremo una trattazione completa degli errori sistematici; qui basta dire che per “sistematico” si intende un qualunque effetto, uguale per tutti i punti, che fa scostare la curva di *best fit* dalla media vera, indipendentemente dal numero delle osservazioni.

Consideriamo, ad esempio, il caso semplice di n misure indipendenti $y_i \pm \sigma_i$ tutte affette da uno stesso errore sistematico costante s incognito dovuto ad una cattiva calibrazione dello strumento utilizzato. Supponiamo inoltre di aver stimato, da misure indipendenti, che S è una variabile gaussiana con $\langle S \rangle = 0$ e varianza σ_S^2 (in questo paragrafo, S non indica lo stimatore della varianza campionaria).

In questa situazione, che nel linguaggio degli statistici è ricompresa nei cosiddetti modelli lineari misti [Dav08], occorre incorporare tutti gli errori, con le rispettive correlazioni, nella matrice V delle covarianze. Sotto queste condizioni, ogni variabile aleatoria y_i può essere espressa come somma di due componenti

$$Y_i = Y_i^R + Y^S$$

dove Y_i^R è una variabile aleatoria con valore medio pari al valore incognito da determinare con la procedura di *best fit* mentre Y^S è lo scostamento aleatorio, uguale per tutti i punti, che tiene conto dell'errore sistematico.

Applicando l'equazione di propagazione degli errori (5.65) e ricordando che per costruzione Y_i^R ed Y^S sono scorrelate, otteniamo:

$$\text{Var}[Y_i] = \text{Var}[Y_i^R] + \text{Var}[Y^S] = \sigma_i^2 + \sigma_S^2. \quad (10.78)$$

Per le ipotesi fatte, le variabili Y^S hanno invece tra di loro il grado di correlazione massima. Applicando il teorema di Cauchy-Schwartz (vedere eq. 4.28) ricaviamo quindi:

$$\text{Cov}[Y_i, Y_j] = \sigma_S^2 \quad \text{per } i \neq j. \quad (10.79)$$

Generalizzando la (10.70), possiamo così costruire la matrice V : lungo la diagonale abbiamo la somma quadratica degli errori statistici e sistematici mentre negli altri termini è presente il quadrato dell'errore sistematico:

$$V = \begin{pmatrix} \sigma_1^2 + \sigma_S^2 & \sigma_S^2 & \dots & \sigma_S^2 \\ \sigma_S^2 & \sigma_2^2 + \sigma_S^2 & \dots & \sigma_S^2 \\ \dots & \dots & \dots & \dots \\ \sigma_S^2 & \sigma_S^2 & \dots & \sigma_n^2 + \sigma_S^2 \end{pmatrix}. \quad (10.80)$$

Sia ora Δ il vettore delle differenze tra i valori teorici e sperimentali:

$$\Delta = \begin{pmatrix} \mu_1(\boldsymbol{\theta}) - y_1 \\ \dots \\ \dots \\ \mu_n(\boldsymbol{\theta}) - y_n \end{pmatrix}. \quad (10.81)$$

La funzione da minimizzare, in notazione matriciale, diventa

$$\chi^2(\boldsymbol{\theta}) = \Delta^T V^{-1} \Delta. \quad (10.82)$$

Come esempio, consideriamo una procedura di *best fit* lineare, con $\mu_i(\boldsymbol{\theta}) = \mu_i(a, b) = a + bx_i$, applicata ai punti riportati in Fig. 10.6, dove abbiamo posto $\sigma_i = \sigma_S = 0.5$. Le barre continue rappresentano, come al solito gli errori statistici mentre le linee tratteggiate indicano gli errori sistematici che sono stati aggiunti linearmente a quelli statistici, come conseguenza della (10.78).

Se eseguiamo il *best fit* considerando solo gli errori casuali, otteniamo i valori riportati nella prima riga della tabella sottostante

Errore	\hat{a}	\hat{b}
stat.	1.86 ± 0.47	0.44 ± 0.06
stat.+sys	1.86 ± 0.68	0.44 ± 0.06

Se teniamo conto anche degli errori sistematici attraverso la (10.80) otteniamo invece il risultato riportato nella seconda riga.

La soluzione al problema di minimizzazione è ricavabile in modo del tutto simile al par. 10.4:

$$\hat{a} = \frac{1}{D} [(\mathbf{x}^\dagger V^{-1} \mathbf{x})(\mathbf{1}^\dagger V^{-1} \mathbf{y}) - (\mathbf{1}^\dagger V^{-1} \mathbf{x})(\mathbf{x}^\dagger V^{-1} \mathbf{y})], \quad (10.83)$$

$$\hat{b} = \frac{1}{D} [(\mathbf{1}^\dagger V^{-1} \mathbf{1})(\mathbf{x}^\dagger V^{-1} \mathbf{y}) - (\mathbf{1}^\dagger V^{-1} \mathbf{x})(\mathbf{1}^\dagger V^{-1} \mathbf{y})], \quad (10.84)$$

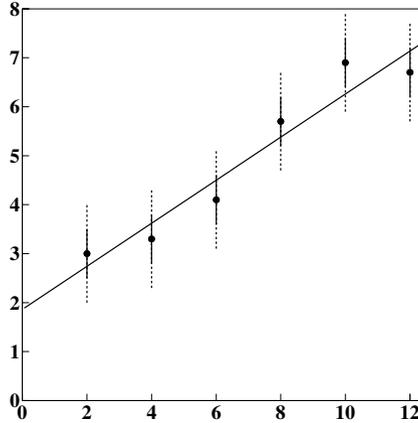


Fig. 10.6. Fit con errori sia casuali che sistematici. La linea continua rappresenta il risultato della procedura di *best fit*

dove $D = (\mathbf{1}^\dagger V^{-1} \mathbf{1})(\mathbf{x}^\dagger V^{-1} \mathbf{x}) - (\mathbf{1}^\dagger V^{-1} \mathbf{x})^2$ e $\mathbf{1}$ è un vettore colonna di elementi unitari di dimensione compatibile. Se $\sigma_S^2 = 0$ e $\sigma_i^2 = \sigma_z^2$, allora $V = \sigma_z^2 I$ e ritroviamo esattamente le equazioni (10.25) e (10.26); se $\sigma_S^2 = 0$ otteniamo le stime dei minimi quadrati pesati specifiche per la retta di regressione. Nel nostro esempio numerico σ_S^2 è positiva, ma $\sigma_i = \sigma_z = 0.5$ e questo fa sì che \hat{a} e \hat{b} siano esattamente le stesse del caso non pesato. Quanto appena affermato è verificabile nel modo seguente: se $\sigma_i = \sigma_z$ per ogni $i = 1, \dots, n$ e $\sigma_S \geq 0$, allora l'elemento di posto ij di V^{-1} si può scrivere così: $(\delta_{ij}\tau_z^2 - \tau_S^2)$, con $\tau_z^2 = 1/\sigma_z^2$ e

$$\tau_S^2 = \frac{\sigma_S^2}{\sigma_z^2} \frac{1}{n\sigma_z^2 + \sigma_S^2}.$$

Sviluppando le (10.83)-(10.84) e semplificando, si riottengono di nuovo le (10.25)-(10.26).

Tuttavia, l'errore delle stime cambia: come si può anche capire intuitivamente, l'aggiunta di uno stesso errore costante per tutti i punti non ha alcuna influenza sull'errore della pendenza della retta, ma solo sul parametro a , la cui stima è affetta da un errore maggiore rispetto al puro caso statistico. Applicando agli stimatori delle (10.25)-(10.26) la (5.72), cioè generalizzando al caso di matrice V non diagonale le formule di propagazione degli errori ricavate nei paragrafi 10.4 e 10.5, è facile verificare (vedere anche [Bar89]) che l'errore sul parametro \hat{b} rimane invariato mentre la varianza di \hat{a} è pari a

$$\frac{\sigma_z^2}{D^2} \sum_i [S_{xx} - S_x x_i]^2 + \frac{\sigma_S^2}{D^2} \sum_i \sum_j [S_{xx} - S_x x_i] [S_{xx} - S_x x_j]$$

(nella quale $D = nS_{xx} - S_x^2$), cioè alla somma tra la varianza ricavata dal *best fit* con le sole incertezze statistiche e l'ulteriore fonte di variazione dovuta all'errore sistematico.

L'applicazione della (5.72) dà risultati esatti per i minimi quadrati lineari, mentre è valida in generale solo nel caso di “piccoli” errori sistematici

$$\frac{\sigma_S^2}{\sigma_z^2 + \sigma_S^2} \ll 1 .$$

Per risolvere in maniera statisticamente corretta anche il caso generale di grandi errori sistematici, vi consigliamo la consultazione di [D'A94, HL07]. Infine, per un approfondimento sui modelli lineari misti e per le procedure di stima con σ_z^2 e σ_S^2 incognite, il riferimento è ancora [Dav08].

10.8 Proprietà delle stime dei minimi quadrati

Qui vi illustriamo i tre teoremi fondamentali (tra cui quello di Gauss-Markov) su cui si basano le stime dei minimi quadrati lineari. Se non vi piacciono le gioie della matematica, potete leggere solo l'enunciato dei teoremi (e le loro conseguenze!) e passare al prossimo paragrafo.

Teorema 10.1 (Stime corrette). *Nel caso di dipendenza lineare dai parametri, le stime dei minimi quadrati (LS) sono non distorte (“unbiased”).*

Dimostrazione. Applicando l'operatore media alla (10.59) e tenendo presente la (10.54) si ha immediatamente:

$$\langle \hat{\boldsymbol{\theta}} \rangle = (X^\dagger X)^{-1} X^\dagger \langle \mathbf{Y} \rangle = (X^\dagger X)^{-1} X^\dagger X \boldsymbol{\theta} = \boldsymbol{\theta} ,$$

in accordo con la (9.14). □

Teorema 10.2 (di Gauss-Markov). *Con riferimento al modello (10.54), tra tutti gli stimatori di $\boldsymbol{\theta}$ lineari in Y e non distorti lo stimatore LS è quello a varianza minima, ovvero lo stimatore più efficiente.*

Dimostrazione. Dobbiamo dimostrare che, se $\boldsymbol{\theta}^*$ è una stima lineare in \mathbf{Y} e non distorta dei parametri, si ha;

$$a^\dagger V(\hat{\boldsymbol{\theta}})a \leq a^\dagger V(\boldsymbol{\theta}^*)a , \quad (10.85)$$

dove a è un qualsiasi vettore di dimensione $p+1$. In particolare, se a contiene tutti zeri e 1 nella posizione i -esima, la (10.85) comprende anche i casi

$$V(\hat{\theta}_{ii}) \leq V(\theta_{ii}^*) , \quad i = 0, 1, \dots, p . \quad (10.86)$$

Consideriamo pertanto una generica stima lineare in \mathbf{Y} e non distorta dei parametri $\boldsymbol{\theta}$:

$$\boldsymbol{\theta}^* = U\mathbf{Y} .$$

Nel caso dello stimatore dei minimi quadrati, $U = (X^\dagger X)^{-1} X^\dagger$. Dato che la stima è per ipotesi non distorta, dalla (10.54) si ha:

$$\langle \boldsymbol{\theta}^* \rangle = U \langle \mathbf{Y} \rangle = UX\boldsymbol{\theta} = \boldsymbol{\theta} ,$$

da cui:

$$UX = I , \quad (UX)^\dagger = I . \quad (10.87)$$

È cruciale notare che questa proprietà non implica affatto $U = X^{-1}$, perché U ed X non sono matrici quadrate. In base alla (10.60) abbiamo:

$$V(\boldsymbol{\theta}^*) = \sigma_z^2 U U^\dagger .$$

Vale ora l'identità:

$$U U^\dagger = C + (U - CX^\dagger)(U - CX^\dagger)^\dagger , \quad (10.88)$$

dove C è la matrice LS degli errori divisa per σ_z^2 ,

$$C = (X^\dagger X)^{-1} .$$

Questa identità si può verificare facilmente utilizzando la (10.87) e il fatto che C , V e le loro inverse sono matrici simmetriche coincidenti con le corrispondenti trasposte:

$$\begin{aligned} & C + U U^\dagger - C X^\dagger U^\dagger - U X C^\dagger + C X^\dagger X C^\dagger \\ &= C + U U^\dagger - C X^\dagger U^\dagger - U X C^\dagger + C (X^\dagger X) C \\ &= C + U U^\dagger - C - C + C (X^\dagger X) C \\ &= C + U U^\dagger - C - C + C = U U^\dagger . \end{aligned}$$

Tenendo presente la (10.60) possiamo scrivere la (10.88) come:

$$V(\boldsymbol{\theta}^*) = V(\hat{\boldsymbol{\theta}}) + \sigma_z^2 (U - CX^\dagger)(U - CX^\dagger)^\dagger ,$$

la quale mostra che la matrice delle covarianze di $\boldsymbol{\theta}^*$ è uguale alla matrice delle covarianze delle stime LS $\hat{\boldsymbol{\theta}}$ più una matrice simmetrica definita positiva, in quanto essa è scritta nella forma HH^\dagger come nella (4.69). Questo dimostra il teorema.

Ovviamente, l'uguaglianza $V(\boldsymbol{\theta}^*) = V(\hat{\boldsymbol{\theta}})$ si ha quando $U = CX^\dagger = (X^\dagger X)^{-1} X^\dagger$, come è facile verificare.

È importante notare come questo teorema non implichi alcuna assunzione né sulla distribuzione delle popolazioni né sulla dimensione del campione \mathbf{y} . Ciò che viene richiesto è soltanto che i valori medi $\langle \mathbf{Y} \rangle$ siano funzioni lineari dei parametri. \square

Nel caso dei minimi quadrati pesati del par. 10.6, i teoremi appena dimostrati continuano a valere sostituendo rispettivamente \tilde{X} a X e $\tilde{\mathbf{Y}}$ a \mathbf{Y} .

Abbiamo visto che, nel caso di variabili gaussiane e modelli lineari, le stime LS forniscono anche intervalli di confidenza per i parametri, per $\langle Y|\mathbf{x} \rangle$ e per Y stessa, oltre a consentire di svolgere un test χ^2 di adattamento del modello tramite la statistica $\chi^2(\hat{\boldsymbol{\theta}})$ della (10.2). Il $\chi^2(\hat{\boldsymbol{\theta}})$, in questo caso, ha infatti distribuzione $\chi^2(n-p-1)$ e il *fit* del modello viene giudicato insoddisfacente al livello α se

$$SL = P \left\{ Q_R \geq \frac{\chi^2(\hat{\boldsymbol{\theta}})}{n-p-1} \right\} \leq \alpha, \quad (10.89)$$

dove Q_R ha distribuzione χ^2 ridotto con $n-p-1$ gradi di libertà.

Un importante teorema [SW89] permette di estendere queste proprietà, sotto opportune condizioni, anche quando i modelli non sono lineari e le variabili non sono gaussiane. Quanto diremo si applica ai χ^2 che rispettano la definizione 10.1, quindi non al caso con l'approssimazione della varianza effettiva. I modelli considerati nel teorema sono infatti del tipo

$$Y_i = \mu_i(\mathbf{x}_i, \boldsymbol{\theta}) + Y_{Ri}, \quad i = 1, \dots, n \quad (10.90)$$

dove le \mathbf{x}_i sono fissate e le Y_{Ri} sono indipendenti e identicamente distribuite con media nulla e varianze σ_i^2 . La definizione 10.1 rispetta la (10.90) perché l'assunzione $\langle Y_i|\mathbf{x}_i \rangle = \mu_i(\boldsymbol{\theta})$, coincide con la (10.90) e comprende sia il caso di predittori non aleatori, sia quello di predittori aleatori quando ci si condiziona ad $\underline{\mathbf{x}}$. Per il teorema assumeremo che $\mu_i(\boldsymbol{\theta}) = f(\mathbf{x}_i, \boldsymbol{\theta})$ per ogni i e useremo le notazioni

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = (\mu_1(\boldsymbol{\theta}), \dots, \mu_n(\boldsymbol{\theta}))^\dagger$$

$$F_{ij} = \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \theta_j}$$

dove le F_{ij} sono gli elementi di una matrice F di dimensione $n \times (p+1)$, generalizzazione della (10.53) vista nel caso lineare.

Teorema 10.3 (Stime dei minimi quadrati). *Dato il modello (10.90) con Y_{R1}, \dots, Y_{Rn} indipendenti e identicamente distribuite con media zero e varianze $\sigma_1^2, \dots, \sigma_n^2$, abbiamo approssimativamente:*

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \sim N_{p+1}(\mathbf{0}, \Sigma^{-1}), \quad \Sigma = F^\dagger W F, \quad (10.91)$$

dove W è la matrice diagonale con $1/\sigma_i^2$ nella posizione (i, i) e $\hat{\boldsymbol{\theta}}$ è lo stimatore dei minimi quadrati ([SW89], par. 12.2).

Se le Y_{Ri} sono anche gaussiane, la variabile aleatoria $\chi^2(\hat{\boldsymbol{\theta}})$ ha distribuzione asintotica $\chi^2(\nu)$, dove $\nu = n-p-1$, con $p+1$ pari alla dimensione di $\boldsymbol{\theta}$ ([SW89], teorema 2.1).

In pratica, allora, possiamo applicare le (10.62-10.66) al caso non lineare sostituendo ovunque $f(\mathbf{x}, \hat{\boldsymbol{\theta}})$ a $\mathbf{x}^\dagger \hat{\boldsymbol{\theta}}$, e il vettore delle derivate

$$\left. \frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\dagger} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}$$

Tabella 10.1. Proprietà dello stimatore dei minimi quadrati nel caso di errori noti. Il simbolo (*) indica che la proprietà vale sotto le condizioni del teorema 10.3. L'efficienza si riferisce al limite inferiore di Rao-Cramér per la prima riga (con dimensione campionaria finita) e per la seconda riga, dove le stime sono di massima verosimiglianza, mentre nella terza riga è limitata agli stimatori corretti e lineari nelle variabili osservate per il teorema di Gauss-Markov

Tipo di problema		Proprietà		
dati gaussiani?	modello lineare?	stimatore efficiente?	stime gaussiane?	test χ^2 possibile?
sì	sì	sì	sì	sì
sì	no	sì	sì (*)	sì (*)
no	sì	sì	sì (*)	no
no	no	?	sì (*)	no

a ciascuna riga \mathbf{x}_i^\dagger di X , quando X non è moltiplicata per $\boldsymbol{\theta}$.

Il teorema 10.3 giustifica in parte l'abitudine diffusa di applicare la legge 3σ agli intervalli di stima forniti dagli algoritmi dei minimi quadrati non lineari e di condurre il test χ^2 di *best fit* (si veda anche la Tab. 10.1). Vi consigliamo però, nei casi importanti e qualora vi siano dei dubbi, di simulare la procedura con dati artificiali e verificare direttamente le distribuzioni dei parametri stimati e dei valori del χ^2 con i metodi del par. 7.10.

10.9 Verifica del modello e determinazione di forme funzionali

I risultati che abbiamo esposto finora, sono validi se il modello ipotizzato per il valore atteso della risposta è corretto. Quando non si conosce a priori la forma funzionale del modello lineare, si comincia a stimarne uno ipotetico scegliendolo nella classe definita dalla (10.47). Consideriamo ad esempio una sottoclasse interessante di modelli di regressione nella quale ciascuno dei p predittori è funzione della stessa variabile x :

$$f(x, \boldsymbol{\theta}) = \theta_0 + \sum_{k=1}^p \theta_k f_k(x) \quad , \quad f(x, \boldsymbol{\theta}) = \theta_0 + \sum_{k=1}^p \theta_k x^k \quad , \quad (10.92)$$

da adattare ad n punti sperimentali con errore σ_z ignoto.

Ricordiamo che il problema consiste nel trovare le curva rispetto alla quale le *fluttuazioni dei punti sono casuali*, ovvero la curva che rappresenta la relazione funzionale tra \mathbf{x} e il valore medio di Y . Questa curva *non deve passare esattamente per i punti* (interpolazione) e pertanto deve avere un numero

di parametri inferiore alla dimensione del campione; per questo motivo essa è detta anche curva di regressione (anche se storicamente questo termine è stato introdotto in un contesto diverso).

La scelta iniziale della curva nella sottoclasse (10.92) avviene in base alla conoscenza del problema in esame e in base all'andamento dei grafici delle coppie (x_i, y_i) , quando ciò è possibile, come nelle figure 10.1 e 10.2. Ottenuta la stima dalla minimizzazione del χ^2 , la prima cosa da fare è controllare che i residui

$$\hat{z}_i = y_i - \hat{y}(\mathbf{x}_i) = y_i - \hat{y}_i, \quad (10.93)$$

che di fatto costituiscono una stima delle fluttuazioni casuali z_i , si comportino appunto come fluttuazioni casuali. Il grafico delle coppie (\hat{y}_i, \hat{z}_i) è molto informativo in questo senso. Infatti i residui hanno sempre media nulla e si dimostra che, se il modello include il termine costante θ_0 , la correlazione campionaria tra i valori previsti \hat{y}_i e i residui \hat{z}_i è nulla e non sussiste tra essi relazione lineare. Se le ipotesi fatte su $f(x, \boldsymbol{\theta})$ e su media (nulla) e varianza (costante) delle fluttuazioni casuali z_i sono compatibili coi dati, dovremmo dunque osservare nel grafico una nuvola di punti racchiusa in una fascia di ampiezza costante² e collocata attorno allo zero. In particolare, se i residui sono anche gaussiani, la fascia risulterà approssimativamente simmetrica attorno allo zero.

Nel caso pesato si devono rappresentare i residui standard (cioè divisi per l'errore), calcolati per ogni predittore \mathbf{x}_i :

$$t(x_i) = \frac{\hat{z}_i}{\sigma_i} = \frac{y_i - \hat{y}(\mathbf{x}_i)}{\sigma_i}. \quad (10.94)$$

L'andamento deve apparire casuale ed i valori al di fuori della fascia $|t_i| < 3$ devono essere rari, in accordo con la legge 3σ . (Questa è una verifica approssimativa, infatti la varianza di \hat{Z}_i è quella della nota 2, con \tilde{X} al posto di X). Se σ_z^2 non è nota i residui vanno comunque pesati e nel grafico si rappresentano le coppie $(\hat{y}(\mathbf{x}_i), \sqrt{w_i}\hat{z}_i)$. La Fig. 10.7 mostra tre possibili casi. Nel caso a) il *fit* è soddisfacente, nel caso b) i punti indicano un andamento che descrive una curva che viola la legge 3σ , ad indicazione di una inadeguata forma funzionale introdotta nel *fit*; il caso c) invece si verifica quando gli errori sui valori grandi di y sono stati sottostimati. Frequentemente questo avviene quando si esegue un *fit* non pesato, con errori assoluti tenuti costanti, di dati che invece hanno errore *relativo* costante. L'errore assoluto cresce allora con y e il grafico dei residui è come nella Fig. 10.7c); i matematici chiamano questo comportamento col nome di *eteroschedasticità*.

Tuttavia, diversi modelli possono produrre un grafico dei residui visivamente soddisfacente. Lo si affianca allora a dei metodi iterativi, con i quali

² Attenzione però che $\text{Var}[\hat{Z}_i] = \sigma_z^2(1 - l_{ii})$, dove l_{ii} sta tra 0 e 1 ed è l'elemento di posto ii della matrice $X(X^\dagger X)^{-1}X^\dagger$. Può accadere che, per certe matrici X , alcuni l_{ii} siano sensibilmente diversi dalla maggioranza degli altri.

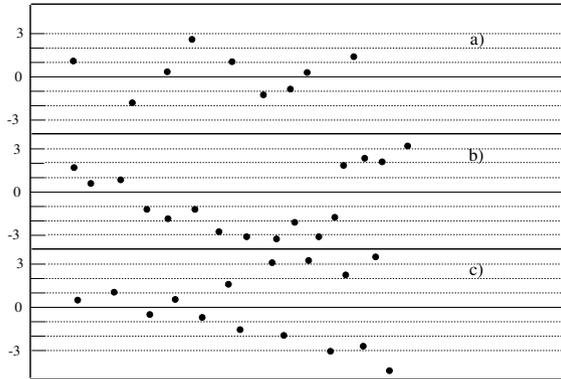


Fig. 10.7. Possibili grafici dei residui a seguito di un *best fit*: *fit* buono a); forma funzionale insoddisfacente b); eteroschedasticità (si veda il testo) c)

si confrontano in successione coppie di modelli diversi, ottenuti l'uno dall'altro aggiungendo o togliendo predittori. Nel caso del polinomio della (10.92), potremmo aumentarne il grado di un'unità per volta, minimizzare il χ^2 e decidere quando fermarci. Il metodo si basa sul *fit* dei modelli e sulla risposta alla domanda: “Il peggioramento del *fit* che ottengo togliendo determinati predittori $f_k(x)$ è statisticamente significativo o no?”

Per rispondere, partiamo dall'esame della Fig. 10.8, che rappresenta i vettori coinvolti nel nostro problema dei minimi quadrati nello spazio \mathbb{R}^n . Per capirne il senso, notiamo che la stima della risposta in corrispondenza di ciascuna riga di X , cioè il vettore delle $\hat{y}(x_i)$, corrisponde a

$$\hat{\mathbf{y}} = X \hat{\boldsymbol{\theta}} = X(X^\dagger X)^{-1} X^\dagger \mathbf{y} = P \mathbf{y} . \tag{10.95}$$

Si dimostra che la matrice P non fa altro che attuare la proiezione ortogonale di \mathbf{y} sullo spazio vettoriale generato dalle colonne di X , cioè $\hat{\boldsymbol{\theta}}$ è quel vettore di parametri che avvicina il più possibile \mathbf{y} ai vettori generati da combinazioni lineari delle colonne di X . Consideriamo ora la matrice X_0 , ottenuta da X eliminando un certo numero di colonne e tenendo p_0 predittori (quindi X_0 ha dimensione $n \times (p_0 + 1)$). La stima della risposta con questo modello ridotto sarà naturalmente $\hat{\mathbf{y}}_0 = P_0 \mathbf{y}$, con $P_0 = X_0(X_0^\dagger X_0)^{-1} X_0^\dagger$, matrice di proiezione sullo spazio delle colonne di X_0 . Ciò equivale ad aver posto $p - p_0$ coefficienti θ_i pari a zero.

Applicando il teorema di Pitagora al triangolo rettangolo tratteggiato in Fig. 10.8, otteniamo un'importante relazione che interessa i residui della regressione con il modello completo e con quello ridotto. Ricordando che $\sum_i \hat{z}_i^2 = \|\hat{\mathbf{z}}\|^2 = \sum_i (y_i - \hat{y}(x_i))^2$ è il quadrato della norma del vettore $\hat{\mathbf{z}}$, abbiamo

$$\|\hat{\mathbf{z}}_0\|^2 = \|\hat{\mathbf{z}} - \hat{\mathbf{z}}_0\|^2 + \|\hat{\mathbf{z}}\|^2 . \tag{10.96}$$

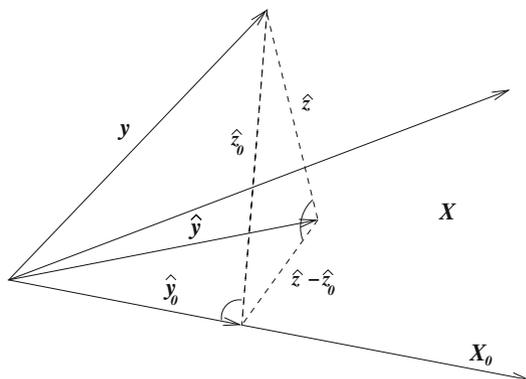


Fig. 10.8. Rappresentazione vettoriale nello spazio \mathbb{R}^n . Regressione con tutti i predittori (matrice X) e con un sottoinsieme di predittori (matrice X_0). La stima della risposta $\hat{y}(X) \equiv \hat{y}$ in corrispondenza di ciascuna riga di X è la proiezione ortogonale di y sullo spazio lineare generato dalla colonne di X . Un discorso simile vale per \hat{y}_0 . È dunque immediato ricavare i vettori dei residui delle due regressioni e la relazione tra essi

Se $y(x)$ è dato dalla prima delle (10.92) e le funzioni $f_k(x)$ soddisfano alla proprietà di ortogonalità (10.67), scrivendo esplicitamente i residui si può verificare facilmente che la (10.96) equivale alla condizione

$$\sum_i [y_i - \hat{y}(x_i)][\hat{y}(x_i) - \hat{y}_0(x_i)] = \sum_{k=p_0+1}^p \sum_i [y_i - \hat{y}(x_i)] \hat{\theta}_k f_k(x_i) = 0, \quad (10.97)$$

la quale è verificata grazie alle equazioni normali (10.57).

Dalla (10.35) è chiaro che gli elementi dei vettori dei residui sono combinazioni lineari delle risposte y_i . Se la risposta ha distribuzione gaussiana, le due norme al secondo membro, sono relative a vettori ortogonali (quindi incorrelati) che, essendo a loro volta combinazione lineare di vettori gaussiani, sono indipendenti. Inoltre, se vale il modello ridotto con p_0 predittori, si dimostra che

$$\frac{1}{\sigma_z^2} \|\hat{z} - \hat{z}_0\|^2 \sim \chi^2(p - p_0), \quad \frac{1}{\sigma_z^2} \|\hat{z}\|^2 \sim \chi^2(n - p - 1). \quad (10.98)$$

Questi risultati sull'indipendenza e la distribuzione delle norme al quadrato, ci danno la risposta alla domanda che ci siamo posti poco fa, perché, se è valido il modello ridotto, allora

$$\frac{\|\hat{z} - \hat{z}_0\|^2 / (p - p_0)}{\|\hat{z}\|^2 / (n - p - 1)} = \frac{(\|\hat{z}_0\|^2 - \|\hat{z}\|^2) / (p - p_0)}{\|\hat{z}\|^2 / (n - p - 1)} \quad (10.99)$$

segue la densità F di Snedecor con $(p - p_0, n - p - 1)$ gradi di libertà definita dalla (5.46) e già utilizzata nell'esercizio 6.12. Questa statistica è infatti il

rapporto tra due distribuzioni χ^2 ridotto indipendenti. Spesso la (10.99) è scritta con una notazione diversa

$$F = \frac{(RSS_0 - RSS)/(p - p_0)}{RSS/(n - p - 1)} \tag{10.100}$$

e $RSS = \|\hat{z}\|^2$ è l'acronimo di *residual sum of squares*. La statistica (10.100) misura il peggioramento del *fit* (aumento della somma dei quadrati dei residui) del modello ridotto rispetto a quello del modello completo. Se F è significativamente grande vuol dire che i predittori eliminati dal modello completo erano importanti, dunque rifiuteremo al livello α l'ipotesi che i relativi coefficienti siano nulli se

$$F > F_{1-\alpha}(p - p_0, n - p - 1) . \tag{10.101}$$

Un caso particolare è quello nel quale il modello ridotto si ottiene eliminando un solo predittore, ad esempio ponendo $\theta_p = 0$ in una delle (10.92), così $p_0 = p - 1$. La F allora, sotto l'ipotesi $H_0 : \theta_p = 0$, ha distribuzione $F(1, n - p - 1)$ e non è difficile vedere che sussiste l'uguaglianza $F_{1-\alpha}(1, n - p - 1) = t_{1-\alpha/2}^2(n - p - 1)$, il percentile della t di Student con $(n - p - 1)$ gradi di libertà. In realtà anche F è il quadrato di

$$T = \frac{|\hat{\theta}_p|}{s(\hat{\theta}_p)} \tag{10.102}$$

dove $s(\hat{\theta}_p)$ è come nella (10.63), e T ha distribuzione t di Student con $(n - p - 1)$ gradi di libertà. Potete verificarlo leggendo la discussione che precede gli intervalli di confidenza per i parametri della retta dei minimi quadrati (10.39), dove $p = 1$, $a = \theta_0$ e $b = \theta_1$. Detto ciò, rifiuteremo H_0 accettando il nuovo parametro $(p_0 + 1)$ -esimo se T supera $t_{1-\alpha/2}(n - p - 1)$. Visto in altro modo, un valore grande di T indica un errore relativo piccolo e quindi l'importanza del parametro.

Tabella 10.2. Risultati di una serie di *best fit* con polinomi del tipo $y(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$ per i dati della (10.104). RSS è la somma dei quadrati dei residui pesati, corrispondente a $\sigma_z^2 \chi^2$

	<i>FIT1</i>	<i>FIT2</i>	<i>FIT3</i>	<i>FIT4</i>	<i>FIT5</i>
θ_0	-5.1 ± 2.7	-15.1 ± 3.0	-6.1 ± 5.1		1.8 ± 1.1
θ_1	12.8 ± 1.4	23.2 ± 2.8	10.6 ± 6.7	2.7 ± 1.4	
θ_2		-1.5 ± 0.4	2.8 ± 2.1	5.1 ± 0.8	6.0 ± 0.5
θ_3			-0.4 ± 0.2	-0.5 ± 0.1	-0.6 ± 0.1
<i>RSS</i>	0.380	0.097	0.049	0.067	0.079

Con riferimento al problema della scelta del grado del polinomio nella (10.92), potremmo ad esempio partire dal modello con il solo θ_0 e verificare l'ipotesi $H_0 : \theta_1 = 0$. Se questa viene rifiutata, aggiungeremo il termine di secondo grado e verifichiamo l'ipotesi $H_0 : \theta_2 = 0$ e così via. Se k è l'indice del primo test che non porta a rifiutare H_0 , porremo $p = k - 1$. Questo è solo uno dei possibili procedimenti. Se per caso il termine di secondo grado non risultasse significativo, ma lo fosse quello di terzo grado, con questo procedimento non ce ne accorgeremo.

Tutto quello che abbiamo visto finora non basta ad assicurare che la determinazione della forma funzionale sia avvenuta correttamente, poiché sono condizioni necessarie ma non sufficienti. In altre parole, vi sono molte forme funzionali (non solo quella "vera") che soddisfano ai criteri di buon *fit* sopra elencati. Questa è una ambiguità di fondo, *non risolvibile solo per via statistica*, della quale occorre sempre essere consci. È quindi molto importante introdurre, nel modello di *fit*, tutte le informazioni a priori che si hanno del problema e cercare di ridurre gli errori sui dati per quanto possibile.

Esemplifichiamo questi concetti simulando un caso realistico. Partendo dal polinomio

$$y = \theta_0 + \theta_2 x^2 + \theta_3 x^3 = 3 + 5x^2 - 0.5x^3, \quad (10.103)$$

generiamo dei dati simulati di tipo gaussiano con una deviazione standard relativa piuttosto grande, di $\pm 15\%$:

$$y_i = (1 + 0.15 g_i)(3 + 5x_i^2 - 0.5x_i^3),$$

dove g_i è un numero casuale estratto dalla gaussiana standard, secondo le tecniche illustrate nel cap. 7. I dati ottenuti sono riportati nella (10.104):

x	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	
y	6.9	22.4	40.8	64.4	60.0	81.0	78.0	70.5	,
σ	1.1	2.9	5.2	7.7	9.8	12.1	11.7	10.6	(10.104)

dove $\sigma = \sigma_z / \sqrt{w} = 0.15 / \sqrt{w}$ e $1 / \sqrt{w} = 3 + 5x^2 - 0.5x^3$.

I polinomi di regressione di vario grado ricavati da questi dati sono mostrati nella Tab. 10.2, assieme ai valori di *RSS*. Essi sono stati ottenuti col nostro codice `Pol14fit` che fa uso della routine `Linfit`.

Il *fit* con la retta

$$y(x) = \theta_0 + \theta_1 x,$$

detto FIT1 in Tab. 10.2, non supera la verifica con il grafico dei residui: la Fig. 10.9 ha infatti un andamento simile a quello di Fig. 10.7b.

Consideriamo allora il polinomio di secondo grado

$$y(x) = \theta_0 + \theta_1 x + \theta_2 x^2,$$

ottenendo il risultato FIT2.

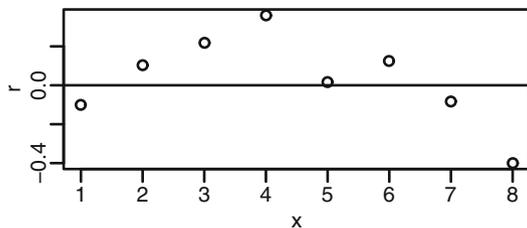


Fig. 10.9. Residui della regressione FIT1 di Tab. 10.2

Il rapporto (10.99) per il test F tra FIT2 e FIT1 vale:

$$F(1, 5) = \frac{0.380 - 0.097}{0.097} 5 = 14.6 .$$

Consultando la Tab. D.5 per i livelli al 5% e all'1%, otteniamo

$$14.6 > F_{0.95}(1, 5) = 6.61 , \quad 14.6 < F_{0.99}(1, 5) = 16.3 .$$

Il test mostra che l'ipotesi nulla (cioè la inutilità di FIT2) può essere rifiutata al livello del 5%, ma non a quello dell'1%. Se consideriamo che la soluzione è al limite del livello all'1%, è lecito ritenere significativa la soluzione a 3 parametri di FIT2.

A questo punto potremmo fermarci, ma vediamo cosa succede con FIT3, in cui si utilizza la cubica con 4 parametri

$$y(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 .$$

In questo caso l' F test, che vale $3.9 < F_{0.99}(1, 4) = 21.20$, mostra che il quarto parametro è inutile e tutti i 4 parametri di FIT3 presentano un piccolo valore (10.102) di T e sono pertanto compatibili con lo zero. Tutto questo ci dice che quattro parametri liberi sono eccessivi.

In FIT4 e FIT5 si ritenta allora il *fit* con delle cubiche, ma sopprimendo rispettivamente i parametri θ_0 e θ_1 . Il confronto di FIT4 e FIT5 rispetto a FIT3 con il test F , indica di nuovo che i quattro parametri liberi di FIT3 sono troppi rispetto ai tre di FIT4 e FIT5, indicando che i *fit* risultanti sono in pratica equivalenti a FIT2.

I risultati FIT2, FIT4 e FIT5 di Tab. 10.2 sono dunque tutti egualmente compatibili, anche se noi sappiamo (ma solo perché i dati sono stati simulati dalla (10.103)) che la soluzione più vicina a quella vera è quella di FIT5.

La conclusione corretta è quindi che, partendo da un polinomio di primo grado e aggiungendo progressivamente termini di grado superiore, gli 8 valori di Tab. 10.104, affetti da un errore considerevole di $\pm 15\%$, sono compatibili con una forma funzionale quadratica o cubica, avente non più di tre parametri. Tra questa classe di soluzioni rientra la funzione vera (10.103). L'uso

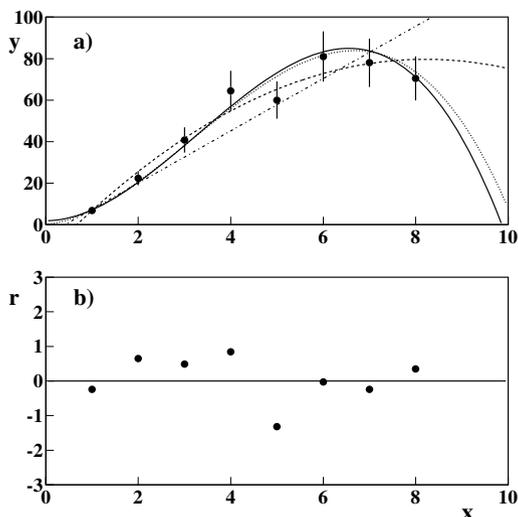


Fig. 10.10. a) Punti sperimentali della (10.104) e polinomi di *fit* ottenuti in Tab. 10.2; linea tratto-punto: FIT1; linea a tratti: FIT2; linea a punti: FIT4; linea piena: FIT5. A parte FIT1, tutti i polinomi hanno 3 parametri liberi. b) Residui (10.93) tra i dati e la curva piena FIT5; si confronti con la Fig. 10.7

di funzioni ortogonali con la proprietà (10.67) ottimizza notevolmente la procedura di minimizzazione, perchè i parametri sono indipendenti e quelli già trovati non cambiano valore aggiungendone dei nuovi, ma in genere non aiuta a risolvere le ambiguità.

La situazione è riassunta in modo efficace in Fig. 10.10, dove si vede immediatamente che polinomi di secondo o terzo grado sono statisticamente compatibili con le fluttuazioni dei dati. La figura mostra anche come *l'estrapolazione* delle curve al di fuori della zona di esistenza dei dati sia estremamente pericolosa: il valore estrapolato corretto, per $x = 10$, è quello di FIT5, che è la curva piena; risultati completamente diversi si ottengono con le altre curve, che pure rappresentano bene i dati entro i valori di x misurati.

10.10 Ricerca delle correlazioni

Il metodo per la ricerca del modello (con l'esempio specifico della ricerca delle forme funzionali) del paragrafo precedente si fonda sul confronto tra coppie di modelli annidati l'uno nell'altro, ma non ci dice se il *fit* di ciascun modello, preso singolarmente, è buono. Se confrontassimo ogni modello ragionevole con uno stesso modello base di riferimento, potremmo associare a ciascuno di essi una "cifra di merito" che ne riflette il *fit* ai dati. Una tecnica per far questo e che ora illustreremo prende il nome di *ricerca delle correlazioni*.

Prendiamo in esame la decomposizione dei quadrati dei residui dal modello ridotto (10.96) con $p_0 = 0$, ovvero quello con il solo termine costante θ_0 e la matrice X_0 coincidente col vettore colonna di tutti uni. Dalla (10.59) ricaviamo $\hat{\theta}_0 = \langle y \rangle = \hat{y}_{0i}$ e quindi $\hat{z}_{0i} = y_i - \langle y \rangle$. Pertanto

$$\|\hat{z}_0\| = \sum_{i=1}^n (y_i - \langle y \rangle)^2 = (n - 1)s_y^2 .$$

Inoltre, dalla Fig 10.8, vediamo che $\|\hat{z} - \hat{z}_0\| = \|\hat{y} - \hat{y}_0\|$. La decomposizione (10.96) diviene in questo caso una decomposizione della varianza campionaria di Y e assume la forma:

$$\underbrace{\sum_{i=1}^n (y_i - \langle y \rangle)^2}_{(RSS_0)} = \underbrace{\sum_{i=1}^n (\hat{y}_i - \langle y \rangle)^2}_{(RSS_0 - RSS)} + \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{(RSS)} \tag{10.105}$$

che viene detta a parole come:

$$\text{devianza totale} = \text{devianza spiegata} + \text{devianza residua} . \tag{10.106}$$

Questa eguaglianza, che è la più semplice forma di *analisi della varianza*, ha un'interpretazione molto interessante, cioè che la dispersione delle y_i attorno a $\langle y \rangle$ è decomposta in due parti tra loro loro incorrelate: la prima è quella colta dal modello di regressione lineare con p predittori, la seconda è il residuo, cioè quello che il modello non può spiegare della variazione presente nei dati. Se il modello interpolasse i dati, si avrebbe che $y_i - \hat{y}_i = 0$ per ogni i e la componente residua sarebbe nulla.

Dividendo la devianza spiegata della (10.106) per la devianza totale otteniamo il *coefficiente di determinazione* :

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \langle y \rangle)^2}{\sum_{i=1}^n (y_i - \langle y \rangle)^2} \tag{10.107}$$

$$= 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \langle y \rangle)^2} = 1 - \frac{RSS}{(n - 1)s_y^2} . \tag{10.108}$$

Vista la (10.105), $0 \leq R^2 \leq 1$, raggiunge 1 nel caso in cui il modello interpola i dati e vale zero solo se $\hat{\theta}_1 = \hat{\theta}_2 = \dots = \hat{\theta}_p = 0$, il che accade se y non dipende linearmente da nessuno dei predittori inseriti nel modello e le sue fluttuazioni sono esclusivamente "residue". In pratica non accade mai che R^2 sia esattamente zero, ma può accadere che sia molto vicino a zero.

Ancora grazie alla Fig. 10.8, si mostra che R^2 è il quadrato del coefficiente di correlazione campionario tra le y_i e le \hat{y}_i . Tale coefficiente, si dice *coefficiente di correlazione multipla* e può essere visto come una misura della correlazione campionaria tra le y_i e le righe \mathbf{x}_i della matrice X , in quanto ciascun \hat{y}_i è la previsione di Y in corrispondenza di \mathbf{x}_i . Nel caso specifico della retta

dei minimi quadrati (con $p = 1$), sostituendo la (10.29) al numeratore della (10.107) si ha

$$\sum_{i=1}^n (\hat{y}_i - \langle y \rangle)^2 = r^2 \frac{s_y^2}{s_x^2} \sum_{i=1}^n (x_i - \langle x \rangle)^2 = r^2 (n-1) s_y^2$$

e si vede immediatamente che $R^2 = r^2$, cioè il quadrato del coefficiente di correlazione lineare tra le x_i e le y_i .

Per completare l'analisi della decomposizione (10.105) menzioniamo il coefficiente di determinazione corretto con i gradi di libertà, se non altro perché compare nell'output di molti programmi di stima dei minimi quadrati. Esso è definito da :

$$R_a^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-p-1)}{\sum_{i=1}^n (y_i - \langle y \rangle)^2 / (n-1)} = 1 - \frac{RSS / (n-p-1)}{s_y^2} .$$

La ricerca delle correlazioni ha come obiettivo *trovare la funzione $f(x)$ che massimizza R^2* , cioè la funzione che massimizza la devianza spiegata e minimizza quella residua. Schematizziamo la procedura per forme funzionali del tipo (10.92), tenendo presente che resta valida per qualunque modello di regressione lineare multipla:

- a) data una funzione $f(x, \theta)$, lineare nei parametri, si trova la stima dei minimi quadrati $\hat{\theta}$, minimizzando la quantità

$$\sigma_z^2 \chi^2(\theta) = \sum_i [y_i - f(x_i, \theta)]^2 ; \quad (10.109)$$

- b) una volta eseguita la stima, si calcolano la devianza spiegata e il coefficiente di determinazione R^2 ;
- c) tra tutte le forme funzionali ipotizzate al punto a), si cerca quella che ha il massimo R^2 . Se vi sono due soluzioni pressoché equivalenti, si sceglie quella col minor numero di parametri. (Il risultato non cambia se, anziché massimizzare R^2 , si minimizza $RSS = \sigma_z^2 \chi^2(\hat{\theta})$);
- d) non dimenticate di verificare che il grafico dei residui della soluzione scelta abbia un andamento casuale, senza strutture evidentemente attribuibili ad una relazione funzionale non colta da $f(x, \hat{\theta})$.

Applichiamo queste regole ad un insieme di dati simulati con l'algoritmo:

$$\begin{aligned} x &= x_0 + x_R = 10 + 2g_1 \\ f(x) &= 2 + x^2 \\ y &= f(x) + Z = 2 + x^2 + 5g_2 , \end{aligned} \quad (10.110)$$

dove g_1 e g_2 sono valori estratti da curve gaussiane standard. La variabile X è dunque di tipo normale con $\mu_x = 10$, $\sigma_x = 2$, mentre la Y non è di tipo normale perché la correlazione $f(X)$ non è lineare; tuttavia, le fluttuazioni

intorno alla funzione di correlazione sono gaussiane, con $\sigma_z = 5$. Si sono così ottenute le venti coppie (x_i, y_i) riportate nella (10.111).

x	6.28	6.62	7.10	7.46	7.54
y	40.2	47.3	47.4	48.7	52.2
x	8.11	8.62	8.95	9.00	9.62
y	68.3	79.8	88.3	86.4	97.7
x	9.92	10.10	10.11	10.25	10.34
y	98.7	102.2	101.7	111.2	105.4
x	11.09	12.23	12.46	13.31	13.82
y	118.5	154.3	164.0	182.5	191.1

(10.111)

I risultati ottenuti con la nostra routine `Pol3fit` minimizzando il χ^2

$$\chi^2 = \sum_i (y_i - \theta_0 - \theta_1 x_i - \theta_2 x_i^2 - \theta_3 x_i^3)^2$$

rispetto a polinomi fino al terzo grado sono riportati in Tab. 10.3 e Fig. 10.11. Gli errori dei parametri sono calcolati dalla routine `Linfit` con le (10.74, 10.75).

Se ora applichiamo le regole a)-d) appena discusse, arriviamo a scegliere come migliore soluzione la FIT3:

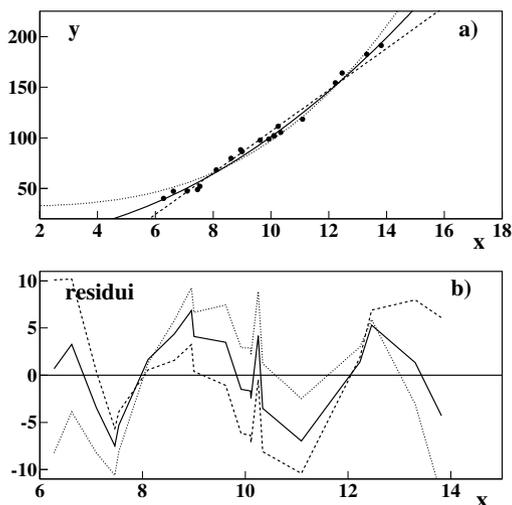
$$f(x) = 0.95 + 1.03 x^2 .$$

Questa soluzione massimizza la varianza spiegata al 99.1% e minimizza la fluttuazione gaussiana intorno alla curva di regressione al valore $s_z^2 = 4.4$. La soluzione FIT2 fornisce gli stessi risultati, ma con un parametro libero in più e con errori dei parametri troppo grandi, cioè con valori T della (10.102) troppo piccoli. La conclusione più probabile è quindi che i dati abbiano una funzione di correlazione di tipo parabolico $f(x) \simeq x^2$, con un termine costante piccolo rispetto alla media dei valori di y o forse trascurabile, dato che l'errore sul parametro è grande ($\theta_0 \in -1.02 \pm 2.45$). Le fluttuazioni assolute intorno a questa curva sono $\simeq 4.4$. Come vedete, questa conclusione è abbastanza vicina alla “verità” rappresentata dalle (10.110). Anche il grafico dei residui, riportato in Fig. 10.11b), mostra che la soluzione FIT3 ha un grafico dei residui con fluttuazioni più regolari rispetto alle altre soluzioni.

Notiamo infine che, se aumentassimo di molto il numero dei parametri, ad un certo punto troveremmo certamente valori di R^2 ancora più grandi e valori di s_z^2 ancora più piccoli. Al limite, con venti parametri passeremmo esattamente attraverso tutti i punti, ottenendo $R^2 = 1$ e $s_z^2 = 0$! Tuttavia, questi *fit* di ricerca delle correlazioni vanno fatti con funzioni lontane dal limite dell'interpolazione, rispetto alle quali le fluttuazioni dei dati devono apparire come statistiche. Le forme funzionali da provare, come ad esempio

Tabella 10.3. Risultati di una serie di *best fit* con polinomi del tipo $y(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$ per i dati della (10.111)

	<i>FIT1</i>	<i>FIT2</i>	<i>FIT3</i>	<i>FIT4</i>
θ_0	-98.7 ± 6.6	-12.2 ± 20.2	-1.02 ± 2.45	32.4 ± 2.8
θ_1	20.5 ± 0.7	2.3 ± 4.2		
θ_2		0.91 ± 0.21	1.03 ± 0.02	
θ_3				0.065 ± 0.002
$s_z = \sqrt{\frac{\chi^2}{n-p-1}}$	6.3	4.4	4.4	6.9
R^2	98.11%	99.12%	99.10%	97.72%

**Fig. 10.11.** a) Punti sperimentali della (10.111) e polinomi di *fit* ottenuti in Tab. 10.3; linea a tratti: *FIT1*; linea piena: *FIT3*; linea a punti: *FIT4*. La soluzione *FIT2* fornisce risultati graficamente indistinguibili da quelli di *FIT3*. b) Residui (10.93) corrispondenti alle tre soluzioni riportate in Fig. a)

il grado massimo dei polinomi, vanno quindi decise a priori sulla base delle informazioni che si hanno sul problema, per raggiungere un compromesso tra il miglior *fit* possibile e un modello con il minimo numero di parametri liberi.

A conclusione di questo paragrafo, rileviamo la relazione tra il test F per verificare l'ipotesi nulla $H_0 : \theta_1 = \dots = \theta_p = 0$ e l' R^2 . Confrontando la

(10.105) e la (10.107), troviamo che

$$R^2 = \frac{RSS_0 - RSS}{RSS_0} = \frac{(RSS_0 - RSS)/RSS}{(RSS_0 - RSS)/RSS + 1} = \frac{pF}{pF + (n - p - 1)}$$

dove $F \sim F(p, n - p - 1)$. Poiché R^2 è funzione crescente di F , scegliere la forma funzionale tra quelle con i più elevati valori di R^2 significa cercarla tra quelle che, sottoposte alla verifica di H_0 , portano a rifiutarla con i più piccoli p -value.

10.11 Test di adattamento del modello

Consideriamo, per generalità, i minimi quadrati pesati, che comprendono anche gli altri casi quando gli errori sono costanti. Se le precedenti stime degli errori σ_i^2 sono considerate affidabili e la risposta è gaussiana, si ha, quando il modello ipotizzato è corretto

$$\chi^2(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \frac{[y_i - \mathbf{x}_i^\dagger \hat{\boldsymbol{\theta}}]^2}{\sigma_i^2} \sim \chi^2(n - p - 1).$$

Dunque è lecito eseguire il test χ^2 di adattamento. In genere la bontà del *fit* è controllata col test a una coda della (6.123), che scarta il modello se il livello di significatività osservato SL (o p -value) per la coda *a destra* è più piccolo del livello del test α :

$$SL = P \left\{ Q_R \geq \frac{\chi^2(\hat{\boldsymbol{\theta}})}{n - p - 1} \right\} \leq \alpha. \quad (10.112)$$

Meno usato, ma più sicuro, è il test a due code delle (6.124), che come sapete dalla discussione del cap. 6 (pagine 234 e seguenti), tutela anche dall'uso di modelli con troppi parametri, che tendono ad interpolare i dati (sovrapparametrizzazioni).

Se gli errori sono affidabili ma le variabili osservate non sono gaussiane, si ottengono in genere valori del χ^2 elevati, perché di solito i punti sono più dispersi di quanto previsto dalla curva normale. In questo caso alcune volte si accettano valori del χ^2 corrispondenti a livelli di significatività piccoli dal punto di vista puramente statistico (anche del per mille o meno), ma giustificabili dal punto di vista pratico (ovvero: si attribuisce il χ^2 cattivo più ai dati non gaussiani che alla forma del modello utilizzato).

Una visione approssimata ma immediata del risultato la si può anche vedere ad occhio prima di effettuare il test χ^2 : se i punti “toccano” la curva $y = f(x, \hat{\boldsymbol{\theta}})$ entro una, due o tre barre d'errore in accordo con la legge 3σ , allora il χ^2 probabilmente sarà buono. La Fig. 10.12 esemplifica la situazione; ricordate sempre che, per convenzione, la barra d'errore è sempre pari ad una σ .

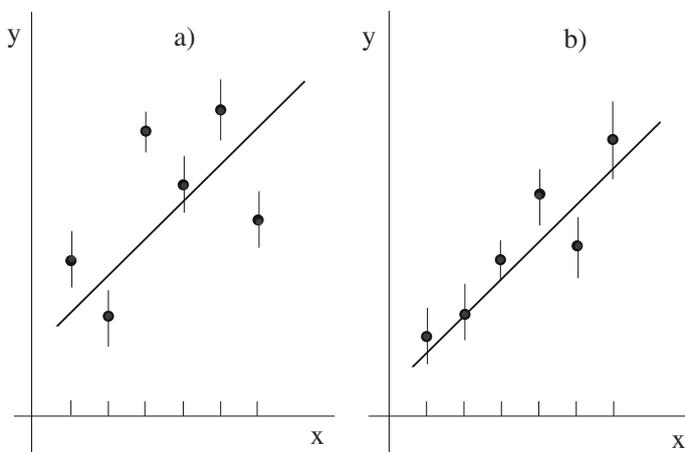


Fig. 10.12. Esempio di *fit* cattivo a) e di *fit* buono b); nel secondo caso il 68% circa dei punti tocca la curva di regressione entro una barra d'errore

La maggior parte dei codici di minimizzazione, se gli errori non vengono specificati, assume tacitamente che essi siano tutti uguali e che $\sigma_z = 1$, senza comunicare all'utente... i rischi che corre! Se ci troviamo effettivamente nella condizione di poter assumere $\sigma_i^2 = \sigma_z^2$ per ogni i , senza conoscere σ_z^2 , dovremo poi fornire una stima s_z tramite la (10.62) per poter calcolare gli errori delle stime. Questo procedimento si chiama riaggiustamento dell'errore. Ricordiamo che questo procedimento è valido solo se non vi sono dubbi né sul fatto che gli errori siano costanti, né sul modello funzionale $\mu_i(\mathbf{x}_i, \boldsymbol{\theta}) = \mathbf{x}_i^\dagger \boldsymbol{\theta}$ utilizzato. Inoltre, non ha senso fare, alla fine del *fit*, il test χ^2 , perché, inserendo s_z^2 nell'espressione del $\chi^2(\hat{\boldsymbol{\theta}})$ al posto di σ_z^2 , avremo sempre $\chi_{\sigma_z=s_z}^2(\hat{\boldsymbol{\theta}}) = n - p - 1$ (una costante!) per qualunque modello funzionale. Se non ci accorgessimo dell'errore, potremmo valutare come buono un *fit* che non lo è: il valore atteso $(n - p - 1)$ del $\chi^2(n - p - 1)$ si trova infatti ben al di sotto del valore critico. Per visualizzare il problema ricorriamo alla Fig. 10.13, dove supponiamo che i dati provengano da una parabola e che abbiamo errore costante, come in Fig. 10.13 a). Un *fit* che utilizzi la retta e che operi il riaggiustamento dell'errore fornirà un risultato come in Fig. 10.13 b), dove il χ^2 finale sarà ovviamente buono, ma solo perché avremo usato un errore più grande di quello vero. Il risultato giusto lo si ottiene solo rappresentando $f(x, \boldsymbol{\theta})$ come una parabola $\theta_0 + \theta_1 x + \theta_2 x^2$, il che presuppone naturalmente la conoscenza a priori della forma funzionale del modello.

Un altro grossolano errore è quello di fornire ai programmi solo i dati senza gli errori anche nel caso di errori variabili; il più delle volte questa procedura fornisce risultati sbagliati, come mostrato in Fig. 10.13 c).

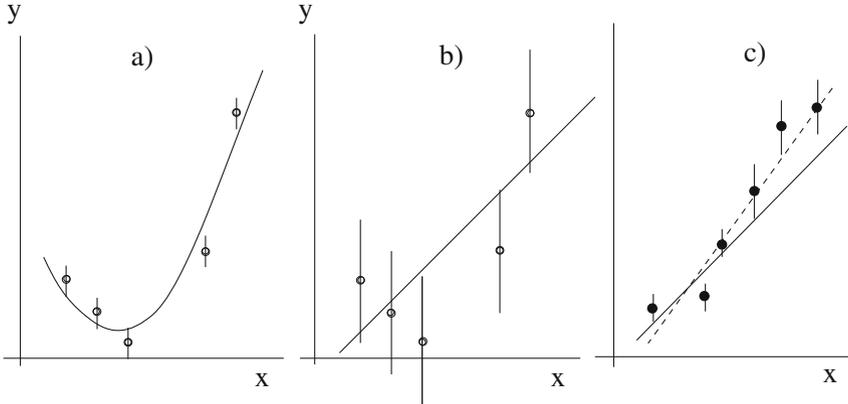


Fig. 10.13. Se i dati fluttuano intorno ad una parabola con errore costante come in a), un *fit* che assume come curva la retta e opera il riaggiustamento degli errori troverà un χ^2 buono sovrastimando scorrettamente l'errore come in b). In c) si mostra un *fit* lineare pesato di punti con errori diversi, che fornisce come risultato la retta in linea piena; la soluzione di un *fit* non pesato, che suppone errori tutti uguali, attribuisce la stessa importanza a tutti i punti e fornisce un risultato errato, dato dalla linea tratteggiata

10.12 Minimizzazione non lineare

Come abbiamo visto nel paragrafo precedente, a volte la funzione che si introduce nel χ^2 non è lineare nei parametri. Ne abbiamo avuto un esempio anche negli esercizi 9.6 e 9.7, dove il modello introdotto nella funzione da minimizzare (10.2) era rappresentato dalle densità gaussiane e binomiale.

Abbiamo a suo tempo già risolto questi esercizi, utilizzando la nostra routine di minimizzazione non lineare `Nlinfit`, commentandone dal punto di vista statistico i risultati. In questo paragrafo intendiamo descrivere sommariamente gli algoritmi utilizzati in questo codice per trovare il minimo della funzione $\chi^2(\boldsymbol{\theta})$ in uno spazio a $(p+1)$ -dimensioni, senza però entrare in eccessivi dettagli di calcolo numerico, che sono trattati in altri testi come [BR92] o [PFTW92] ed anche nelle nostre pagine web [RPP].

Gli algoritmi più efficienti sono basati sul metodo del gradiente negativo, perché *l'opposto del gradiente è un vettore che punta sempre verso il minimo*. Operando quindi per passi successivi, si va in un punto dove la derivata seconda è positiva; si continua così finché la funzione non riprende a crescere. Nell'intorno del minimo il χ^2 viene sviluppato in approssimazione parabolica:

$$\chi^2 \simeq \chi_0^2 + \sum_j \frac{\partial \chi^2}{\partial \theta_j} \theta_j + \frac{1}{2} \sum_{jk} \frac{\partial^2 \chi^2}{\partial \theta_j \partial \theta_k} \theta_j \theta_k . \quad (10.113)$$

Pe rffissare le idee, supponiamo che la funzione χ^2 sia del tipo della (10.6):

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_i^2} \equiv \sum_i H_i^2,$$

dove l'indice i si riferisce ai punti misurati; le derivate seconde assumono la forma:

$$\begin{aligned} \frac{\partial^2 \chi^2}{\partial \theta_j \partial \theta_k} &= \frac{\partial}{\partial \theta_j} \frac{\partial}{\partial \theta_k} \sum_i H_i^2 = \frac{\partial}{\partial \theta_j} \sum_i 2 H_i \frac{\partial H_i}{\partial \theta_k} \\ &= 2 \sum_i \frac{\partial H_i}{\partial \theta_j} \frac{\partial H_i}{\partial \theta_k} + 2 \sum_i H_i \frac{\partial^2 H_i}{\partial \theta_j \partial \theta_k}. \end{aligned} \quad (10.114)$$

Se si trascura il secondo termine, si ottiene l'approssimazione:

$$\frac{\partial^2 \chi^2}{\partial \theta_j \partial \theta_k} \simeq 2 \sum_i \frac{\partial H_i}{\partial \theta_j} \frac{\partial H_i}{\partial \theta_k} = 2 \sum_i \frac{1}{\sigma_i^2} \frac{\partial f_i}{\partial \theta_j} \frac{\partial f_i}{\partial \theta_k}. \quad (10.115)$$

L'esperienza ha mostrato che con questa approssimazione vi sono molti vantaggi: l'algoritmo risulta più rapido, accurato quanto gli algoritmi che ricorrono a sviluppi di ordine superiore e la *matrice delle derivate seconde è sempre definita positiva* [EDJ⁺71, Jam92].

È importante notare che annullare le derivate seconde rispetto ai parametri θ della funzione

$$H_i(\boldsymbol{\theta}) = \frac{y_i - f(x_i, \boldsymbol{\theta})}{\sigma_i}$$

corrisponde a supporre che le derivate seconde della funzione $f(x, \boldsymbol{\theta})$ siano nulle, ovvero che f sia esprimibile, nell'intorno di un valore di partenza $\boldsymbol{\theta}^*$, con uno sviluppo di Taylor al primo ordine *lineare nei parametri*:

$$f(x, \boldsymbol{\theta}) = f(x, \boldsymbol{\theta}^*) + \sum_k \left. \frac{\partial f}{\partial \theta_k} \right|_{\theta_k^*} \Delta \theta_k, \quad (10.116)$$

dove $\Delta \theta = (\boldsymbol{\theta} - \boldsymbol{\theta}^*)$. In questo caso la (10.113) è una relazione esatta, in quanto le derivate di χ^2 di ordine superiore al secondo sono nulle. Se ora teniamo presente le formule viste nel par. 10.5 e operiamo nelle (10.54-10.57) le sostituzioni:

$$\theta_k \rightarrow \Delta \theta_k = \theta_k - \theta_k^*, \quad y_i \rightarrow y_i - f(x_i, \boldsymbol{\theta}^*), \quad x_{ik} \rightarrow \left. \frac{\partial f(x_i)}{\partial \theta_k} \right|_{\theta_k^*},$$

vediamo che le equazioni risolutive, passo per passo, sono identiche alle (10.56-10.59). La matrice X della (10.53) contiene ora le $(p+1)$ derivate prime di $f(x, \boldsymbol{\theta})$ calcolate nel punto $\boldsymbol{\theta}^*$ e negli n punti sperimentali x_i , e la matrice $(X^\dagger W X)$ della (10.58), che in questo caso viene detta matrice di curvatura,

contiene le derivate seconde del χ^2 espresse come prodotti delle derivate prime di f e dei pesi $1/\sigma_i^2$. Una volta trovato il minimo rispetto ai valori $\Delta\theta_k$, il *fit* si sposta nel nuovo punto θ_k . Nel punto di minimo, gli errori sui parametri vengono ottenuti attraverso la matrice degli errori (10.61), come nel caso lineare. Per ulteriori dettagli, potete guardare la routine `Nlinfit` e le pagine di commento ad essa collegate.

Quando il χ^2 da minimizzare proviene da una funzione di verosimiglianza dalla (9.45) risulta:

$$\chi^2(\boldsymbol{\theta}) = -2 \ln L(\boldsymbol{\theta}) + C \equiv -2 \sum_i \ln f(x_i, \boldsymbol{\theta}) + C . \tag{10.117}$$

La ripetizione del calcolo appena fatto mostra che, linearizzando la funzione f , la derivata seconda del χ^2 vale:

$$\frac{\partial^2 \chi^2}{\partial \theta_j \partial \theta_k} \simeq 2 \sum_i \frac{1}{f_i^2} \frac{\partial f_i}{\partial \theta_j} \frac{\partial f_i}{\partial \theta_k} . \tag{10.118}$$

Si può mostrare che anche in questo caso la matrice delle derivate seconde è definita positiva [EDJ⁺71, Jam92].

Problemi

10.1. Trovare il coefficiente di correlazione lineare (4.31) quando $f(X) = a + bX$, X ed Z sono indipendenti.

10.2. Due grandezze X e Y , tra le quali esiste una relazione lineare, sono misurate con uno strumento di accuratezza pari al 10%:

x	10	20	30	40	50	60	70
y	21.4	38.8	52.2	88.1	99.5	120.4	158.3

Intendendo l'accuratezza come un intervallo che contiene il valore vero con $CL = 100\%$ e probabilità date dalla densità uniforme, determinare la retta di regressione utilizzando la routine `Rettamq` delle nostre pagine web. Calcolare il valore del χ^2 e commentare.

10.3. Due grandezze X e Y sono affette da un errore gaussiano:

x	10	20	30	40	50	60	70
s_x	0.3	0.6	0.9	1.2	1.4	1.7	2.0
y	20.5	40.0	63.6	86.7	104.3	123.3	144.7
s_y	0.6	1.1	1.9	2.5	3.1	3.5	4.1

Determinare la retta di regressione con la routine `Rettamq`. Calcolare il valore del χ^2 e commentare.

10.4. Il problema del vertice: determinare col metodo dei minimi quadrati, da un insieme di rette dei minimi quadrati $y = a_i + b_i x$, il vertice comune (x_0, y_0) . (*Suggerimento*: considerare l'equazione di un fascio di rette passanti per un punto: $(y - y_0)/(x - x_0) = b$).

10.5. Scrivere la matrice V (10.80) nel caso di errore sistematico proporzionale al valore y_i misurato ($y_i^S = \epsilon \cdot y_i$).

10.6. Una variabile gaussiana Y , misurata in funzione di X ha fornito i valori:

x	1.0	1.1	1.4	1.5	1.8	2.0	2.2	2.3	2.4	2.5
y	5.16	5.96	6.29	7.41	7.31	8.26	9.15	9.51	9.96	9.03

Sapendo che Y ha deviazione standard costante, determinare se è statisticamente compatibile un legame polinomiale fino al secondo grado tra X e Y . Si può utilizzare la routine `Pol2fit`.

10.7. Il campionamento di due variabili gaussiane X e Y ha dato il risultato:

x	1.271	0.697	2.568	2.400	2.879	2.465	2.472	2.039	2.277	1.392
y	6.05	3.57	13.88	13.77	15.77	13.61	13.86	11.30	12.77	6.38

Utilizzando nella routine `Pol2fit` un polinomio fino al secondo grado, determinare la funzione di correlazione tra le variabili.

10.8. Sono date le seguenti 5 misure y_i in funzione dei valori esatti x_i :

x	2	4	6	8	10
y	7.9	11.9	17.0	25.5	23.8

I valori di Y sono affetti da un errore relativo gaussiano pari a $\pm 10\%$. Determinare i coefficienti della relazione funzionale $Y = a + bX$ e studiare il risultato ottenuto simulando il *fit* 20 000 volte.

10.9. Sono date le seguenti 5 misure y_i in funzione dei valori x_i :

x	1.85	3.77	5.74	7.56	9.66
y	8.87	13.90	17.70	22.91	23.59

I valori di X e Y sono affetti da fluttuazioni relative di tipo uniforme di $\pm 10\%$. Determinare i coefficienti a e b della relazione funzionale $Y = a + bX$ e studiare il risultato ottenuto simulando il *fit* 20 000 volte.

Analisi dei dati sperimentali

Difatti, le conclusioni poggiavano su valori corrispondenti a uno o due punti situati in fondo alla curva, mentre esiste il principio secondo cui gli ultimi punti in fondo – l'ultimo in particolare – non valgono molto, altrimenti sarebbero seguiti da altri punti.

Richard P. Feynman, *Sta scherzando, Mr. Feynman!*

11.1 Introduzione

La parte tecnica e più estesa di questo capitolo riguarda l'applicazione delle nozioni di statistica e calcolo delle probabilità ai vari tipi di misure e di esperienze che vengono effettuate nei laboratori scientifici.

Lo scopo principale che ci proponiamo è quello di mettere in grado il ricercatore o il tecnico di laboratorio di riconoscere il tipo di misura che sta eseguendo e di calcolarne in modo appropriato l'accuratezza e la precisione.

Non affronteremo qui l'importante problema della ricerca di leggi fisiche tramite le tecniche di *best fit* dei dati, perché questo importante aspetto è stato ampiamente trattato nel cap. 10.

Accanto agli argomenti tecnici, affronteremo anche alcuni aspetti concettuali e metodologici molto importanti, che riguardano i fondamenti del metodo scientifico, cioè del metodo che ha permesso la nascita e lo sviluppo della scienza moderna.

Questo metodo è basato sull'*osservazione* dei fenomeni naturali, ovvero sulla raccolta e sulla analisi dei dati secondo quei principi e procedure che sono stati per la prima volta adottati in modo sistematico da Galileo Galilei e che si sono poi consolidati e perfezionati nel corso degli ultimi quattro secoli. L'analisi matematica e l'*analisi statistica dei dati* rivestono, in questo contesto, un ruolo di primaria importanza.

Oggi discipline come la fisica e la medicina sono considerate scienze, in quanto basate sui fatti sperimentali, mentre ciò non è più vero per l'astrologia, perché quest'ultima è basata sulle aspettative delle persone ed è totalmente

smentita dai fatti quando questi vengono analizzati, come nell'esercizio 3.17, col metodo scientifico.

Più sottile è la distinzione tra le scienze in senso lato, come la medicina, e le cosiddette scienze esatte, come la fisica. Più correttamente, la distinzione dovrebbe essere fatta tra scienze totalmente o parzialmente quantitative.

Una buona definizione di scienza esatta, se proprio vogliamo usare questo termine, potrebbe essere la seguente:

Affermazione 11.1 (Scienza esatta). *Una scienza si dice esatta quando è sempre in grado di associare un errore alle proprie previsioni e risultati.*

In sostanza, il termine “esatto” non implica che i risultati debbano essere affetti da errore nullo o trascurabile; esso è invece sinonimo del termine “quantitativo”, che sta ad indicare un metodo che fornisce risultati prevedibili con certezza o *con livelli di confidenza affidabili*. Come già sapete, per la determinazione dei livelli di confidenza il calcolo dell'incertezza di misura è di fondamentale importanza.

Dal punto di vista teorico, gli errori sono talvolta presenti quando si utilizzano modelli semplificati del fenomeno che si sta studiando oppure si eseguono i calcoli con metodi numerici approssimati; tuttavia, qui non approfondiremo questo aspetto della questione.

Dal punto di vista sperimentale, che è invece quello di cui vogliamo occuparci, gli errori derivano dagli effetti casuali o sistematici legati alle operazioni di misura. Spesso la parte più difficile e laboriosa di un esperimento è proprio la valutazione degli errori. In questa fase lo sperimentatore non è guidato tanto da teoremi o regole precise, quanto piuttosto dall'esperienza e da una serie di assunzioni, a volte anche un po' arbitrarie. Esiste tuttavia un vincolo importante: queste assunzioni e regole pratiche *devono sempre essere in accordo con le leggi fondamentali del calcolo delle probabilità e della statistica*. Notiamo anche che alcune delle assunzioni soggettive e arbitrarie che faremo riguardano *la forma delle distribuzioni statistiche da associare a priori al comportamento degli strumenti o alle operazioni di misura che vengono fatte*. Per un approfondimento delle conseguenze legate a queste scelte si può leggere [D'A99].

Ricordiamo infine che in questo capitolo useremo, per gli intervalli di confidenza, la notazione (6.16) di pagina 188, e che il significato dei livelli di confidenza associati a questi intervalli è quello frequentista ampiamente discusso nel par. 6.2 a pagina 181. Queste sono le convenzioni attualmente usate nella letteratura scientifica internazionale per i risultati di esperimenti di laboratorio.

11.2 Terminologia

L'operazione di misura di una grandezza con un apparato può essere schematizzata come in Fig. 11.1. Come si vede, le incertezze che caratterizzano la misura possono essere riferite all'oggetto, allo strumento oppure all'interazio-

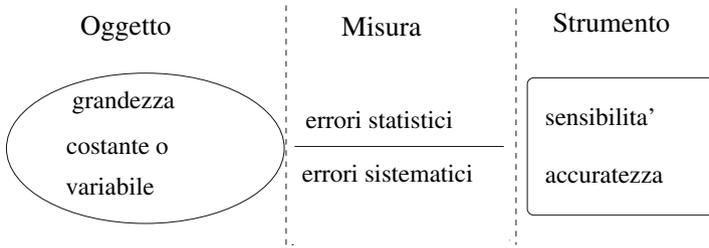


Fig. 11.1. Descrizione schematica dell'operazione di misura

ne tra oggetto e strumento. Attualmente non esiste una terminologia unica per la descrizione di tali incertezze. Le raccomandazioni ISO [fSI93] sono di classificarle in due tipi: quelle che si possono trattare con metodi solamente statistici e quelle (dette sistematiche) che vanno trattate con altri metodi. La nomenclatura di uso corrente fa uso dei sinonimi:

incertezza $\begin{cases} \nearrow \text{incertezza statistica, errore statistico, errore casuale} \\ \searrow \text{incertezza sistematica, errore sistematico, errore strumentale} \end{cases}$

Noi operiamo la scelta seguente, alla quale ci manterremo fedeli per tutto il capitolo:

$$\text{incertezza} \begin{cases} \nearrow \text{errore statistico} \\ \searrow \text{errore sistematico} \end{cases} \quad (11.1)$$

Esamineremo ora in dettaglio tutti i casi che si possono presentare.

11.3 Grandezze fisiche costanti e variabili

Quando si inizia un esperimento, prima di tutto occorre chiedersi se la grandezza che si vuole misurare è una costante oppure se è una variabile aleatoria.

Nel primo caso, se durante l'esperimento si osservano delle fluttuazioni (risultati diversi ad ogni prova, pur mantenendo costanti le condizioni sperimentali), esse sono da attribuire alle operazioni di misura oppure al funzionamento degli strumenti impiegati.

Nel secondo caso, le fluttuazioni osservate conterranno anche quelle dell'oggetto in misura. Questa componente delle fluttuazioni contiene le informazioni fisiche sulla legge statistica (2.7) della grandezza che si sta misurando.

La situazione si riassume nelle due definizioni operative:

Affermazione 11.2 (Grandezza fisica costante). *Una grandezza fisica viene detta costante quando essa è una costante fisica universale, cioè una grandezza che ha lo stesso valore in tutti i sistemi di riferimento (costante*

di Planck, carica dell'elettrone, massa a riposo di una particella stabile), oppure una grandezza che ragionevolmente si può considerare costante e stabile rispetto alle operazioni della misura che si sta effettuando.

Affermazione 11.3 (Grandezza fisica variabile). *Una grandezza fisica si dice variabile quando essa presenta delle fluttuazioni e variazioni misurabili che sono intrinseche al processo fisico che si sta studiando. Molto spesso le fluttuazioni sono puramente statistiche, ed allora la grandezza è una variabile aleatoria che possiede una certa distribuzione. Scopo della misura è proprio la determinazione di questa distribuzione.*

Alcuni esempi di variabili fisiche di tipo aleatorio sono:

- la velocità della molecola di un gas (densità χ^2 con 3 gradi di libertà, detta anche densità di Maxwell) (si veda l'esercizio 3.10);
- il numero di raggi cosmici al secondo (circa cento) che attraversano il vostro corpo mentre state leggendo questa pagina (legge di Poisson);
- il numero di elettroni che attraversano, entro un tempo prefissato, la sezione di un filo percorso da corrente;
- in genere, tutte le grandezze oggetto di studio della meccanica e della fisica statistica.

11.4 Sensibilità e accuratezza degli strumenti

Il comportamento di uno strumento è definito da due importanti caratteristiche, la sensibilità (detta anche risoluzione) e l'accuratezza.

Affermazione 11.4 (Sensibilità (risoluzione) di uno strumento). *Se uno strumento fornisce il valore x nella misura di una grandezza, si definisce come intervallo di sensibilità o risoluzione e si indica con Δx , la minima quantità necessaria per spostare il risultato della misura dal valore x ad uno contiguo. Si dice sensibilità dello strumento il rapporto*

$$S = \frac{1}{\Delta x} .$$

In uno strumento ideale, ad alta sensibilità, $\Delta x \simeq 0$ ed $S \gg 1$. Se Δy è la variazione nella lettura in corrispondenza di una variazione Δx della quantità da misurare, la sensibilità si può definire anche come $S = \Delta y / \Delta x$.

Negli strumenti digitali l'intervallo di sensibilità è definibile chiaramente, poiché in questo caso esso non è altro che un errore di arrotondamento. Ad esempio, se un multimetro digitale costruito a regola d'arte indica una tensione di 2.15 V, il valore vero sarà compreso tra 2.145 e 2.155 V (vedi Fig. 11.2a), poiché è lecito presumere che le operazioni di arrotondamento vengano effettuate correttamente. L'intervallo di sensibilità è in questo caso $\Delta = (2.155 - 2.145) V = 10 mV$.

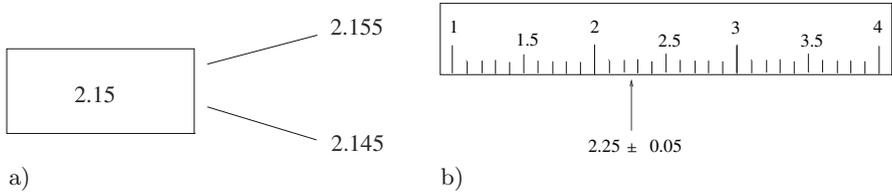


Fig. 11.2. a) Negli strumenti digitali l'intervallo di sensibilità diventa un errore di arrotondamento. b) Negli strumenti analogici l'intervallo di sensibilità è dato dall'ampiezza della tacca minima sulla scala. Il valore misurato viene assunto come il punto medio della tacca entro cui cade la misura

Poiché in genere non c'è correlazione tra l'intervallo di sensibilità e la localizzazione del valore vero all'interno di questo intervallo, possiamo dire che, se x è il valore letto e Δx è l'intervallo di sensibilità, il valore vero sarà compreso, con legge di probabilità uniforme, nell'intervallo $x \pm \Delta x/2$, con un livello di confidenza del 100%. In sintesi, si suppone che il valore vero sia localizzato in modo equiprobabile entro l'intervallo:

$$\text{valore vero} = x \pm \frac{\Delta x}{2}, \quad (CL = 100\%) . \quad (11.2)$$

Nel caso di Fig. 11.2a), diremo che il valore vero sta nell'intervallo:

$$2.150 \pm 0.005 \text{ V} ,$$

che l'intervallo di sensibilità è di 10 mV e che la sensibilità è $S = 100 \text{ V}^{-1}$, cioè che 1 V sposta la lettura di 100 posizioni.

Negli strumenti analogici ad ago, l'intervallo di sensibilità ha una definizione meno chiara: l'ago si sposta in modo continuo, ma lo spessore stesso dell'ago e la larghezza delle tacche di misura, disegnate sul quadrante dello strumento, definiscono un intervallo minimo di lettura al di sotto del quale non ha senso procedere. Generalmente il risultato viene ancora riportato nella forma (11.2), dove il valore della misura è il punto medio delle tacca entro cui si legge il valore e l'errore è la ampiezza della tacca minima, centrata sul valore letto (vedere Fig. 11.2b).

Spesso si tende ad interpolare ad occhio la lettura, ed a ridurre così l'errore. La procedura è lecita, ma in questo caso si deve essere consci che si sostituisce, alla scala dello strumento, una scala soggettiva (più fine) ottenuta interpolando visivamente tra le tacche segnate sul quadrante dello strumento. In questo caso si può associare all'intervallo (11.2) non la legge uniforme, ma quella triangolare (5.35), centrata sul valore letto e di larghezza pari all'intervallo della tacca minima "ideale" trovata ad occhio. Al posto della interpolazione visiva, è comunque meglio in questi casi cercare di utilizzare, se disponibile, uno strumento digitale a più alta sensibilità.

Oltre alla sensibilità, l'altro parametro fondamentale che caratterizza uno strumento è l'accuratezza. Essa è definita come:

Affermazione 11.5 (Accuratezza o errore sistematico). *Per uno strumento di sensibilità ideale, si definisce accuratezza, o errore sistematico, lo scostamento tra il valore misurato dallo strumento ed il valore vero. Questo effetto non è di tipo casuale ed ha origine da una taratura imperfetta dello strumento. Nel seguito esso verrà indicato con δ .*

Veniamo ora ad un punto importante: come si combinano, in uno strumento, gli errori dovuti alla sensibilità e quelli di accuratezza?

Se il multimetro digitale immaginato in Fig. 11.2a) avesse una taratura perfetta, se cioè fosse $\delta \ll \Delta x$, il valore vero sarebbe certamente localizzato nell'intervallo (11.2). Se invece vi fosse un difetto di taratura (errore di accuratezza), poniamo di 30 mV , allora $\delta \gg \Delta x$ e l'intervallo (11.2) sarebbe privo di significato.

In genere, negli strumenti scientifici professionali in buono stato e prodotti a regola d'arte, si fa in modo che l'intervallo di accuratezza sia sempre minore o al più dello stesso ordine di quello di sensibilità. Questi strumenti sono corredati da una tavola di accuratezza (*table of accuracy*), dove vengono date le norme per definire un intervallo globale Δ , comprensivo degli errori di sensibilità e accuratezza, per il quale è valida l'equazione (11.2). In questo caso Δ indica un intervallo globale

$$\delta + \Delta \simeq \Delta(\text{syst}) , \quad (11.3)$$

che viene chiamato errore strumentale o errore sistematico (*systematic error*). Questo tipo di errore non può essere eliminato a priori dalla misura, come in genere avviene per gli errori sistematici di altro tipo, che discuteremo nel prossimo paragrafo.

Se vi sono errori di sensibilità, è ragionevole supporre il valore vero localizzato in modo equiprobabile nell'intervallo (11.2); se invece è presente anche una componente di accuratezza, questo a rigore non è più vero, perché il difetto di taratura provoca in genere uno scostamento costante e correlato tra il valore vero e quello misurato. Tuttavia, in mancanza di informazioni più dettagliate, all'errore sistematico (11.3) si associa in genere la densità uniforme.

In sede internazionale gli strumenti sono stati divisi in classi di accuratezza, definite in base all'errore sistematico relativo:

<i>Classi di Accuratezza</i>					
CLASSE	0.2	0.5	1	1.5	2.5
$\Delta x/x$	0.2 %	0.5 %	1 %	1.5 %	2.5 %

Diremo quindi che uno strumento è di classe 1 se il suo errore sistematico totale non supera l'1%.

11.5 Incertezza nelle operazioni di misura

Proseguendo nell'analisi dello schema di Fig. 11.1, ci occupiamo ora dell'analisi delle operazioni di misura.

In questa fase, che riguarda l'interazione tra tutto l'apparato sperimentale (che può comprendere anche l'osservatore) e le grandezze che si stanno misurando, intervengono due tipi di errori, statistici e sistematici.

Gli errori statistici sono stati ampiamente trattati nel cap. 6; nelle misure di laboratorio essi si presentano quando vengono impiegati strumenti di sensibilità molto elevata, tali da rendere critica la stabilizzazione delle operazioni sperimentali o le operazioni stesse di misura. Se si misura la lunghezza di un banco da officina con un metro da falegname si ottiene, ripetendo più volte la misura, sempre il medesimo valore e non vi sono errori statistici. Se invece si usano strumenti ottici ad altissima sensibilità, la sovrapposizione di numerose fluttuazioni (nel posizionamento, nella taratura od altro) fa sì che ad ogni misura si ottenga un valore leggermente diverso. Abbiamo in questo caso uno spettro di valori sperimentali, ed il valore della lunghezza misurata diventa una variabile aleatoria.

Se le fluttuazioni che influenzano il processo di misura sono numerose, si sovrappongono linearmente e nessuna di esse è preponderante sulle altre, allora valgono le condizioni del teorema Limite Centrale 3.1 e le misure tendono a distribuirsi secondo la curva di Gauss.

Solitamente si suppone che, in presenza dei soli errori statistici, la media delle misure debba tendere al valore vero (badate bene: questa è una assunzione!). L'errore statistico che, come sappiamo, è la stima della deviazione standard della distribuzione delle misure, caratterizza la *precisione* della misura.

Gli errori sistematici sono invece dovuti in questo caso ad operazioni errate oppure ad ipotesi sbagliate sul modello fisico su cui si basa la misura (ad esempio, trattare le grandi oscillazioni del pendolo con un modello lineare). L'effetto degli errori sistematici è quello di introdurre uno scostamento *non casuale* tra valore misurato e valore vero.

In linea di principio, tutte le possibili fonti di errore sistematico devono essere eliminate *prima* di iniziare la misura. Se questo non è possibile, alla *fine* della misura, ma *prima* di effettuare l'eventuale analisi statistica dei dati, si devono apportare le correzioni sistematiche ai valori osservati, ricorrendo ad equazioni basate su modelli fisici o ad altri metodi. Se, ad esempio, ci siamo accorti, alla fine di una misura, che gli strumenti hanno avuto effetti di deriva dovuti alla temperatura, dovremo studiare quantitativamente questi effetti ed elaborare delle equazioni con le quali correggere tutti i dati osservati, oppure dovremo ripetere l'esperimento stabilizzando termicamente tutto l'apparato.

La distinzione tra errori statistici e sistematici non è netta; se leggiamo uno strumento ad ago e cerchiamo correttamente di minimizzare l'errore di parallasse, otterremo una serie di letture diverse, la cui media sarà più o meno coincidente, per grandi campioni, col valore vero della misura. In questo

caso l'errore di parallasse è di tipo statistico. Se invece leggiamo sempre lo strumento di sbieco, la media delle misure si discosterà sempre dal valore vero, dando luogo alla presenza di un errore sistematico.

Tutta questa discussione sugli errori statistici e sistematici può essere condensata in una serie di definizioni fondamentali:

Affermazione 11.6 (Errore statistico). *È quel tipo di errore, dovuto alle operazioni di misura, che fa variare i risultati secondo una certa distribuzione statistica. La media di questa distribuzione (media vera) coincide col valore vero della grandezza fisica che si misura. La deviazione standard della distribuzione è l'errore di misura. Queste due grandezze sono stimate dalla media e dalla deviazione standard del campione sperimentale delle misure.*

Affermazione 11.7 (Precisione). *La precisione è determinata dall'errore statistico della misura, che è dato dalla deviazione standard stimata dal campione delle misure effettuate. Una misura è tanto più precisa quanto più è piccolo l'errore statistico.*

Affermazione 11.8 (Errore sistematico). *È quel tipo di errore che fa scostare la media dei valori misurati dal valore vero, indipendentemente dal numero di misure che si effettuano. In altre parole, gli errori sistematici provocano uno scostamento tra la media della distribuzione delle misure e il valore vero della grandezza.*

Affermazione 11.9 (Accuratezza). *L'accuratezza è determinata dagli errori sistematici della misura. Una misura è tanto più accurata quanto più sono piccoli gli errori sistematici.*

Precisione e accuratezza si possono rappresentare in modo schematico ma efficace indicando, come in Fig. 11.3, l'esperimento come un bersaglio il cui centro rappresenta il valore vero della misura. I risultati delle misure si possono allora rappresentare come dei punti caduti sul bersaglio. Una misura né precisa né accurata si può rappresentare come un insieme di punti dispersi, con un baricentro (media) differente dal valore vero (il centro del bersaglio); una misura precisa ma poco accurata vede i risultati della misura addensarsi intorno ad un valore, che tuttavia può differire di molto dal valore vero; una misura accurata ma poco precisa presenta un insieme di dati notevolmente dispersi ma aventi la media vicina al valore vero; infine, una misura precisa e accurata presenta poca dispersione nei dati, e questi si raggruppano intorno al valore vero. È evidente che solo una misura accurata (precisa o meno) è una buona misura, in quanto il valor medio dei dati è una stima corretta della media vera. In questo caso, come mostreremo tra poco nel par. 11.9, l'errore sulla media può essere ridotto aumentando il numero di misure.

L'accuratezza è definibile anche per una sola misura come la differenza tra il singolo valore misurato e quello vero, mentre la precisione, che è data dalla dispersione tra misure ripetute, perde in questo caso di significato. Pertanto, una singola misura sarà accurata se, nell'unità di misura scelta, risulta vicina

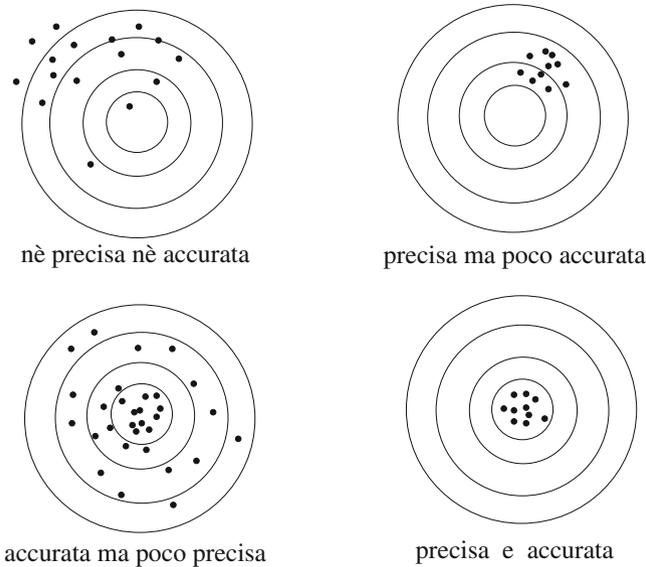


Fig. 11.3. Rappresentazione degli effetti degli errori sistematici (accuratezza) e statistici (precisione) in una misura. Il valore vero è rappresentato dal centro del bersaglio, mentre le misure sono rappresentate dai punti

al valore vero, non accurata se ne è distante. Un buon esempio per distinguere tra accuratezza e precisione è quello dell'orologio al quarzo: se l'orologio viene regolato con "l'ora esatta", nei giorni seguenti differirà di poco da questo valore ed avremo una misura del tempo precisa ed accurata. Se invece regoliamo l'orologio cinque minuti più avanti dell'ora esatta, avremo una misura precisa ma non accurata.

Supponete ora di non conoscere l'ora esatta oppure, il che è lo stesso, di togliere gli anelli concentrici centrati sul valore vero in Fig. 11.3. In questo caso siamo in grado di giudicare se la misura è precisa, *ma non se è accurata*; in altri termini, le situazioni della prima e della seconda riga di Fig. 11.3 ci appariranno indistinguibili.

Sapere quanto una misura è accurata richiede la conoscenza del valore vero, che però è lo scopo della misura! Questa, come ben sanno tutti gli sperimentatori, è la difficoltà più grande che si incontra nelle misure di laboratorio. Occorre quindi conoscere bene il proprio apparato ed i metodi di trattamento dei dati che si utilizzano, in modo da essere ragionevolmente certi di aver eliminato *a priori* gli errori sistematici, oppure di saperli valutare.

Nel corso del capitolo daremo alcuni esempi ed approfondiremo ulteriormente questi importanti aspetti.

11.6 Misure indirette e propagazione degli errori

Una grandezza si dice misurata in modo indiretto quando è una funzione $z = f(x, y, w, \dots)$ di una o più grandezze misurate direttamente e affette da incertezza.

La determinazione dell'incertezza su z a partire da quella delle grandezze misurate si chiama propagazione degli errori. Nel seguito partiremo il caso semplice $z = f(x, y)$ che si può estendere senza difficoltà ad un numero qualunque di variabili.

Se x e y sono affette solo da errori statistici e sono indipendenti, per effettuare la propagazione degli errori basta utilizzare la formula (5.65) o la sua generalizzazione in n variabili (5.71), sostituendo alle deviazioni standard gli errori di misura s_x e s_y , come indicato nell'affermazione 11.7:

$$s_f^2 = \left(\frac{df}{dx}\right)^2 s_x^2 + \left(\frac{df}{dy}\right)^2 s_y^2 . \quad (11.4)$$

Ovviamente, con n misure tra loro indipendenti, si ha:

$$s_f^2 = \sum_{i=1}^n \left(\frac{df}{dx_i}\right)^2 s^2(x_i) . \quad (11.5)$$

Questa equazione, che è la famosa legge di propagazione degli errori di misura, è valida solo in approssimazione lineare, come ampiamente discusso nel cap. 5 e, in particolare, nel par. 5.4. Inoltre, la stima della deviazione standard risultante definisce un intervallo di confidenza gaussiano solo se tutte le variabili sono gaussiane, come è stato mostrato nell'esercizio 5.3. Tuttavia, gli intervalli tendono ad essere approssimativamente gaussiani anche quando la funzione f dipende da un numero grande ($> 5-10$) di variabili aleatorie qualunque.

Veniamo ora alle incertezze strumentali, alle quali spesso si attribuisce, come mostrato nella (11.2), la densità uniforme. In questo caso a volte si usa, come legge di propagazione degli errori di due misure lo sviluppo di Taylor al primo ordine con le derivate prese in valore assoluto:

$$\Delta_f = \left|\frac{\partial f}{\partial x}\right| \Delta_x + \left|\frac{\partial f}{\partial y}\right| \Delta_y . \quad (11.6)$$

La generalizzazione al caso di n misure è ovviamente:

$$\Delta_f = \sum_{i=1}^n \left|\frac{\partial f}{\partial x_i}\right| \Delta_i . \quad (11.7)$$

Queste formule non rappresentano la deviazione standard della distribuzione risultante, ma la sua larghezza totale quando la composizione delle misure è lineare. Per questo motivo esse in genere vengono usate per una stima del limite superiore dell'errore, il che può essere utile nel caso di grandezze correlate in modo complicato. Il valore assoluto delle derivate assicura che l'effetto

della propagazione venga sempre preso nel modo più “sfavorevole”, quello corrispondente ad un incremento positivo dell’errore di misura.

Se le variabili non sono correlate, allora un modo più corretto di procedere è quello di applicare le (11.4, 11.5) al caso della densità uniforme, la cui varianza, in base alla (3.82), vale $\Delta^2/12$. Per due ed n variabili si ha rispettivamente:

$$s_f^2 = \gamma \left[\left(\frac{\partial f}{\partial x} \right)^2 \Delta_x^2 + \left(\frac{\partial f}{\partial y} \right)^2 \Delta_y^2 \right], \quad \gamma = \frac{1}{12}, \quad (11.8)$$

$$s_f^2 = \gamma \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \Delta_i^2, \quad \gamma = \frac{1}{12}, \quad (11.9)$$

da cui si vede che, per ottenere la varianza totale, basta *sommare quadraticamente gli errori sistematici e poi dividere per 12*. La varianza è stata qui indicata con lettere latine, in quanto si tratta pur sempre di una quantità stimata dai dati nell’assunzione *che sia valida l’ipotesi a priori della densità uniforme per gli errori sistematici*.

Queste varianze non vanno però associate alla densità gaussiana, anche nel caso che la funzione $f(x, y)$ combini linearmente le variabili. Infatti, mentre la combinazione lineare di errori gaussiani mantiene intervalli di confidenza gaussiani, in questo caso la somma di due o più errori sistematici, cioè di due o più variabili uniformi, dà luogo a densità differenti, che dipendono dal numero di errori sommati. Sappiamo però, dal teorema Limite Centrale 3.1, che questa densità tende rapidamente ad una gaussiana. Si pone ora il problema pratico di stabilire a partire da quale numero di misure combinate linearmente sia lecito usare l’approssimazione gaussiana. Il risultato che troveremo probabilmente vi sorprenderà.

Affrontiamo il problema trattando in modo completo il caso istruttivo della somma di due errori sistematici di pari valore.

La semplice generalizzazione dell’esercizio 5.2 di pagina 157 indica che la somma di due errori sistematici uguali definiti in $[0, \Delta/2]$ segue la densità triangolare in $[0, \Delta]$:

$$p(x) = \begin{cases} \frac{4}{\Delta^2}x & \text{per } 0 \leq x \leq \frac{\Delta}{2} \\ \frac{4}{\Delta^2}(\Delta - x) & \text{per } \frac{\Delta}{2} < x \leq \Delta \\ 0 & \text{altrimenti,} \end{cases} \quad (11.10)$$

di parametri

$$\mu = \frac{\Delta}{2}, \quad \sigma^2 = \frac{\Delta^2}{24}, \quad \sigma = \frac{\Delta}{2\sqrt{6}}. \quad (11.11)$$

Essa è rappresentata in Fig. 11.4 per $\Delta/2 = 1$. Se i due errori sistematici che si sommano sono definiti in $[0, 1]$, la (11.6) fornisce il valore

$$\Delta_f = \Delta_x + \Delta_y = 2,$$

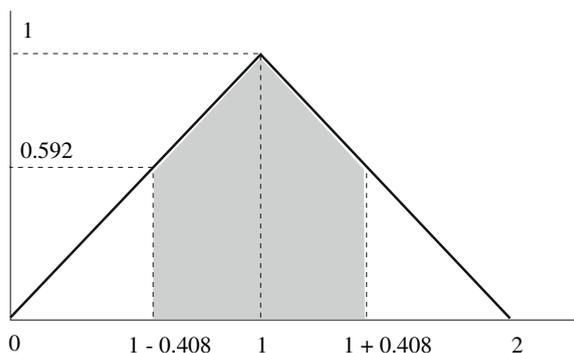


Fig. 11.4. Densità triangolare della somma di due misure affette da errori sistematici uguali compresi in $[0, 1]$. L'area ombreggiata è il livello di confidenza corrispondente all'intervallo $\mu \pm \sigma$ e vale $\simeq 65\%$

che è proprio la larghezza totale della distribuzione. Invece, dalla (11.8) abbiamo:

$$s_f = \frac{1}{\sqrt{12}} \sqrt{\Delta_x^2 + \Delta_y^2} = \frac{1}{\sqrt{6}} = 0.408, \quad (11.12)$$

che è un risultato esatto in accordo con la (11.11), perché stiamo considerando la somma di due errori. L'area corrispondente all'intervallo $|x - \mu| \leq K\sigma$, con K numero reale, può essere trovata in modo elementare dalla Fig. 11.4, oppure, in modo più formale:

$$\begin{aligned}
 P\{|X - \mu| \leq K\sigma\} &= \int_{\mu - K\sigma}^{\mu} x \, dx + \int_{\mu}^{\mu + K\sigma} (2 - x) \, dx \quad (11.13) \\
 &= \begin{cases} K\sigma(2 - K\sigma) & \text{per } K\sigma \leq 1 \\ 1 & \text{per } K\sigma > 1 \end{cases} = \begin{cases} 0.649 & \text{per } K = 1 \\ 0.966 & \text{per } K = 2 \\ 1 & \text{per } K = 3 \end{cases}
 \end{aligned}$$

dove nell'ultimo passaggio si è utilizzato il valore della (11.12), $\sigma \simeq s_f = 0.408$. Questi valori sono molto vicini agli intervalli di probabilità gaussiani della (3.35). Il fatto sorprendente che vi avevamo preannunciato è che in genere gli sperimentatori, sulla base di questo risultato, *considerano intervalli di confidenza gaussiani già a partire da errori ottenuti dalla combinazione lineare di due soli errori sistematici*.

Questo risultato giustifica anche l'assunzione di una densità triangolare, che talvolta viene fatta per certi tipi di errori sistematici, che provengono da test sperimentali o simulazioni che combinano tra loro pochi errori sistematici. In questo caso, in base alla (11.11), nelle (11.8, 11.9) si deve porre $\gamma = 1/24$.

Veniamo ora al caso della combinazione di errori statistici e sistematici. Per trovare la densità di probabilità del risultato, nel caso più semplice dato dalla sovrapposizione lineare di due misure, una con errore gaussiano e l'altra

con errore sistematico uniforme entro un intervallo, basta rifarsi alla formula (5.34) ottenuta nell'esercizio 5.1.

Se immaginiamo l'errore sistematico come un intervallo ($a = -\Delta/2, b = +\Delta/2$), ed introduciamo la variabile standard $t = (z - \mu)/\sigma$, dove σ è in questo caso l'errore statistico vero, tenendo conto che $dz = \sigma dt$, la formula (5.34) diventa:

$$p(t) = \frac{\sigma}{\Delta} \left[\Phi \left(\frac{\Delta}{2\sigma} - t \right) - \Phi \left(-\frac{\Delta}{2\sigma} - t \right) \right] .$$

In base alle (3.40, 3.44) possiamo anche scrivere:

$$p(t) = \frac{\sigma}{\Delta} \left[E \left(t + \frac{\Delta}{2\sigma} \right) - E \left(t - \frac{\Delta}{2\sigma} \right) \right] . \tag{11.14}$$

Questa densità è una funzione pari rispetto all'origine, poiché è la convoluzione di due funzioni pari. Essa ha una deviazione standard che vale:

$$\sigma_m = \sqrt{\sigma^2 + \Delta^2/12} . \tag{11.15}$$

I livelli di probabilità si possono trovare integrando la (11.14). La Tab. 11.1 riporta i risultati ottenuti con la nostra routine `Stasys(f,sigma,delta)`, che calcola con la regola trapezoidale

$$\int_{-a}^a p(t) dt , \text{ dove } a = f \sqrt{\text{sigma}^2 + \text{delta}^2/12} \equiv f \sigma_m .$$

La tabella mostra che i risultati, per $\sigma = \Delta$, cioè per $\sigma > \Delta/\sqrt{12}$, coincidono con i livelli gaussiani (prima riga), mentre, per $\sigma \ll \Delta$ essi tendono a quelli della distribuzione uniforme (ultima riga). Per i casi intermedi, tutto sommato *il risultato non si discosta molto dai livelli standard gaussiani.*

Tabella 11.1. Probabilità degli intervalli di confidenza per misure con errori statistici combinati linearmente con errori sistematici. Gli intervalli sono parametrizzati in funzione del rapporto Δ/σ e della deviazione standard $\sigma_m = \sqrt{\sigma^2 + \Delta^2/12}$, dove σ è l'errore statistico e Δ è la larghezza totale dell'errore sistematico della misura

Δ/σ	$\pm\sigma_m$	$\pm 2\sigma_m$	$\pm 3\sigma_m$
1.0	68.3	95.4	99.7
3.0	67.1	95.8	99.8
5.0	64.9	96.7	100.0
10.0	60.9	98.6	100.0
100.0	57.8	100.0	100.0

I risultati calcolati con le (11.13, 11.14) possono essere ottenuti con semplici codici di simulazione, come nel caso della nostra routine `Sigdel(sigma, delta)`:

```
function Sigdel(sigma,delta)

// Studio della legge 3-sigma nel caso della somma di
// errori statistici gaussiani di deviazione standard 'sigma'
// e di errori sistematici di ampiezza totale 'delta'

// misure w somma di una misura gaussiana y e una uniforme z
Nev=50000;
y=rand(1,Nev,'norm')*sigma;
z=-0.5*delta + delta*rand(1,Nev,'unif');
w=y+z;
// display dell'istogramma
xbase(); Histplote(20,w,stat=1);
// conteggio misure entro i sigma-intervalli
stot = sqrt(sigma^2+delta^2/12);
cum1=0; cum2=0; cum3=0;
for j=1:Nev,
    if(- stot<w(j) & w(j)< stot) then cum1=1+cum1; end;
    if(-2*stot<w(j) & w(j)<2*stot) then cum2=1+cum2; end;
    if(-3*stot<w(j) & w(j)<3*stot) then cum3=1+cum3; end;
end;
// stampa dei risultati con errore statistico
write(%io(2),'dev. standard calcolata = '+string(stot));
write(%io(2),'entro 1 sigma = ('+string(cum1*100/Nev)+'+'..
    +part(string(sqrt(cum1*(1-cum1/Nev))*100/Nev),1:4)+'')%');
write(%io(2),'entro 2 sigma = ('+string(cum2*100/Nev)+'+'..
    +part(string(sqrt(cum2*(1-cum2/Nev))*100/Nev),1:4)+'')%');
write(%io(2),'entro 3 sigma = ('+string(cum3*100/Nev)+'+'..
    +part(string(sqrt(cum3*(1-cum3/Nev))*100/Nev),1:4)+'')%');
endfunction
```

I risultati che si ottengono coincidono, entro l'errore statistico, con quelli di Tab. 11.1.

In genere il metodo della propagazione lineare fornisce buoni risultati anche nel caso di prodotti o rapporti di misure. In questo caso la (11.5) dà luogo alla *propagazione lineare degli errori percentuali*, che viene sovente utilizzata. Infatti, ricordando la (5.67), se sostituiamo alle deviazioni standard vere gli errori statistici, possiamo scrivere, con ovvia notazione:

$$Z = X_1 X_2, \quad Z = \frac{X_1}{X_2}, \quad Z = \frac{X_2}{X_1} \quad \implies \quad \frac{s_z^2}{z^2} = \frac{s_1^2}{x_1^2} + \frac{s_2^2}{x_2^2}, \quad (11.16)$$

la quale può anche essere agevolmente ricavata dalla (11.4), ed è immediatamente generalizzabile al caso di n variabili. In sostanza, nel prodotto e nella divisione si sommano le *varianze percentuali o relative*.

La stessa proprietà, come è facile dimostrare dalla (11.6), vale anche per gli errori massimi:

$$Z = X_1 X_2, \quad Z = \frac{X_1}{X_2}, \quad Z = \frac{X_2}{X_1} \quad \implies \quad \frac{\Delta_z}{z} = \frac{\Delta_1}{x_1} + \frac{\Delta_2}{x_2}. \quad (11.17)$$

Tuttavia, la propagazione quadratica degli errori statistici relativi (11.16) e quella lineare degli errori massimi sistematici (11.17) vanno usate con cautela, perché sono valide solo nel caso in cui le misure x_1 e x_2 siano indipendenti. Ad esempio, nella propagazione dell'errore nel rapporto $x/(x+y)$, è facile vedere che la (11.4) e la (11.16) danno risultati diversi. In questo caso solo la (11.4) è corretta, perché la presenza della variabile x sia al numeratore sia al denominatore crea, pur essendo le misure x e y indipendenti, una correlazione nel rapporto.

Per verificare la bontà della (11.16) in un caso particolare, consideriamo la variabile $Z = XY$ prodotto di due variabili uniformi $0 \leq X, Y \leq 1$. In questo caso, ponendo $\mu = x = y = 1/2$ e $\sigma^2 = s_x^2 = s_y^2 = 1/12$ la (11.16) fornisce il risultato:

$$s_z^2 = y^2 s_x^2 + x^2 s_y^2 = \frac{1}{4} \left(\frac{1}{12} + \frac{1}{12} \right) = \frac{1}{24} = 0.042. \quad (11.18)$$

Si ha pertanto:

$$\mu_z \pm \sigma_z = \frac{1}{2} \frac{1}{2} \pm \sqrt{0.042} = 0.25 \pm 0.20, \quad 0 \leq z \leq 1. \quad (11.19)$$

Si può subito notare come il modello gaussiano sia in questo caso approssimato, perché la media non è al centro dell'intervallo di definizione di z e a sinistra della media esso è largo poco più di una deviazione standard.

Per procedere in modo rigoroso, occorre utilizzare la soluzione del problema 5.7 di pagina 177, dalla quale risulta che la variabile $Z = XY$ ha densità logaritmica $p_Z(z) = -\ln z$, di media e deviazione standard pari a:

$$\mu_z = 0.25, \quad \sigma_z = \sqrt{0.049} = 0.22. \quad (11.20)$$

In termini numerici, questo risultato è molto simile alle (11.18, 11.19). I livelli di confidenza si possono trovare integrando la densità negli intervalli $(\mu \pm K\sigma)$, dove K è un numero reale. Definendo

$$\alpha = \begin{cases} \mu - K\sigma & \text{se } \mu - K\sigma \geq 0 \\ 0 & \text{se } \mu - K\sigma < 0 \end{cases}, \quad \beta = \begin{cases} \mu + K\sigma & \text{se } \mu + K\sigma \leq 1 \\ 1 & \text{se } \mu + K\sigma > 1 \end{cases},$$

si ottiene:

$$\int_{\alpha}^{\beta} -\ln z = [z - z \ln z]_{\alpha}^{\beta} = \beta - \beta \ln(\beta) - \alpha + \alpha \ln(\alpha) = \begin{cases} 0.689 & \text{per } K = 1 \\ 0.946 & \text{per } K = 2 \\ 1 & \text{per } K = 3 \end{cases}. \quad (11.21)$$

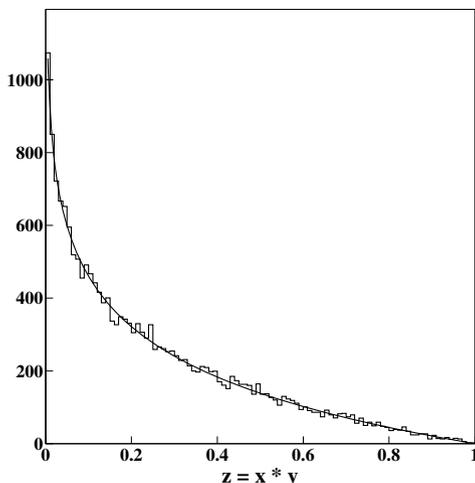


Fig. 11.5. Simulazione al computer di 20 000 variabili aleatorie prodotto di due variabili uniformi standard. La curva rappresenta il *fit* con la (11.20): $n(z) = -a \log z$ (con a parametro di normalizzazione)

Anche questo risultato è numericamente vicino ai livelli di confidenza gaussiani. Tuttavia, esso si riferisce ad una densità asimmetrica completamente differente dalla curva normale. Il buon accordo tra i valori numerici giustifica però, almeno ad una deviazione standard, l'uso diffuso della approssimazione gaussiana nella propagazione degli errori nel prodotto di due misure.

La routine `Prodx` consente di ottenere l'istogramma di Fig. 11.5, a verifica della trattazione analitica.

Esercizio 11.1

La misura dei lati di una piastrina metallica con un metro commerciale con tacca millimetrica ha fornito i valori:

$$b = 25.5 \pm 0.5 \text{ mm} , \quad h = 34.5 \pm 0.5 \text{ mm} .$$

Trovare il valore dell'area della piastrina.

Risposta. Alle misure dei lati è stato giustamente assegnato un errore sistematico di 1 mm, centrato sul valore medio dell'intervallo entro il quale è caduta la misura. Poiché in questo caso, in accordo con la (11.2), si assume che il valore vero delle misure sia entro l'intervallo riportato con un $CL=100\%$ e probabilità uniforme di larghezza $2 \cdot 0.5 = 1$ mm, la deviazione standard da assegnare alle misure sarà:

$$s_b = s_h = \frac{2 \cdot 0.5}{\sqrt{12}} = 0.289 \text{ mm} .$$

Le deviazioni percentuali valgono allora:

$$\frac{s_b}{b} = \frac{0.289}{25.5} = 0.011 \simeq 1.1\% , \quad \frac{s_h}{h} = \frac{0.289}{34.5} = 0.008 \simeq 0.8\% .$$

In base alla (11.16) l'errore percentuale sull'area $A = bh$ vale:

$$\frac{s(A)}{A} = \sqrt{(0.011)^2 + (0.008)^2} = 0.014 = 1.4\% .$$

Poiché l'area vale ovviamente $A = bh = 25.5 \cdot 34.5 = 879.75 \text{ mm}^2$, il risultato richiesto è

$$A = 880 \pm 0.014 \cdot 880 = 880 \pm 12 \text{ mm}^2 .$$

Vi sarete accorti che la propagazione degli errori con metodi analitici, già nei casi semplici che abbiamo sviluppato in dettaglio, è molto laboriosa; vi sarete pure resi conto di come essa possa venire sostituita, più semplicemente, dalla simulazione diretta al calcolatore, come nel caso di Fig. 11.5. La tecnica più diffusa è quella approssimata descritta nel par. 7.10 come metodo *bootstrap*: si assumono come valori centrali e deviazioni standard i valori misurati ed i loro errori, si estraggono a caso delle variabili aleatorie, che vengono combinate come prescritto dalla misura, per ottenere l'istogramma simulato delle grandezze finali (o gli istogrammi nel caso di misure complesse). La forma dell'istogramma fornisce un'idea della densità di probabilità della grandezza misurata, mentre gli errori di misura vengono ricavati direttamente *come limiti delle aree corrispondenti, sull'istogramma, ai livelli di confidenza assegnati*. I valori centrali in genere coincidono o sono molto vicini, entro l'errore statistico, ai valori misurati e quindi non forniscono nuove informazioni. Per quanto riguarda i valori centrali (cioè le medie) risultanti dalla simulazione, va notato che, se gli istogrammi simulati provengono da densità asimmetriche, il valor medio simulato differisce dal valore misurato. Come dare allora il risultato della misura? Il modo corretto di procedere è quello di *riferirsi comunque al valore misurato e di calcolare l'errore di misura centrandosi sempre su di esso*.

Nei casi più complicati, che coinvolgono densità (in genere multidimensionali) di forma non invariante rispetto ai parametri, è necessario abbandonare il metodo *bootstrap*, approssimato, ed utilizzare il metodo rigoroso, che prevede la generazione di dati simulati per tutti i possibili valori dei parametri e la valutazione delle regioni di confidenza secondo quanto descritto nel par. 7.10. Queste tecniche sono discusse in dettaglio in [FC98, JLPe00].

L'esempio che segue ha in sé tutti gli elementi per capire bene la tecnica Monte Carlo applicata alla propagazione degli errori. Il fatto che il caso trattato sia semplice non è affatto limitativo, poiché la tecnica di simulazione ha il grande pregio di mantenere più o meno lo stesso livello di complessità, indipendentemente dalla difficoltà del problema affrontato.

Esercizio 11.2

Uno sperimentatore getta un sasso in un pozzo e con un cronometro manuale misura il tempo di caduta, che risulta di 3 secondi. Attribuendo alla misura manuale del tempo un'incertezza di ± 0.25 secondi, determinare la profondità del pozzo.

Risposta. Il valore centrale della profondità del pozzo si trova con la nota legge della caduta dei gravi:

$$l = \frac{1}{2}gt^2 = 0.5 \cdot 9.81 \frac{m}{s^2} \cdot 9 s^2 = 44.14 m , \quad (11.22)$$

dove $g = 9.81 m/s^2$ è la accelerazione di gravità.

Poiché l'errore che si fa con un cronometro manuale è di tipo sistematico, un'incertezza di ± 0.25 secondi corrisponde ad una deviazione standard da associare alla densità uniforme secondo la legge

$$s_t = \frac{0.5}{\sqrt{12}} = 0.144 \simeq 0.14 s .$$

L'errore di misura si può trovare allora con la (11.4) applicata al caso di una sola variabile:

$$s_l = \left(\frac{dl}{dt} \right) s_t = g t s_t = 9.81 \cdot 3 \cdot 0.144 = 4.24 m .$$

Secondo la propagazione analitica degli errori il valore della profondità del pozzo sta perciò nell'intervallo:

$$l = (39.90, 48.38) = 44.1 \pm 4.2 \text{ metri} . \quad (11.23)$$

Qual è livello di confidenza associare a questa misura? Se indichiamo con τ il valore vero del tempo di caduta, avendo misurato $t = 3 s$ con $\Delta = 0.5 s$, sappiamo che $2.75 \leq \tau \leq 3.25 s$. Questo valore è legato alla lunghezza vera λ dalla relazione:

$$\tau = \sqrt{2\lambda/g} \implies 37.09 \leq \lambda \leq 51.81 .$$

Questo intervallo rappresenta l'insieme di valori del parametro λ da esplorare per risolvere gli integrali (6.7), che forniscono i limiti (λ_1, λ_2) , dell'intervallo di confidenza. Se cerchiamo l'intervallo simmetrico con $CL = 68.3\%$, le (6.7) diventano:

$$\int_{t_0}^{\infty} p_t(t; \lambda_1) dt = 0.158 , \quad \int_{-\infty}^{t_0} p_t(t; \lambda_2) dt = 0.158 , \quad (11.24)$$

dove $t_0 = 3 s$ è il tempo misurato. Questi integrali possono essere facilmente risolti se si individua la cumulativa di $p_t(t)$. Supponendo le misure dei tempi

$T \sim U(\tau - 0.25, \tau + 0.25)$, otteniamo subito:

$$P\{T < t\} = 2(t - \tau + 0.25) = 2(t - \sqrt{2\lambda/g} + 0.25) ,$$

$$\frac{g(\tau - 0.25)^2}{2} \leq \lambda \leq \frac{g(\tau + 0.25)^2}{2} .$$

EsPLICITANDO questa relazione rispetto a λ , possiamo risolvere gli integrali (11.24) ponendo semplicemente:

$$\lambda_\alpha = \frac{g}{2} \left(t_0 - \frac{\alpha}{2} + 0.25 \right)^2 , \tag{11.25}$$

dove i valori $\alpha = 0.158$ e $\alpha = 0.841$ danno gli estremi dell'intervallo (λ_1, λ_2) con $CL = 68.3\%$. Indicando il valore vero della lunghezza con l (come si fa in laboratorio) otteniamo quindi l'intervallo:

$$l = (39.28, 49.29) = 44.1_{-4.8}^{+5.2} , \quad CL = 68.3\% , \tag{11.26}$$

sensibilmente più ampio dell'intervallo approssimato (11.23).

È anche utile il confronto con il risultato che si ottiene dalla simulazione. Se si usa il metodo bootstrap, si estraggono a caso dei tempi dalla densità uniforme avente come valor medio quello misurato e si genera un istogramma delle lunghezze misurate (11.22), che è mostrato in Fig. 11.6.

In questo semplice caso la simulazione non sarebbe necessaria, perché è facile, dalla cumulativa dei tempi, passare a quella delle lunghezze:

$$F_l(l) = P\{L < l\} = P\left\{ \frac{1}{2}gT^2 < l \right\} = P\{T < \sqrt{2l/g}\}$$

$$= 2(\sqrt{2l/g} - \tau + 0.25) , \tag{11.27}$$

$$\frac{1}{2}g(\tau - 0.25)^2 \leq l \leq \frac{1}{2}g(\tau + 0.25)^2 ,$$

e trovare poi, per derivazione, la densità di probabilità di l :

$$p_l(l; \tau) = \sqrt{\frac{2}{g}} \frac{1}{\sqrt{l}} , \quad \frac{1}{2}g(\tau - 0.25)^2 \leq l \leq \frac{1}{2}g(\tau + 0.25)^2 . \tag{11.28}$$

La forma di questa funzione dipende da τ attraverso i limiti dell'intervallo di definizione. Procediamo ora ad analizzare i risultati della simulazione. Dato che l'istogramma contiene 20 000 eventi, dalla (6.46) e dai dati riportati in figura otteniamo l'intervallo di stima della media:

$$\mu = 44.19 \pm \frac{4.22}{\sqrt{20000}} = 44.19 \pm 0.03 ,$$

che è un risultato in accordo col valore misurato di 44.14 metri. Data l'asimmetria della densità di Fig. 11.6 e la non linearità della legge oraria utilizzata,

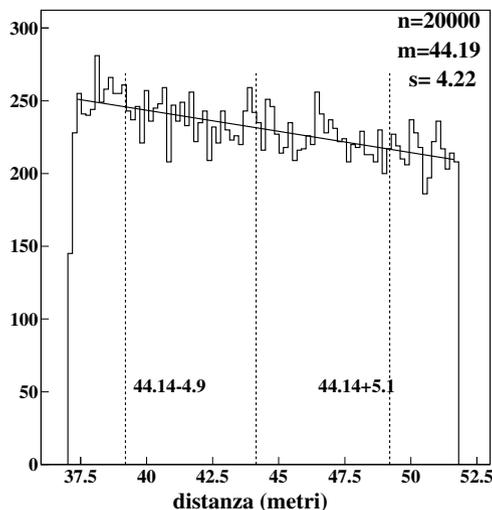


Fig. 11.6. Simulazione al computer di $N = 20000$ lunghezze $l = (1/2)gt^2$ dove $g = 9.81 \text{ m/s}^2$ e t è un tempo misurato pari a 3 secondi, disperso secondo la densità uniforme con $\Delta = 0.5$ secondi. L'area tra le due linee tratteggiate, a destra e a sinistra del valore misurato contiene il 68% degli eventi (lunghezze) dell'istogramma. L'andamento (11.28) della densità è ben approssimato da una retta di best fit $f(l) = 343 - 2.5l$ indicata come linea continua. Dividendo questa funzione per il passo dell'istogramma $\Delta l = 0.17$, si ottiene la densità puntuale $Np(l) = 2018 - 14.70l$, che permette di calcolare le aree sottese dalla curva

è ragionevole comunque aspettarsi uno scostamento, anche piccolo, tra il valor medio e quello misurato.

Se ci si centra sul valore misurato e si analizzano le aree dell'istogramma, risulta che il 68.3% delle profondità, come mostrato in figura, è compreso tra i limiti:

$$l = [39.24, 49.24] \simeq 44.1_{-4.9}^{+5.1} \text{ metri} \quad (CL = 0.68), \quad (11.29)$$

che è un risultato praticamente coincidente col risultato corretto (11.26). La Fig. 11.6 mostra che i livelli di confidenza non sono affatto gaussiani. Potremmo dire in questo caso che, più che dalla (11.29), il risultato della misura è rappresentato proprio dalla Fig. 11.6.

La lunga discussione di questo paragrafo può essere dunque riassunta nei punti seguenti:

- nel caso di errori statistici occorre stimare le deviazioni standard dai campioni sperimentali ed applicare la (11.5). Se tutti gli errori di partenza sono gaussiani, il risultato fornisce la deviazione standard della densità

gaussiana risultante (poiché abbiamo supposto che gli effetti si combinino linearmente) e vale ancora la legge 3σ (3.35);

- nel caso di errori sistematici, va applicata la (11.9) e la deviazione standard risultante quasi sempre definisce livelli di confidenza approssimativamente gaussiani. La formula (11.7) definisce invece un errore massimo, che non va combinato con altre grandezze che rappresentano stime di deviazioni standard;
- la combinazione lineare di errori statistici e sistematici va fatta sempre in quadratura, tramite la (11.5), dove la varianza degli effetti strumentali va calcolata con la (11.9) (non dimenticatevi i fattori $1/12$ oppure $1/24$). La deviazione standard risultante non segue la legge 3σ , perché la densità coinvolta non è gaussiana (si veda la 11.14). Tuttavia, spesso in pratica si assumono i livelli di confidenza gaussiani. Questa assunzione è in genere tanto più vera quanto più è elevato il numero degli errori combinati;
- i casi in cui vi siano errori grandi, correlati, oppure da combinarsi non linearmente, la procedura analitica può portare a notevoli imprecisioni. Si ricorre allora ai metodi di simulazione, che in genere permettono di risolvere in modo completo e soddisfacente qualunque problema di propagazione degli errori. Grazie ai metodi di simulazione, *negli anni recenti i risultati forniti dalla fisica sperimentale hanno compiuto un notevole salto di qualità in precisione, accuratezza ed affidabilità.*

11.7 Definizione dei tipi di misura

Lo schema di Fig. 11.1 suggerisce di classificare i vari tipi di misura secondo la notazione illustrata in Fig. 11.7. Questa è una nostra notazione personale, non la troverete su altri testi. Gli esempi che seguono ne chiariscono il significato:

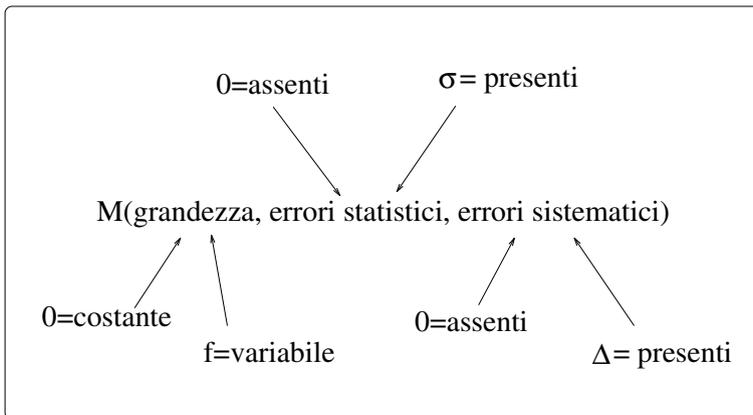


Fig. 11.7. Classificazione dei vari tipi di misura possibili

- $M(0, \sigma, 0)$ = misura di una grandezza fisica costante con errori statistici;
- $M(0, 0, \Delta)$ = misura di una grandezza fisica costante con errori sistematici;
- $M(f, \sigma, \Delta)$ = misura di una grandezza fisica variabile in presenza di errori sia statistici sia sistematici.

Poiché ognuno dei tre simboli della notazione può assumere due valori, in totale si hanno ($2 \times 2 \times 2 = 8$) tipi differenti di misura. Tuttavia, il caso $M(0, 0, 0)$, che si riferisce alla misura senza errori di una grandezza fissa, rappresenta un caso ideale privo di interesse. Pertanto, *in pratica vanno considerati sette tipi differenti di misura*, che andiamo ad analizzare in dettaglio nei prossimi paragrafi.

11.8 Misure del tipo $M(0, 0, \Delta)$

In questo caso si misura una grandezza costante e gli errori statistici non appaiono, perché assenti oppure molto minori dell'errore sistematico Δ .

La misura non presenta fluttuazioni e misure ripetute forniscono tutte lo stesso valore x .

Il risultato della misura va presentato nella forma (11.2):

$$x \pm \frac{\Delta(x)}{2}, \quad (CL = 100\%). \quad (11.30)$$

L'errore ha qui il significato di errore massimo, ed il valore vero è compreso nell'intervallo con probabilità del 100 %, cioè con certezza.

Spesso, in base agli argomenti che vi abbiamo illustrato nel par. 11.4, agli errori sistematici si attribuisce una densità uniforme. In questo caso la varianza della misura è data dalla (3.82):

$$s^2(x) = \frac{\Delta^2(x)}{12}, \quad (11.31)$$

mentre la deviazione standard viene solitamente scritta come:

$$s(x) = \frac{\Delta(x)}{2} \frac{1}{\sqrt{3}} \equiv \frac{U}{\sqrt{3}}, \quad (11.32)$$

dove $U = \Delta/2$ viene talvolta chiamata incertezza della misura.

L'errore derivante dalla combinazione lineare di più misure di questo tipo va propagato con le (11.8, 11.9), e gli intervalli di confidenza tendono rapidamente a diventare gaussiani all'aumentare del numero di misure.

Esercizio 11.3

La misura di una resistenza con un multimetro digitale ha fornito il valore

$$R = 235.4 \quad \Omega .$$

Inoltre, sulla “table of accuracy” allegata allo strumento, si legge che, per le misure di resistenza, l’accuratezza è:

$$\pm (0.1\% \text{ rdg} + 1 \text{ dgt}) .$$

Determinare il valore della resistenza.

Risposta. Questa è una misura in cui è presente solo l’errore sistematico poiché le fluttuazioni statistiche relative alle variazioni della resistenza nel tempo sono al di sotto della sensibilità del multimetro digitale.

L’accuratezza dello strumento è pari allo 0.1% del valore letto (“reading”, ovvero rdg) con l’aggiunta di una unità dell’ultima cifra a destra evidenziata dal display (“last digit”, ovvero dgt), che nel nostro caso vale 0.1 Ω . Abbiamo pertanto:

$$\Delta(R) = \pm(235.4 \cdot 0.001 + 0.1) = \pm(0.2 + 0.1) = \pm 0.3 \Omega .$$

Il valore della resistenza è allora:

$$R = 235.4 \pm 0.3 \Omega ,$$

con un livello di confidenza del 100%.

11.9 Misure del tipo $M(0, \sigma, 0)$

In questo caso si misura ancora una grandezza fisica costante, ma con un apparato di sensibilità molto elevata rispetto agli errori statistici in gioco: $s \gg \Delta$.

Misure ripetute danno valori diversi, che in genere, ma non sempre, si distribuiscono secondo la densità gaussiana. Si ottiene pertanto un campione di N misure, di cui si calcola la media m e la deviazione standard s . Poiché si assume che il valore vero della grandezza fisica coincida con la media della distribuzione delle misure (media vera), in base alla (6.46) il risultato della misura va presentato nella forma:

$$x = m \pm \frac{s}{\sqrt{N}} \quad (CL=68 \% \text{ se } N > 10) , \quad (11.33)$$

a cui va associato un intervallo di confidenza gaussiano se $N > 10$. Per $N \rightarrow \infty$ una misura di questo tipo tende ad avere errore nullo. L’ipotesi di partenza è però che lo strumento sia perfetto, cioè che $\Delta = 0$. In pratica, i risultati vengono presentati nella forma (11.33) quando la dimensione del campione non è sufficiente ad ottenere una precisione che possa competere con l’accuratezza dell’apparato, cioè quando

$$\frac{s}{\sqrt{N}} \gg \Delta .$$

Vogliamo insistere su un punto importante: l'intervallo (11.33) non va mai confuso con l'intervallo

$$m \pm s ,$$

il quale è una stima dell'intervallo $\mu \pm \sigma$, che dà la probabilità di localizzazione di una singola misura entro lo spettro dei valori misurati.

Ricordiamo ancora che, in base alla definizione 11.7, la deviazione standard s è una stima della precisione di una singola misura (errore di misura), mentre la quantità s/\sqrt{N} rappresenta la precisione della misura globale.

Dalla (11.33) risulta anche che se due misure M_1 ed M_2 hanno errori diversi è possibile ottenere la stessa precisione finale da entrambe le misure se N_1 ed N_2 soddisfano la relazione:

$$\frac{N_1}{N_2} = \frac{s_1^2}{s_2^2} .$$

Se ad esempio $s_1 > s_2$, allora deve essere $N_1 > N_2$, cioè il numero di misure dell'esperimento con errore più grande deve essere più grande (si veda ancora la Fig. 11.3).

Se le x_i misure di cui si deve fare la media provengono da esperimenti diversi, esse avranno precisioni s_i differenti, ed allora si deve utilizzare la formula della media pesata (9.69) di pagina 378 con $m_i = x_i$, $n_i = 1$. In pratica, qui si deve considerare una funzione di verosimiglianza prodotto di N misure gaussiane, che hanno fornito i risultati $x_i \pm s_i$. Con la approssimazione $s_i \simeq \sigma_i$ possiamo scrivere la probabilità di ottenere il risultato osservato secondo la verosimiglianza (9.49):

$$L(\mu; \underline{x}) = \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi} s_i} \exp \left(-\frac{1}{2} \frac{(x_i - \mu)^2}{s_i^2} \right) \right] , \quad (11.34)$$

La massimizzazione di questa verosimiglianza equivale alla minimizzazione della verosimiglianza logaritmica (9.6), di pagina 349, che in questo caso coincide col metodo dei minimi quadrati:

$$-\ln L(\mu) \equiv \mathcal{L}(\mu) = -\sum_{i=1}^n \ln \left(\frac{1}{\sqrt{2\pi} s_i} \right) + \frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{s_i^2} . \quad (11.35)$$

Per trovare il minimo $\hat{\mu}$ rispetto alla media generale (pesata) μ si pone $d\mathcal{L}(\mu)/d\mu = d\chi^2/d\mu = 0$ e, ripetendo la procedura di pagina 378, si ottiene il risultato:

$$x = \hat{\mu} \pm \text{Var}[\hat{\mu}] = \frac{\sum_{i=1}^N x_i p_i}{\sum_{i=1}^N p_i} \pm \sqrt{\frac{1}{N \sum_{i=1}^N p_i}} , \quad p_i = \frac{1}{s_i^2} . \quad (11.36)$$

Giova ricordare che questa formula, vale a rigore solo per misure gaussiane, dove l'intervallo (11.36) è un intervallo con copertura del 68%. Tuttavia, anche se i dati non fossero gaussiani, per $N > 10$ vale il teorema Limite Centrale, come nel caso della (11.33).

Esercizio 11.4

Una serie di misure della velocità della luce da parte di vari sperimentatori ha fornito le seguenti medie:

$$\begin{aligned} c_1 &= 2.99 \pm 0.03 \cdot 10^{10} \text{ cm/s} \\ c_2 &= 2.996 \pm 0.001 \quad " \\ c_3 &= 2.991 \pm 0.005 \quad " \\ c_4 &= 2.97 \pm 0.02 \quad " \\ c_5 &= 2.973 \pm 0.002 \quad " . \end{aligned}$$

Determinare, a partire da questi dati, la miglior stima della velocità della luce.

Risposta. Poiché il quinto dato è incompatibile con i primi tre, l'ipotesi più probabile è che lo sperimentatore abbia commesso un errore e che soltanto i primi 4 dati siano corretti.

Alle prime 4 misure corrispondono i pesi

$$\begin{aligned} p_1 &= 1\,111. \\ p_2 &= 10^6 \\ p_3 &= 40\,000. \\ p_4 &= 2\,500. \end{aligned}$$

$$\sum_{i=1}^4 p_i = 1\,043\,611.$$

Notate l'elevato valore dei pesi associati ai dati più precisi. Applicando la (11.36) otteniamo:

$$c = (2.99574 \pm 0.00098) \cdot 10^{10} \text{ cm/s} . \quad (11.37)$$

Qui vale la pena notare un fatto generale: l'errore della media pesata è sempre minore dell'errore più piccolo presente nei dati. Il risultato può anche essere presentato nella forma arrotondata:

$$c = (2.996 \pm 0.001) \cdot 10^{10} \text{ cm/s} ,$$

che è identica al risultato c_2 della seconda misura.

Se non avessimo usato la media pesata, avremmo potuto usare (sbagliando) la formula (11.33) della media normale. Poiché la deviazione standard dei primi 4 dati, in base alla seconda delle (6.50), vale:

$$s(c) = \sqrt{\frac{\sum_{i=1}^4 (c_i - c)^2}{3}} = 0.01147 \cdot 10^{10} \text{ cm/s} ,$$

avremmo ottenuto:

$$c = 2.98675 \pm \frac{s(c)}{\sqrt{4}} = (2.9867 \pm 0.0057) 10^{10} \text{ cm/s} ,$$

che è un risultato completamente diverso da quello corretto (11.37).

11.10 Misure del tipo $M(0, \sigma, \Delta)$

Qui una grandezza fisica costante viene misurata con un apparato in cui sono rilevanti sia gli errori statistici sia sistematici. Misure ripetute forniscono valori diversi ed il campione dei dati viene in genere presentato in forma di istogramma. Per ovvie ragioni, non ha senso scegliere il passo dell'istogramma inferiore all'errore sistematico Δ . Gli istogrammi possono avere tanti o pochi canali, a seconda che sia $s \gg \Delta$, $s \simeq \Delta$ oppure $s \ll \Delta$, dove s è la deviazione standard associata agli effetti casuali (Fig. 11.8). In questa classe di misure rientrano anche quelle del tipo $M(0, \sigma, 0)$ quando i dati sono riportati in un istogramma con passo Δ .

Supponiamo di aver raccolto un campione di N misure e di aver calcolato m ed s . La varianza del campione osservato è ora la somma di due varianze, quella dovuta agli errori statistici, che chiamiamo $s(\text{stat})$, e quella dovuta agli errori sistematici di intervallo massimo $\Delta(\text{syst})$. In base alla (11.15) possiamo scrivere la relazione

$$s^2 = s^2(\text{stat}) + \frac{\Delta^2(\text{syst})}{12} , \quad (11.38)$$

da cui è possibile estrarre la parte delle fluttuazioni dovuta ai soli effetti casuali:

$$s^2(\text{stat}) = s^2 - \frac{\Delta^2(\text{syst})}{12} . \quad (11.39)$$

Questa formula è nota come correzione di Sheppard. Essa dà buoni risultati nel caso di misure $M(0, \sigma, 0)$ istogrammate con passo Δ . In questo caso infatti la misura osservata è la somma di due fluttuazioni, quella dovuta agli

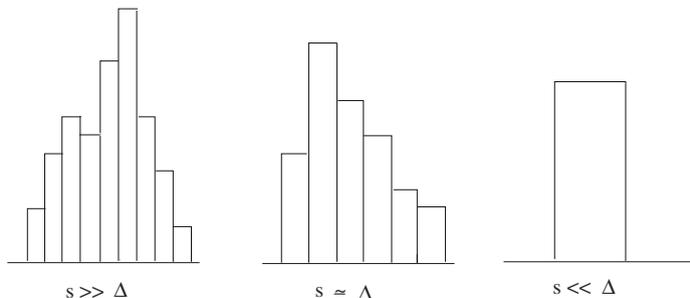


Fig. 11.8. Tipi di istogrammi che si possono presentare in presenza sia di errori sistematici sia statistici

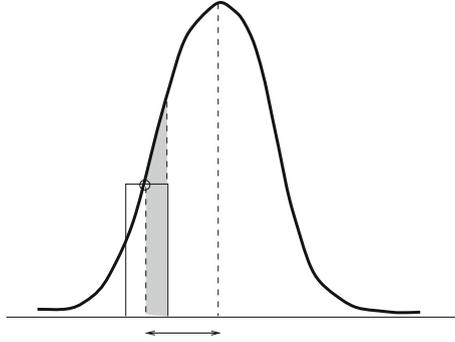


Fig. 11.9. L'istogrammazione di misure a spettro continuo sovrastima la dispersione dei dati quando la popolazione ha densità di tipo regolare come in figura. Infatti a tutti i dati contenuti nella zona ombreggiata (che sono la maggioranza all'interno del canale) viene attribuita l'ascissa del punto medio del canale, che ha una distanza dalla media maggiore di quella effettiva

effetti casuali e quella dovuta alla discretizzazione delle misure, che consiste nell'attribuire ad ogni misura il punto medio del canale dove essa è andata a cadere. In questo caso, come mostrato in Fig. 11.9, la varianza dei dati è sovrastimata e la (11.39) tende a correggere questa sovrastima.

Effetti di istogrammazione a parte, nel caso di esperimenti affetti da errori sia statistici sia sistematici, non c'è tra i ricercatori completo accordo sull'uso o meno della correzione di Sheppard (anzi, crediamo che molti ne ignorino addirittura l'esistenza ...). Infatti, in molte misure gli effetti casuali e strumentali si combinano in modo non lineare, ed in genere la correzione di Sheppard tende a sottostimare gli errori statistici quando $s \simeq \Delta$. Essa dà invece buoni risultati quando $s \gg \Delta$, ma in questi casi la correzione è poco importante.

Per questi motivi nella letteratura scientifica recente le misure $M(0, \sigma, \Delta)$ vengono riportate nella forma:

$$x = m \pm \frac{s}{\sqrt{N}} (\text{stat}) \pm \frac{\Delta}{2} (\text{syst}) , \quad (11.40)$$

dove gli errori sistematici e statistici vengono mantenuti ben separati. La deviazione standard s viene stimata dai dati direttamente oppure con la correzione di Sheppard (11.39). Per trovare $s(\text{stat})$ e Δ a volte vengono utilizzati metodi più sofisticati, quali quelli di simulazione, appena descritti nel par. 11.6.

Nella (11.40) l'intervallo di confidenza non è ben determinato, perché gli errori sistematici sono presentati con un $CL = 100\%$, quelli statistici con $CL = 68\%$ (se sono gaussiani). Il modo di combinare i due errori è in qualche misura arbitrario e viene lasciato a chi deve utilizzare il risultato. In linea di massima, sono possibili due scelte. La prima è di definire un errore mas-

simo globale, moltiplicando per tre gli errori statistici e sommandoli a quelli sistematici:

$$x = m \pm \left(3 \frac{s}{\sqrt{N}} + \frac{\Delta}{2} \right) \quad (CL \simeq 100\%) . \quad (11.41)$$

La seconda scelta è quella di attribuire agli effetti strumentali una densità uniforme e utilizzare la (11.15):

$$x = m \pm \sqrt{\frac{s^2}{N} + \frac{\Delta^2}{12}} . \quad (11.42)$$

In questo caso gli intervalli di confidenza sono approssimativamente gaussiani e sono dati dalla (11.14) e dalla Tab. 11.1.

Come vedete, la procedura per combinare gli errori statistici con quelli sistematici non è univoca né esente da ambiguità. Tuttavia, vi sono casi in cui la situazione non presenta difficoltà. Un buon esempio è costituito dalla scoperta della particella nucleare Z^0 , che fruttò ad alcuni suoi scopritori il premio Nobel per la fisica nel 1984. I due esperimenti (denominati UA1 e UA2), che scoprirono contemporaneamente la particella presso i laboratori europei del CERN di Ginevra, trovarono i seguenti valori di massa, espressi in Giga elettron-volt (GeV) [UA183], [UA283]:

$$\begin{aligned} M_Z &= 95.2 \pm 2.5 \text{ (stat)} \pm 2.8 \text{ (syst)} \text{ GeV} \quad (\text{UA1}) \\ M_Z &= 91.9 \pm 1.3 \text{ (stat)} \pm 1.4 \text{ (syst)} \text{ GeV} \quad (\text{UA2}) , \end{aligned}$$

dove, in entrambi i casi, l'errore sistematico derivava dalla incertezza nella calibrazione assoluta in energia dell'apparato. D'altro canto, la teoria prevedeva il valore

$$92.3 \pm 0.7 \text{ GeV} ,$$

dove l'errore dipendeva in questo caso dalle approssimazioni utilizzate nei calcoli.

È evidente in questo caso l'ottimo accordo tra teoria ed esperimenti, indipendentemente dalle tecniche di analisi dei dati e trattamento degli errori.

11.11 Misure del tipo $M(f, 0, 0)$

Qui ci si riferisce al caso di una variabile aleatoria misurata senza errore alcuno.

Lo scopo della misura è in questo caso *la determinazione della distribuzione o legge statistica che determina il fenomeno fisico*. A volte può essere sufficiente la determinazione delle medie e dei parametri di dispersione di questa distribuzione, altre volte è fondamentale conoscerne la forma funzionale precisa. Si pensi, per quest'ultimo caso, alle densità di Maxwell e Boltzmann, che stanno alla base della meccanica statistica.

A questa classe di misure ed osservazioni appartengono in pratica tutti i fenomeni stocastici discussi nei capitoli precedenti, ed i metodi per la determinazione di medie, varianze, forme funzionali sono proprio quelli che abbiamo ampiamente discusso fino ad ora. Come esempio significativo, basta ricordare l'esperimento delle 10 monete, la cui analisi finale è stata effettuata nell'esercizio 9.7.

Nel campo della fisica, rientrano in questo tipo di misure tutti gli esperimenti di conteggio, per i quali, come discusso nel par. 3.8, la distribuzione di Poisson riveste un ruolo fondamentale.

Spesso capita di avere contato N_{s+b} eventi da una sorgente entro un tempo t_s ; rimossa la sorgente, vengono registrati N_b eventi di fondo per un tempo in genere più lungo t_b . Il rapporto segnale/fondo si può stimare come numero n_σ di deviazioni standard del segnale sul fondo (cioè come variabile standard), riportando i conteggi di fondo all'intervallo di tempo t_s . In questo passaggio bisogna fare attenzione al calcolo dell'errore: $\sigma[N_b] = \sqrt{N_b} t_s/t_b$, cioè *prima* si deve trovare l'errore poissoniano dei conteggi originali e *poi* si deve moltiplicare per la costante t_s/t_b , come prescrive la propagazione degli errori. Si ottiene quindi:

$$n_\sigma = \frac{N_{s+b} - N_b t_s/t_b}{\sqrt{N_{s+b} + N_b t_s^2/t_b^2}}. \quad (11.43)$$

Nella ricerca di nuovi fenomeni, se $n_\sigma \simeq 3$ i fisici parlano di "forte evidenza", mentre se $n_\sigma > 5$ si ritiene avvenuta la scoperta. Per una applicazione di questa formula si veda il problema 11.3.

Discuteremo ora un tipico esperimento di conteggio di fisica nucleare.

Esercizio 11.5

Nel 1930 il fisico L.F. Curtiss eseguì un esperimento per determinare la legge statistica di emissione delle particelle nel decadimento di un nucleo radioattivo. Egli registrò con un contatore Geiger il numero di particelle α emesse da una sottile pellicola di polonio. Durante l'esperimento fu registrato il numero di particelle contate in 3455 intervalli di tempo tutti uguali e della durata di alcuni minuti. Se definiamo

x = numero di particelle emesse (conteggi Geiger)

n_x = numero di intervalli con x conteggi

i risultati dell'esperimento, riportati in [Eva55], possono essere riassunti nella tabella:

x	0	1	2	3	4	5	6	7
n_x	8	59	177	311	492	528	601	467
x	8	9	10	11	12	13	14	15
n_x	331	220	121	85	24	22	6	3

(11.44)

dalla quale risulta che ci furono 8 intervalli senza conteggi, 59 intervalli con un solo conteggio, 177 intervalli con 2 conteggi e così via, fino ad un limite di 3 intervalli con 15 conteggi.

Eseguire l'analisi completa dell'esperimento.

Risposta. Il tempo di dimezzamento del polonio è di circa 4 mesi, per cui l'intensità di emissione della sorgente si può ritenere costante durante l'esecuzione dell'esperimento.

Il contatore Geiger è un tubo a gas sotto tensione, nel quale avviene una scarica in corrispondenza del passaggio di una particella α ionizzante. Il tempo di ricarica del tubo, necessario affinché la tensione tra gli elettrodi risalga ad un valore sufficiente per una nuova scarica, è dell'ordine del millesimo di secondo. Poiché l'esperimento ha registrato meno di una ventina di conteggi nell'arco di alcuni minuti, gli effetti di alterazione dei conteggi dovuti a particelle α penetrate nel Geiger durante il tempo di ricarica (tempo morto) è assolutamente trascurabile. In questo caso, se la strumentazione è ben funzionante, nella lettura dei conteggi ovviamente non vi sono errori. Siamo pertanto nel caso di un esperimento $M(f, 0, 0)$, che conta il numero di particelle emesse senza errori nelle operazioni di misura.

Supponendo quindi che l'intensità della sorgente di polonio resti costante, e che tutti i nuclei del campione emettano particelle α in modo indipendente e con la stessa probabilità elementare costante nel tempo (ipotesi che stanno alla base del modello nucleare del decadimento radioattivo), la legge statistica dei conteggi deve essere di Poisson. Verifichiamo dunque questa ipotesi.

Per prima cosa notiamo che, avendo raggruppato i dati come in Tab. (11.44), lo spettro della variabile aleatoria in esame è dato dai valori di x , cioè dal numero di conteggi registrati, mentre le frequenze degli eventi sono date dal numero di intervalli di tempo con un dato numero di conteggi. Le frequenze dello spettro sono pertanto date da:

$$f_x = \frac{n_x}{N}, \quad N = \sum_x n_x = 3455.$$

Utilizzando le (2.52, 2.54, 2.57), possiamo allora calcolare media varianza e momento del quart'ordine dai dati (potete utilizzare la vostra calcolatrice tascabile, meglio se programmabile):

$$\begin{aligned} m &= \sum_{x=0}^{15} x f_x = 5.877 \\ s^2 &= \frac{N}{N-1} \sum_x (x-m)^2 f_x = 5.859 \\ D_4 &\simeq \frac{1}{N} \sum_x (x-m)^4 f_x = 107.07. \end{aligned}$$

Appare subito evidente, in accordo con la proprietà (3.16) della distribuzione di Poisson, l'uguaglianza tra media e varianza. Infatti, la stima statistica della media è data dalla (6.46):

$$\mu = m \pm \frac{s}{\sqrt{N}} = 5.877 \pm 0.041 \simeq 5.87 \pm 0.04,$$

mentre quella sulla varianza è data dalla (6.59):

$$\sigma^2 = s^2 \pm \sqrt{\frac{D_4 - s^4}{N}} = 5.859 \pm 0.145 \simeq 5.86 \pm 0.14 .$$

Per la varianza abbiamo usato la formula generale, in quanto la media dei conteggi, essendo < 10 , non permette a rigore l'uso della formula (6.74), valida per variabili gaussiane (si veda la Tab.6.3 di pag. 213). Tuttavia, anche la (6.62) (o la (6.74)) forniscono un risultato statisticamente identico:

$$s^2 \pm s^2 \sqrt{\frac{2}{N-1}} = 5.859 \pm 0.141 \simeq 5.86 \pm 0.14 .$$

Gli intervalli di confidenza per media e varianza possono essere associati a livelli di probabilità gaussiani, perché $N = 3455 \gg 100$. Utilizzando la (6.99) possiamo allora procedere al test della differenza:

$$t = \frac{|m - s^2|}{\sqrt{s^2(m) + s^2(s^2)}} = \frac{|5.877 - 5.859|}{\sqrt{0.041^2 + 0.145^2}} = 0.12 ,$$

consistente con l'ipotesi $\mu = \sigma^2$ entro 0.12 unità di errore.

Da questa analisi preliminare abbiamo ottenuto una prima prova importante a favore della distribuzione di Poisson. Possiamo quindi procedere oltre, ed introdurre i dati in un programma di minimizzazione rispetto a μ della funzione χ^2 data dalla (6.125):

$$\chi^2 = \sum \frac{(n_x - Np(x; \mu))^2}{n_x} ,$$

dove

$$p(x; \mu) = \frac{\mu^x}{x!} e^{-\mu}$$

è la densità di Poisson di media μ da determinare.

Abbiamo usato la routine `Nlinfit`, che impiega il metodo del gradiente negativo, ottenendo per la stima della media e per il χ^2 di minimo i valori:

$$\mu = 5.866 \pm 0.041 , \quad \chi_{min}^2 = 18.64 .$$

Veniamo ora alla stima dei gradi di libertà. L'istogramma ha 16 canali, ed è stato ottenuto con un numero totale $N = 3455$ di eventi predeterminato, poiché questo non è altro che il numero di prove di conteggio effettuate da Curtiss. Questo diminuisce di una unità i gradi di libertà (se avete dei dubbi, rileggete i paragrafi 6.16, 6.17). Inoltre, dai dati si è stimato il parametro μ , il che diminuisce di un'altra unità i gradi di sistema, che sono pertanto pari a $\nu = 16 - 2 = 14$. Il chi quadrato ridotto χ_ν^2 vale allora:

$$\chi_{14}^2 = \frac{18.64}{14} = 1.33 .$$

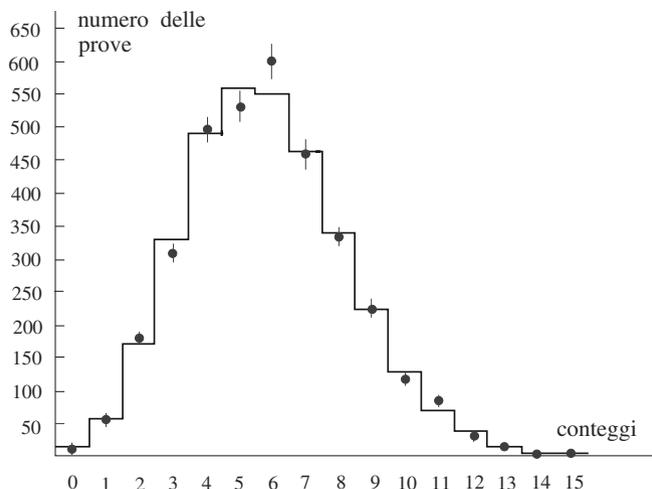


Fig. 11.10. Confronto tra dati (punti con barra d'errore) e densità di Poisson (istogramma con linea continua) per l'esperimento del decadimento radioattivo del polonio

Dalla Tab. D.3 risulta un livello di significatività di circa il 15 %, il che rivela un buon accordo tra l'esperimento e la distribuzione di Poisson.

I dati finali possono essere condensati in una tabella oppure, in modo più sintetico, in un grafico, riportando, per ogni valore di x , il valore di n_x , il suo errore statistico (la barra d'errore) e la predizione della legge di Poisson. Ad esempio, per il valore $x = 5$, dalle (3.14, 6.116) si ottiene:

$$n_x \pm \sqrt{n_x} = 528 \pm 23, \quad Np(5; 5.866) = 3455 \frac{5.866^5}{5!} e^{-5.866} = 568.$$

Il risultato del fit è riportato in Fig. 11.10, dove si vede un ottimo accordo globale tra dati e teoria.

Il valor medio trovato dall'esperimento, una volta normalizzato per unità di tempo e diviso per il numero dei nuclei radioattivi presenti, fornisce la costante di decadimento per radiazione α del nucleo di polonio (con il relativo errore). Questa costante, indicata solitamente come λ , è una caratteristica fisica intrinseca fondamentale del nucleo di polonio, che ha un raggio dell'ordine di 10^{-12} cm.

Esercizio 11.6

Due esperimenti di conteggio dello stesso fenomeno hanno fornito rispettivamente 92 conteggi in 100 secondi e 1025 conteggi in 1000 secondi. Fare la media pesata dei due risultati.

Risposta. *Non si può fare la media pesata dei dati grezzi, perché essi hanno media diversa. Tuttavia, poiché i conteggi provengono dalla stessa sorgente e differiscono per il tempo di conteggio, che nella seconda misura è 10 volte più grande, si può procedere alla normalizzazione dei dati come segue:*

$$m_1 \in 92 \pm \sqrt{92} = 92 \pm 9.6$$

$$m_2 \in \frac{1}{10} 1025 \pm \frac{1}{10} \sqrt{1025} = 102.5 \pm 3.2$$

dove si è utilizzata la (6.116) e si è normalizzata la seconda misura (col relativo errore) alla prima. Notate che la seconda misura fornisce un risultato più preciso, perché, per eventi poissoniani, l'errore statistico relativo, come dovrete sapere avendo letto il par. 6.16, va come \sqrt{n}/n e tende ad annullarsi per grandi campioni.

Le variabili, essendo poissoniane di media $\gg 10$, visto che l'ordine di grandezza dei conteggi registrati è maggiore di 100, si possono considerare gaussiane. Possiamo pertanto applicare la formula della media pesata.

I pesi di m_1 ed m_2 sono dati da:

$$p_1 = \frac{1}{9.6^2} = 0.011, \quad p_2 = \frac{1}{3.2^2} = 0.098$$

e la stima pesata della media vera vale:

$$\mu \in \frac{92 \cdot 0.011 + 102.5 \cdot 0.098}{0.011 + 0.098} \pm \sqrt{\frac{1}{0.011 + 0.098}}$$

$$= 101.4 \pm 3.0 \quad \text{conteggi in 100 secondi.}$$

Per concludere, vale la pena di notare due fatti del tutto generali: il dato finale è più vicino alla misura più precisa (la seconda, che ha un peso 10 volte maggiore della prima); l'inclusione della prima misura, anche se molto meno precisa, riduce comunque l'errore, anche se di poco.

11.12 Misure del tipo $M(f, \sigma, 0)$, $M(f, 0, \Delta)$, $M(f, \sigma, \Delta)$

In questa classe di esperimenti, molto complicata da analizzare, le fluttuazioni dei valori assunti dall'oggetto fisico si combinano con le fluttuazioni e le incertezze dovute alla misura. L'obiettivo della analisi è quello di determinare la funzione $f(x)$, dipendente da una o più variabili, che rappresenta la legge statistica della grandezza fisica in esame. Tuttavia, ciò che viene osservato è una densità g che risente delle fluttuazioni fisiche come di quelle di misura. È indispensabile allora misurare, in via preliminare, la risposta dell'apparato di misura, detta funzione dell'apparato (*instrumental function* o *apparatus function*), che ha il seguente significato:

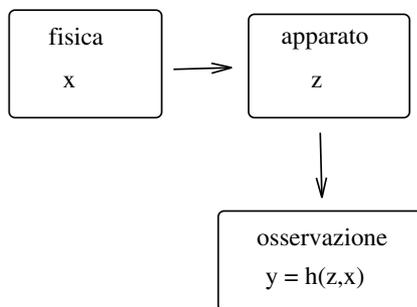


Fig. 11.11. Effetto di folding: la grandezza fisica x e l'effetto dell'apparato z si combinano in una funzione h , per dare un valore y misurato funzione di queste due variabili

Definizione 11.10 (Funzione dell'apparato). Si dice *funzione dell'apparato* o *funzione strumentale*, e si indica in genere con $\delta(x, y)$, la funzione che dà la probabilità $\delta(x, y) dx dy$ di misurare un valore in $[y, y + dy]$ quando il valore della variabile è in $[x, x + dx]$.

In sostanza, la funzione dell'apparato fa conoscere la probabilità che lo strumento risponda con y quando il valore di ingresso è x .

La funzione osservata $g(y)$ dipende quindi da due variabili (Z, X) (apparato e fisica), e i valori misurati di Y sono legati a queste variabili dalla relazione (schematizzata anche in Fig. 11.11):

$$y = h(z, x), \quad z = h^{-1}(y, x),$$

dove la funzione h rappresenta il legame tra z e x creato dall'operazione di misura (una somma $x + z$, un prodotto $x \cdot z$, od altro).

Se ora $p_Z(z)$ e $p_X(x) \equiv f(x)$ sono le densità di probabilità associate rispettivamente all'apparato e al processo fisico, Z ed X sono indipendenti, perché è ragionevole supporre che la fisica dell'apparato sia indipendente da quella del processo fisico e vale la (5.28) di pagina 156, che per comodità riscriviamo con la nuova notazione data alle variabili:

$$g(y) = \int p_Z(h^{-1}(y, x)) f(x) \frac{\partial h^{-1}}{\partial y} dx.$$

La funzione dell'apparato è ora identificabile con la quantità

$$p_Z(h^{-1}(y, x)) \frac{\partial h^{-1}}{\partial y} \equiv \delta(y, x),$$

che non è altro che la densità dell'apparato $p_Z(z)$ valutata per $z = h^{-1}(y, x)$ legata a x e y dall'operazione di misura. La derivata $\partial h^{-1}/\partial y$ è lo jacobiano della trasformazione.

Riassumendo possiamo dire che, in base alle leggi delle probabilità composte e totali, riassunte nella (5.28), le tre funzioni $f(x)$ (fisica), $g(y)$ (osservazione) e $\delta(y, x)$ (apparato) sono legate dalla relazione:

$$g(y) = \int f(x) \delta(y, x) dx \equiv f * \delta, \tag{11.45}$$

che si può interpretare come segue: la probabilità di osservare un valore y è data dalla probabilità che la variabile fisica assuma un valore x , per la probabilità che lo strumento, in presenza di un valore x , fornisca un valore y . Queste probabilità vanno sommate (integrate) su tutti i valori veri x dello spettro che possono avere y come valore osservato.

La relazione (11.45) è detta integrale di folding, viene talvolta indicata col simbolo dell'asterisco ed è di notevole importanza sia in fisica sia in ingegneria. Poiché $g(y)$ è misurata e $\delta(x, y)$ deve essere nota, l'esperimento deve determinare la funzione $f(x)$ che è, per così dire, intrappolata o avvolta (in inglese *folded*) nell'integrale di folding.

Se, come succede spesso, la risposta dell'apparato è lineare,

$$y = h(z, x) = z + x, \quad z = h^{-1}(y, x) = y - x, \quad \frac{\partial h^{-1}}{\partial y} = 1$$

l'integrale di folding si trasforma nell'integrale di convoluzione (5.27):

$$g(y) = \int f(x) \delta(y - x) dx, \tag{11.46}$$

che può essere schematizzato graficamente come in Fig. 11.12.

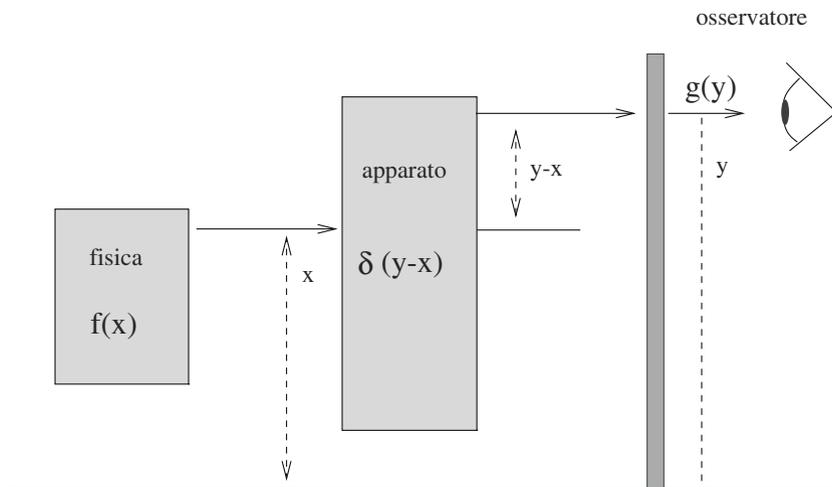


Fig. 11.12. Rappresentazione grafica della convoluzione

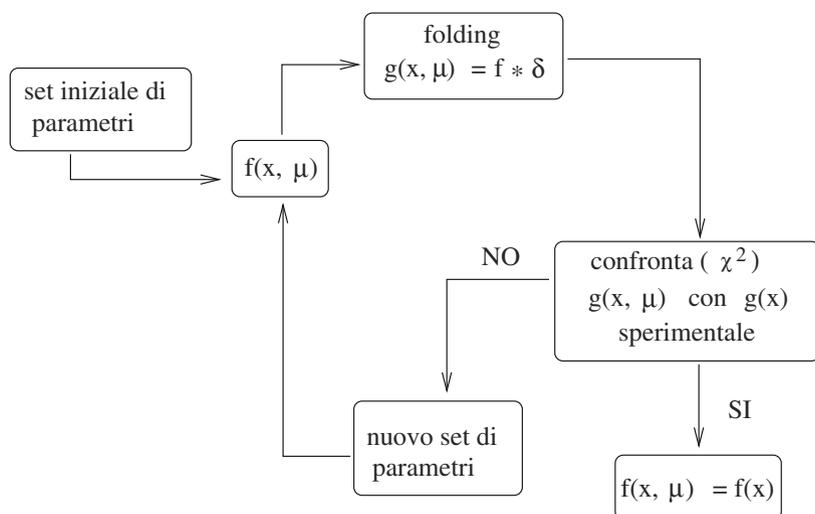


Fig. 11.13. Il metodo iterativo per la risoluzione dell'integrale di folding

Le tecniche che estraggono $f(x)$ dagli integrali (11.45, 11.46) prendono il nome di tecniche di unfolding e deconvoluzione. In linea di principio, la tecnica delle trasformate di Fourier o Laplace consente di trasformare questi integrali in equazioni algebriche facilmente risolvibili. Tuttavia, la necessità di propagare gli errori attraverso questi algoritmi crea sovente notevoli difficoltà.

Spesso si usa la tecnica iterativa che è schematizzata in Fig. 11.13. Una funzione $f(x, \mu)$ dipendente da uno o più parametri μ viene “iniettata” nell'integrale di folding, che viene risolto con metodi numerici o di simulazione; la soluzione $g(x, \mu)$ che si ottiene, dipendente anch'essa dallo stesso insieme di parametri, viene confrontata con la funzione $g(x)$ effettivamente osservata. Se si usano i metodi di *best fit* (come quelli usati nella routine `Nlinfit`) che consistono nel metodo del gradiente negativo per la determinazione dei parametri μ di minimo e nel test χ^2 per la determinazione del grado di accordo con le osservazioni, alla fine del ciclo si ottiene la funzione cercata $f(x, \mu_0)$, con i valori ottimali dei parametri μ_0 e relativo errore. Questo metodo, sebbene piuttosto dispendioso dal punto di vista del calcolo, ha il pregio di essere abbastanza flessibile e generale.

Non intendiamo qui entrare in ulteriore dettaglio in questi argomenti, sui quali esiste una letteratura tecnica specializzata molto vasta (vedere, ad esempio, [Blo84]). A nostro avviso è sufficiente che siate consci del problema e che sappiate che esistono dei modi per risolverlo. Se intraprenderete la strada della ricerca scientifica sperimentale, allora c'è di certo un integrale di folding o convoluzione che vi sta aspettando. Quando lo incontrerete, non scoraggiatevi e cercate nella letteratura il metodo migliore per risolvere quel problema specifico: quasi sicuramente esiste. Se il calcolo degli errori o della stabilità

della soluzione è importante, ricordatevi di essere particolarmente attenti al tipo di algoritmo da usare; nella scelta, un po' di nozioni imparate sul nostro libro vi saranno probabilmente di aiuto.

Vi abbiamo mostrato che la determinazione completa e affidabile della densità $f(x)$ è in genere complicata a causa della presenza della funzione dell'apparato $\delta(y, x)$, che tiene conto degli effetti strumentali (Δ) e/o casuali (σ) presenti nell'operazione di misura.

Nel caso della convoluzione, dove

$$\text{osservazione} = \text{fisica} + \text{apparato} \implies y = x + (y - x) \implies y = x + d$$

se non la struttura completa, almeno la media $\langle X \rangle$ e la varianza $\text{Var}[X]$ (o la deviazione standard) possono essere stimate dalle quantità $m(x)$ e $s^2(x)$ deducibili dai dati con le note regole (5.67, 6.61, 6.64).

Notiamo prima di tutto che la deviazione media $m(d)$ e la dispersione causata dalle operazioni di misura, (quest'ultima caratterizzata dalla deviazione standard $s^2(d)$ e/o dall'errore sistematico Δ), devono essere quantità note con errore trascurabile, altrimenti l'esperimento non è fattibile.

Inoltre, anche $m(y)$ e $s(y)$ sono quantità note con errore, perché si ottengono dal campione dei dati osservati.

Allora, per le misure $M(f, \sigma, 0)$ abbiamo:

$$\begin{aligned} s^2(x) &= s^2(y) - s^2(d) \\ m(x) &= m(y) - m(d) \\ \mu &= m(x) \pm \frac{s(x)}{\sqrt{N}} \\ \sigma &\simeq s(x) \pm \frac{s(x)}{\sqrt{2N}}. \end{aligned} \tag{11.47}$$

Nel caso invece delle misure $M(f, 0, \Delta)$ abbiamo:

$$\begin{aligned} s^2(x) &= s^2(y) - \frac{\Delta^2}{12} \\ m(x) &= m(y) - m(d) \\ \mu &= m(x) \pm \frac{s(x)}{\sqrt{N}} \pm \frac{\Delta}{2} \\ \sigma &\simeq s(x) \pm \frac{s(x)}{\sqrt{2N}}. \end{aligned} \tag{11.48}$$

Infine, per le misure $M(f, \sigma, \Delta)$ si ha:

$$\begin{aligned} s^2(x) &= s^2(y) - s^2(d) - \frac{\Delta^2}{12} \\ m(x) &= m(y) - m(d) \end{aligned}$$

e per μ e σ valgono ancora le (11.48).

11.13 Studio di un caso: gli esperimenti di Millikan

Riteniamo ora abbastanza istruttivo applicare alcuni dei concetti fin qui introdotti alla analisi dei famosi esperimenti di Millikan sulla carica dell'elettrone. Questi esperimenti hanno recentemente attirato l'attenzione di alcuni storici e critici della scienza, aprendo alcune polemiche sul modo di operare degli scienziati. Vediamo di che si tratta.

Robert Millikan fu un fisico americano divenuto famoso per le sue misure sulla carica dell'elettrone. Per queste ricerche ottenne il premio Nobel nel 1923 ed altri numerosi riconoscimenti, tra cui 20 lauree ad honorem.

Nel 1910, quando era professore all'università di Chicago, Millikan pubblicò i primi risultati dei suoi esperimenti, ottenuti studiando la caduta di goccioline di olio in un campo elettrico. L'apparato utilizzato è mostrato in Fig. 11.14. Un flusso d'aria *A* cattura per effetto Venturi delle minuscole goccioline dalla ampolla d'olio *O*; queste goccioline cadono per gravità entro un foro praticato in un condensatore *C*, ove trovano un campo elettrico statico prodotto dal generatore *P*. L'apparato è tenuto in depressione e controllato da un manometro. Le goccioline si elettrizzano negativamente per contatto col flusso d'aria, ed il loro moto è quindi influenzato dalla presenza del campo elettrico del condensatore, la cui polarità è tale da rallentare il moto della goccia, essendo il piatto inferiore a potenziale negativo.

È noto che la caduta di un grave in un fluido viscoso (l'aria) è a regime un moto di tipo uniforme, con velocità costante, quando la forza di attrito viscoso, che secondo la legge di Stokes dipende linearmente dalla velocità, diventa uguale alla forza peso. Questo è il moto delle gocce di pioggia, oppure quello di un paracadutista, sia in caduta libera sia col paracadute aperto. Utilizzando dunque la legge di Stokes e l'equazione del campo elettrico di un condensatore, Millikan scrisse le equazioni del moto della goccia con e senza

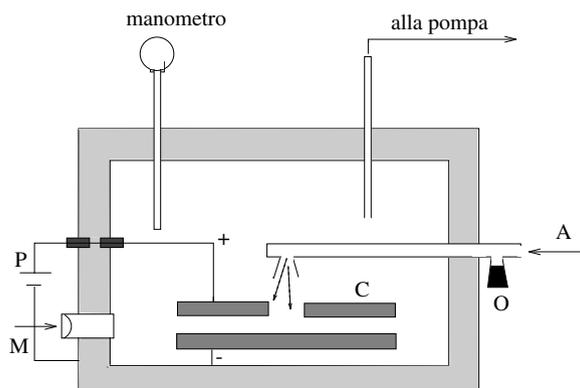


Fig. 11.14. L'apparato di Millikan per la misura della carica elementare. *A* = flusso d'aria, *C* = condensatore, *M* = microscopio, *P* = generatore di tensione, *O* = ampolla d'olio

campo come:

$$6\pi\eta v_0 r = \frac{4}{3}\pi r^3 \rho g \quad (\text{senza campo elettrico}) , \quad (11.49)$$

$$6\pi\eta v r = \frac{4}{3}\pi r^3 \rho g - \frac{V}{h} q \quad (\text{con campo elettrico}) , \quad (11.50)$$

dove η è il coefficiente di viscosità dell'aria, v e v_0 le velocità di caduta della goccia con e senza campo rispettivamente, r il raggio della goccia, ρ la densità dell'olio, g l'accelerazione di gravità, V la tensione del generatore, h la distanza tra i piatti del condensatore e q la carica della goccia, che è l'oggetto della misura.

Eseguendo due misure della velocità di caduta v e v_0 con e senza campo, dalle equazioni (11.49, 11.50) è possibile ricavare le incognite r e q . La misura della velocità veniva effettuata misurando col microscopio, guardando attraverso una lente munita di tacche, il tempo di caduta di ogni goccia. Questo poteva essere regolato a piacere cambiando la tensione V o la pressione dell'aria. Era anche possibile fermare le goccioline in sospensione nell'aria ($v = 0$) applicando un opportuno potenziale negativo al piatto inferiore del generatore.

In questo modo Millikan notò che la carica q assumeva sempre un valore minimo q_0 oppure un multiplo di questo valore. Egli scrisse tutte le sue osservazioni in un registro che è disponibile ancora oggi. L'errore di misura era assunto di tipo puramente statistico, derivante dall'incertezza nella determinazione manuale dei tempi di caduta con il microscopio e un orologio.

Siamo pertanto in presenza di una misura del tipo $M(0, \sigma, 0)$.

Millikan effettuò la media di un primo gruppo di 23 misure, calcolò l'errore della media come prescritto nella (11.33) e pubblicò nel 1913 il valore della carica elementare dell'elettrone:

$$q_0 \equiv e = 4.778 \pm 0.002 \cdot 10^{-10} \quad \text{esu} . \quad (11.51)$$

Una polemica riguardante la stabilità della misura ed il piccolo errore assegnato al valore di e , spinse Millikan a pubblicare, nel 1923, un nuovo studio della caduta di 58 gocce, nel quale sottolineava che i dati non erano relativi "a un gruppo selezionato di cadute, ma rappresentavano tutte le cadute verificate in 60 giorni consecutivi". Il valore di questa seconda misura era:

$$e = 4.780 \pm 0.002 \cdot 10^{-10} \quad \text{esu} . \quad (11.52)$$

Dagli appunti di Millikan, recentemente riesaminati, risulta però che nella prima misura (quella delle 23 gocce), egli escluse dalla pubblicazione 7 misure giudicate "cattive", mentre nella seconda misura, (quella delle 58 gocce), ne escluse addirittura 82.

Sulla base di queste vicende, spesso Millikan è citato come un esempio negativo di disonestà scientifica. Tuttavia, poiché egli riposa in pace ormai da

molti anni, non ci è dato sapere se le gocce furono scartate sulla base di buoni motivi (misure non effettuate a regola d'arte da lui o dai suoi assistenti) o solo con l'intenzione di minimizzare l'errore. Il riesame dei dati di Millikan ha tuttavia dimostrato che le gocce scartate non modificano il risultato pubblicato in modo statisticamente significativo [Ber85].

A questo punto vi sarete già chiesti qual è il valore vero della carica dell'elettrone. Ovviamente, non conosciamo il valore vero, ma la media pesata di tutte le misure effettuate fino ad oggi, che risulta essere

$$e = 4.803\,206\,8 \pm 0.000\,001\,5 \cdot 10^{-10} \text{ esu} . \quad (11.53)$$

Questo valore è noto con un errore relativo di $3 \cdot 10^{-7}$ (0.3 parti per milione) e nella nostra discussione può essere considerato come valore “vero” di riferimento. Il confronto tra questo valore e quello di Millikan (11.51) può essere fatto in termini di variabile standard:

$$\frac{|\text{valore di Millikan} - \text{valore vero}|}{\text{errore di Millikan}} = \frac{|4.778 - 4.803|}{0.002} = 12.5 .$$

Il valore di Millikan dista di 12.5 unità di errore da quello vero, in assoluta violazione della legge 3σ . Da un punto di vista formale e metodologico, Millikan ha pertanto sbagliato la misura, perché il valore vero *non è compreso con legge 3σ nell'intervallo (11.51)*.

Tuttavia va notato, a favore dello scienziato, che Millikan dimostrò per la prima volta che la carica elettrica è quantizzata, determinandone il valore con un errore relativo

$$\frac{|4.778 - 4.803|}{4.803} \simeq 0.005 = 0.5\% ,$$

il quale ci sembra un risultato di tutto rispetto, considerando anche l'epoca, l'originalità della misura ed il tipo di apparato usato.

Resta comunque da chiarire la differenza anomala, molto maggiore dell'errore statistico, tra risultati di Millikan (sia di quelli originali che di quelli corretti) ed il valore vero. La ragione sta nel modello fisico usato per la misura, rappresentato dalle equazioni (11.49, 11.50): fu utilizzato un coefficiente η di viscosità dell'aria approssimato e, nella determinazione del valore finale, Millikan non tenne conto dell'errore sistematico derivante da questa approssimazione.

Concludiamo con le parole di un altro premio Nobel per la fisica, Richard Feynman, che cita [Fey92] le misure di Millikan a proposito del cosiddetto effetto di trascinamento, di cui parleremo nel prossimo paragrafo:

L'esperienza ci ha insegnato come evitare di autoingannarci. Faccio un esempio: Millikan ha misurato la carica di un elettrone con un esperimento in cui cadevano delle gocce d'olio, ed ha ottenuto una risposta che ora noi sappiamo essere non del tutto esatta. Non lo è perché Millikan non aveva i dati esatti della viscosità dell'aria. È molto interessante studiare la storia delle misure della carica dell'elettrone,

dopo Millikan. Se riportaste quelle misure su una curva temporale, vi accorgeteste che la prima è leggermente superiore a quella di Millikan, la successiva ancora di poco superiore e così via fino a raggiungere una cifra parecchio più grande.

Perché mai non si è trovata subito la cifra più alta? È una storia di cui gli scienziati si vergognano, eppure il motivo è chiaro. Ogni volta che uno scienziato otteneva un valore molto superiore a quello di Millikan, pensava di aver sbagliato e cercava l'errore. Se la misura era più vicina a quella di Millikan, non tentava così accuratamente di rintracciare l'errore: eliminava i valori troppo discordanti. Ci è servito da lezione, non soffriamo più di questa malattia.

11.14 Alcune note sul metodo scientifico

Dopo aver illustrato i contenuti tecnici delle operazioni di misura, vogliamo concludere il capitolo (e il libro) con alcune osservazioni di metodo e di principio.

Come abbiamo già detto all'inizio, il metodo scientifico è basato sull'osservazione della natura, ovvero sulla misura delle osservabili fisiche con i metodi e le tecniche descritte nei paragrafi precedenti.

Alla base di questo approccio sta un principio senza il quale l'intera scienza moderna non potrebbe esistere: il *postulato di oggettività*. In altre parole, la natura esiste di per sé, con sue proprie leggi, e non è una proiezione della mente dell'uomo. Non è lecito quindi attribuire alla natura finalità e scopi propri del mondo soggettivo e culturale dello sperimentatore.

L'oggettività va considerata come un postulato, dato che non può essere dimostrata sulla base di assunzioni più ovvie od elementari. Tuttavia, quando questo principio non è valido, come succede in alcune sperimentazioni mediche, *gli scienziati sono in grado di accorgersene*. Ci riferiamo al cosiddetto *effetto placebo*, che consiste nella guarigione spontanea di pazienti ai quali viene fatto credere di essere curati con un farmaco efficace, mentre in realtà viene somministrata loro una sostanza (come acqua o zucchero), detta placebo, priva di effetto terapeutico. Si può verificare sperimentalmente che nei pazienti fiduciosi si determinano condizioni psicologiche che portano, per alcune patologie, ad una percentuale di guarigioni per nulla trascurabile [Bro98].

Come si fa allora a verificare l'efficienza di un farmaco? Il metodo è proprio quello delle tabelle di contingenza illustrato nell'esercizio 6.14 di pagina 244: al campione A di ammalati si somministra un placebo, al campione B il farmaco; il test χ^2 dirà se l'azione del farmaco dà effetti statisticamente diversi dal placebo. Questo tipo di esperimenti è detto a doppio cieco, perché né il medico né il paziente sanno se fanno parte del gruppo che sta sperimentando il farmaco o il placebo. Tra le pratiche alla moda che non hanno mai superato il test a doppio cieco ricordiamo l'omeopatia e la pranoterapia. È quindi corretto affermare che queste "terapie" guariscono; tuttavia si dovrebbe sempre

ricordare che in questi casi non si sono mai misurati effetti terapeutici diversi in modo significativo da quelli del placebo.

A parte questi problemi legati alla sperimentazione biomedica, i risultati scientifici sono assolutamente indipendenti dai voleri consci od inconsci di chi sperimenta. Non è escluso che un giorno possa accadere di trovarsi di fronte ad una situazione differente, ma fino a questo momento, in tutte le misure nel campo della chimica, della fisica e della biologia condotte a regola d'arte, ciò non è mai avvenuto.

In linguaggio non tecnico, comprensibile a tutte le persone di buon senso, potremmo dire che il metodo scientifico può essere condensato in due principi equivalenti, del tutto ovvi anche se purtroppo poco utilizzati: sono le teorie che crollano davanti ai fatti (e non viceversa); oppure: bisogna analizzare la realtà così com'è e non come si vorrebbe che fosse.

Tecnicamente, questa metodologia corrisponde a quella che nella scienza è nota come *procedura di falsificazione*: se avete una teoria, dovete in tutti i modi cercare di dimostrare che è falsa; se anche un solo fatto sperimentale è in disaccordo con essa la teoria va rivista; se tutti i risultati sperimentali sono in accordo, la teoria va (temporaneamente) accettata. Non esistono pertanto teorie vere, ma solo teorie errate (o valide limitatamente ad un ambito ristretto di fenomeni), in quanto falsificate da uno o più esperimenti, e teorie valide, ovvero non ancora falsificate. Queste ultime, come la teoria della relatività e la meccanica quantistica, fanno parte del patrimonio scientifico corrente.

A questo punto dovrebbe esservi del tutto chiaro che il confronto statistico tra un risultato sperimentale ed un modello, attraverso il calcolo del livello di significatività e la valutazione della probabilità di sbagliare rigettando (falsificando) l'ipotesi (errore del I tipo) è *un aspetto tecnico particolare ma fondamentale del procedimento generale di falsificazione della scienza moderna*. A titolo di esempio, potete rivedere gli esercizi 3.17, 6.11 e 11.5.

Caratteristica del processo di falsificazione, ovvero di chi possiede una mentalità e una cultura scientifica, è quella di porre l'attenzione sui fallimenti, sui rigetti e sugli insuccessi piuttosto che sui successi di una teoria o di una ipotesi. Questo punto è fondamentale e non è per niente ovvio. Supponete di fare un oroscopo e di predire ai nati di un dato segno che entro una certa settimana foreranno una gomma dell'auto; su (poniamo) 5000 persone che vi leggono questo capiterà per caso (diciamo) a 10 persone. Se queste 10 persone non hanno una mentalità scientifica esse saranno impressionate dalla vostra predizione, diventeranno vostri seguaci e diffonderanno ad altri la loro positiva esperienza. Continuando a fare oroscopi, le vostre predizioni saranno sempre coronate da successo (in senso statistico); in questo modo a poco a poco accumulerete un numero ragguardevole di fan e potrete magari diventare un astrologo famoso (e ricco).

Il procedimento di falsificazione non va inteso in senso rigido e schematico: se un esperimento falsifica una teoria che ha resistito a centinaia o migliaia di esperimenti precedenti, è forse più probabile che l'esperimento sia sbagliato e la teoria valida che non il contrario. Periodicamente scienziati dilettanti

comunicano clamorose scoperte, come quella del moto perpetuo o della non conservazione della carica elettrica; se questi esperimenti fossero veri, l'intero edificio della fisica moderna crollerebbe. Occorre pertanto una certa prudenza nel giudicare e valutare risultati scientifici nuovi e clamorosi.

Affinché il procedere della scienza possa avvenire secondo lo schema che vi abbiamo illustrato, occorre quindi che negli scienziati vi sia buona fede e controllo reciproco e che vi siano meccanismi di verifica dell'informazione scientifica. Le regole che a questo proposito la comunità degli scienziati si è data prevedono che i risultati scientifici, per essere considerati tali, debbano essere pubblicati su riviste "serie", cioè gestite da alcuni tra i massimi esperti mondiali della materia (ogni addetto ai lavori sa quali sono quelle del campo in cui opera). Le teorie possono così essere controllate e gli esperimenti ripetuti: si innesca allora a livello internazionale una fase transitoria, che potremmo definire di validazione, alla fine della quale il risultato scientifico, teorico o sperimentale, si può considerare acquisito. L'analisi statistica dei dati e dei risultati di esperimenti differenti gioca a questo punto quasi sempre un ruolo di primo piano. Gli scienziati operano dunque in una sorta di zona franca, autogestendo l'informazione scientifica in assoluta autonomia. Questo, a nostro avviso, è una delle caratteristiche positive più significative dei nostri tempi. È allora molto importante, per salvaguardare questa autonomia che la scienza ha saputo conquistarsi, che ogni ricercatore sia conscio di quanto sia importante pubblicare su una rivista o comunicare ad un congresso un risultato scientifico, e che operi pertanto con scrupolo e correttezza.

Dopo aver descritto la fisiologia del sistema, vogliamo ora illustrarne alcune possibili patologie. Essenzialmente, per quanto riguarda l'esecuzione degli esperimenti, esse sono di tre tipi:

- utilizzare la frode;
- trattare i dati in modo scorretto (*cooking, trimming*);
- subire l'effetto di trascinamento (effetto *bandwagon*).

Passeremo ora brevemente in rassegna queste tre patologie.

L'utilizzo della frode nella scienza, cioè l'esistenza di scienziati imbrogliatori, si è fino ad ora rivelato un aspetto fortunatamente secondario. I meccanismi di selezione nelle università ed il controllo incrociato tra scienziati, che si realizza nei congressi e sulle riviste scientifiche, hanno finora consentito di rivelare tempestivamente frodi e trucchi. Essi si sono verificati principalmente nel settore biomedico, dove più spesso si verifica uno stretto intreccio tra scienza e interessi economici. Spesso viene citato come caso limite quello di W.T. Summerlin, che a metà degli anni '70 dipinse chiazze nere sul pelo di topi bianchi per simulare innesti [DT95]. Nel campo della fisica è famoso il caso di E. Rupp che, come successivamente ammise lui stesso, si inventò letteralmente, all'inizio degli anni '30, alcuni risultati sulla polarizzazione nella doppia diffusione di elettroni. È da notare che i risultati di Rupp falsificavano senza ombra di dubbio la teoria relativistica dell'elettrone di Dirac, che era già ricca di conferme. Tuttavia la comunità dei fisici, allora già particolarmente attiva

e vivace, scoprì in breve tempo l'imbroglio, costringendo Rupp a ritrattare i suoi risultati.

La seconda patologia, il trattamento scorretto dei dati, consiste in operazioni di ripulitura (*trimming*) e manipolazione (*cooking*) dei risultati, ed è molto più insidiosa della precedente, in quanto molto più difficile da scoprire. Essa non è dovuta tanto alla malafede di chi opera quanto all'ignoranza di tecniche di analisi e trattamento dei dati. I casi più frequenti riguardano la mancata correzione di effetti sistematici, oppure *l'uso scorretto della statistica e delle tecniche di calcolo e propagazione degli errori di misura*. Gli esperimenti di Millikan, trattati nel paragrafo precedente, rientrano proprio in questa casistica. Non sappiamo infatti se Millikan scartò alcune misure perché in malafede; certo è che non tenne in considerazione effetti sistematici importanti.

Un altro aspetto di questa patologia è la esagerazione degli errori di misura, a conferma di teorie alla moda. In questo modo possibili discrepanze significative tra teoria ed esperimento vengono nascoste, vanificando totalmente il principio di falsificazione. Anche questo dimostra, ancora una volta, quanto sia importante per uno sperimentatore *saper calcolare bene le incertezze di misura*.

La ripetizione degli esperimenti ed il controllo tra scienziati sono le armi utilizzate con successo dalla comunità scientifica per individuare esperimenti compiuti non correttamente o teorie tecnicamente sbagliate. Talvolta però esse non sono sufficienti. A titolo di esempio, esaminate le varie determinazioni della costante di gravitazione universale G , tratte da da [Gil97] e riportate in Fig. 11.15; qui le barre d'errore, a differenza della convenzione standard, *sono*

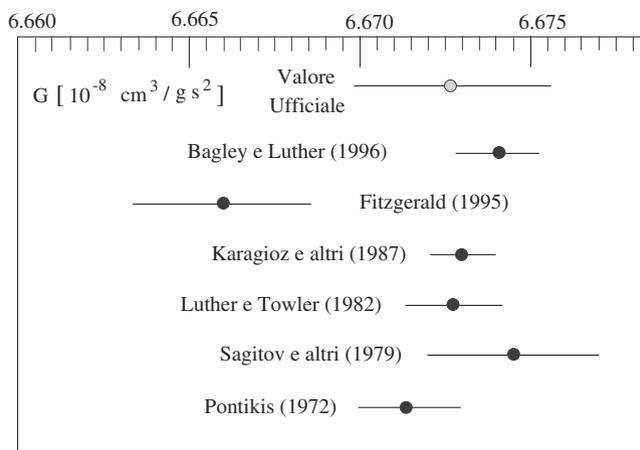


Fig. 11.15. Alcune misure della costante di gravitazione universale G col metodo del pendolo a torsione. Le barre d'errore sono riportate a tre deviazioni standard, per evidenziare l'incompatibilità tra le due misure più recenti

disegnate a tre deviazioni standard invece che ad una, per mostrare l'incompatibilità delle due misure più recenti. Si tratta di errori mal calcolati o di una dipendenza di G dal tipo di materiale usato negli esperimenti? Questo è un punto tuttora aperto, legato alla possibile esistenza, nella forza gravitazionale, di una componente sconosciuta di intensità molto debole (la cosiddetta "quinta forza"), che priverebbe G del carattere di universalità.

E veniamo ora all'effetto più subdolo e pericoloso di tutti, l'effetto di trascinamento, spesso nella letteratura inglese citato come effetto *bandwagon*, che si può forse tradurre efficacemente anche come "effetto diligenza".

Questo effetto è già stato chiaramente descritto nella citazione di Feynman che abbiamo riportato a pagina 484. Come avrete capito, si tratta dell'influenza psicologica sullo sperimentatore della situazione scientifica preesistente. Infatti, se un esperimento fornisce risultati in netto contrasto con teorie accettate o esperimenti precedenti, lo sperimentatore si affanna a ricercare possibili cause di queste discrepanze, finché non arriva ad ottenere un risultato vicino a quello atteso; a questo punto si ritiene soddisfatto e diventa meno attento. In questo modo, però, si possono generare situazioni che vedono i risultati sperimentali tutti in accordo tra loro *e tutti sbagliati*. Un buon esempio è costituito dalla Fig. 11.16, tratta da [Mor92], che riporta le misure successive del parametro ρ di Michel, che caratterizza la distribuzione in energia degli elettroni emessi in certi decadimenti nucleari. Ogni misura è in accordo con la precedente, ma le prime sono incompatibili col valore di 0.75, che è il valore teorico atteso, verso il quale le misure si avvicinano in modo monotono crescente. Con tutta probabilità qui si sono verificati effetti di trascinamento. Un altro esempio, riportato in [Ber85], è la misura del parametro η_{+-} , che è un rapporto tra i decadimenti di alcune particelle nucleari instabili, i mesoni K . Essendo questo un parametro molto difficile da determinare ma di cruciale importanza per la fisica delle particelle, esso è stato oggetto di numerose misure

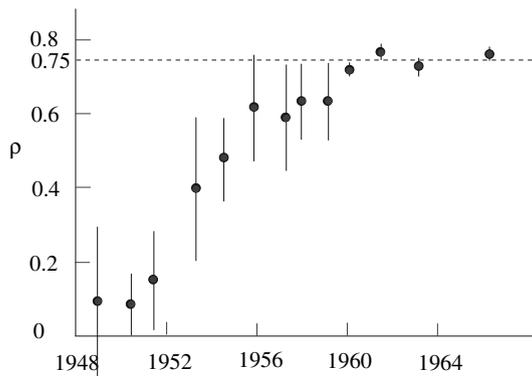


Fig. 11.16. Misure del parametro ρ di Michel riportate in corrispondenza dell'anno in cui sono state effettuate. Ogni misura, entro gli errori, è in accordo con la precedente

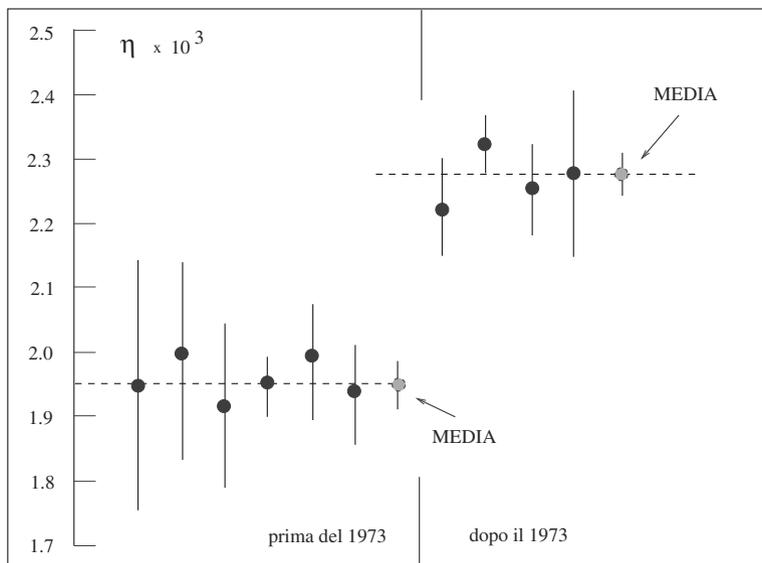


Fig. 11.17. Misure del parametro η_{+-} in ordine di data di pubblicazione

succedutesi nel corso degli anni. Esse sono riportate in Fig. 11.17: le medie prima e dopo il 1973 sono incompatibili. Sembra anche qui evidente un effetto di trascinamento nei due gruppi di misure. Il valore ufficiale (2000) è la media pesata del secondo gruppo di misure:

$$\eta_{+-} = (2.276 \pm 0.017) 10^{-3} .$$

A differenza degli altri effetti, quello di trascinamento non implica la mala-fede o l'ignoranza dello sperimentatore. Al contrario, si tratta di un effetto psicologico del quale possono cadere vittime ottimi ricercatori, ben preparati tecnicamente e molto esperti. Probabilmente il rimedio migliore contro questo effetto è, come sottolinea Feynman nella citazione di pagina 484, sapere che esso esiste.

Crediamo con questa discussione di avervi mostrato che gli scienziati, a differenza di quanto si crede, non sono infallibili, né si ritengono tali. Tuttavia, pur consci che stiamo per fare una affermazione molto impegnativa, ci sentiamo di dire che il *metodo scientifico*, così come ve lo abbiamo descritto nei suoi aspetti tecnici e di principio, alla lunga è infallibile e riesce ad essere immune dagli errori che *i singoli scienziati* possono commettere.

Problemi

11.1. Se la massa m e l'accelerazione a sono misurate indipendentemente con un errore statistico del 4% e del 5% rispettivamente, qual è l'errore statistico relativo della forza $F = ma$?

11.2. Una sorgente radioattiva di vita media $\tau = 5$ giorni, avente inizialmente una attività di $I_0 = 1000$ decadimenti/s, di incertezza trascurabile, ha una attività residua di $I = 10$ decadimenti/s, con una incertezza di $s_I = \pm 1$ decadimento/s. Sapendo che la legge del decadimento è $I/I_0 = \exp(-t/\tau)$, determinare il tempo trascorso con il relativo errore.

11.3. Un monitor di radiazione registra $I = 157$ conteggi/s in presenza di una sorgente e $F = 620$ conteggi di fondo in 10 secondi in assenza di sorgente. Determinare il numero di conteggi a fondo sottratto con il relativo errore ed il rapporto segnale/fondo.

11.4. Il fondo atteso di un esperimento di conteggio (misurato accuratamente in fase di calibrazione) è di 10 conteggi/s. In una prova (fondo più eventuale segnale) si registrano in un secondo 25 conteggi. Utilizzando l'approssimazione gaussiana, trovare il limite superiore dei conteggi per il solo segnale con $CL = 95\%$.

11.5. Un risultato scientifico viene riportato come 5.05 ± 0.04 , con la specifica che si tratta della media di 4 misure e che all'intervallo va associato un $CL = 95\%$. Trovare la precisione della singola misura.

11.6. Il volume di un cilindro $V = (\pi R^2)L$ è ottenuto misurando il raggio di base R e l'altezza L . Per la precisione della misura indiretta di V , è più importante ridurre l'errore percentuale su R o su L ?

11.7. La misura di un angolo fornisce il valore $\theta = (30 \pm 2)^0$. Calcolare $\sin \theta$ con il relativo errore.

11.8. Un fascio polarizzato di particelle viene fatto diffondere su un bersaglio e la polarizzazione viene misurata come $P = (N_+ - N_-)/(N_+ + N_-)$, dove N_+ è il numero di particelle deviate "in su", N_- il numero di quelle deviate "in giù" ed $N = N_+ + N_-$ è il numero totale di particelle diffuse. È ben noto ai fisici che la polarizzazione misurata va scritta come $P \pm s(P) = P \pm \sqrt{(1 - P^2)/N}$. Verificate la formula dell'errore, sia nel caso di N fisso che variabile. Utilizzate i risultati del par. 6.16 di pagina 228.

11.9. Una tensione è misurata con uno strumento di classe 1 come $V = 10.00 \pm 0.05$ V, dove l'errore è l'intervallo di sensibilità. Determinare il valore della funzione $f(V) = \exp(0.1V)$ col relativo errore. Quali sono i livelli di confidenza?

11.10. La tensione di uscita E_2 di un partitore con due resistenze R_1 ed R_2 e tensione di ingresso E_1 è data da $E_2 = E_1 R_1 / (R_1 + R_2)$. Se $R_1 = R_2 = 1000 \pm 50 \Omega$ ed E_1 è misurato all'1% come $E = 10.00 \pm 0.05$ V, calcolare E_2 . Verificare con una simulazione i livelli di confidenza. La misura all'1% della tensione di uscita è $E_2 = 4.91 \pm 0.02$ V. Dire se la misura è in accordo con le previsioni. Nota: tutte le incertezze sono errori massimi di sensibilità ($CL = 100\%$).

11.11. La velocità costante di una slitta a cuscino d'aria su un piano orizzontale è misurata come $v = s/t$. Lo spazio percorso è $s = 2$ m, con una incertezza dovuta alla sensibilità del metro usato stimata in ± 1 mm. Il tempo t , misurato ripetendo la prova 20 volte, risulta variare in modo aleatorio; l'istogramma dei tempi ha parametri $m_t = 5.35$ ed $s_t = 0.05$ s. Determinare la velocità e commentare.

11.12. Un contatore piano rettangolare fa registrare 1750 conteggi in un secondo. Le dimensioni del contatore sono $a = 30 \pm 0.5$ cm e $b = 50 \pm 0.5$ cm, dove l'errore è di tipo sistematico. Trovare la frequenza di conteggio nelle unità conteggi/(m² s) con errore.

11.13. L'energia dei fotoni provenienti da una transizione atomica viene trasformata in impulsi elettrici, la cui tensione di picco viene misurata con un analizzatore multicanale. La riga spettrale appare sul display con un profilo gaussiano di $\sigma \simeq 20$ canali. Tarando l'apparato con un generatore di impulsi elettrici di tensione rigorosamente costante, si ottiene sull'analizzatore una curva gaussiana di $\sigma_0 \simeq 5$ canali, che rappresenta la dispersione introdotta dalla misura. Determinare la larghezza effettiva σ_r della riga.

Appendice A

Tabella dei simboli

<i>Simbolo</i>	<i>Significato</i>
\mathbb{R}, \mathbb{R}^1	numeri reali
\mathbb{R}^n	n-ple reali
X, Y, Z, \dots	variabili aleatorie prima di una prova
x, y, z, \dots	valore assunto dalle variabili dopo una prova
\underline{X}	campione casuale di dimensione n di X : $\underline{X} = (X_1, X_2, \dots, X_n)$
\mathbf{X}	vettore di m variabili aleatorie: $\mathbf{X} = (X_1, X_2, \dots, X_m)$
$\underline{\mathbf{X}}$	campione casuale di dimensione n di \mathbf{X} : $\underline{\mathbf{X}} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$
$\underline{x}, \underline{x}$	osservazione di un campione casuale
$\langle X \rangle, E[X]$	operatore media o valore atteso di X
μ_x	valore dell'operatore $\langle X \rangle$
$\text{Var}[X]$	operatore varianza di X
σ_x^2	valore dell'operatore $\text{Var}[X]$
$\sigma[X]$	deviazione standard di X : $\sqrt{\text{Var}[X]} \equiv \sigma[X]$
σ_x	valore dell'operatore $\sigma[X]$
$\text{Cov}[X, Y]$	operatore covarianza di X e Y
σ_{xy}	valore dell'operatore $\text{Cov}[X, Y]$
M	media campionaria
$m, \langle x \rangle$	valore di M dopo una prova
S^2	varianza campionaria
s^2	valore di S^2 dopo una prova

<i>Simbolo</i>	<i>Significato</i>
S	deviazione standard campionaria
s	valore della deviazione standard stimata dopo una prova o errore statistico (a volte detto scarto quadratico medio)
$s_{xy}, s(x, y)$	valore di $\text{Cov}[X, Y]$ dopo una prova
$\rho(x, y)$	valore vero del coefficiente di correlazione
$r(x, y)$	valore campionario del coefficiente di correlazione
\simeq	approssimativamente uguale
\sim	distribuito come
\implies	implica
$p_X(x)$	funzione di densità di probabilità di X
$F_X(x)$	funzione cumulativa o di ripartizione di X
$f(x)$	espressione matematica contenente x
$f(\cdot)$	forma funzionale di una variabile
$N(\mu, \sigma^2)$	distribuzione normale di parametri μ e σ
$g(x; \mu, \sigma)$	funzione densità di $X \sim N(\mu, \sigma^2)$
$\Phi(x)$	funzione cumulativa di $X \sim N(\mu, \sigma^2)$
t_α	quantile della densità di Gauss o di Student
$U(a, b)$	distribuzione uniforme in $[a, b]$
$\chi^2(\nu)$	distribuzione chi-quadrato (o chi-quadro) con ν gradi di libertà
$Q(\nu)$	variabile distribuita come $\chi^2(\nu)$
χ^2	valori assunti da $Q(\nu)$ dopo una prova
χ_α^2	quantile della densità di χ^2
$\chi_R^2(\nu)$	distribuzione chi-quadrato ridotto con ν gradi di libertà
$Q_R(\nu)$	variabile distribuita come $\chi_R^2(\nu)$; $Q_R(\nu) = Q(\nu)/\nu$
χ_R^2	valori assunti da $Q_R(\nu)$ dopo una prova
$\chi_{R\alpha}^2$	quantile della densità di χ_R^2
$L(\theta; \underline{x})$	funzione di verosimiglianza
$\hat{\theta}$	stima puntuale di massima verosimiglianza (ML) o dei minimi quadrati (LS) del parametro θ

Appendice B

Funzioni generatrici

In questa appendice accenniamo brevemente ad uno dei metodi di analisi più importanti del calcolo delle probabilità, il metodo delle funzioni generatrici, che consente di risolvere in modo sintetico ed elegante molti problemi, senza ricorrere a complicate somme od integrazioni.

Il metodo associa alla variabile aleatoria X una nuova variabile, definita dalla funzione e^{tX} , dove t è una variabile reale che non ha un particolare significato statistico. Il valor medio $\langle e^{tX} \rangle$ è una funzione $G_X(t)$ che prende il nome di funzione generatrice dei momenti o Mgf (*Moment generating function*):

$$G_X(t) = \langle e^{tX} \rangle = \begin{cases} \sum_i p_i e^{tx_i} & \text{variabili discrete} \\ \int e^{tx} p(x) dx & \text{variabili continue.} \end{cases} \quad (\text{B.1})$$

Per $t = 0$ la somma e l'integrale sono ovviamente sempre convergenti. Per $t \neq 0$ le (B.1) possono non risultare convergenti. Qui tratteremo alcuni casi, relativi alle distribuzioni più importanti, in cui si ha la convergenza per valori $|t| < M$, dove M è un numero positivo arbitrario. Per variabili continue la funzione generatrice, a meno di un segno nell'esponente che compare nella seconda delle (B.1), è la trasformata di Laplace della densità di X .

Sviluppando in serie G_X e sfruttando le proprietà di linearità della media si ha:

$$G_X(t) = 1 + t \langle X \rangle + \frac{t^2}{2} \langle X^2 \rangle + \frac{t^3}{6} \langle X^3 \rangle + \dots \quad (\text{B.2})$$

e quindi l'importante relazione:

$$\langle X^n \rangle = \left. \frac{d^n G_X}{dt^n} \right|_0 = G_X^{(n)}(0), \quad (\text{B.3})$$

che lega la derivata n -esima di G calcolata nell'origine al momento di ordine n di X .

Risolvendo gli integrali o le somme della (B.1), si possono ottenere abbastanza facilmente le funzioni generatrici delle distribuzioni di probabilità più importanti. Alcune di queste sono riportate in Tab. B.1.

Tabella B.1. Funzioni generatrici (Mgf) di alcune distribuzioni di probabilità

<i>Densità</i>	<i>Funzione generatrice $G_X(t)$</i>
binomiale $b(x; n, p)$	$[pe^t + (1 - p)]^n$
poissoniana $p(x; \mu)$	$\exp[\mu(e^t - 1)]$
esponenziale $e(x; \lambda)$	$\lambda/(\lambda - t)$, $t < \lambda$
gaussiana $g(x; \mu, \sigma)$	$\exp(\mu t + \sigma^2 t^2/2)$
chi-quadrato $\chi^2(\nu)$	$(1 - 2t)^{-\nu/2}$

L'importanza della funzione generatrice sta nella proprietà seguente: se $Y = X_1 + X_2$ e le variabili aleatorie X_1 e X_2 sono indipendenti, dalla (4.9) di pagina 119 segue che la Mgf di Y

$$G_Y(t) = \langle e^{t(X_1+X_2)} \rangle = \langle e^{tX_1} \rangle \langle e^{tX_2} \rangle = G_{X_1}(t)G_{X_2}(t), \quad (\text{B.4})$$

è il prodotto delle Mgf di X_1 e X_2 .

Utilizzando la Mgf, si può dimostrare facilmente il teorema Limite Centrale (3.1) di pagina 86 per una somma di n variabili X_i indipendenti aventi tutte la stessa densità di parametri μ e σ . Infatti, se

$$Y = \frac{\sum_i X_i/n - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}} \frac{\sum_i (X_i - \mu)}{\sigma} = \frac{1}{\sqrt{n}} \sum_i Z_i, \quad Z_i = \frac{X_i - \mu}{\sigma}, \quad (\text{B.5})$$

dalla (B.4) si ottiene:

$$\begin{aligned} G_Y(t) &= \langle e^{tY} \rangle = \prod_i \left\langle \exp\left(\frac{t}{\sqrt{n}} Z_i\right) \right\rangle \\ &= \left[1 + \frac{t}{\sqrt{n}} \langle Z \rangle + \left(\frac{t}{\sqrt{n}}\right)^2 \frac{\langle Z^2 \rangle}{2!} + \left(\frac{t}{\sqrt{n}}\right)^3 \frac{\langle Z^3 \rangle}{3!} + \dots \right]^n. \end{aligned} \quad (\text{B.6})$$

Dato che $\langle Z \rangle = 0$ e $\langle Z^2 \rangle = 1$, passando al limite si trova il risultato

$$\lim_{n \rightarrow \infty} G_Y(t) = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n} t^2 \right)^n = e^{t^2/2}, \quad (\text{B.7})$$

che mostra come, per n grande, la Mgf di Y tenda a quella di una densità gaussiana standard (si veda la Tab. B.1).

Dimostriamo anche, facendo uso della Mgf, il teorema di additività (3.4) delle variabili χ^2 di pagina 97. Se

$$Y = Q(\nu_1) + Q(\nu_2) \quad (\text{B.8})$$

e $Q(\nu_1) \sim \chi^2(\nu_1)$, $Q(\nu_2) \sim \chi^2(\nu_2)$ sono indipendenti, allora $G_{\nu_1}(t)G_{\nu_2}(t) = G_Y(t)$ e quindi

$$G_Y(t) = (1 - 2t)^{-(\nu_1 + \nu_2)/2} \implies Y \sim \chi^2(\nu_3) \quad \nu_3 = \nu_1 + \nu_2 \quad (\text{B.9})$$

dalla Tab. B.1. Se invece $Q(\nu_1) \sim \chi^2(\nu_1)$ e $Q(\nu_2) \sim \chi^2(\nu_2)$, è facile mostrare, invertendo la (B.4), che se Y e $Q(\nu_2)$ sono indipendenti, vale la proprietà:

$$Q(\nu_1) = Y + Q(\nu_2) \implies Y \sim \chi^2(\nu_1 - \nu_2) . \quad (\text{B.10})$$

Appendice C

Soluzioni dei problemi

Capitolo 1.

1.1. Se C è il cambio di porta, A l'auto dietro la prima porta scelta e \bar{C} e \bar{A} gli eventi complementari, dalla partizione dell'evento certo si ha: $P(C) = P(C|A)P(A) + P(C|\bar{A})P(\bar{A}) = 0 \cdot 1/3 + 1 \cdot 2/3 = 2/3$, $P(\bar{C}) = P(\bar{C}|A)P(A) + P(\bar{C}|\bar{A})P(\bar{A}) = 1 \cdot 1/3 + 0 \cdot 2/3 = 1/3$. Conviene cambiare porta.

1.2. Dalla (1.30) si ricava che il numero di possibili partite è $(4! \cdot 53!)/(13!)^4 \simeq 1.287 \cdot 10^{30}$. Poiché il numero di partite giocate è $\simeq 5.475 \cdot 10^{14}$, $P \simeq 4.25 \cdot 10^{-16}$.

1.3. $P = p_1[1 - (1 - p_2)(1 - p_3)] = 0.776$.

1.4. a) Gli elementi 1 e 2 sono in parallelo, in serie col parallelo degli elementi 3 e 4. b) $P = [1 - (1 - p)^2]^2 = 0.922$.

1.5. $P = 1 - (5/6)^3 = 0.421$.

1.6. a) $P = 1 - 7/10 = 0.30$; b) $P = 1 - 120/720 = 0.83$.

1.7. Delineiamo la non semplice soluzione: si può immaginare l'arrivo di uno dei due amici come un punto x o y localizzato a caso entro un intervallo lungo 60 minuti. X e Y non si incontrano se accade l'evento $\{X \notin [y, y + 12], Y \notin [x, x + 10]\}$. Quindi $P = 1 - (48/60)(50/60) = 1 - 2/3 = 1/3$.

1.8. Dalla formula delle probabilità totali: $P(T) = 0.14$. Dal teorema di Bayes: $P(B|T) = 0.678$. Si può usare anche il metodo grafico di Fig. 1.6.

1.9. $P\{X \leq Y\} = 1/2$. Infatti, è ragionevole assumere la probabilità come il rapporto tra l'area sopra la diagonale e l'area totale di un quadrato di lato unitario.

1.10. Con ovvia notazione:

$$P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C) = 0.165 \rightarrow 16.5\% .$$

1.11. $P\{C \geq R|C \geq 150, R \leq 155\}P\{C \geq 150, R \leq 155\} = (1/2)(5/15)(5/20) = 0.0417$.

1.12. Si ottiene $P(H_1) \simeq P(H_2) \simeq 0$, $P(H_3) = 0.03$, $P(H_4) = 0.22$, $P(H_5) = 0.47$, $P(H_6) = 0.28$. Confrontate i risultati con quelli della Tab. 1.2.

1.13. Dalla (1.38), se V_n è l'evento in cui l'amico vince n volte consecutive ($V_1 \equiv V$), definendo le ipotesi $B = \text{baro}$ e $O = \text{onesto}$, assumendo $P(B) = P(O) = 0.5$, e $P(V|O) = 0.5$, $P(V|B) = 1$, si ottiene la formula ricorsiva: $P(B|V_n) = P(B|V_{n-1})/[P(B|V_{n-1}) + 0.5(1 - P(B|V_{n-1}))]$. Si ottiene $P(B|V_5) = 0.97$, $P(B|V_{10}) = 0.999$, $P(B|V_{15}) = 0.99997$. Verificate come cambiano i risultati cambiando $P(B)$ e $P(O)$.

1.14. 1) vera, 2) falsa, 3) vera, 4) falsa.

1.15. $P(21|20)P(20) = 0.2 \cdot 0.4 = 0.08$, $P(21|21)P(21) = 0.6 \cdot 0.4 = 0.24$, $P(21|22)P(22) = 0.2 \cdot 0.1 = 0.02$, da cui $P(20|21) = 0.24$, $P(21|21) = 0.70$, $P(22|21) = 0.06$ dalla formula di Bayes.

Capitolo 2.

2.1. $b(2; 3, 1/6) = 5/72$.

2.2. Il numero di casi favorevoli è $10!/(3!7!) = 120$, quello dei casi possibili $2^{10} = 1024$. La probabilità vale $720/1024 = 0.117$, in accordo con la densità binomiale.

2.3. Dalla densità binomiale: $1 - b(0; 10, 0.1) - b(1; 10, 0.1) - b(2; 10, 0.1) = 0.0702$, pari a circa il 7%.

2.4. $\langle X + Y \rangle = \langle X \rangle + \langle Y \rangle = 100 \cdot 18/37 + 0 \cdot 19/37 + 360 \cdot 1/37 + 0 \cdot 36/37 = 58.4$. Se si immagina un insieme molto grande (al limite infinito) di N giocatori, ognuno con un capitale iniziale di 60 euro, dopo una puntata il capitale medio di questo insieme sarà sceso a 58.4 euro, con un guadagno medio per giocatore a favore del banco di 1.6 euro.

2.5. $\langle X \rangle = 6$, $\sigma[X] = 2.83$, $\langle Y \rangle = 13$, $\sigma[Y] = 5.66$.

2.6.

r	0	1	2	3
$p(r)$	$6/720$	$126/720$	$378/720$	$210/720$

Si noti che la densità è normalizzata. $\langle R \rangle = 0.9$, $\sigma[R] = 0.7$.

2.7. $p(x) = 2x$, $F(x) = x^2$, $\langle X \rangle = 2/3$, $\sigma[X] = 1/(3\sqrt{2})$.

2.8. $F(x_{0.25}) = x_{0.25}^2 = 0.25$, da cui $x_{0.25} = 0.5$. La probabilità di osservare valori ≤ 0.5 è pari al 25%.

2.9. Il viaggio dura 6 ore, 4 in andata e 2 al ritorno. Se si considera la velocità V come una variabile a due valori, si ha: $\langle V \rangle = 25 \cdot 4/6 + 50 \cdot 2/6 = 200/6 = 33.3$ Km/h. Si noti che il risultato $\langle V \rangle = (50 + 25)/2 = 37.5$ Km/h è errato.

2.10. Le prove sono 1000 (invece che 100) e i valori dello spettro sono due $x = 0, 1$ (testa o croce, appunto, invece degli 11 valori $x = 0, 1, \dots, 10$). Dalla Tab. 2.2, sommando i prodotti della prima colonna per la seconda ($0 \cdot 0 + 1 \cdot 0 + 2 \cdot 5 + 3 \cdot 13 + \dots$), otteniamo 521 teste, e, ovviamente, 479 croci, a fronte di un valore atteso di $1000 \cdot b(1, 0.5, 0) = 1000 \cdot b(1, 0.5, 1) = 1000 \cdot 0.5 = 500$.

Capitolo 3.

3.1. $\sum_x P\{Y \geq X|X = x\}P\{X = x\} \rightarrow \int_0^1 (1 - x) dx = 1/2$.

3.2. La densità di X è binomiale (gaussiana), quindi $\langle X \rangle = 500 \cdot 0.5 = 250$, $\sigma[X] = \sqrt{500 \cdot 0.25} = 11.18$. Vale la legge 3-sigma.

3.3. Dopo n passi, il percorso X ha uno spettro $X = n, n-2, \dots, -n$ di valori discreti che differiscono tra loro di 2 unità. La variabile $Y = (X + n)/2$ è allora binomiale e assume valori interi nell'intervallo $[0, n]$. Si ha quindi: $\langle Y \rangle = (\langle X \rangle + n)/2 = np = n/2$ e $\text{Var}[Y] = \text{Var}[X]/4 = np(1 - p) = 125$. Risulta allora: $\langle X \rangle = 0$ e $\sigma[X] = 2\sqrt{125} = 22.36$. Vale la legge 3-sigma.

3.4. a) 0.01576; b) 0.016.

3.5. $\mu = 200 \cdot 0.2 = 40$, $\sigma = \sqrt{200 \cdot 0.2 \cdot 0.8} = 5.66$. Utilizzando l'approssimazione gaussiana e la Tab. D.1 si ha $P\{-1.77 < t < 1.77\} = 0.923$.

3.6.

$$p(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left[-\frac{1}{2\sigma^2}(\ln y - \mu)^2\right], \quad y > 0.$$

Basta notare che $\ln Y$ è gaussiana e tenere presente che $d \ln y = dy/y$. Una distribuzione avente questa densità è detta lognormale.

3.7. FWHM=2.355 σ .

3.8. a) 0.0695 \simeq 7% ; b) 0.7165 \simeq 70%, indipendentemente dagli 8 mesi precedenti.

3.9. $\chi^2/\nu = 7/10 = 0.7$, corrispondente ad una probabilità (livello di significatività) \simeq 28%, dalla Tab. D.3.

3.10. $(8500 - 8100)/\sqrt{8500} = 4.34$. La riduzione è significativa.

3.11. $1 - \exp(-100 \cdot 0.001) = 0.0952 \simeq 9.5\%$.

3.12. Dalla densità di Poisson: $P\{X = 0\} = \exp(-10/2) = 0.0067$. La probabilità di sbagliarsi è allora \simeq 7%.

3.13. Se $\{X_1 + X_2 = n\}$ è la somma dei conteggi registrati in intervalli di tempo disgiunti, utilizzando la formula del binomio di Newton si dimostra facilmente che $P\{X_1 + X_2 = n\} = \sum_k [P\{X_1 = k\}P\{X_2 = n - k\}]$ quando le probabilità sono calcolate con la densità di Poisson (3.47).

3.14. $[-0.675, 0.675]$, interpolando linearmente dalla Tab. D.1.

3.15. Dalla Tab. D.1 si ottiene interpolando: $P\{t > 0.806\} = 0.21$ e $P\{t > 1.55\} = 0.06$: Dalle relazioni $(4.41 - \mu)/\sigma = 0.806$ e $(6.66 - \mu)/\sigma = 1.55$ si ottiene poi $\mu = 2$ e $\sigma = 3$. Potete fare anche qualche verifica numerica con la funzione `cdfnor` di SCILAB.

3.16. Assumendo l'ora come unità di misura, $\lambda t = 100/120 = 0.8333$ e quindi, dalla distribuzione di Poisson: $p(4) = (\lambda t)^4 \exp(-\lambda t)/4! = 0.00873$. Il tempo in ore è dato da $1/0.00873 = 114.5$ ore pari a $4.77 \simeq 5$ giorni.

3.17. La funzione cumulativa è $F(x) = \int_1^x (2x - 2) dx = x^2 - 2x + 1$. Dal teorema 3.5, $X^2 - 2X + 1 = R$ dove $R \sim U(0, 1)$. L'equazione di secondo grado, per $1 \leq x \leq 2$, fornisce la soluzione $X = 1 + \sqrt{R}$. Per generare X secondo la densità assegnata basta l'istruzione $X = 1 + \sqrt{\text{Random}} \leq 1$ ha una distribuzione uniforme.

3.18. a) 4.6% dalla Tab. D.1; b) dato che $\text{Var}[\sum_{i=1}^{10} x_i] = 10\sigma^2[x] = 250$ si ha $\sigma[\sum_{i=1}^{10} x_i] = 15.8$ e $t = |1050 - 1000|/15.8 = 3.16$ pari a una probabilità di $8 \cdot 10^{-4}$; c) $< 25\%$ dalla disuguaglianza di Tchebychev.

3.19. $\langle X \rangle = 1.5$, $\langle X^2 \rangle = 2.9$. $\langle 30X \rangle = 45$, $\sigma[30X] = 24$; $(80 - 45)/24 = 1.46$ pari a una probabilità del 7.2% dalla Tab. D.1.

3.20. Il numero dei difetti è $np \pm \sqrt{np(1-p)} = 10.0 \pm 3.1$, per cui $t = |16 - 10|/3.1 = 1.93$. Dalla Tab. D.1 si vede che a questo valore corrisponde una coda a destra del 2.7%, che è la percentuale richiesta.

Capitolo 4.

4.1. $p(x, y) dx dy = P\{x \leq X \leq x + dx, y \leq Y \leq y + dy\} = dx dy/(ab)$ per $0 \leq x \leq a, 0 \leq y \leq b, p(x, y) = 0$ altrimenti.

4.2. a) $(X_1 - \mu_1)^2/\sigma_1^2 = 2.71$; b) $\sum_{i=1}^2 (X_i - \mu_i)^2/\sigma_i^2 = 4.61$; c) $\sum_{i=1}^3 (X_i - \mu_i)^2/\sigma_i^2 = 6.25$. I valori si ricavano dalla Tab. D.4 di pagina 524.

4.3. L'uguaglianza è valida, in base al teorema 4.1 ed alle (4.14, 4.16).

4.4. Si tracciano le tangenti come in Fig. 4.5 e si individuano i punti di intersezione (x_2, y_2) e (x_1, y_1) . Dalla (4.55) si ha poi $(\sigma_x/\sigma_y)(y_2 - y_1)/(x_2 - x_1) = \rho$.

4.5. $\text{Cov}[Z, U] = \langle ZU \rangle - \langle Z \rangle \langle U \rangle = \langle (aX + b)(cY + d) \rangle - (a \langle X \rangle + b)(c \langle Y \rangle + d) = ac(\langle XY \rangle - \langle X \rangle \langle Y \rangle)$. Dato che $\sigma[Z] = a\sigma[X]$ e $\sigma[U] = c\sigma[Y]$, si ha $\rho[Z, U] = (\langle XY \rangle - \langle X \rangle \langle Y \rangle) / (\sigma[X]\sigma[Y]) = \rho[X, Y]$.

4.6. Basta rifarsi alla (4.77) cambiando i limiti: $P\{X \geq 180, Y \geq 170\} = [0.5 - E(0.625)] \cdot [0.5 - E(0.833)] = 0.266 \cdot 0.202 = 0.0537 \simeq 5.4\%$. I valori della funzione $E(\dots)$ si ricavano dalla Tab. D.1.

4.7. Occorrerebbe integrare per via numerica la (4.40) nella regione $X \in [180, +\infty)$, $Y \in [170, +\infty)$. Si noti anche che l'ellisse di concentrazione non fornisce in questo caso la soluzione giusta. Per una soluzione col metodo Monte Carlo si veda anche il problema 7.11 a pagina 305.

4.8. La densità p_{XY} è data dalla tabella:

(x, y)	p_{XY}	(x, y)	p_{XY}	(x, y)	p_{XY}
(1, 1)	4/36	(3, 1)	1/36	(5, 1)	1/36
(1, 3)	1/36	(3, 3)	4/36	(5, 3)	1/36
(1, 5)	1/36	(3, 5)	1/36	(5, 5)	4/36
(2, 1)	1/36	(4, 1)	1/36	(6, 1)	1/36
(2, 2)	3/36	(4, 3)	1/36	(6, 3)	1/36
(2, 3)	1/36	(4, 4)	3/36	(6, 5)	1/36
(2, 5)	1/36	(4, 5)	1/36	(6, 6)	3/36

Le densità marginali p_X e p_Y valgono:

(x, y)	1	2	3	4	5	6
p_X	1/6	1/6	1/6	1/6	1/6	1/6
p_Y	1/4	1/12	1/4	1/12	1/4	1/12

da cui risulta che la variabile X , corrispondente al primo dado, è uniformemente distribuita, mentre la Y , che è correlata alla X , ha una densità diversa. Dalle (4.17, 4.18) è possibile ricavare i valori delle medie e delle deviazioni standard: $\mu_x = 3.5$, $\mu_y = 3.25$, $\sigma_x = 1.708$, $\sigma_y = 1.726$. Calcolando infine la covarianza tramite la (4.21) e i valori della tabella di p_{XY} , otteniamo il valore $\sigma_{xy} = 1.458$.

Capitolo 5.

5.1. $F_Z(z) = P\{-\ln X \leq z\} = P\{X \geq e^{-z}\} = 1 - e^{-z}$, da cui $dF/dz = p_Z(z) = e^{-z}$, $z \geq 0$.

5.2. Per $x \geq 0$ c'è una sola radice $x = \sqrt{z}$. Si ottiene pertanto: $p_Z(z) = 2(1 - \sqrt{z}) / (2\sqrt{z}) = 1/\sqrt{z} - 1$, $0 < z \leq 1$.

5.3. Definendo la variabile ausiliaria $W = Y$, si ha $X = ZW$ e $Y = W$; le derivate $\partial f_1^{-1} / \partial Z = W$, $\partial f_1^{-1} / \partial W = z$, $\partial f_2^{-1} / \partial Z = 0$, $\partial f_2^{-1} / \partial W = 1$, permettono di calcolare la jacobiano $|J| = |W|$. Il risultato è quindi $p_Z = \int |w| p_{XY}(zw, w) dw$.

5.4. Dalla (5.32) si ha $p_Z(z) = \int_{-\infty}^{+\infty} p_Y(z-x)p_X(x) dx$, con $x \geq 0$, $(z-x) \geq 0$, da cui $p_Z(z) = \int_0^z p_Y(z-x)p_X(x) dx = z \exp[-z]$.

5.5. Si deve estendere per induzione la procedura del problema precedente e si trova la distribuzione di Erlang o gamma $\Gamma(n, \lambda)$, avente densità: $e_k(t) = \lambda(\lambda t)^{n-1} e^{-\lambda t} / (n-1)!$.

5.6. Valgono le condizioni $0 \leq Z \leq 1$, $W \geq 0$, e le funzioni inverse sono: $X = ZW$ e $Y = W(1-Z)$. Lo jacobiano vale $|J| = |W| = W$ e la densità cercata è: $p_{ZW}(z, w) dz dw = w \exp[-w] dz dw$. In base al teorema 4.1, Z e W sono indipendenti e Z è una variabile uniforme.

5.7. Dalla (5.27), tenendo presente che $z = f(x, y) = xy$, $f^{-1}(z, y) = z/y$, $\partial f^{-1} / \partial z = 1/y$ e che nel rapporto $x = z/y$ si deve avere $y \geq z$ per assicurare i limiti $0 \leq x \leq 1$, otteniamo: $p_Z(z) = \int_z^1 (1/y) dy = -\ln z$, $0 \leq z \leq 1$. Dall'integrale notevole $\int z^n \ln z dz = \ln z z^{n+1} / (n+1) - z^{n+1} / (n+1)^2$ si ottiene poi facilmente $\langle Z \rangle = 1/4 = 0.25$ e $\text{Var}[Z] = \langle Z^2 \rangle - \langle Z \rangle^2 = (1/9) - (1/16) \simeq 0.049$.

5.8. Se T_1 e T_2 sono i tempi di guasto, il sistema cessa di funzionare ad un tempo $T = T_1 + T_2$. La densità di probabilità è quindi data dalla convoluzione delle due densità. Si ha pertanto $\lambda^2 \int \exp[-\lambda(t-u)] \exp(-\lambda u) du = \lambda^2 t \exp(-\lambda t)$, che è la densità gamma (3.57). Il tempo medio di vita del sistema è $\langle T \rangle = 2/\lambda$. Si confronti il risultato con quello dell'esercizio 4.1.

5.9. Caso a): $\langle Z_1 \rangle = 0$, $\langle Z_2 \rangle = 0$, $\text{Var}[Z_1] = 13$ e, dalla (5.68), $\text{Var}[Z_2] = 1$. Si noti che in approssimazione lineare si avrebbe, dalla (5.66), $\text{Var}[Z_2] = 0$. Utilizzando il metodo della (5.83) si ottiene facilmente $\text{Cov}[Z_1, Z_2] = 0$. La covarianza e il coefficiente di correlazione lineare sono nulli, pur esistendo una relazione di dipendenza tra le variabili. Caso b): $\langle Z_1 \rangle = 5$, $\langle Z_2 \rangle = 1$, $\text{Var}[Z_1] = 13$, $\text{Var}[Z_2] = 3$, $\text{Cov}[Z_1, Z_2] = 5$, $\rho[Z_1, Z_2] = 0.8$.

5.10. Dato che X e Y sono gaussiane, in base all'esercizio 5.3 lo è anche $Y - X$. Dalla (5.66) risulta $\sigma[Y - X] = \sqrt{0.020^2 + 0.020^2} = 0.0283$. Utilizzando il metodo della variabile standard otteniamo i valori $t_1 = (0.050 - 0.100)/0.0283 = -1.768$, $t_2 = (0.150 - 0.100)/0.0283 = 1.768$, corrispondenti ad un intervallo di probabilità di $P\{0.050 \leq (Y - X) \leq 0.150\} = 2 \cdot 0.4614 = 0.923$, interpolando dalla Tab. D.1. La percentuale richiesta è pertanto: $1 - 0.923 = 0.077 \simeq 7.7\%$.

5.11. In base all'esercizio 5.4, la densità del numero totale di veicoli è poissoniana di media $\mu + \lambda$. Dalla densità binomiale si ha poi: $P(k) = n! / [k!(n-k)!] [\mu / (\mu + \lambda)]^k [\lambda / (\mu + \lambda)]^{(n-k)}$.

Capitolo 6.

6.1. a) $13 \leq R \leq 120$, b) $26 \leq R \leq 133$, c) $17 \leq R \leq 116$.

6.2. La soluzione è il valore μ che soddisfa $\mu - 1.28\sqrt{\mu} - 20 = 0$ con $\mu > 20$, quindi $\mu = 26.6$.

6.3. Il valore atteso dello scarto quadratico medio (deviazione standard) coincide col valore vero: $\langle S \rangle = 12 \text{ Kg}$. Si ha invece: $\sigma[S] \simeq \sigma/\sqrt{2N} = 0.6 \text{ Kg}$.

6.4. Dalla relazione $Np = 215 \pm \sqrt{215(1-0.2)} = 215 \pm 13$ segue $N = 215/0.2 \pm 1.65 \cdot 13/0.2 = 1075 \pm 108$, $CL = 90\%$.

6.5. Il risultato segue dal teorema 6.1, dato che la varianza campionaria S^2 è funzione dello scarto, e dal teorema 4.3. Si può anche mostrare direttamente che $\text{Cov}[M, (X_i - M)] = \langle X_i^2 \rangle / N - \langle M^2 \rangle = \sigma^2 / N - \sigma^2 / N = 0$, poiché in un campione casuale le estrazioni sono indipendenti: $\langle X_i X_j \rangle = 0$ se $i \neq j$.

6.6. a) 9.8 ± 0.2 ; b) $(x-245)/5 \simeq 1.645$, da cui $x = 253.2$, pari ad un limite superiore di 10.13 per la media. Si può anche scrivere $(\mu - 9.8)/0.2 \simeq 1.645$ da cui $\mu < 10.13$, $CL = 95\%$.

6.7. Occorre applicare la (6.105), con $s_1 = s_2 = 0.05\sqrt{10} = 0.16$. Risulta $t_s = 2.12$ con 18 gradi di libertà, corrispondente, dalla Tab. D.2, ad un SL osservato (p -value) $\simeq 2.5\%$. Questo valore piuttosto piccolo autorizza a nutrire qualche dubbio sulla omogeneità delle due macchine, che andrebbero tenute sotto controllo. Confrontare il risultato col test gaussiano delle medie.

6.8. Supponendo che le differenze siano dovute al caso (ipotesi nulla), le (6.132, 6.133) forniscono il valore $\chi_R^2(1) = 0.993$, corrispondente, interpolando dalla Tab. D.3, ad un $SL \simeq 34\%$. Non è lecito scartare l'ipotesi, quindi non si può affermare che il farmaco è efficace.

6.9. Si applica la (6.121), dove $Np_i = 100 m^{x_i} \exp[-m]/x_i!$ ed $m = 4.58$ è il valore della media campionaria ricavata dai dati. Occorre raggruppare i primi due e gli ultimi tre canali, per avere $Np_i > 5$. Si trova $\chi^2 = 2.925$ con 6 gradi di libertà, pari a $\chi_R^2(6) = 0.49$. Interpolando da Tab. D.3 si trova $P\{Q_R \geq 0.49\} \simeq 0.82$, pari a un p -value di $2(1 - 0.82) \simeq 0.36$. Risulta che in media arrivano 4.6 autobus in 5 minuti e che i dati sono compatibili con la distribuzione di Poisson, perché se scartiamo questa ipotesi abbiamo un'alta probabilità (36%) di sbagliare.

6.10. $\chi^2(1) = (356 - 375)^2/375 + (144 - 125)^2/125 = 3.85$. $P\{Q > 3.85\} < 0.05$ da Tab. D.3. Il modello è quindi rifiutato.

6.11. a) Il metodo non parametrico delle tabelle di contingenza fornisce, con 5 gradi di libertà, $\chi_R^2 = \chi^2(5)/5 = 0.84/5 \simeq 0.2$, $P\{Q_R \geq \chi_R^2\} \simeq 0.96$ dalla Tab. D.3. Il test di compatibilità è superato, ma i dati vanno guardati con sospetto, per via del χ^2 troppo piccolo. b) Applicando la (6.121) con $Np_i = 100$ per tutti i 12 canali, si ottiene $\chi^2 = 2.52$. Ogni dado contribuisce con 5 gradi di libertà, per cui $\chi_R^2(10) = 2.52/10 = 0.252$, $P\{Q_R \geq \chi_R^2\} \simeq 0.99$. Se affermiamo che i dadi o i lanci sono truccati, abbiamo una probabilità di sbagliarci $\leq 1\%$. In effetti, con 100 eventi attesi per canale, le fluttuazioni statistiche a 1σ valgono $\simeq 100 \pm \sqrt{100} = 100 \pm 10$, e i 12 valori osservati sono tutti entro 1σ . È quindi ragionevole scartare l'ipotesi, perché i dati "fluttuano troppo poco".

6.12. $\rho \in [-0.013, 0.305]$. Dalla formula approssimata (6.145) si ottiene $\rho \in [-0.011, 0.311]$.

6.13. Dal test della differenza si ottiene $t = (60 - 33)/\sqrt{60 + 33} = 2.8$, da cui, in approssimazione gaussiana, $P\{T > 2.8\} = 2.6 \cdot 10^{-3}$.

6.14. Dalla propagazione degli errori e dalla formula $(x - b)/\epsilon$, con $b = 5.30 \pm 0.23$ conteggi/s, $\epsilon = 0.90 \pm 0.10$, $x = 16.7 \pm 1.3$ conteggi/s, si ottiene $\nu = 12.7 \pm 2.0$ conteggi/s.

6.15. Il χ^2 vale: $(29 - 19)^2/19 + (18 - 19)^2/19 + \dots = 11.7$ con 5 gradi di libertà, pari ad un livello $< 5\%$. Il test non è superato.

6.16. Dal test della differenza si ottiene $(10500 - 2400 - 670)/\sqrt{10500 + 90601} = 4.4$. I due risultati non sono compatibili.

6.17. Il test χ^2 fornisce il risultato:

$\chi^2 = (22 - 21.2)^2/21.2 + (12 - 11.6)^2/11.6 + (7 - 6.37)^2/6.37 + (6 - 3.49)^2/3.49 = 1.91$ con 3 gradi di libertà. Dato che $SL \simeq 8\%$, i dati sono in accordo col modello.

6.18. Integrando la legge esponenziale sugli intervalli di tempo dell'esperimento, si ottengono i valori: 393, 239, 233, 151. Il test χ^2 fornisce:

$$\frac{(370 - 393)^2}{393} + \frac{(268 - 239)^2}{239} + \dots = 7.6 .$$

I gradi di libertà sono 4, corrispondenti a $SL \simeq 25\%$. Il modello va accettato.

6.19. Dato che $\chi^2 = 2.5$ con 3 gradi di libertà, $SL \simeq 40\%$, le misure sono compatibili.

6.20. la variabile standard è $(58 - 55)/(10/\sqrt{3}) \simeq 1$. Per non superare il test si richiede un valore di almeno 1.65, quindi il dato non supera i limiti al livello richiesto.

6.21. Il χ^2 vale $\chi^2 = (13 - 16)^2/16 + (25 - 34)^2/34 + (44 - 34)^2/34 + (16 - 16)^2/16 = 5.88$, corrispondente a $SL \simeq 15\%$. L'ipotesi va accettata.

6.22. La distribuzione esponenziale dei tempi ha $\mu = \sigma = \tau$. La somma media dei tempi è $N\tau$, la deviazione standard è $\sigma = \sqrt{N}\tau$: Dalla disuguaglianza $(1000 - 100N)/(\sqrt{N}100) \leq 1.64$ segue l'equazione di secondo grado $N^2 - 22.7N + 100 = 0$ e la soluzione $N = 11.6 \simeq 12$.

6.23. Il χ^2 vale:

$$(18 - 15)^2/15 + (14 - 15)^2/15 + \dots = 6.7$$

pari ad un $SL = 50\%$, i dati sono in accordo.

6.24. Il limite superiore con $CL = 95\%$ è dato dalla equazione $25 + 1.65\sqrt{\mu} = \mu$, da cui $\mu = 34.7 \simeq 35$. Il limite per il segnale vale perciò $35 - 10 = 25$ conteggi/s. Nota: è importante sottrarre il fondo *dopo* i calcoli.

Capitolo 7.¹

7.1. No.

7.2. Il valore vero della probabilità è $2/3$.

7.3. Occorre generare due variabili uniformi $X = 60 * \text{rndm}$, $Y = 60 * \text{rndm}$ e contare il numero di volte che $-10 \leq X - Y \leq 12$. Il risultato deve essere statisticamente compatibile col valore di $1/3$.

7.4. Il risultato esatto, che si può ricavare con considerazioni geometriche non banali, è: $1 - 3\sqrt{3}/(4\pi) = 0.586503$. Utilizzando il metodo del problema 7.9 abbiamo ottenuto una frazione di coppie pari a $3\,614\,289/6\,164\,195$, corrispondente a una probabilità di 0.58634 ± 0.00020 .

7.5. La massima efficienza di generazione ($\simeq 78\%$) viene ottenuta in corrispondenza del minimo valore di k che soddisfa la relazione $ke^x \geq p(x) \quad \forall x \geq 0$, che è verificata per $k \geq \sqrt{2e/\pi}$. Per ottenere una variabile $Z \sim N(0, 1)$ è necessario utilizzare un nuovo numero ξ e porre $z = -x$ se $0 \leq \xi \leq 0.5$; $z = x$ se $0.5 < \xi \leq 1$. In alternativa, si può anche utilizzare il metodo descritto in [Mor84].

7.6. Si esegue un ciclo di $n = 5$ tentativi in cui si ottiene la frequenza $f = x/n$, dove x è il numero di volte che si verifica la condizione $0 \leq \text{rndm} \leq 0.25$. Si incrementano di 1 i contatori t_1, t_2, t_3 quando p è contenuto nell'intervallo della 6.21, calcolato col valore trovato per f e con $t = 1, 2, 3$ rispettivamente. Questo ciclo va ripetuto per un numero grande N di volte e i valori finali di $t_1/N, t_2/N$ e t_3/N forniscono i livelli cercati. Con $N = 10\,000$ abbiamo ottenuto: $t_1 = 6570, t_2 = 9852, t_3 = 9994$. Calcolate le frequenze risultanti col relativo errore e confrontatele con i livelli di probabilità della legge 3-sigma.

7.7. Per variabili uniformi $\mu = 0.5$ e la media campionaria ha incertezza $\sigma = 1/\sqrt{12N}$. Se $N = 12$ si ottiene il valore standard $t = \sum_{i=1}^{12} \xi_i - 6$. Questo algoritmo è usato dalla routine **Gauss1** presente sul nostro sito.

7.8. $Y = \rho X + \sqrt{1 - \rho^2} Y_R$, dove X e Y_R sono variabili standard generate con la routine **gauss2(0, 1)**.

7.9. Si estraggono a caso in modo uniforme $-R \leq X \leq R, -R \leq Y \leq R$, e si accettano le coordinate quando $\sqrt{X^2 + Y^2} \leq R$.

¹ I valori che il lettore ottiene con codici di simulazione devono essere statisticamente compatibili con le soluzioni qui riportate. La compatibilità va verificata con i test statistici descritti in particolare nei paragrafi 6.12, 6.14 e 6.15.

7.10. L'istogramma proviene da una popolazione di densità $p_Z(z) = \frac{1}{\pi(1+z^2)}$, detta di Cauchy, che non possiede media ed ha varianza infinita.

7.11. Bisogna generare due variabili gaussiane di media, deviazione standard e correlazione assegnata: $X = 8 \cdot g_1 + 175$, $Y = (\rho \cdot g_1 + \sqrt{1-\rho^2} \cdot g_2) \cdot 6 + 165$, dove g_1 e g_2 sono valori normali standard dati dalle routine `gauss` o `gauss2`. Si trova poi, tra le coppie generate, la percentuale di quelle in cui $X \geq 180$ e $Y \geq 170$. Con 10 milioni di coppie abbiamo ottenuto $(10.71 \pm 0.01)\%$.

7.12. I risultati devono essere compatibili con la soluzione 5.9 di pagina 504.

7.13. Si generano due numeri $x = \xi_1$ e $\phi = 2\pi\xi_2$. Si registra il numero n di successi $x + \cos(\phi) < 0$ oppure $x + \cos(\phi) > 1$ su un totale di N tentativi. La stima di π è $2N/n \pm (2N/n^2)\sqrt{n(1-n/N)}$. Si veda la routine `Buffon`.

7.14. Una possibile soluzione è la nostra routine `Asimm`. Si ricordi che l'errore sulla deviazione standard di N_e eventi vale $\simeq \sigma/\sqrt{2N_e}$ (capitolo 6).

7.15. Se si escudono i valori limite $p < 0.01$ e $p > 99$, la condizione si raggiunge intorno a $n > 200$.

7.16. La condizione si raggiunge intorno a $\mu > 70$.

7.17. $P\{r^* > 0\} \simeq 6\%$ ($r =$ coefficiente di correlazione campionario). Il modello (anche se per pogo) va rifiutato.

Capitolo 8.

8.1. Una soluzione è il nostro codice `Scarti`, che genera $n = 100$ valori gaussiani e calcola la differenza $\Delta = x_{\max} - x_{\min}$ tra il valore massimo e minimo. Il procedimento viene ripetuto un numero N molto grande di volte per ottenere, alla fine del ciclo, l'istogramma della variabile Δ . Appare un istogramma asimmetrico, proveniente da una popolazione di densità di forma analitica non nota. Con $N = 50\,000$, abbiamo ottenuto, per $n = 100$, un campione di media e deviazione standard $m = 2.508 \pm 0.001$, $s = 0.302 \pm 0.001$. Dallo studio grafico dell'istogramma abbiamo poi determinato il valore quantile $\Delta_{0.99} = 3.32$. Il controllo deve scartare il lotto quando $\Delta > 3.32$. Si noti che il metodo è insensibile allo spostamento della media $\langle X \rangle$.

8.2. Si estraggono a caso i 5 tempi con legge esponenziale e si determinano $m_1 = \min(t_1, t_3, t_5)$, $m_2 = \min(t_2, t_3, t_4)$, $m_3 = \min(t_1, t_4)$, $m_4 = \min(t_2, t_5)$. La macchina si ferma dopo un tempo $t = \max(m_1, m_2, m_3, m_4)$. Ripetendo il ciclo 10000 volte abbiamo ottenuto un istogramma asimmetrico di parametri $m = 2.52 \pm 0.02$ e $s = 1.70 \pm 0.01$ giorni. L'istogramma contiene nell'intervallo $m \pm s$ il 73% circa dei tempi di funzionamento.

8.3. Occorre modificare la funzione `Funint` della routine `MCinteg` e porre `Mode3=0` per escludere il campionamento ad importanza. L'estremo superiore x è la variabile `i2` che va modificata al valore voluto. La soluzione è la routine `Mcinteg1` del nostro sito.

8.4. Il valore esatto di I , calcolabile analiticamente, è $I = 2 \log 2 - 1 = 0.38629 \dots$. Per $N = 1000$, con il metodo della media si ottiene $\sigma \simeq 6.25 \cdot 10^{-3}$; con il metodo "vinci o perdi" $\sigma \simeq 1.1 \cdot 10^{-2}$; con il campionamento ad importanza (scegliendo $g(x) = x$) $\sigma \simeq 1 \cdot 10^{-3}$; con il campionamento stratificato ($k = 20$ strati) $\sigma \simeq 0.3 \cdot 10^{-3}$.

8.5. Il risultato deve essere statisticamente compatibile con il valore $I = 8$. Con il metodo della media ed $N = 1000$ abbiamo ottenuto, ad esempio, $I = 8.01 \pm 0.13$. Con il metodo "vinci o perdi" e delimitando la funzione da integrare entro l'iperparallelepipedo definito dalle condizioni $-1 \leq x_1, x_2, x_3 \leq 1$, $0 \leq y \leq 3$, abbiamo invece ottenuto (sempre con $N = 1000$) $I = 7.89 \pm 0.36$.

8.6. L'equazione rappresenta una ellisse con centro nell'origine e semiassi lungo x e y di valore $\sqrt{2}$ e 1 rispettivamente. Il rettangolo che circoscrive l'ellisse ha un'area $A = 4\sqrt{2}$. Occorre scrivere un codice che estragga a caso un punto entro il rettangolo e lo accetti se $x^2/2 + y^2 \leq 1$. Il rapporto tra i punti accettati N_S e quelli generati N fornisce l'area dell'ellisse con errore secondo le (8.34, 8.35). Col programma `Ellisse` e 100 000 punti abbiamo ottenuto un'area pari a 4.433 ± 0.007 .

8.7. Le prove possono essere fatte con la nostra routine `Metrop`. Generando 10000 variabili ed utilizzando le ultime 5000 abbiamo ottenuto $m = 0.029 \pm 0.018$, $s = 0.875 \pm 0.010$ con $a = 2$ e $m = 0.017 \pm 0.024$, $s = 1.002 \pm 0.016$ con $a = 3$. Si vede che con $a = 3$ l'errore aumenta, ma la stima distorta di σ viene corretta. I risultati non cambiano sostanzialmente con $a > 3$, ma sono necessarie sequenze più lunghe per stabilizzare il risultato. Il valore $a = 3.2$ sembra ottimale. Con 5000 variabili standard ottenute con la routine `grand` abbiamo ottenuto $m = 0.023 \pm 0.014$, $s = 0.994 \pm 0.010$, che è un risultato più accurato di quello ottenuto con Metropolis con $a = 3$.

8.8. Con riferimento alla figura, si deve scrivere un codice che estrae il punto di emissione con le formule $x = -2 + 4\xi$, $y = -3 + 6\xi$ e la direzione di volo come $\cos \theta = 1 - 2\xi$, $\phi = 2\pi \xi$. Si ha poi $a = h \operatorname{tg} \theta$, $r = \sqrt{x^2 + y^2}$, $R = \sqrt{r^2 + a^2 - 2ra \cos \phi}$. Se R è minore del raggio della finestra R_d , la particella è contata dal rivelatore.. Se n è il numero di particelle contate su N generate, l'efficienza è data da $\epsilon = n/N \pm \sqrt{n(1-n/N)}/N$. Con i dati del problema e la nostra routine `Detector` abbiamo ottenuto $\epsilon = 0.059 \pm 0.002$.

8.9. Una possibile soluzione è la nostra routine `Vonmises`, che campiona uniformemente in $-\pi \leq x \leq \pi$. I valori di c sono introdotti da terminale.

8.10. La soluzione dipende dalla valutazione soggettiva del tracciato. Il lettore dovrebbe essere in grado di verificare che, per n troppo elevato, l'algoritmo non è affidabile nel predire il numero atteso di atomi con spin 1.

Capitolo 9.

9.1. La probabilità di estrarre una biglia nera vale $1/3$ o $2/3$. Dalla densità binomiale si ottiene la tabella:

	$x = 0$	$x = 1$	$x = 2$	$x = 3$	$x = 4$
$b(x; 4, p = 1/3)$	16/81	32/81	24/81	8/81	1/81
$b(x; 4, p = 2/3)$	1/81	8/81	24/81	32/81	16/81

La stima ML di p vale pertanto: $\hat{p} = 1/3$ se $0 \leq x \leq 1$, $\hat{p} = 2/3$ se $3 \leq x \leq 4$. Se $x = 2$ la funzione di verosimiglianza (densità binomiale) non ha massimo, pertanto lo stimatore ML non è definito in questo caso.

9.2.

	$x = 0$	$x = 1$	$x = 2$	$x = 3$
\hat{p}	0.1	0.3	0.7	0.9

9.3. \hat{x}_α è tale che $\int_{-\infty}^{\hat{x}_\alpha} p(x; \hat{\mu}, \hat{\sigma}) dx = \alpha$, dal teorema 9.2.

9.4. $d\mathcal{L}/d\lambda = n/\hat{\lambda} - \sum_i t_i = 0$, da cui $1/\hat{\lambda} = \sum_i t_i/n = m$.

9.5. a) Dalla (1.32) di pag. 25 risulta $P(x; N) = A(x)[(N-n)!]/[N!(N-2n+x)!]$, dove $A(x)$ contiene fattori indipendenti da N . Dato che $P(x; N) \equiv L(N)$, lo studio della funzione per valori discreti di N mostra che $L(N+1) \geq L(N)$ finché $N \leq n^2/x-1$. Il massimo di $L(N)$ si ha quindi per $\hat{N} = 1 + \text{int}(n^2/x-1)$, dove int è la parte intera dell'argomento. Dai dati si ottiene $\hat{N} = 609$. b) Utilizzando l'approssimazione di Stirling e la (3.21) di pag. 77, si ottiene $d\mathcal{L}/dN = -d \ln L(N)/dN = \ln[N(N-2n+x)/(N-n)^2]$, da cui $\hat{N} = n^2/x = 608$. Dato che N è una variabile discreta, il limite di Cramér-Rao non è utilizzabile. Anche con l'approssimazione di Stirling, supponendo N variabile continua, la (9.29) è di difficile applicazione. Si può simulare il processo, generando da $N = 600$ una serie di valori x e studiando l'istogramma degli \hat{N} stimati. Ripetendo per 5000 volte l'esperimento, si ottiene un istogramma asimmetrico (con una lunga coda a destra) di parametri $m = 618 \pm 1$, $s = 80.0 \pm 0.8$. Il valore massimo è ad $\hat{N} = 600$. Nell'intervallo 608 ± 80 è contenuto il 72% dei valori. Una stima ragionevole è pertanto: $N \in 608 \pm 80$, $CL = 72\%$. Dato che x ha densità binomiale, in approssimazione gaussiana ($x > 10$, $n-x > 10$) si può anche usare la (5.57) di pag. 167, che dà il risultato: $\text{Var}[\hat{N}] = \text{Var}[n^2/x] \simeq (n^4/x^3)[(1-x/n) + (2/x)(1-x/n)^2]$ (si noti che il secondo termine è trascurabile). Si ottiene $N \in 608 \pm 88$, in buon accordo con la simulazione.

9.6. La funzione di verosimiglianza di due prove vale $L = p^{x_1+x_2} (1-p)^{2-(x_1+x_2)} = L(p; S)$. S è sufficiente, P non lo è. Infatti, p può essere stimato solo dalla somma dei successi, non dal prodotto.

9.7. Tralasciando i fattori costanti, $\ln L = -(n/2)[\ln \sigma^2 + s^2/\sigma^2]$, dove $\hat{\sigma}^2 \equiv s^2 = w/n$ (la media è nota). Si noti che $\langle s^2 \rangle = \sigma^2$ e che w è una statistica sufficiente. Calcolando la derivata prima e seconda di $\ln L$ si trova che l'informazione attesa vale $nI(\theta) = n/(2\sigma^4)$. Applicando i metodi a) e b) suggeriti, da $|\ln L|/\sqrt{nI} \leq t_\alpha$ e $|[s^2 - \sigma^2]/\sqrt{(nI)^{-1}}| \leq t_\alpha$, dove t_α è il quantile, si trova lo stesso risultato, identico alla 6.63 di pagina 208.

9.8. Se $w = \sum_i x_i^2$ ed $n = 6$, il logaritmo della verosimiglianza che si ottiene dalla gaussiana con media nulla vale $\ln L(\sigma^2; w) = -(n/2) \ln \sigma^2 - w/(2\sigma^2)$. Il massimo si ha per $\hat{\sigma}^2 = w/n = 39.0/6 = 6.5$, che rappresenta la stima ML di σ^2 . Per applicare la (9.45) occorre ridefinire L , in modo che $\Delta[\ln L(\hat{\sigma}^2; x)] = 0$. Si ha pertanto: $\Delta[\ln L(\sigma^2; w)] = +(n/2)[\ln \sigma^2 + w/(n\sigma^2) - (n/2)[\ln(w/n) + 1]]$. Il valore $CL = 95.4\%$ si ha quando $2\Delta[\ln L(\sigma^2; x)] = 4$. Lo studio numerico della funzione fornisce i valori: $\sigma^2 \in [2.5, 27.1] = 6.5_{-4}^{+20.6}$.

Dato che la media è supposta nota, nella (6.72) i gradi di libertà di χ^2_R sono 6. Interpolando da tab (D.3) si ottengono i valori: $\chi^2_{R0.023}(6) = 2.46$ e $\chi^2_{R0.977}(6) = 0.20$. L'intervallo di confidenza con $CL = 95.4\%$ vale: $\sigma^2 \in [2.6, 32.6]$. Questa stima è migliore della precedente, che vale solo asintoticamente. I dati dell'esercizio sono stati estratti da una gaussiana con $\sigma^2 = 9$.

9.9. I due metodi forniscono numericamente lo stesso risultato: $\sigma^2 \in [8.0, 14.0]$. Anche questi dati sono stati estratti da una gaussiana con $\sigma^2 = 9$.

9.10. La stima per intervallo delle due medie vale: $\mu_1 \in 2.08 \pm 0.016$ cm, $\mu_2 \in 2.05 \pm 0.011$ cm. Le medie sono compatibili, perché il test della differenza (6.102) dà un valore standard $t = 1.58$. La stima ML è la media pesata: $\hat{\mu} = 2.0596$. La stima per intervallo è data dalla (9.69): $\mu \in 2.0596 \pm 0.0091$.

9.11. Se la variabile aleatoria X rappresenta il numero di campioni negativi su un numero di N campioni esaminati fissato a priori, la funzione di verosimiglianza è binomiale: $L(\mu; x) = N!/[x!(N-x)!] [\exp(-\mu)]^x [1 - \exp(-\mu)]^{N-x}$, dove $\exp(-\mu)$ è la probabilità poissoniana di non osservare eventi quando la media è μ . La stima ML di μ fornisce $\exp(-\hat{\mu}) = x/N$, da cui $\hat{\mu} = \ln(N/x) = \ln(50/5) = 2.3$. Si confronti il risultato con la Tab. 6.2 di pag. 202 e con le (6.39).

9.12. Occorre raggruppare gli ultimi 3 canali per avere $n(t) \geq 5$: 10 tempi tra 14 e 20 secondi. I valori di probabilità attesi entro ogni canale $\Delta t = (t_1, t_2)$ sono dati da $p_i(\theta) \equiv p_i(\lambda) = \int_{\Delta t} e(t) dt = \exp(-\lambda t_1) - \exp(-\lambda t_2)$. Utilizzando la (9.57) ed un programma di fit non lineare si ottiene: $\lambda \in 0.337 \pm 0.011$, $\chi^2_R = 9.55/7 = 1.36$. Il χ^2 va diviso per 7 gradi di libertà: 8 canali meno un parametro stimato (λ). Il livello SL osservato è $\simeq 24\%$, interpolando dalla Tab. D.3. Utilizzando la (9.59) si ottiene invece: $\lambda \in 0.339 \pm 0.011$, $\chi^2_R = 10.32/7 = 1.47$, $SL \simeq 19\%$. I dati sono stati campionati dalla densità esponenziale con $\lambda = 1/3$.

9.13. Livello del test: $\alpha = P\{X = 1; H_0\} = \varepsilon$. Potenza del test: $1 - \beta = 1 - P\{X = 0; H_1\} = P\{X = 1, H_1\} = 1 - \varepsilon$.

9.14. Livello del test: $\alpha = P\{X_1 = 1, X_2 = 1; H_0\} + (1/2)P\{X_1 + X_2 = 1; H_0\} = \varepsilon$. Potenza del test: $1 - \beta = 1 - P\{X_1 = 0, X_2 = 0; H_1\} - (1/2)P\{X_1 + X_2 = 1; H_1\} = 1 - \varepsilon$. Il risultato è identico al test con una sola prova. Tuttavia, conviene effettuare due prove se è possibile, perché nei casi estremi il livello di significatività osservato e la corrispondente potenza assumono valori più favorevoli. Ad esempio: $\alpha = P\{X_1 = 1, X_2 = 1; H_0\} = \varepsilon^2$, $1 - \beta = 1 - P\{X_1 = 0, X_2 = 0; H_1\} = 1 - \varepsilon^2$.

9.15. Dato che $p = (0.050 - 0.029)/0.057 - 0.029 = 0.750$, se $S = 16$ si estrae un numero $0 \leq \text{random} \leq 1$ e si accetta l'ipotesi se $0.75 \leq \text{random} \leq 1$. In pratica si accetta un lotto su 4.

9.16. I livelli del test richiesti sono $\alpha = 0.01$ e $\beta = 0.05$. a) Se $\lambda_0 = 1/100$, $\lambda_1 = 1/110$ e $t_n = (\sum_1^n t_i)/n$ è la media dei tempi osservati, dalle equazioni $(t_n - 100)/(100/\sqrt{n}) = 2.326$ e $(t_n - 110)/(110/\sqrt{n}) = -1.645$ si ricavano i valori $n = 1710$ e $t_n = 105.6$. Occorre quindi sacrificare un campione di ben 1710 sospensioni, calcolare la media dei tempi di guasto e accettare la dichiarazione del fornitore se la media supera 105.6 ore.

b) Dal rapporto $L_0/L_1 = [\lambda_0 \exp(-\lambda_0 t_n)]/[\lambda_1 \exp(-\lambda_1 t_n)]$ ed utilizzando i logaritmi si può scrivere l'intervallo di indecisione (9.98) come $-4.55/n - \ln(\lambda_0/\lambda_1) < (\lambda_1 - \lambda_0)t_n < 2.98/n - \ln(\lambda_0/\lambda_1)$. Si sceglie H_1 (H_0) quando la condizione non è soddisfatta a sinistra (destra). La simulazione mostra che, se le sospensioni sono effettivamente migliori, basta in media un campione di $n \in 934 \pm 6$ sospensioni e una media osservata di $t_n \in 112.7 \pm 0.1$ ore per confermare l'affermazione del costruttore. Se le nuove sospensioni sono come le vecchie, per la decisione giusta basta in media un campione di $n = 670$ sospensioni e una media di $t_n \in 97.1 \pm 0.1$ ore. Attenzione: questi sono valori medi e può capitare, con un solo test, di superare il valore $n = 1710$ ottenuto in precedenza con il campione di dimensione fissa e l'approssimazione normale. La regione di indecisione è compresa, nel piano (n, t_n) , tra due iperboli. Vi consigliamo di esaminare con attenzione gli istogrammi simulati! Sarebbe conveniente iniziare col test sequenziale e utilizzare il metodo a) quando il numero dei pezzi supera 1710. Tuttavia, questa procedura è scorretta, perché decidere il tipo di test da usare *dopo* avere visto (in tutto o in parte) i dati altera i livelli di probabilità stabiliti a priori. Il tipo di test a) o b) va quindi deciso prima della prova.

9.17. La variabile $Q_{\lambda_1} = 2\lambda_1 n T_n$, dove T_n è la media di n tempi esponenziali, ha una distribuzione $\chi^2(2n)$. Si ha pertanto: $1 - \beta(\lambda_1) = 1 - P\{Q \leq q_{\lambda_1}\}$, dove Q ha distribuzione $\chi^2(2n)$. Dato che n è molto grande, si può usare l'approssimazione normale e ricondursi al test sulla media: $1 - \beta(\lambda_1) = 1 - \Phi[(q_{\lambda_1} - 2n)/(2\sqrt{n})] = 1 - \Phi[(t_n - 1/\lambda_1)/(1/(\lambda_1\sqrt{n}))]$.

9.18. Si guardi la Fig. 9.6, supponendo che alle ipotesi $H_0 : p = 0.5$ e $H_1 : p = 0.3$ corrispondano densità gaussiane. Dato che dalla Tab. D.1 risulta $P\{t \leq -(0.5 - \alpha) = -0.4\} = -1.28$, $P\{t \geq (0.5 - \beta) = 0.4\} = +1.28$ ed il test è a una coda, valgono i limiti: $(x - 0.5n)/\sqrt{n} \cdot 0.5 - 0.5 = -1.28$ e $(x - 0.3n)/\sqrt{n} \cdot 0.3 - 0.7 = 1.28$. Le equazioni forniscono i valori, arrotondati agli interi, $x = 15$ e $n = 38$. Occorre un campione di 38 elementi, se $x \leq 15$ si accetta H_1 , altrimenti si accetta H_0 .

9.19. Il rapporto (9.87) vale: $R = L_0/L_1 = (0.5^x \cdot 0.5^{n-x})/(0.3^x \cdot 0.7^{n-x})$. Dalle (9.92, 9.94), dato che $(1 - \alpha)/\beta = 9$ e $\alpha/(1 - \beta) = 1/9$, passando ai logaritmi si ottiene il risultato seguente: si accetta $p = 0.5$ se $x \geq 0.397n + 2.593$, si sceglie $p = 0.3$ se $x \leq 0.397n - 2.593$. La fascia di piano compresa tra queste due rette è la regione di incertezza.

Con un codice di simulazione che ha ripetuto per 10000 volte il test, con una moneta non truccata ($p = 0.5$), si è ottenuto un istogramma dei valori di n di tipo

esponenziale, con i seguenti parametri: $m \in 23.3 \pm 0.2$, $\sigma \in 17.3 \pm 0.1$. In 786 casi su 10 000 è stata scelta l'ipotesi sbagliata. Il numero medio di tentativi (23) è minore di quello che risulta dal problema 9.18 (38). Tuttavia, in 1484 casi il test sequenziale simulato ha richiesto un numero di tentativi > 38 .

9.20. Si trova facilmente $\langle X \rangle = (\theta+1)/(\theta+2)$, $\text{Var}[X] = (\theta+1)/[(\theta+2)^2(\theta+3)]$, e la funzione di ripartizione $F_X(x) = x^{\theta+1}$. Il rapporto di Neyman-Pearson individua la miglior regione critica come $\{-\ln r_\alpha + n \ln[(1+\theta_0)/(1+\theta_1)]\} / (\theta_1 - \theta_0) \leq \sum_i \ln x_i \leq 0$. Se $Z = \ln X$, $F_z(z) = P\{Z \leq z\} = P\{X \leq e^z\} = e^{(1+\theta)z}$. La densità di Z è allora $dF_z/dz = p_Z(z) = (1+\theta) \exp[(1+\theta)z]$ e quindi $\langle \sum_i \ln X_i \rangle = -n/(1+\theta)$, $\text{Var}[\sum_i \ln X_i] = n/(1+\theta)^2$. Dato che $n = 100$, possiamo usare l'approssimazione normale; per $\alpha = 0.05$, $t_{1-\alpha/2} = 1.64$ e, se $\theta = \theta_0 = 1$, il rigetto dell'ipotesi nulla si ha per $[\sum_i \ln x_i - (-n/2)]/(\sqrt{n}/2) \geq 1.64$, cioè $\sum_i x_i \geq -41.8$. Questo test su $\sum_i \ln x_i$ è di potenza massima per $\alpha = 0.05$. Segue pure che $r_\alpha = 3.49$, ma questo valore non è necessario per il test, in approssimazione normale. La curva di potenza è $1 - \beta = 1 - \Phi[(-41.8 + n/(1+\theta_1))/(\sqrt{n}/(1+\theta_1))]$, $\theta_1 > \theta_0$. Per $n = 100$ e $\theta_1 = 2$ si ha $1 - \beta = 0.996$. È utile verificare che il test sulla somma $\sum x_i$ fornisce, in approssimazione normale, una potenza solo di poco inferiore, $1 - \beta = 0.989$.

Capitolo 10.

10.1. Dalla (5.65) e dall'ipotesi $\text{Cov}[X, Z] = 0$ otteniamo: $\sigma_y^2 = b^2 \text{Var}[X] + \text{Var}[Z] + 2b \text{Cov}[X, Z] = b^2 \text{Var}[X] + \text{Var}[Z]$. Definendo $\Delta X = X - \mu_x$ e $\Delta Y = Y - \mu_y$, dalla (5.82) possiamo calcolare anche la covarianza tra le variabili come nell'esercizio 4.3: $\text{Cov}[X, Y] = \langle \Delta X \Delta Y \rangle = b \langle \Delta^2 X \rangle = b \sigma_x^2$. Otteniamo pertanto il risultato: $\rho = \pm \sigma[f(X)]/\sigma[Y] = b\sigma_x/\sigma_y = b\sigma_x^2/\sigma_x \sigma_y = \sigma_{xy}/(\sigma_x \sigma_y)$.

10.2. Dalla relazione $s_x = 0.10x/\sqrt{12}$ (e analogamente per y) e utilizzando la formula della varianza effettiva per il calcolo di s_E , si ottiene la tabella:

x	10	20	30	40	50	60	70
s_x	0.3	0.6	0.9	1.2	1.4	1.7	2.0
y	21.4	38.8	52.2	88.1	99.5	120.4	158.3
s_y	0.6	1.1	1.5	2.6	2.9	3.5	4.6

La minimizzazione della (10.16), dove $f(x; a, b) = a + bx$ e la varianza effettiva al denominatore vale $s_y^2 + b^2 s_x^2$, fornisce i valori $a \in -0.13 \pm 1.15$, $b \in 2.04 \pm 0.05$. Con un *fit* lineare iterato due volte si ottiene invece: $a \in 0.23 \pm 1.12$, $b \in 2.02 \pm 0.05$. I dati sono stati generati con una simulazione partendo da $Y = 1 + 2X$. Si ottiene anche $\chi^2(\nu)/\nu = 28.0/5 = 5.6$. Il test χ^2 non è però significativo, perchè i dati non sono gaussiani.

10.3. Un *fit* non lineare con la retta $f(x; a, b) = a + bx$ e varianza effettiva $s_y^2 + b^2 s_x^2$ fornisce i valori: $a \in -0.64 \pm 1.14$, $b = 2.10 \pm 0.05$. Con un *fit* lineare iterato in due passi si ottiene: $a \in -0.62 \pm 1.12$, $b \in 2.10 \pm 0.05$. I dati sono stati generati con una simulazione partendo da $Y = 1 + 2X$. Il test χ^2 è significativo, perchè i dati sono gaussiani. Si ottiene: $\chi^2(\nu)/\nu = 2.05/5 = 0.41$, corrispondente nella Tab. D.3 a $SL \simeq 20\%$ nella coda a sinistra, che qualifica la retta come un buon modello di relazione funzionale.

10.4. Le rette passanti per (x_0, y_0) devono soddisfare il vincolo $y_0 = a_i + b_i x_0$. Se a_i e b_i sono noti, si considera il piano duale, dove una retta è rappresentata da un punto di coordinate (b, a) , e si può fare il *fit* dei punti (b_i, a_i) con una retta $a = (-x_0)b + y_0$, determinando il vertice (x_0, y_0) con errore. Questo metodo è usato in fisica per determinare il vertice di tracce di particelle nucleari nei rivelatori. Se le tracce sono curve, si può procedere per passi, utilizzando tratti di rette tangenti alle tracce.

10.5. In questo caso, $\text{Var}[Y_i] = \sigma_i^2 + \epsilon^2 x_i^2$ e $\text{Cov}[Y_i, Y_j] = \epsilon^2 x_i x_j$ per $i \neq j$. Otteniamo pertanto:

$$V = \begin{pmatrix} \sigma_1^2 + \epsilon^2 x_1^2 & \epsilon^2 x_1 x_2 & \dots & \epsilon^2 x_1 x_n \\ \epsilon^2 x_1 x_2 & \sigma_2^2 + \epsilon^2 x_2^2 & \dots & \epsilon^2 x_2 x_n \\ \dots & \dots & \dots & \dots \\ \epsilon^2 x_1 x_n & \epsilon^2 x_2 x_n & \dots & \sigma_n^2 + \epsilon^2 x_n^2 \end{pmatrix}.$$

10.6. I dati sono stati ottenuti da una simulazione ponendo $Y = 5 + 0.8X^2 + Y_R$, dove $\sigma[Y_R] = 0.5$.

10.7. X e Y hanno entrambe $\sigma = 0.5$ e tra loro è stata simulata una correlazione $f(X) = 5X + 0.2X^2$.

10.8. L'errore $\sigma[Y]$ si stima come $s(y_i) = 0.10 \cdot y_i$. Si ottengono in questo modo i valori:

x	2	4	6	8	10
y	7.9	11.9	17.0	25.5	23.8
$s(y)$	0.8	1.2	1.7	2.5	2.4

Questi dati sono passati al codice di minimizzazione lineare che determina i coefficienti \hat{a} e \hat{b} della relazione funzionale $y = \hat{a} + \hat{b}x$. Il programma fornisce il risultato $\hat{a} \pm s(\hat{a}) = 3.31 \pm 1.08$, $\hat{b} \pm s(\hat{b}) = 2.26 \pm 0.24$, $r(\hat{a}, \hat{b}) = -0.845$, $\chi^2 = 3.66$. Le stime LS sono compatibili con i valori veri $a = 5$ e $b = 2$, che sono quelli utilizzati per generare i dati artificiali con una simulazione. Il χ^2 ridotto $\chi_R^2(3) = 1.22$, corrisponde, in Tab. D.3, ad un livello di significatività ad una coda di circa il 30%. La covarianza vale $s(\hat{a}\hat{b}) = r(\hat{a}, \hat{b}) s(\hat{a}) s(\hat{b}) = -0.845 \cdot 1.08 \cdot 0.24 = -0.22$. Verifichiamo il risultato del *fit* con una simulazione, attraverso il metodo descritto nel par. 7.10. Utilizzando le stime LS \hat{a} e \hat{b} e i dati sperimentali x_i possiamo generare dati artificiali y'_i ed errori $s(y'_i)$ secondo l'algoritmo: $y'_i = (1 + 0.10 \cdot \xi_i)(\hat{a} + \hat{b}x_i)$, $s(y'_i) = 0.10 \cdot y'_i$, dove ξ è estratto da una gaussiana standard. Notate che la retta è calcolata coi valori stimati e che l'errore è sempre calcolato in modo approssimato in percentuale sui dati osservati, come nel caso di un esperimento reale. Ripetendo 20 000 volte il *fit* si ottengono gli istogrammi di \hat{a} , \hat{b} , $s(\hat{a})$, $s(\hat{b})$ e di χ^2 . *Le larghezze degli istogrammi*, permettono di verificare direttamente che gli errori $s(\hat{a})$ e $s(\hat{b})$ ottenuti dal *fit* coincidono entro 3 cifre decimali con le deviazioni standard degli istogrammi di \hat{a} e \hat{b} ottenuti con la simulazione. Le densità delle stime di a e b sono gaussiane praticamente perfette, come era lecito attendersi. Anche i 20 000 χ^2 si distribuiscono perfettamente come la densità di χ^2 con tre gradi di libertà.

10.9. Per prima cosa è necessario valutare gli errori $s(y_i) = (\Delta/\sqrt{12})y_i = 2(0.10/3.46)y_i = 0.058y_i$ e in modo analogo $s(x_i) = 0.058x_i$. Anche in questo caso la stima è approssimata, perché è ottenuta come percentuale dei valori osservati, non di quelli veri. Fornendo i valori x_i, y_i e la varianza effettiva $s_{Ei}^2 = b^2s^2(x_i) + s^2(y_i)$ ad un codice di *fit non* lineare otteniamo: $\hat{a} \pm s(\hat{a}) = 5.07 \pm 0.78$, $\hat{b} \pm s(\hat{b}) = 2.17 \pm 0.19$, $r(\hat{a}, \hat{b}) = -0.841$, $\chi^2 = 3.26$. Il χ^2 ridotto $\chi^2/3 = 1.09$ mostra un buon accordo tra dati e modello. Le stime sono compatibili con i valori veri $a = 5$ e $b = 2$, che sono quelli utilizzati per generare i dati simulati. Come nell'esercizio precedente, ripetiamo ora il *fit* per 20000 volte. Utilizzando i dati sperimentali x_i e le stime LS dei parametri, possiamo generare dati artificiali x'_i, y'_i ed errori $s(x'_i)$ e $s(y'_i)$ secondo l'algoritmo: $y'_i = [1 + 0.10(1 - 2\xi_i)](\hat{a} + \hat{b}x_i)$, $x'_i = x_i + 0.10(1 - 2\xi_i)x_i$, $s(x'_i) = 0.058x'_i$, $s(y'_i) = 0.058y'_i$. Ad ogni iterazione, si calcola la varianza s_E^2 e si esegue il *fit*. Dalla simulazione si vede che *le deviazioni standard degli istogrammi di a e b coincidono con i valori stimati col metodo LS*, fatto che accresce la nostra fiducia negli algoritmi dei minimi quadrati, anche con dati non gaussiani. La simulazione ci mostra un fatto importante: *gli scostamenti dal modello gaussiano sono piccoli*. In definitiva, si commette un errore trascurabile applicando gli intervalli della legge 3σ ai risultati della regressione.

Capitolo 11.

11.1. $\sigma[F]/F \simeq \sqrt{0.04^2 + 0.05^2} = 0.064$, pari al 6.4%.

11.2. Passando ai logaritmi si ha: $t = -\tau \ln(I/I_0) = 23$ giorni. Propagando l'errore su I si ottiene: $s_t = \pm \tau s_I/I = \pm 0.5$ giorni.

11.3. Dato che il fenomeno è poissoniano, il valore del conteggio è una stima della varianza. Dalla legge di propagazione dell'errore il valore del fondo vale $F = 620/10 = 62$ conteggi/s e $\sigma[F] = \sqrt{620}/10 = 2.5$ conteggi/s. Allora $(I - F) \pm \sigma[I - F] \simeq (157 - 62) \pm \sqrt{157 + 620/100} = 95 \pm 13$, conteggi/s, $CL \simeq 68\%$. Il rapporto segnale fondo vale "7 sigma": $n_\sigma = 95/13 \simeq 7.3$.

11.4. Il limite superiore con $CL = 95\%$ è dato dalla equazione $25 + 1.65\sqrt{\mu} = \mu$, da cui $\mu = 34.7 \simeq 35$. Il limite per il segnale vale perciò $35 - 10 = 25$ conteggi/s. Nota: è importante sottrarre il fondo *dopo* i calcoli.

11.5. Con 4 valori si devono utilizzare i quantili di Student con 3 gradi di libertà; dalla Tab. D.2 (per una probabilità di 0.975) si deduce che l'errore statistico della media è $0.04/3.18 = 0.012$. La precisione vale $0.012\sqrt{4} \simeq 0.02$.

11.6. L'errore percentuale di R si propaga su quello di V in misura doppia dell'errore percentuale di L . Occorre pertanto ottimizzare la misura di R .

11.7. $\text{Var}[\text{sen } \theta] \simeq (\cos^2 \theta) s_\theta^2$, da cui $\text{sen } \theta = 0.50 \pm 0.03$. Gli angoli vanno trasformati in radianti!

11.8. È immediato vedere che $(1 - P^2)/N = (4N_+N_-)/N^3$. Se N è aleatorio e poissoniano, N_{\pm} sono valori di variabili poissoniane indipendenti, $s^2(N_{\pm}) = N_{\pm}$ e quindi, dato che $(\partial P/\partial N_{\pm}) = \pm 2N_{\mp}/N^2$, propagando l'errore si ha $s^2(P) = (4N_+N_-)/N^3$. Se N è fissato prima della misura, N_{\pm} sono valori di variabili binomiali correlate. Dato che $N_- = N - N_+$, si ha $P = 2(N_+/N) - 1$, $\partial P/\partial N_+ = 2/N$, $s^2(N_+) = N_+(1 - N_+/N)$ e quindi $s^2(P) = 4(N_+/N^2)(1 - N_+/N) = 4N_+N_-/N^3$. In conclusione, l'incertezza $s(P) = \sqrt{(1 - P^2)/N}$ vale per qualunque condizione sperimentale.

11.9. $f(V) = 2.718 \pm 0.008$, dove l'errore è la deviazione standard. La distribuzione di $f(V)$ è praticamente uniforme, come è facile verificare sia analiticamente sia con una simulazione. È lecito quindi scrivere: $f(V) = 2.718 \pm 0.008 \sqrt{12}/2 = 2.718 \pm 0.013$, $CL = 100\%$.

11.10. Dato che $s^2(E_2) = (0.05)^2(\partial E_2/\partial E_1)^2 + (100/\sqrt{12})^2(\partial E_2/\partial R_1)^2 + (100/\sqrt{12})^2(\partial E_2/\partial R_2)^2 = (0.103)^2$, si ha $E_2 = 5.00 \pm 0.10$ V. La simulazione genera a caso valori di E_1, R_1, R_2 dalla distribuzione uniforme (ad es. $R_1 = 1000 + (2 \text{ randm} - 1) 50$) e fornisce l'istogramma di E_2 . Appare una forma vicina a quella triangolare, con deviazione standard $s = 0.103$, coincidente con quella calcolata. Nell'intervallo $m \pm s$ è contenuto il 65% dei valori. Il $CL = 68\%$ si raggiunge per $s = 0.110$. La misura diretta di E_2 è in accordo col calcolo.

11.11. $v = 2.00/5.35 = 0.3738$ m/s. $s_t = 0.05/\sqrt{20} = 0.011$ s; $s_l = 0.002/\sqrt{12} = 5.8 \cdot 10^{-4}$ m. Dato che $s_v^2 = v^2[s_t^2/l^2 + s_l^2/t^2] = 6.2 \cdot 10^{-7}$ si ottiene $v = 0.3738 \pm 0.0008$ m/s. L'errore è dominato dall'incertezza sui tempi. Dato che v è espressa come rapporto tra i valori di una variabile uniforme ed una gaussiana, conviene verificare il CL con una simulazione. Appare un istogramma di forma gaussiana, di deviazione standard coincidente con quella calcolata. Al risultato si può quindi associare un $CL \simeq 68\%$.

11.12. La frequenza si ottiene dalla formula $f = c/(ab)$ dove $c = 1750$, $a = 0.3$ m e $b = 0.5$ m. I corrispondenti errori valgono: $\sigma_c = \sqrt{1750} = 42$, $\sigma_a = \sigma_b = 2 \cdot 0.005/\sqrt{12} = 0.00289$ m. Dalla propagazione degli errori

$$\sigma_f^2 = [1/(ab)]^2 \sigma_c^2 + [c/(a^2b)]^2 \sigma_a^2 + [c/(ab^2)]^2 \sigma_b^2 = (306)^2$$

da cui $f = 1750 \pm 306$ conteggi/(m²s).

11.13. $\sigma_r \simeq \sqrt{20^2 - 5^2} \simeq 19$ canali, per le proprietà della convoluzione.

Appendice D

Tabelle

D.1 Integrale della densità gaussiana

La densità di probabilità gaussiana $g(x; \mu, \sigma)$ di media μ e deviazione standard σ segue l'espressione (vedere par. 3.4):

$$g(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right). \quad (\text{D.1})$$

Quando x viene espressa in modo standard:

$$t = \frac{x - \mu}{\sigma} \quad (\text{D.2})$$

l'equazione precedente viene sostituita dalla densità gaussiana standard, discussa nel par. 3.6:

$$g(t; 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right). \quad (\text{D.3})$$

I valori della probabilità integrale:

$$E(t) = \int_0^t g(t; 0, 1) dt \quad (\text{D.4})$$

di ottenere, in un campionamento casuale, un valore compreso nell'intervallo $[0, t]$ sono riportati nella Tab. D.1 per t compreso nell'intervallo $[0.0, 3.99]$ con incrementi di 0.01. Le prime due cifre di t si leggono nelle righe in grassetto a sinistra, la terza cifra nella colonna in grassetto in alto. I valori dell'integrale (D.4) si leggono all'incrocio delle righe con le colonne.

Sfruttando la simmetria di $E(t)$ attorno al valore $t = 0$ (vedere eq. 3.40) e servendosi di questa tavola, è possibile trovare l'area compresa in un intervallo qualsiasi.

L'integrale della densità con sette cifre decimali può anche essere ottenuto, per qualunque valore di t , con l'istruzione SCILAB

$$0.5 * \text{erf}(t/\text{sqrt}(2))$$

dove a t è stato precedentemente assegnato il valore richiesto.

I valori del quantile $t_{1-\alpha/2}$ si ottengono aggiungendo 0.5 ai valori della tabella e ricavando il corrispondente valore di t . Ad esempio, ad $1 - \alpha = CL = 0.99\%$ corrisponde $1 - \alpha/2 = 0.995$ e quindi $t \simeq 2.57$, in corrispondenza del valore integrale $0.4949 \simeq 0.495$.

D.2 Valori quantili della densità di Student

La densità di probabilità della variabile T di Student con ν gradi di libertà viene data dall'espressione ricavata nell'esercizio 5.5:

$$s_\nu(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{1}{2}) \Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\nu}} \left(\frac{t^2}{\nu} + 1 \right)^{-\frac{\nu+1}{2}}. \quad (\text{D.5})$$

I valori quantili t_P della variabile T corrispondenti ai differenti valori della probabilità integrale:

$$P = P\{T \leq t_P\} = \int_{-\infty}^{t_P} s_\nu(t) dt \quad (\text{D.6})$$

di non superare un dato valore di t_P sono riportati nella Tab. D.2 per differenti valori di ν . L'ultima riga della tabella, dove $\nu = \infty$, fornisce i valori quantili esatti della densità gaussiana.

I valori quantili corrispondenti ad una probabilità p per df gradi di libertà possono anche essere ottenuti, con sette cifre decimali, con l'istruzione SCILAB

$$\text{cdf}t('T', df, p, 1-p)$$

D.3 Integrale della densità χ^2 ridotto

La densità di probabilità della variabile $Q(\nu)$ con ν gradi di libertà segue l'espressione (3.67):

$$p_\nu(\chi^2) = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} e^{-\chi^2/2} (\chi^2)^{\nu/2-1}. \quad (\text{D.7})$$

I valori della variabile $Q_R(\nu) = Q(\nu)/\nu$ corrispondenti ai differenti valori della probabilità integrale:

$$P = P\{Q_R(\nu) > \chi_{R\nu}^2\} = \int_{\chi_{\nu}^2}^{\infty} p_\nu(\chi_\nu^2) d\chi_\nu^2 \quad (\text{D.8})$$

di superare un valore dato di χ^2 ridotto sono riportati nella Tab. D.3 per valori di ν compresi tra 0 e 100. Quando $\nu > 100$, $p_\nu(\chi^2)$ tende a diventare una gaussiana di parametri $\mu = 1$, $\sigma = \sqrt{2/\nu}$.

I valori quantili corrispondenti ad una probabilità p per df gradi di libertà possono anche essere ottenuti, con sette cifre decimali, con l'istruzione SCILAB

$$\text{cdfchi}('X', df, 1-p, p) / df$$

D.4 Valori quantili di χ^2 non ridotto

La Tab. D.4, che riporta i valori quantili di χ^2 *non ridotto* tali che:

$$P = P\{0 \leq Q(\nu) \leq \chi^2\} = \int_0^{\chi^2} p_\nu(\chi^2) d\chi^2,$$

permette di effettuare il test χ^2 in luogo della Tab. D.3. L'uso della Tab. D.3 di χ^2 ridotto corrispondente a certi livelli di significatività è equivalente all'uso della Tab. D.4 dei valori quantili di χ^2 non ridotto. Dipende dai "gusti" del lettore o dal tipo di problema la scelta dell'una o dell'altra. Quando $\nu > 100$, la densità $\chi^2(\nu)$ tende ad assumere una forma gaussiana di parametri $\mu = \nu$ e $\sigma = \sqrt{2\nu}$.

I valori quantili della tavola (D.4) sono più indicati ad essere usati per calcolare il livello di confidenza corrispondente al valore di $\Delta\chi^2$ dell'equazione:

$$\chi^2 = \chi_{min}^2 + \Delta\chi^2, \quad (D.9)$$

dove χ_{min}^2 è il χ^2 calcolato con i valori veri (o stimati con il best fit) dei parametri della popolazione modello. In questo caso i gradi di libertà ν rappresentano il numero di parametri presenti nel χ^2 .

Ad esempio, se il χ^2 ha dieci parametri liberi, l'ellissoide di concentrazione che ha per contorno $\Delta\chi^2 \simeq 16$ contiene il 90% dei punti, ovvero comprende il 90% dell'iperspazio, corrispondente a $CL = 0.90$.

I valori quantili corrispondenti ad una probabilità p per df gradi di libertà possono anche essere ottenuti, con sette cifre decimali, con l'istruzione SCILAB

$$\text{cdfchi}('X', df, p, 1-p)$$

D.5 Valori quantili della densità F

La densità di probabilità della variabile F di Snedecor viene data dall'espressione ottenuta nell'esercizio 5.6:

$$p_{\mu\nu}(F) = \frac{\Gamma\left(\frac{\mu+\nu}{2}\right)}{\Gamma\left(\frac{\mu}{2}\right) \Gamma\left(\frac{\nu}{2}\right)} \mu^{\frac{\mu}{2}} \nu^{\frac{\nu}{2}} \frac{F^{(\nu-2)/2}}{(\mu + F\nu)^{(\nu+\mu)/2}}. \quad (D.10)$$

La probabilità di ottenere un valore $\{F \leq F_\alpha\}$ in un campione casuale con ν_1 e ν_2 gradi di libertà viene data dall'integrale:

$$P\{F \leq F_\alpha\} = \int_0^{F_\alpha} p_{\nu_1, \nu_2}(F) dF . \quad (\text{D.11})$$

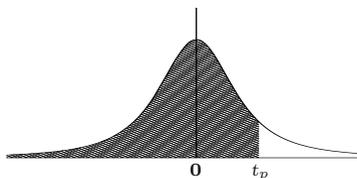
I valori di F_α corrispondenti ai livelli di significatività del 5% ($\alpha = 0.05$) e dell'1% ($\alpha = 0.99$) a destra della media sono rispettivamente riportati nelle Tab. D.5 e D.6 per differenti valori di ν_1 e ν_2 . I corrispondenti valori $F_{1-\alpha}$ per la coda a sinistra della media possono essere ricavati dalla relazione (5.49):

$$F_\alpha(\nu_1, \nu_2) = \frac{1}{F_{1-\alpha}(\nu_2, \nu_1)} . \quad (\text{D.12})$$

I valori quantili corrispondenti ad una probabilità p per df1 e df2 gradi di libertà possono anche essere ottenuti, con sette cifre decimali, con l'istruzione SCILAB

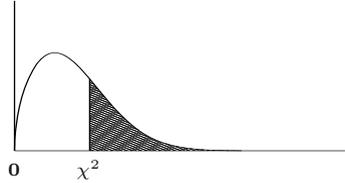
```
cdf('F', df1, df2, p, 1-p)
```


Tabella D.2. Valori quantili t_P della variabile di Student per ν gradi di libertà



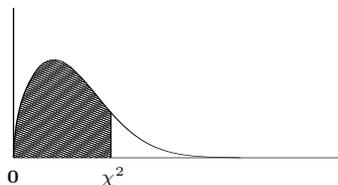
$\nu \backslash P$	0.60	0.70	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995	0.9995
1	0.325	0.727	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.65	632.0
2	0.289	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.60
3	0.277	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.92
4	0.271	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.267	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.265	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.263	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.262	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.261	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.260	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.260	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.259	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.259	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.258	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.258	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.258	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.257	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.257	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.257	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.257	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.849
21	0.257	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.256	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.256	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.256	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.256	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.256	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.256	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.256	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.256	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.256	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.255	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
50	0.255	0.528	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.496
60	0.254	0.527	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.460
70	0.254	0.527	0.678	0.847	1.044	1.294	1.667	1.994	2.381	2.648	3.435
80	0.254	0.526	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.416
90	0.254	0.526	0.677	0.846	1.042	1.291	1.662	1.987	2.368	2.632	3.402
100	0.254	0.526	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.390
∞	0.253	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Tabella D.3. Valori della variabile χ^2/ν aventi probabilità P di essere superati in un campionamento



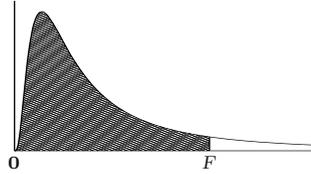
$\nu \backslash P$	0.005	0.01	0.025	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.975	0.99	0.995
1	7.88	6.63	5.02	3.84	2.71	1.32	0.45	0.10	0.02	0.004	0.001	0.000	0.000
2	5.30	4.61	3.69	3.00	2.30	1.39	0.69	0.29	0.11	0.05	0.03	0.01	0.005
3	4.28	3.78	3.12	2.61	2.08	1.37	0.79	0.40	0.19	0.12	0.07	0.04	0.02
4	3.71	3.32	2.79	2.37	1.95	1.35	0.84	0.48	0.27	0.18	0.12	0.07	0.05
5	3.35	3.02	2.57	2.21	1.85	1.33	0.87	0.53	0.32	0.23	0.17	0.11	0.08
6	3.09	2.80	2.41	2.10	1.77	1.31	0.89	0.58	0.37	0.27	0.21	0.15	0.11
7	2.90	2.64	2.29	2.01	1.72	1.29	0.91	0.61	0.40	0.31	0.24	0.18	0.14
8	2.74	2.51	2.19	1.94	1.67	1.28	0.92	0.63	0.44	0.34	0.27	0.21	0.17
9	2.62	2.41	2.11	1.88	1.63	1.27	0.93	0.66	0.46	0.37	0.30	0.23	0.19
10	2.52	2.32	2.05	1.83	1.60	1.25	0.93	0.67	0.49	0.39	0.32	0.26	0.22
11	2.43	2.25	1.99	1.79	1.57	1.25	0.94	0.69	0.51	0.42	0.35	0.28	0.24
12	2.36	2.18	1.94	1.75	1.55	1.24	0.95	0.70	0.53	0.44	0.37	0.30	0.26
13	2.29	2.13	1.90	1.72	1.52	1.23	0.95	0.72	0.54	0.45	0.39	0.32	0.27
14	2.24	2.08	1.87	1.69	1.50	1.22	0.95	0.73	0.56	0.47	0.40	0.33	0.29
15	2.19	2.04	1.83	1.67	1.49	1.22	0.96	0.74	0.57	0.48	0.42	0.35	0.31
16	2.14	2.00	1.80	1.64	1.47	1.21	0.96	0.74	0.58	0.50	0.43	0.36	0.32
17	2.10	1.97	1.78	1.62	1.46	1.21	0.96	0.75	0.59	0.51	0.44	0.38	0.34
18	2.06	1.93	1.75	1.60	1.44	1.20	0.96	0.76	0.60	0.52	0.46	0.39	0.35
19	2.03	1.90	1.73	1.59	1.43	1.20	0.97	0.77	0.61	0.53	0.47	0.40	0.36
20	2.00	1.88	1.71	1.57	1.42	1.19	0.97	0.77	0.62	0.54	0.48	0.41	0.37
21	1.97	1.85	1.69	1.56	1.41	1.19	0.97	0.78	0.63	0.55	0.49	0.42	0.38
22	1.95	1.83	1.67	1.54	1.40	1.18	0.97	0.78	0.64	0.56	0.50	0.43	0.39
23	1.92	1.81	1.66	1.53	1.39	1.18	0.97	0.79	0.65	0.57	0.51	0.44	0.40
24	1.90	1.79	1.64	1.52	1.38	1.18	0.97	0.79	0.65	0.58	0.52	0.45	0.41
25	1.88	1.77	1.63	1.51	1.38	1.17	0.97	0.80	0.66	0.58	0.52	0.46	0.42
26	1.86	1.76	1.61	1.50	1.37	1.17	0.97	0.80	0.67	0.59	0.53	0.47	0.43
27	1.84	1.74	1.60	1.49	1.36	1.17	0.98	0.81	0.67	0.60	0.54	0.48	0.44
28	1.82	1.72	1.59	1.48	1.35	1.17	0.98	0.81	0.68	0.60	0.55	0.48	0.45
29	1.80	1.71	1.58	1.47	1.35	1.16	0.98	0.81	0.68	0.61	0.55	0.49	0.45
30	1.79	1.70	1.57	1.46	1.34	1.16	0.98	0.82	0.69	0.62	0.56	0.50	0.46
40	1.67	1.59	1.48	1.39	1.30	1.14	0.98	0.84	0.73	0.66	0.61	0.55	0.52
50	1.59	1.52	1.43	1.35	1.26	1.13	0.99	0.86	0.75	0.70	0.65	0.59	0.56
70	1.49	1.43	1.36	1.29	1.22	1.11	0.99	0.88	0.79	0.74	0.70	0.65	0.62
80	1.45	1.40	1.33	1.27	1.21	1.10	0.99	0.89	0.80	0.75	0.71	0.67	0.64
90	1.43	1.38	1.31	1.26	1.20	1.10	0.99	0.90	0.81	0.77	0.73	0.69	0.66
100	1.40	1.36	1.30	1.24	1.18	1.09	0.99	0.90	0.82	0.78	0.74	0.70	0.67

Tabella D.4. Valori quantili χ^2_P della variabile χ^2 non ridotto per ν gradi di libertà



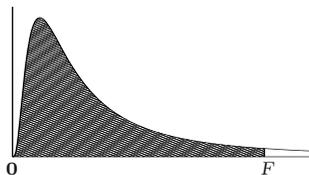
$\nu \backslash P$	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.975	0.99	0.995
1	0.06	0.15	0.27	0.45	0.71	1.07	1.64	2.71	3.84	5.02	6.63	7.88
2	0.45	0.71	1.02	1.39	1.83	2.41	3.22	4.61	5.99	7.38	9.21	10.60
3	1.00	1.42	1.87	2.37	2.95	3.67	4.64	6.25	7.82	9.35	11.34	12.84
4	1.65	2.19	2.75	3.36	4.04	4.88	5.99	7.78	9.49	11.14	13.28	14.86
5	2.34	3.00	3.66	4.35	5.13	6.06	7.29	9.24	11.07	12.83	15.09	16.75
6	3.07	3.83	4.57	5.35	6.21	7.23	8.56	10.64	12.59	14.45	16.81	18.55
7	3.82	4.67	5.49	6.35	7.28	8.38	9.80	12.02	14.07	16.01	18.47	20.28
8	4.59	5.53	6.42	7.34	8.35	9.52	11.03	13.36	15.51	17.53	20.09	21.95
9	5.38	6.39	7.36	8.34	9.41	10.66	12.24	14.68	16.92	19.02	21.67	23.59
10	6.18	7.27	8.30	9.34	10.47	11.78	13.44	15.99	18.31	20.48	23.21	25.19
11	6.99	8.15	9.24	10.34	11.53	12.90	14.63	17.27	19.68	21.92	24.72	26.76
12	7.81	9.03	10.18	11.34	12.58	14.01	15.81	18.55	21.03	23.34	26.22	28.30
13	8.63	9.93	11.13	12.34	13.64	15.12	16.98	19.81	22.36	24.74	27.69	29.82
14	9.47	10.82	12.08	13.34	14.69	16.22	18.15	21.06	23.68	26.12	29.14	31.32
15	10.31	11.72	13.03	14.34	15.73	17.32	19.31	22.31	25.00	27.49	30.58	32.80
16	11.15	12.62	13.98	15.34	16.78	18.42	20.47	23.54	26.30	28.85	32.00	34.27
17	12.00	13.53	14.94	16.34	17.82	19.51	21.61	24.77	27.59	30.19	33.41	35.72
18	12.86	14.44	15.89	17.34	18.87	20.60	22.76	25.99	28.87	31.53	34.81	37.16
19	13.72	15.35	16.85	18.34	19.91	21.69	23.90	27.20	30.14	32.85	36.19	38.58
20	14.58	16.27	17.81	19.34	20.95	22.77	25.04	28.41	31.41	34.17	37.57	40.00
21	15.44	17.18	18.77	20.34	21.99	23.86	26.17	29.62	32.67	35.48	38.93	41.40
22	16.31	18.10	19.73	21.34	23.03	24.94	27.30	30.81	33.92	36.78	40.29	42.80
23	17.19	19.02	20.69	22.34	24.07	26.02	28.43	32.01	35.17	38.08	41.64	44.18
24	18.06	19.94	21.65	23.34	25.11	27.10	29.55	33.20	36.42	39.36	42.98	45.56
25	18.94	20.87	22.62	24.34	26.14	28.17	30.68	34.38	37.65	40.65	44.31	46.93
26	19.82	21.79	23.58	25.34	27.18	29.25	31.79	35.56	38.89	41.92	45.64	48.29
27	20.70	22.72	24.54	26.34	28.21	30.32	32.91	36.74	40.11	43.19	46.96	49.64
28	21.59	23.65	25.51	27.34	29.25	31.39	34.03	37.92	41.34	44.46	48.28	50.99
29	22.48	24.58	26.48	28.34	30.28	32.46	35.14	39.09	42.56	45.72	49.59	52.34
30	23.36	25.51	27.44	29.34	31.32	33.53	36.25	40.26	43.77	46.98	50.89	53.67
40	32.34	34.87	37.13	39.34	41.62	44.16	47.27	51.81	55.76	59.34	63.69	66.77
50	41.45	44.31	46.86	49.33	51.89	54.72	58.16	63.17	67.50	71.42	76.15	79.49
60	50.64	53.81	56.62	59.33	62.13	65.23	68.97	74.40	79.08	83.30	88.38	91.95
70	59.90	63.35	66.40	69.33	72.36	75.69	79.71	85.53	90.53	95.02	100.4	104.2
80	69.21	72.92	76.19	79.33	82.57	86.12	90.41	96.58	101.88	106.6	112.3	116.3
90	78.56	82.51	85.99	89.33	92.76	96.52	101.1	107.6	113.2	118.1	124.1	128.3
100	87.95	92.13	95.81	99.33	103.0	106.9	111.7	118.5	124.3	129.6	135.8	140.2

Tabella D.5. Valori quantili al 95% della variabile $F(\nu_1, \nu_2)$ di Snedecor



$\nu_1 \backslash \nu_2$	1	2	3	4	6	8	10	15	20	30	40	60	120	∞
1	161	200	216	225	234	239	242	246	248	250	251	252	253	254
2	18.6	19.0	19.2	19.2	19.3	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	8.94	8.85	8.79	8.70	8.66	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.16	6.04	5.96	5.86	5.80	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	4.95	4.82	4.74	4.62	4.56	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.28	4.15	4.06	3.94	3.87	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.87	3.73	3.64	3.51	3.45	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.58	3.44	3.35	3.22	3.15	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.37	3.23	3.14	3.01	2.94	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.22	3.07	2.98	2.85	2.77	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.09	2.95	2.85	2.72	2.65	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.00	2.85	2.75	2.62	2.54	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	2.92	2.77	2.67	2.53	2.46	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.85	2.70	2.60	2.46	2.39	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.79	2.64	2.54	2.40	2.33	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.74	2.59	2.49	2.35	2.28	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.70	2.55	2.45	2.31	2.23	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.66	2.51	2.41	2.27	2.19	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.63	2.48	2.38	2.23	2.16	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.60	2.45	2.35	2.20	2.12	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.57	2.42	2.32	2.18	2.10	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.55	2.40	2.30	2.15	2.07	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.53	2.38	2.27	2.13	2.05	1.96	1.91	1.87	1.81	1.76
24	4.26	3.40	3.01	2.78	2.51	2.36	2.25	2.11	2.03	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.49	2.34	2.24	2.09	2.01	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.47	2.32	2.22	2.07	1.99	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.46	2.31	2.20	2.06	1.97	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.45	2.29	2.19	2.04	1.96	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.43	2.28	2.18	2.03	1.94	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.42	2.27	2.16	2.02	1.93	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.34	2.18	2.08	1.92	1.84	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.25	2.10	1.99	1.84	1.75	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.18	2.02	1.91	1.75	1.66	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.10	1.94	1.83	1.67	1.57	1.46	1.39	1.32	1.22	1.00

Tabella D.6. Valori quantili al 99% della variabile $F(\nu_1, \nu_2)$ di Snedecor



$\nu_1 \backslash \nu_2$	1	2	3	4	6	8	10	15	20	30	40	60	120	∞
1	4052	5000	5403	5625	5859	5981	6056	6157	6209	6261	6287	6313	6339	6366
2	98.5	99.0	99.3	99.2	99.3	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	27.9	27.5	27.2	26.9	26.7	26.5	26.4	26.3	26.2	26.1
4	21.2	18.0	16.7	16.0	15.2	14.8	14.6	14.2	14.0	13.8	13.8	13.7	13.6	13.5
5	16.3	13.3	12.1	11.4	10.7	10.3	10.1	9.72	9.55	9.38	9.29	9.20	9.11	9.02
6	13.7	10.9	9.78	9.15	8.47	8.10	7.87	7.56	7.40	7.23	7.14	7.06	6.97	6.88
7	12.2	9.55	8.45	7.85	7.19	6.84	6.62	6.31	6.16	5.99	5.91	5.82	5.74	5.65
8	11.3	8.65	7.59	7.01	6.37	6.03	5.81	5.52	5.36	5.20	5.12	5.03	4.95	4.86
9	10.6	8.02	6.99	6.42	5.80	5.47	5.26	4.96	4.81	4.65	4.57	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.39	5.06	4.85	4.56	4.41	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.07	4.75	4.54	4.25	4.10	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.96	5.41	4.82	4.50	4.30	4.01	3.86	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.62	4.30	4.10	3.82	3.66	3.51	3.43	3.34	3.25	3.17
14	8.86	6.52	5.56	5.04	4.46	4.14	3.94	3.66	3.51	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.41	4.89	4.32	4.00	3.81	3.52	3.37	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.20	3.89	3.69	3.41	3.26	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.19	4.67	4.10	3.79	3.59	3.31	3.16	3.00	2.92	2.84	2.75	2.65
18	8.29	6.01	5.09	4.58	4.02	3.71	3.51	3.23	3.08	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	3.94	3.63	3.43	3.15	3.00	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	3.87	3.56	3.37	3.09	2.94	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	3.81	3.51	3.31	3.03	2.88	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.76	3.45	3.26	2.98	2.83	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.71	3.41	3.21	2.93	2.78	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.67	3.36	3.17	2.89	2.74	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.63	3.32	3.13	2.85	2.70	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.59	3.29	3.09	2.82	2.66	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.56	3.26	3.06	2.78	2.63	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.53	3.23	3.03	2.75	2.60	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.05	3.50	3.20	3.00	2.73	2.57	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.47	3.17	2.98	2.70	2.55	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.29	2.99	2.80	2.52	2.37	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.12	2.82	2.63	2.35	2.20	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	2.96	2.66	2.47	2.19	2.03	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	2.80	2.51	2.32	2.04	1.88	1.70	1.59	1.47	1.32	1.00

Bibliografia

- [Azz01] A. Azzalini. *Inferenza Statistica*. Springer-Verlag Italia, Milano, 2001.
- [Bal98] P. Baldi. *Calcolo delle Probabilità e Statistica*. McGraw Hill Italia, Milano, 1998.
- [Bar89] R. Barlow. *Statistics*. J.Wiley and Sons, New York, 1989.
- [BCD01] L.D. Brown, T.T. Cai, and A. DasGupta. Interval estimation for a binomial proportion. *Statistical Science*, 16:101–133, 2001.
- [BD02] E. Bee Dagum. *Analisi delle serie storiche: modellistica, previsione e scomposizione*. Springer-Verlag Italia, Milano, 2002.
- [Ber85] C. Bernardini. L'onestà degli scienziati. *Sapere*, 1/2:45–46, 1985.
- [BFS87] P. Bratley, B.L. Fox, and L.E Schrage. *A Guide to Simulation*. Academic Press, New York, second edition, 1987.
- [Blo84] V. Blobel. Unfolding methods in high-energy physics experiments. Technical Report DESY 84-118, DESY, Amburgo, 1984.
- [Bol62] M. Boldrini. *Statistica (Teoria e Metodi)*. A. Giuffrè, 1962.
- [BR92] P.R. Bevington and D.K. Robinson. *Data reduction and error analysis for the physical sciences*. McGraw-Hill, New York, 1992.
- [Bro98] W.A. Brown. L'effetto placebo. *Le Scienze*, 335:84–89, 1998.
- [BS91] M. Berblinger and C. Schlier. Monte Carlo integration with quasi-random numbers: some experience. *Computer Physics Communications*, 66:157–166, 1991.
- [Buc84] S.T. Buckland. Monte Carlo confidence intervals. *Biometrics*, 40:811–817, 1984.
- [Bun86] B.D. Bunday. *Basic Queueing Theory*. Edward Arnold, London, 1986.
- [BV03] C. S. Bertuglia and F. Vaio. *Non linearità, caos, complessità*. Bollati Boringhieri, 2003.
- [Car85] J. Carlson. A double blind test for astrology. *Nature*, 318:419–425, 1985.
- [CB90] G. Casella and R.L. Berger. *Statistical Inference*. Wadsworth & Brook/Cole, Pacific Grove, 1990.

- [Cha75] G.J. Chaitin. Casualità e dimostrazione matematica. *Le Scienze*, 85:30–35, settembre 1975.
- [Coc77] W.G. Cochran. *Sampling techniques*. J.Wiley and Sons, New York, third edition, 1977.
- [Com91] A. Compagner. Definitions of randomness. *American Journal of Physics*, 59:700–705, 1991.
- [Cou95] R.D. Cousins. Why isn't every physicist a bayesian? *American Journal of Physics*, 63:398–410, 1995.
- [Cra51] H. Cramer. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, 1951.
- [CS61] D.R. Cox and W.C. Smith. *Queues*. Chapman and Hall, London, 1961.
- [D'A94] G. D'Agostini. On the use of the covariance matrix to fit the correlated data. *Nuclear Instruments and Methods in Physics Research A*, 346:306–311, 1994.
- [D'A99] G. D'Agostini. Bayesian reasoning in high-energy physics: Principles and applications. Technical Report CERN 99-03, CERN, Ginevra, 1999.
- [Dav08] A.C. Davison. *Statistical Models*. Cambridge University Press, New York, 2008.
- [DE83] P. Diaconis and B. Efron. Statistica al calcolatore. *Le Scienze*, 179:116–123, 1983.
- [DH99] A.C. Davison and D.V. Hinkley. *Bootstrap Methods and their Applications*. Cambridge University Press, Cambridge, 1999.
- [DMM+75] L. Daboni, P. Malesani, P. Manca, G. Ottaviani, F. Ricci, and G. Sommi. *Ricerca Operativa*. Zanichelli, Bologna, 1975.
- [DT95] F. Di Trocchio. *Le Bugie della Scienza*. Mondadori, Milano, 1995.
- [EDJ+71] W.T. Eadie, D. Drijard, F. James, M. Roos, and B. Sadoulet. *Statistical Methods in experimental Physics*. North Holland, Amsterdam, 1971.
- [Efr79] B. Efron. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7:1–26, 1979.
- [Efr82] B. Efron. *The Jackknife, the Bootstrap and Other Resampling plans*. SIAM, Filadelfia, 1982.
- [ET93] B. Efron and R.J. Tibshirani. *An introduction to the bootstrap*. Chapman and Hall, London, 1993.
- [Eva55] R.D. Evans. *The Atomic Nucleus*. Mc Graw-Hill, New York, 1955.
- [FC98] G.J. Feldman and R.D. Cousins. A unified approach to the classic statistical analysis of small signals. *Physical Review D*, 57:5873–5889, 1998.
- [Fel47] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley and Sons, New York, second edition, 1947.
- [Fey92] R.P. Feynman. *Sta scherzando, Mr Feynman!* Zanichelli, Bologna, 1992.

- [Fis41] R.A. Fisher. *Statistical Methods for Research Workers*. Eighth, London, 1941.
- [Fis96] G.S. Fishman. *Monte Carlo Concepts, Algorithms, and Applications*. Springer-Verlag, New York, 1996.
- [FLJW92] A.M. Ferrenberg, D.P. Landau, and Y. Joanna Wong. Monte Carlo simulation: hidden errors from “good” random numbers generators. *Physical Review Letters*, 69:3382–3384, 1992.
- [fSI93] International Organization for Standardization (ISO). Guide to the expression of uncertainty in measurement. Technical report, ISO, Ginevra, 1993.
- [Gal84] P. Galeotti. *Probabilità e Statistica*. Levrotto e Bella, Torino, 1984.
- [Gil97] G.T. Gillies. The newtonian gravitational constant: recent measurement and related studies. *Reports on Progress in Physics*, 60:151–225, 1997.
- [Gne76] B.V Gnedenko. *The Theory of Probability*. Mir, Mosca, 1976.
- [Gro02] Particle Data Group. *Review of Particle Physics*, volume 66. 2002.
- [GS92] G.R. Grimmet and D.R. Stirzaker. *Probability and Random Processes*. Clarendon Press, Oxford, 1992.
- [HH64] J.M. Hammersley and D.C. Handscomb. *Monte Carlo Methods*. Chapman and Hall, London, 1964.
- [HL07] J. Heinrich and L. Lyons. Systematic errors. *Annual Review of Nuclear and Particle Science*, 57:145–169, 2007.
- [Jam80] F. James. Monte Carlo theory and practice. *Reports on Progress in Physics*, 43:1145–1189, 1980.
- [Jam90] F. James. A review of pseudorandom number generators. *Computer Physics Communications*, 60:329–344, 1990.
- [Jam92] F. James. Minit reference manual. Technical Report CERN D506, CERN, Ginevra, 1992.
- [JLPe00] F. James, Lyons L., and Y. Perrin (editori). Atti del “Workshop on confidence limits”. Technical Report CERN 2000-005, CERN, Ginevra, 2000.
- [Kah56] H. Kahn. Use of different Monte Carlo sampling techniques. In H.A. Meyer, editor, *Symposium on Monte Carlo Methods*, pages 146–190. J.Wiley and Sons, New York, 1956.
- [KH96] K. Kacperski and A. Holyst. Phase transition and hysteresis in a cellular automata-based model of opinion formation. *Journal of Statistical Physics*, 84:168–189, 1996.
- [Knu81] D.E. Knuth. *The Art of Computer Programming*, volume 2: Seminumerical Algorithms, chapter 2. Addison-Wesley, Reading, second edition, 1981.
- [KR89] B.W. Kernighan and D.M. Ritchie. *Linguaggio C*. Jackson Libri, Milano, 1989.
- [KS73] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics*, volume II. Griffin, London, 1973.

- [KW86] M.H. Kalos and P. Whitlock. *Monte Carlo Methods*, volume One: Basics. J.Wiley and Sons, New York, 1986.
- [Lam66] J.R. Lamarsh. *Nuclear Reactor Theory*. Addison-Wesley, Reading (Mass.), 1966.
- [Lyb84] M. Lybanon. Comment on least squares when both variables have uncertainties. *American Journal of Physics*, 52:276–278, 1984.
- [MGB91] A.M. Mood, F.A. Graybill, and D.C. Boes. *Introduzione alla Statistica*. McGraw-Hill Italia, Milano, 1991.
- [MNZ90] G. Marsaglia, B. Narashimhan, and A. Zaman. A random number generator for pc's. *Computer Physics Communications*, 60:345–349, 1990.
- [Mor84] B.J.T. Morgan. *Elements of Simulations*. Chapman and Hall, London, 1984.
- [Mor92] G. Morpurgo. *Introduzione alla Fisica delle Particelle*. Zanichelli, Bologna, 1992.
- [Ney37] J. Neyman. Outline of a theory of statistical estimation based on the classical theory of probability. *Phil. Trans. Royal Society London*, 236:333–380, 1937.
- [Ore82] J. Orear. Least squares when both variables have uncertainties. *American Journal of Physics*, 50:912–916, 1982.
- [Pap77] A. Papoulis. *Probabilità, Variabili Aleatorie e Processi Stocastici*. Boringhieri, Torino, 1977.
- [PFTW92] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Wetterling. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, second edition, 1992.
- [Pic96] L. Piccinato. *Metodi per le decisioni statistiche*. Springer-Verlag Italia, Milano, 1996.
- [RC99] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Verlag, New York, 1999.
- [Rip86] B. Ripley. *Stochastic Simulation*. J.Wiley and Sons, New York, 1986.
- [Ros96] S.B. Ross. *Simulation*. Academic Press, Londra, second edition, 1996.
- [Rot10] A. Rotondi. On frequency and efficiency measurements in counting experiments. *Nuclear Instruments and Methods in Physics Research A*, 614:106–119, 2010.
- [RPP] A. Rotondi, P. Pedroni, and A. Pievatolo. Sito web del testo: <http://www.mi.imati.cnr.it/~marco/springer/>.
- [Rub81] R.Y. Rubinstein. *Simulation and Monte Carlo Method*. J.Wiley and Sons, New York, 1981.
- [Rue87] D. Ruelle. *Caso e caos*. Boringhieri, Torino, 1987.
- [RvN63] R.D. Richmeyer and J. von Neumann. Statistical methods in neutron diffusion. In A.H. Taub, editor, *John von Neumann collected works*, volume V, pages 751–767. Pergamon Press, Oxford, 1963.
- [SCI] Consorzio SCILAB. <http://www.scilab.org>.

- [Spi76] M.R. Spiegel. *Statistica*. ETAS Libri, Milano, 1976.
- [Ste93] I. Stewart. *Dio gioca a dadi?* Boringhieri, Torino, 1993.
- [SW89] G.A.F. Seber and C.J. Wild. *Nonlinear Regression*. Wiley Interscience, New York, 1989.
- [TC93] R. Toral and A. Chakrabarti. Generation of gaussian distributed random numbers by using a numerical inversion method. *Computer Physics Communications*, 74:327–334, 1993.
- [UA183] Collaborazione UA1. Experimental observations of lepton pairs of invariant mass around 95 GeV/c at the CERN SPS collider. *Physics Letters B*, 126:398–410, 1983.
- [UA283] Collaborazione UA2. Evidence of $Z^0 \rightarrow e^+e^-$ at the CERN collider. *Physics Letters B*, 129:130–140, 1983.
- [Voi03] J. Voit. *The Statistical Mechanics of Financial Markets*. Springer-Verlag, Berlin, second edition, 2003.

Indice analitico

- accuratezza, 450, 452
 - classi di, 450
 - tavola di, 450, 467
- aleatorio
 - sistema o processo, 2
- algebra- σ , 13, 14
- algoritmo
 - del rigetto elementare, 275, 279
 - del rigetto ottimizzato, 277, 280
 - del rigetto pesato, 278, 281
 - della trasf. inversa, 271
 - di Box e Muller, 283
 - di generazione, 267
 - di generazione gaussiana, 284
 - di generazione poissoniana, 287
 - di Metropolis, 328
 - di ricerca lineare, 282
- approccio
 - bayesiano, 12, 31
 - frequentista, 12, 184
- astrologia, 112

- batch means, metodo delle, 320, 330
- Bernoulli, J., 6, *vedi* distribuzione di
- best fit, 347, 402
 - con errori sistematici, 421
- bias, 206, 250, 355
- bin, 44
- Boltzmann, H.
 - costante di, 100
 - distribuzione di, 100
 - equazione di, 308

- bootstrap
 - *bias*, 302
 - non parametrico, 301
 - parametrico, 293, 461, 463
- Buffon, conte di, 261

- campionamento
 - con reimmissione, 218
 - definizione di, 9
 - per importanza, 339
 - stratificato, 339
 - tecniche di, 192
- campione
 - analisi del, 228
 - definizione di, 9, 63
- caos, 2
- caotico, fenomeno, 2
- carta di controllo, 217
- catene di Markov, 328
- centro di massa, 311
- coefficiente
 - R^2 corretto, 436
 - di determinazione, 435
 - di variazione, 169
- combinazioni, 24
- confidenza
 - fascia di, 182
 - intervallo di, 183
- convergenza
 - in legge, 66
 - in probabilità, 65
 - quasi certa, 65
- convoluzione, 156, 479

- copertura, 183, 296
- correlazione, 125, 403
 - coefficiente di, 126, 139, 173, 249, 256, 301, 405
 - definizione di, 126
 - funzione di, 404
 - lineare, 409
 - ricerca della, 434
- correlazione multipla
 - coefficiente di, 435
- correzione
 - di continuità, 193, 201
 - di Sheppard, 470
- coseni direttori, 311
- covarianza, 119, 122, 125, 168, 232
 - campionaria, 249, 251
 - tra parametri LS, 412
- coverage, 183
- Curtiss, L.F., 473
- curva di regressione (o di best fit), 402, 428

- dati epidemiologici, 29
- decile, 43
- densità, *vedi* distribuzione
 - condizionata, 119
 - congiunta, 118
 - definizione di, 47, 54
 - gaussiana bivariata, 143
 - gaussiana condizionata, 135
 - gaussiana standard, 82, 517
 - marginale, 119
 - stabile, 159
- determinazione, coefficiente di, 435
- devianza
 - totale, 435
- deviazione standard
 - definizione di, 55, 57
 - stima della, 208
- diagnosi automatica, 31
- diffusione
 - di neutroni, 308
 - di particelle, 307
 - elastica, 309
 - lunghezza di, 315
- disposizioni, 24
- distribuzione, 39
 - $N(\mu, \sigma^2)$, 84
 - χ^2 , 94, 208, 518
 - binomiale o di Bernoulli, 48, 71, 110, 229, 350
 - campionaria, 222
 - del coefficiente di correlazione, 251
 - di Boltzmann, 100
 - di Cauchy, 508
 - di Erlang, *vedi* esponenziale negativa
 - di Maxwell, 99
 - di Rayleigh, 144
 - di Snedecor, 164, 227, 430, 519
 - di Student, 161, 211, 518
 - di von Mises, 345
 - di Weibull, 92
 - esponenziale negativa, 89
 - gamma, 93, 153, 504
 - gaussiana bivariata, 129, 143, 283
 - gaussiana multivariata, 141, 147
 - gaussiana o di Gauss, 76
 - geometrica, 73, 89
 - logaritmica, 459, 504
 - logonormale, 501
 - multinomiale, 147, 176, 229
 - normale, *vedi* gaussiana
 - poissoniana o di Poisson, 74, 88, 389, 473
 - semi-gaussiana, 304
 - studio Monte Carlo di una, 288
 - triangolare, 158, 449, 457
 - uniforme, 100, 466
- disuguaglianza
 - di Cauchy-Schwarz, 125, 361
 - di Tchebychev, 104, 213

- effetto
 - di trascinamento, 489
 - farfalla, 4
 - placebo, 486
- efficienza
 - di generazione, 276
 - di simulazione, 337
- ellissi di concentrazione, 133, 142
- energia delle molecole, 100
- equazione
 - del bilancio dettagliato, 328
 - del trasporto, 308
 - di Boltzmann, 308
- equazioni
 - di Clopper-Pearson, 193

- errore
 – barra di, 235, 402, 476
 – casuale, *vedi* statistico
 – cifre significative dell', 188
 – di I tipo, 221, 237, 381, 391, 486
 – di II tipo, 221, 382, 388, 391
 – di misura, 468
 – di parallasse, 451
 – massimo, 466
 – percentuale, 458
 – propagazione dell', 292, 454
 – riaggiustamento dell', 440
 – simulazione dell', 461, 465
 – sistematico, 421, 450
 – statistico, 179, 452, 464
 – strumentale, *vedi* sistematico
 – sulla polarizzazione, 491
 esperimenti indipendenti, 22
 esperimento
 – definizione di, 9
 – delle monete, 44, 52, 62, 73, 80, 239, 264, 269, 375
 – di Millikan, 482
 estrapolazione, 434
 eteroschedasticità, 428
 eventi
 – incompatibili, 20
 – indipendenti, 19
 evento, definizione di, 8
- falsificazione, 486
 falso allarme, 217
 fascia di confidenza, 182
 fattore di distorsione, 206
 Feigenbaum, M., 3
 fenomeni
 – di attesa, 315
 – stocastici, 87
 Fermat, P., 6
 Fermi, E., 262
 Feynman, R., 484
 Fisher, R.A., 349
 – trasformazione di, 252
 fluttuazione statistica, 51, 106, 222, 230, 405
 formula
 – della media pesata, 378
 – di Clopper-Pearson, 295
 – di partizione dell'evento certo, 26
 – di Stirling, 77
 – di Wald, 190, 296
 – di Wilson, 189, 295
 frequenza, 45, 188
 – analisi della, 247
 – cumulativa, 45
 – di arrivo, 91
 – di decadimento, 474
 – di emissione, 88
 – di una malattia, 244
 – in un canale, 148
 – limite della, 13
 – misurata, 190, 204, 240
 – osservata, 351
 – sperimentale, 55
 – test della, 243
 – test di confronto, 224
 funzione
 – cumulativa, 42, 84, 91
 – degli errori, 84
 – dell'apparato, 478
 – di potenza, 389
 – di ripartizione, *vedi* funz. cumulativa
 – di variabile aleatoria, 41
 – di verosimiglianza, 349
 – generatrice (Mgf), 495
 – test, 386
- Galilei, G., 445
 Gauss, K., *vedi* distribuzione di
 gradi di libertà, 207, 246, 371
 grandezza
 – costante, 447
 – variabile, 447
- incertezza, 447, *vedi* errore
 – sistematica, 447
 – statistica, 179
 indice di rifrazione, 294
 indipendenza
 – di eventi, 19
 – di variabili, 41, 118, 119
 – di variabili gaussiane, 131
 – e correlazione, 126
 – stocastica, 41
 inferenza statistica, 9
 informazione, 359

integrale

- con metodo “vinci o perdi”, 335
- con metodo della media (o grezzo), 336
- di convoluzione, 156, 479
- di folding, 478
- multidimensionale, 343

interpolazione

- lineare, 282

intervallo

- di probabilità , 106
- di confidenza, 181, 291
- di sensibilità, 448

ipotesi

- alternativa, 381
- composte, 389
- nulla, 109, 220, 381
- semplice, 389

Ising, modello di, 331

isotropia, 273

istogramma, 44

- best fit di un, 372
- canale dell', 44, 46, 230
- delle frequenze, 45
- delle frequenze cumulative, 46
- normalizzato, 45

Kolmogorov

- probabilità di, 13, 15
- disuguaglianza di, 66, 213

Laplace, P.S., 6, 399

Legendre, A., 399

legge, 39, 47

- 3-sigma (3σ), 82, 105, 143, 190, 235, 464, 484
- dei grandi numeri, 13, 214
- di propagazione degli errori, 454
- empirica del caso, 13
- esponenziale negativa, 89
- ipergeometrica, 24
- logistica, 3

limite

- di Cramér-Rao, 366
- frequentista, 12, 64
- in probabilità, 65
- inferiore (stima del), 184
- poissoniano, 201
- quasi certo, 65

- superiore (stima del), 184

livello

- del test, 221, 382
- di probabilità , 82, 142
- di confidenza, 181
- di significatività, 110, 181, 222, 224, 237, 238, 382, 439
- lotto di controllo, 387

mappa logistica, 3

Markov

- catene di, 328

matrice

- dei pesi, 419
- del trasporto, 170, 172, 175
- delle covarianze, 139, 170, 171, 175, 421
- delle derivate seconde, 442
- di correlazione, 140, 175
- di curvatura, 443
- positiva definita, 140

media

- campionaria, 55, 64, 216
- campionaria (popolazione finita), 219
- come operatore, 60
- condizionata, 121
- definizione di, 55, 57
- di distribuzioni, 59
- di istogrammi, 56
- pesata, 378, 468
- stima della, 203, 209
- stimatore della, 64, 67

metodo

- “vinci o perdi”, 335
- dei minimi quadrati, 371, 400
- del gradiente, 441
- della griglia, 292
- della massima verosimiglianza, 350
- dello scarto massimo, 344
- di best fit, 347
- Monte Carlo, 261
- scientifico, 485
- Millikan, R., 482
- minimi quadrati
 - generali, 416
 - metodo dei, 400
 - pesati, 419

- misura
 – definizione di, 9
 – del decadimento radioattivo, 473
 – del parametro di Michel, 489
 – del tempo di caduta, 462
 – della carica, 482
 – della vel. della luce, 469
 – errore di, 452
 – indiretta, 454
 – operazioni di, 446
 – precisione della, 451
 – tipi di, 466
 momenti, 58, 59, 207
 Monte Carlo, metodo, 6, 103, 261
- neutroni termici, 309
 – assorbimento dei, 309
 – diffusione elastica dei, 309
 Neyman, J., 184
 normalizzazione, 45
- omeopatia, 485
 ortogonalizzazione, 418
- p-value, 222, 238, 439, 505
 parametri
 – campionari, 55
 – veri, 55
 partizione dell'evento certo, 26
 passeggiata a caso, 114
 Pearson, K., 97
 – teorema di, 97, 142
 percentile, 43, 164, 230
 percezione extrasensoriale, 225
 perimetro toracico, 256, 414
 pivot, quantità, 186
 Poincaré, H., 2
 Poisson, *vedi* distribuzione di
 – processo di, 88
 polarizzazione
 – misura della, 491
 popolazione
 – definizione di, 9
 – di un campione, 63
 – tipi di, *vedi* densità, distribuzioni
 postulato di oggettività, 485
 potenza
 – del test, 382
 – funzione di, 389
- pranoterapia, 485
 precisione, 452
 predittore, 400
 – non osservato, 405
 – osservato, 402
 probabilità
 – a priori, 12
 – assiomatica, 13
 – composta, 17
 – condizionata, 18
 – definizione di, 9
 – frequentista, 12
 – limite in, 65
 – soggettiva, 11
 – spazio di, 15
 – stima della, 188, 193
 – totali, 26
 problema
 – dell'incontro, 34, 304
 – di Monty, 33, 304
 processo
 – di urto, 311
 – stocastico, 310
 processo di Poisson, 88
 prova
 – definizione di, 9
 – ripetuta, 20
- quantile, 43, 164, 227, 383, 518, 519
 quantità pivotale, 186, 190, 210, 211
- regione critica, 221, 381–383
 regressione
 – curva di, 402
 – lineare, 135
 – multipla, 416
 residui, 410, 428
 – pesati, 428
 resistenze elettriche, 215
 retta
 – dei minimi quadrati stimata, 409
 – di regressione, 135, 405
 – di regressione, prevista, 412
 ricerca
 – binaria, *vedi* dicotomica
 – dicotomica, 268
 – operativa, 315
 – sequenziale, 268
 risoluzione, *vedi* sensibilità

- risultati elettorali, 191
- routine
 - Asimm, 508
 - Bayes, 32
 - Binocov, 296, 305
 - Bootcor, 301
 - Bootcor20, 303
 - Bootrif, 294
 - Buffon, 508
 - Chisq, 239, 248
 - Conguf, 157
 - Corcov, 123, 129
 - Covardat, 140, 251
 - Dadi, 269
 - Detector, 509
 - Ellisse, 509
 - Gauss1, 507
 - Gauss2, 285
 - Gaussfit, 374
 - Gridrif, 294
 - Hist3d, 137
 - Histfreqe, 46, 269
 - Histplote, 46
 - Linfit, 401, 416, 418, 420, 432, 437
 - Logiplot, 4
 - Logist, 4
 - Mcinteg, 337, 342, 344
 - Mcinteg1, 509
 - MCmoneta, 263
 - Metrop, 509
 - Monetafit, 375
 - Nlinfit, 374, 401, 406, 441, 443, 475
 - Optasinc, 326, 327
 - Optsinc, 321
 - per la simulazione sincrona, 321
 - Poiss, 287
 - Poisscov, 300, 305
 - Pol2fit, 444
 - Pol3fit, 437
 - Pol4fit, 432
 - Prodx, 460
 - Rettamq, 443
 - Rifraz, 289
 - Scarti, 508
 - Sequen, 394
 - Sigdel, 458
 - Simpoiss, 287
 - Stasys, 458
 - Stimvs, 214, 215
 - Stiprob, 194
 - Tabcont, 247
 - Test2v, 227, 228
 - Vonmises, 509
- routine SCILAB
 - binomial, 50
 - cdfbin, 82, 194, 296
 - cdfchi, 98, 519
 - cdff, 164, 227, 520
 - cdfpoi, 112, 201
 - cdft, 162, 518
 - clf, 47
 - convol, 157
 - correl, 256
 - covar, 123, 256
 - dsearch, 269
 - erf, 84, 517
 - grand, 262, 283, 296, 345
 - histplote, 46
 - mean, 60
 - meanf, 60
 - nand2med, 227
 - plot, 296
 - rand, 262
 - sum, 248
 - variance, 60
 - variancef, 60
- scarto, 56
- scarto quadratico medio, *vedi* deviazione standard
- scienza esatta, 446
- SCILAB, 37
- score function, 359
- sensibilità, 448
 - intervallo di, 448
- sezione d'urto macroscopica, 309
- Sheppard, correzione di, 470
- simulazione
 - asincrona, 316
 - discreta, 316
 - sincrona, 316
- somme aleatorie, 314
- sovra-copertura, 183
- spazio
 - campionario, 8
 - di probabilità, 15
- speranza matematica, *vedi* valore atteso

- spettro
 - continuo, 42
 - definizione di, 9, 42
 - discreto, 42
- statistica
 - definizione di, 64
 - sufficiente, 356
- statistiche congiuntamente sufficienti, 356
- statura, 256
- stima, 64
 - per intervalli, 366
- stimatore, 64, 67, 180
 - asintoticamente corretto, 206
 - BAN, 366
 - consistente, 354
 - corretto, 68, 355
 - della media, 67
 - della varianza, 67
 - distorto, 206
 - non ammissibile, 361
 - più efficiente, 355
- stocastico
 - sistema o processo, 2
- stocastico, caso o fenomeno, 1
- strumento
 - accuratezza di uno, 450
 - analogico, 449
 - digitale, 448
 - sensibilità di uno, 448
- tabella di contingenza, 243, 246
- Tchebychev, P.L., disuguaglianza di, 104
- tempo
 - di arrivo, 89
 - di caduta, 462
 - morto, 89, 474
- teorema
 - della z di Fisher, 251
 - della media, 229
 - della somma di Pearson, 147
 - delle probabilità composte, 18
 - di additività, 16
 - di additività del χ^2 , 97
 - di Bayes, 25, 27, 30
 - di Cauchy-Schwarz, 125
 - di Cramér-Rao, 361
 - di DeMoivre-Laplace, 78
 - di efficienza, 363
 - di fattorizzazione, 356
 - di Gauss-Markov, 424
 - di indipendenza stocastica, 94
 - di indipendenza tra media e varianza, 209
 - di Leibnitz, 151
 - di Neyman-Pearson, 383
 - di normalità asintotica, 366
 - di Pearson, 97, 236
 - Limite Centrale, 86, 159
 - sul cambiamento di variabile, 154
 - sul campione di Metropolis, 329
 - sull'indipendenza di variabili, 120, 131
 - sulla varianza campionaria, 210
 - sulle forme quadratiche, 142
 - sulle funzioni di parametri, 357
 - sulle statistiche sufficienti, 357
 - sulle stime corrette, 424
 - sulle stime dei minimi quadrati, 426
 - sulle variabili cumulative, 102
- test
 - χ^2 , 133, 148, 236, 237, 241, 243, 246
 - a doppio cieco, 485
 - a due code, 110, 221, 238, 439
 - a una coda, 110, 221, 238, 439
 - della differenza, 223, 475
 - delle cinghie, 248
 - delle resistenze, 215
 - di χ^2 , 373
 - di Neyman-Pearson, 383
 - diagnostici, 27
 - F di Snedecor, 227
 - non parametrici, 242
 - più potente, 386
 - potenza del, 382, 384
 - randomizzato, 222, 388
 - sequenziale, 393
 - sui vaccini, 244
 - uniformemente più potente, 389
- Ulam, S., 262
- valore
 - atteso, 26, 52, 61, 72, 190, 231, 265
 - campionario, 55
 - sperimentale, 179
 - vero, 55, 179

- variabile
 - $N(\mu, \sigma^2)$, 84
 - Z di Fisher, 252
 - ξ , 263
 - χ^2 ridotto, 211
 - aleatoria, 5, 38
 - casuale, *vedi* v. aleatoria
 - continua, 42, 52
 - cumulativa, 102, 267
 - di Student, 161
 - differenza, 223
 - discreta, 42
 - F di Snedecor, 163, 430
 - indicatrice, 40
 - poissoniana, 231
 - prodotto, 169, 459
 - standard, 83, 223, 484
 - uniforme, 102, 271
- variabili
 - indipendenti, 41, 118, 119
 - stocasticamente indipendenti, 41
 - vincolate, 141
- varianza
 - analisi della, 435
 - campionaria, 55, 64
 - definizione di, 55, 57
 - del prodotto, 169
 - della media, 204
 - della media (popolazione finita), 220
 - di distribuzioni, 59
 - di istogrammi, 56
 - effettiva, 407
 - nel campionamento per importanza, 339
 - nel campionamento stratificato, 340
 - percentuale o relativa, 169, 458
 - stima della, 205, 209
 - stimatore della, 63, 67
- verifica
 - di ipotesi, 9
 - di più ipotesi, 381
 - di una ipotesi, 220
- verosimiglianza
 - funzione di, 349
 - logaritmo della, 349
 - massima, 349
 - rapporto di, 383
- vertice, determinazione del, 444
- von Mises
 - distribuzione di, 345
 - probabilità di, 12
- Von Neumann, J., 262, 277

Unitext – Collana di Statistica e Probabilità Applicata

A cura di:

F. Battaglia (Editor-in-Chief)

C. Carota

P.L. Conti

L. Piccinato

M. Riani

Editor in Springer:

F. Bonadei

francesca.bonadei@springer.com

C. Rossi, G. Serio

La metodologia statistica nelle applicazioni biomediche
1990, XVIII, 353 pp, ISBN 978-3-540-52797-8

A. Azzalini

Inferenza statistica

Una presentazione basata sul concetto di verosimiglianza
2a ed., 2001, XIV, 367 pp, ISBN 978-88-470-0130-5

E. Bee Dagum

Analisi delle serie storiche: modellistica, previsione e scomposizione
2002, XI, 302 pp, ISBN 978-88-470-0146-6

B. Luderer, V. Nollau, K. Velters

Formule matematiche per le scienze economiche
2003, X, 212 pp, ISBN 978-88-470-0224-1

A. Azzalini, B. Scarpa

Analisi dei dati e data mining

Corr. 2a ed., 2004, X, 232 pp. ISBN 978-88-470-0272-2

F. Battaglia

Metodi di previsione statistica

2007, X, 323 pp, ISBN 978-88-470-0602-7

L. Piccinato
Metodi per le decisioni statistiche
2a ed., 2009, XIV, 474, ISBN 978-88-470-1077-2

E. Stanghellini
Introduzione ai metodi statistici per il credit scoring
2009, IX, 179 pp, ISBN 978-88-470-1080-2

A. Rotondi, P. Pedroni, A. Pievatolo
Probabilità, Statistica e Simulazione
3a ed., 2011, XIV, 540 pp, ISBN 978-88-470-2363-5

La versione online dei libri pubblicati nella serie è disponibile
su SpringerLink. Per ulteriori informazioni, visitare il sito:
<http://www.springer.com/series/1380>