

# Introduzione all'analisi dei dati

*Per il laboratorio di Fisica*

LUCA BALDINI ([luca.baldini@pi.infn.it](mailto:luca.baldini@pi.infn.it))

NICCOLÒ DI LALLA ([niccolo.dilalla@pi.infn.it](mailto:niccolo.dilalla@pi.infn.it))

MASSIMILIANO RAZZANO ([massimiliano.razzano@unipi.it](mailto:massimiliano.razzano@unipi.it))

CARMELO SGRÒ ([carmelo.sgro@pi.infn.it](mailto:carmelo.sgro@pi.infn.it))

Versione 0.19.1 compilata il 26 maggio 2017

web: <https://bitbucket.org/lbaldini/statnotes>

Git sha1: 7adddda67c1f75970005a38fba29ofdof7db9fa





This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



---

## INDICE

---

1	PRIMI PASSI: MISURE ED INCERTEZZE	1
1.1	Il concetto di errore massimo	1
1.1.1	Errore assoluto ed errore relativo	2
1.1.2	Breve digressione: precisione ed accuratezza	2
1.2	Cifre significative e convenzioni di scrittura	2
1.3	Unità di misura e ordini di grandezza	3
1.3.1	Il Sistema Internazionale di unità di misura	3
1.3.2	Unità derivate ed unità riconosciute	5
1.3.3	Multipli, sottomultipli ed ordini di grandezza	5
1.3.4	Cenni di calcolo dimensionale	6
1.4	Stima dell'errore massimo in condizioni di ripetitività	8
1.4.1	Misure in regime di dispersione nulla: la risoluzione strumentale	8
1.4.2	Un semplice esperimento: l'incertezza di misura con il metro a nastro	8
1.4.3	Fluttuazioni casuali: stima a posteriori dell'errore massimo	9
1.4.4	Si può ridurre l'incertezza di misura al di sotto della risoluzione strumentale?	10
1.5	Propagazione dell'errore massimo	11
1.5.1	Propagazione dell'errore massimo nelle somma e nella differenza	12
1.5.2	Propagazione dell'errore massimo nel prodotto e nel quoziente	12
1.5.3	Propagazione dell'errore massimo nel caso generale	13
1.5.4	Scelta delle grandezze per le misure dirette ed indipendenza	15
1.6	Errori sistematici	17
1.6.1	Errori sistematici strumentali	18
1.6.2	Incerezze di modellizzazione	18
1.6.3	Identificazione degli effetti sistematici	20
1.6.4	Propagazione degli effetti sistematici	20
1.7	In breve...	20
2	RAPPRESENTAZIONE DEI DATI	23
2.1	Tabelle	23
2.2	Grafici di dispersione	23
2.2.1	Fit grafici lineari	24
2.3	Grafici a barre ed istogrammi	26
2.4	Esponenziali e logaritmi	27
2.4.1	Un tipico processo esponenziale: la divisione cellulare	28
2.4.2	La ricerca binaria	29
2.4.3	Ancora sui logaritmi: il concetto di decade	29
2.5	Una breve digressione: il suono	30
2.5.1	L'altezza del suono: l'ottava	30
2.5.2	Il volume sonoro	31
2.6	Scale non lineari	32
2.6.1	Di nuovo sulle decadi: scale logaritmiche	33
2.6.2	Leggi di potenza e grafici bilogaritmici	34
2.6.3	Fit grafici in scala bilogaritmica	35
2.6.4	Funzioni esponenziali e grafici semilogaritmici	36
2.6.5	Fit grafici in scala semilogaritmica	38
2.7	In breve...	38
3	ELEMENTI DI TEORIA DELLE PROBABILITÀ	41
3.1	Definizione assiomatica della probabilità	41
3.1.1	Leggi elementari della probabilità	42
3.2	Definizioni operative di probabilità	44

3.2.1	Definizione combinatoriale	44
3.2.2	Definizione frequentista	45
3.2.3	Definizione soggettiva	46
3.2.4	Uno spunto di riflessione: esistono monete e dadi equi?	46
3.3	Digressione: elementi di calcolo combinatorio	47
3.3.1	Permutazioni e funzione fattoriale	47
3.3.2	Coefficienti binomiali	48
3.3.3	Una applicazione elementare: il gioco del lotto	49
3.3.4	Coefficienti multinomiali	49
3.4	Probabilità condizionata	50
3.4.1	Eventi indipendenti: un esempio elementare	51
3.5	Il teorema di Bayes	52
3.5.1	Un semplice problema di probabilità inversa	52
3.5.2	Ancora sul teorema di Bayes: un po' di nomenclatura	53
3.6	Variabili casuali e funzioni di distribuzione	54
3.6.1	Variabili casuali discrete	54
3.6.2	Variabili casuali continue	56
3.7	Valore di aspettazione	56
3.8	Tendenza centrale e dispersione intorno alla media	57
3.8.1	Media, mediana e moda	58
3.8.2	Varianza e deviazione standard	59
3.8.3	La larghezza a metà altezza	60
3.8.4	La disuguaglianza di Chebyshev ed il significato della deviazione standard	61
3.9	Momenti di una distribuzione	62
3.10	La funzione cumulativa	63
3.11	Breve riepilogo: un semplice problema di calcolo delle probabilità	63
3.11.1	Calcolo della funzione di distribuzione	64
3.11.2	Media, mediana e moda	64
3.11.3	Varianza, deviazione standard e larghezza a metà altezza	65
3.11.4	Coefficiente di asimmetria	65
3.11.5	Distribuzione cumulativa e distribuzione inversa	66
3.12	Digressione: alcuni paradossi probabilistici	66
3.12.1	Quanti compleanni lo stesso giorno?	67
3.12.2	Breve digressione: somma della serie geometrica	68
3.12.3	Maschio o femmina?	68
3.12.4	Una strategia di gioco interessante	70
3.13	Cenni alle variabili multivariate: indipendenza, covarianza e correlazione	70
3.13.1	L'indipendenza tra variabili casuali	71
3.13.2	Covarianza e correlazione	72
3.14	In breve...	74
4	VARIABILI CAMPIONE E PROPAGAZIONE DELL'ERRORE STATISTICO	75
4.1	Campionamenti singoli	76
4.2	Somma di variabili casuali indipendenti	76
4.2.1	Media e varianza della somma di variabili casuali indipendenti	77
4.2.2	Una applicazione interessante: il random walk unidimensionale	78
4.3	Misure ripetute: media e varianza campione	80
4.3.1	Digressione: l'imparzialità è davvero irrinunciabile?	81
4.3.2	La deviazione standard della media	82
4.3.3	Un esempio concreto: la statura dei coscritti	83
4.4	Calcolo della media e varianza campione in pratica	84
4.4.1	Un algoritmo alternativo con un ciclo singolo	85
4.4.2	L'algoritmo di Welford	86
4.4.3	E se io fossi veramente pigro?	86
4.5	Covarianza e correlazione campione	88

4.5.1	Correlazioni e rapporti di causa effetto	89
4.6	Media e varianza di una funzione di variabili casuali	89
4.6.1	Il caso unidimensionale	90
4.6.2	Il caso generale	90
4.7	La propagazione dell'errore statistico	92
4.7.1	Un esempio di propagazione degli errori per variabili correlate	93
4.8	In breve...	94
5	DISTRIBUZIONI UNIVARIATE DI USO COMUNE	97
5.1	La distribuzione binomiale	97
5.1.1	Normalizzazione, media e varianza	98
5.1.2	Momenti di ordine superiore	100
5.2	Digressione: distribuzione binomiale e misure di efficienza	100
5.2.1	La soluzione classica	101
5.2.2	Una soluzione basata sul teorema di Bayes	101
5.3	La distribuzione multinomiale	102
5.4	La distribuzione di Poisson	103
5.4.1	Processi Poissoniani	103
5.4.2	La distribuzione di Poisson come limite della binomiale	105
5.4.3	Due esempi di interesse storico ed alcune considerazioni	106
5.4.4	Normalizzazione, media e varianza	108
5.4.5	Momenti di ordine superiore	109
5.4.6	Distribuzione di Poisson e distanza tra eventi successivi	110
5.4.7	Somma di variabili Poissoniane	110
5.5	La distribuzione uniforme	112
5.5.1	Normalizzazione, media, varianza e coefficiente di asimmetria	112
5.5.2	Funzione cumulativa e quantili	113
5.6	La distribuzione esponenziale	113
5.6.1	Normalizzazione, media, varianza e coefficiente di asimmetria	114
5.6.2	Funzione cumulativa e quantili	115
5.6.3	Assenza di memoria	115
5.7	La distribuzione di Gauss	116
5.7.1	La distribuzione di Gauss come limite della Poissoniana	117
5.7.2	Normalizzazione, media e varianza	118
5.7.3	L'integrale normale degli errori	120
5.7.4	Alcuni commenti sul passaggio al continuo	123
5.7.5	Una derivazione alternativa della distribuzione di Gauss	123
5.8	Il teorema centrale del limite	124
5.8.1	Il teorema centrale del limite e la media campione	127
5.9	In breve...	127
6	INTRODUZIONE AI METODI MONTE CARLO	129
6.1	Introduzione: due semplici esempi	129
6.2	Che cosa è un "numero casuale"?	132
6.2.1	Generatori casuali e pseudo-casuali	133
6.3	Digressione: l'aritmetica modulare	133
6.3.1	Il logaritmo discreto	134
6.4	Sequenze pseudo-casuali	134
6.4.1	Aspetti generali dei generatori pseudo-casuali	135
6.4.2	Lo schema lineare congruenziale	136
6.5	Breve digressione: bitmap casuali	138
6.6	Generazione di numeri pseudo-casuali con distribuzione non uniforme	138
6.7	In breve...	138
7	FUNZIONI DI VARIABILI CASUALI	139
7.1	Variabili discrete	139
7.2	Variabili continue	139

7.3	Un esempio significativo: il quadrato di una variabile casuale	141
7.3.1	Il quadrato di una distribuzione uniforme	141
7.3.2	Il quadrato di una Gaussiana	142
7.4	La distribuzione del $\chi^2$	143
7.5	Un caso peculiare: la distribuzione di Cauchy	145
7.5.1	Alcune proprietà della distribuzione di Cauchy	146
7.5.2	La distribuzione di Cauchy ed il teorema del limite centrale	147
7.6	Ancora sul <i>random walk</i>	147
7.7	In breve...	149
8	METODI DI FIT	151
8.1	Descrizione generale del problema	151
8.2	Introduzione informale al fit dei minimi quadrati	152
8.2.1	Fit dei minimi quadrati nel caso costante: la media pesata	153
8.2.2	Fit dei minimi quadrati nel caso lineare	155
8.2.3	La covarianza tra i parametri nel fit lineare	157
8.2.4	Fit dei minimi quadrati con un polinomio di grado arbitrario	158
8.2.5	Fit dei minimi quadrati non pesato	159
8.3	Il test del $\chi^2$	159
8.3.1	Il contenuto fisico del $\chi^2$ : quanti gradi di libertà?	160
8.3.2	Il test del $\chi^2$ per una serie di dati	161
8.3.3	Il grafico dei residui	163
8.3.4	Fit dei minimi quadrati e test del $\chi^2$ per una distribuzione	164
8.3.5	Poissoniana o binomiale? Un semplice esempio	165
8.3.6	Il concetto di <i>p-value</i>	167
8.3.7	Fit dei minimi quadrati e test del $\chi^2$ con errori non Gaussiani	168
8.4	Il fit in pratica: metodi numerici	169
8.4.1	Digressione: il metodo di Newton in una dimensione	170
8.4.2	Minimi quadrati non lineari e metodi iterativi	171
8.4.3	Il problema dei valori iniziali	174
8.4.4	Un esempio di fit complesso	175
8.5	Fit di tipo generale	177
8.5.1	L'algoritmo di orthogonal distance regression (ODR)	178
8.5.2	Una possibile modifica al metodo dei minimi quadrati	179
8.6	In breve...	182
9	CENNI AL PRINCIPIO DI MASSIMA VEROSIMIGLIANZA	183
9.1	Di nuovo sul concetto di verosimiglianza	183
9.1.1	Un esempio concreto: il campionamento singolo di una variabile Poissoniana	184
9.2	Stime di massima verosimiglianza	185
9.2.1	Stime di massima verosimiglianza nel caso Poissoniano	185
9.2.2	Stime di massima verosimiglianza nel caso Gaussiano	186
9.2.3	Stime di massima verosimiglianza nel caso uniforme	187
9.3	Digressione: la distribuzione di Gauss e gli errori di misura	187
9.4	Incerezze nelle stime di massima verosimiglianza	188
9.4.1	Incerezze nel caso unidimensionale	188
9.4.2	Incerezze nel caso generale	189
9.5	Il $\chi^2$ come applicazione del principio di massima verosimiglianza	189
9.6	In breve...	189
10	COMPLEMENTI DI STATISTICA	191
10.1	La funzione generatrice dei momenti	191
10.2	La distribuzione <i>t</i> di Student	191
10.3	In breve...	191
A	TAVOLE NUMERICHE	195
A.1	Integrale normale degli errori—I	196
A.2	Integrale normale degli errori—I	197



A.3	Integrale normale degli errori—II	198
A.4	Integrale normale degli errori—III	199
A.5	Integrale normale degli errori—III	200
A.6	Funzione cumulativa del $\chi^2$	201
B	LA RELAZIONE DI LABORATORIO	203
B.1	Considerazioni generali	203
B.2	Carta, penna e calamaio, oppure il computer?	203
B.3	La struttura della relazione	204
B.3.1	Informazioni generali	204
B.3.2	Scopo dell'esperienza	204
B.3.3	Cenni teorici	204
B.3.4	Apparato sperimentale e strumenti	205
B.3.5	Descrizione delle misure	205
B.3.6	Analisi dei dati	205
B.3.7	Conclusioni	205
B.4	Cosa non dimenticare, e cosa non fare	205
B.4.1	Cose da non dimenticare	205
B.4.2	Cose da non fare	206
B.5	In conclusione...	206
C	ALCUNI STRUMENTI IN USO IN LABORATORIO	207
C.1	Il calibro ventesimale	207
C.1.1	Il nonio	207
C.1.2	Lettura del calibro	208
C.2	Il micrometro Palmer	209
C.2.1	Lettura del calibro Palmer e consigli pratici	209
D	SISTEMA BINARIO E RAPPRESENTAZIONE DEI NUMERI	211
D.1	Sistemi di numerazione in base diversa da 10	211
D.1.1	Il sistema binario	211
D.2	La rappresentazione dei numeri in un calcolatore	213
D.2.1	Numeri interi	213
D.2.2	Numeri in virgola mobile	213



---

## PREFAZIONE

---

Queste dispense sono state compilate a supporto del corso di Laboratorio 1 per gli studenti del primo anno del corso di Laurea in Fisica presso l'Università di Pisa durante l'anno accademico 2016–2017. Anche se l'impostazione è in parte mutuata dalle dispense della Prof.ssa Liana Martinelli (che per anni sono state utilizzate come testo di riferimento per il corso, e sono adesso liberamente disponibili alla pagina web <https://bitbucket.org/lbaldini/dispenselab1>) credo si possa dire senza tema di smentita che i due manoscritti siano abbastanza diversi tra loro da poterli considerare, a tutti gli effetti, indipendenti.

Il tema centrale e filo conduttore delle dispense è quello della rappresentazione e dell'analisi dei dati. Accanto agli argomenti *classici* che si trovano virtualmente in qualsiasi testo di carattere introduttivo (elementi di teoria delle probabilità e di statistica, propagazione degli errori, metodi di *fit*) si è cercato di dare spazio ad elementi appena più avanzati ma indispensabili nel bagaglio di conoscenze del Fisico (e.g., i metodi Monte Carlo), nella speranza che possano suscitare interesse e rendere meno arido il materiale.

Il livello della trattazione è il più elementare possibile—compatibilmente con la sofisticatezza di alcuni concetti che è pur necessario assimilare in un corso di laboratorio del primo anno. Si dà per scontato che lo studente sia familiare con il calcolo differenziale ed integrale, principalmente per funzioni di una variabile, ma, a parte questo, le dispense hanno l'obiettivo di essere per quanto possibile auto-contenute ed i concetti necessari sono via via richiamati al momento opportuno.

I punti qualificanti di queste dispense, che derivano direttamente dall'esperienza di insegnamento accumulata con gli studenti del primo anno, sono tre:

- ▶ i calcoli sono svolti dall'inizio alla fine, cercando di non saltare nessun passaggio intermedio, ed ampiamente commentati—a costo di sacrificare a tratti la sintesi e la scorrevolezza del testo;
- ▶ i concetti rilevanti sono illustrati da un gran numero di esempi concreti, sviluppati in modo dettagliato, a diversi livelli di difficoltà;
- ▶ il testo è corredato da frammenti di codice (in Python) che illustrano come si possano implementare in pratica alcuni dei concetti che descritti in astratto.

I frammenti di codice sono particolarmente importanti, perché il laboratorio è materia che si impara sul campo—ed una comprensione solida della teoria è utile solo nel limite in cui si riesce a tradurla ed applicarla in pratica. Gli esempi di codice non sono pensati per essere particolarmente efficienti o eleganti, ma sono, per la maggior parte, direttamente utilizzabili. Lo studente ha accesso alla versione elettronica su web di ciascun frammento di codice attraverso un *hyperlink* situato nella parte superiore della cornice corrispondente, e l'*output* fornito dall'esecuzione è riportato nella sua interezza in calce al frammento stesso, di modo che sia chiara non solo l'intenzione, ma anche il risultato.

```
https://bitbucket.org/.../hello_world.py
1 print('Hello world!')
2
3 [Output]
4 Hello world!
```

FRAMMENTO 1. Illustrazione del più semplice programma realizzabile in Python. Si notino l'*hyperlink* alla versione elettronica nella parte superiore della cornice e l'*output* completo dell'esecuzione nella parte inferiore.



Di tanto in tanto troverete sezioni che differiscono dalla maggior parte del resto del contenuto delle dispense—perché focalizzate su argomenti specifici più avanzati o per il livello leggermente più sofisticato degli strumenti matematici necessari. Queste sezioni sono indicate con il simbolo di curva pericolosa e, in una prima lettura, possono essere ignorate senza che questo pregiudichi la comprensione del resto. (Ma non fatelo sistematicamente, perché gli argomenti sono interessanti.)

Che dire di più? Buona lettura!



---

## RINGRAZIAMENTI

---

Queste dispense, ve ne accorgete, sono ancora un lavoro in corso. Per facilitare la segnalazioni di omissioni ed inesattezze e la raccolta di suggerimenti abbiamo una pagina web dedicata

<https://bitbucket.org/lbaldini/statnotes/issues>

in cui chiunque, anche in forma anonima, può contribuire a migliorare il manoscritto.

Quella che segue è una lista, potenzialmente incompleta, di tutte le persone (principalmente studenti del corso) che, nel tempo, hanno fornito consigli e suggerimenti, o più semplicemente hanno fatto domande che hanno generato cambiamenti nel testo—si spera per il meglio: Anonymous, Elia Bianucci, Edoardo Bogina, Benedetta Bruno, Andrea Canzio, Davide Carducci, Eva Castagna, Alberto Cilli, Gaia Da Prato, Mario Di Luca, Francesco Geroni, Tommaso Giovannelli, Kosimos, Andrei Lipovanu, Giovanni Megliola, Antonio A. Mele, Francesco Pellegrini, Gianluca Scardaci, Giovanni Spanu, Alessandro Tauro, Maria Paola Vaccaro.

(Se vi riconoscete in un alias e volete che appaia il vostro nome completo non esitate a mandare un messaggio a [luca.baldini@pi.infn.it](mailto:luca.baldini@pi.infn.it).)



---

## PRIMI PASSI: MISURE ED INCERTEZZE

---

Il concetto di *incertezza* o *errore* di misura gioca un ruolo centrale nelle scienze sperimentali. Visto che non è concettualmente possibile misurare una grandezza fisica con accuratezza infinita, il risultato di una misura manca di una parte importante del proprio contenuto quantitativo in assenza di una stima (implicita o esplicita) dell'incertezza ad essa associata—quest'ultima è elemento imprescindibile per poter confrontare una misura con una predizione teorica o con un'altra misura.

Nel linguaggio comune è piuttosto inusuale fare menzione esplicita delle incertezze di misura: vi immaginate comperare mezzo kilogrammo di pane al supermercato e chiedere all'inserviente “potrebbe tagliarlo con una tolleranza di 10 grammi, per cortesia?”. Eppure il fatto che, a fronte di una richiesta di mezzo kilogrammo, 600 grammi sarebbero probabilmente accettabili dalla maggior parte dei clienti—mentre 2 kilogrammi no—ci dice che in fondo la nozione di errore di misura deve essere in qualche modo connaturata con il nostro senso comune.

In Fisica, e nelle scienze sperimentali in genere, si cerca tipicamente di evitare ambiguità ed usare le parole in un modo per quanto possibile quantitativo e ben definito. Ciò nonostante anche in letteratura scientifica si trovano occasionalmente diciture in cui l'errore di misura è omesso—di solito volontariamente e consapevolmente. Anche in questo caso, come vedremo, è il bagaglio di conoscenze comuni (che cominceremo a sviluppare in questo capitolo) a guidarci.

### 1.1 IL CONCETTO DI ERRORE MASSIMO

Detto nei minimi termini, il concetto di *errore massimo* è legato alla domanda (apparentemente lecita ed innocua ma, come vedremo, patologicamente mal posta): “qual è il più piccolo intervallo che contiene con certezza il valore numerico della quantità che sto misurando?”. In altre parole, se scriviamo il risultato della misura di una generica grandezza fisica  $x$  come:

$$x = \hat{x} \pm \Delta x \text{ [unità di misura]} \quad (1)$$

intendendo  $\Delta x$ <sup>1</sup> come errore massimo, allora in effetti stiamo dicendo che l'intervallo  $[\hat{x} - \Delta x, \hat{x} + \Delta x]$  è il più piccolo intervallo possibile che ci dia la certezza di includere il valore (incognito) di  $x$ . Se scriviamo, ad esempio,  $l = 12.4 \pm 0.2$  cm, stiamo dicendo di essere certi che  $l$  sia compreso tra 12.2 e 12.6 cm—e l'intervallo  $[12.2, 12.6]$  cm è il più piccolo che ci possa dare tale certezza. (Quest'ultima condizione è ovviamente necessaria per definire univocamente il concetto.)

Gli ingredienti della (1) hanno ciascuno un significato ben preciso che è bene fissare il prima possibile:  $x$  è il valore (incognito) della grandezza che vogliamo misurare, che chiameremo *misurando* (cercando deliberatamente di evitare l'espressione *valore vero*, che pure si trova usata in letteratura);  $\hat{x}$  è la *miglior stima* di  $x$  che possiamo fornire a partire dai dati a nostra disposizione—che chiameremo anche *valore centrale* della misura;  $\Delta x$  è l'*errore massimo* di misura, nel senso che abbiamo precisato sopra (notiamo che, per definizione,  $\Delta x$  rappresenta la lunghezza di un intervallo ed è perciò una grandezza definita positiva); notiamo infine che il risultato di una misura non ha senso se non sono indicate le unità.

Anticipiamo che, quando il risultato di una misura fluttua, la nozione di errore massimo porta in sé una contraddizione logica dovuta al fatto che, da un punto di vista operativo, non è possibile garantire a priori che una nuova misura della stessa grandezza non fornisca un valore al di fuori dell'intervallo iniziale di incertezza. Il che porta come corollario un ulteriore problema insanabile: se questo accade

---

<sup>1</sup> Per evitare qualsiasi ambiguità, cercheremo di essere consistenti nell'indicare l'errore massimo con la lettera  $\Delta$  e l'errore statistico, una volta che l'avremo introdotto, con la lettera  $\sigma$ .

siamo costretti ad allargare tale intervallo—ci troviamo, cioè, nella situazione assurda in cui acquisire nuova informazione può solo peggiorare (o, al massimo, lasciare invariato) il nostro stato di conoscenza, nella misura in cui esso è determinato dall'incertezza di misura.

Formalizzeremo più avanti una nozione più sofisticata di incertezza di misura (quello che, in modo non molto felice, viene tipicamente chiamato *errore statistico*) che risolverà questi problemi—ed altri che non siamo attrezzati per affrontare in questo momento. E a quel punto abbandoneremo il concetto di errore massimo.

### 1.1.1 Errore assoluto ed errore relativo

Definiamo *errore relativo* (detto anche *errore percentuale*) il rapporto tra l'incertezza della misura ed il suo valore centrale. In particolare, utilizzando la nostra notazione, l'errore massimo relativo è

$$\text{Errore massimo relativo} := \frac{\Delta x}{|\hat{x}|}. \quad (2)$$

Poiché l'incertezza ha, per ovvie ragioni, le stesse dimensioni fisiche della misura a cui si riferisce, l'errore relativo è una quantità adimensionale (e definita positiva).

► **ESEMPIO 1.1.** Si misura una lunghezza  $l$  ottenendo il risultato  $l = 12.4 \pm 0.2$  cm. L'errore relativo è in questo caso 0.016, vale a dire 1.6%.

### 1.1.2 Breve digressione: precisione ed accuratezza

In letteratura si trova la distinzione tra i concetti di *accuratezza*, intesa come accordo tra il valore della misura e quello del misurando, e di *precisione*, intesa come consistenza tra risultati di successive misure della stessa quantità nelle medesime condizioni.

Si tratta di una distinzione molto rilevante, poiché un dato strumento può essere accurato ma poco preciso (se fornisce in media la risposta giusta, ma con fluttuazioni rilevanti tra misurazioni successive) o—il che è potenzialmente più pericoloso—molto preciso ma poco accurato (se fornisce in modo estremamente riproducibile la risposta sbagliata). Avremo tempo di tornare sull'argomento in seguito, ma diciamo fin da subito che cercheremo scrupolosamente di usare questi due termini nell'accezione appena indicata.

## 1.2 CIFRE SIGNIFICATIVE E CONVENZIONI DI SCRITTURA

Prima ancora di capire come si stima in pratica l'incertezza massima di misura, ci soffermiamo per un attimo su alcune questioni, semplici ma importanti, di notazione. In un articolo scientifico non troverete mai scritto (almeno sperabilmente) che la massa di un oggetto è  $74.562572 \pm 0.024894$  kg. Ed il motivo è che, se l'errore di misura è dell'ordine di 25 g, è assurdo scrivere il valore centrale fino al mg. Le ultime tre cifre (almeno) sono irrilevanti—semplice rumore numerico.

► **ESEMPIO 1.2.** Negli esempi che seguono indichiamo con una sottolineatura la cifra meno significativa, con una sopralineatura quella più significativa e tra parentesi quadre il numero di cifre significative:

$\overline{3}11\overline{5}$	[4]
$\overline{3}212.\overline{5}$	[5]
$0.0\overline{3}2\overline{5}$	[3]
$0.0\overline{3}0\overline{0}$	[3]
$\overline{3}00.\overline{00}$	[5]

Facciamo un passo indietro. La questione si può inquadrare quantitativamente attraverso la nozione di *cifra significativa*, che si definisce mediante le tre semplici regole. Dato un numero generico: (i) la cifra più



significativa è quella più a sinistra e diversa da zero; (ii) la cifra meno significativa è quella più a destra (fatta eccezione per gli zeri nei numeri interi); (iii) tutte le cifre comprese tra la più significativa e la meno significativa sono significative. L'esempio 1.2 illustra il concetto.

Torniamo allora al nostro problema originale. La nozione di cifra significativa è strettamente connessa con quella di arrotondamento del valore centrale della misura ed il numero di cifre significative con cui si deve scrivere tale valore è determinato dall'incertezza associata. Più precisamente si scrive l'incertezza di misura (arrotondando opportunamente) con una o, al massimo, due cifre significative, e si arrotonda il valore centrale in modo da essere consistente, in termini di cifre decimali, con l'incertezza associata. Così  $74.563 \pm 0.025$  kg e  $74.56 \pm 0.02$  kg sono entrambe scritture accettabili per il peso dell'oggetto con cui abbiamo aperto questa sezione. (Va da sé che, quando i risultati di misure, dirette o indirette, vengono usati per stimare grandezze derivate, è buona norma avere cura che gli arrotondamenti effettuati nei passaggi intermedi non introducano errori significativi nel risultato finale.)

► **ESEMPIO 1.3.** La seguente è una breve lista di errori tipici che si possono compiere; i valori tra parentesi quadre indicano la (o una) possibile versione corretta.

$3.436 \pm 0.542$	errore con troppe (tre) cifre significative	$[3.44 \pm 0.54]$
$3.4 \pm 0.54$	errore con una cifra significativa di troppo	$[3.4 \pm 0.5]$
	oppure valore con una cifra significativa mancante	$[3.40 \pm 0.54]$
$3.44 \pm 0.5$	valore con una cifra significativa di troppo	$[3.4 \pm 0.5]$
	oppure errore con una cifra significativa mancante	$[3.44 \pm 0.50]$
$3000 \pm 100$	ambigua per gli zeri meno significativi	$[(3.0 \pm 0.1) \times 10^3]$

Per completezza notiamo che non è inconsueto in letteratura che l'incertezza di misura venga (consapevolmente) omessa. In questi casi la precisione del risultato sperimentale è, almeno parzialmente, implicita nel modo in cui il risultato stesso è scritto, e si può assumere che l'errore sia da considerarsi pari ad una (o, talvolta, 5) unità della cifra meno significativa del valore centrale. Così i valori 5.0 m, e 5.00 m hanno significati diversi—il primo è da leggersi come  $5.0 \pm 0.1$  m o  $5.0 \pm 0.5$  m, mentre il secondo come  $5.00 \pm 0.01$  m o  $5.00 \pm 0.05$  m. Inutile a dirsi, riportare esplicitamente l'errore di misura è il modo migliore (e l'unico) per evitare ambiguità di sorta.

### 1.3 UNITÀ DI MISURA E ORDINI DI GRANDEZZA

L'accuratezza delle misure raggiunta in alcune aree della Fisica è niente di meno che stupefacente. In elettrodinamica quantistica, una delle teorie Fisiche meglio verificata sperimentalmente, il momento magnetico dell'elettrone è misurato sperimentalmente a meglio di una parte su  $10^{12}$ —e la misura è in accordo con i calcoli teorici, che hanno a loro volta raggiunto un livello di sofisticazione tale da consentire di avere incertezze dello stesso ordine di grandezza. Tanto per fissare le idee: sarebbe come misurare la distanza tra Londra e New York con una precisione di 1  $\mu$ m (un millesimo di mm): non sfuggirà la portata dei problemi operativi connessi con il raggiungimento di un tale livello di riproducibilità, che è una delle aree di studio della moderna metrologia.

#### 1.3.1 Il Sistema Internazionale di unità di misura

Il primo sforzo sistematico dell'età moderna in direzione di un sistema standard di unità di misura è costituito dalla definizione, nella Francia della fine del XVIII secolo, del sistema metrico decimale—basato sulle unità standard di *metro* e *kilogrammo*. Definiti inizialmente come la decimilionesima parte della lunghezza del meridiano terrestre e come la massa di 1000 cm<sup>3</sup> di acqua alla temperatura di fusione del ghiaccio, il metro ed il kilogrammo sono stati ridefiniti circa un secolo più tardi sulla base di due manufatti campione (il metro campione ed il kilogrammo campione) realizzati in una lega di platino-iridio e conservati a Parigi (con un certo numero di repliche identiche distribuite ai principali istituti di metrologia). È chiaro che l'utilizzo di manufatti fisici come unità fondamentali costituisce di per sé un limite alla riproducibilità delle misure fisiche.

Ormai adottato quasi universalmente (sia pure con importanti eccezioni), il Sistema Internazionale (SI) di unità di misura costituisce ad oggi lo stato dell'arte nella materia. Il SI include 7 unità di base per la misura di altrettante grandezze fisiche: lunghezza, massa, tempo, corrente elettrica, temperatura termodinamica, quantità di sostanza ed intensità luminosa—come mostrato nella tabella 1.

Quantità	Simbolo	Unità	Simbolo
Lunghezza	l	metro	m
Massa	m	kilogrammo	kg
Tempo	t	secondo	s
Corrente elettrica	I, i	ampere	A
Temperatura	T	kelvin	K
Quantità di sostanza	n	mole	mol
Intensità luminosa	I <sub>v</sub>	candela	cd

TABELLA 1. Tabella delle 7 grandezze fisiche di base, con relative unità di misura, su cui è fondato il Sistema Internazionale. (Adattato da [8].)

Come elencato di seguito, tre delle 7 unità fondamentali sono definite indipendentemente da tutte le altre: il secondo in base ad una transizione atomica specifica dell'atomo di  $^{133}\text{Cs}$ , il kilogrammo in base ad un manufatto campione, e la temperatura in base al punto triplo dell'acqua. Le altre quattro unità sono invece dipendenti da queste tre—in particolare il metro è definito a partire dal secondo utilizzando la velocità della luce nel vuoto come costante fondamentale.

- Lunghezza [m]: il metro è la lunghezza percorsa dalla luce nel vuoto in un intervallo di tempo pari a  $1/299\,792\,458$  s. (Ne segue che la velocità della luce nel vuoto è *esattamente*  $299\,792\,458$  m/s).
- Massa [kg]: il kilogrammo è per definizione uguale alla massa del kilogrammo campione internazionale, un cilindro di una lega di platino ed iridio realizzato nel 1889.
- Tempo [s]: il secondo corrisponde alla durata di  $9\,192\,631\,770$  periodi della transizione tra i due livelli iperfini dello stato fondamentale dell'atomo di  $^{133}\text{Cs}$ .
- Corrente elettrica [A]: l'ampere corrisponde alla corrente costante che, se stabilita in due conduttori rettilinei paralleli di lunghezza infinita e sezione trascurabile, posti ad 1 m di distanza nel vuoto, produce tra i due conduttori una forza di  $2 \times 10^{-7}$  newton per metro di lunghezza.
- Temperatura termodinamica [K]: il kelvin corrisponde a  $1/273.16$  della temperatura termodinamica del punto triplo dell'acqua.
- Quantità di sostanza [mol]: la mole è la quantità di sostanza che contiene un numero di costituenti elementari pari al numero di atomi in  $0.012$  kg di  $^{12}\text{C}$ .
- Intensità luminosa [cd]: la candela è l'intensità luminosa, in una direzione fissata, di una sorgente luminosa che emette radiazione monocromatica di frequenza  $540 \times 10^{12}$  Hz con una intensità radiante, in quella direzione, di  $1/683$  W sr $^{-1}$ .

È importante sottolineare come il Sistema Internazionale sia in continua evoluzione, nel tentativo di rispondere alle esigenze sempre più stringenti dalla scienza e della tecnologia moderna. Uno degli obiettivi fondamentali è quello di arrivare ad un sistema in cui tutte le unità fondamentali siano definite in termini di *invarianti* della natura—costanti fisiche fondamentali o proprietà degli atomi. La frequenza della radiazione emessa in corrispondenza di una specifica transizione atomica e la velocità della luce nel vuoto sono due ottimi esempi in questo senso. Il kilogrammo è, ad oggi, l'unica unità di misura ad essere ancora definita sulla base di un manufatto, e la definizione di ampere, mole e candela dipendono da quella del kilogrammo. Questo implica un certo numero di limitazioni sostanziali, la più seria delle quali è il fatto che la stabilità della massa del kilogrammo campione non è garantita nel lungo periodo<sup>2</sup>.

È probabile che in un futuro non troppo lontano il kilogrammo sia ridefinito, all'interno del SI, sulla base della costante di Planck  $h$ . Ove ciò accadesse,  $h$  diverrebbe, come la velocità della luce nel vuoto,

<sup>2</sup> Il kilogrammo campione, conservato insieme a sei copie di riferimento presso il *Bureau International des Poids et Mesures* (BIPM), è soggetto ad una contaminazione superficiale al livello di  $\sim 1$   $\mu\text{g}$  all'anno. Questa contaminazione, largamente reversibile, rende periodicamente necessarie sofisticate operazioni di pulizia. Nonostante tutti gli accorgimenti del caso, tra 1889 al 1989 si è osservata una variazione relativa di massa del kilogrammo campione rispetto alle sei copie al livello di alcune decine di  $\mu\text{g}$  [7].

una costante esatta della natura, e la massa del kilogrammo campione diverrebbe invece una quantità misurata sperimentalmente. È altresì probabile che, nel più lungo termine, l'ampere venga ridefinito in base alla carica elementare dell'elettrone  $e$ , il kelvin in base al valore della costante di Boltzmann  $k$  e la mole in termini del numero di Avogadro  $N_A$ . (A quel punto  $e$ ,  $k$  ed  $N_A$  diverrebbero a loro volta costanti della natura definite esattamente.)

### 1.3.2 Unità derivate ed unità riconosciute

Alcune unità di misura derivate hanno, nel Sistema Internazionale, nomi speciali, che corrispondono ad una forma convenientemente compatta di combinazioni specifiche delle 7 unità di base. Una lista parziale di queste unità derivate è presentata nella tabella 2.

Quantità	Unità	Simbolo	Espressione SI
Frequenza	hertz	Hz	$s^{-1}$
Forza	newton	N	$m\ kg\ s^{-2}$
Pressione	pascal	Pa	$m^{-1}\ kg\ s^{-2}$
Energia	joule	J	$m^2\ kg\ s^{-2}$
Potenza	watt	W	$m^2\ kg\ s^{-3}$
Carica elettrica	coulomb	C	sA
Potenziale elettrico	volt	V	$m^2\ kg\ s^{-3}\ A^{-1}$
Capacità	farad	F	$m^{-2}\ kg^{-1}\ s^4\ A^2$
Resistenza	ohm	$\Omega$	$m^2\ kg\ s^{-3}\ A^{-2}$

TABELLA 2. Lista (non esaustiva) delle unità derivate che hanno nomi speciali nel Sistema Internazionale. Le corrispondenti espressioni in termini delle 7 unità fondamentali del Sistema Internazionale sono mostrate esplicitamente nell'ultima colonna. (Adattato da [8].)

In generale l'espressione, in termini delle unità fondamentali del SI, per una generica grandezza derivata si ricava, più o meno banalmente, da una qualsiasi legge fisica che coinvolga l'unità derivata. Così, dalla legge di Newton

$$F = ma \quad (3)$$

sappiamo che la forza è una massa (kg) per un'accelerazione ( $m\ s^{-2}$ ), per cui il newton corrisponde a  $m\ kg\ s^{-2}$ ; una pressione, d'altra parte, è una forza per unità di superficie ( $m^2$ ) per cui il pascal corrisponde a  $m^{-1}\ kg\ s^{-2}$ ; l'energia (o il lavoro) è il prodotto di una forza per uno spostamento (m), per cui si misura in  $m^2\ kg\ s^{-2}$ , e così via.

Vi è infine un certo numero di unità di misura che, pur non facendo parte del Sistema Internazionale, sono così comunemente usate (ad esempio il litro per le misure di volume) o sono così profondamente radicate nella nostra cultura (ad esempio le ore, i minuti ed i secondi per le misure di tempo) da essere ufficialmente riconosciute ed accettate. (Per inciso: l'atmosfera, che corrisponde a 101325 Pa, non fa parte delle unità riconosciute.)

Quantità	Unità	Simbolo	Conversione SI
Tempo	minuto	min	1 min = 60 s
	ora	h	1 h = 3600 s
	giorno	d	1 d = 86400 s
Volume	litro	l	1 l = 1 dm <sup>3</sup>
Massa	tonnellata	t	1 t = 1000 kg
	electronvolt	eV	1 eV $\approx 1.602 \times 10^{-19}$ J
Pressione	erg	erg	1 erg = 10 <sup>-7</sup> J
	bar	bar	1 bar = 10 <sup>5</sup> Pa
	mm di Hg	mmHg	1 mmHg $\approx 133.3$ Pa

TABELLA 3. Lista (non esaustiva) delle unità che, pur non facendo parte del Sistema Internazionale, sono riconosciute in quanto culturalmente o storicamente rilevanti. (Adattato da [8].)

### 1.3.3 Multipli, sottomultipli ed ordini di grandezza

Le grandezze oggetto di studio in Fisica possono differire tra loro di numerosi ordini di grandezza. Così, mentre il raggio di un protone è dell'ordine di  $10^{-15}$  m, le sorgenti astronomiche più lontane si trovano a

distanze di svariati Gpc (gigaparsec), cioè dell'ordine di  $10^{26}$  m. Analogamente, mentre la vita media dei bosoni  $W$  e  $Z_0$  è dell'ordine di  $10^{-25}$  s, l'età del nostro universo è dell'ordine di 10 miliardi di anni, o  $10^{17}$  s. In entrambi i casi la gamma dinamica corrisponde a più di 40 ordini di grandezza. A questo scopo il Sistema Internazionale definisce una serie di prefissi per facilitare la scrittura di grandezze il cui valore sia molto più piccolo o molto più grande della relativa unità di misura, come mostrato in tabella 4.

Fattore	Nome	Simbolo	Fattore	Nome	Simbolo
$10^1$	deca	da	$10^{-1}$	deci	d
$10^2$	hecto	h	$10^{-2}$	centi	c
$10^3$	kilo	k	$10^{-3}$	milli	m
$10^6$	mega	M	$10^{-6}$	micro	$\mu$
$10^9$	giga	G	$10^{-9}$	nano	n
$10^{12}$	tera	T	$10^{-12}$	pico	p
$10^{15}$	peta	P	$10^{-15}$	femto	f
$10^{18}$	exa	E	$10^{-18}$	atto	a
$10^{21}$	zetta	Z	$10^{-21}$	zepto	z
$10^{24}$	yotta	Y	$10^{-24}$	yocto	y

TABELLA 4. Prefissi per i multipli e sottomultipli decimali delle unità di misura del Sistema Internazionale. Il kg si trova nella situazione peculiare di rappresentare una delle unità fondamentali pur essendo preceduto dal prefisso k. (Adattato da [8].)

#### 1.3.4 Cenni di calcolo dimensionale

Cogliamo l'occasione per accennare ad un argomento di estrema rilevanza legato alle dimensioni fisiche, ovvero l'analisi dimensionale. Quando si ha a che fare con un sistema fisico descritto da grandezze non adimensionali, le equazioni che lo descrivono debbono soddisfare il requisito di base di essere *dimensionalmente corrette*—cioè le dimensioni fisiche dei due membri devono essere omogenee. In caso contrario l'equazione è semplicemente sbagliata—e controllare sistematicamente le dimensioni fisiche in ogni passaggio di un calcolo permette spesso di evitare errori banali.

Nel seguito indicheremo le dimensioni fisiche di una generica grandezza con il simbolo corrispondente in tabella 1 racchiuso in parentesi quadre. Così, e.g., diremo che il raggio della Terra ha le dimensioni di una lunghezza, che indicheremo con  $[l]$ . Le grandezze derivate seguono le regole del prodotto, e le dimensioni fisiche di una velocità, e.g., sono quelle di una lunghezza per l'inverso di un tempo, o  $[l\ t^{-1}]$ .

Supponiamo allora di essere interessati alla relazione che lega il periodo  $T$  di un pendolo alla sua lunghezza  $l$ . Se qualcuno scrivesse

$$T = 2\pi \frac{l}{g} \quad (\text{dimensionalmente non corretta}) \quad (4)$$

potremmo smentire l'affermazione, anche ignorando le leggi fondamentali della meccanica, sulla base del fatto che le dimensioni fisiche del membro di destra dell'equazione sono quelle di un tempo al quadrato e non quelle di un tempo

$$[t] \neq [l] \times [l\ t^{-2}]^{-1} = [t^2] \quad (5)$$

L'analisi dimensionale non è solo utile come verifica minima (che dovrebbe sempre essere fatta, ricordate!) dei risultati di un calcolo, ma può anche avere potere predittivo. Supponiamo di essere interessati alla relazione funzionale che lega una generica grandezza  $y$  ad una serie di grandezze (dimensionate)  $x_1, x_2 \dots x_n$ . Anche se in generale non vi è garanzia che questo funzioni, possiamo provare a scrivere

$$y = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}. \quad (6)$$

Il requisito che l'equazione scritta sopra sia dimensionalmente corretta può essere trasformato in un sistema di equazioni lineari per gli esponenti  $\alpha_1, \alpha_2 \dots \alpha_n$  che a sua volta può fornire informazioni non banali sul sistema che stiamo studiando, come illustrato negli esempi 1.4 e 1.5.

► **ESEMPIO 1.4.** Torniamo all'esempio del pendolo. È plausibile ipotizzare che il periodo  $T$  possa dipendere dalla lunghezza  $l$  del pendolo e dalla sua massa  $m$ , oltre che dall'accelerazione di gravità  $g$ . Se assumiamo una dipendenza funzionale del tipo  $T = l^{\alpha_1} m^{\alpha_2} g^{\alpha_3}$ , da un punto di vista dimensionale possiamo scrivere

$$[t] = [l]^{\alpha_1} \times [m]^{\alpha_2} \times [l \, t^{-2}]^{\alpha_3} = [l]^{\alpha_1 + \alpha_3} \times [m]^{\alpha_2} \times [t]^{-2\alpha_3},$$

che a sua volta può essere trasformato nel sistema di equazioni

$$\begin{cases} \alpha_1 + \alpha_3 = 0 \\ \alpha_2 = 0 \\ -2\alpha_3 = 1 \end{cases} \quad \text{ossia} \quad \begin{cases} \alpha_1 = 1/2 \\ \alpha_2 = 0 \\ \alpha_3 = -1/2. \end{cases}$$

L'analisi dimensionale ci permette dunque di concludere che il periodo del pendolo non può dipendere dalla massa (a meno che non inseriamo nel problema dall'esterno un'ulteriore grandezza che abbia la massa nelle sue dimensioni fisiche) e che l'equazione cercata deve essere nella forma

$$T \propto \sqrt{\frac{l}{g}}.$$

(Per completezza: la conclusione è corretta ed il fattore di proporzionalità è, nel limite di piccole oscillazioni,  $2\pi$ . Nel caso di ampiezza iniziale  $\theta_0$  arbitrariamente grande si ha un ulteriore fattore moltiplicativo adimensionale, funzione di  $\theta_0$  che non si può ovviamente ricavare per via dimensionale.)

► **ESEMPIO 1.5 (TERZA LEGGE DI KEPLERO).** Ci proponiamo di ricavare per via dimensionale la terza legge di Keplero, ossia la relazione che lega il periodo orbitale  $T$  di un pianeta al semiasse maggiore  $a$  della sua orbita. Assumeremo che gli ulteriori parametri che entrano nel problema siano la costante di gravitazione universale  $G$  e la massa del Sole  $m_\odot$  e che possiamo scrivere la relazione cercata come  $T = G^{\alpha_1} m_\odot^{\alpha_2} a^{\alpha_3}$ . (Se a questo punto vi state chiedendo perché non abbiamo incluso la massa del pianeta nella lista... si tratta di un'ottima domanda, cui l'analisi dimensionale da sola non fornisce una risposta soddisfacente, ma procediamo ugualmente e vediamo dove ci porta il nostro ragionamento.) Le unità di misura di  $G$  si possono ricavare, e.g. dalla legge di gravitazione universale

$$F = G \frac{m_1 m_2}{r^2} \quad \text{da cui} \quad G = \frac{F r^2}{m_1 m_2} \quad (7)$$

e sono, nel sistema internazionale  $m^3 \, kg^{-1} \, s^{-2}$ . Ragionando come prima si ha allora

$$[t] = [l^3 \, m^{-1} \, t^{-2}]^{\alpha_1} \times [m]^{\alpha_2} \times [l]^{\alpha_3} = [l]^{3\alpha_1 + \alpha_3} \times [m]^{\alpha_2 - \alpha_1} \times [t]^{2\alpha_1},$$

da cui

$$\begin{cases} 3\alpha_1 + \alpha_3 = 0 \\ \alpha_2 - \alpha_1 = 0 \\ -2\alpha_1 = 1 \end{cases} \quad \text{ossia} \quad \begin{cases} \alpha_1 = -1/2 \\ \alpha_2 = -1/2 \\ \alpha_3 = 3/2. \end{cases}$$

Ne possiamo trarre la conclusione, tutt'altro che ovvia

$$T \propto \frac{a^{\frac{3}{2}}}{\sqrt{G m_\odot}}. \quad (8)$$

(E, per completezza, la conclusione è di nuovo corretta ed il fattore di proporzionalità è di nuovo  $2\pi$ —ma non fateci l'abitudine, perché si tratta di una semplice coincidenza.)

## 1.4 STIMA DELL'ERRORE MASSIMO IN CONDIZIONI DI RIPETITIVITÀ

Torniamo adesso alla discussione delle incertezze di misura che abbiamo temporaneamente messo da parte dopo la sezione 1.1. Quando una misura è fatta più volte in condizioni di ripetitività si hanno due casi tipici: (i) la dispersione delle misure è nulla—cioè otteniamo sempre lo stesso risultato—oppure (ii) il valore ottenuto fluttua—cioè misure successive forniscono in generale risultati diversi tra loro.

## 1.4.1 Misure in regime di dispersione nulla: la risoluzione strumentale

La *risoluzione* di uno strumento è la più piccola *variazione* della quantità da misurare che è possibile apprezzare con lo strumento stesso<sup>3</sup>. Così la risoluzione di un metro a nastro è 1 mm, quella di un calibro ventesimale 0.05 mm e quella di un calibro cinquantessimale 0.02 mm, tanto per fare alcuni esempi.

Quando eseguiamo una misura in condizioni di dispersione nulla (cioè misure successive della stessa quantità forniscono sempre lo stesso risultato), almeno nei casi più semplici si può assumere la risoluzione strumentale (oppure la metà della risoluzione strumentale) come errore massimo di misura.

► **ESEMPIO 1.6.** Se misuriamo la lunghezza  $l$  di un tavolo con un metro a nastro (con una risoluzione di 1 mm) ottenendo il valore 189.4 cm, scriveremo la nostra misura come  $l = 189.4 \pm 0.1$  cm. Alternativamente possiamo scrivere  $l = 189.40 \pm 0.05$  cm, sulla base del fatto che, se abbiamo scelto la tacca corrispondente a 189.4 come risultato della misura, implicitamente stiamo dicendo che il misurando è più vicino a 189.4 che a 189.3 o 189.5—e dunque compreso tra 189.35 e 189.45.

► **ESEMPIO 1.7.** Misuriamo la temperatura  $T$  dell'acqua in un recipiente con un termometro digitale ed il *display* indica 34.5°C. In mancanza di informazioni aggiuntive (ad esempio fornite dal *data sheet* dello strumento), dobbiamo assumere che tutte le cifre siano significative e che la risoluzione dello strumento sia 0.1°C. Scriveremo dunque il risultato della misura come  $T = 34.5 \pm 0.1$ °C o come  $T = 34.50 \pm 0.05$ °C.

Per quanto apparentemente banali, i due esempi precedenti non devono trarre in inganno: la vita è spesso più complicata e a volte identificare l'accuratezza di una misura (nel senso della distanza dal misurando) con la risoluzione dello strumento utilizzato può portare ad errori apprezzabili. In una determinata situazione potremmo avere difficoltà pratiche nella lettura del metro a nastro o nella definizione stessa della lunghezza da misurare (ad esempio per problemi di messa a fuoco nelle misure di ottica). Oppure il nostro termometro potrebbe essere *scalibrato* di un fattore che non conosciamo per cui, dato lo stesso oggetto, ci fornisce sempre la stessa temperatura, ma quest'ultima è sistematicamente sbagliata. Torneremo brevemente sulla questione dei cosiddetti errori sistematici nella sezione 1.6.

## 1.4.2 Un semplice esperimento: l'incertezza di misura con il metro a nastro

A corredo della discussione sulla stima dell'incertezza di misura, in figura 1.1 mostriamo i risultati di un semplice esperimento in cui si è chiesto ad un gruppo di studenti di Fisica del primo anno di misurare (indipendentemente) con un metro a nastro, *interpolando tra le divisioni al meglio delle loro capacità* le lunghezze  $l_1$  ed  $l_2$  dei due lati di una lastrina di ottone.

I due istogrammi (se non sai cosa è un istogramma puoi saltare per un attimo alla sezione 2.3) sono interessanti sotto diversi aspetti. Per prima cosa questa è una situazione in cui, se arrotondassimo il risultato della misura alla divisione più vicina, ci aspetteremmo di essere in regime di dispersione nulla—ci aspetteremmo cioè che (modulo errori di lettura) tutte le misure distino meno di mezzo millimetro dalla divisione del metro a nastro più vicina al misurando. Ora: benché non sia possibile conoscere il misurando con precisione infinita, possiamo misurare le lunghezze dei due lati della nostra lastrina con un calibro centesimale—cosa che abbiamo fatto ed è indicata dalle linee verticali tratteggiate in figura. È

<sup>3</sup> Notiamo, per completezza, che la risoluzione in generale non coincide con la più piccola quantità che uno strumento può apprezzare (che a volte si dice *sensibilità*). Notiamo anche che la risoluzione non coincide necessariamente con l'unità di formato dello strumento (e.g., la divisione più piccola di una scala graduata).

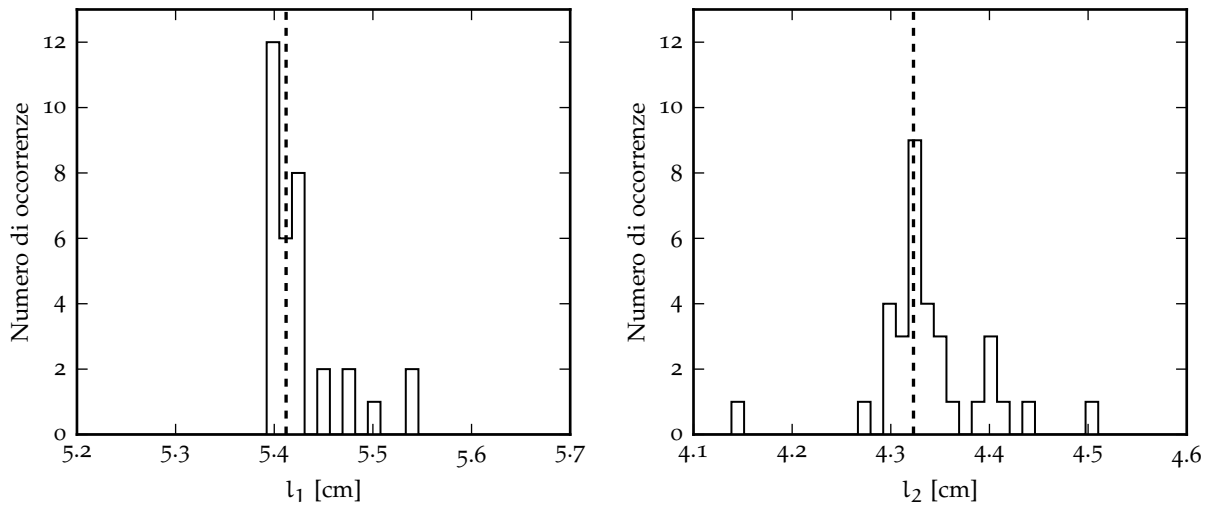


FIGURA 1.1. Istogramma delle lunghezze  $l_1$  ed  $l_2$  dei due lati di una lastrina di ottone misurate indipendentemente da un gruppo di studenti, interpolando ad occhio tra le divisioni di un metro a nastro. Le linee verticali indicano le stesse lunghezze misurate con un calibro centesimale.

ovvio allora che le nostre aspettative sono disilluse: vi sono misure che distano 2 mm (il doppio della risoluzione strumentale) o più dal misurando. La prima conclusione è dunque che se non stiamo attenti a non fare errori di lettura, la nostra stima dell'errore massimo è errata anche se prendiamo come incertezza di misura la risoluzione dello strumento.

Se guardiamo meglio gli istogrammi in figura 1.1, però, ci accorgiamo che la maggioranza delle misure in effetti è contenuta in un intervallo la cui semilarghezza (diciamo  $\sim 2/10$  mm) è significativamente più piccola della risoluzione strumentale. Questo ci dice che misurando con attenzione può aver senso interpolare tra le divisioni del metro a nastro—e, come abbiamo visto, possiamo valutare quantitativamente il nostro grado di attenzione utilizzando uno strumento relativamente più preciso come il calibro centesimale.

#### 1.4.3 Fluttuazioni casuali: stima a posteriori dell'errore massimo

Il caso in cui i valori misurati sono soggetti a fluttuazioni statistiche è chiaramente, dal nostro punto di vista, più interessante—e, come abbiamo già detto, è questo il caso in cui la nozione di errore massimo diventa problematica. Le fluttuazioni possono essere dovute alle caratteristiche dello strumento di misura (ad esempio il rumore in un dispositivo elettronico), alle proprietà intrinseche del sistema fisico sotto studio (ad esempio effetti quanto-meccanici), oppure ad una combinazione dei due fattori.

In situazioni di questo tipo possiamo stimare a posteriori l'incertezza in base alla dispersione attorno al valor medio di misure successive eseguite in condizioni di ripetibilità. Se eseguiamo, cioè,  $n$  misure  $x_1, x_2, \dots, x_n$  della grandezza  $x$  cui siamo interessati, è ragionevole a questo livello prendere la media aritmetica delle misure stesse come valore centrale e la semidispersione come errore massimo:

$$\begin{cases} \hat{x} &= \frac{1}{n} \sum_{i=1}^n x_i \\ \Delta x &= \frac{x_{\max} - x_{\min}}{2}. \end{cases} \quad (9)$$

Per quanto apparentemente sensata, si tratta di nuovo di una ricetta che possiamo utilizzare per avere un'idea dell'incertezza di misura in alcune situazioni semplici, ma che è afflitta dai problemi insanabili descritti alla fine della sezione 1.1. Il primo è che, data una serie finita di misure, nessuno ci assicura che misure successive non possano cadere al di fuori della semidispersione iniziale—il che, a rigore, contraddice la nostra definizione di errore massimo. Il secondo è che all'aumentare del numero  $n$  di misure la semidispersione non può che aumentare, il che ci lascia apparentemente nella situazione assurda

► **ESEMPIO 1.8.** Vogliamo misurare il periodo  $\tau$  di un pendolo ed abbiamo a disposizione un cronometro digitale con risoluzione di 0.01 s. Si tratta di un caso interessante, in cui l'apparato di misura è costituito, per così dire, dal combinato del cronometro e della persona che fisicamente lo utilizza. Non è banale quantificare a priori l'incertezza di misura, ma possiamo sospettare che le fluttuazioni del tempo di reazione umano rappresentino il contributo più rilevante.

Eseguiamo dunque 5 misurazioni del periodo ottenendo i valori: 2.12 s, 2.22 s, 2.16 s, 2.10 s e 2.15 s. (Come ci aspettavamo la risoluzione dello strumento è più piccola del tempo di reazione e di fatto non gioca alcun ruolo nella misura.) Seguendo la (9), abbiamo  $\tau = 2.15 \pm 0.06$  s.

in cui eseguire nuove misure (cioè aggiungere informazione) causa un incremento dell'incertezza, mentre intuitivamente ci aspetteremmo il contrario.

#### 1.4.4 Si può ridurre l'incertezza di misura al di sotto della risoluzione strumentale?

Il titolo di questa sezione può sembrare provocatorio, ma la risposta, in generale, è affermativa—ed in effetti abbiamo già visto con l'esperimento descritto nella sezione 1.4.2 che interpolare tra le tacche di un metro a nastro è una via possibile. C'è un'altra situazione tipica in cui si può battere (per così dire) la risoluzione strumentale—quella cioè in cui si ha a disposizione un certo numero di copie *identiche* (o abbastanza simili da poter essere considerate tali) dell'oggetto o della grandezza che vogliamo misurare. In tal caso si può fare una misura diretta della somma di queste copie e *spalmare* la risoluzione strumentale su di esse, dividendo il valore centrale e l'errore per il loro numero, come illustrato negli esempi

► **ESEMPIO 1.9.** Si vuole misurare lo spessore  $s$  di un foglio di carta disponendo solo di un metro a nastro. A tale scopo si misura lo spessore  $h$  di una risma di 500 fogli, ottenendo il valore  $h = 40 \pm 1$  mm. Dividendo tutto per 500, possiamo scrivere  $s = 0.080 \pm 0.002$  mm, con un errore massimo 500 volte più piccolo della risoluzione del metro a nastro. (Per completezza è importante notare come questo risultato dipenda dall'assunzione implicita, che può essere verificata o meno in pratica, che i 500 fogli abbiano lo stesso spessore nei limiti dell'incertezza sperimentale.)

► **ESEMPIO 1.10.** Si vuole misurare il peso di uno spillo con una bilancia con risoluzione di 1 g. Analogamente all'esempio precedente, se si dispone di un gran numero di spilli uguali, possiamo pesarli tutti insieme e dividere il valore centrale e l'incertezza di misura sul peso complessivo per questo numero. (Esattamente come prima, l'assunzione che gli spilli possano essere considerati identici è da verificarsi in pratica.)

A complemento di questa sezione descriviamo il risultato di un semplice esperimento in cui si è chiesto ad un gruppo di studenti di misurare ripetutamente (per dieci volte) la durata di (i) un singolo periodo e (ii) di 10 periodi del moto armonico di un cerchietto colorato simulato su un calcolatore portatile e proiettato su uno schermo. (Il vantaggio di utilizzare una simulazione anziché un oggetto reale è la possibilità di controllare il periodo con una precisione maggiore di tutte le incertezze in gioco nell'esperimento—che nel nostro linguaggio significa essenzialmente conoscere il valore del misurando). Va da sé che l'idea di base è che, dividendo le misure di 10 periodi per 10, ci aspettiamo di ottenere una misura più precisa che non misurando il periodo singolo direttamente.

In figura 1.2 sono mostrate (sulla stessa scala orizzontale per evidenziare la differenza) le distribuzioni su un campione di 93 studenti (per un totale di 930 valori) delle misure di 1 e 10 periodi—queste ultime ovviamente divise per 10. Le prime variano di qualche decimo di secondo (cioè qualche decina di volte la risoluzione strumentale, a conferma che la misura è dominata dalle fluttuazioni del tempo di reazione dello sperimentatore) e sono anche sistematicamente sottostimate rispetto al valore *vero*, che è 1.76 s come indicato dalla linea verticale tratteggiata. Le altre tendono, almeno per la maggior parte, a fluttuare significativamente di meno (diciamo qualche centesimo di secondo) intorno al valore vero—esattamente come ci aspettavamo.



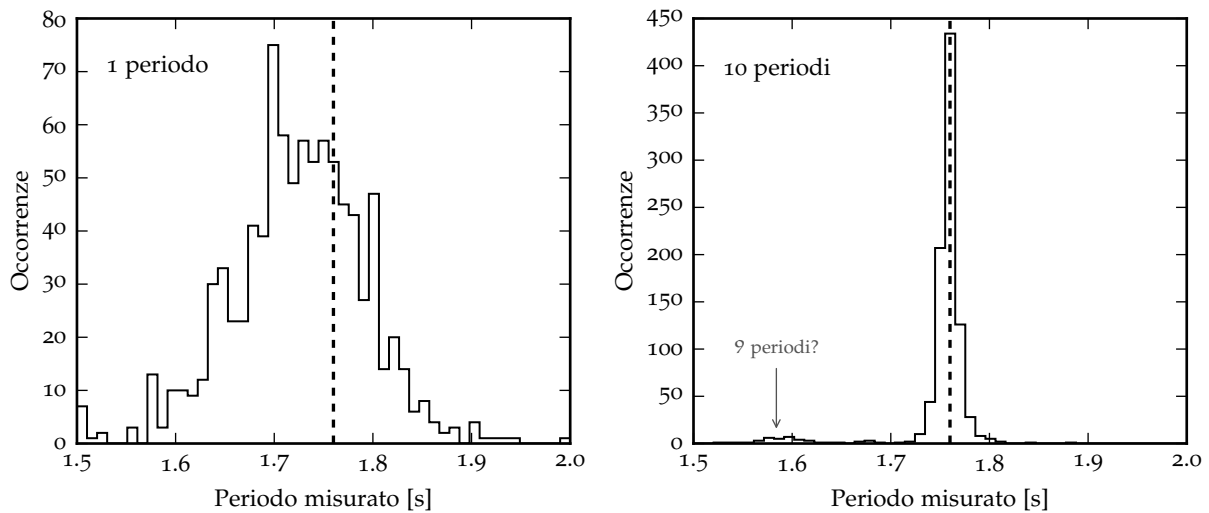


FIGURA 1.2. Distribuzione dei valori del periodo di un moto armonico di un cerchietto proiettato su uno schermo, ottenuti misurando direttamente un periodo singolo (a sinistra) e dividendo per 10 la misura di 10 periodi (a destra). Nel secondo caso l'effetto delle fluttuazioni del tempo di reazione umano è significativamente ridotto (e, incidentalmente, appare anche un piccolo picco secondario a  $9/10$  del valore vero, dovuto evidentemente ad un banale errore di conteggio).

Questo semplice esempio va estrapolato con un minimo di cura, e caso per caso, a situazioni diverse da quella descritta. In particolare se dovessimo misurare in questo modo il periodo di un pendolo, il cui moto non è armonico per angoli sufficientemente grandi, dovremmo fare attenzione ai possibili effetti dello smorzamento, che potrebbe provocare una diminuzione apprezzabile del periodo su tempi abbastanza lunghi. Purtroppo l'idea di base è buona ed utilizzabile in pratica.

Concludiamo notando come, nell'istogramma di destra in figura 1.2 appaia un piccolo picco secondario, contenente poche decine di misurazioni, a sinistra di quello principale. La posizione del picco è intorno a 1.58 s, che corrisponde con buona precisione a  $9/10$  del valore vero 1.76 s. Si tratta banalmente di un errore di conteggio (alcuni studenti hanno misurato il periodo corrispondente a 9 periodi anziché 10), che ci ricorda come sia sempre importante fare attenzione, perché di fronte agli errori (non intesi come incertezze) tutta la nostra teoria crolla.

## 1.5 PROPAGAZIONE DELL'ERRORE MASSIMO

Adesso che abbiamo nel nostro bagaglio di conoscenze il concetto di errore massimo e sappiamo come stimarlo (per lo meno nelle situazioni più semplici) per le misure dirette, siamo pronti ad affrontare il problema generale di come l'errore massimo si propaghi nelle misure indirette.

Dato un certo numero di grandezze *indipendenti*<sup>4</sup>  $x, y, z \dots$  (con errori massimi  $\Delta x, \Delta y, \Delta z \dots$ ) ed una funzione  $f(x, y, z \dots)$  delle grandezze stesse, vorremmo scrivere il risultato della nostra misura indiretta come  $f = \hat{f} \pm \Delta f$  in modo da essere certi che l'intervallo  $[\hat{f} - \Delta f, \hat{f} + \Delta f]$  contenga il misurando  $f$ —così come siamo certi che l'intervallo  $[\hat{x} - \Delta x, \hat{x} + \Delta x]$  contenga il valore di  $x$  e così via. È ragionevole prendere come valore centrale della misura indiretta il valore della funzione calcolata in corrispondenza dei valori centrali delle grandezze di partenza

$$\hat{f} = f(\hat{x}, \hat{y}, \hat{z} \dots).$$

Per stimare l'errore associato, invece, dovremmo calcolare i valori minimo e massimo che la funzione  $f(x, y, z \dots)$  assume al variare di  $x, y, z \dots$  nei rispettivi intervalli di incertezza massima—e prenderne la semidispersione. Nel caso generale si tratta di un problema niente affatto banale. Fortunatamente, come

<sup>4</sup> La richiesta dell'indipendenza delle grandezze, che in questo momento non siamo attrezzati per discutere in maniera quantitativa, è di fondamentale importanza per tutta la discussione sviluppata in questa sezione. Si tratta di un aspetto su cui torneremo in seguito ma che è bene tenere a mente sin dall'inizio.

vedremo tra un attimo, nella maggior parte dei casi di interesse pratico la soluzione del corrispondente problema linearizzato fornisce una risposta sufficientemente accurata.

### 1.5.1 Propagazione dell'errore massimo nella somma e nella differenza

Cominciamo con il caso più semplice, ovvero la propagazione dell'errore massimo sulla somma  $S = x + y$  di due grandezze indipendenti affette da errore. I casi estremi sono quelli in cui le incertezze su  $x$  e  $y$  contribuiscono nello stesso verso:

$$\begin{aligned} S_{\max} &= (\hat{x} + \Delta x) + (\hat{y} + \Delta y) = (\hat{x} + \hat{y}) + (\Delta x + \Delta y) \\ S_{\min} &= (\hat{x} - \Delta x) + (\hat{y} - \Delta y) = (\hat{x} + \hat{y}) - (\Delta x + \Delta y). \end{aligned}$$

Ne segue che possiamo scrivere  $S = (\hat{x} + \hat{y}) \pm (\Delta x + \Delta y)$ , ovvero che in una somma gli errori massimi si sommano

$$\Delta S = \frac{S_{\max} - S_{\min}}{2} = \Delta x + \Delta y. \quad (10)$$

Per analogia potremmo allora essere tentati di pensare che in una differenza gli errori si sottraggano. Invece, se proviamo a fare lo stesso esercizio sulla quantità  $D = x - y$

$$\begin{aligned} D_{\max} &= (\hat{x} + \Delta x) - (\hat{y} - \Delta y) = (\hat{x} - \hat{y}) + (\Delta x + \Delta y) \\ D_{\min} &= (\hat{x} - \Delta x) - (\hat{y} + \Delta y) = (\hat{x} - \hat{y}) - (\Delta x + \Delta y) \end{aligned}$$

scopriamo che possiamo scrivere  $D = (\hat{x} - \hat{y}) \pm (\Delta x + \Delta y)$ , ovvero che gli errori massimi si sommano anche in una differenza:

$$\Delta D = \frac{D_{\max} - D_{\min}}{2} = \Delta x + \Delta y. \quad (11)$$

(Se siete veramente tentati di scrivere  $\Delta D = \Delta x - \Delta y$  pensate a cosa succederebbe nel caso  $\Delta x = \Delta y$ . Quando le grandezze misurate sono indipendenti gli errori purtroppo non si sottraggono mai.)

► **ESEMPIO 1.11.** Abbiamo misurato direttamente con un calibro ventesimale le due lunghezze  $a = 2.50 \pm 0.05$  mm e  $b = 2.35 \pm 0.05$  mm e vogliamo propagare l'errore sulla somma  $S$  e sulla differenza  $D$ . Applicando direttamente la (10) e la (11) si ha  $S = 4.85 \pm 0.10$  mm e  $D = 0.15 \pm 0.10$  mm. È interessante notare come l'errore relativo sulla differenza sia del  $\sim 65\%$ , nonostante le due grandezze di partenza siano misurate al  $\sim 2\%$ . Si tratta di un problema tipico quando si sottraggono misure vicine tra loro.

### 1.5.2 Propagazione dell'errore massimo nel prodotto e nel quoziente

Il problema della propagazione dell'errore sulla somma e sulla differenza è peculiare per il fatto che la soluzione esatta si può scrivere in modo compatto. Come abbiamo già detto questo non è vero in generale. Nel caso del prodotto  $P = xy$  (assumendo, senza perdere in generalità che  $x$  ed  $y$  siano positivi) i casi estremi si scrivono come

$$\begin{aligned} P_{\max} &= (\hat{x} + \Delta x)(\hat{y} + \Delta y) = \hat{x}\hat{y} + (\hat{x}\Delta y + \hat{y}\Delta x + \Delta x\Delta y) \\ P_{\min} &= (\hat{x} - \Delta x)(\hat{y} - \Delta y) = \hat{x}\hat{y} - (\hat{x}\Delta y + \hat{y}\Delta x - \Delta x\Delta y) \end{aligned}$$

e la novità rispetto al caso precedente è che adesso appaiono termini contenenti il prodotto  $\Delta x\Delta y$  degli errori sulle due grandezze—che con un leggero abuso di notazione potremmo chiamare infinitesimi di ordine superiore. Nella situazione (tipica, ma da verificare caso per caso) in cui gli errori relativi sono piccoli, cioè  $\Delta x \ll |\hat{x}|$  e  $\Delta y \ll |\hat{y}|$ , questi termini possono essere trascurati per ottenere la relazione approssimata  $P = \hat{x}\hat{y} \pm (\hat{x}\Delta y + \hat{y}\Delta x)$ , ovvero:

$$\Delta P = \frac{P_{\max} - P_{\min}}{2} \approx |\hat{x}| \Delta y + |\hat{y}| \Delta x \quad (12)$$

(abbiamo ripristinato i valori assoluti, che sono necessari nel caso generale in cui i valori centrali di  $x$  o  $y$  sono negativi per garantire che  $\Delta P$  sia positivo). Per coloro che si sentissero insoddisfatti da questa approssimazione apparentemente brutale, vale la pena ricordare che  $\Delta P$  rappresenta un errore di misura, e come tale non si scrive con più di due cifre significative—per cui conoscerlo al livello di qualche % è tipicamente più che sufficiente. In molti casi i termini di tipo  $\Delta x \Delta y$  sono molto più piccoli della prima cifra affetta da errore e costituiscono essenzialmente rumore numerico, senza alcun contenuto reale di informazione, come illustrato nell'esempio 1.12.

► **ESEMPIO 1.12.** Vogliamo propagare l'errore sul prodotto  $P = ab$  delle due grandezze dell'esempio 1.11. I due termini della (12) sono  $\hat{a}\Delta b \approx 0.12 \text{ mm}^2$  e  $\hat{b}\Delta a \approx 0.12 \text{ mm}^2$ ; per confronto, il termine che abbiamo trascurato è  $\Delta a \Delta b \approx 0.0025 \text{ mm}^2$ , ossia circa 50 volte più piccolo. Possiamo dunque scrivere  $P = 5.87 \pm 0.24 \text{ mm}^2$ . (Notiamo esplicitamente che il termine  $\Delta a \Delta b$  non influisce sulle cifre significative con cui abbiamo scritto la misura.)

La propagazione dell'errore massimo sul quoziente  $Q = x/y$  è simile e, assumendo come prima che  $x$  ed  $y$  siano positivi, i casi estremi si scrivono come

$$Q_{\max} = (\hat{x} + \Delta x)/(\hat{y} - \Delta y)$$

$$Q_{\min} = (\hat{x} - \Delta x)/(\hat{y} + \Delta y)$$

Analizziamo in dettaglio  $Q_{\max}$  e, come per il prodotto, trascuriamo i termini in cui gli errori appaiono al quadrato

$$Q_{\max} = \frac{\hat{x} + \Delta x}{\hat{y} - \Delta y} = \frac{\hat{x} + \Delta x}{\hat{y} - \Delta y} \times \frac{\hat{y} + \Delta y}{\hat{y} + \Delta y} = \frac{\hat{x}\hat{y} + \hat{x}\Delta y + \hat{y}\Delta x + \Delta x\Delta y}{\hat{y}^2 - (\Delta y)^2} \approx \frac{\hat{x}}{\hat{y}} + \frac{\hat{x}\Delta y + \hat{y}\Delta x}{\hat{y}^2}.$$

Questo (e l'espressione analoga per  $Q_{\min}$ ) portano alla formula approssimata (in cui, di nuovo, abbiamo ripristinato i valori assoluti)

$$\Delta Q = \frac{Q_{\max} - Q_{\min}}{2} \approx \frac{|\hat{x}| \Delta y + |\hat{y}| \Delta x}{\hat{y}^2}. \quad (13)$$

Una conseguenza interessante dell'espressione appena scritta è che, nel caso in cui  $\Delta y = 0$ , cioè stiamo dividendo per un numero senza errore, la (13) si semplifica in  $\Delta Q = \Delta x/\hat{y}$ . Questo è il principio alla base della possibilità di fare misure (indirette) con un errore massimo inferiore alla risoluzione strumentale, come già illustrato nella sezione 1.4.4.

Notiamo anche che, quando si propagano gli errori su un prodotto o un quoziente è comodo lavorare con gli errori relativi, poiché, nella nostra notazione

$$\frac{\Delta P}{|\hat{P}|} = \frac{|\hat{x}| \Delta y + |\hat{y}| \Delta x}{|\hat{x}| |\hat{y}|} = \frac{\Delta x}{|\hat{x}|} + \frac{\Delta y}{|\hat{y}|} \quad (14)$$

$$\frac{\Delta Q}{|\hat{Q}|} = \frac{|\hat{x}| \Delta y + |\hat{y}| \Delta x}{\hat{y}^2} \bigg/ \frac{|\hat{x}|}{|\hat{y}|} = \frac{|\hat{x}| \Delta y + |\hat{y}| \Delta x}{|\hat{x}| |\hat{y}|} = \frac{\Delta x}{|\hat{x}|} + \frac{\Delta y}{|\hat{y}|}, \quad (15)$$

ovvero nei prodotti e nei quozienti gli errori massimi relativi si sommano. Vedremo tra un attimo che le equazioni scritte sopra si possono generalizzare ad un numero arbitrario di prodotti e quozienti di grandezze elevate ad un esponente generico.

### 1.5.3 Propagazione dell'errore massimo nel caso generale

Generalizziamo quanto detto fino ad ora per le quattro operazioni elementari, cominciando dal caso di una funzione generica  $f(x)$  di una sola grandezza  $x$  (misurata) con errore massimo associato  $\Delta x$ . Possiamo sviluppare  $f$  in serie di Taylor al prim'ordine attorno al punto  $\hat{x}$

$$f(\hat{x} \pm \Delta x) \approx f(\hat{x}) \pm \frac{df(\hat{x})}{dx} \Delta x$$

da cui si ottiene banalmente

$$\Delta f = \frac{f_{\max} - f_{\min}}{2} \approx \left| \frac{df(\hat{x})}{dx} \right| \Delta x. \quad (16)$$

Il simbolo  $\approx$  ci ricorda che la (16) è una relazione che vale approssimativamente nell'ipotesi in cui i termini di ordine superiore nello sviluppo di Taylor siano trascurabili. Geometricamente questo equivale a richiedere, come mostrato in figura 1.3, che la retta tangente ad  $f$  nel punto  $\hat{x}$  sia una buona approssimazione di  $f$  nell'intervallo  $[\hat{x} - \Delta x, \hat{x} + \Delta x]$ .

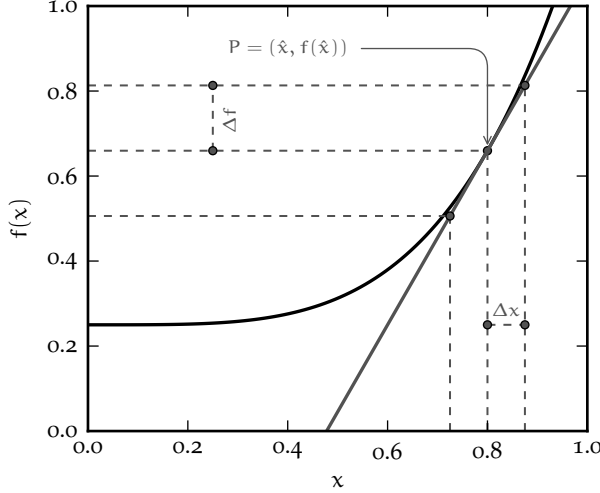


FIGURA 1.3. Rappresentazione grafica della linearizzazione della propagazione dell'errore su una generica funzione di una variabile (16). La derivata della funzione  $f(x)$  nel punto  $\hat{x}$  coincide con il coefficiente angolare della retta tangente ad  $f(x)$  e passante per il punto  $P = (\hat{x}, f(\hat{x}))$  e l'approssimazione lineare è tanto più accurata quanto più la tangente è vicina alla funzione nell'intervallo  $[\hat{x} - \Delta x, \hat{x} + \Delta x]$ .

Come detto in precedenza questa assunzione, che dipende sia dal valore di  $\Delta x$  che dalla forma della funzione  $f(x)$  in un intorno di  $\hat{x}$ , va verificata caso per caso. (Un caso banale in cui in nostro sviluppo in serie troncato al prim'ordine non funziona è quello in cui  $f'(\hat{x}) = 0$ , come illustrato nell'esempio 1.14.)

I risultati precedenti si estendono banalmente (modulo uno sviluppo di Taylor al prim'ordine per una funzione di più variabili) al caso generale di una funzione  $f(x, y, z \dots)$  delle grandezze  $x, y, z \dots$

$$\Delta f \approx \left| \frac{\partial f(\hat{x}, \hat{y}, \hat{z} \dots)}{\partial x} \right| \Delta x + \left| \frac{\partial f(\hat{x}, \hat{y}, \hat{z} \dots)}{\partial y} \right| \Delta y + \left| \frac{\partial f(\hat{x}, \hat{y}, \hat{z} \dots)}{\partial z} \right| \Delta z + \dots \quad (17)$$

Il lettore può verificare facilmente che la (17) permette di derivare le formule che abbiamo ricavato mediante calcolo diretto per la propagazione dell'errore massimo nelle quattro operazioni elementari. Essa permette anche di ricavare una formula compatta per la propagazione dell'errore massimo relativo su una grandezza del tipo  $G(x, y, z \dots) = x^\alpha y^\beta z^\gamma \dots$

$$\frac{\Delta G}{|G|} = \frac{|\alpha \hat{x}^{\alpha-1} \hat{y}^\beta \hat{z}^\gamma| \Delta x + |\hat{x}^\alpha \beta \hat{y}^{\beta-1} \hat{z}^\gamma| \Delta y + |\hat{x}^\alpha \hat{y}^\beta \gamma \hat{z}^{\gamma-1}| \Delta z \dots}{|\hat{x}^\alpha| |\hat{y}^\beta| |\hat{z}^\gamma| \dots} = |\alpha| \frac{\Delta x}{|\hat{x}|} + |\beta| \frac{\Delta y}{|\hat{y}|} + |\gamma| \frac{\Delta z}{|\hat{z}|} + \dots \quad (18)$$

Quando cioè si ha a che fare con prodotti e quozienti di grandezze elevate ad un esponente generico, gli errori relativi si sommano pesati con il modulo del proprio esponente.

Per quanto (apparentemente) banale, l'esempio 1.15 ha in sé gli ingredienti fondamentali che caratterizzano una misura di laboratorio: un fenomeno fisico che può essere studiato quantitativamente (il moto del pendolo) ed un modello teorico—nel nostro caso la (19)—di cui vogliamo stimare un parametro ( $g$ ) mediante un processo di riduzione dei dati raccolti tramite misure dirette.

Il lettore più attento avrà notato che non abbiamo specificato le condizioni in cui il nostro modello è applicabile, cioè non ci siamo chiesti esplicitamente che cosa significhi lavorare in approssimazione di piccole oscillazioni. Nel caso del pendolo semplice la condizione di piccole oscillazioni viene spesso espressa (detto  $\theta_0$  l'ampiezza massima di oscillazione) nelle forme  $\theta_0 \ll 1$  o  $\sin \theta_0 \approx \theta_0$ . Il concetto di incertezza di misura ci permette di precisare quantitativamente questa condizione: un modello è valido quando fornisce una descrizione accurata della realtà entro l'errore di misura. Ne segue che la bontà o meno di un modello (nel senso dell'accordo con una misura) non si può definire astrattamente, ma

► **ESEMPIO 1.13.** Dato un angolo (misurato)  $\theta = \hat{\theta} \pm \Delta\theta = (30 \pm 1)^\circ$  si vuole propagare l'errore sulla funzione  $\sin \theta$ . Possiamo utilizzare direttamente la (16), che nel caso specifico fornisce:

$$\Delta(\sin \theta) \approx \left| \frac{d \sin(\hat{\theta})}{d\theta} \right| \Delta\theta = |\cos \hat{\theta}| \Delta\theta.$$

Con l'accortezza necessaria di convertire  $\Delta\theta$  in radianti (le derivate delle funzioni trigonometriche sono quelle note solo se gli angoli sono misurati in radianti) possiamo scrivere  $\sin \theta = 0.500 \pm 0.015$ .

► **ESEMPIO 1.14.** Dato un angolo (misurato)  $\theta = \hat{\theta} \pm \Delta\theta = (90 \pm 5)^\circ$  si vuole propagare l'errore sulla funzione  $\sin \theta$ . In questo caso la derivata prima si annulla nel punto  $\hat{\theta}$  per cui il nostro formalismo è inutilizzabile (si tratta di un caso in cui, per ovvie ragioni, i termini di ordine superiore nello sviluppo di Taylor non sono trascurabili). In situazioni come queste possiamo stimare esplicitamente i casi estremi:

$$\sin(\hat{\theta} - \Delta\theta) = \sin(\hat{\theta} + \Delta\theta) \approx 0.996$$

per cui possiamo scrivere  $\sin \theta = 1.000 \pm 0.004$ . Alternativamente possiamo estendere lo sviluppo in serie di Taylor all'ordine successivo

$$\Delta(\sin \theta) \approx \frac{1}{2} \left| \frac{d^2 \sin(\hat{\theta})}{d\theta^2} \right| (\Delta\theta)^2 = \frac{1}{2} |\sin \hat{\theta}| (\Delta\theta)^2 \approx 0.004$$

(che è in accordo con il risultato ottenuto per calcolo diretto.) Notiamo anche che questo esempio è peculiare per un altro motivo: dato che  $\sin \theta \leq 1$ , ci troviamo in un caso in cui l'intervallo di incertezza sulla nostra grandezza derivata è asimmetrico. Più correttamente dovremmo allora scrivere  $\sin \theta = 1.000^{+0.000}_{-0.004}$ .

dipende dal contesto specifico in cui la misura stessa è effettuata. (Per cui va da sé che affermazioni tipo "siamo in regime di piccole oscillazioni per  $\theta_0 < 5^\circ$ " non hanno alcun senso.) Torneremo brevemente sulla questione nella sezione 1.6, quando discuteremo gli errori sistematici.

#### 1.5.4 Scelta delle grandezze per le misure dirette ed indipendenza

Abbiamo avuto modo di sottolineare più volte che la (17) vale solo se le grandezze di partenza sono indipendenti, ma non abbiamo ancora precisato la questione in modo quantitativo. L'esempio che segue illustra il concetto con gli strumenti limitati che abbiamo a disposizione a questo punto ed illustra anche come, ove ciò sia possibile, una scelta oculata delle variabili indipendenti da misurare permetta di minimizzare l'errore propagato su una grandezza derivata—che è un elemento di ovvia rilevanza pratica nella progettazione di un esperimento.

Supponiamo allora di voler misurare la superficie  $A$  di una corona circolare come quella mostrata nella figura 1.4. Il valore della superficie cercata può essere scritta in funzione di una coppia qualsiasi scelta tra le tre variabili indipendenti  $r$  (raggio interno),  $R$  (raggio esterno) e  $d$  (spessore della corona). In particolare possiamo scegliere di misurare  $r$  ed  $R$ , come mostrato nel pannello a), oppure  $R$  e  $d$ , come mostrato nel pannello d).

Supponiamo dunque di misurare il raggio interno ed il raggio esterno della corona e di ottenere i valori  $R = \hat{R} \pm \Delta R = 15.0 \pm 0.1$  cm e  $r = \hat{r} \pm \Delta r = 6.0 \pm 0.1$  cm. Dalla relazione

$$A = \pi(R^2 - r^2),$$

possiamo calcolare il valore centrale della misura  $\hat{A} = 594$  cm<sup>2</sup>, e l'errore si scrive utilizzando la (17)

$$\Delta A = \left| \frac{\partial A(\hat{R}, \hat{r})}{\partial R} \right| \Delta R + \left| \frac{\partial A(\hat{R}, \hat{r})}{\partial r} \right| \Delta r = 2\pi(\hat{R}\Delta R + \hat{r}\Delta r) \approx 13$$
 cm<sup>2</sup>,

► **ESEMPIO 1.15.** Vogliamo misurare l'accelerazione di gravità  $g$  al livello del suolo utilizzando un pendolo semplice e sfruttando la relazione che lega il periodo alla lunghezza del pendolo stesso in approssimazione di piccole oscillazioni

$$T = 2\pi\sqrt{\frac{l}{g}}. \quad (19)$$

Misuriamo dunque direttamente  $l = 1.005 \pm 0.006$  m e  $T = 2.02 \pm 0.01$  s. La relazione scritta sopra si può invertire per ottenere un'espressione per  $g$  in funzione di  $l$  e  $T$

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

da cui possiamo calcolare il valore centrale  $\hat{g} = 9.72$  m s<sup>-2</sup>. Non ci rimane che propagare l'errore su  $g$  utilizzando la (17)—cosa che facciamo pedissequamente, calcolando le derivate

$$\frac{\partial g}{\partial l} = \frac{4\pi^2}{T^2} \quad \text{e} \quad \frac{\partial g}{\partial T} = -\frac{8\pi^2 l}{T^3},$$

e mettendo tutto insieme

$$\Delta g = \left| \frac{dg(\hat{l}, \hat{T})}{d\hat{l}} \right| \Delta l + \left| \frac{dg(\hat{l}, \hat{T})}{d\hat{T}} \right| \Delta T = \frac{4\pi^2}{\hat{T}^2} \left( \Delta l + 2\hat{l} \frac{\Delta T}{\hat{T}} \right) \approx 0.15 \text{ m s}^{-2}.$$

Scriveremo dunque il risultato come  $g = 9.72 \pm 0.15$  m s<sup>-2</sup>, che è compatibile con il valore noto di 9.81 m s<sup>-2</sup>. Notiamo esplicitamente che, dividendo per il valore centrale, otteniamo

$$\frac{\Delta g}{\hat{g}} = \left( \frac{\Delta l}{\hat{l}} + 2 \frac{\Delta T}{\hat{T}} \right).$$

che, come è lecito aspettarsi, è in accordo con la (18).

per cui scriveremo la nostra misura come  $A = 594 \pm 13$  cm<sup>2</sup>. Si può verificare per sostituzione diretta che l'errore massimo propagato coincide (approssimativamente) con la semidispersione delle superfici massima e minima mostrate in figura 1.4 (pannelli b e c):

$$A_{\max} = \pi \left[ (\hat{R} + \Delta R)^2 - (\hat{r} - \Delta r)^2 \right] \approx 607 \text{ cm}^2$$

$$A_{\min} = \pi \left[ (\hat{R} - \Delta R)^2 - (\hat{r} + \Delta r)^2 \right] \approx 581 \text{ cm}^2.$$

Fino a qui niente di nuovo, ma proviamo adesso a ripetere l'esercizio nell'ipotesi in cui misuriamo il raggio esterno  $R = \hat{R} \pm \Delta R = 15.0 \pm 0.1$  cm e lo spessore della corona  $d = \hat{d} \pm \Delta d = 9.0 \pm 0.1$  cm. Prima di procedere fermiamoci un attimo ad osservare la figura 1.4: i valori centrali di  $R$  e  $d$  corrispondono allo stesso oggetto fisico della situazione precedente, ma le configurazioni di area massima e minima definite dagli involucri degli errori sono diverse nei due casi.

Potremmo pensare di trarre vantaggio dalla propagazione degli errori che abbiamo appena calcolato ricavando il valore del raggio interno  $r$  a partire da  $R$  e  $d$  (con l'opportuna propagazione dell'errore massimo)

$$r = \hat{r} \pm \Delta r = 6.0 \pm 0.2 \text{ cm}$$

e riducendoci al caso precedente. Così facendo otterremo un errore propagato pari a  $\approx 17$  cm<sup>2</sup>, ma se calcoliamo esplicitamente le superfici massima e minima in figura 1.4 e) ed f), si ha

$$A_{\max} = \pi \left[ (\hat{R} + \Delta R)^2 - (\hat{R} + \Delta R - \hat{d} - \Delta d)^2 \right] \approx 603 \text{ cm}^2$$

$$A_{\min} = \pi \left[ (\hat{R} - \Delta R)^2 - (\hat{R} - \Delta R - \hat{d} + \Delta d)^2 \right] \approx 584 \text{ cm}^2,$$

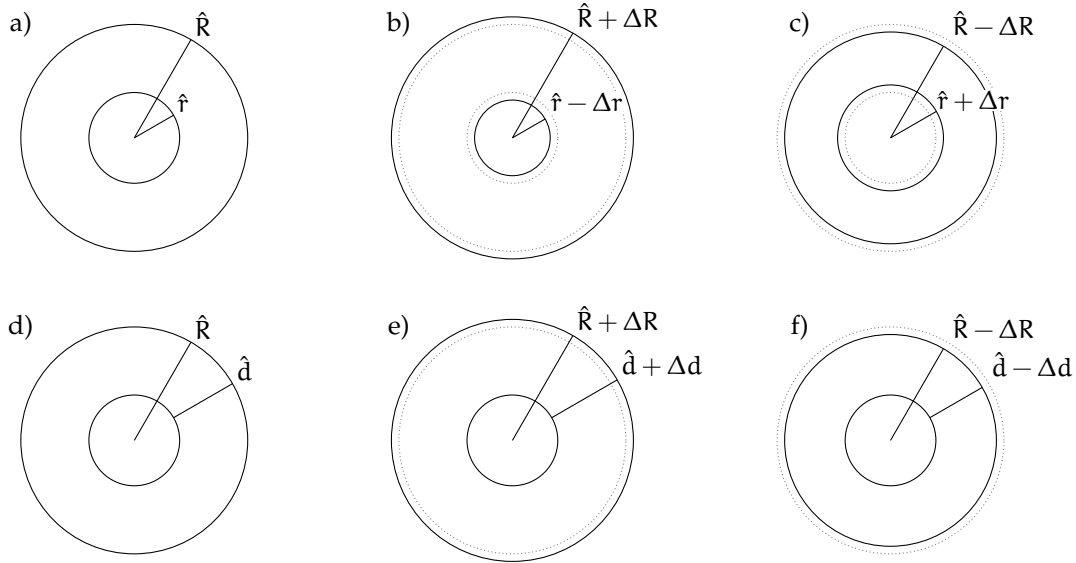


FIGURA 1.4. Illustrazione grafica della propagazione dell'errore massimo sulla misura della superficie di una corona circolare. Se misuriamo il raggio esterno  $R$  ed il raggio interno  $r$ , la configurazione di area massima si ha quando  $R$  è massimo e  $r$  è minimo (entro i rispettivi errori di misura), mentre quella di area minima si ha quando  $R$  è minimo e  $r$  è massimo. Viceversa, se misuriamo  $R$  e  $d$ , la configurazione di area massima (minima) si ha quando entrambe le grandezze sono massime (minime).

la cui semidispersione è  $\approx 9 \text{ cm}^2$ . Chiaramente qualcosa non va, ed il problema è che  $R$  ed  $r$  in questo caso non sono indipendenti poiché  $r$  è stato calcolato a partire da  $R$ .

Il modo corretto di inquadrare il problema in questo secondo scenario è invece quello di scrivere la superficie in funzione delle due grandezze (misurate) indipendenti

$$A = \pi [R^2 - (R - d)^2] = \pi(R^2 - R^2 + d^2 + 2Rd) = \pi(2Rd - d^2)$$

e propagare l'errore di conseguenza

$$\Delta A = \left| \frac{\partial A(\hat{R}, \hat{d})}{\partial R} \right| \Delta R + \left| \frac{\partial A(\hat{R}, \hat{d})}{\partial d} \right| \Delta d = 2\pi [\hat{d}\Delta R + (\hat{R} - \hat{d})\Delta d] \approx 9 \text{ cm}^2.$$

L'errore è, in questo caso, più piccolo di quello stimato nello scenario precedente, il che ci dice che, nel caso specifico, a parità di risoluzione strumentale, è più conveniente misurare raggio esterno e spessore piuttosto che raggio esterno e raggio interno.

## 1.6 ERRORI SISTEMATICI

Nella sezione 1.1 abbiamo introdotto la distinzione tra le nozioni di precisione ed accuratezza. Torniamo adesso su una domanda che abbiamo toccato brevemente più volte, ovverosia: immaginando di poter migliorare a piacimento la precisione di una misura, possiamo affermare che siamo in grado di migliorare a piacimento anche la sua accuratezza? In altre parole, possiamo avvicinarci arbitrariamente al valore del misurando semplicemente facendo in modo di ridurre le fluttuazioni statistiche della misura (dove possibile) e migliorando la risoluzione dello strumento? La risposta, in generale, è no, ed il motivo di fondo è l'esistenza degli *errori sistematici*.

Gli errori sistematici possono essere causati da svariati fattori, tra cui: problemi di calibrazione degli strumenti di misura, errori di lettura o utilizzo improprio degli strumenti stessi, condizioni ambientali o fattori esterni non completamente sotto controllo, errori ed incertezze di modellizzazione nelle misure indirette. Corrispondentemente, la zoologia degli errori sistematici è molto variegata e in questa sezione ci limiteremo a fare alcune considerazioni generali riesaminando criticamente alcuni degli esempi visti in precedenza.

## 1.6.1 Errori sistematici strumentali

Spesso gli errori sistematici hanno origine strumentale, ad esempio una non perfetta calibrazione dell'apparato di misura. Questo è particolarmente vero negli esperimenti di frontiera, in cui gli strumenti sono estremamente complessi e sono utilizzati al limite delle proprie possibilità.

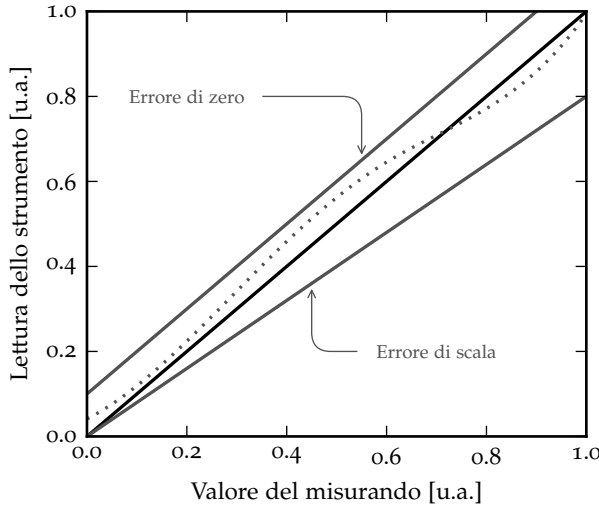


FIGURA 1.5. Esempio illustrativo della funzione di risposta di uno strumento generico. La diagonale indica la situazione in cui la lettura è identica al valore (incognito) del misurando su tutto l'intervallo operativo (cioè lo strumento è perfettamente calibrato). In pratica la risposta non è mai ideale, e si possono dare errori di zero, errori di scala o, più in generale deviazioni dalla diagonale di forma arbitraria. La funzione fondamentale della calibrazione di uno strumento è di fare in modo che queste deviazioni siano il più piccole possibili e sotto controllo.

Ogni strumento è caratterizzato da una funzione di risposta, che rappresenta essenzialmente il valore di lettura dello strumento stesso in funzione del valore (incognito) del misurando (figura 1.5). Idealmente vorremmo che la risposta dello strumento fosse l'identità su tutto l'intervallo operativo. In pratica essa non può essere misurata con precisione infinita, ed il massimo che si può fare, attraverso la calibrazione, è cercare di avvicinarsi il più possibile alla condizione ideale e quantificare più o meno precisamente le deviazioni residue. Si parla allora di *errore di zero* quando la lettura di uno strumento è traslata sistematicamente di una costante additiva (o *offset*) rispetto al valore del misurando. Si parla viceversa di un *errore di scala* quando il rapporto tra lettura e misurando è pari ad una costante moltiplicativa diversa da 1. Nel caso più generale la risposta di uno strumento può essere affetta contemporaneamente da errori di zero ed errori di scala, e questi ultimi possono variare in grandezza all'interno dell'intervallo di misura dello strumento stesso.

## 1.6.2 Incertezze di modellizzazione

Abbiamo già avuto modo di dire che l'utilizzo di un *modello* inadeguato può essere fonte di errori sistematici nella stima dei parametri del modello stesso. Riesaminiamo per un attimo l'esempio 1.15—ossia la stima del valore dell'accelerazione di gravità  $g$  attraverso la misura del periodo di un pendolo semplice. Avevamo lasciato in sospeso la questione dell'ampiezza di oscillazione oltre la quale il nostro modello semplicistico cessava di essere appropriato.

Partiamo dunque dallo sviluppo in serie di Taylor dell'espressione generale per il periodo del pendolo semplice in funzione dell'ampiezza massima di oscillazione  $\theta_0$ :

$$T = T_0 \left( 1 + \frac{1}{16} \theta_0^2 + \frac{11}{3072} \theta_0^4 + \dots \right) \quad \text{con} \quad T_0 = 2\pi \sqrt{\frac{l}{g}}. \quad (20)$$

Il primo termine che abbiamo trascurato è  $T_0 \theta_0^2 / 16$ , e ci chiediamo adesso se eravamo legittimati a farlo oppure no. La risposta è implicita nella domanda: la condizione è che ciò che abbiamo trascurato fosse molto più piccolo dell'errore di misura

$$\frac{T_0 \theta_0^2}{16} \ll \Delta T,$$



► **ESEMPIO 1.16.** Riesaminiamo brevemente l'esempio 1.7 del nostro termometro digitale che, immerso in acqua, fornisce una lettura  $T_1 = 34.5^\circ\text{C}$  (cui associamo un errore di  $0.1^\circ\text{C}$ ). Acquistiamo adesso un secondo termometro identico, lo immergiamo nello stesso contenitore, e notiamo che quest'ultimo segna  $T_2 = 33.2^\circ\text{C}$  (se l'esempio vi pare pretestuoso, comprate due termometri digitali a basso prezzo e provate voi stessi: non mancherete di rimanere sorpresi). Aspettiamo qualche ora e notiamo che i due termometri segnano  $T_1 = 26.5^\circ\text{C}$  e  $T_2 = 25.2^\circ\text{C}$ —notate che la differenza è rimasta invariata ( $1.3^\circ\text{C}$ ), ed è molto più grande dell'errore di misura nominale. In queste condizioni è ragionevole concludere che i due termometri siano in grado di apprezzare variazioni di  $0.1^\circ\text{C}$  (nel senso della risoluzione strumentale) ma che allo stesso tempo abbiano un errore di zero (incognito, perché a questo livello non possiamo dire quale dei due sia più vicino al misurando) dell'ordine  $1.3^\circ\text{C}$ . (Se avessimo più termometri a disposizione potremmo costringere meglio l'entità di questo errore di zero.) Scriveremo allora l'ultima misura fornita (ad esempio) dal primo termometro come  $T_1 = 26.5 \pm 0.1 \pm 1.3(\text{sys})^\circ\text{C}$ , indicando separatamente il contributo dell'errore sistematico di zero.

► **ESEMPIO 1.17.** Il coefficiente di dilatazione termica lineare dell'acciaio (materiale con cui sono tipicamente realizzati i metri a nastro metallici) è dell'ordine di  $\lambda \approx 1.7 \times 10^{-5}^\circ\text{C}^{-1}$ . Questo significa che, passando da  $T_1 = 0^\circ\text{C}$  a  $T_2 = 30^\circ\text{C}$  la lunghezza reale di un metro a nastro di lunghezza nominale  $L = 2\text{ m}$  aumenta di un fattore  $\lambda L(T_2 - T_1) \approx 1\text{ mm}$ —che è pari alla risoluzione del metro stesso. (Nel caso di un calibro cinquantiesimale l'effetto può essere, in termini relativi, significativamente più importante.) Una contrazione o dilatazione di questo tipo introduce chiaramente un errore di scala. Questo dovrebbe darci alcuni spunti di riflessione: a che temperatura è stato calibrato il metro (o il calibro)? Qual è il coefficiente di dilatazione termica lineare dell'oggetto che stiamo misurando? E cosa significa misurare la lunghezza di un oggetto se essa non è una grandezza caratteristica dell'oggetto stesso, ma dipende dalla temperatura?

che ci dà una definizione operativa di cosa significhi “piccole oscillazioni” nel contesto della nostra misura:

$$\theta_0 \ll 4\sqrt{\frac{\Delta T}{T}}.$$

Il contenuto dell'equazione che abbiamo appena scritto è chiaro: l'angolo oltre il quale le oscillazioni cessano di essere piccole dipende dalla risoluzione dello strumento che uso. Se l'errore relativo di misura è dell'ordine dell'1% (che è un valore tipico se si usa un cronometro manuale), si cominciano ad apprezzare i termini anarmonici per  $\theta_0 \approx 20^\circ$ ; ma se è di  $10^{-5}$  (che non è impossibile da ottenere con un microcontrollore relativamente poco sofisticato), allora  $\theta_0 \approx 1^\circ$  è più che sufficiente.

Ma torniamo adesso alla questione originaria dell'errore sistematico connesso con l'uso di un modello inadeguato. Se misuriamo abbastanza accuratamente (nel senso dell'accuratezza, non solo della precisione)  $l$  e  $T$  possiamo dire che misureremo altrettanto accuratamente  $g$ ? Sicuramente l'errore propagato secondo la prescrizione dell'esempio 1.15 sarà piccolo se  $\Delta l$  e  $\Delta T$  sono piccoli, ma ad un certo punto entreremo nelle condizioni in cui non sarà più lecito utilizzare la (19) per ricavare  $g$  dalle misure di  $l$  e  $T$  (e, vale la pena ripeterlo ancora una volta, non perché queste ultime non siano accurate).

Ci si aprono dunque due strade: (i) ci assicuriamo che il valore di  $\theta_0$  che scegliamo per eseguire la misura sia abbastanza piccolo da garantire che i termini anarmonici dello sviluppo in serie del periodo siano molto più piccoli dell'errore di misura; oppure (ii) si utilizzano a posteriori tutti i termini della (20) che è necessario usare per evitare errori sistematici.

(L'esempio appena fatto è parzialmente fuorviante per un motivo che vale la pena illustrare brevemente. In una situazione in cui, almeno in linea di principio, si ha a disposizione un modello arbitrariamente accurato, utilizzare un modello palesemente inadeguato non costituisce un errore sistematico, ma semplicemente un *errore*—non nel senso di incertezza, ma nel senso di sbaglio vero e proprio. Vi sono casi, tuttavia, in cui tale modello non è disponibile; ed in quei casi l'inadeguatezza del modello a disposizione costituisce il limite ultimo all'accuratezza della misura.)

### 1.6.3 Identificazione degli effetti sistematici

Identificare e trattare opportunamente tutte le possibili sorgenti di errori sistematici in una misura, specialmente in esperimenti complessi, è spesso il compito più difficile con cui chi fa analisi dati si trova a misurarsi. Una trattazione esaustiva del problema a partire dai principi primi è estremamente difficile, ma in linea generale alcune strategie possibili sono: ripetere la misura con uno o più strumenti indipendenti—possibilmente più accurati; verificare la calibrazione dello strumento utilizzato attraverso la misura di una o più grandezze di riferimento note; verificare la consistenza interna del proprio campione di dati (ad esempio suddividendolo in sottocampioni o intervalli temporali); verificare possibili incongruenze con il modello utilizzato per interpretare i dati.

► **ESEMPIO 1.18.** Supponiamo di prendere i due termometri dell'esempio 1.16 e di immergerli in un (abbondante) misto di acqua e ghiaccio—cioè in un bagno termico a  $0^{\circ}\text{C}$ . Il primo termometro segna  $0.2^{\circ}\text{C}$  ed il secondo  $1.5^{\circ}\text{C}$ . Data questa nuova informazione potremmo essere tentati di assumere che il primo termometro sia più accurato del secondo, in quanto si avvicina di più al nostro valore di riferimento. Potremmo addirittura essere tentati di usare questo procedimento per correggere lo zero di entrambi. (Prima di fare questo, dovremmo cercare di analizzare criticamente il nostro procedimento di taratura, però. Se immergiamo il termometro in un punto diverso del bagno termico, ad esempio, la lettura cambia? E in tal caso quali sono le implicazioni?)

► **ESEMPIO 1.19.** Data una serie di oggetti dello stesso materiale, misuriamo il volume  $V$  e la massa  $m$  per ciascuno di essi. Mettiamo poi i valori ottenuti su di un grafico cartesiano (di  $m$  in funzione di  $V$ ) per stimare la densità  $\rho$  del materiale dalla retta che meglio li approssima—secondo la relazione

$$m = \rho V.$$

Se non vincoliamo la nostra retta a passare per l'origine ci accorgiamo però che l'intercetta stimata è significativamente diversa da 0. In questo caso abbiamo ottime ragioni fisiche per credere che (come previsto dal nostro modello) la massa tenda a 0 quando il volume tende a 0. Il fatto che i nostri dati non si dispongano su una retta passante per l'origine indica che, con ogni probabilità, siamo in presenza di un errore sistematico sulla misura delle masse o dei volumi—o tutti e due.

### 1.6.4 Propagazione degli effetti sistematici

Una caratteristica peculiare degli errori sistematici è costituita dal fatto che, per loro natura, non seguono le regole di propagazione dell'errore massimo. La differenza di base è che gli errori sistematici non sono indipendenti tra misura e misura: se un termometro, ad esempio, ha un problema di zero, tutte le misure saranno troppo grandi o troppo piccole della stessa quantità. Questo porta a situazioni interessanti—e a volte sorprendenti—nella propagazione delle incertezze di questo tipo.

► **ESEMPIO 1.20.** Supponiamo di misurare le temperature  $T_1$  e  $T_2$  ai due estremi di una sbarretta conduttrice e di essere interessati *solo* alla loro differenza—ad esempio per stimare la conducibilità termica del materiale di cui la sbarretta è costituita. Questo è un caso tipico in cui un possibile errore di zero del termometro (assumendo di utilizzare lo stesso termometro per le due misure) è ininfluenza.

## 1.7 IN BREVE...

► Il risultato di una misura è composto da: (i) valore centrale; (ii) incertezza ad esso associata; (iii) unità di misura. Mai omettere alcuna delle tre per nessuna ragione.

► L'incertezza di misura si scrive con 1 o, al massimo, 2 cifre significative. Il valore centrale si scrive, di conseguenza, con un numero di cifre decimali consistente con quelle dell'incertezza.

- Controllare sempre le dimensioni fisiche delle equazioni che scriviamo. (E l'analisi dimensionale è anche uno strumento predittivo.)
- La propagazione dell'errore massimo per misure indipendenti si fa seguendo la (17)—e la parola “indipendenti” nella frase precedente non è superflua.
- L'identificazione e la trattazione degli errori sistematici sono una parte fondamentale del processo di misura ed analisi dei dati.



---

## RAPPRESENTAZIONE DEI DATI

---

In Fisica si ha spesso a che fare con enormi quantità di dati che non sono facilmente interpretabili nella loro forma *grezza*. Rappresentare opportunamente tali dati ed estrarre l'informazione rilevante in una forma intelligibile è uno degli obiettivi fondamentali dell'analisi dei dati. In questo capitolo illustriamo brevemente alcune idee di fondo che svilupperemo ulteriormente nel seguito.

### 2.1 TABELLE

Supponiamo di voler stimare la velocità di un oggetto—vincolato a muoversi su una retta—attraverso la misura della sua posizione ad intervalli di tempo regolari. (Per fissare le idee possiamo pensare di fotografare l'oggetto ad istanti di tempo fissati e di determinarne poi la posizione analizzando le immagini così acquisite. Notiamo esplicitamente che, in questo schema, gli errori sui tempi di misura sono *trascurabili*, in un senso che avremo modo di precisare in seguito, e quindi non rappresentati esplicitamente.) Detto  $n$  il numero di fotogrammi, il nostro insieme di dati consiste in  $n$  coppie ordinate di numeri (con i rispettivi errori), che possono essere efficacemente rappresentate in forma tabellare come mostrato nella tabella 5.

Tempo [s]	Posizione [cm]
1.0	$20.5 \pm 2.5$
2.0	$28.7 \pm 2.5$
3.0	$35.4 \pm 2.5$
4.0	$43.1 \pm 2.5$
5.0	$51.8 \pm 2.5$
6.0	$54.6 \pm 2.5$
7.0	$64.1 \pm 2.5$
8.0	$69.7 \pm 2.5$
9.0	$77.5 \pm 2.5$

TABELLA 5. Esempio di rappresentazione di una serie di dati in forma tabellare. L'intestazione (cioè la prima riga) contiene un nome descrittivo per ogni colonna, con le rispettive unità di misura, di modo che la tabella risulti auto-contenuta. Gli errori sono da intendere come errori massimi. (In questo caso, visto che l'errore sulle posizioni è lo stesso per tutte le righe, avremmo potuto riportarlo una volta sola nell'intestazione oppure nella didascalia.)

L'efficacia della formattazione di una tabella dipende in modo cruciale dal contenuto, ed è argomento difficile da discutere in astratto. È tuttavia buona norma, in generale, dare nomi il più possibile descrittivi alle colonne ed indicare sempre le unità di misura, di modo che la tabella sia per quanto possibile auto-esplicativa, ed il suo contenuto possa essere compreso senza la necessità di informazioni esterne.

### 2.2 GRAFICI DI DISPERSIONE

Un *grafico di dispersione* (il nome Italiano, traduzione dell'Inglese *scatter plot*, non è molto usato)—o *grafico xy*, o semplicemente *grafico*—è un diagramma in cui coppie ordinate di dati (ad esempio misure dirette) sono visualizzate come punti, eventualmente con le rispettive barre d'errore, su un piano cartesiano. Si tratta di una delle più semplici rappresentazioni possibili di una serie di dati—eppure è spesso utile, in pratica, per mettere in evidenza proprietà rilevanti dei dati stessi che non sono immediatamente riconoscibili nella rappresentazione in forma tabellare.

Il grafico mostrato in figura 2.1 corrisponde ai dati nella tabella 5, ed ha tutte le caratteristiche di base di un grafico realizzato correttamente: gli assi cartesiani sono suddivisi ad intervalli regolari in modo che le coordinate dei punti possano essere lette agevolmente; le grandezze rappresentate in ascissa

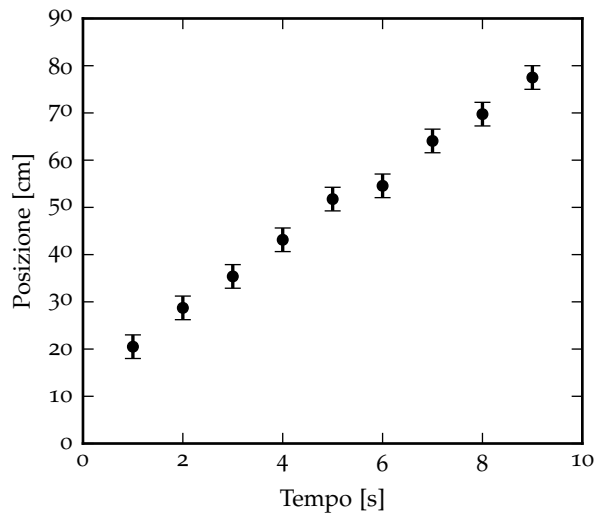


FIGURA 2.1. Esempio di rappresentazione attraverso un grafico di dispersione dei dati contenuti nella tabella 5. Si notino: la suddivisione regolare degli assi cartesiani, le etichette di testo che identificano le variabili rappresentate e le loro unità di misura, e le barre di errore—ove rappresentabili.

ed ordinata sono chiaramente identificate con una etichetta di testo e le rispettive unità di misura (tra parentesi quadre); i dati sono rappresentati come punti e le barre di errore (in questo caso solo sulle ordinate) indicano graficamente l'intervallo di errore della misura.

### 2.2.1 Fit grafici lineari

Una cosa che appare evidente dalla figura 2.1 (senza dubbio più che dalla tabella 5), è che i nostri dati tendono a disporsi su una retta. Avremo modo di precisare in seguito il contenuto quantitativo di questa affermazione, ma se torniamo per un attimo alla Fisica del problema, potremmo essere tentati di concludere che il nostro oggetto si sta muovendo con moto rettilineo uniforme e che il coefficiente angolare della retta che meglio rappresenta i nostri punti altro non è che una misura della velocità  $v_0$ .

Tracciamo allora (con il righello, ad esempio) la retta che meglio ci sembra rappresentare i nostri dati e prendiamo due punti  $P_1 = (x_1, y_1)$  e  $P_2 = (x_2, y_2)$  su tale retta, come indicato in figura 2.2. (La scelta dei punti è largamente arbitraria, ma per minimizzare gli errori relativi sulle grandezze da misurare conviene sceglierli in modo da sfruttare l'intera gamma dinamica dei dati a disposizione.) A questo punto abbiamo tutti gli ingredienti per stimare i due parametri della nostra retta, che scriviamo nella forma  $y = mx + q$ . Il coefficiente angolare è dato dal rapporto

$$m = \frac{y_2 - y_1}{x_2 - x_1}, \quad (21)$$

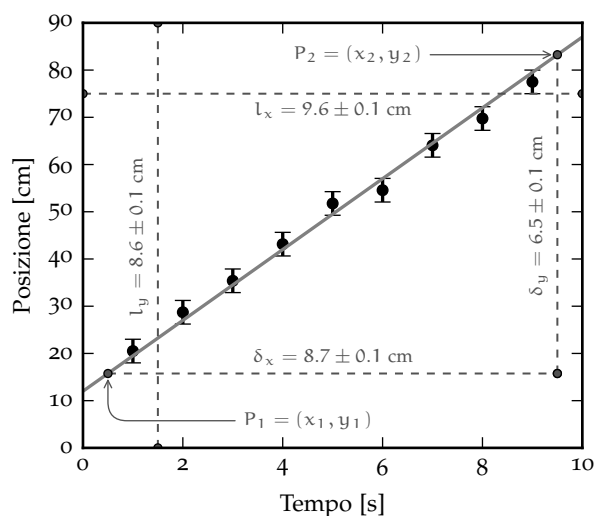


FIGURA 2.2. Esempio di *fit* grafico lineare dei dati contenuti nella tabella 5. Una volta tracciata la retta che meglio ci sembra rappresentare i nostri dati, e presi due punti su tale retta come mostrato in figura, possiamo stimare la velocità del nostro oggetto a partire dal rapporto incrementale. Le annotazioni sul grafico sono state ottenute misurando le lunghezze rilevanti sullo schermo di un computer con un semplice righello, e sono sufficienti per stimare la velocità del nostro oggetto e l'incertezza da associare alla misura, come illustrato passo passo nell'esempio 2.1.

e l'intercetta  $q$  si legge direttamente dal grafico—i.e., coincide con il valore delle ordinate per cui la retta intercetta l'asse  $x = 0$ .

Il procedimento di identificare, all'interno di una famiglia parametrica di funzioni, i valori dei parametri che meglio si adattano ad una serie di dati sperimentali, prende il nome (Inglese) di *fit*. I valori che tali parametri assumono in corrispondenza della soluzione ottimale del problema si dicono valori di *best fit*. Vi sono metodi analitici di *fit* ben definiti e statisticamente fondati che avremo occasione di discutere molto più in dettaglio nel seguito. La variante puramente grafica che abbiamo appena descritto, e che si dice solitamente *fit grafico*, presenta invece numerosi aspetti di arbitrarietà. Il concetto di “retta che meglio ci sembra rappresentare i nostri dati” non è ben definito e sperimentatori diversi troveranno in generale rette diverse partendo dalla stessa serie di dati. (Questo problema può essere mitigato seguendo le regole dettate dal buon senso—ad esempio cercare di far passare la retta di *fit* attraverso tutte le barre d'errore e cercando di fare in modo che il numero di punti sperimentali che stanno sotto la retta sia più o meno uguale al numero di punti che stanno sopra.) Anche la scelta dei punti  $P_1$  e  $P_2$  sulla retta è arbitraria—e può influenzare la stima degli errori sui parametri. Purtroppo si tratta di un esercizio utile (che, se eseguito con cura, fornisce tipicamente risultati ragionevoli) e di un primo esempio significativo di riduzione dei dati, in cui abbiamo riassunto il contenuto quantitativo di una serie di dati in due parametri fisicamente significativi: la velocità e la posizione iniziale del nostro oggetto.

► **ESEMPIO 2.1.** Stimiamo graficamente il coefficiente angolare della retta di *fit* in figura 2.2. Abbiamo a disposizione le lunghezze (misurate con il righello)  $\delta_y$  e  $\delta_x$ , e per convertirle nelle unità di misura appropriate dobbiamo dividerle per le lunghezze (di nuovo, misurate con il righello) dell'unità di misura  $u_y$  e  $u_x$  sull'asse delle ordinate e su quello delle ascisse, rispettivamente. A tal proposito, per minimizzare gli errori relativi, è conveniente misurare le lunghezze complete dei due assi  $l_x$  ed  $l_y$  e poi dividere per il numero totale di divisioni

$$u_x = \frac{l_x}{10} = 0.96 \pm 0.01 \text{ cm} \quad u_y = \frac{l_y}{90} = 0.0956 \pm 0.001 \text{ cm}.$$

Il coefficiente angolare cercato è dato da

$$v_0 = \frac{\delta_y/u_y}{\delta_x/u_x}$$

e l'errore relativo massimo (trattandosi di una serie di prodotti e quozienti) non è altro che la somma degli errori relativi delle 4 grandezze in gioco:

$$\frac{\Delta v_0}{v_0} = \frac{\Delta \delta_x}{\delta_x} + \frac{\Delta u_x}{u_x} + \frac{\Delta \delta_y}{\delta_y} + \frac{\Delta u_y}{u_y} = \frac{\Delta \delta_x}{\delta_x} + \frac{\Delta l_x}{l_x} + \frac{\Delta \delta_y}{\delta_y} + \frac{\Delta l_y}{l_y}.$$

Mettendo tutto insieme si ha  $v_0 = 7.5 \pm 0.4 \text{ cm/s}$ . (L'intercetta e l'incertezza di misura ad essa associata dovrebbero essere, a questo punto, banali da stimare e sono lasciate per esercizio.)

Attenzione! Spesso capita di eseguire *fit* su grafici in carta millimetrata disegnati da noi stessi. In questo caso  $\delta_y$  e  $\delta_x$  vengono misurati con la carta stessa e, più importante, le scale degli assi  $u_y$  e  $u_x$  sono già riferite alla carta millimetrata: non sono quindi “misurate”, ma “scelte”, di conseguenza sono senza incertezza e  $\Delta u_x = \Delta u_y = 0$

La nostra stima degli errori sul coefficiente angolare di *best fit* nell'esempio 2.1 si basa sull'assunzione implicita che gli errori di misura siano abbastanza piccoli da definire univocamente la retta di *best fit*. Nella pratica si danno due casi estremi, come illustrato in figura 2.3, con tutte le sfumature intermedie. Se gli errori sono così piccoli da non permettere nessun margine nella determinazione della retta di *fit* (o addirittura non rappresentabili), allora l'errore sul coefficiente angolare è determinato dalla propagazione degli errori sul rapporto (21), mentre l'errore sull'intercetta è fissato dall'errore di lettura sul grafico, come già illustrato nell'esempio 2.1. Se invece le incertezze di misura sono abbastanza ampie, si può tracciare le rette di massima e minima pendenza ed utilizzare la media e la semidispersione tra i due casi per determinare il valore centrale e l'errore sui parametri, come illustrato nel grafico a destra in figura 2.3.

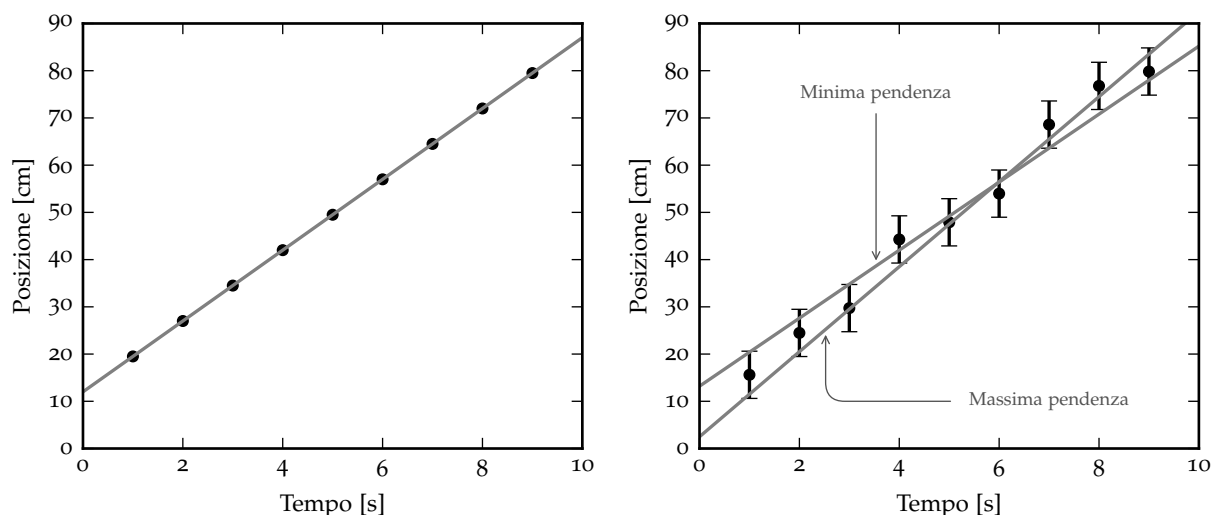


FIGURA 2.3. Esempi di fit grafici in due condizioni limite: quando gli errori non sono rappresentabili (sinistra) e quando gli errori sono grandi abbastanza da permettere di tracciare le rette di minima e massima pendenza (destra).

### 2.3 GRAFICI A BARRE ED ISTOGRAMMI

Non tutte le misure di laboratorio possono essere ridotte a situazioni in cui si misura una variabile dipendente in funzione di una indipendente, come nel caso illustrato nella sezione precedente. Supponiamo di avere a disposizione i voti d'esame di una sessione in cui siano stati valutati 83 studenti. Per rappresentare in una forma espressiva questa serie di dati possiamo realizzare un *grafico a barre*, come quello mostrato in figura 2.4 in cui, per ogni voto d'esame, l'altezza della barra corrispondente rappresenta il numero di studenti che hanno ottenuto quel voto. Va da sé che si tratta di un modo molto più immediato di rappresentare i dati rispetto ad una tabella di 83 numeri: si vede immediatamente, ad esempio, che il voto 29 non è mai stato assegnato. Sottolineiamo anche il fatto che, nell'assunzione che l'ordine temporale degli esami non sia rilevante (il che, in pratica, può verificarsi o meno), il grafico in figura 2.4 contiene tutta l'informazione originaria—non abbiamo operato, cioè, alcuna riduzione dei dati.

Vi sono casi in cui i valori possibili della grandezza che stiamo misurando sono troppi (o addirittura infiniti, se la grandezza può variare con continuità) per essere rappresentati con un grafico a barre. In tal caso si ricorre ad un *istogramma*, cioè ad una rappresentazione grafica in cui l'intervallo di variabilità della grandezza misurata è suddiviso in un numero fissato di sottointervalli (che si dicono canali, o *bin* dell'istogramma), e si conta il numero di misure che cadono in ciascuno di questi sottointervalli.

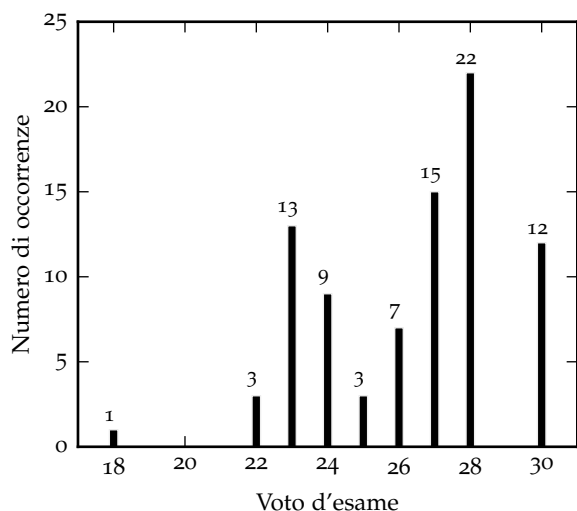


FIGURA 2.4. Rappresentazione, nella forma di un grafico a barre, dei voti degli 83 studenti in una determinata sessione di esame. (Ogni riferimento a persone o cose realmente esistenti è puramente casuale.) Per ogni voto d'esame, l'altezza della barra corrispondente rappresenta il numero di studenti che hanno ottenuto quel voto. L'indicazione esplicita delle occorrenze in corrispondenza di ogni barra contribuisce ad aumentare la leggibilità del grafico.



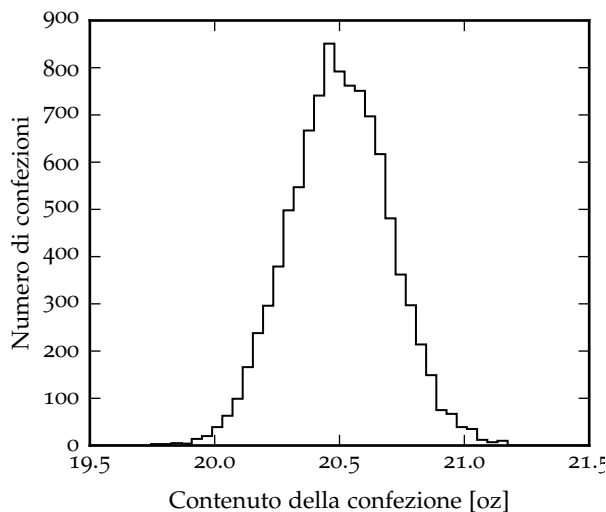


FIGURA 2.5. Istogramma del peso (misurato) del contenuto di 10,000 scatole di cereali *Chocolate Frosted Sugar Bombs*, prodotte dalla *General Junk-foods Corporation*—a fronte di un peso nominale di 20 oz. (1 oz, o oncia, equivale a 28.3 g e, non essendo un'unità di misura del Sistema Internazionale, non dovrebbe essere utilizzata.) I dati sono disponibili all'indirizzo <http://www2.stetson.edu/~jrasp/data.htm> e l'informazione rilevante può essere reperita all'indirizzo [http://www2.stetson.edu/~jrasp/stat201/activities/cereal\\_boxes.htm](http://www2.stetson.edu/~jrasp/stat201/activities/cereal_boxes.htm).

In figura 2.5 è mostrato, a titolo di esempio, l'istogramma del peso (misurato) del contenuto di 10,000 scatole di cereali—a fronte di un peso nominale di 20 oz. L'altezza di ciascun canale dell'istogramma rappresenta il numero di scatole di cereali con un peso contenuto nel sottointervallo corrispondente dell'asse delle ascisse. Chiaramente questo istogramma è estremamente più espressivo di una lista di 10,000 numeri: possiamo dire immediatamente, ad esempio, che il peso medio è attorno a 20.5 oz e che solo qualche scatola su 1,000 ha un peso inferiore a quello nominale. (Il che, a sua volta, ci dice che il processo produttivo è evidentemente tarato, per motivi legali, per garantire una probabilità ragionevolmente piccola che la quantità del prodotto nella scatola sia inferiore a quello pubblicizzato).

Benché superficialmente (e, talvolta, anche graficamente) simili, il grafico a barre e l'istogramma sono due rappresentazioni dei dati profondamente diverse da un punto di vista concettuale. Il primo, come abbiamo avuto occasione di notare poco fa, non presuppone nessuna riduzione dei dati (e quindi nessuna perdita di informazione), mentre l'altro sì. L'istogramma in figura 2.5 ha 50 canali—di cui alcuni vuoti—per cui implica una riduzione effettiva dell'informazione originale (10,000 numeri) di più di un fattore 200. Se questa perdita di informazione sia rilevante o meno va giudicato caso per caso. Certo è che, se avessimo scelto una suddivisione in canali diversa, avremmo ottenuto un istogramma diverso da cui, potenzialmente avremmo potuto trarre conclusioni diverse. (In altre parole: gli istogrammi sono oggetti molto utilizzati ed estremamente utili in pratica, ma bisogna sempre tenere in mente che essi implicano una compressione dell'informazione e deve essere cura dello sperimentatore assicurarsi che essa non abbia implicazioni rilevanti).

## 2.4 ESPONENZIALI E LOGARITMI

Sebbene dovrebbe essere un argomento ben noto, è bene fermarsi un attimo per riassumere—ed inquadrare meglio dal nostro punto di vista—le proprietà fondamentali di esponenziali ed logaritmi.

Per prima cosa esponenziale e logaritmo sono uno l'inverso dell'altro, nel senso delle composizione delle funzioni, e.g.,

$$\begin{aligned}\ln(e^x) &= e^{\ln x} = x \\ \log(10^x) &= 10^{\log x} = x.\end{aligned}$$

(Qui e nel seguito indicheremo con  $\ln$  il logaritmo Neperiano, cioè in base  $e$ , e con  $\log$  il logaritmo in base 10, indicando esplicitamente la base in tutti gli altri casi.)

Sappiamo anche che la funzione esponenziale cresce molto velocemente al crescere del suo argomento—tanto che l'aggettivo “esponenziale” e l'avverbio “esponenzialmente” sono entrati nell'uso comune per indicare genericamente (e, nella maggior parte dei casi, impropriamente) fenomeni che procedono con una progressione molto rapida. Di converso il logaritmo, che è l'inverso dell'esponenziale, cresce molto lentamente al crescere dell'argomento, come illustrato in figura 2.6. Ma cosa intendiamo esattamente per “velocemente” e “lentamente”?

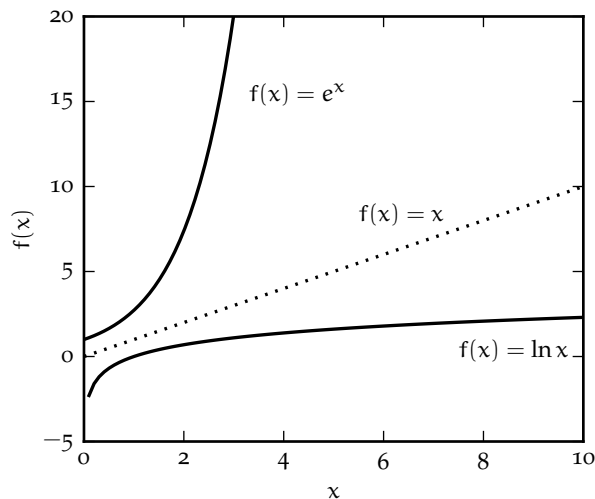


FIGURA 2.6. Grafico delle funzioni esponenziale e logaritmo. Per confronto, la retta tratteggiata tra le due rappresenta la funzione  $f(x) = x$ .

#### 2.4.1 Un tipico processo esponenziale: la divisione cellulare

Consideriamo il modello semplificato per la crescita di una popolazione batterica in cui: (i) ogni batterio si divide in due su un tempo scala fissato  $t_{\text{gen}}$ ; e (ii) i due batteri figli hanno una probabilità pari ad 1 di sopravvivere e dare luogo a loro volta ad una divisione binaria, come illustrato in figura 2.7. Sotto queste due ipotesi, la popolazione totale  $N$  di batteri raddoppia ad ogni generazione per cui all' $n$ -esima generazione (cioè dopo un tempo pari a  $n$  volte  $t_{\text{gen}}$ ) è data da

$$N(n) = 2^n. \quad (22)$$

Per fissare le idee: alla prima generazione abbiamo 2 batteri; alla seconda 4; alla decima 1024; alla ventesima più di un milione. (Se il tutto vi sembra non rilevante per la Fisica, sappiate che il modello più semplice per lo sviluppo di uno “sciame elettromagnetico” nella materia non è molto diverso).

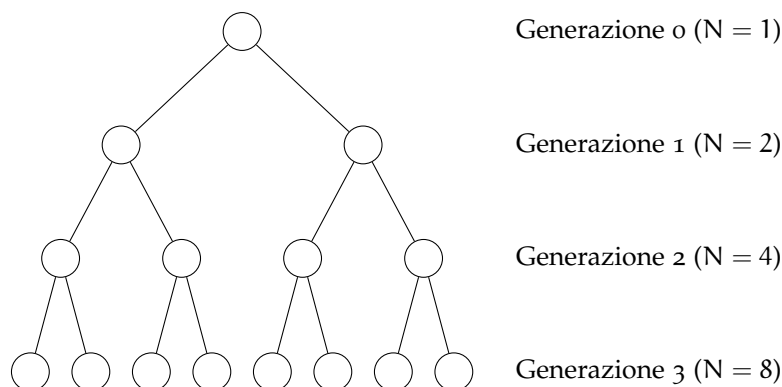


FIGURA 2.7. Schematizzazione del nostro modello semplificato per la crescita di una popolazione di batteri. Sotto le nostre ipotesi il numero totale  $N$  di batteri cresce esponenzialmente con il numero di generazioni.

Possiamo anche farci la domanda opposta, ovverosia: quante generazioni dobbiamo attendere per arrivare ad un numero fissato  $N$  di batteri? La risposta è ovvia:

$$n = \log_2 N. \quad (23)$$

Gli esempi 2.2 e 2.3 illustrano a dovere cosa intendiamo quando diciamo che l'esponenziale cresce “velocemente”. Per completezza dobbiamo notare anche come essi siano completamente irrealistici. Il nostro modello per la popolazione batterica non tiene in conto che le risorse per lo sviluppo cellulare non sono infinite per cui in realtà la popolazione stessa, dopo una fase iniziale di sviluppo esponenziale, tenderà necessariamente ad andare in equilibrio. D'altra parte non è operativamente possibile piegare un foglio di carta 42 volte (potete fare la prova—quanto dovrebbe essere l'area di base del nostro immaginario parallelepipedo finale?). In un certo senso potremmo dire che la crescita esponenziale è così veloce che in natura non si può sostenere a lungo.

► **ESEMPIO 2.2.** Nel nostro modellino di popolazione batterica, assumendo un tempo di generazione di  $t_{\text{gen}} = 20$  min, quanto tempo (e quante generazioni) dobbiamo attendere perché il numero di batteri ecceda il numero di protoni nell'Universo (diciamo  $10^{80}$ )? La risposta segue direttamente dalla (23) ed è 266, per un tempo pari a meno di 4 giorni.

► **ESEMPIO 2.3 (IL FOGLIO E LA LUNA).** Una storiella che gira tra i Fisici da tempo immemore. Quante volte devo piegare un foglio di carta per arrivare fino alla Luna? Non è difficile vedere l'analogia con l'esempio precedente: lo spessore del nostro foglio raddoppia ad ogni piegatura, per cui in effetti segue l'analogo della legge (22). Il rapporto tra la distanza Terra-Luna ( $\sim 380\,000$  km) e lo spessore di un foglio di carta ( $\sim 0.1$  mm) è dell'ordine di  $\sim 3.8 \times 10^{12}$ , e la risposta cercata è 42.

### 2.4.2 La ricerca binaria

Inquadriamo il problema da un punto di vista differente. Supponiamo di avere un elenco telefonico con  $N$  elementi ordinati alfabeticamente e di essere interessati al numero di telefono di una persona specifica—quale strategia useremmo?

Potremmo partire dall'inizio, scorrere nome per nome l'elenco e fermarci quando arriviamo alla persona desiderata—operare, cioè, una *ricerca sequenziale*. Così facendo dovremo controllare in media  $N/2$  numeri, ed  $N$  nel caso più sfavorevole (quello cioè in cui la persona cercata è l'ultima nell'elenco).

Oppure possiamo partire da metà dell'elenco e vedere se il nome cercato è prima o dopo nella lista: in un passo eliminiamo metà dell'elenco, e possiamo ripetere la procedura bisecando ogni volta la parte rimanente, che in questo modo si dimezza ad ogni passo. Non è difficile convincersi che il processo converge in un numero di passi minore o uguale a  $\log_2 N$ . (E per  $N$  grande,  $\log_2 N$  o  $N$  fa un'enorme differenza.) L'algoritmo che abbiamo appena delineato si dice *ricerca binaria* ed è ottimale per la ricerca di un elemento in una lista ordinata.

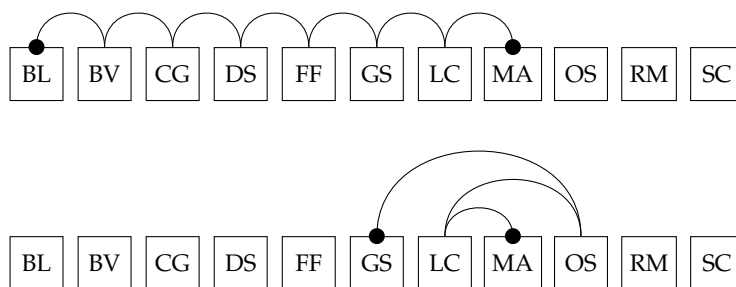


FIGURA 2.8. Illustrazione di due strategie di ricerca delle iniziali “MA” in una lista di 11 iniziali (fittizie) ordinate alfabeticamente. La ricerca sequenziale termina in 8 passi, quella binaria in 4 (o 3, a seconda di come definiamo l'algoritmo di bisezione). I cerchi neri indicano i punti di inizio e fine della ricerca.

► **ESEMPIO 2.4.** In un elenco telefonico con 11 persone una ricerca sequenziale richiede al massimo 11 passi ed una ricerca binaria 4. (Un esempio specifico è mostrato in figura 2.8, in cui per brevità si è usato iniziali fittizie al posto dei nomi completi.) In questo caso la differenza non è impressionante, ma con un elenco di 10 000 000 di persone una ricerca sequenziale richiede al massimo 10 000 000 di passi, mentre una ricerca binaria ne richiede al massimo 24.

### 2.4.3 Ancora sui logaritmi: il concetto di decade

Vi è un'altra proprietà dei logaritmi che utilizzeremo sovente in seguito: quella di trasformare prodotti in somme e quozienti in differenze. In formule:

$$\ln(x_1 x_2) = \ln x_1 + \ln x_2$$

$$\ln\left(\frac{x_1}{x_2}\right) = \ln x_1 - \ln x_2.$$

È in base alla seconda di queste identità che si dà una definizione operativa di decade—e, in contesti diversi, ottave, toni e semitoni.

Tecnicamente una decade è un intervallo di valori numerici i cui estremi stiano tra loro nel rapporto di 1 a 10. Dati allora due numeri  $x_1$  ed  $x_2$ , la loro distanza in decadi è data da

$$\text{dex}(x_1, x_2) = \log \left( \frac{x_2}{x_1} \right) \quad (24)$$

È banale verificare che, se  $x_2 = 10x_1$ ,  $\text{dex}(x_1, x_2) = 1$ , se  $x_2 = 100x_1$ ,  $\text{dex}(x_1, x_2) = 2$ , e così via.

► ESEMPIO 2.5. Il numero di decadi comprese tra i valori  $10^n$  e  $10^{n+m}$  si calcola banalmente come

$$\log \left( \frac{10^{n+m}}{10^n} \right) = \log(10^m) = m.$$

Ci sono cioè 3 decadi tra, e.g., 10 e 10000 oppure 1 e 1000 e due decadi, e.g., tra 0.1 e 10.

## 2.5 UNA BREVE DIGRESSIONE: IL SUONO

Il suono è essenzialmente un'onda meccanica che si propaga in un mezzo (ad esempio l'aria) e che noi percepiamo come una variazione di pressione sul timpano. Come tutte le onde, è caratterizzato dalla sua *frequenza* (e dalla lunghezza d'onda, ad essa legata dalla velocità di propagazione) e dalla sua *intensità*. Il suono è profondamente legato ai logaritmi in almeno due modi diversi, per cui approfittiamo del bagaglio di conoscenze che abbiamo appena acquisito per una breve digressione.

### 2.5.1 L'altezza del suono: l'ottava

L'*altezza* è ciò che distingue un suono acuto da un suono grave, ed è determinato essenzialmente dalla frequenza del suono in questione. L'orecchio umano è sensibile in un intervallo di frequenze compreso tra circa 20 Hz e circa 20 kHz—e questo intervallo (di tre decadi) si dice spettro udibile. Tanto per fissare le idee: il DO più grave del pianoforte moderno è a 32.7032 Hz e quello più acuto a 4186.01 Hz; il LA dell'ottava centrale è, per definizione, a 440 Hz e corrisponde anche all'altezza del diapason che si usa per accordare gli strumenti.

In modo del tutto analogo alla decade, l'*ottava* è un intervallo di frequenze i cui estremi stanno tra loro nel rapporto 1 : 2. Nel linguaggio dei logaritmi, il numero di ottave tra due frequenze  $\nu_1$  e  $\nu_2$  è

$$\text{ottave}(\nu_1, \nu_2) = \log_2 \left( \frac{\nu_2}{\nu_1} \right). \quad (25)$$

Nel sistema musicale occidentale l'ottava è a sua volta divisa in 12 semitoni, di solito equispaziati logaritmicamente secondo il cosiddetto temperamento equabile. Nella nostra notazione il numero di semitoni tra due frequenze è dato semplicemente da

$$\text{semitoni}(\nu_1, \nu_2) = 12 \log_2 \left( \frac{\nu_2}{\nu_1} \right). \quad (26)$$

► ESEMPIO 2.6. Il DO più grave e quello più acuto del pianoforte distano tra loro esattamente  $\log_2(4186.01/32.7032) = 7$  ottave o, equivalentemente, 84 semitoni.

La (26) può essere banalmente invertita per trovare la costante moltiplicativa che lega due frequenze che distino tra di loro esattamente un semitono

$$\frac{\nu_2}{\nu_1} = 2^{\frac{1}{12}} = \sqrt[12]{2} \approx 1.0594630943592953. \quad (27)$$

Nota	Espressione	Frequenza [Hz]
DO <sub>4</sub>	$440 \times 2^{-\frac{9}{12}}$	261.626
DO# <sub>4</sub>	$440 \times 2^{-\frac{8}{12}}$	277.183
RE <sub>4</sub>	$440 \times 2^{-\frac{7}{12}}$	293.665
RE# <sub>4</sub>	$440 \times 2^{-\frac{6}{12}}$	311.127
MI <sub>4</sub>	$440 \times 2^{-\frac{5}{12}}$	329.628
FA <sub>4</sub>	$440 \times 2^{-\frac{4}{12}}$	349.228
FA# <sub>4</sub>	$440 \times 2^{-\frac{3}{12}}$	369.994
SOL <sub>4</sub>	$440 \times 2^{-\frac{2}{12}}$	391.995
SOL# <sub>4</sub>	$440 \times 2^{-\frac{1}{12}}$	415.305
LA <sub>4</sub>	-	440.000
LA# <sub>4</sub>	$440 \times 2^{\frac{1}{12}}$	466.164
SI <sub>4</sub>	$440 \times 2^{\frac{2}{12}}$	493.883

FIGURA 2.9. Mappa delle frequenze corrispondenti ai dodici semitoni dell'ottava centrale del pianoforte moderno, derivate mediante la (26) a partire dalla frequenza (440 Hz) del LA centrale (LA<sub>4</sub>).

Il numero irrazionale  $\sqrt[12]{2}$  è cioè quello che definisce il rapporto tra le frequenze che identificano il semitono temperato. A partire da esso (e dai 440 Hz del LA di centro) possiamo costruire la mappa delle frequenze dell'ottava centrale del pianoforte (o del pianoforte tutto intero), come mostrato nella tabella 2.9.

Incidentalmente, la frequenza fondamentale di una corda vibrante può essere espressa in funzione della lunghezza  $\ell$ , della tensione  $\tau$  e della densità lineare di massa  $\mu$  della corda stessa attraverso la legge di Marsenne

$$v = \frac{1}{2\ell} \sqrt{\frac{\tau}{\mu}} \quad (28)$$

ed è inversamente proporzionale ad  $\ell$ . Le tastiere degli strumenti a corda (e.g., la chitarra) saranno dunque costruite in modo che le distanze dal ponticello di due tasti contigui seguano una legge simile alla (26)

$$\frac{\ell_2}{\ell_1} = \frac{v_1}{v_2} = 2^{-\frac{1}{12}} \approx 0.9438743126816935. \quad (29)$$

### 2.5.2 Il volume sonoro

Adesso la parte più interessante. Il volume di un suono è determinato dalla pressione  $P$  che l'onda sonora esercita sul timpano, ed il volume percepito  $L_p$  è misurato in decibel (dB) rispetto ad una pressione di riferimento  $P_0$

$$L_p = 20 \log \left( \frac{P}{P_0} \right) \text{ [dB]}. \quad (30)$$

La pressione di riferimento viene anche detta *soglia di udibilità* ed è convenzionalmente posta a 20  $\mu\text{Pa}$ . Una pressione sonora di 20  $\mu\text{Pa}$  corrisponde ad un livello di 0 dB, che è, appunto, il volume sonoro più basso che l'orecchio umano è in grado di percepire in aria.

Come è possibile che il nostro orecchio sia sensibile ad una pressione di 20  $\mu\text{Pa}$  quando la pressione atmosferica media è di 101 325 Pa—cioè più grande di quasi dieci ordini di grandezza? Il punto fondamentale è che la  $P$  nella (30) si riferisce alla componente *variabile nel tempo* della pressione—e, più precisamente, al suo scarto quadratico medio, ma per questo dovremo aspettare ancora un po'. La pressione atmosferica è approssimativamente costante nel tempo, almeno sui tempi scala delle frequenze acustiche, ed è applicata su entrambe le facce della membrana del timpano; le variazioni di pressione sono quelle che fanno vibrare la membrana stessa e ci danno la percezione del suono. (Detto questo, la sensibilità dell'orecchio umano rimane niente meno che impressionante. In termini circuitali l'orecchio è un condensatore perfetto che filtra la componente continua della pressione.)

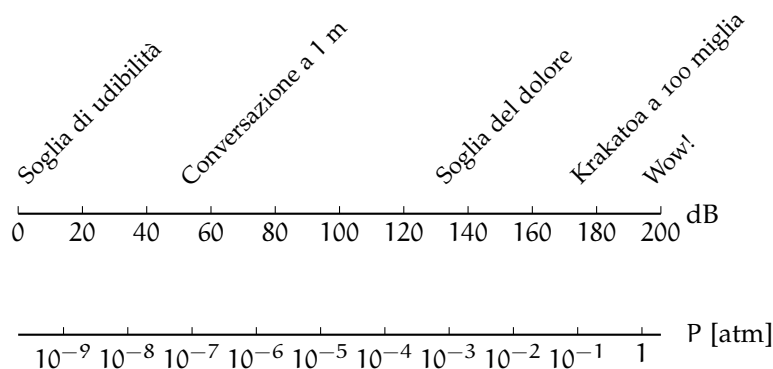


FIGURA 2.10. Illustrazione della scala dei decibel e relazione con la pressione atmosferica. La soglia di udibilità (20  $\mu\text{Pa}$ ) corrisponde ad una variazione di pressione nel tempo di circa  $2 \times 10^{-10}$  atm, una normale conversazione (40–60 dB) si attesta a  $\sim 10^{-7}$  atm e la soglia del dolore (130 dB) è a circa  $10^{-3}$  atm.

Torniamo alla (30). La scelta di questa unità di misura curiosa, il decibel, riflette il fatto che la nostra percezione del suono non scala in modo lineare ma, semmai, logaritmico con la pressione sonora, come illustrato in figura 2.10. (Il nostro orecchio è in grado di percepire la differenza di 100  $\mu\text{Pa}$  fra 100  $\mu\text{Pa}$  e 200  $\mu\text{Pa}$ , ma la stessa differenza tra 10.0 e 10.1 mPa risulta del tutto impercettibile.) Così, a fronte di una soglia di udibilità di 20  $\mu\text{Pa}$  ( $\sim 2 \times 10^{-10}$  atm), una normale conversazione tra due persone ad 1 m di distanza corrisponde ad una pressione sonora di 40–60 dB ( $\sim 10^{-7}$  atm) e la soglia del dolore è convenzionalmente fissata a 130 dB, che è ancora più piccolo di un millesimo di atm. Vedremo nella sezione 2.6 che la scala logaritmica è quella più appropriata per rappresentare grandezze di questo tipo, ed in effetti la figura 2.10 ne è un primo esempio.

Vista in termini della pressione (statica) atmosferica, la variazione di pressione corrispondente ad un qualsiasi fenomeno che classifichiamo come “suono” su scala umana è piccola. Se facessimo il grafico corrispondente della pressione sul timpano in funzione del tempo, sarebbe un grafico noioso—virtualmente indistinguibile da una linea orizzontale a  $\sim 100$  kPa, a meno di non sopprimere gli zeri. Udito distintamente fino a 3000 km di distanza, e con un livello sonoro stimato di 172 dB a 100 km, l’esplosione del vulcano dell’isola di Krakatoa (avvenuta il 27 agosto 1883) è considerato il suono più intenso (in aria) registrato sulla Terra nella storia moderna. Ma persino 172 dB corrispondono ad una variazione di pressione di meno del 10% della pressione atmosferica. (Per completezza, i KISS detengono il record di livello sonoro ad un concerto rock—un misero 136 dB misurato ad Ottawa nel 2009.<sup>1</sup>)

Il che ci porta ad un’ultima cosa interessante. Esiste un limite fisico al livello sonoro? La risposta è sì: quando la componente variabile della pressione è uguale alla pressione atmosferica, l’estremo inferiore della pressione totale è zero—il vuoto

$$L_p^{\max} = 20 \log \left( \frac{P_{\text{atm}}}{P_0} \right) = 194 \text{ [dB]}. \quad (31)$$

Un’onda sonora non può propagarsi in aria senza distorsioni a più di 194 dB picco-picco. (Se vi chiedevate il significato del “Wow!” in figura 2.10, questa è la risposta.)

## 2.6 SCALE NON LINEARI

Riprendiamo il filo del nostro discorso. Per l’occhio umano è relativamente semplice percepire deviazioni da un andamento lineare. Non è così semplice, di contrasto, distinguere immediatamente i grafici cartesiani di funzioni i cui andamenti siano qualitativamente simili—ad esempio  $x^2$  e  $x^4$ . Questo è uno dei motivi per cui ci siamo limitati alle funzioni lineari nella discussione dei *fit* grafici.

A titolo illustrativo, la tabella 6 contiene alcuni dati orbitali (semiasse maggiore dell’orbita e periodo di rivoluzione) degli 8 pianeti del Sistema Solare (Plutone escluso). Un grafico xy dei dati nella tabella, come quello in figura 2.11, suggerisce che le due quantità siano legate tra di loro da una qualche relazione funzionale—ma chiaramente questa relazione non è di tipo lineare, per cui, arrivati a questo punto, non sapremmo bene come procedere, dato il nostro bagaglio di conoscenze. Quale tipo di legge potrebbe legare il periodo dell’orbita al semiasse? (Un ulteriore problema, per certi versi secondario, risiede nel

<sup>1</sup> [https://en.wikipedia.org/wiki/Loudest\\_band\\_in\\_the\\_world](https://en.wikipedia.org/wiki/Loudest_band_in_the_world)

Pianeta	$a$ [UA]	$T$ [a]
Mercurio	0.3870993	0.2408467
Venere	0.723336	0.61519726
Terra	1.000003	1.0000174
Marte	1.52371	1.8808158
Giove	5.2029	11.862615
Saturno	9.537	29.447498
Urano	19.189	84.016846
Nettuno	30.0699	164.79132

TABELLA 6. Dati orbitali relativi agli 8 pianeti del Sistema Solare, da [http://www.princeton.edu/~willman/planetary\\_systems/Sol/](http://www.princeton.edu/~willman/planetary_systems/Sol/). Il semiasse maggiore  $a$  è espresso in unità astronomiche ( $1 \text{ UA} = 149\,597\,870\,700 \text{ m}$ ) ed il periodo di rivoluzione  $T$  in anni Giuliani ( $1 \text{ a} = 31\,557\,600 \text{ s}$  o 365.25 giorni). Assumeremo che tutte le cifre siano significative, anche se non è banale associare incertezze ai dati, poiché i parametri orbitali sono soggetti a complicate variazioni secolari.

fatto che i nostri dati variano su diversi ordini di grandezza, per cui nel grafico i 4 pianeti più vicini al sole tendono ad essere schiacciati sull'origine degli assi. Ci torneremo in seguito.)

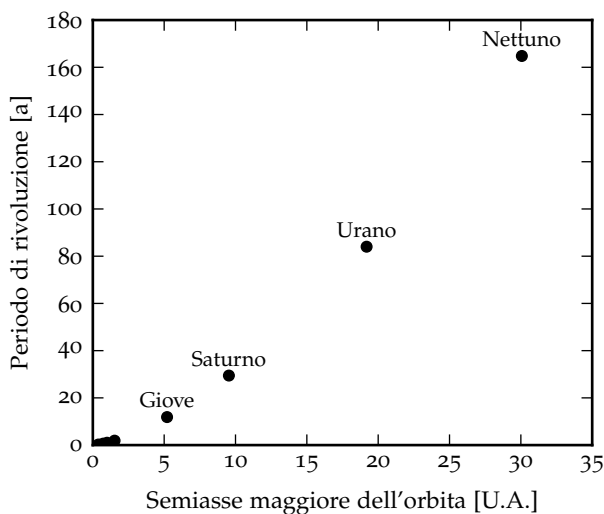


FIGURA 2.11. Grafico xy dei parametri orbitali (semiasse maggiore dell'orbita e periodo di rivoluzione) degli 8 pianeti del Sistema Solare, come mostrati nella tabella 6. Per la terza legge di Keplero (8), che abbiamo derivato per via dimensionale nell'esempio 1.5, ci aspettiamo che i punti si dispongano su una legge di potenza con esponente  $3/2$ .

Spesso funzioni complesse possono essere *linearizzate* mediante opportuni cambiamenti di variabili. È possibile in tal modo evidenziare caratteristiche che altrimenti sarebbero estremamente difficili da cogliere dalla semplice rappresentazione grafica. In questa sezione ci occuperemo di due classi di funzioni estremamente utili in Fisica—le leggi di potenza e gli esponenziali. Vedremo che in entrambi i casi si può linearizzare il problema usando scale logaritmiche.

### 2.6.1 Di nuovo sulle decadi: scale logaritmiche

Come abbiamo visto nel caso del suono, in natura esistono fenomeni che scalano in modo intrinsecamente logaritmico. Questo tipo di grandezze è convenientemente rappresentato in scala logaritmica.

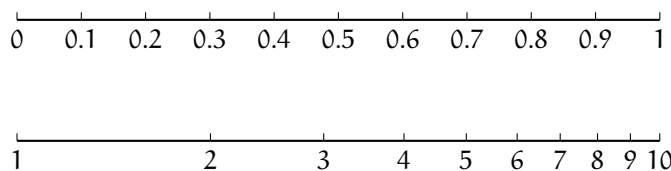


FIGURA 2.12. Costruzione di un asse in scala logaritmica. Assumendo una lunghezza unitaria per la decade, la distanza tra, e.g., 1 e 2 è pari a  $\log(2/1) \approx 0.301$ .

Torniamo per un attimo alla definizione di decade che abbiamo dato nella sezione 2.4.3. La (24) è importante perché ci permette di costruire scale graduate non lineari in cui la distanza (misurata in cm) tra due punti sia proporzionale alla distanza in decadi tra i punti corrispondenti, come illustrato in figura 2.12. Una scala di questo tipo si dice *scala logaritmica*. Così in scala logaritmica la distanza fisica tra 1 e 10 è identica a quella tra 10 e 100 o 0.01 e 0.1. Va da sé che lo 0 non può essere rappresentato su scala

logaritmica in quanto la distanza da esso un qualsiasi valore positivo divergerebbe—che è sostanzialmente una conseguenza del fatto che

$$\lim_{x \rightarrow 0^+} \log(x) = -\infty.$$

### 2.6.2 Leggi di potenza e grafici bilogaritmici

Si dice *legge di potenza* una funzione del tipo

$$y(x) = Cx^\Gamma, \quad (32)$$

in cui i parametri  $C$  e  $\Gamma$  si dicono rispettivamente costante ed esponente della legge di potenza (a seconda del contesto  $\Gamma$  viene detto talvolta indice spettrale o semplicemente indice). In Fisica gli esempi di legge di potenza sono innumerevoli.

► ESEMPIO 2.7. La legge spazio tempo per un corpo in caduta libera sotto l'azione della gravità terrestre (con le opportune condizioni iniziali)

$$s(t) = \frac{1}{2}gt^2$$

è una legge di potenza con  $C = g/2$  ed esponente  $\Gamma = 2$ .

► ESEMPIO 2.8. La relazione tra il periodo e la lunghezza di un pendolo (in approssimazione di piccole oscillazioni)

$$T(l) = 2\pi\sqrt{\frac{l}{g}}$$

è una legge di potenza con  $C = 2\pi/\sqrt{g}$  e  $\Gamma = 1/2$ .

► ESEMPIO 2.9. Una retta passante per l'origine  $y(x) = Cx$  è una legge di potenza con esponente  $\Gamma = 1$ .

► ESEMPIO 2.10. La terza legge di Keplero (8) è una legge di potenza con esponente  $\Gamma = 3/2$ .

► ESEMPIO 2.11. Il reddito procapite annuo è (approssimativamente) descritto da una distribuzione di probabilità a legge di potenza. (Come osservato da Pareto alla fine del XIX secolo, la maggior parte della ricchezza è posseduto da un piccolo numero di persone—che va anche sotto il nome di legge 80/20.)

► ESEMPIO 2.12. Lo spettro in energia dei raggi cosmici può essere ragionevolmente descritto da una legge di potenza con indice  $\Gamma \approx -2.75$  su quasi dieci ordini di grandezza in energia (si tratta di uno degli esempi più spettacolari in natura).

Calcolando il logaritmo in base 10 di entrambi i membri della (32), essa può essere convenientemente riscritta come:

$$\log(y) = \log(Cx^\Gamma) = \log(C) + \Gamma \log(x).$$

Mediante il cambiamento di variabile

$$\begin{cases} x' = \log(x) \\ y' = \log(y) \end{cases}$$

e ponendo  $C' = \log(C)$ , la (32) diviene:

$$y'(x') = C' + \Gamma x'. \quad (33)$$



Nelle nuove variabili la (32) è una retta il cui coefficiente angolare coincide con l'esponente della legge di potenza iniziale. Questo ci suggerisce che, data una serie di punti sperimentali  $(x_i, y_i)$ , un grafico dei valori di  $\log(y_i)$  in funzione di  $\log(x_i)$  sia un modo immediato per verificare se i dati sono legati tra di loro da una relazione funzionale di tipo legge di potenza. In realtà, in pratica, si preferisce usare grafici con assi in scala logaritmica, come mostrato in figura 2.12.

Un grafico in cui sia l'asse delle  $x$  che quello delle  $y$  sono in scala logaritmica si dice *grafico bilogaritmico*. Un foglio di carta prestampato con una coppia di assi ortogonali graduati in scala logaritmica (tipicamente corredati di griglie) si dice *carta bilogaritmica*. La carta bilogaritmica offre due ovvi vantaggi. Il primo è che in fase di creazione del grafico non è necessario calcolare i logaritmi dei valori sperimentali: è sufficiente usare gli assi graduati per mettere i punti sul grafico (questo in effetti non fa necessariamente una grande differenza se si utilizza un programma al calcolatore per realizzare il grafico). Il secondo, più importante, è che in fase di lettura del grafico è immediato ricavare i valori delle grandezze fisiche che i punti rappresentano—senza bisogno di elevamenti a potenza per invertire la trasformazione iniziale.

Torniamo allora al nostro esempio dei dati orbitali relativi ai pianeti del Sistema Solare, che abbiamo già illustrato nella tabella 6 e nella figura 2.11. Il grafico corrispondente in scala bilogaritmica è mostrato in figura 2.13, da cui è evidente che i punti si dispongono su di una retta. Questo ci permette di concludere che il legame funzionale tra il periodo di rivoluzione ed il semiasse maggiore dell'orbita è del tipo legge di potenza, come ci aspettavamo dalla terza legge di Keplero (cosa che non era affatto ovvio né dalla rappresentazione in forma tabellare, né dal grafico di dispersione in scala lineare). Vedremo nella prossima sezione come si possono stimare graficamente i parametri della legge di potenza in questione.

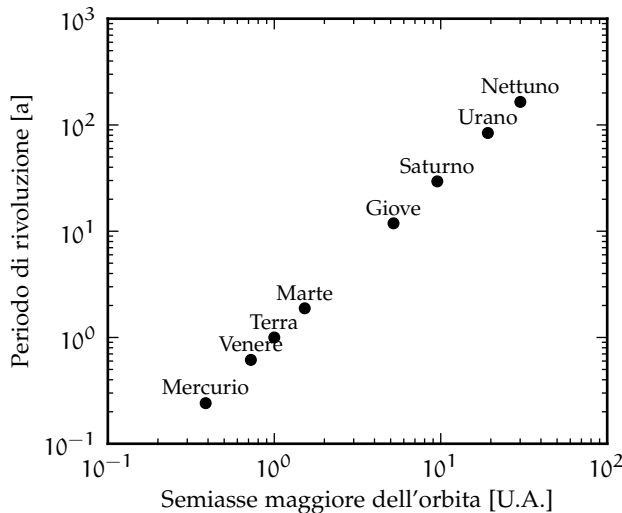


FIGURA 2.13. Grafico  $xy$  in scala bilogaritmica dei parametri orbitali (semiasse maggiore dell'orbita e periodo di rivoluzione) degli 8 pianeti del Sistema Solare, come mostrati nella tabella 6 e nella figura 2.11 (su scala lineare). In questa rappresentazione i punti si dispongono su di una retta, il che indica che le due grandezze sono legate tra di loro da una relazione funzionale di tipo legge di potenza, in accordo con la (8).

### 2.6.3 Fit grafici in scala bilogaritmica

I grafici in carta bilogaritmica permettono di estendere enormemente la classe di funzioni che è possibile fittare graficamente (e che su scala lineare si limitano essenzialmente a quelle lineari). Il procedimento è analogo a quello descritto per i grafici in scala lineare, con due peculiarità degne di nota.

Presi due punti sulla nostra retta di fit (figura 2.14), il coefficiente angolare si ricava dalla 33 come

$$\Gamma = \frac{y'_2 - y'_1}{x'_2 - x'_1} = \frac{\log(y_2) - \log(y_1)}{\log(x_2) - \log(x_1)} = \frac{\log(y_2/y_1)}{\log(x_2/x_1)}. \quad (34)$$

L'esponente della legge di potenza è dunque una misura di quante decadi ci spostiamo sull'asse delle  $y$  quando ci spostiamo di una decade sull'asse delle  $x$ . Se leggiamo i valori dal grafico,  $\Gamma$  si calcola usando la (34) (al solito propagando gli errori ove necessario). Se invece misuriamo direttamente i cateti del triangolo rettangolo in figura 2.14, dobbiamo usare la lunghezza di una decade (su ciascuno dei due assi separatamente) come unità di misura, come illustrato nell'esempio 2.13. Ne segue banalmente, per inciso, che se usiamo una carta bilogaritmica con le decadi sui due assi di uguale lunghezza l'esponente della legge di potenza è semplicemente il rapporto tra i cateti misurati in cm.

Se poi, nella (33), poniamo  $x' = 0$ , si ha  $y' = C'$ , ovvero  $y = C$ . Ma la condizione  $x' = 0$  corrisponde, nelle variabili non trasformate, a  $x = 1$ , per cui in scala bilogarithmica l'intercetta non è l'intercetta con l'asse  $x = 0$  (che per altro non esiste), ma l'intercetta con l'asse  $x = 1$ .

► **ESEMPIO 2.13.** Stimiamo graficamente l'indice della legge di potenza di *best fit* in figura 2.14. L'esercizio procede in modo sostanzialmente analogo a quello descritto nell'esempio 2.1, con la differenza sostanziale che  $\delta_x$  e  $\delta_y$  debbono adesso essere convertiti in decadi. Le lunghezze di una decade sui due assi sono rispettivamente

$$d_x = \frac{l_x}{3} \quad d_y = \frac{l_y}{4}$$

(ci sono 3 decadi sull'asse delle ascisse e 4 su quello delle ordinate). Procedendo come prima si ha

$$\Gamma = \frac{\delta_y/d_y}{\delta_x/d_x} = 1.51 \pm 0.08,$$

in accordo con la terza legge di Keplero (8).

#### 2.6.4 Funzioni esponenziali e grafici semilogaritmici

Si dice *esponenziale* una funzione del tipo

$$y(x) = Ce^{-\lambda x}. \quad (35)$$

(In Fisica gli esponenziali sono spesso decrescenti, da cui il segno  $-$  davanti a  $\lambda$ .) Nel caso, tipico, in cui la variabile indipendente  $x$  rappresenti un tempo, il parametro  $\lambda$  prende il nome di costante di decadimento ed il suo reciproco  $\tau = 1/\lambda$  si dice vita media. La vita media si interpreta come l'intervallo di tempo dopo il quale il valore della funzione si è ridotto ad un fattore

$$\frac{y(\tau)}{y(0)} = \frac{Ce^{-\lambda\tau}}{C} = \frac{Ce^{-\lambda/\lambda}}{C} = \frac{1}{e} \quad (36)$$

rispetto al valore iniziale. Una grandezza correlata alla vita media è il tempo di dimezzamento, definito come il tempo dopo il quale l'ampiezza si è ridotta ad  $1/2$  del valore iniziale, che si dimostra banalmente essere

$$T_{1/2} = \tau \ln 2. \quad (37)$$

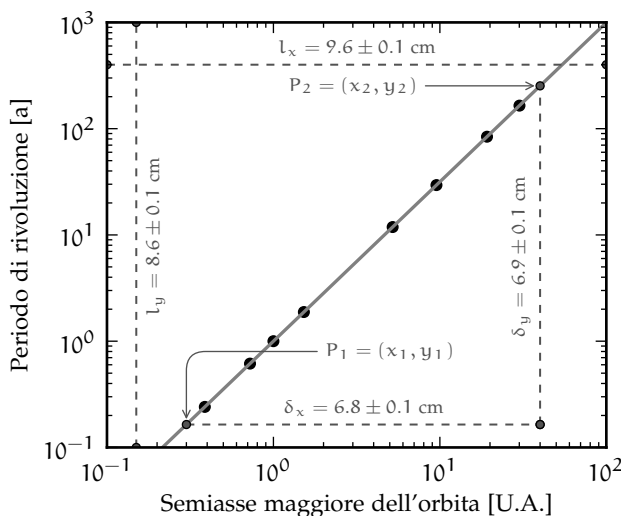


FIGURA 2.14. Grafico xy in scala bilogarithmica dei parametri orbitali (semiasse maggiore dell'orbita e periodo di rivoluzione) degli 8 pianeti del Sistema Solare, come mostrati nella tabella 6. Le annotazioni sul grafico sono state ottenute misurando le lunghezze rilevanti sullo schermo di un computer con un semplice righello, e sono sufficienti per stimare l'indice della legge di potenza, come illustrato passo passo nell'esempio 2.13.

► **ESEMPIO 2.14.** Si immerge un termometro in un misto di acqua e ghiaccio—per definizione a  $0^\circ$ . Assumendo la capacità termica del termometro trascurabile (cioè assumendo di avere abbastanza acqua e ghiaccio), ci aspettiamo che la temperatura  $T$  di quest'ultimo, per la legge del raffreddamento di Newton, si porti esponenzialmente nel tempo dalla sua temperatura iniziale  $T_0$  a quella ( $0^\circ$ ) del bagno termico

$$T[t] = T_0 e^{-\lambda t}.$$

Come per le leggi di potenza, calcoliamo il logaritmo in base dieci di entrambi i membri della 35:

$$\log(y) = \log(Ce^{\lambda x}) = \log(C) + \lambda x \log(e).$$

Con il cambiamento di variabile<sup>2</sup>

$$\begin{cases} x' = x \log(e) \\ y' = \log(y) \end{cases}$$

e ponendo, come prima,  $C' = \log(C)$ , la 35 diviene:

$$y' = C' + \lambda x' \quad (38)$$

cioè una retta. A questo punto non stupirà il fatto che esistono grafici *semilogaritmici* e carte prestampate semilogaritmiche.

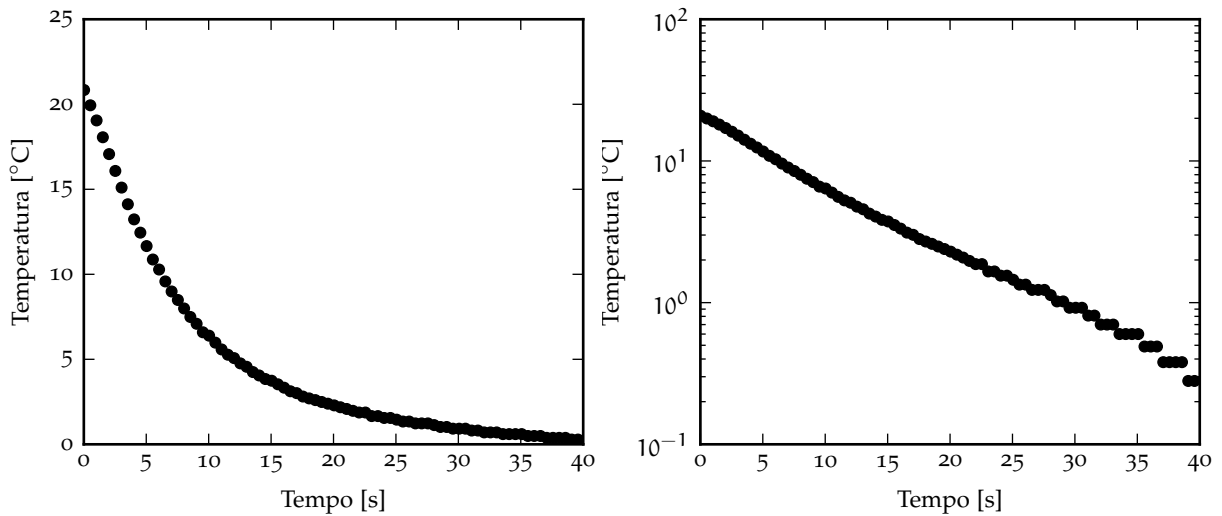


FIGURA 2.15. Grafici temperatura-tempo relativi all'esempio 2.14 in scala lineare (a sinistra) e semilogaritmica (a destra).

In figura 2.15 è mostrato un insieme di dati relativo all'esempio 2.14. In scala lineare non è ovvio riconoscere un esponenziale—inoltre è difficile leggere dal grafico i valori di  $T$  per tempi grandi, che tendono ad essere schiacciati nella parte inferiore del grafico. Nel grafico in carta semilogaritmica, viceversa, è immediato vedere che il decadimento delle temperatura è approssimativamente esponenziale, e si vede anche chiaramente una deviazione dall'andamento rettilineo, che indica che il nostro modello non è perfetto.

<sup>2</sup> Uno potrebbe, altrettanto correttamente, operare il cambiamento di variabile  $x' = x$  e, contestualmente,  $\lambda' = \lambda \log(e)$ . È facile verificare che l'unica equazione rilevante, la (39), rimane invariata.

## 2.6.5 Fit grafici in scala semilogaritmica

La discussione segue sostanzialmente quella della sezione 2.6.3. Presi due punti sulla funzione di fit, il coefficiente angolare si ricava dalla (35) come

$$\lambda = \frac{y'_2 - y'_1}{x'_2 - x'_1} = \frac{\log(y_2) - \log(y_1)}{(x_2 - x_1) \log(e)} = \frac{\log(y_2/y_1)}{(x_2 - x_1) \log(e)} \quad (39)$$

e l'intercetta è questa volta l'intercetta con l'asse  $x = 0$ .

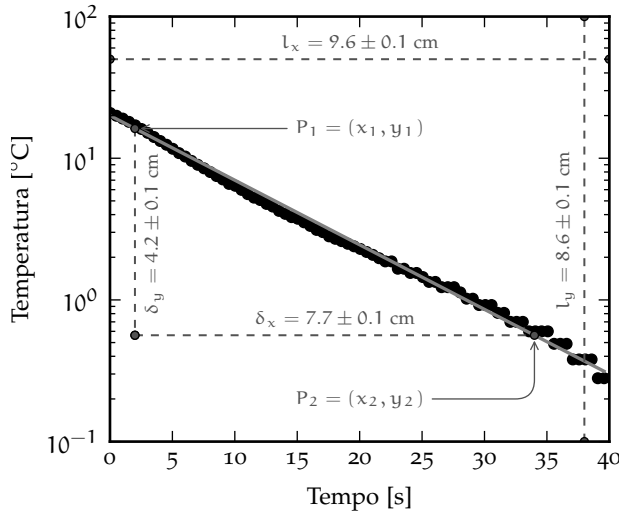


FIGURA 2.16. Fit grafico in scala semilogaritmica relativo ai dati mostrati in figura 2.15. Le annotazioni sul grafico sono state ottenute misurando le lunghezze rilevanti sullo schermo di un computer con un semplice righello, e sono sufficienti per stimare la costante di decadimento del processo, come illustrato passo passo nell'esempio 2.15.

Un metodo alternativo per stimare graficamente  $\lambda$  è quello di utilizzare la (37): si traccia la retta orizzontale con ordinata pari ad  $1/2$  del valore iniziale ed il valore delle ascisse per cui essa intercetta con la funzione di fit corrisponde al tempo di dimezzamento.

► **ESEMPIO 2.15.** Stimiamo graficamente il parametro  $\lambda$  per il fit relativo ai dati dell'esempio 2.14 mostrato in figura 2.16. Per la (39) si tratta di una sorta di ibrido tra i procedimenti utilizzati negli esempi 2.1 e 2.13. Le quantità rilevanti sono

$$u_x = \frac{l_x}{40} \quad d_y = \frac{l_y}{3}$$

ed il parametro cercato si scrive come

$$\lambda = \frac{\delta_y/d_y}{\log(e)\delta_x/u_x} = 0.105 \pm 0.006 \text{ s}^{-1},$$

che corrisponde ad una vita media di  $\tau = 9.5 \pm 0.6 \text{ s}$ .

## 2.7 IN BREVE...

- Tabelle e grafici sono fondamentali per ordinare e rendere comprensibili i vostri dati e devono essere il più possibile chiari ed auto-esplicativi.
- Dare sempre nomi chiari alle colonne delle tabelle ed agli assi dei grafici, ed indicate sempre le unità di misura.
- Le incertezze devono essere mostrate in modo chiaro ed esplicito:
  - Nelle tabelle devono essere sempre espresse

- Nei grafici, per quando possibile, scegliete le scale degli assi in modo che le incertezze siano ben visibili.
- È bene riportare le divisioni, regolari, degli assi dei grafici, in modo da poter leggere le coordinate di un qualunque punto in modo immediato. Non è necessario che gli assi partano sempre da zero.
- Annotate sul grafico le informazioni che possono essere utili al lettore, ad esempio:
  - In un fit con una retta può servire annotare i punti (sulla retta) e le distanze utilizzate per stimare il coefficiente angolare
  - Se gli assi sono ben leggibili, è inutile annotare le coordinate dei punti stessi
- La scelta dei bin in un istogramma è sempre un po' arbitraria, bisogna cercare il giusto compromesso per avere bin abbastanza popolati senza ridurne troppo il numero.
- In un fit grafico (a mano, con un righello) cercate la retta che passa "vicino" a tutti i punti, compatibilmente con le incertezze sperimentali.
- Le scale logaritmiche sono vostre amiche: calcolano i logaritmi al posto vostro e vi aiutano a rappresentare facilmente leggi di potenza (scala bilogaritmica) ed esponenziali (scala semilogaritmica).



## ELEMENTI DI TEORIA DELLE PROBABILITÀ

Cominciamo questo capitolo con una domanda e due problemi. La domanda è la seguente: qual è la differenza (o le differenze) tra la *teoria delle probabilità* e la *statistica*? Ed i due problemi, connessi con la nostra domanda, sono:

- *Problema 1*: abbiamo un'urna contenente 5 palline, di cui 3 rosse e 2 blu. Se estraiamo (bendati) una pallina qual è la probabilità che essa sia blu?
- *Problema 2*: abbiamo un'urna contenente 5 palline—alcune rosse ed alcune blu (ma non conosciamo la proporzione tra i colori). Estraiamo una pallina, che risulta essere blu, e la mettiamo da parte. Estraiamo una seconda pallina, che stavolta è rossa. Qual è il numero di palline rosse nell'urna?

La prima domanda è facile, e la risposta è univocamente determinata dalle condizioni del problema (e, per completezza, è  $2/5$ ). La seconda, invece, sembra apparentemente senza speranza. Ma, a pensarci meglio, abbiamo sicuramente informazione per dire *qualcosa*; possiamo dire, ad esempio, che le palline non sono tutte blu e non sono tutte rosse. E il fatto di aver estratto una pallina rossa ed una pallina blu potrebbe farci pensare che l'eventualità che il numero di palline blu sia 1 o 4 sia relativamente poco probabile, anche se non impossibile.

Proviamo a rifrasare il tutto. Il primo è un tipico problema di probabilità, in cui abbiamo un sistema descritto da un modello ben preciso e possiamo trarre conclusioni in modo deduttivo. Il secondo, che è più vicino al mestiere del Fisico, è invece un problema di statistica, in cui a partire da un numero finito di osservazioni, cerchiamo (induttivamente) di trarre conclusioni sul modello che regola il nostro sistema. È chiaro che i due problemi sono legati tra loro e, in un certo senso, il secondo è l'inverso del primo—in effetti potremmo chiamarlo un problema di *probabilità inversa*.

Benché non universalmente accettata (si veda [10] come esempio di uno schema logico alternativo in cui le due cose sono unificate), questa distinzione di fondo tra teoria delle probabilità, come branca della matematica pura, e statistica, come mezzo di indagine dei fenomeni fisici, ci pare utile per inquadrare, almeno in una fase iniziale, i problemi che incontreremo.

## 3.1 DEFINIZIONE ASSIOMATICA DELLA PROBABILITÀ

Si deve a Kolmogorov [12] la prima costruzione rigorosa della teoria della probabilità, in una struttura che sostanzialmente sopravvive nei manuali moderni. Quella che segue non è una esposizione rigorosa della teoria assiomatica della probabilità, ma solo un accenno superficiale ad alcune idee di base ad essa connesse, in cui daremo per noti i concetti di insieme ed operazioni tra insiemi.

La struttura di base su cui si fonda la teoria assiomatica della probabilità *a la* Kolmogorov è data dal cosiddetto *spazio campionario*  $\Omega$ , cioè l'insieme (che per semplicità assumeremo numerabile) di tutte le possibili realizzazioni elementari di un dato fenomeno. Sulla base di questo si definisce poi lo spazio degli eventi  $\mathcal{F}$  come *l'insieme di tutti i sottoinsiemi* di  $\Omega$ , che si dice anche *l'insieme di potenza* di  $\Omega$ . L'idea di base è che la probabilità è definita proprio sullo spazio degli eventi—cioè si può assegnare una probabilità non solo ad un qualsiasi elemento dello spazio campionario, ma anche ad uno qualsiasi dei suoi sottoinsiemi. Siccome l'unione e l'intersezione di sottoinsiemi dello spazio campionario fanno ancora parte dello spazio degli eventi, avrà senso parlare anche della probabilità dell'unione e dell'intersezione di eventi. Anziché addentrarci troppo nei dettagli tecnici cercheremo di chiarire il concetto con i due esempi 3.1 e 3.2.

► **ESEMPIO 3.1.** Nel lancio di una moneta lo spazio campionario  $\Omega$  è dato semplicemente dalle due possibili uscite  $\{T, C\}$  (testa o croce). L'insieme di potenza di  $\Omega$  contiene, oltre all'insieme nullo  $\emptyset$  ed allo spazio campionario stesso, i due sottoinsiemi propri di  $\Omega$   $\{T\}$  e  $\{C\}$ . Le quattro domande che possiamo farci a proposito del lancio di una moneta sono dunque a proposito degli eventi

$\{T\}$  : qual è la probabilità che esca testa?

$\{C\}$  : qual è la probabilità che esca croce?

$\{T, C\}$  : qual è la probabilità che esca testa oppure croce?

$\{\}$  : qual è la probabilità che non esca né testa né croce?

► **ESEMPIO 3.2.** Nel lancio di un dado a sei facce lo spazio campionario  $\Omega$  è dato dai sei valori possibili delle uscite  $\{1, 2, 3, 4, 5, 6\}$ . L'insieme di potenza di  $\Omega$  contiene tutti gli eventi di cui sia lecito chiedersi quale sia la probabilità—ad esempio:

$\{3\}$  : qual è la probabilità che esca il numero 3?

$\{2, 4, 6\}$  : qual è la probabilità che l'uscita sia pari?

$\{1, 2, 3\}$  : qual è la probabilità che l'uscita sia  $< 4$ ?

$\{1, 2, 3, 4, 5, 6\}$  : qual è la probabilità che esca una faccia qualsiasi?

$\{\}$  : qual è la probabilità che non esca nessuna faccia?

In entrambi gli esempi gli ultimi due casi corrispondono all'intero spazio campionario  $\Omega$  ed all'insieme vuoto  $\emptyset$ , e le risposte alle domande corrispondenti sembrano semplici.

Si definisce probabilità una misura  $P$  su  $\mathcal{F}$  che associi univocamente ad ogni elemento  $E$  di  $\mathcal{F}$  un numero reale  $P(E)$  che soddisfa le seguenti tre proprietà (o assiomi di Kolmogorov):

1.  $0 \leq P(E) \leq 1 \forall E \in \mathcal{F}$ ;
2.  $P(\Omega) = 1$ ;
3.  $P(E_1 \cup E_2) = P(E_1) + P(E_2)$  se  $E_1 \cap E_2 = \emptyset$ .

Notiamo che il terzo assioma si estende banalmente all'unione numerabile di eventi disgiunti. Per completezza, due eventi che abbiano intersezione nulla si dicono disgiunti o *incompatibili*—nel senso che la probabilità che essi accadano contemporaneamente è nulla. Gli assiomi di Kolmogorov costituiscono una sorta di *grammatica* di base su cui sviluppare il linguaggio delle probabilità.

► **ESEMPIO 3.3.** Nel caso del lancio di una moneta possiamo associare una probabilità allo spazio campionario definendo, ad esempio,  $P(T) = P(C) = 1/2$  (con questa prescrizione la moneta è per definizione equa). Per il primo assioma di Kolmogorov la probabilità che non esca né testa né croce è  $P(\emptyset) = 0$  e per il secondo e terzo assioma, visto che i due eventi elementari sono mutuamente esclusivi, la probabilità che esca testa oppure croce è  $P(\Omega) = 1$ .

► **ESEMPIO 3.4.** Nel lancio di un dado a sei facce la misura che associa  $1/6$  alle 6 possibili realizzazioni elementari  $\{1\} \dots \{6\}$  è una probabilità, come si può verificare banalmente. Si noti in particolare che

$$\Omega = \{1\} \cup \{2\} \cup \{3\} \cup \{4\} \cup \{5\} \cup \{6\}$$

e  $P(\Omega) = 1$  per l'assioma 3. Va da sé che, con questa prescrizione, il dado è equo.

### 3.1.1 Leggi elementari della probabilità

Gli assiomi di Kolmogorov possono essere utilizzati per dimostrare un certo numero di risultati elementari.



**TEOREMA 3.1 (DELLA PROBABILITÀ COMPLEMENTARE).** Dato un evento  $E$ , e detto  $\bar{E}$  il suo complementare in  $\Omega$ , si ha

$$P(\bar{E}) = 1 - P(E). \quad (40)$$

*Dimostrazione.* Segue banalmente dal fatto che  $\bar{E} \cup E = \Omega$  e  $\bar{E} \cap E = \emptyset$  per cui, utilizzando gli assiomi 2. e 3. si ha

$$1 = P(\Omega) = P(E \cup \bar{E}) = P(E) + P(\bar{E}),$$

da cui la tesi.  $\square$

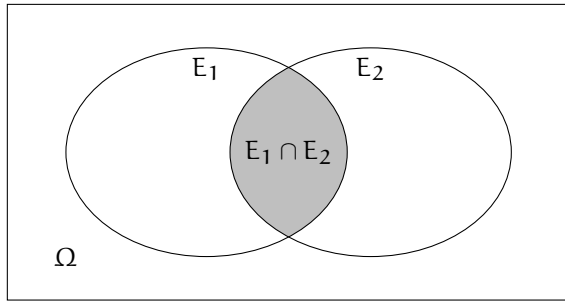
► **ESEMPIO 3.5.** Nel lancio di un dado equo a sei facce la probabilità che l'uscita non sia un 3, cioè dell'evento  $\{1, 2, 4, 5, 6\}$ , è data dalla somma  $P(1) + P(2) + P(4) + P(5) + P(6) = 5/6$ , che si può scrivere più facilmente come

$$P(\bar{3}) = 1 - P(3) = 5/6.$$

Per completezza, la sezione 3.12.1 illustra un'applicazione più interessante del teorema della probabilità complementare.

**COROLLARIO 3.1.**  $P(\emptyset) = 0$ .

**COROLLARIO 3.2.** Se  $E_1 \subset E_2$ , allora  $P(E_1) \leq P(E_2)$ .



**FIGURA 3.1.** Diagramma di Venn per il teorema di addizione delle probabilità 3.2. Il diagramma rende intuitivo il contenuto dell'ultimo termine della formula: sommando semplicemente  $P(E_1)$  e  $P(E_2)$  conteremmo due volte l'intersezione. Il diagramma illustra anche i due fatti elementari utilizzati nella dimostrazione:  $E_2 = (E_2 \cap E_1) + (E_2 \cap \bar{E}_1)$  e  $E_1 \cup E_2 = E_1 \cup (\bar{E}_1 \cap E_2)$ .

**TEOREMA 3.2 (DI ADDIZIONE DELLE PROBABILITÀ).** Dati due eventi  $E_1$  ed  $E_2$ , si ha

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2). \quad (41)$$

*Dimostrazione.* Possiamo scrivere  $E_2$  come la somma di due insiemi disgiunti notando semplicemente che, come illustrato nel diagramma in figura 3.1

$$E_2 = E_2 \cap \Omega = E_2 \cap (E_1 \cup \bar{E}_1) = (E_2 \cap E_1) \cup (E_2 \cap \bar{E}_1),$$

che, nel nostro linguaggio, equivale a dire

$$P(E_2) = P((E_2 \cap E_1) \cup (E_2 \cap \bar{E}_1)) = P(E_2 \cap E_1) + P(E_2 \cap \bar{E}_1).$$

D'altra parte si ha anche

$$\begin{aligned} E_1 \cup E_2 &= (E_1 \cup E_2) \cap \Omega = (E_1 \cup E_2) \cap (E_1 + \bar{E}_1) = (E_1 \cap E_1) \cup (E_1 \cap \bar{E}_1) \cup (E_2 \cap E_1) \cup (E_2 \cap \bar{E}_1) = \\ &= E_1 \cup \emptyset \cup (E_2 \cap E_1) \cup (E_2 \cap \bar{E}_1) = E_1 \cup (E_2 \cap \bar{E}_1), \end{aligned}$$

ovverosia

$$P(E_1 \cup E_2) = P(E_1) + P(E_2 \cap \bar{E}_1).$$

Mettendo insieme le due equazioni sopra si ottiene la tesi.  $\square$

Il diagramma di Venn in figura 3.1 illustra in modo intuitivo il contenuto della dimostrazione. Se gli eventi  $E_1$  e  $E_2$  sono disgiunti (cioè  $E_1 \cap E_2 = \emptyset$ ) allora la formula generale di addizione delle probabilità (41) si riduce banalmente al terzo assioma di Kolmogorov.

► ESEMPIO 3.6. Nel lancio di un dado equo a 6 facce la probabilità che esca il 2 oppure il 3 è data da  $P(2 \text{ o } 3) = P(2) + P(3) = 1/6 + 1/6 = 1/3$ .

► ESEMPIO 3.7. La probabilità di estrarre un re oppure una carta di cuori da un mazzo di 52 carte è data da  $P(K \cup \heartsuit) = P(K) + P(\heartsuit) - P(K \cap \heartsuit) = 4/52 + 13/52 - 1/52 = 16/52$  (dove l'ultimo termine della somma serve a non contare due volte il re di cuori).

### 3.2 DEFINIZIONI OPERATIVE DI PROBABILITÀ

Come vedremo nel seguito, la definizione assiomatica di probabilità che abbiamo appena enunciato permette di derivare in modo rigoroso un certo numero di proprietà praticamente rilevanti di cui la probabilità stessa gode. Essa non dice niente, però, su come si possa *calcolare* la probabilità di un dato evento in casi concreti. (Non a caso, nell'esempio 3.4 abbiamo detto che la nostra misura era *una* probabilità.) In questa sezione esaminiamo brevemente alcune definizioni *operative* di probabilità che ci saranno utili nel seguito.

#### 3.2.1 Definizione combinatoriale

Nella sua definizione *combinatoria* la probabilità di un evento  $E$  coincide con il rapporto tra il numero di casi favorevoli  $n$  ed il numero di casi possibili  $N$ , a condizione che questi ultimi siano tutti ugualmente probabili:

$$P(E) = \frac{n}{N}. \quad (42)$$

Notiamo che questa definizione soddisfa gli assiomi di Kolmogorov: il primo assioma discende dalla ovvia condizione  $0 \leq n \leq N$  ed il secondo dal fatto che se  $n = N$ , allora  $P(E) = 1$ . Si ha inoltre che, se  $E_1$  ed  $E_2$  sono due eventi disgiunti con un numero di casi favorevoli pari a  $n_1$  ed  $n_2$ , rispettivamente, si ha

$$P(E_1 \cup E_2) = \frac{n_1 + n_2}{N} = \frac{n_1}{N} + \frac{n_2}{N} = P(E_1) + P(E_2).$$

Il lettore più attento si sarà accorto che si tratta di una definizione circolare, nel senso che richiediamo l'equiprobabilità dei casi nella definizione stessa di probabilità. È altresì chiaro che il campo di applicabilità della (42) è limitato ai problemi più elementari—ad una domanda come “qual è la probabilità che un certo dispositivo elettronico subisca un guasto nel primo anno di funzionamento?” è estremamente difficile rispondere nel quadro di questo schema logico poiché non è affatto ovvio come si protrebbero definire i casi favorevoli ed i casi possibili. E, purtuttavia, si tratta di una nozione utile che utilizzeremo occasionalmente in seguito—negli esempi 3.8–3.10 e, soprattutto, nella derivazione della distribuzione binomiale (sezione 5.1).

► ESEMPIO 3.8. Il lancio di un dado equo a sei facce ha sei possibili esiti (equiprobabili). La probabilità che esca un numero fissato, ad esempio il 3, è dunque  $P(3) = 1/6$ .

► ESEMPIO 3.9. Supponiamo di avere un mazzo di 52 carte. La probabilità di estrarre un re è  $P(K) = 4/52 = 1/13$ ; la probabilità di estrarre una carta di cuori è  $P(\heartsuit) = 13/52 = 1/4$ .

► ESEMPIO 3.10. Supponiamo di lanciare due dadi equi a sei facce ed essere interessati alla somma delle uscite. Gli esiti possibili sono in questo caso 11 (i numeri interi da 2 a 12) ma essi non sono equiprobabili, per cui  $P(3) \neq 1/11$ . Il modo corretto di affrontare il problema dalla prospettiva della definizione combinatoriale di probabilità è il seguente: vi sono esattamente  $6 \times 6 = 36$  configurazioni (equiprobabili) distinte in cui i dadi possono atterrare e, tra queste, esattamente due danno come somma 3—per cui  $P(3) = 2/36 = 1/18$ . Torneremo sulla questione nell'esempio 3.21.

### 3.2.2 Definizione frequentista

Quando è possibile ripetere un esperimento in condizioni controllate, possiamo definire la probabilità di un evento  $E$  come il limite della frequenza relativa dell'evento stesso quando il numero di ripetizioni dell'esperimento è molto grande. In altre parole possiamo pensare di eseguire l'esperimento un numero (arbitrariamente grande) di volte  $N$ , contare il numero di volte  $n$  in cui  $E$  si verifica, e definire  $P(E)$  come il limite del rapporto  $n/N$  per  $N \rightarrow \infty$ :

$$P(E) = \lim_{N \rightarrow \infty} \frac{n}{N}. \quad (43)$$

(Vale la pena sottolineare come, benché formalmente simili, la (42) e (43) siano completamente diverse dal punto di vista concettuale.)

La (43) si dice generalmente *definizione frequentista* di probabilità e, seppure non particolarmente adatta ad una costruzione matematica rigorosa della teoria della probabilità, può essere operativamente utile—specialmente in Fisica, ove la ripetizione di un esperimento controllato è tipica. In particolare lo schema frequentista è alla base del metodo Monte Carlo che discuteremo (molto brevemente) in seguito.

Ovviamente non si ha nessuna garanzia che un numero fissato di ripetizioni sia in generale sufficiente perché la frequenza relativa sia una stima *sufficientemente buona* della probabilità cercata. Il limite non è cioè da intendersi nel senso usuale dell'analisi matematica

$$\forall \epsilon > 0 \exists \tilde{N} > 0 : N \geq \tilde{N} \rightarrow \left| \frac{n}{N} - P(E) \right| \leq \epsilon$$

ma piuttosto in quello (più debole) della convergenza statistica, che possiamo definire formalmente come

$$\forall \epsilon, \delta > 0 \exists \tilde{N} > 0 : N > \tilde{N} \rightarrow P\left(\left| \frac{n}{N} - P(E) \right| \geq \delta\right) \leq \epsilon.$$

In pratica il punto (sottile) è che non esiste un numero  $N$  di ripetizioni del nostro esperimento che mi permetta di affermare con certezza che la differenza tra la frequenza registrata  $n/N$  e la probabilità  $P(E)$  cui essa tende sia al di sotto di un valore prefissato  $\epsilon$ . In generale, se eseguo due serie di  $N$  ripetizioni dello stesso esperimento otterrò frequenze relative  $n/N$  diverse anche se  $P(E)$  è la stessa, come illustrato in figura 3.2. Quello che posso dire è che se  $N$  è abbastanza grande, allora posso rendere piccola a piacere la probabilità che  $n/N$  si discosti da  $P(E)$  di più di un valore prefissato  $\epsilon$ .

► **ESEMPIO 3.11.** Se lanciamo  $N$  volte un dado equo a sei facce e registriamo (al crescere di  $N$ ) il numero  $n$  di volte in cui esce, ad esempio, il numero 3, per  $N$  molto grande il rapporto  $n/N$  tenderà a  $P(3) = 1/6$  (nel senso della convergenza statistica), come mostrato in figura 3.2.

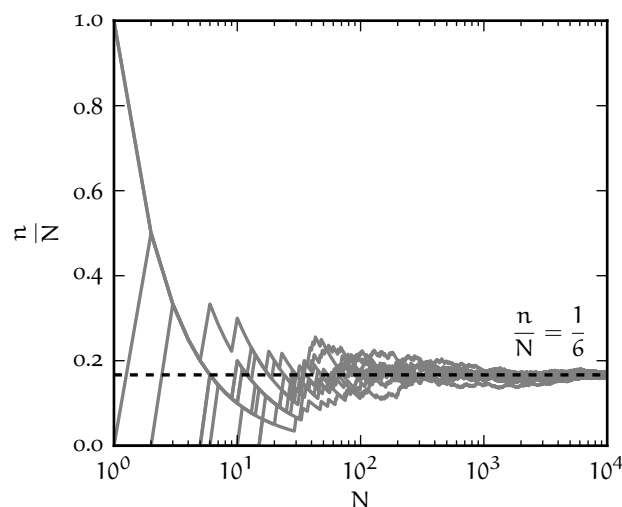


FIGURA 3.2. 10 diverse realizzazioni di 10 000 lanci di un dado equo a sei facce, in cui si registra, al variare del numero parziale  $N$  di lanci, la frazione  $n/N$  delle volte in cui esce il numero 3. La linea orizzontale tratteggiata indica la probabilità  $p = 1/6$  dell'evento  $e$ , come si vede, la frequenza tende a questa probabilità per  $N$  grande—nel senso che le fluttuazioni attorno a questo valore si riducono al crescere di  $N$ .

### 3.2.3 Definizione soggettiva

Nella sua definizione *soggettiva* la probabilità di un evento  $E$  si identifica con la *misura del grado di fiducia che un individuo attribuisce al verificarsi di  $E$ , sulla base dell'informazione a sua disposizione*. (La seconda parte della definizione è importante, perché persone diverse, con informazioni diverse, in generale assoceranno una probabilità diversa allo stesso evento—da cui il termine “soggettivo” nel titolo di questa sezione.)

Una volta superato lo sbigottimento iniziale—causato dall'attaccamento dei Fisici al concetto di oggettività—questa definizione è tutto sommato naturale e vicina al modo in cui operiamo nella vita di tutti i giorni. Ogni volta che facciamo una scelta tra due o più possibilità diverse, di fatto associamo implicitamente delle probabilità agli eventi che ne conseguono—e persone diverse, in base alle informazioni a loro disposizione, fanno scelte diverse. È altresì chiaro che la definizione soggettiva di probabilità ha un campo di applicabilità estremamente più vasto di quelle combinatoriale o frequentista: all'interno di questo schema concettuale possiamo parlare della probabilità di qualsiasi evento o proposizione.

All'interno della scuola soggettivista vi sono diversi approcci distinti per derivare le regole fondamentali della probabilità in un modo logicamente consistente, il più popolare dei quali è probabilmente l'idea del principio della scommessa coerente: *una volta assegnate la probabilità ad un evento dovremmo essere disposti ad accettare scommesse sul verificarsi dell'evento stesso con un rapporto tra puntata e vincita determinato dalla probabilità stessa*. In altre parole, se diciamo che due eventi sono equiprobabili, allora dobbiamo essere pronti ad accettare scommesse 1:1; in caso contrario la nostra assegnazione di probabilità non è coerente. (Riflettete: si tratta di una formulazione meno banale di quanto non possa sembrare a prima vista.) La cosa interessante è che, una volta derivate, le regole della probabilità soggettiva sono le stesse di quella combinatoriale e frequentista—essenzialmente gli assiomi di Kolmogorov che abbiamo visto nella sezione 3.1 (con la differenza che, in questo, schema, esse non sono assiomi ma, appunto, regole che si derivano formalmente da un principio più fondamentale).

La definizione soggettiva di probabilità ed il teorema di Bayes, che vedremo nella sezione 3.5, sono alla base della cosiddetta scuola Bayesiana, che tradizionalmente si contrappone (e spesso in modo veemente) alla scuola classica o frequentista. Il livello della nostra esposizione sarà abbastanza rudimentale da poter evitare quasi completamente questa contrapposizione, ma vale la pena sottolineare come l'impostazione Bayesiana permetta di derivare entro un quadro logico e coerente, e sotto condizioni ben definite, tutte le applicazioni elementari della probabilità e della statistica che deriveremo nel seguito. Per una introduzione pedagogica all'argomento si rimanda a [3].

### 3.2.4 Uno spunto di riflessione: esistono monete e dadi equi?

Fino a questo momento abbiamo parlato disinvoltamente di monete eque e dadi equi come se la probabilità di uscita di una data faccia (ad esempio, testa per una moneta o il numero 3 per un dado) fosse una proprietà intrinseca dell'oggetto. A pensarci meglio si tratta di una assunzione non banale che merita una riflessione più approfondita.

Trattandosi di sistemi fisici macroscopici, il lancio di una moneta o di un dado sono fenomeni regolati da leggi deterministiche: se conoscessimo esattamente le condizioni iniziali del lancio (posizione, velocità iniziale del centro di massa, asse istantaneo di rotazione e momento angolare) e tutti i fattori che determinano l'evoluzione temporale (forza di gravità, attrito dell'aria, vincoli fisici) potremmo in linea di principio predire con esattezza l'esito di ciascun lancio. Nella realtà quando lanciamo una moneta o un dado non misuriamo nessuna di queste quantità—e anche se volessimo, sarebbe estremamente complicato farlo con l'accuratezza necessaria. È proprio questo il motivo per cui trattiamo con le leggi della probabilità un fenomeno che è intrinsecamente deterministico.

Detto questo: che cosa vuol dire che una moneta è equa? Per definizione vuol dire che  $P(T) = P(C) = 1/2$ . Ma su quale base assumiamo che questa sia una proprietà della moneta e non dipenda, ad esempio, da *come* la lanciamo? Siamo sicuri che la stessa moneta, assumendo che sia equa se lanciata da una certa persona, continui ad essere equa anche quando lanciata da una persona diversa? Siamo sicuri che la stessa moneta, assumendo che sia equa quando lasciata cadere da 1 m di altezza, continui ad essere equa se lasciata cadere (dalla stessa persona) da 10 cm di altezza?

Se ci pensiamo per un secondo, la risposta più logica a tutte queste domande è: no. È noto che esistono bari ai tavoli da gioco, per cui è chiaro che, con un po' di pratica, si può imparare a lanciare un dado in

modo che le probabilità di uscita delle singole facce siano significativamente differenti da  $1/6$ . Nel seguito, dunque, quando parleremo di monete e dadi equi, avremo bene in mente che si tratta di astrazioni (sia pure non irragionevoli) che facciamo a scopo illustrativo, ma che una definizione operativa rigorosa di questi concetti richiederebbe un'analisi estremamente più approfondita di quella che possiamo fare in questo contesto.

### 3.3 DIGRESSIONE: ELEMENTI DI CALCOLO COMBINATORIO

Il calcolo combinatorio è quella branca della matematica che si occupa del *contare*. Si tratta di un argomento ovviamente connesso con la definizione combinatoriale di probabilità e in questa sezione introdurremo brevemente alcuni dei concetti fondamentali.

#### 3.3.1 Permutazioni e funzione fattoriale

Dato un numero intero non negativo  $n$  definiamo il suo fattoriale (che scriveremo come  $n!$ ) come il prodotto di tutti gli interi positivi minori o uguali ad  $n$

$$n! = \prod_{k=1}^n k = n(n-1) \dots 1, \quad (44)$$

con l'ulteriore convenzione che  $0! = 1$ . Equivalentemente, la funzione fattoriale può essere definita per ricorrenza come

$$\begin{cases} 0! = 1 \\ n! = n(n-1)! \end{cases} \quad (45)$$

Il fattoriale  $n!$  rappresenta il numero di permutazioni di un insieme di  $n$  elementi, cioè il numero di modi in cui si possono disporre  $n$  elementi. La convenzione  $0! = 1$  è consistente con il fatto che vi è un solo modo di permutare zero oggetti.

► **ESEMPIO 3.12.** In quanti modi diversi possiamo disporre 6 cd in un caricatore a 6 posti di un autoradio? Possiamo ragionare come segue: siamo liberi di mettere un cd qualsiasi tra i 6 da caricare nella posizione 1, dopodiché ci rimangono 5 cd tra cui scegliere per la posizione 2, 4 per la posizione 3 e così via. Giunti all'ultimo cd, non abbiamo più scelta perché l'unica posizione libera è la sesta. La risposta è dunque  $6! = 720$ .

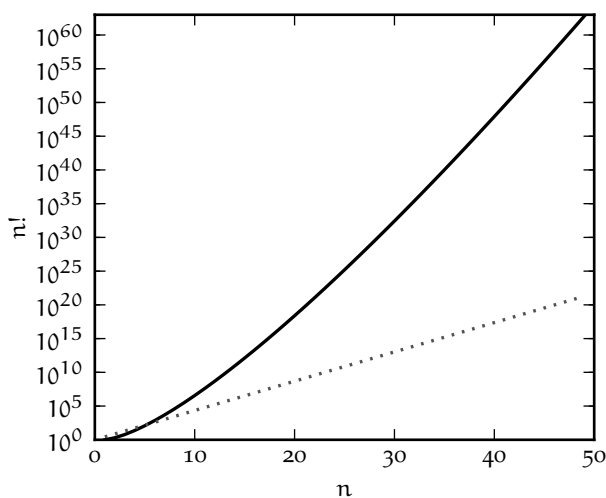


FIGURA 3.3. Grafico della funzione  $n!$  per  $n = 0 \dots 50$ . Per confronto, la linea grigia tratteggiata rappresenta la funzione  $y = e^n$ , che in scala semilogaritmica è una retta. Come è ovvio, la funzione fattoriale cresce più che esponenzialmente al crescere di  $n$ .

La funzione fattoriale cresce molto velocemente al crescere di  $n$ —molto più velocemente, e.g., di un esponenziale, come illustrato in figura 3.3. Tanto per fare un esempio,  $50! \approx 3 \times 10^{64}$ . Per  $n$  grande, il fattoriale può essere approssimato con la formula di Stirling

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n, \quad (46)$$

o, equivalentemente

$$\ln n! \approx \frac{1}{2} \ln(2\pi n) + n \ln n - n. \quad (47)$$

Per fissare le idee, l'approssimazione (46) è accurata a meglio dell'1% per  $n \geq 10$ , e l'errore relativo diminuisce al crescere di  $n$ . Dato che il logaritmo è una funzione che per grandi valori del suo argomento varia piuttosto lentamente, lo sviluppo del  $\ln n!$  può, a seconda della situazione, essere utilizzato trascurando il primo termine, visto che  $\ln n \ll n$ —cioè

$$\ln n! \approx n \ln n - n. \quad (48)$$

(Più precisamente: nello sviluppo di  $\ln n!$  (47) il primo termine è una costante additiva che diventa irrilevante nel limite  $n \rightarrow \infty$ , mentre nello sviluppo di  $n!$  (46) esso corrisponde alla costante moltiplicativa  $\sqrt{2\pi n}$  che, ovviamente, non può essere omessa.)

Il fattoriale è, come è noto, utilizzato nello sviluppo in serie di Taylor. In particolare, sviluppando in serie la funzione  $e^x$  si ottiene la relazione interessante

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad (49)$$

da cui anche

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} = 2 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} \dots \quad (50)$$

### 3.3.2 Coefficienti binomiali

Il numero di modi possibili per scegliere  $k$  elementi non ordinati da un insieme di  $n$  elementi è dato dal *coefficiente binomiale*  $n$  su  $k$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\dots(n-k+1)}{k(k-1)(k-2)\dots 1}. \quad (51)$$

In breve, il ragionamento procede nel modo seguente: per la scelta del primo dei  $k$  elementi abbiamo  $n$  possibilità, per il secondo  $n-1$ , per il terzo  $n-2$  e così via fino a  $n-(k-1)$ ; il  $k!$  al denominatore serve per non contare più di una volta scelte che differiscono solo per l'ordine.

Il coefficiente binomiale è strettamente connesso al triangolo di Pascal e può anche essere caratterizzato attraverso la formula della potenza di binomio (o teorema binomiale), che utilizzeremo nel seguito:

$$(x_1 + x_2)^n = \sum_{k=0}^n \binom{n}{k} x_1^k x_2^{n-k}. \quad (52)$$

Nel caso in cui  $x_1 = x_2 = 1$ , la (52) si riduce alla formula notevole

$$\sum_{k=0}^n \binom{n}{k} = 2^n \quad (53)$$

in cui, nel membro a sinistra, si riconosce il numero di elementi dell'insieme di potenza di un generico insieme di  $n$  elementi—scritto come la somma del numero di sottoinsiemi di  $0, 1 \dots n$  elementi. Così, come abbiamo visto negli esempi 3.1 e 3.2, lo spazio degli eventi corrispondenti al lancio di una moneta ha  $2^2 = 4$  elementi e quello corrispondente al lancio di un dado ha  $2^6 = 64$  elementi.

► ESEMPIO 3.13. Dato un insieme di tre elementi  $\{A, B, C\}$ , vi sono esattamente tre sottoinsiemi non ordinati di 2 elementi:  $\{A, B\}$ ,  $\{A, C\}$  e  $\{B, C\}$  ( $\{A, B\}$  e  $\{B, A\}$  sono lo stesso insieme). La risposta si può scrivere anche come

$$\binom{3}{2} = \frac{3!}{2!1!} = \frac{6}{2} = 3.$$

► ESEMPIO 3.14. Quanti *cin-cin* ci sono in un brindisi fra 10 amici? Possiamo ragionare come segue: date  $n$  persone, ognuna brinda con le altre  $n - 1$ , ma dobbiamo poi dividere per 2 per non contare doppio ogni singolo *cin-cin*. La risposta è dunque  $n(n - 1)/2$  o, nel nostro caso, 45. In termini dei coefficienti binomiali, la risposta è semplicemente

$$\binom{10}{2} = \frac{10!}{2!8!} = \frac{10 \times 9}{2} = 45.$$

### 3.3.3 Una applicazione elementare: il gioco del lotto

Come applicazione elementare dei coefficienti binomiali ci chiediamo quale sia la probabilità di vincere giocando un ambo secco al lotto su una singola ruota—ad esempio quella di Napoli. Il calcolo è leggermente più complicato di quelli che abbiamo visto sino ad ora, ma gli ingredienti sono essenzialmente due: il numero di ambi (cioè di coppie non ordinate) possibili con 90 numeri

$$\binom{90}{2} = \frac{90 \times 89}{2} = 4005$$

(che sono *tutti* gli ambi possibili) ed il numero di ambi generato dai 5 numeri estratti

$$\binom{5}{2} = \frac{5 \times 4}{2} = 10.$$

La probabilità che il nostro ambo sia estratto è dunque data da

$$\binom{5}{2} / \binom{90}{2} = \frac{10}{4005} \approx 0.25\%,$$

a fronte della quale il lotto paga un ambo secco 250 : 1, cioè poco più della metà del valore 400.5 : 1 che renderebbe il gioco equo. Il calcolo si può generalizzare facilmente a terne, quaterne e cinquine, come mostrato nella tabella che segue.

Giocata	Probabilità di vincita	Premio
Ambo	1 : 401	250 : 1
Terno	1 : 11 748	4 500 : 1
Quaterna	1 : 511 038	120 000 : 1
Cinquina	1 : 43 949 268	6 000 000 : 1

TABELLA 7. Probabilità di vincita e premio corrisposto per alcune possibili giocate al lotto (da intendersi su una singola ruota). Come è ovvio, il rapporto tra questi due numeri è sempre sfavorevole per il giocatore.

Il gioco è sempre sfavorevole per il giocatore, con un rapporto che va da meno di 2 per l'ambo a più di 7 per la cinquina.

### 3.3.4 Coefficienti multinomiali



Il concetto di coefficiente binomiale si può generalizzare in modo naturale all'idea della suddivisione di un insieme di  $n$  elementi in  $m$  sottoinsiemi disgiunti, ciascuno con un numero  $k_i$  di elementi, la cui unione costituisca l'insieme di partenza. È ovvio che, sotto queste condizioni i  $k_i$  soddisfano la relazione

$$\sum_{i=1}^m k_i = n.$$

Il numero di modi in cui si può effettuare tale suddivisione prende il nome di *coefficiente multinomiale*  $n$  su  $k_1, \dots, k_m$  e non è difficile da calcolare—essenzialmente si hanno  $n$  su  $k_1$  (binomiale) modi per scegliere il primo sottoinsieme, dopo di che si hanno  $n - k_1$  su  $k_2$  (binomiale) modi di scegliere il secondo e così via fino all'ultimo

$$\begin{aligned} \binom{n}{k_1, \dots, k_m} &= \binom{n}{k_1} \binom{n-k_1}{k_2} \dots \binom{n-k_1-\dots-k_{m-1}}{k_m} = \\ &= \frac{n!}{k_1!(n-k_1)!} \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} \dots \frac{(n-k_1-\dots-k_{m-1})!}{k_m!0!} = \frac{n!}{k_1!k_2!\dots k_m!}. \end{aligned} \quad (54)$$

► **ESEMPIO 3.15.** In quanti modi si possono disporre 20 invitati ad un banchetto di matrimonio su tre tavoli, rispettivamente da 4, 6 e 10 persone? La risposta è

$$\binom{20}{4, 6, 10} = \frac{20!}{4!6!10!} = 38798760$$

ossia più di 38 milioni di modi distinti.

Per inciso, notiamo che i coefficienti multinomiali permettono di scrivere in modo compatto la formula per la potenza della somma di un numero arbitrario di monomi attraverso il cosiddetto teorema multinomiale, che è la generalizzazione del teorema binomiale (52)

$$(x_1 + x_2 + \dots + x_m)^n = \sum \binom{n}{k_1, \dots, k_m} x_1^{k_1} x_2^{k_2} \dots x_m^{k_m}, \quad (55)$$

in cui la somma è estesa a tutte le sequenze  $(k_1, k_2, \dots, k_m)$  di interi non negativi la cui somma sia  $n$ .

### 3.4 PROBABILITÀ CONDIZIONATA

Dati due eventi  $E_1$  ed  $E_2$ , con  $P(E_2) \neq 0$ , definiamo la probabilità *condizionata*  $P(E_1 | E_2)$  di  $E_1$  dato  $E_2$  (cioè la probabilità che si verifichi l'evento  $E_1$  nel caso sappiamo già che si è verificato l'evento  $E_2$ ) come

$$P(E_1 | E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}. \quad (56)$$

Intuitivamente, come si vede in figura 3.1, la probabilità condizionata  $P(E_1 | E_2)$  rappresenta la misura dell'intersezione tra  $E_1 \cap E_2$  *relativamente* ad  $E_2$ . Si può dimostrare banalmente che la probabilità condizionata soddisfa gli assiomi della probabilità—e di fatto ogni probabilità  $P(E)$  può essere vista come una probabilità condizionata  $P(E | \Omega)$ .

► **ESEMPIO 3.16.** Qual è la probabilità che in un lancio di un dado equo a sei facce esca il numero 2, condizionata al fatto che l'uscita sia pari? Intuitivamente risponderemmo  $1/3$ , poiché ci sono tre numeri pari tra 1 e 6—ossia  $\{2, 4, 6\}$ . Formalmente dobbiamo applicare la (56);  $\{2\} \cap \{2, 4, 6\} = \{2\}$  per cui il numeratore è  $1/6$ , mentre il denominatore è  $P(\text{pari}) = 1/2$ , da cui segue che la nostra probabilità condizionata è effettivamente  $1/3$ .

► **ESEMPIO 3.17.** In un mazzo da 52 carte la probabilità di estrarre un re condizionata ad aver estratto una carta di cuori è  $P(K | \heartsuit) = 1/13$ . Di converso la probabilità di estrarre una carta di cuori condizionata all'aver estratto un re è  $P(\heartsuit | K) = 1/4$ .

Il concetto di probabilità condizionata è utile, operativamente, per calcolare la probabilità che due eventi si verifichino contemporaneamente (cioè per calcolare la misura della loro intersezione): la (56) si riscrive banalmente come

$$P(E_1 \cap E_2) = P(E_1 | E_2) P(E_2) = P(E_2 | E_1) P(E_1) \quad (57)$$

(dove la doppia uguaglianza deriva banalmente dal fatto che  $E_1 \cap E_2 = E_2 \cap E_1$ .)



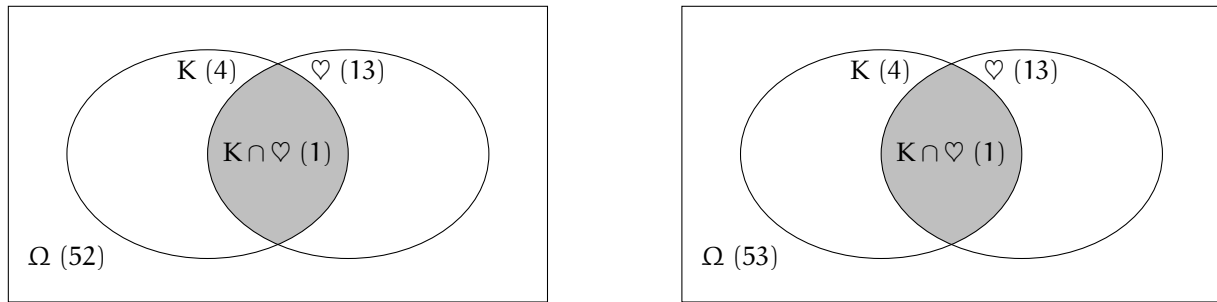


FIGURA 3.4. Diagrammi di Venn relativi agli esempi che illustrano l'indipendenza tra eventi. I numeri tra parentesi indicano il numero di elementi che compongono gli insiemi corrispondenti e consentono di calcolare in modo immediato le probabilità condizionate necessarie. I diagrammi illustrano come la vera differenza causata dall'aggiunta del jolly sia l'aumentare il numero di elementi dello spazio campionario da 52 a 53.

### 3.4.1 Eventi indipendenti: un esempio elementare

Si dice che due eventi  $E_1$  ed  $E_2$  sono *indipendenti* se il fatto che si sia verificato  $E_2$  non influenza la probabilità che si verifichi  $E_1$  (e viceversa), cioè se

$$P(E_1 | E_2) = P(E_1) \quad \text{e} \quad P(E_2 | E_1) = P(E_2). \quad (58)$$

Per la definizione di probabilità condizionata (57) la condizione di indipendenza tra eventi si può riscrivere nella forma più espressiva

$$P(E_1 \cap E_2) = P(E_1) P(E_2). \quad (59)$$

(Notiamo, per inciso, che il lettore non deve confondere la nozione di eventi indipendenti con quella di eventi disgiunti—eventi, cioè, che hanno intersezione nulla in  $\Omega$ . In quel caso gli eventi si dicono, come abbiamo già avuto modo di dire, incompatibili.)

Il concetto di indipendenza di eventi è più sottile di quanto non possa sembrare in apparenza—cosa che illustriamo con un semplice esempio. Consideriamo il nostro solito mazzo di 52 carte. Come sappiamo la probabilità di estrarre un re è  $P(K) = 1/13$  e la probabilità di estrarre una carta di cuori è  $P(H) = 1/4$ . Qual è la probabilità di estrarre il re di cuori? Sappiamo già la risposta, perché esiste un solo re di cuori nel mazzo per cui la probabilità di estrarlo è  $P(K \cap H) = 1/52$ . Il fatto che

$$P(K) P(H) = \frac{1}{13} \times \frac{1}{4} = \frac{1}{52} = P(K \cap H)$$

dimostra che i due eventi (“la carta estratta è un re” e “la carta estratta è di cuori”) sono indipendenti.

Proviamo adesso a complicare appena il nostro problema ed aggiungiamo un jolly al nostro mazzo (che diventa così di 53 carte). La probabilità di estrarre un re diviene  $P(K) = 4/53$  mentre quella di estrarre una carta di cuori è adesso  $P(H) = 13/53$ . La probabilità di estrarre il re di cuori è banalmente  $P(K \cap H) = 1/53$  per cui

$$P(K) P(H) = \frac{4}{53} \times \frac{13}{53} = \frac{52}{2809} \neq P(K \cap H),$$

cioè i nostri due eventi non sono più indipendenti!

Che cosa è successo? Se analizziamo i due eventi più da vicino notiamo che nel primo caso: se estraiamo una carta di cuori la probabilità che essa sia un re è  $1/13$ ; se la carta estratta *non* è di cuori, allora la probabilità che essa sia un re è  $3/39$  (ci sono  $13 \times 3 = 39$  carte non di cuori e tra queste 3 sono re), cioè di nuovo  $1/13$ —i due eventi sono indipendenti. Nel secondo caso: se estraiamo una carta di cuori la probabilità che essa sia un re è, come prima,  $1/13$ ; se però la carta estratta *non* è di cuori, allora la probabilità che essa sia un re è  $3/40 \neq 1/13$ . In altre parole, se sappiamo che la carta estratta è di fiori, allora sappiamo che non può essere un jolly, e questo cambia la probabilità che essa sia un re—i due

eventi non sono indipendenti. Ma se riesaminiamo brevemente l'esempio del mazzo di 53 carte alla luce di quanto sappiamo sulla probabilità condizionata, tutto diventa perfettamente consistente:

$$P(K \cap \heartsuit) = \begin{cases} P(K | \heartsuit) P(\heartsuit) = \frac{1}{13} \times \frac{13}{53} = \frac{1}{53} \\ P(\heartsuit | K) P(K) = \frac{1}{4} \times \frac{4}{53} = \frac{1}{53} \end{cases}$$

### 3.5 IL TEOREMA DI BAYES

La (57) ha come conseguenza immediata e banale il cosiddetto teorema di Bayes, che lega tra loro le probabilità condizionate  $P(E_1 | E_2)$  e  $P(E_2 | E_1)$

$$P(E_1 | E_2) = \frac{P(E_2 | E_1) P(E_1)}{P(E_2)}. \quad (60)$$

Sfruttando il fatto che  $E_1 \cup \overline{E_1} = \Omega$ , la (60) si può scrivere anche nella forma esplicita (ma leggermente meno compatta)

$$P(E_1 | E_2) = \frac{P(E_2 | E_1) P(E_1)}{P(E_2 | E_1) P(E_1) + P(E_2 | \overline{E_1}) P(\overline{E_1})}. \quad (61)$$

Più in generale, se si dispone di una partizione dello spazio campionario in un insieme  $\{E_i\}$  di insiemi disgiunti ( $E_i \cap E_j = \emptyset$ ) la cui unione coincida con lo spazio stesso ( $\cup_i E_i = \Omega$ ), allora il teorema di Bayes si può anche esprimere nella forma

$$P(E_1 | E_2) = \frac{P(E_2 | E_1) P(E_1)}{\sum_i P(E_2 | E_i) P(E_i)}. \quad (62)$$

La (60) è di fondamentale importanza perché collega un problema di probabilità diretta al problema corrispondente di probabilità inversa. Torniamo per un attimo alla nostra moneta ideale. Se sappiamo esattamente la probabilità  $p$  che esca testa (ad esempio  $p = 0.5$ ) in un singolo lancio è facile calcolare la probabilità di ottenere esattamente  $n$  teste in  $N$  lanci—si tratta della distribuzione binomiale che vedremo nella sezione 5.1. Ma non è questo che fa il Fisico, nella sua professione. Tutto il contrario: il Fisico lancia la moneta  $N$  volte e, a partire dal numero di volte  $n$  in cui esce testa, cerca di inferire la probabilità (incognita)  $p$ . È chiaro che i due problemi sono legati tra di loro, ed in questa sezione cominceremo a sviscerare questo collegamento.

#### 3.5.1 Un semplice problema di probabilità inversa

A questo punto in un qualsiasi articolo o libro di statistica trovereste il famoso problema del test per l'HIV [3]. Noi preferiamo invece una variante meno cruenta—quella del rilevatore automatico di banconote false. Le specifiche fornite dalla casa produttrice dicono che l'oggetto ha un'efficienza del 100% nel segnalare banconote contraffatte ma, allo stesso tempo, ha un tasso di falsi positivi del 5%—cioè nel 5% dei casi segnala come contraffatta una banconota che invece è autentica. (In fondo non esistono dispositivi perfetti.) Prendiamo dunque una banconota e la passiamo dal rilevatore, che la segnala come contraffatta. Qual è la probabilità che la banconota sia effettivamente falsa?

Prima di rispondere 95% (come uno potrebbe ingenuamente essere tentato di fare), fermiamoci per un attimo a pensare e convinciamoci che la domanda, così come è scritta, è mal posta. Manca un ingrediente fondamentale, ossia la probabilità *a priori* che, presa una banconota a caso, essa sia contraffatta. Dopo tutto in un mondo ideale in cui la contraffazione non esistesse (e quindi non esistessero banconote false) se il nostro dispositivo scattasse non potremmo far altro che imputare il fatto al 5% dei falsi positivi, non vi pare? (E in quel caso la probabilità che cerchiamo sarebbe identicamente nulla.)

Torniamo dunque un attimo indietro e cerchiamo di formalizzare meglio il problema. Per snellire la notazione definiamo i seguenti eventi:

- A: la banconota è autentica
- C: la banconota è contraffatta
- ✓: il rilevatore segnala la banconota come autentica
- ✗: il rilevatore segnala la banconota come contraffatta

Notiamo che questi eventi sono a due a due disgiunti e partizionano in due modi diversi il nostro spazio degli eventi  $\Omega$ , che possiamo vedere come il prodotto cartesiano dei due insiemi  $\{A, C\}$  e  $\{\checkmark, \times\}$

$$\Omega = \{(A, \checkmark), (A, \times), (C, \checkmark), (C, \times)\}.$$

Le specifiche del costruttore si possono allora riscrivere (e precisare) come

$$\begin{aligned} P(\times | C) &= 1 \\ P(\times | A) &= 0.05 \end{aligned}$$

Assumeremo inoltre che la probabilità a priori che una banconota sia contraffatta sia  $P(C) = 10^{-3}$  (cioè che una banconota su mille in circolazione sia, in media, falsa).

Torniamo dunque al nostro problema iniziale, cioè la stima della probabilità che una banconota segnalata come contraffatta dal nostro rilevatore sia effettivamente contraffatta. Per la (60) possiamo scrivere

$$P(C | \times) = \frac{P(\times | C) P(C)}{P(\times)}.$$

L'unica cosa che ci manca per calcolare la probabilità cercata è  $P(\times)$ , che non abbiamo direttamente, ma che possiamo calcolare sfruttando il fatto che  $A \cup C = \Omega$

$$P(\times) = P(\times \cap A) + P(\times \cap C) = P(\times | A) P(A) + P(\times | C) P(C).$$

Questo ci permette di scrivere la risposta cercata nella forma meno compatta ma esplicita (che è anche la forma in cui si trova spesso espresso in letteratura il teorema di Bayes)

$$P(C | \times) = \frac{P(\times | C) P(C)}{P(\times | A) P(A) + P(\times | C) P(C)} = \frac{1 \times 0.001}{0.05 \times 0.999 + 1 \times 0.001} \approx 2\%$$

Sorprendentemente (?) la probabilità cercata è dell'ordine del 2%, nonostante l'efficienza del nostro rilevatore sia il 100% ed il tasso di falsi positivi sia solo del 5%. Il commento sopra a proposito del mondo ideale (senza contraffazione) può essere qui precisato notando che se  $P(C) = 0$ , allora  $P(C | \times) = 0$  come anticipato (vedi figura 3.5).

(Pensate un attimo alla vostra esperienza di tutti i giorni: perché il cassiere del supermercato non chiama immediatamente la sicurezza ogni volta che una banconota viene segnalata come contraffatta dal rilevatore? Quante volte vi è capitato di vedere la stessa banconota segnalata come autentica al secondo o terzo passaggio?)

### 3.5.2 Ancora sul teorema di Bayes: un po' di nomenclatura

Torniamo per un attimo sul teorema di Bayes in un contesto più vicino al mestiere del Fisico. Supponiamo di aver raccolto un campione di dati  $d$  e di avere un modello del nostro sistema fisico dipendente da un parametro  $\theta$ , che vogliamo stimare a partire dai dati stessi. Possiamo riscrivere la (60) come

$$P(\theta | d) = \frac{P(d | \theta) P(\theta)}{P(d)}. \quad (63)$$

Il primo termine dell'equazione  $P(\theta | d)$  è detto *probabilità a posteriori*, o *posterior*, ed è esattamente ciò a cui siamo interessati da Fisici: la probabilità (inversa) che il parametro del nostro modello assuma un

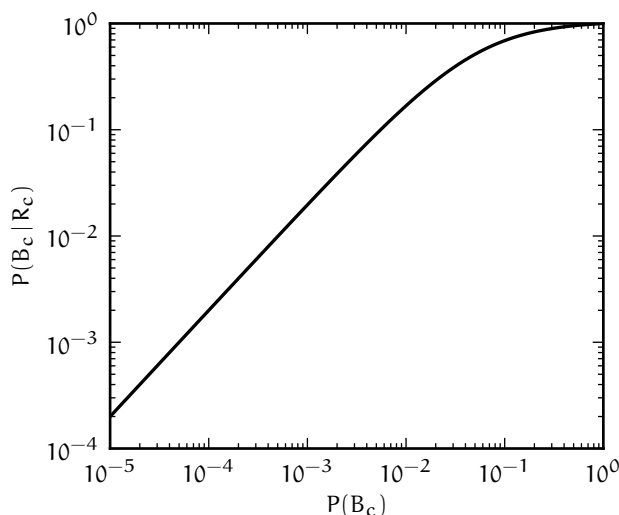


FIGURA 3.5. Probabilità condizionata  $P(C|X)$  che la banconota segnalata dal nostro rilevatore sia effettivamente contraffatta in funzione della probabilità a priori che una banconota scelta a caso tra quelle in circolazione sia contraffatta  $P(C)$ , dati  $P(X|C) = 1$  e  $P(X|A) = 0.05$ . Per  $P(C) = 10^{-3}$  si ha il  $\approx 2\%$  calcolato sopra.

certo valore, date le nostre misure.  $P(d|\theta)$ , che si dice *verosimiglianza* o *likelihood*, è invece la probabilità (diretta) di misurare  $d$ , dato un generico valore di  $\theta$ , ed in pratica è specificata completamente dal nostro modello.  $P(\theta)$  prende il nome di *probabilità a priori*, o *prior*, ed è il termine del teorema di Bayes su cui le diverse scuole statistiche discutono più veementemente.

E  $P(d)$ ? A prima vista potrebbe sembrare assurdo parlare della probabilità che le nostre misure  $d$  assumano una serie di valori dati—in fondo una volta che le abbiamo fatte le misure sono fissate, no? (Ma ricordatevi: nello schema soggettivista possiamo assegnare una probabilità a qualsiasi evento o proposizione, per cui questa non è una difficoltà insuperabile.) Ma se guardiamo meglio la (63) la cosa fondamentale è che stiamo facendo una affermazione (probabilistica) su  $\theta$ , mentre il denominatore del membro di destra dell'equazione dipende solo da  $d$  per cui, dal nostro punto di vista, esso è solo una costante di normalizzazione. Allora la (63) può essere riscritta in modo espressivo nella forma

$$\text{probabilità a posteriori} \propto \text{verosimiglianza} \times \text{probabilità a priori}, \quad (64)$$

che si trova frequentemente in letteratura e che utilizzeremo, sia pure solo occasionalmente, nel seguito.

### 3.6 VARIABILI CASUALI E FUNZIONI DI DISTRIBUZIONE

Si dice *variabile casuale* o *variabile aleatoria* una variabile che rappresenta la realizzazione numerica di un processo casuale, per cui il suo valore è soggetto a fluttuazioni casuali e non è noto a priori. Esistono casi di variabili casuali discrete, cioè variabili che possono assumere un numero finito o numerabile di valori, e casi di variabili continue, ed entrambe possono essere definite su intervalli finiti o infiniti.

► **ESEMPIO 3.18.** L'uscita del lancio di un dado a sei facce (ossia la faccia del dado rivolta verso l'alto) è una variabile casuale discreta che può assumere esattamente sei valori:  $1 \dots 6$ .

► **ESEMPIO 3.19.** Il tempo necessario per arrivare da casa al luogo di lavoro è un esempio di variabile casuale continua.

#### 3.6.1 Variabili casuali discrete

Consideriamo una variabile discreta  $x$  che può assumere  $n$  valori distinti possibili  $x_1 \dots x_n$ . Indicando con  $P(x_k)$  la probabilità che  $x$  assuma il valore  $x_k$ , definiamo *funzione di distribuzione*<sup>1</sup> di  $x$  (o semplicemente

<sup>1</sup> Il termine "funzione di distribuzione" è talvolta usato in letteratura per indicare la funzione cumulativa di una distribuzione, di cui ci occuperemo nella sezione 3.10. Un altro termine molto usato per indicare quello che noi chiamiamo funzione di distribuzione nel caso continuo è l'Inglese *probability density function*, che si trova anche abbreviato come *pdf*.

distribuzione di  $x$ ) la funzione che associa ad ogni valore  $x_k$  della variabile  $x$  la sua probabilità  $P(x_k)$ . La funzione di distribuzione per una variabile discreta può essere rappresentata in forma tabellare o, più efficacemente, con un grafico a barre, come mostrato in figura 3.6.

► **ESEMPIO 3.20.** Sia  $x$  l'uscita del lancio di un dado a sei facce. Se il dado è equo, ciascuno dei valori  $x_k = 1 \dots 6$  ha probabilità  $1/6$  di uscire. La funzione di distribuzione corrispondente è mostrata in figura 3.6 (a sinistra).

► **ESEMPIO 3.21.** Consideriamo il lancio di due dadi e sia  $x$  la somma delle due uscite.  $x$  può assumere tutti i valori interi tra 2 e 12, ma stavolta gli  $x_k$  non sono equiprobabili anche se i dadi sono equi. I casi possibili sono in totale  $6 \times 6 = 36$  ed ognuno può essere univocamente identificato con la coppia  $(u_1, u_2)$ , in cui  $u_1$  è l'uscita del primo dado e  $u_2$  quella del secondo. Mentre il 2 ha solamente un modo per realizzarsi,  $(1, 1)$ , il 4, ad esempio, può realizzarsi in tre modi diversi:  $(1, 3)$ ,  $(2, 2)$  o  $(3, 1)$ , come illustrato nella tabella 8. La probabilità di uscita può allora essere calcolata come rapporto tra il numero di casi favorevoli ed il numero totale di casi possibili, ed ha la forma triangolare mostrata in figura 3.6 (a destra).

$x$	Combinazioni possibili	$P(x)$
2	(1, 1)	$1/36$
3	(1, 2), (2, 1)	$2/36$
4	(1, 3), (3, 1), (2, 2)	$3/36$
5	(1, 4), (4, 1), (2, 3), (3, 2)	$4/36$
6	(1, 5), (5, 1), (2, 4), (4, 2), (3, 3)	$5/36$
7	(1, 6), (6, 1), (2, 5), (5, 2), (4, 3), (3, 4)	$6/36$
8	(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)	$5/36$
9	(3, 6), (6, 3), (4, 5), (5, 4)	$4/36$
10	(4, 6), (6, 4), (5, 5)	$3/36$
11	(5, 6), (6, 5)	$2/36$
12	(6, 6)	$1/36$

TABELLA 8. Funzione di distribuzione per la somme delle uscite di due dadi equi. Per completezza tutte le combinazioni di uscite che danno luogo ad un dato valore della somma sono elencate esplicitamente.

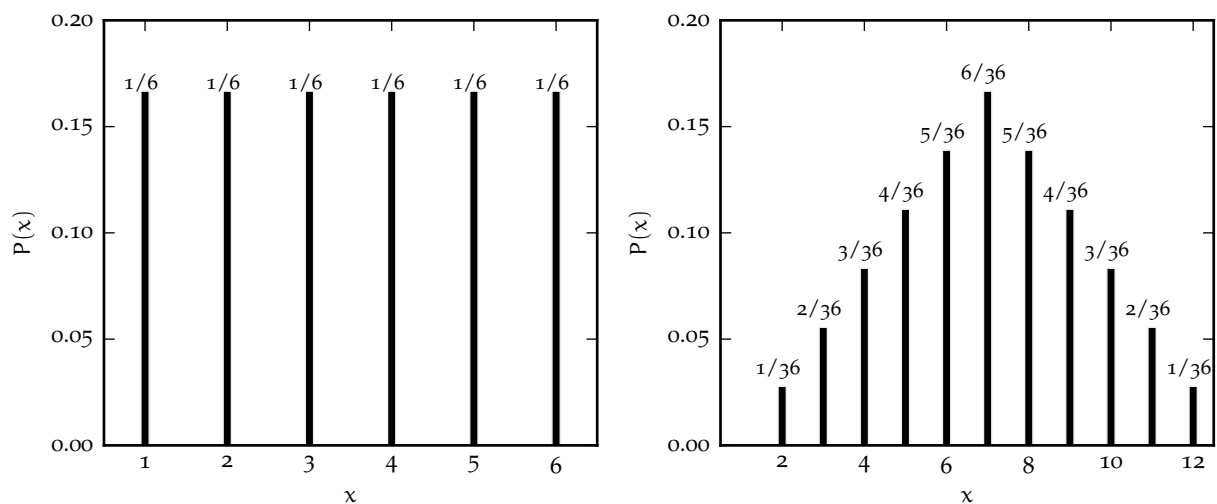


FIGURA 3.6. Funzione di distribuzione per l'uscita di un dado equo a sei facce (a sinistra) e per la somma delle uscite di due dadi equi a sei facce (a destra)—vedi gli esempi 3.20 e 3.21.

In questo contesto il secondo assioma di Kolmogorov si scrive nella forma di una *condizione di normalizzazione*, che tutte le funzioni di distribuzione debbono soddisfare

$$\sum_k P(x_k) = 1 \quad (65)$$

(la sommatoria è estesa a tutti i valori  $x_k$  che la variabile  $x$  può assumere). È banale verificare direttamente che le funzioni di distribuzione illustrate negli esempi 3.20 e 3.21 ed in figura 3.6 sono normalizzate.

### 3.6.2 Variabili casuali continue

Nel caso di una variabile continua la definizione di funzione di distribuzione data per una variabile discreta non è più applicabile poiché la probabilità che la variabile  $x$  assuma un valore *esattamente* definito è identicamente zero (vedremo tra un attimo che si tratta essenzialmente di un integrale su un dominio di misura nulla). È invece sensato chiedersi quale sia la probabilità che la  $x$  assuma un valore compreso in un generico intervallo compreso tra  $x_0$  ed  $x_0 + dx$

$$P(x_0, dx) = P(x_0 \leq x < x_0 + dx).$$

A questo punto, se dividiamo per la larghezza dell'intervallo e prendiamo il limite per  $dx \rightarrow 0$ , otteniamo una sorta di probabilità specifica o probabilità per unità di intervallo che chiamiamo *densità di probabilità*

$$p(x_0) = \lim_{dx \rightarrow 0} \frac{P(x_0, dx)}{dx} \quad (66)$$

La (66) è una sorta di rapporto incrementale, per cui la densità di probabilità è, in un certo senso, la derivata della funzione probabilità. La probabilità (infinitesima) che la variabile casuale assuma un valore compreso nell'intervallo  $dx$  centrato attorno al valore  $x_0$  si scrive come

$$P(x_0, dx) = p(x_0)dx$$

e, corrispondentemente, la probabilità che essa assuma un valore compreso nell'intervallo *finito*  $[x_1, x_2]$  è data da:

$$P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} p(x)dx. \quad (67)$$

(Ne segue, per inciso, che la densità di probabilità di una variabile casuale continua  $x$  ha le dimensioni fisiche dell'inverso di  $x$ .) La condizione di normalizzazione per una funzione di distribuzione di variabile continua si scrive allora come

$$\int_{-\infty}^{\infty} p(x)dx = 1. \quad (68)$$

(Formalmente l'integrale andrebbe calcolato sul supporto della funzione di distribuzione ma, se assumiamo che la funzione di distribuzione sia identicamente nulla al di fuori del supporto stesso, possiamo integrare su tutta la retta reale senza perdere di generalità.)

► **ESEMPIO 3.22.** Il più semplice esempio di funzione di distribuzione di variabile continua è quello in cui la densità di probabilità è costante su un dato intervallo  $[a, b]$  (vedi figura 3.7 e sezione 5.5). La normalizzazione è fissata dalla (68) e vale  $1/(b - a)$ .

## 3.7 VALORE DI ASPETTAZIONE

Sia data una funzione  $f(x)$  di una variabile casuale  $x$  (continua o discreta). Definiamo il *valore di aspettazione*, o *speranza matematica*, o semplicemente *speranza* di  $f(x)$  come

$$E[f(x)] = \begin{cases} \sum f(x_k)P(x_k) & \text{per variabili discrete} \\ \int_{-\infty}^{\infty} f(x)p(x)dx & \text{per variabili continue.} \end{cases} \quad (69)$$

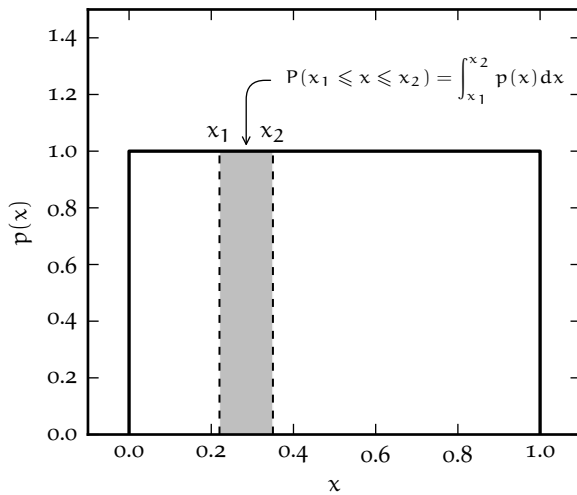


FIGURA 3.7. Esempio di funzione di distribuzione di variabile continua, in cui la densità di probabilità è costante tra 0 e 1. (Per ovvi motivi questa distribuzione prende il nome di distribuzione uniforme.) La distribuzione è correttamente normalizzata e la regione ombreggiata illustra il concetto di probabilità come integrale della densità di probabilità su un intervallo finito.

Notiamo esplicitamente che il valore di aspettazione trasforma una funzione di  $x$  in un *numero*. Si tratta di una sorta di procedura di media in cui i valori di  $f(x)$  sono pesati con il valore della funzione di distribuzione di  $x$ . È uno strumento fondamentale che, come vedremo in seguito, permette di definire molte delle proprietà di base comuni alle distribuzioni.

► **ESEMPIO 3.23.** Torniamo per un attimo alla sezione 3.3.3 e supponiamo di giocare 1 euro su un ambo secco sulla ruota di Napoli. Consideriamo la nostra vincita come una variabile casuale e calcoliamo il suo valore di aspettazione. La probabilità di vincere è  $10/4005$ , ed in tal caso il guadagno netto è  $250 - 1 = 249$  euro. In caso contrario semplicemente perdiamo 1 euro (che possiamo considerare una vincita negativa di  $-1$  euro). Il valore di aspettazione cercato è dunque

$$249 \times \frac{10}{4005} - 1 \times \frac{3995}{4005} \approx -0.38 \text{ euro},$$

che è come dire che in media perdiamo circa 38 centesimi per ogni euro giocato. (Come è ovvio dalla tabella nella sezione 3.3.3 la situazione è anche peggio per il terno, la quaterna e la cinquina.)

Per la linearità di integrale e sommatoria, il valore di aspettazione è un *operatore lineare*, nel senso che

$$E[c_1 f(x) + c_2 g(x)] = c_1 E[f(x)] + c_2 E[g(x)], \quad (70)$$

come si verifica facilmente per sostituzione. Notiamo anche che, per la condizione di normalizzazione, il valore di aspettazione di una *costante* (cioè di una espressione che non dipende dalla variabile casuale  $x$ ) è uguale alla costante stessa

$$E[c] = c. \quad (71)$$

### 3.8 TENDENZA CENTRALE E DISPERSIONE INTORNO ALLA MEDIA

Nelle applicazioni pratiche è comune condensare l'informazione contenuta nella funzione di distribuzione di una variabile casuale in pochi parametri significativi. Tipicamente siamo interessati a sapere quale valore ci aspettiamo che la variabile assuma *in media* e quanto la variabile stessa si discosti *in media* da questo valore.

## 3.8.1 Media, mediana e moda

Definiamo il *valor medio* (o semplicemente la *media*) di una variabile casuale  $x$  (continua o discreta) come il valore di aspettazione di  $x$

$$\mu = E[x] = \begin{cases} \sum x_k P(x_k) & \text{per variabili discrete} \\ \int_{-\infty}^{\infty} xp(x) dx & \text{per variabili continue.} \end{cases} \quad (72)$$

Nel seguito indicheremo il valor medio di  $x$  con  $\mu$ ,  $\mu_x$  o  $E[x]$  a seconda del contesto, cercando di conciliare sintesi e chiarezza. Notiamo esplicitamente che se  $c$  è una costante

$$E[cx] = cE[x]. \quad (73)$$

► **ESEMPIO 3.24.** Sia la variabile casuale  $x$  l'uscita di un dado equo a sei facce. La media vale

$$\mu = \sum_{k=1}^6 x_k P(x_k) = \frac{1}{6} \sum_{k=1}^6 k = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2}.$$

Notiamo esplicitamente che in questo caso il valor medio non coincide con nessuno dei valori che  $x$  può assumere.

► **ESEMPIO 3.25.** Ripetiamo l'esercizio precedente nel caso in cui  $x$  sia la somma delle uscite nel lancio di due dadi.

$$\begin{aligned} \mu &= \sum_{k=2}^{12} x_k P(x_k) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} + \\ &+ 7 \times \frac{6}{36} + 8 \times \frac{5}{36} + 9 \times \frac{4}{36} + 10 \times \frac{3}{36} + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} = \frac{252}{36} = 7 \end{aligned}$$

che è esattamente il doppio di quanto calcolato poco fa per un singolo dado. (Vedremo nel seguito che non si tratta di un caso.)

► **ESEMPIO 3.26.** Sia  $x$  una variabile casuale continua distribuita uniformemente tra 0 ed 1 (cioè  $p(x) = 1$  in  $[0, 1]$  e  $p(x) = 0$  fuori). La media vale:

$$\mu = \int_{-\infty}^{\infty} xp(x) dx = \int_0^1 x dx = \left. \frac{x^2}{2} \right|_0^1 = \frac{1}{2} - 0 = \frac{1}{2},$$

che coincide con il punto medio dell'intervallo  $[0, 1]$ .

Nel caso particolare di una variabile casuale discreta per cui si abbiano  $n$  uscite possibili  $x_k$  equiprobabili (cioè  $P(x_1) = \dots = P(x_n) = 1/n$ ) si ha

$$\mu = \sum_{k=1}^n x_k P(x_k) = \frac{1}{n} \sum_{k=1}^n x_k,$$

cioè la nostra definizione (72) di media coincide con la media aritmetica delle uscite possibili, come è ragionevole aspettarsi. Nel caso generale in cui i valori  $x_k$  non siano equiprobabili, essi vengono correttamente *pesati* in modo nella (72) a seconda della probabilità corrispondente di modo che la media si sposti verso i valori più probabili.

Il valor medio non è l'unica stima possibile di tendenza centrale. Definiamo *mediana* di una distribuzione quel valore  $\mu_{1/2}$  della variabile casuale tale che

$$P(x \leq \mu_{1/2}) = P(x \geq \mu_{1/2}). \quad (74)$$



Per una variabile casuale continua la mediana è dunque definita dalla condizione

$$\int_{-\infty}^{\mu_{1/2}} p(x) dx = \int_{\mu_{1/2}}^{\infty} p(x) dx = \frac{1}{2}.$$

Per una variabile discreta non è detto che questo valore esista e sia univocamente determinato (motivo per il quale la mediana è rilevante soprattutto per distribuzioni continue). Se una funzione di distribuzione è simmetrica rispetto al valor medio, media e mediana coincidono.

La *moda* (o *valore più probabile*) è semplicemente il valore della variabile casuale (se esiste ed è unico) in corrispondenza del quale la funzione di distribuzione ha un massimo.

### 3.8.2 Varianza e deviazione standard

Caratterizzare la *dispersione* attorno alla media di una variabile casuale significa definire una funzione il cui valore di aspettazione sia una misura di quanto la variabile  $x$  tenda a discostarsi dal proprio valor medio. Potremmo essere tentati di utilizzare la funzione  $f(x) = x - \mu$ , ma è facile verificare che

$$E[x - \mu] = E[x] - \mu = \mu - \mu = 0.$$

Questo valore di aspettazione non fornisce alcuna informazione utile in quanto fluttuazioni per eccesso e fluttuazione per difetto tendono, statisticamente, a compensarsi. Per ovviare a questo inconveniente si usa  $f(x) = (x - \mu)^2$  e si definisce la *varianza* come<sup>2</sup>

$$\sigma^2 = E[(x - \mu)^2] = \begin{cases} \sum (x_k - \mu)^2 P(x_k) & \text{per variabili discrete} \\ \int_{-\infty}^{\infty} (x - \mu)^2 p(x) dx & \text{per variabili continue} \end{cases} \quad (75)$$

(nel seguito indicheremo la varianza di  $x$  con  $\sigma^2$ ,  $\sigma_x^2$  o  $\text{Var}(x)$  a seconda del contesto). Definiamo inoltre la *deviazione standard* come radice quadrata della varianza

$$\sigma = \sqrt{\sigma^2}. \quad (76)$$

È proprio la deviazione standard (che, al contrario della varianza, ha le stesse dimensioni fisiche della variabile casuale di partenza) a rappresentare la misura della dispersione attorno alla media cercata. Esamineremo dettagliatamente in seguito la connessione con l'errore statistico. Notiamo, per inciso, che se  $c$  è una costante

$$\text{Var}(cx) = E[(cx - E[cx])^2] = E[(cx - c\mu)^2] = E[c^2(x - \mu)^2] = c^2 E[(x - \mu)^2] = c^2 \text{Var}(x). \quad (77)$$

Questa è una relazione molto importante: una costante passa fuori dall'operatore varianza *al quadrato*—da cui, se facciamo la radice quadrata di entrambi i membri, si vede che la deviazione standard di  $cx$  è, come uno si aspetterebbe,  $c$  volte la deviazione standard di  $x$ .

Partendo dalla definizione di varianza (75), attraverso una semplice manipolazione algebrica

$$\begin{aligned} \sigma^2 &= E[(x - \mu)^2] = E[x^2 - 2x\mu + \mu^2] = E[x^2] - E[2x\mu] + E[\mu^2] = \\ &= E[x^2] - 2\mu E[x] + \mu^2 = E[x^2] - 2\mu^2 + \mu^2 \end{aligned}$$

si arriva alla relazione utile che utilizzeremo spesso nel seguito

$$\sigma^2 = E[x^2] - \mu^2. \quad (78)$$

<sup>2</sup> La quantità  $f(x) = |x - \mu|$  potrebbe servire altrettanto bene allo scopo, ma come vedremo tra breve,  $(x - \mu)^2$  ha alcune proprietà che la rendono più comoda da utilizzare nella pratica.

► ESEMPIO 3.27. Torniamo al nostro dado a sei facce. Utilizzando il valor medio calcolato nell'esempio 3.24 la varianza si scrive come

$$\sigma^2 = \sum_{k=1}^6 (x_k - \mu)^2 P(x_k) = \frac{1}{6} \sum_{k=1}^6 \left(k - \frac{7}{2}\right)^2 = \frac{1}{6} \left(\frac{25}{4} + \frac{9}{4} + \frac{1}{4} + \frac{1}{4} + \frac{9}{4} + \frac{25}{4}\right) = \frac{35}{12},$$

e la deviazione standard è banalmente  $\sigma = \sqrt{35/12}$ .

► ESEMPIO 3.28. Calcoliamo adesso la varianza per la somma delle uscite di due dadi equi

$$\begin{aligned} \sigma^2 &= \sum_{k=2}^{12} (x_k - \mu)^2 P(x_k) = \sum_{k=2}^{12} (x_k - 7)^2 P(x_k) = 25 \times \frac{1}{36} + 16 \times \frac{2}{36} + 9 \times \frac{3}{36} + 4 \times \frac{4}{36} + 1 \times \frac{5}{36} + \\ &+ 0 \times \frac{6}{36} + 1 \times \frac{5}{36} + 4 \times \frac{4}{36} + 9 \times \frac{3}{36} + 16 \times \frac{2}{36} + 25 \times \frac{1}{36} = \frac{210}{36} = \frac{35}{6}. \end{aligned}$$

(Che è esattamente il doppio della varianza per l'uscita di un dado a sei facce calcolata nell'esercizio precedente.)

► ESEMPIO 3.29. Consideriamo la variabile casuale  $x$  distribuita uniformemente tra 0 ed 1 introdotta nell'esempio 3.26. La varianza si calcola come

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 p(x) dx = \int_0^1 \left(x - \frac{1}{2}\right)^2 dx = \int_0^1 \left(x^2 - x + \frac{1}{4}\right) dx = \left(\frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{4}\right) \Big|_0^1 = \\ &= \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12}, \end{aligned}$$

e la deviazione standard è  $\sigma = \sqrt{1/12}$ . Vedremo nel seguito che questo esempio non è solo di interesse accademico.

### 3.8.3 La larghezza a metà altezza

Un concetto utile che si applica alle funzioni di distribuzione di variabile continua è quello di *semilarghezza a metà altezza*, illustrato in figura 3.8. Se assumiamo che la distribuzione sia unimodale—cioè che abbia un solo massimo—la retta orizzontale che interseca l'asse delle ordinate in corrispondenza della metà del valore del punto di massimo della distribuzione intersecherà la distribuzione stessa esattamente in due punti  $x_a$  ed  $x_b$ . Definiamo allora la larghezza a metà altezza FWHM (*full width at half maximum* in Inglese) come

$$\text{FWHM} = |x_b - x_a|, \quad (79)$$

e la semilarghezza a metà altezza HWHM (*half width at half maximum* in Inglese) come

$$\text{HWHM} = \frac{|x_b - x_a|}{2}. \quad (80)$$

Nella maggior parte delle distribuzioni continue di interesse pratico la semilarghezza a metà altezza è una stima ragionevole delle deviazione standard, nel senso che

$$\text{HWHM} = c\sigma \quad (81)$$

con  $c$  dell'ordine dell'unità—e ne vedremo molti esempi concreti nel seguito. Questo offre un'ulteriore interpretazione geometrica del concetto di deviazione standard come *larghezza* di una distribuzione (vedi figura 3.8).

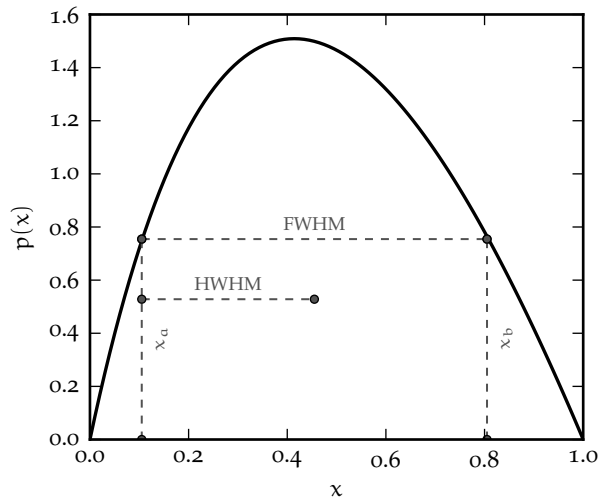


FIGURA 3.8. Significato geometrico della larghezza a metà altezza. Notiamo esplicitamente che, come in questo caso, non è necessario che la distribuzione sia simmetrica rispetto al valor medio.

### 3.8.4 La disuguaglianza di Chebyshev ed il significato della deviazione standard

Il significato profondo della deviazione standard è sostanzialmente che, se misuriamo le deviazioni dal valor medio in unità di  $\sigma$ , è *poco probabile* che esse siano molto grandi. Si tratta di una affermazione piuttosto generica che si può precisare in un certo numero di modi alternativi, uno dei quali è il teorema di Chebyshev.

**TEOREMA 3.3 (DI CHEBYSHEV).** Sia  $x$  una variabile casuale tale che esistano finiti la media  $\mu$  e la varianza  $\sigma^2$ ; preso un numero reale positivo  $c > 0$  si ha:

$$P(|x - \mu| \geq c\sigma) \leq \frac{1}{c^2} \quad (82)$$

*Dimostrazione.* Per comodità illustriamo la dimostrazione nel caso di una variabile continua (il caso discreto è del tutto analogo):

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 p(x) dx \geq \int_{|x - \mu| \geq c\sigma} (x - \mu)^2 p(x) dx \geq \int_{|x - \mu| \geq c\sigma} c^2 \sigma^2 p(x) dx = c^2 \sigma^2 \int_{|x - \mu| \geq c\sigma} p(x) dx = c^2 \sigma^2 P(|x - \mu| \geq c\sigma),$$

da cui segue banalmente la tesi.  $\square$

Il teorema di Chebyshev vale sotto ipotesi molto generali (il che è la ragione principale della sua importanza), per cui il limite superiore fornito è tipicamente piuttosto blando. In tutti i casi di interesse pratico, se conosciamo la forma della distribuzione, o anche solo alcune delle due proprietà, la probabilità in (82) si può calcolare esplicitamente oppure limitare in modo significativamente più stringente.

► **ESEMPIO 3.30.** Nel caso di una variabile casuale  $x$  distribuita uniformemente tra 0 ed 1 i punti a  $\mu - 2\sigma$  e  $\mu + 2\sigma$  cadono al di fuori dell'intervallo di variabilità di  $x$  per cui  $P(|x - \mu| \geq 2\sigma) = 0$ . Il teorema di Chebyshev fornisce il limite  $P(|x - \mu| \geq 2\sigma) \leq 1/4$ .

Se poniamo  $c = 1$  nella (82), il teorema di Chebyshev che abbiamo appena dimostrato si traduce nella condizione banale

$$P(|x - \mu| \geq \sigma) \leq 1 \quad \text{ovvero} \quad P(|x - \mu| \leq \sigma) \geq 0.$$

Cioè la probabilità che la variabile casuale assuma un valore nell'intervallo  $[\mu - \sigma, \mu + \sigma]$  è maggiore di zero. Il che è come non dire niente. Se però ci fermiamo un attimo a guardare i pochi esempi di distribuzione che abbiamo incontrato sino a questo momento ci rendiamo conto che in realtà la probabilità  $P(|x - \mu| \leq \sigma)$  è sostanziale—tipicamente più grande del 60%, come mostrato in tabella 9 e come vedremo più in dettaglio nel seguito.

Distribuzione	Esempi	$P( x - \mu  \leq \sigma)$
Un dado	3.27	$4/6 \approx 67\%$
Somma di due dadi	3.28	$24/36 \approx 67\%$
Distribuzione uniforme	3.29	$2/\sqrt{12} \approx 58\%$

TABELLA 9. Valore della probabilità  $P(|x - \mu| \leq \sigma)$  che una variabile casuale disti di meno di una deviazione standard dalla media per tre distribuzioni particolari.

In un certo senso potremmo dire che, nella maggior parte dei casi di interesse pratico, e a meno di non scegliere distribuzioni multimodali a scopo di contresempio, l'intervallo  $[\mu - \sigma, \mu + \sigma]$  *racchiude* la maggior parte della distribuzione, nel senso che  $P(|x - \mu| \leq \sigma)$  è dell'ordine del 60–70%.

### 3.9 MOMENTI DI UNA DISTRIBUZIONE

Definiamo il *momento di ordine*  $n$  di una variabile casuale  $x$  attorno ad un punto  $x_0$  come:

$$\mathcal{M}_n(x_0) = E[(x - x_0)^n] = \begin{cases} \sum (x_k - x_0)^n P(x_k) & \text{per variabili discrete} \\ \int_{-\infty}^{+\infty} (x - x_0)^n p(x) dx & \text{per variabili continue.} \end{cases} \quad (83)$$

Hanno particolare rilevanza i *momenti algebrici*  $\lambda_n$ , ossia i momenti di ordine generico attorno al punto  $x_0 = 0$

$$\lambda_n = \mathcal{M}_n(0) \quad (84)$$

ed i *momenti centrali*  $\mu_n$ , ossia i momenti di ordine generico attorno al valor medio  $\mu$  di  $x$ :

$$\mu_n = \mathcal{M}_n(\mu) \quad (85)$$

In particolare la media è il momento algebrico di ordine 1 e la varianza è il momento centrale di ordine 2 (e, cosa meno interessante, tutti i momenti di ordine 0 valgono 1 per la condizione di normalizzazione).

Quello di ordine tre è il primo momento centrale di ordine dispari che non si annulla ed è importante perché, pesando con il segno le code a destra e sinistra della media, misura l'eventuale asimmetria della funzione di distribuzione. Sfruttando la linearità del valore di aspettazione, il momento centrale di ordine tre si può scrivere come

$$\begin{aligned} \mu_3 &= E[(x - \mu)^3] = E[x^3 - 3\mu x^2 + 3\mu^2 x - \mu^3] = E[x^3] - 3\mu E[x^2] + 3\mu^2 E[x] - E[\mu^3] = \\ &= E[x^3] - 3\mu(\sigma^2 + \mu^2) + 3\mu^3 - \mu^3. \end{aligned}$$

(Nell'ultimo passaggio abbiamo sfruttato il fatto che  $\sigma^2 = E[x^2] - \mu^2$ .) Si ha dunque la relazione notevole

$$\mu_3 = E[x^3] - 3\mu\sigma^2 - \mu^3, \quad (86)$$

che è in un qualche senso l'equivalente della 78. A partire dal momento centrale di ordine tre si può definire il *coefficiente di asimmetria* o *skewness*  $\gamma_1$

$$\gamma_1 = \frac{\mu_3}{\sigma^3}. \quad (87)$$

La *skewness* è una quantità adimensionale che vale zero per le distribuzioni simmetriche rispetto al valor medio (non è in generale vero il viceversa) e che è diversa da zero se la funzione di distribuzione presenta una coda più *lunga* dell'altra—nel qual caso una *skewness* positiva (negativa) indica che la coda a destra (sinistra) è più lunga.

## 3.10 LA FUNZIONE CUMULATIVA

Data una variabile casuale  $x$  la *funzione cumulativa* o *cumulative density function* (cdf) è definita come

$$F(x') = P(x \leq x'). \quad (88)$$

Dato che la funzione cumulativa ha lo stesso dominio della funzione di distribuzione, il suo argomento si indica generalmente con lo stesso nome della variabile casuale. Operativamente:

$$F(x) = \begin{cases} \sum_{x_k \leq x} P(x_k) & \text{per variabili discrete} \\ \int_{-\infty}^x p(t) dt & \text{per variabili continue} \end{cases} \quad (89)$$

Poiché la probabilità (o la densità di probabilità) è non negativa, la funzione cumulativa è monotona crescente, e, per la condizione di normalizzazione, assume valori nell'intervallo  $[0, 1]$ :

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{e} \quad \lim_{x \rightarrow +\infty} F(x) = 1 \quad (90)$$

(nel caso la densità di probabilità abbia supporto finito, la condizione implica che la funzione cumulativa assume i valori 0 e 1 negli estremi del supporto stesso). Si ha inoltre l'uguaglianza ovvia

$$P(x_1 < x \leq x_2) = F(x_2) - F(x_1). \quad (91)$$

Se la funzione cumulativa è strettamente crescente e continua, allora per ogni numero reale  $0 \leq q \leq 1$  esiste uno ed un solo valore di  $x$  per cui  $F(x) = q$ . L'inverso della funzione cumulativa nel senso della composizione delle funzioni

$$F^{-1}(q) \quad (92)$$

si dice *funzione di distribuzione inversa* o *percent point function*.

► **ESEMPIO 3.31.** Nel caso della variabile casuale  $x$  distribuita uniformemente tra 0 ed 1 introdotta nell'esempio 3.26 la distribuzione cumulativa è semplicemente

$$F(x) = \int_0^x p(t) dt = t \Big|_0^x = (x - 0) = x.$$

Notiamo che  $F(0) = 0$  e  $F(1) = 1$ , come deve essere. La funzione di distribuzione inversa è invece

$$F^{-1}(q) = q.$$

## 3.11 BREVE RIEPILOGO: UN SEMPLICE PROBLEMA DI CALCOLO DELLE PROBABILITÀ

Fermiamoci per un attimo e passiamo in rassegna alcune delle cose che abbiamo visto in questa sezione nel contesto di un (semplice) problema concreto. Supponiamo di avere due variabili casuali continue indipendenti  $x_1$  ed  $x_2$ , entrambe distribuite uniformemente tra 0 ed 1. Vogliamo calcolare la funzione di distribuzione della variabile

$$x = \max(x_1, x_2)$$

e stimare i parametri fondamentali (e.g., media e deviazione standard) caratteristici della distribuzione stessa. (Prima di proseguire intendiamoci sulle regole del gioco: la realizzazione elementare del nostro ipotetico esperimento consiste nel campionare due volte una distribuzione uniforme e prendere il massimo tra i due campionamenti. Un possibile equivalente discreto sarebbe lanciare due dadi e considerare l'uscita più alta come variabile causale.)

3.11.1 *Calcolo della funzione di distribuzione*

Cominciamo dal calcolo della funzione di distribuzione di  $x$ . La prima cosa che possiamo dire senza alcun dubbio è che, se  $x_1$  ed  $x_2$  sono comprese tra 0 e 1, allora anche  $x = \max(x_1, x_2)$  sarà compresa tra 0 ed 1. Intuitivamente ci aspettiamo anche che, se prendiamo il massimo tra due campionamenti di una variabile distribuita uniformemente tra 0 ed 1, valori relativamente più grandi (cioè vicino ad 1) saranno più probabili di valori relativamente più piccoli (cioè vicino a 0).

La figura 3.9 mostra una possibile costruzione geometrica utile per il calcolo della funzione di distribuzione che stiamo cercando. Se  $x_1$  ed  $x_2$  sono indipendenti, allora la realizzazione elementare del nostro esperimento consiste essenzialmente nello scegliere un punto distribuito uniformemente nel piano  $x_1$ - $x_2$ . Prendere il massimo tra  $x_1$  ed  $x_2$  è come dire che scegliamo  $x_1$  se il punto scelto è sotto la diagonale, mentre scegliamo  $x_2$  se il punto è sopra la diagonale. È chiaro allora che il luogo geometrico dei punti, nel piano  $x_1$ - $x_2$ , che corrisponde ad un valore fissato di  $x = \max(x_1, x_2)$  è una linea costituita da due segmenti di uguale lunghezza, paralleli agli assi cartesiani, che si incontrano sulla diagonale, come mostrato in figura 3.9.

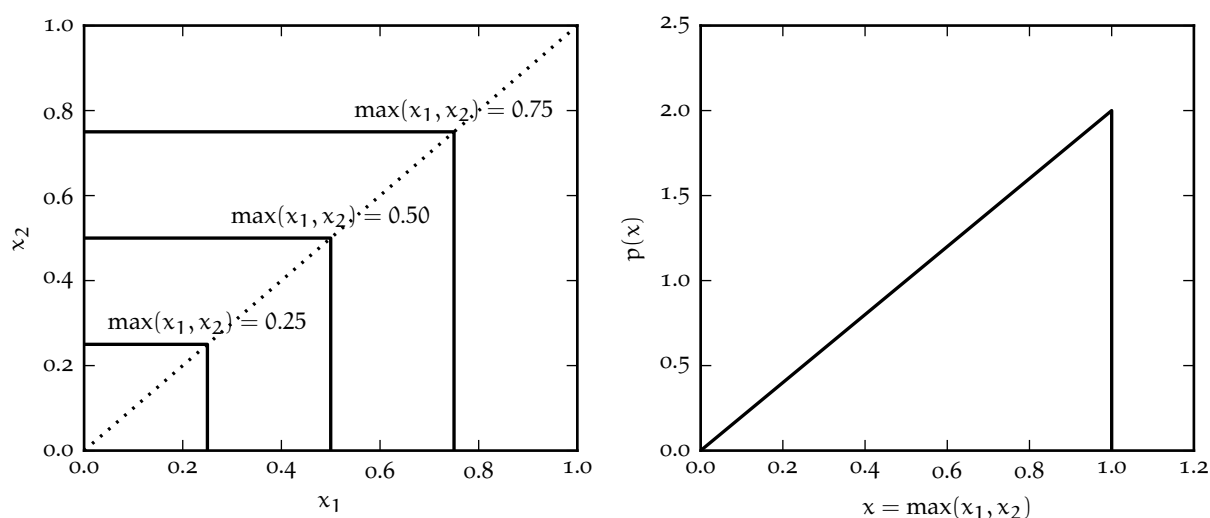


FIGURA 3.9. Costruzione geometrica per il calcolo della funzione di distribuzione del massimo  $x = \max(x_1, x_2)$  di due variabili casuali continue distribuite uniformemente tra 0 e 1 (a sinistra). Il luogo geometrico dei punti che corrisponde ad un valore fissato di  $x = \max(x_1, x_2)$  è dato da due segmenti di uguale lunghezza, paralleli agli assi cartesiani, che si incontrano sulla diagonale. Corrispondentemente, il valore della densità di probabilità  $p(x)$  è proporzionale alla lunghezza di tali segmenti, e cresce linearmente da 0 ad 1, come mostrato a destra.

Tornando alla funzione di distribuzione, possiamo dire che la densità di probabilità  $p(x)$  in un generico punto  $0 \leq x \leq 1$  sarà proporzionale alla lunghezza del segmento corrispondente nella nostra costruzione geometrica. Essa sarà nulla in  $x = 0$  (dove i due segmenti degenerano in un punto) e aumenterà linearmente con  $x$  fino a raggiungere il massimo in  $x = 1$ . La forma analitica sarà cioè  $p(x) = cx$ —con la costante moltiplicativa  $c$  fissata a 2 dalla condizione di normalizzazione

$$p(x) = 2x, \quad (93)$$

come si può verificare facilmente per calcolo diretto. Adesso che abbiamo la forma analitica delle funzione di distribuzione possiamo calcolare facilmente tutte le proprietà derivate.

3.11.2 *Media, mediana e moda*

Cominciamo dunque dalle stime di tendenza centrale. La media si calcola al solito secondo la definizione

$$\mu = E[x] = \int_{-\infty}^{\infty} xp(x)dx = \int_0^1 2x^2 dx = \left. \frac{2x^3}{3} \right|_0^1 = \frac{2}{3}.$$

Non dovrebbe sorprendere che  $\mu > 1/2$ , poiché la densità di probabilità è più alta nella parte destra dell'intervallo di definizione della variabile rispetto alla parte sinistra. La mediana è definita dall'uguaglianza

$$\frac{1}{2} = \int_0^{\mu_{1/2}} p(x) dx = \int_0^{\mu_{1/2}} 2x dx = x^2 \Big|_0^{\mu_{1/2}} = \mu_{1/2}^2,$$

da cui  $\mu_{1/2} = 1/\sqrt{2} \approx 0.707$ . La moda, infine, coincide con il massimo della funzione di distribuzione  $x = 1$ . Il fatto che media, mediana e moda non coincidano, come illustrato in figura 3.10, non dovrebbe sorprendere poiché la funzione di distribuzione non è simmetrica.

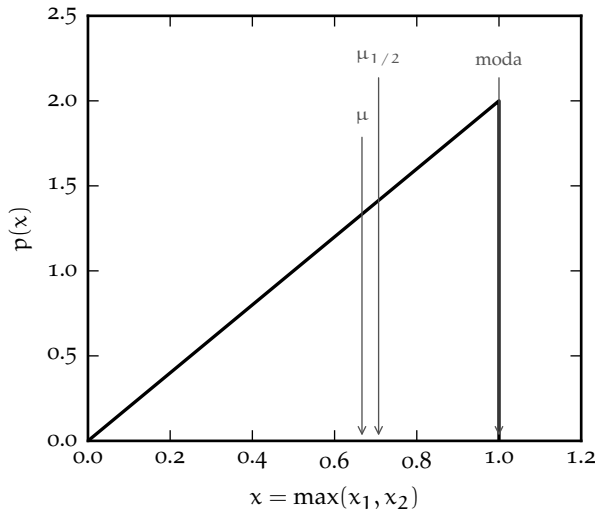


FIGURA 3.10. Illustrazione delle posizioni della media ( $\mu = 2/3$ ), della mediana ( $\mu_{1/2} = 1/\sqrt{2}$ ) e della moda (1) per la funzione di distribuzione triangolare (93). Non essendo la distribuzione simmetrica, non sorprende il fatto che media, mediana e moda non coincidano. (Incidentalmente, questo è anche un esempio di distribuzione per la quale, a causa della peculiarità della forma, il valore più probabile non è molto interessante come stima di tendenza centrale.)

### 3.11.3 Varianza, deviazione standard e larghezza a metà altezza

Procediamo con il calcolo delle metriche di dispersione attorno alla media per la nostra distribuzione. La varianza vale

$$\sigma^2 = E[(x - \mu)^2] = E[x^2] - \mu^2 = \int_{-\infty}^{\infty} x^2 p(x) dx - \mu^2 = \int_0^1 2x^3 dx - \frac{4}{9} = \frac{x^4}{2} \Big|_0^1 - \frac{4}{9} = \frac{1}{2} - \frac{4}{9} = \frac{1}{18},$$

da cui  $\sigma = 1/\sqrt{18} \approx 0.236$ . La retta  $p(x) = 1$  interseca il grafico della funzione di distribuzione nei punti  $x = 1/2$  e  $x = 1$ , come mostrato in figura 3.11. La larghezza intera a metà altezza è dunque  $\text{FWHM} = 1/2$  e la semilarghezza a metà altezza  $\text{HWHM} = 1/4$ . Il rapporto tra semilarghezza a metà altezza e deviazione standard vale dunque

$$\frac{\text{HWHM}}{\sigma} = \frac{\sqrt{18}}{4} \approx 1.061$$

che, come ci aspettiamo, è dell'ordine dell'unità. La frazione della funzione di distribuzione contenuta entro una deviazione standard della media è infine

$$P(|x - \mu| \leq \sigma) = \int_{\mu - \sigma}^{\mu + \sigma} p(x) dx = \int_{\mu - \sigma}^{\mu + \sigma} 2x dx = x^2 \Big|_{\mu - \sigma}^{\mu + \sigma} = (\mu + \sigma)^2 - (\mu - \sigma)^2 = 4\mu\sigma = \frac{8}{9\sqrt{2}} \approx 0.629,$$

in linea con i valori riportati nella tabella 9.

### 3.11.4 Coefficiente di asimmetria

La funzione di distribuzione (93) non è simmetrica rispetto al valor medio, per cui ci aspettiamo che il momento centrale di ordine tre  $\mu_3$  e, di conseguenza, il coefficiente di asimmetria  $\gamma_1$  siano non nulli.

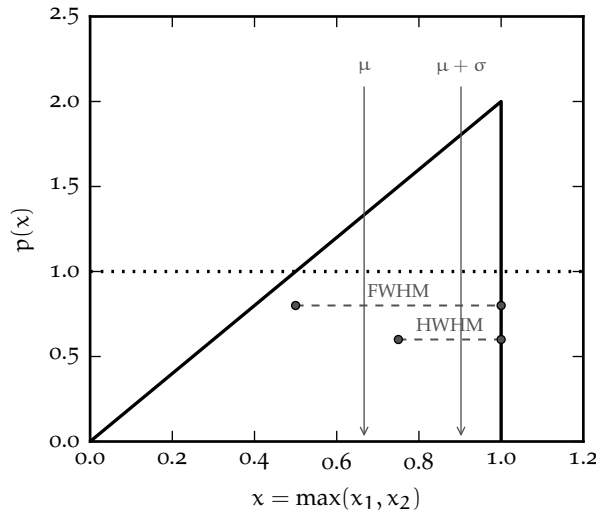


FIGURA 3.11. Illustrazione della deviazione standard  $\sigma = 1/\sqrt{18}$  e della larghezza a metà altezza  $\text{FWHM} = 1/2$  per la funzione di distribuzione triangolare (93). Come abbiamo avuto occasione di dire la deviazione standard e la semilarghezza a metà altezza sono vicine tra loro (nel senso che il loro rapporto è dell'ordine di 1. Inoltre l'intervallo  $[\mu - \sigma, \mu + \sigma]$  racchiude una frazione sostanziale (circa il 63%) della funzione di distribuzione.)

Per la precisione, dato che la coda più pronunciata è quella a sinistra delle media, ci aspettiamo che la *skewness* sia negativa. Procediamo allora per gradi e calcoliamo il momento terzo utilizzando la (86)

$$\mu_3 = E[x^3] - 3\mu\sigma^2 - \mu^3 = \int_{-\infty}^{\infty} x^3 p(x) dx - \frac{1}{9} - \frac{8}{27} = \int_0^1 2x^4 dx - \frac{11}{27} = \frac{2x^5}{5} \Big|_0^1 - \frac{11}{27} = \frac{2}{5} - \frac{11}{27} = -\frac{1}{135},$$

da cui, come ci aspettavamo qualitativamente

$$\gamma_1 = -\frac{\sqrt{18^3}}{135} = -\frac{2\sqrt{2}}{5} \approx -0.566.$$

### 3.11.5 Distribuzione cumulativa e distribuzione inversa

Completiamo la nostra panoramica con il calcolo esplicito della distribuzione cumulativa

$$F(x) = \int_{-\infty}^x p(t) dt = \int_0^x 2t dt = t^2 \Big|_0^x = x^2,$$

da cui, tra le altre cose, possiamo derivare in modo leggermente più semplice il risultato trovato prima

$$P(|x - \mu| \leq \sigma) = F(\mu + \sigma) - F(\mu - \sigma) = (\mu + \sigma)^2 - (\mu - \sigma)^2.$$

La funzione di distribuzione inversa si scrive infine come

$$F^{-1}(q) = \sqrt{q}.$$

## 3.12 DIGRESSIONE: ALCUNI PARADOSSI PROBABILISTICI

Per *paradosso* si intende letteralmente una proposizione logicamente corretta ma in apparente contraddizione con l'esperienza comune. Se inserite in un motore di ricerca le due parole chiave "paradosso" e "probabilità" (in qualsiasi lingua) non rimarrete delusi dalla quantità di informazione sull'argomento disponibile su web. Non c'è che l'imbarazzo della scelta.

I paradossi probabilistici si possono rozzamente suddividere in due macro-categorie: (i) conseguenze dirette (e logicamente corrette) delle leggi della probabilità che, applicate ad una specifica situazione, sono sorprendenti perché anti-intuitive e (ii) problemi in cui il paradosso in effetti ha origine da una formulazione incompleta o volutamente ambigua del problema. Questi ultimi hanno spesso più a che vedere con la semantica che con la probabilità, per cui non sono molto interessanti in questo contesto (ma se siete interessati potete dare un'occhiata, tra le altre cose, a [6] e [4]). I veri paradossi, invece, possono essere utili ad illustrare i fondamenti del ragionamento probabilistico, per cui ne esamineremo alcuni tra i più famosi.



## 3.12.1 Quanti compleanni lo stesso giorno?

Dato un insieme di  $n$  persone scelte casualmente, qual è la probabilità che almeno due di esse abbiano lo stesso compleanno? (Per semplicità trascureremo gli anni bisestili ed assumeremo che le nascite siano uniformemente distribuite sui 365 giorni dell'anno—escluso il 29 febbraio. Assumeremo anche che nel gruppo di persone non vi siano gemelli.)

Una breve nota a margine: ogni volta che la parola *almeno* compare in un problema di probabilità dovete riflettere attentamente su come essa si traduca effettivamente in termini quantitativi. In questo caso il nostro evento si può realizzare con una singola coppia di persone con lo stesso compleanno, oppure due coppie, oppure con 3 o 4 persone nate lo stesso giorno e così via—addirittura con la (improbabile) situazione in cui tutte le  $n$  persone sono nate lo stesso giorno. Al crescere di  $n$  il numero di possibilità (che si dice anche genericamente il *combinatoriale* del problema) cresce rapidamente e spesso è più semplice calcolare la probabilità dell'evento complementare all'evento di partenza ed usare il teorema della probabilità complementare (40).

Attacchiamo allora il nostro paradosso chiedendoci quale sia la probabilità che i compleanni delle  $n$  persone che abbiamo scelto siano tutti diversi. La prima persona avrà il compleanno in un determinato giorno dell'anno (tra i 365). Per la seconda persona rimangono 364/365 disponibili diversi dal compleanno della prima, per cui la probabilità che il compleanno non cada lo stesso giorno è proprio 364/365. Per la terza persona la probabilità che il compleanno cada in un giorno diverso dalle prime due è 363/365 e così via—fino all'ultima, per cui questa probabilità è  $(365 - n + 1)/365$ . Visto che le date di nascita sono tutte indipendenti tra loro, la probabilità che esse siano *tutte* diverse sarà

$$P(\text{compleanni tutti diversi}) = 1 \times \frac{364}{365} \times \frac{363}{365} \times \cdots \times \frac{365 - (n - 1)}{365} = \frac{365!}{365^n (365 - n)!} \quad (94)$$

e la probabilità che almeno due persone abbiano il compleanno lo stesso giorno è, per il teorema (40)

$$P(\text{almeno due compleanni nello stesso giorno}) = 1 - \frac{365!}{365^n (365 - n)!}. \quad (95)$$

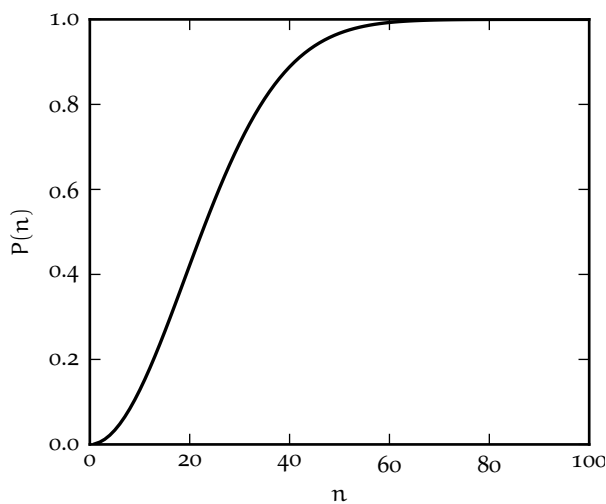


FIGURA 3.12. Andamento della (95) per  $n \leq 100$ . La probabilità che in un gruppo di  $n$  persone almeno due festeggino il compleanno lo stesso giorno supera il 50% per  $n = 23$ .

La (95) non è molto espressiva, ma la figura 3.12 illustra l'andamento della (95) per  $n \leq 100$ . È interessante notare come la funzione superi il 50% per  $n = 23$  ed il 99% per  $n = 60$ . In un gruppo di 23 persone, dunque, la probabilità che ve ne siano almeno due che festeggino il compleanno lo stesso giorno è maggiore del 50%—che è proprio il contenuto del nostro paradosso. (Cosa suggeriva il vostro senso comune prima di fare il conto?)

## 3.12.2 Breve digressione: somma della serie geometrica

Dato un generico numero reale  $r$ , con  $|r| < 1$  si dice serie geometrica la somma (infinita)

$$\sum_{k=0}^{\infty} r^k \quad (96)$$

(è ovvio che per  $|r| \geq 1$  i termini della serie crescono o al più rimangono costanti al crescere di  $k$  per cui la serie diverge—il caso  $r = -1$  è interessante, e di nuovo la (96) non ha somma definita).

Si tratta di uno dei casi in cui la somma parziale dei primi  $n$  termini si può scrivere esplicitamente in forma chiusa

$$\begin{aligned} \sum_{k=0}^n r^k &= 1 + r + r^2 + \dots + r^n = (1 + r + r^2 + \dots + r^n) \times \frac{1-r}{1-r} = \\ &= \frac{1 + r + r^2 + \dots + r^n - r - r^2 - r^3 - \dots - r^{n+1}}{1-r} = \frac{1 - r^{n+1}}{1-r} \end{aligned}$$

a partire dalla quale la somma della serie si calcola banalmente come limite per  $n \rightarrow \infty$

$$\sum_{k=0}^{\infty} r^k = \lim_{n \rightarrow \infty} \frac{1 - r^{n+1}}{1-r} = \frac{1}{1-r}. \quad (97)$$

Con lo stesso metodo si può calcolare anche la somma parziale (che è diversa, ma correlata alla precedente)

$$\begin{aligned} \sum_{k=0}^n kr^k &= 0 + r + 2r^2 + \dots + nr^n = (r + 2r^2 + \dots + nr^n) \times \frac{1-r}{1-r} = \\ &= \frac{r + 2r^2 + \dots + nr^n - r^2 - 2r^3 - (n-1)r^n - nr^{n+1}}{1-r} = \frac{r + r^2 + \dots + r^n - nr^{n+1}}{1-r} = \\ &= \frac{\sum_{k=1}^n r^k - nr^{n+1}}{1-r} = \frac{\sum_{k=0}^n r^k - 1 - nr^{n+1}}{1-r} = \frac{(1 - r^{n+1})/(1-r) - 1 - nr^{n+1}}{1-r} = \\ &= \frac{1 - r^{n+1} - 1 - nr^{n+1} + r + nr^{n+2}}{(1-r)^2} = \frac{r - (n+1)r^{n+1} + nr^{n+2}}{(1-r)^2}, \end{aligned}$$

da cui, prendendo come prima il limite per  $n \rightarrow \infty$  si ottiene

$$\sum_{k=0}^{\infty} kr^k = \lim_{n \rightarrow \infty} \frac{r - (n+1)r^{n+1} + nr^{n+2}}{(1-r)^2} = \frac{r}{(1-r)^2}. \quad (98)$$

Cogliamo l'occasione per notare che lo stesso risultato si poteva ottenere, più semplicemente, derivando la (97) rispetto ad  $r$

$$\sum_{k=0}^{\infty} kr^k = r \frac{d}{dr} \sum_{k=0}^{\infty} r^k = \frac{r}{(1-r)^2},$$

che illustra come spesso, con un minimo di creatività, lo stesso calcolo può diventare da laborioso a banale.

## 3.12.3 Maschio o femmina?

Riprendiamo il filo della nostra discussione sui paradossi della probabilità. Una coppia di genitori decide di avere figli fino alla nascita della prima femmina—avvenuta la quale semplicemente smetterà di procreare. Quanti maschi (e quanti figli in totale) avrà *in media* la coppia?

Assumeremo per semplicità che la probabilità che un figlio sia maschio (o femmina) sia esattamente il 50%—il che è notoriamente non vero, ma è di fatto irrilevante per la nostra discussione. Si tratta

banalmente di un processo binomiale in cui, detti F ed M la nascita di un maschio e di una femmina, rispettivamente, siamo interessati alle sequenze F, MF, MMF, MMMF e così via. La probabilità di avere  $k$  figli, cioè  $n_M = (k-1)$  maschi seguiti da  $n_F = 1$  femmina, è ovviamente  $1/2^k$ , come illustrato nella tabella 10.

Sequenza	$n$	$n_F$	$n_M$	Probabilità
F	1	1	0	$1/2$
MF	2	1	1	$1/4$
MMF	3	1	2	$1/8$
MMMF	4	1	3	$1/16$
...				
$(k-1)MF$	$k$	1	$k-1$	$1/2^k$
$kMF$	$k+1$	1	$k$	$1/2^{k+1}$
...				

TABELLA 10. Illustrazioni delle sequenze possibili e delle probabilità corrispondenti per il paradosso delle nascite (maschio o femmina?).

Nelle nostre ipotesi il numero di femmine è fisso a 1—anche se, in effetti, può essere considerato come una variabile casuale che può assumere un solo valore (1, appunto) la cui probabilità di uscita è correttamente normalizzata

$$P(n_F = 1) = \sum_{k=0}^{\infty} \frac{1}{2^{k+1}} = \frac{1}{2} \sum_{k=0}^{\infty} \frac{1}{2^k} = \frac{1}{2} \times 2 = 1,$$

ed il cui valore di aspettazione è, ovviamente, 1. Il valore di aspettazione per il numero di maschi  $n_M$  si scrive invece come

$$E[n_M] = \sum_{k=0}^{\infty} \frac{k}{2^{k+1}} = \frac{1}{2} \sum_{k=0}^{\infty} \frac{k}{2^k} = \frac{1}{2} \times 2 = 1 \quad (99)$$

ed è anch'esso pari ad 1, nonostante  $n_M$  possa essere arbitrariamente grande. In altre parole il numero medio (totale) atteso di figli è esattamente due—un maschio ed una femmina.

È interessante chiedersi come cambi il tutto se introduciamo un limite  $n_{\max}$  al numero totale di figli—cioè se diciamo che, nell'ipotesi che una femmina non sia nata nei primi  $n_{\max}$  tentativi, la coppia decida di darsi pace e non riprovare. In questo nuovo schema la tabella 10 cessa di essere di lunghezza infinita e termina con la sequenza  $n_{\max}M$ , per la quale  $n = n_{\max}$ ,  $n_F = 0$  e  $n_M = n_{\max}$ , con probabilità  $1/2^{n_{\max}}$ . (Notate che le ultime due righe di questa tabella *finita* sono equiprobabili, perché corrispondono a due sequenze di lunghezza identica in cui l'unica differenza è l'ultimo elemento—M o F.)

Sequenza	$n$	$n_F$	$n_M$	Probabilità
F	1	1	0	$1/2$
MF	2	1	1	$1/4$
MMF	3	1	2	$1/8$
MMMF	4	1	3	$1/16$
MMMM	4	0	4	$1/16$

TABELLA 11. Illustrazioni delle sequenze possibili e delle probabilità corrispondenti per il paradosso delle nascite modificato, nel caso particolare  $n_{\max} = 4$ .

La prima conseguenza è che il numero di femmine  $n_F$  è adesso a tutti gli effetti una variabile casuale (che può assumere più di un valore), il cui valore di aspettazione è

$$E[n_F] = 1 \times \left(1 - \frac{1}{2^{n_{\max}}}\right) + 0 \times \frac{1}{2^{n_{\max}}} = 1 - \frac{1}{2^{n_{\max}}},$$

mentre il valore di aspettazione del numero di maschi è questa volta

$$E[n_M] = \sum_{k=0}^{n_{\max}-1} \frac{k}{2^{k+1}} + \frac{n_{\max}}{2^{n_{\max}}} = \frac{1}{2} \sum_{k=0}^{n_{\max}-1} \frac{k}{2^k} + \frac{n_{\max}}{2^{n_{\max}}} = \frac{1}{2} \times \frac{1 - \frac{n_{\max}}{2^{n_{\max}}} + \frac{(n_{\max}-1)}{2^{n_{\max}+1}}}{\frac{1}{4}} + \frac{n_{\max}}{2^{n_{\max}}} = 1 - \frac{1}{2^{n_{\max}}}. \quad (100)$$

Quindi, indipendentemente dal numero massimo di figli  $n_{\max}$ , i valori di aspettazione per il numero di femmine e di maschi sono esattamente uguali ( $e \leq 1$ ). Per completezza, nel caso  $n_{\max} = 4$  illustrato in tabella 11 questi valori di aspettazione sono  $15/16$ .

### 3.12.4 Una strategia di gioco interessante

Un casinò offre un tavolo di gioco di testa (T) o croce (C) con una moneta equa, in cui banalmente si riceve il doppio della puntata in caso di vittoria (e, ovviamente, si perde la posta in caso contrario). Un giocatore decide di adottare la seguente strategia di gioco. Inizialmente scommette una certa somma  $s$  su T; se vince si ritira con il corrispettivo  $2s$  (ed una vincita netta pari ad  $s$ ), mentre se perde continua raddoppiando la puntata ogni volta fino alla prima uscita di T—momento nel quale interrompe il gioco.

La prima cosa che ci chiediamo è la somma netta vinta (o perduta) al termine del gioco. Se T esce per la prima volta al secondo lancio la somma totale giocata è  $S_2 = s + 2s = 3s$  ed il corrispettivo pagato dal casinò è  $4s$ , per una vincita netta di  $s$ . Se T esce per la prima volta al terzo lancio la somma giocata è  $S_3 = s + 2s + 4s = 7s$  ed il corrispettivo è  $8s$ , per una vincita netta che è di nuovo  $s$ . È allora facile convincersi che se T esce per la prima volta dopo  $n$  lanci la somma totale giocata è

$$S_n = \sum_{k=0}^{n-1} s2^k = s(2^n - 1).$$

e, come detto, la vincita netta è sempre la puntata iniziale  $s$ —sembrerebbe una strategia perfetta, in cui si vince sempre.

La seconda domanda interessante a cui rispondere è: quale somma di denaro ci aspettiamo di dover giocare, in media, prima di poter abbandonare il tavolo con la nostra vincita netta  $s$ ? O, in altre parole, quanto è il valore di aspettazione  $E[S]$ ? Per rispondere dobbiamo calcolare la probabilità che T esca per la prima volta dopo esattamente  $n$  lanci, il che è banale, perché si tratta di  $n$  eventi indipendenti con probabilità di successo (e fallimento) pari ad  $1/2$

$$P\left(\overbrace{CC \dots CC}^{n-1 \text{ volte}} T\right) = \frac{1}{2^n}.$$

A questo punto abbiamo tutti gli ingredienti per rispondere alla nostra domanda, e banalmente

$$E[S] = \sum_{n=1}^{\infty} \frac{s(2^n - 1)}{2^n} = \sum_{n=1}^{\infty} \left(s - \frac{s}{2^n}\right), \quad (101)$$

che chiaramente diverge perché il primo termine della serie è costante. Fisicamente questo significa che, benché la strategia, in astratto, garantisca una vincita netta, la somma media necessaria per lasciare il tavolo ed incassare la vincita diverge—la strategia funziona solo se si ha a disposizione una quantità infinita di denaro.

Il punto della questione sta ovviamente nel fatto che la nostra strategia di incremento della puntata è esponenziale—e se non siete convinti ripensate a cosa abbiamo detto a proposito nella sezione 2.4. Per fissare le idee, la probabilità che occorran 30 lanci per vincere è estremamente piccola:  $1/2^{30} \approx 9.3 \times 10^{-10}$ —se giocassimo senza interruzione al ritmo di un lancio al secondo vedremmo in media una occorrenza di questo evento ogni 34 anni circa. Ma nel momento in cui ciò accadesse, assumendo una puntata iniziale di  $s = 1$  euro, dovremmo avere  $2^{31} - 1 \approx 2.15$  miliardi (!) di euro per poter proseguire al lancio successivo.

## 3.13 CENNI ALLE VARIABILI MULTIVARIATE: INDIPENDENZA, COVARIANZA E CORRELAZIONE

La nostra discussione, fino a questo momento, è stata largamente incentrata su variabili casuali *single*. Abbiamo definito l'indipendenza tra eventi, ma non abbiamo dedicato molta attenzione alle relazioni tra variabili casuali diverse, ed in particolare al concetto di indipendenza tra variabili casuali, per cui completiamo il capitolo colmando questa lacuna.

## 3.13.1 L'indipendenza tra variabili casuali

Quando si ha a che vedere con un insieme di variabili casuali  $x_1 \dots x_n$ , il problema è completamente specificato dalla loro funzione di distribuzione congiunta, vale a dire dalla probabilità (o densità di probabilità) che ciascuna delle variabili assuma un valore specifico

$$\begin{cases} P(x_1, \dots, x_n) & \text{per variabili discrete} \\ p(x_1, \dots, x_n) & \text{per variabili continue.} \end{cases}$$

Per semplicità, ma senza perdere di generalità, in questa sezione ci concentreremo sul caso di due variabili casuali continue  $x_1$  ed  $x_2$ , descritte dalla densità di probabilità congiunta  $p(x_1, x_2)$ . Analogamente a quanto abbiamo fatto nel caso di una variabile, richiederemo che la densità di probabilità sia positiva e propriamente normalizzata sul suo supporto di definizione

$$p(x_1, x_2) \geq 0 \quad \text{e} \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) dx_1 dx_2 = 1,$$

col che possiamo calcolare la probabilità che la coppia ordinata  $(x_1, x_2)$  sia contenuta in una regione specifica  $A$  del piano come

$$P((x_1, x_2) \in A) = \iint_A p(x_1, x_2) dx_1 dx_2,$$

e definire il valore di aspettazione di una generica funzione  $f(x_1, x_2)$  come

$$E[f(x_1, x_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) p(x_1, x_2) dx_1 dx_2.$$

► **ESEMPIO 3.32.** La funzione  $p(x_1, x_2) = 4x_1x_2$  (con  $0 \leq x_1 \leq 1$  e  $0 \leq x_2 \leq 1$ ) è una funzione di distribuzione, in quanto è semidefinita positiva ed è correttamente normalizzata

$$\int_0^1 \int_0^1 4x_1x_2 dx_1 dx_2 = 4 \int_0^1 x_1 dx_1 \int_0^1 x_2 dx_2 = 4 \left. \frac{x_1^2}{2} \right|_0^1 \times \left. \frac{x_2^2}{2} \right|_0^1 = 1.$$

Ora, con più di una variabile in gioco, la situazione diventa immediatamente interessante—e questo ci permette di rivisitare in una chiave più concreta alcune delle cose che abbiamo visto nella parte iniziale del capitolo. Possiamo chiederci, ad esempio, quale sia la densità di probabilità di  $p(x_1 | x_2)$  di  $x_1$  condizionata al fatto che  $x_2$  assuma un valore specifico. Intuitivamente possiamo aspettarci che essa sia proporzionale a  $p(x_1, x_2)$ , ma perché essa sia una densità di probabilità unidimensionale dobbiamo imporre la condizione di normalizzazione, cioè:

$$p(x_1 | x_2) = \frac{p(x_1, x_2)}{\int_{-\infty}^{\infty} p(x_1, x_2) dx_1} = \frac{p(x_1, x_2)}{p_2(x_2)} \quad (102)$$

in cui abbiamo implicitamente definito la densità di probabilità marginale, cioè la densità di probabilità unidimensionale di una delle due variabili che si ottiene integrando sull'altra variabile la densità di probabilità congiunta

$$p_1(x_1) = \int_{-\infty}^{\infty} p(x_1, x_2) dx_2 \quad \text{e} \quad p_2(x_2) = \int_{-\infty}^{\infty} p(x_1, x_2) dx_1. \quad (103)$$

La (102) non è altro che l'equivalente della (56) nel caso di variabili casuali. Nel nostro linguaggio, due variabili casuali si dicono indipendenti se la densità di probabilità congiunta può essere fattorizzata nel prodotto delle due densità di probabilità marginali, ovvero sia

$$p(x_1, x_2) = p_1(x_1)p_2(x_2). \quad (104)$$

La definizione non è sorprendente, poiché in questo caso la probabilità condizionata da cui siamo partiti diviene

$$p(x_1 | x_2) = \frac{p_1(x_1)p_2(x_2)}{p_2(x_2)} = p_1(x_1) \quad \text{e} \quad p(x_2 | x_1) = \frac{p_1(x_1)p_2(x_2)}{p_1(x_1)} = p_2(x_2),$$

cioè la funzione di distribuzione unidimensionale di ciascuna delle due variabili non dipende dal particolare valore assunto dall'altra—esattamente ciò che intendiamo per indipendenza. Le due relazioni appena viste sono la riscrittura in questo contesto delle (58).

► **ESEMPIO 3.33.** Consideriamo di nuovo la funzione di distribuzione congiunta dell'esempio 3.32. Le due densità di probabilità marginali sono

$$p_1(x_1) = \int_0^1 p(x_1, x_2) dx_2 = \int_0^1 4x_1 x_2 dx_2 = 2x_1 \quad \text{e analogamente} \quad p_2(x_2) = 2x_2.$$

Le due variabili sono dunque indipendenti in quanto vale la (104).

Se  $x_1$  ed  $x_2$  sono indipendenti, allora si dimostra banalmente che il valore di aspettazione del loro prodotto è uguale al prodotto dei valori di aspettazione

$$E[x_1 x_2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^{\infty} x_1 p_1(x_1) dx_1 \int_{-\infty}^{\infty} x_2 p_2(x_2) dx_2 = E[x_1] E[x_2]. \quad (105)$$

Notiamo, per completezza, che tutto ciò che abbiamo detto in questa sezione vale anche, modulo la sostituzione degli integrali con sommatorie, per variabili casuali discrete, e si estende banalmente ad un numero arbitrario di variabili.

### 3.13.2 Covarianza e correlazione

Date due variabili casuali  $x_1$  ed  $x_2$ , e dette  $\mu_1 = E[x_1]$  e  $\mu_2 = E[x_2]$  le rispettive medie, definiamo la loro covarianza, che chiameremo  $\text{Cov}(x_1, x_2)$  o  $\sigma_{x_1 x_2}$ , come il valore di aspettazione delle rispettive fluttuazioni attorno al valor medio

$$\text{Cov}(x_1, x_2) = \sigma_{x_1 x_2} = E[(x_1 - \mu_1)(x_2 - \mu_2)]. \quad (106)$$

Si dimostra banalmente che

$$\text{Cov}(x_1, x_2) = E[x_1 x_2] - \mu_2 E[x_1] - \mu_1 E[x_2] + \mu_1 \mu_2 = E[x_1 x_2] - \mu_1 \mu_2 = E[x_1 x_2] - E[x_1] E[x_2],$$

che ci permette di riscrivere la (105) nella forma equivalente

$$E[x_1 x_2] = E[x_1] E[x_2] \iff \text{Cov}(x_1, x_2) = 0,$$

Il concetto di covarianza è dunque legato a quello di indipendenza, nel senso che *se due variabili sono indipendenti, la loro covarianza è nulla*. Questo non dovrebbe stupire, poiché se  $x_1$  ed  $x_2$  sono indipendenti, allora non c'è nessun motivo per cui valori di  $x_1$  al di sotto (o al di sopra) della media debbano essere associati preferenzialmente a valori di  $x_2$  al di sotto (o al di sopra della media), per cui valori positivi e negativi dei due fattori della (106) tendono cancellarsi tra di loro.

È interessante notare come l'implicazione inversa non valga in generale, cioè *il fatto che la covarianza tra due variabili sia nulla non implica necessariamente che le due variabili siano indipendenti*, come illustrato nell'esempio 3.34. Il motivo è che la covarianza è sensibile solo ad un tipo specifico di mutua dipendenza, ovvero quello lineare.

Da un punto di vista matematico, la covarianza è una forma bilineare simmetrica, nel senso che gode delle seguenti proprietà elementari

$$\text{Cov}(x_1, x_2) = \text{Cov}(x_2, x_1) \quad (107)$$

$$\text{Cov}(c_1 x_1 + c_2 x_2, x_3) = c_1 \text{Cov}(x_1, x_3) + c_2 \text{Cov}(x_2, x_3) \quad (108)$$

$$\text{Cov}(x_1, c_2 x_2 + c_3 x_3) = c_2 \text{Cov}(x_1, x_2) + c_3 \text{Cov}(x_1, x_3) \quad (109)$$

► **ESEMPIO 3.34.** Consideriamo una variabile casuale continua  $x$  con funzione di distribuzione  $p(x)$  simmetrica rispetto a 0—il che implica che tutti i momenti algebrici di ordine dispari sono nulli. Si ha banalmente

$$\text{Cov}(x, x^2) = E[x^3] - E[x] E[x^2] = 0$$

poiché sia  $E[x^3]$  che  $E[x]$  sono nulli. Dunque la covarianza tra  $x$  ed  $x^2$  è nulla, eppure è ovvio che le due variabili non sono indipendenti, in quanto una è univocamente determinata dall'altra.

Si dimostra anche banalmente che la covarianza di una variabile casuale con una costante è identicamente nulla e che la covarianza di una variabile casuale con se stessa non è altro che la varianza della variabile stessa

$$\text{Cov}(x, c) = 0 \quad (110)$$

$$\text{Cov}(x, x) = \text{Var}(x). \quad (111)$$

Le covarianze  $\text{Cov}(x_i, x_j)$  di  $n$  variabili casuali si possono organizzare in modo naturale in una matrice simmetrica  $n \times n$  che prende il nome di *matrice di covarianza* e che, nel caso  $n = 2$ , si scrive esplicitamente come

$$\Sigma = \begin{bmatrix} \text{Cov}(x_1, x_1) & \text{Cov}(x_1, x_2) \\ \text{Cov}(x_2, x_1) & \text{Cov}(x_2, x_2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \text{Cov}(x_1, x_2) \\ \text{Cov}(x_2, x_1) & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \text{Cov}(x_1, x_2) \\ \text{Cov}(x_1, x_2) & \sigma_2^2 \end{bmatrix}. \quad (112)$$

(La generalizzazione al caso generale dovrebbe essere ovvia.)

A partire dalla covarianza si definisce una sua versione *riscalata* che prende il nome di *correlazione*  $\text{Corr}(x_1, x_2)$  o  $\rho_{x_1 x_2}$

$$\text{Corr}(x_1, x_2) = \rho_{x_1 x_2} = \frac{\text{Cov}(x_1, x_2)}{\sigma_1 \sigma_2}, \quad (113)$$

in cui  $\sigma_1$  e  $\sigma_2$  sono le deviazioni standard di  $x_1$  ed  $x_2$ , rispettivamente. La correlazione possiede banalmente tutte le proprietà della covarianza—in particolare è nulla se le variabili  $x_1$  ed  $x_2$  sono indipendenti (ma, di nuovo, l'implicazione inversa non è vera in generale).

La correlazione è una quantità adimensionale che assume valori compresi tra  $-1$  ed  $1$  (non lo dimostreremo formalmente, ma la covarianza è una sorta di prodotto scalare, mentre la deviazione standard ha il significato di una norma, per cui questa proprietà segue essenzialmente dalla disuguaglianza di Cauchy-Schwarz). Se la correlazione tra due variabili è nulla, esse si dicono *scorrelate*. Se la correlazione è maggiore (minore) di  $0$ , esse si dicono *positivamente* (*negativamente*) *correlate*. Nel caso particolare in cui una variabile casuale dipenda linearmente da un'altra

$$x_2 = mx_1 + q \quad (\text{per cui } \sigma_2 = |m| \sigma_1) \quad (114)$$

allora si ha

$$\text{Corr}(x_1, x_2) = \frac{\text{Cov}(x_1, x_2)}{\sigma_1 \sigma_2} = \frac{\text{Cov}(x_1, mx_1 + q)}{\sigma_1 \sigma_2} = \frac{m \text{Cov}(x_1, x_1)}{|m| \sigma_1^2} = \frac{m}{|m|} = \pm 1 \quad (115)$$

ovverosia la correlazione è  $\pm 1$  a seconda che il segno di  $m$  sia positivo o negativo. Si può anche dimostrare il contrario, ovverosia la correlazione tra due variabili è (in modulo) unitaria se e solo se esse sono legate tra loro da una relazione lineare. Ne segue banalmente che la correlazione di una variabile con se stessa è pari all'unità

$$\text{Corr}(x, x) = 1. \quad (116)$$

In analogia alla matrice di covarianza si definisce la *matrice di correlazione* come

$$R = \begin{bmatrix} \text{Corr}(x_1, x_1) & \text{Corr}(x_1, x_2) \\ \text{Corr}(x_2, x_1) & \text{Corr}(x_2, x_2) \end{bmatrix} = \begin{bmatrix} 1 & \text{Corr}(x_1, x_2) \\ \text{Corr}(x_2, x_1) & 1 \end{bmatrix} = \begin{bmatrix} 1 & \text{Corr}(x_1, x_2) \\ \text{Corr}(x_1, x_2) & 1 \end{bmatrix}. \quad (117)$$

## 3.14 IN BREVE...

- La probabilità è una misura che associa ad ogni elemento dello spazio degli eventi un numero reale che soddisfa gli assiomi di Kolmogorov.
- La probabilità condizionata  $P(E_1 | E_2)$  è la probabilità che si verifichi  $E_1$ , subordinata al fatto che si sia verificato  $E_2$ . I due eventi  $E_1$  ed  $E_2$  si dicono indipendenti se  $P(E_1 | E_2) = P(E_1)$  e  $P(E_2 | E_1) = P(E_2)$ . La probabilità che due eventi indipendenti si verifichino insieme è uguale al prodotto delle probabilità.
- Il teorema di Bayes (60) lega tra di loro le probabilità condizionate  $P(E_1 | E_2)$  e  $P(E_2 | E_1)$  ed è di fondamentale importanza nei problemi di probabilità inversa.
- Data una variabile casuale  $x$ , la funzione di distribuzione è una funzione che associa ad ogni valore di  $x$  la sua probabilità (nel caso discreto) o la sua densità di probabilità (nel caso continuo).
- A partire dal concetto di valore di aspettazione di una funzione di variabile casuale si definiscono tutte le proprietà rilevanti delle distribuzioni, tra cui

$$\mu = E[x] \quad \text{(la media)}$$

$$\sigma^2 = E[(x - \mu)^2] = E[x^2] - \mu^2 \quad \text{(la varianza)}$$

$$\sigma = \sqrt{\sigma^2} \quad \text{(la deviazione standard)}$$

$$\mathcal{M}_n(x_0) = E[(x - x_0)^n] \quad \text{(i momenti di ordine generico).}$$



---

VARIABILI CAMPIONE E PROPAGAZIONE DELL'ERRORE STATISTICO

---

È giunto adesso il momento di fare tesoro dei concetti che abbiamo introdotto nel capitolo precedente per rivedere criticamente la nozione di errore di misura. Quando misuriamo una grandezza fisica—e specialmente quando il valore della misura fluttua—possiamo pensare che il valore stesso sia una variabile casuale con una certa distribuzione (tipicamente incognita a priori) che chiamiamo *distribuzione generatrice*. In questo schema concettuale fare  $n$  misure di una stessa grandezza fisica in condizioni di ripetitività (che era il problema da cui eravamo partiti nel capitolo 1) equivale a *campionare*  $n$  volte la distribuzione generatrice corrispondente. È ovvio, allora, che in generale non potremo conoscere mai perfettamente la forma della distribuzione generatrice, né le sue caratteristiche (e.g., la media o la deviazione standard); ma è altresì ovvio che, in questo processo di campionamento, acquisiamo progressivamente informazioni sulla distribuzione stessa. Se facciamo, ad esempio, un istogramma dei valori delle misure ottenute, ci aspettiamo che per  $n \rightarrow \infty$  la forma del nostro istogramma tenda alla forma della distribuzione generatrice, e che, a partire dal nostro campione, possiamo fornire stime via via più accurate delle sue caratteristiche.

Se torniamo allora al problema iniziale di scrivere il risultato di una misura come il dato di un valore centrale e di un'incertezza di misura ad esso associata, cominciamo ad intravedere una possibilità operativa nel linguaggio della teoria della probabilità che abbiamo imparato. Se è vero che il nostro processo di misura equivale al campionamento ripetuto di una distribuzione, allora potremo prendere come miglior stima della grandezza cui siamo interessati una delle misure di tendenza centrale (tipicamente la media) della distribuzione e come incertezza associata la sua deviazione standard—che, come abbiamo visto, rappresenta l'entità delle fluttuazioni attorno al valor medio e fornisce una misura conveniente delle fluttuazioni stesse (e.g., per il teorema di Chebishev).

Da qui in avanti modificheremo la (1) utilizzando la lettera  $\sigma^1$  al posto di  $\Delta$  per sottolineare che non stiamo più parlando di errore massimo ma di *errore statistico*:

$$x = \hat{x} \pm \sigma_x \text{ [unità di misura].} \quad (118)$$

È chiaro che la (118) ha un significato fondamentalmente diverso dalla (1): essa definisce un intervallo che non ci dà più la certezza, ma solo una probabilità ben definita (che chiameremo *livello di confidenza*, o CL, dall'Inglese *confidence level*) di contenere il valore del misurando. (Il fatto di non avere una certezza non dovrebbe preoccuparci perché abbiamo già avuto modo di sottolineare la problematicità della nozione di errore massimo nella situazione in cui le misure fluttuano.)

Nella nostra prescrizione il livello di confidenza corrisponde all'integrale della distribuzione generatrice entro una deviazione standard dalla media, e si può stimare esplicitamente a partire della distribuzione generatrice stessa. Il valore sarà in generale diverso caso per caso, ma sappiamo già che, nei casi tipici, sarà grossolanamente dell'ordine del 60–70%. Per completezza, e per evitare qualsiasi possibile ambiguità, uno può (e dovrebbe) specificare esplicitamente il livello di confidenza, e.g.

$$x = \hat{x} \pm \sigma_x \text{ [unità di misura] (68\% CL)} \quad (119)$$

(o qualunque sia il livello di confidenza scelto.)

---

<sup>1</sup> La scelta, onnipresente in letteratura, di utilizzare la lettera  $\sigma$  per indicare sia la deviazione standard di una distribuzione che l'errore statistico è largamente infelice, perché le due cose sono concettualmente diverse. La prima è una proprietà matematica che si calcola dai principi primi, mentre il secondo, pur rappresentando una stima di una deviazione standard è, nella maggior parte dei casi, una quantità che si stima sperimentalmente. Purtroppo ci adegueremo, sperando che il contesto sia sufficiente ad evitare ambiguità.

#### 4.1 CAMPIONAMENTI SINGOLI

Se conosciamo a priori la deviazione standard  $\sigma$  della distribuzione generatrice del misurando cui siamo interessati, oppure ne abbiamo una stima  $s$ , allora una singola misura (cioè un singolo campionamento) è sufficiente per definire tutte le componenti della (118)—prenderemo il (singolo) valore misurato come valore centrale e  $\sigma$  (oppure  $s$ ) come incertezza associata.

Se conosciamo (esattamente o approssimativamente) la forma della distribuzione generatrice, allora possiamo anche calcolare il livello di confidenza associato ad una deviazione standard, oppure utilizzare come stima dell'incertezza di misura un multiplo o sottomultiplo della deviazione standard per ottenere un livello di confidenza fissato a priori. Torneremo su questo problema specifico nel seguito, e vedremo che utilizzare una deviazione standard come incertezza (con un livello di confidenza che, di conseguenza, dipende dai dettagli della distribuzione generatrice) è comodo in pratica.

► **ESEMPIO 4.1.** Supponiamo di misurare il peso  $m$  di un oggetto con una bilancia digitale con una risoluzione di 1 g. Se il valore indicato dal display è 58 g possiamo assumere che, in assenza di errori sistematici, la distribuzione generatrice del misurando sia uniforme tra 57.5 g e 58.5 g, e possiamo utilizzare la deviazione standard di questa distribuzione (che dall'esercizio 3.29 sappiamo essere  $\sqrt{1/12} \approx 0.29$  g) come incertezza di misura. Sappiamo inoltre che per una distribuzione uniforme la probabilità che la variabile disti meno di una deviazione standard dalla media è pari a circa il 58%, per cui possiamo scrivere

$$m = 58.00 \pm 0.29 \text{ g (58\% CL)}.$$

► **ESEMPIO 4.2.** Il ragionamento dell'esempio precedente si applica banalmente alla misura di una lunghezza con un metro a nastro e, più in generale a tutti gli strumenti digitali—nel caso in cui decidiamo di non interpolare tra le divisioni e, cosa ancora più importante, siamo sicuri di non commettere errori di lettura.

► **ESEMPIO 4.3.** In generale gli strumenti si possono *calibrare* attraverso misure ripetute di grandezze di riferimento (ad esempio misurate preliminarmente con uno strumento di misura più accurato o con un metodo indipendente). Questo permette, nel nostro linguaggio, di stimare la deviazione standard della distribuzione generatrice (e magari anche la sua forma), che può poi essere utilizzata come incertezza sulla singola misura.

Se la deviazione standard della distribuzione generatrice non è nota a priori, possiamo stimarla a posteriori in base ad una serie di misure ripetute. Vedremo che, in presenza di fluttuazioni statistiche, ripetere la misura più volte è anche un modo per aumentarne la precisione. Da un punto di vista operativo il nostro problema fondamentale diventa: data una serie di  $n$  misure di una stessa grandezza fatte in condizioni di ripetitività, quali sono le migliori stime che possiamo dare della media e della deviazione standard della distribuzione generatrice? Si tratta di un problema che risolveremo in questo capitolo ma, per cominciare, abbiamo bisogno di sapere come si “sommano” le distribuzioni.

#### 4.2 SOMMA DI VARIABILI CASUALI INDIPENDENTI

Date due variabili casuali indipendenti  $x_1$  ed  $x_2$  (continue o discrete) caratterizzate dalle rispettive funzioni di distribuzioni (in generale diverse), la loro somma  $x = x_1 + x_2$  è ovviamente ancora una variabile casuale—ricordate l'esempio 3.21 del lancio di due dadi. Allora è lecito chiedersi quale sia la funzione di distribuzione di  $x$  e come essa si possa calcolare a partire dalle funzioni di distribuzione di  $x_1$  ed  $x_2$ .

In generale si tratta di una domanda non banale, che richiede di identificare e “pesare” opportunamente le diverse configurazioni di  $x_1$  ed  $x_2$  che danno origine allo stesso valore della somma  $x$ . Anche nel caso più semplice possibile, quello in cui  $x_1$  ed  $x_2$  sono variabili casuali continue distribuite entrambe uniformemente tra 0 ed 1, la risposta non è scontata. Possiamo sicuramente dire che  $x$  sarà compresa tra 0 e 2; ma cosa altro? La densità di probabilità di  $x$  sarà uniforme in questo intervallo?

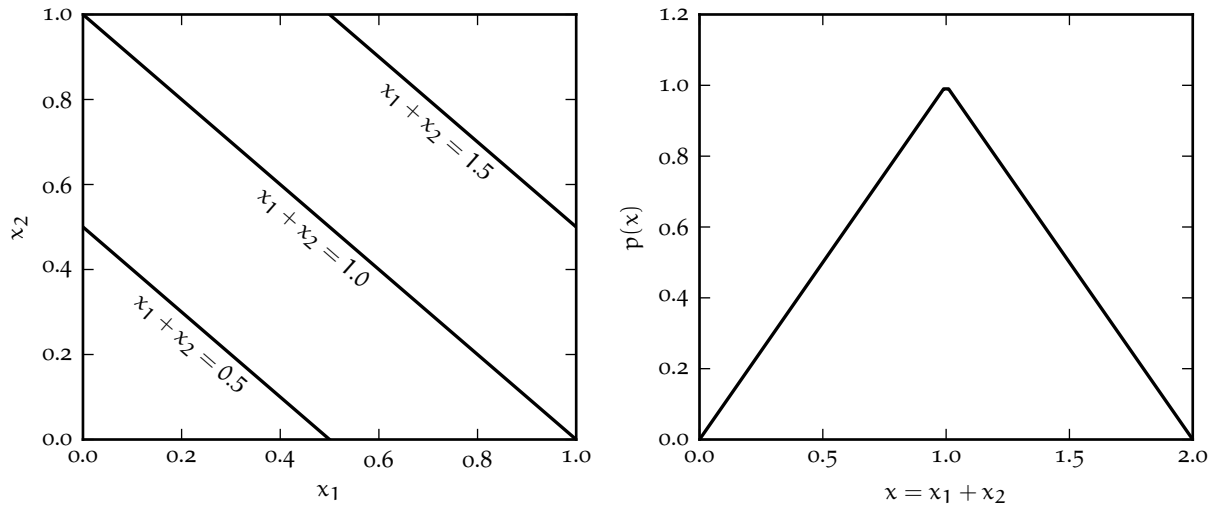


FIGURA 4.1. Rappresentazione geometrica della convoluzione di due funzioni di distribuzione uniformi (a sinistra). Se  $x_1$  ed  $x_2$  sono uniformemente distribuite tra 0 ed 1, allora le regioni del piano cartesiano a somma costante sono segmenti di retta con coefficiente angolare  $-1$  e se  $x_1$  ed  $x_2$  sono indipendenti, la probabilità che  $x_1 + x_2$  assuma un valore fissato è proporzionale alla lunghezza del segmento corrispondente. I segmenti si riducono a punti per i casi limite  $x_1 + x_2 = 0$  e  $x_1 + x_2 = 2$ , ove la densità di probabilità della somma (mostrata a destra) deve tendere a 0.

In questo caso l'intuizione ci viene in aiuto nella forma del grafico mostrato in figura 4.1: le configurazioni nel piano  $x_1$ - $x_2$  che danno origine ad un valore della somma  $x$  fissato sono linee diagonali con pendenza  $-1$ , e la probabilità corrispondente sarà dunque proporzionale alla lunghezza delle linee stesse—identicamente nulla in  $x = 0$  ed  $x = 2$  (ove le linee si riducono a punti), massima in  $x = 1$  e linearmente crescente/decescente nelle due metà dell'intervallo  $[0, 2]$ . La densità di probabilità cercata è dunque triangolare

$$p(x) = \begin{cases} x & \text{se } x \leq 1 \\ (2-x) & \text{se } x > 1, \end{cases}$$

come mostrato in figura 4.1. (Se questo vi sorprende, tornate per un attimo indietro a guardare la figura 3.6 e gli esempi 3.20 e 3.21—si tratta sostanzialmente dello stesso fenomeno, visto questa volta per una variabile casuale continua.)

#### 4.2.1 Media e varianza della somma di variabili casuali indipendenti

Torniamo al nostro problema iniziale, ovvero le proprietà della funzione di distribuzione della somma  $x = x_1 + x_2$  di due variabili casuali indipendenti. Abbiamo visto che calcolare esplicitamente la funzione di distribuzione, nel caso generale, non è banale. Fortunatamente il calcolo della media e della varianza di  $x$  sono problemi più facilmente trattabili. La prima si scrive banalmente come

$$\mu = E[x] = E[x_1 + x_2] = E[x_1] + E[x_2] = \mu_1 + \mu_2. \quad (120)$$

E la varianza  $\sigma^2$  è leggermente più complicata

$$\begin{aligned} \sigma^2 = \text{Var}(x) &= E[(x - \mu)^2] = E[x^2] - \mu^2 = E[(x_1 + x_2)^2] - (\mu_1 + \mu_2)^2 = \\ &= E[x_1^2] + E[x_2^2] + 2E[x_1 x_2] - \mu_1^2 - \mu_2^2 - 2\mu_1 \mu_2 = \sigma_1^2 + \sigma_2^2 + 2E[x_1 x_2] - 2\mu_1 \mu_2. \end{aligned}$$

(Nell'ultimo passaggio  $\sigma_1^2$  e  $\sigma_2^2$  sono le varianze di  $x_1$  ed  $x_2$ , rispettivamente.) Nell'espressione che abbiamo ottenuto riconosciamo la covarianza delle due variabili di partenza, per cui nel caso generale possiamo scrivere

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2\text{Cov}(x_1, x_2) \quad (121)$$

e, nel caso in cui  $x_1$  ed  $x_2$  sono indipendenti, la loro covarianza è per definizione nulla e la (121) si semplifica in

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 \quad \text{da cui} \quad \sigma = \sqrt{\sigma_1^2 + \sigma_2^2} \quad (122)$$

La scrittura nell'ultimo passaggio della (122) prende generalmente il nome di *somma in quadratura*. Notiamo che la somma in quadratura è in generale più piccola della somma propriamente detta

$$\sqrt{\sigma_1^2 + \sigma_2^2} \leq (\sigma_1 + \sigma_2),$$

come si può dimostrare banalmente per via geometrica (pensate al teorema di Pitagora).

A posteriori questo risultato non deve sorprendere—nel capitolo 3 abbiamo visto, e.g., che la media della somma delle uscite di due dadi equi è il doppio della media dell'uscita di un dado equo (esempi 3.24 e 3.25) e che la varianza della somma delle uscite di due dadi equi è il doppio della varianza dell'uscita di un dado equo (esempi 3.27 e 3.28). L'esempio 4.4 mostra che la stessa cosa è vera per la distribuzione triangolare in figura 4.1.

► **ESEMPIO 4.4.** Consideriamo la distribuzione triangolare in figura 4.1. Poiché essa è simmetrica rispetto a  $x = 1$ , la media sarà proprio  $\mu = 1$ . Per la varianza possiamo al solito utilizzare la (78)

$$\begin{aligned} E[x^2] &= \int_0^2 x^2 p(x) dx = \int_0^1 x^3 dx + \int_1^2 x^2(2-x) dx = \left. \frac{x^4}{4} \right|_0^1 + \left. \frac{2x^3}{3} - \frac{x^4}{4} \right|_1^2 = \\ &= \left( \frac{1}{4} - 0 \right) + \left( \frac{16}{3} - \frac{2}{3} \right) - \left( \frac{16}{4} - \frac{1}{4} \right) = \frac{3 - 0 + 64 - 8 - 48 + 3}{12} = \frac{14}{12}, \end{aligned}$$

da cui

$$\sigma^2 = E[x^2] - \mu^2 = \frac{14}{12} - 1 = \frac{2}{12} = \frac{1}{6}.$$

La distribuzione in questione è la “somma” di due distribuzioni uniformi tra 0 ed 1 identiche, con media  $\mu_1 = \mu_2 = 1/2$  e varianza  $\sigma_1^2 = \sigma_2^2 = 1/12$ , come illustrato negli esempi 3.26 e 3.29.

La (120) e la (122) si generalizzano banalmente alla somma di un numero arbitrario di variabili casuali indipendenti  $x = \sum_{i=1}^n x_i$  con medie  $E[x_i] = \mu_i$  e varianze  $\text{Var}(x_i) = \sigma_i^2$ . Dette  $\mu = E[x]$  e  $\sigma^2 = \text{Var}(x)$  si ha

$$\mu = \sum_{i=1}^n \mu_i \quad \text{e} \quad \sigma^2 = \sum_{i=1}^n \sigma_i^2 \quad \text{ovvero} \quad \sigma = \sqrt{\sum_{i=1}^n \sigma_i^2} \quad (123)$$

In altre parole abbiamo il risultato fondamentale che *date  $n$  variabili casuali indipendenti: la media della somma è uguale alla somma delle medie e la varianza della somma è uguale alla somma delle varianze—da cui la deviazione standard della somma è uguale alla somma in quadratura delle deviazioni standard.* (Notiamo, per inciso, come quest'ultima proprietà sia uno dei motivi fondamentali per definire la varianza come abbiamo fatto nella (75). Se avessimo usato, ad esempio, il modulo al posto del quadrato, non esisterebbe un equivalente altrettanto semplice di questa legge di somma.)

#### 4.2.2 Una applicazione interessante: il random walk unidimensionale



Consideriamo un punto materiale vincolato a muoversi su una retta (per fissare le idee fissiamo un sistema di riferimento diretto lungo la retta e la cui origine coincide con la posizione del punto all'istante iniziale). Supponiamo inoltre che il moto del nostro punto avvenga nella forma di una successione (arbitrariamente lunga) di passi elementari di lunghezza prefissata  $\ell$ , nella direzione delle coordinate positive o negative con uguale probabilità. Siamo interessati a studiare la posizione  $x_n$  del punto dopo  $n$  passi—che, va da sé, è una variabile casuale.

La variabile casuale  $s$  che descrive il singolo passo può assumere i valori  $+\ell$  e  $-\ell$  con funzione di distribuzione

$$P(s = +\ell) = P(s = -\ell) = \frac{1}{2}, \quad (124)$$

ed è facile calcolarne media e varianza:

$$E[s] = \ell \times \frac{1}{2} - \ell \times \frac{1}{2} = 0 \quad \text{e} \quad \text{Var}(s) = E[s^2] - \mu_s^2 = E[s^2] = \ell^2 \times \frac{1}{2} + \ell^2 \times \frac{1}{2} = \ell^2.$$

La variabile casuale  $x_n$  (cioè la posizione cercata) sarà data dunque dalla somma di  $n$  variabili casuali identiche distribuite come  $s$ . Ma allora sappiamo immediatamente che

$$E[x_n] = nE[s] = 0 \quad \text{e} \quad \text{Var}(x_n) = E[x_n^2] - \mu_{x_n}^2 = E[x_n^2] = nE[s^2] = n\ell^2, \quad (125)$$

cioè la posizione media del punto dopo un numero arbitrario di passi  $n$  è 0 (il che non dovrebbe sorprendere, dato che, ad ogni passo, ci muoviamo con uguale probabilità verso destra o verso sinistra) e la varianza di  $x_n$  è pari ad  $n\ell^2$ —per cui la sua deviazione standard è proporzionale a  $\sqrt{n}\ell$ .

Il processo che abbiamo appena descritto prende generalmente il nome di *random walk* unidimensionale. La figura 4.2 mostra, a scopo illustrativo, 15 realizzazioni indipendenti dei primi 500 passi di un *random walk* unidimensionale con lunghezza unitaria  $\ell = 1$ .

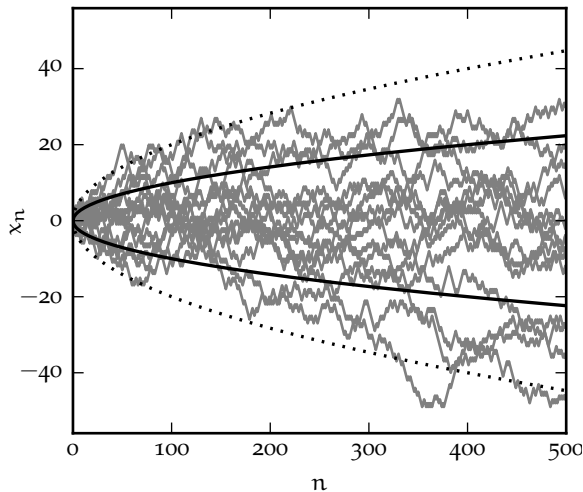


FIGURA 4.2. 15 realizzazioni indipendenti dei primi 500 passi di un *random walk* unidimensionale con lunghezza unitaria  $\ell = 1$  (in unità di misura arbitrarie). La posizione  $x_n$  è mostrata in funzione del numero di passo  $n$ . Le linee continue e tratteggiate indicano i contorni ad 1 e 2 deviazioni standard, cioè  $x_n = \pm\sqrt{n}$  e  $x_n = \pm 2\sqrt{n}$ .

La cosa interessante del *random walk* è che la distanza media  $d_n$  dall'origine al passo  $n$ -esimo, che sarà proporzionale alla deviazione standard di  $x_n$ , scala con la radice quadrata di  $n$  per la (125)

$$d_n \propto \ell\sqrt{n}, \quad (126)$$

(mentre se uno viaggiasse in linea retta si avrebbe banalmente  $d_n = \ell n$ ). La differenza non è da poco, e produce spesso situazioni sorprendenti, come quella descritta nell'esempio 4.5

► **ESEMPIO 4.5.** Un ubriaco esce dal bar e compie un *random walk* fino al suo hotel, che si trova sullo stesso viale a distanza  $d = 1$  km. Assumendo che faccia un passo al secondo (verso destra o verso sinistra con probabilità  $1/2$ ) e che la lunghezza (fissa) di ciascun passo sia  $\ell = 1$  m, quanto tempo impiega (in media) per arrivare a destinazione? Trascurando il fattore di proporzionalità nella (126)

$$n \sim (d/\ell)^2 = 10^6$$

e dunque  $t \sim 10^6$  s (ossia quasi due settimane). Per confronto, ad un sobrio servirebbero  $n \sim d/\ell = 10^3$  passi, ossia circa un quarto d'ora.

## 4.3 MISURE RIPETUTE: MEDIA E VARIANZA CAMPIONE

Riprendiamo il filo della discussione. Abbiamo iniziato questo capitolo con la domanda: data una serie di misure  $x_i$  ( $i = 1 \dots n$ ) indipendenti di una stessa grandezza  $x$  fatte in condizioni di ripetitività, quali sono le migliori stime che possiamo dare della media  $\mu$  e della varianza  $\sigma^2$  della distribuzione generatrice?

Come stima  $m$  della media possiamo prendere la media aritmetica delle misure, che viene generalmente chiamata *media campione*

$$m = \frac{1}{n} \sum_{i=1}^n x_i. \quad (127)$$

La media campione ha la buona proprietà che il suo valore di aspettazione è uguale alla media della distribuzione generatrice

$$E[m] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \frac{1}{n} \sum_{i=1}^n \mu = \frac{n\mu}{n} = \mu.$$

Per completezza, un estimatore che soddisfa questa proprietà (cioè il cui valore di aspettazione coincida con la grandezza da stimare) si dice *imparziale*.

La questione della varianza è leggermente più complicata. Formalmente, se conoscessimo a priori il valore  $\mu$  della media della distribuzione generatrice, l'analogo della (127) sarebbe

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2,$$

il cui valore di aspettazione, come prima, coincide con la quantità che vogliamo stimare ( $s$  così scritto sarebbe, cioè, imparziale):

$$E[s^2] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2\right] = \frac{1}{n} \sum_{i=1}^n E[(x_i - \mu)^2] = \frac{1}{n} \sum_{i=1}^n \sigma^2 = \frac{1}{n} \times n\sigma^2 = \sigma^2.$$

Il problema è che in generale non conosciamo  $\mu$ —abbiamo solo la stima  $m$  data dalla (127). Come miglior stima della varianza della distribuzione generatrice potremmo allora essere tentati di prendere

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2 \quad (128)$$

E qui arriva la parte più interessante. La prima cosa che possiamo chiederci è se il valore di aspettazione della (128) sia ancora pari a  $\sigma^2$ . Come vedremo tra un attimo la risposta è no, ed un modo immediato per rendersene conto è che, detta  $\xi$  una generica stima della media, la derivata

$$\frac{ds_n^2}{d\xi} = \frac{d}{d\xi} \left( \frac{1}{n} \sum_{i=1}^n (x_i - \xi)^2 \right) = -\frac{2}{n} \sum_{i=1}^n (x_i - \xi) = -2 \left( \frac{1}{n} \sum_{i=1}^n x_i - \frac{1}{n} \times n\xi \right) = -2(m - \xi)$$

si annulla per  $\xi = m$ —cioè la (127) è la stima della media che *minimizza* la stima della varianza campione (128). Questo ci dice che, in media,  $s_n^2$  è una sottostima di  $\sigma^2$ . Il valore di aspettazione di  $s^2$  si calcola esplicitamente come

$$E[s_n^2] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i - m)^2\right] = \frac{1}{n} \sum_{i=1}^n E[x_i^2 + m^2 - 2mx_i]$$

e, dato che i valori di aspettazione all'interno dell'ultima sommatoria sono identici poichè gli  $x_i$  hanno tutti la stessa distribuzione generatrice, basta in effetti calcolarne uno solo. Consideriamo dunque i pezzi uno alla volta; il primo si calcola facilmente come

$$E[x_i^2] = \sigma^2 + \mu^2.$$

Il secondo è più complicato

$$\begin{aligned} E[m^2] &= E\left[\frac{1}{n^2} \sum_{j=1}^n x_j \sum_{k=1}^n x_k\right] = \frac{1}{n^2} E\left[\sum_{j,k=1}^n x_j x_k\right] = \frac{1}{n^2} E\left[\sum_{j=1}^n x_j^2 + \sum_{j=1}^n \sum_{k \neq j}^n x_j x_k\right] = \\ &= \frac{1}{n^2} \left( \sum_{j=1}^n E[x_j^2] + \sum_{j=1}^n \sum_{k \neq j}^n E[x_j x_k] \right) = \frac{1}{n^2} \left( n(\sigma^2 + \mu^2) + \sum_{j=1}^n \sum_{k \neq j}^n E[x_j] E[x_k] \right) = \\ &= \frac{1}{n^2} [n(\sigma^2 + \mu^2) + n(n-1)\mu^2] = \frac{1}{n} \sigma^2 + \mu^2. \end{aligned}$$

(Abbiamo spezzato la somma doppia in due termini, a seconda del fatto che gli indici  $i$  e  $j$  siano uguali o diversi. Nel primo caso si ha sostanzialmente il valore di aspettazione di  $x_j^2$ , mentre nel secondo, essendo le misure indipendenti, il valore di aspettazione di  $x_j x_k$  si può fattorizzare come prodotto di valori di aspettazione delle singole variabili.) L'ultimo pezzo vale

$$E[mx_i] = E\left[\frac{1}{n} \sum_{j=1}^n x_j x_i\right] = \frac{1}{n} E\left[x_i^2 + \sum_{j \neq i}^n E[x_i x_j]\right] = \frac{1}{n} (\sigma^2 + \mu^2 + (n-1)\mu^2) = \frac{1}{n} \sigma^2 + \mu^2.$$

Mettendo tutto insieme si ottiene il risultato

$$E[s_n^2] = \sigma^2 + \mu^2 + \frac{1}{n} \sigma^2 + \mu^2 - \frac{2}{n} \sigma^2 - 2\mu^2 = \sigma^2 - \frac{1}{n} \sigma^2 = \frac{(n-1)}{n} \sigma^2.$$

La varianza campione, scritta come la (128), non è imparziale se non asintoticamente, nel senso che il suo valore di aspettazione, in generale, non è uguale alla varianza della distribuzione generatrice se non per  $n \rightarrow \infty$ . Fisicamente questo dipende dal fatto che le  $n$  quantità nella somma (128) non sono tutte indipendenti—sono legate dalla stima della media  $m$  e, se conosco i primi  $n-1$  termini ed  $m$ , posso banalmente calcolare il termine  $n$ -esimo. (Tecnicamente si dice che il fatto di utilizzare  $m$  al posto di  $\mu$  causa la perdita di un grado di libertà.)

Si può ottenere una stima imparziale della varianza della distribuzione generatrice semplicemente moltiplicando per il fattore correttivo  $n/(n-1)$

$$s_{n-1}^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - m)^2 \quad \text{che soddisfa banalmente} \quad E[s_{n-1}^2] = \sigma^2. \quad (129)$$

Per  $n$  grande le due definizioni coincidono, ma in alcuni testi la proprietà di imparzialità è considerata così importante che la varianza del campione è *definita* secondo la (129). Noi utilizzeremo questa definizione, ma mantenendo una posizione pragmatica e non ideologica, come discusso brevemente nella prossima sezione.

#### 4.3.1 Digressione: l'imparzialità è davvero irrinunciabile?

A questo punto la domanda è d'obbligo: è più "giusta" la (128) o la (129)? Nella maggioranza dei casi si tratta di un problema più formale che sostanziale—il rapporto tra le due tende a 1 per  $n$  grande

$$\lim_{n \rightarrow \infty} \frac{s_n^2}{s_{n-1}^2} = \lim_{n \rightarrow \infty} \frac{(n-1)}{n} = 1$$

e la differenza numerica tra  $s_n^2$  e  $s_{n-1}^2$  è entro il 10% per  $n > 10$ . Per piccoli campioni, però, la domanda è rilevante, e la diversa distribuzione delle due stime è mostrata in figura 4.3 per 100 000 campioni di prova con  $n = 8$  elementi estratti da una distribuzione generatrice di varianza (nota)  $\sigma^2 = 1$ .

Notiamo innanzitutto che le due distribuzioni sono asimmetriche ed in entrambi i casi la moda (cioè il valore più probabile) è significativamente diversa dal valore vero di  $\sigma^2$  indicato dalla linea verticale tratteggiata. La stima imparziale  $s_{n-1}^2$  ha (come abbiamo visto) media  $\sigma^2$ , ma il fattore correttivo che garantisce l'imparzialità ha l'effetto collaterale di enfatizzare la coda a destra della distribuzione—cioè di

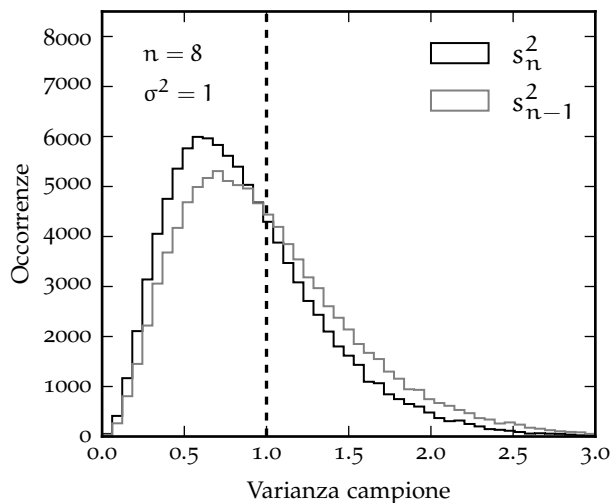


FIGURA 4.3. Varianza campione, calcolata secondo le (128) e (129) per 100 000 campioni di  $n = 8$  elementi estratti da una distribuzione generatrice di varianza  $\sigma^2 = 1$  (indicata della linea verticale tratteggiata). La stima imparziale è mostrata in grigio, mentre quella “non corretta” è in nero. Entrambe le distribuzioni sono evidentemente asimmetriche; la media di  $s_{n-1}^2$  coincide con la varianza  $\sigma^2$  della distribuzione generatrice, ma allo stesso tempo la coda a destra è più pronunciata.

aumentare la frequenza relativa dei valori più distanti da  $\sigma^2$ . Il risultato è che, in questo caso particolare, tra le 100 000 realizzazioni mostrate in figura 4.3 la stima non corretta  $s_n^2$  della varianza è più vicina al valore vero  $\sigma^2$  rispetto alla stima imparziale  $s_{n-1}^2$  in circa il 62% dei casi. Siamo cioè nella situazione interessante in cui una stima è *giusta in media* e l'altra è più vicina al valore vero nella maggior parte dei casi. (Aggiungiamo, per completezza, che il fatto che  $s_{n-1}^2$  sia una stima imparziale di  $\sigma^2$  non implica di per sé che la sua radice  $s_{n-1}$  sia una stima imparziale di  $\sigma$ .)

È tempo di passare oltre—abbiamo discusso il problema abbastanza a lungo e, come detto, si tratta tutto sommato di una questione inessenziale. Utilizziamo pure la (129) ma teniamo a mente il fatto che essa non ha niente di fondamentalmente più giusto della (128).

#### 4.3.2 La deviazione standard della media

Ricapitoliamo ciò che abbiamo imparato fino a questo momento. Abbiamo modellizzato il problema della misura di una grandezza fisica in condizioni di ripetitività come il campionamento di una variabile casuale caratterizzata da una funzione di distribuzione (in generale ignota) che chiamiamo distribuzione generatrice, e abbiamo derivato le relazioni fondamentali per la migliore stima della media e della varianza di questa distribuzione generatrice a partire da un campione finito  $x_1 \dots x_n$  di misure. La domanda ovvia adesso è: qual è la relazione tra queste stime ed il modo in cui scriviamo il risultato di una misura?

La media campione (127) è il candidato ovvio per il valore centrale della misura—questa è la parte semplice. Per l'incertezza associata potremmo essere tentati di prendere la varianza campione (129), ma qui le cose si fanno più complicate, perché la varianza campione è rappresentativa delle fluttuazioni della *singola misura*, mentre ciò a cui siamo interessati sono le fluttuazioni del valore centrale, cioè della media. E, intuitivamente è lecito aspettarsi che le fluttuazioni della media di  $n$  misure siano più piccole di quelle sulla singola misura—e tanto più piccole quanto più grande è  $n$ . (In fondo, se così non fosse, perché le persone si affannerebbero a prendere dati il più a lungo possibile?)

Torniamo dunque per un attimo alla media campione (127). Essendo una media di variabili casuali,  $m$  è essa stessa una variabile casuale anch'essa. Se siamo interessati alle fluttuazioni di  $m$ , la domanda giusta, nel nostro linguaggio, è il valore della deviazione standard di  $m$ :

$$\text{Var}(m) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(x_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}, \quad (130)$$

da cui la deviazione standard della media  $\sigma_m$  è banalmente

$$\sigma_m = \sqrt{\text{Var}(m)} = \frac{\sigma}{\sqrt{n}}. \quad (131)$$

La (131) è un risultato di fondamentale importanza, che ci dice nel dettaglio come le fluttuazioni della media di  $n$  misure si riducono, al variare di  $n$ , rispetto alle fluttuazioni sulla singola misura. Possiamo



utilizzare questo risultato, insieme alla stima della varianza campione, per la stima della varianza e della deviazione standard della media  $s_m$ :

$$s_m^2 = \frac{1}{n(n-1)} \sum_{i=1}^n (x_i - m)^2 \quad \text{e} \quad s_m = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - m)^2}. \quad (132)$$

Questo risponde alla nostra domanda iniziale, e con il nostro nuovo bagaglio di conoscenze (127) e (129) possiamo riscrivere la (118) nella forma che utilizzeremo in pratica

$$x = m \pm s_m [\text{unità di misura}]. \quad (133)$$

Come abbiamo già avuto modo di osservare, il significato di questa scrittura è profondamente diverso da quello della (1)—essa ci dice che abbiamo una probabilità ben definita che l'intervallo  $[m - s_m, m + s_m]$  contenga il misurando. Se conosciamo la funzione di distribuzione di  $m$  possiamo calcolare questa probabilità, e vedremo più avanti che per  $n$  abbastanza grande il teorema centrale del limite ci garantisce questa possibilità.

#### 4.3.3 Un esempio concreto: la statura dei coscritti

Illustriamo la differenza tra la deviazione standard del campione e la deviazione standard della media con un esempio concreto. La tabella 12 riporta alcuni dati (anno di nascita, numero di iscritti, media campione e suddivisione in classi) relativi alla statura degli iscritti di leva nati negli anni 1972–1980, che l'Istituto Nazionale di Statistica mette a disposizione insieme a decine di altre serie storiche interessanti.

Anno di nascita	Iscritti	Statura media	Ripartizione percentuale per classi di statura							
			< 150	150–154	155–159	160–164	165–169	170–174	175–179	≥ 180
1972	445451	173.96	0.0	0.1	1.6	6.8	18.6	29.1	25.2	18.6
1973	440115	174.12	0.1	0.2	1.2	6.5	18.1	29.3	25.8	18.8
1974	429159	174.18	0.1	0.2	1.1	6.1	17.8	29.1	26.0	19.6
1975	382065	174.29	0.1	0.2	1.1	6.0	17.4	29.5	26.0	19.7
1976	366357	174.42	0.1	0.4	1.1	6.2	17.8	28.4	26.0	20.0
1977	324080	174.44	0.1	0.3	1.0	6.3	17.9	28.4	26.0	20.0
1978	366804	174.45	0.1	0.3	1.0	6.1	18.0	28.5	26.0	20.0
1979	368336	174.49	0.1	0.3	1.0	6.1	18.0	28.3	26.1	20.1
1980	369310	174.58	0.1	0.3	1.0	6.1	18.0	28.3	26.1	20.1

TABELLA 12. Statura media e ripartizione percentuale per classi di statura degli iscritti di leva nati negli anni 1972–1980 (adattato dalle serie storiche dell'Istituto Nazionale di Statistica). Tutte le stature sono misurate in cm.

La ripartizione percentuale è interessante perché ci dà un'idea diretta della forma della distribuzione generatrice delle stature. Purtroppo la definizione degli intervalli per la ripartizione in classi è rimasta invariata dal 1854 e, a causa dell'aumento della statura media avvenuta negli ultimi 120 anni, non permette di campionare adeguatamente per anni recenti la distribuzione generatrice al di sopra di 180 cm (ove abbiamo una sola classe). Questo rende difficile valutarne con precisione la deviazione standard, ma sulla base del fatto che la maggior parte delle misure cadono entro quattro intervalli (di 5 cm) attorno al valor medio possiamo dire rozzamente che essa sarà dell'ordine di  $\sim 10$  cm. *Questo è il numero rappresentativo delle fluttuazioni intorno al valor medio per la singola misura.* Se prendiamo una persona a caso nata negli anni settanta ci aspettiamo che la sua statura sia  $174 \pm 10$  cm (a patto che la nostra stima ad occhio della deviazione standard possa essere considerata affidabile).

Tutt'altro discorso vale per la media delle stature su un campione di persone. La terza colonna della tabella indica che le stature medie fluttuano negli anni molto meno di questi 10 cm. Con circa 400 000 misure ci aspettiamo che la deviazione standard della media sia circa 630 volte più piccola di quella del campione—nel nostro caso specifico dell'ordine di  $\sim 0.2$  mm. *Questo è il numero rappresentativo delle*

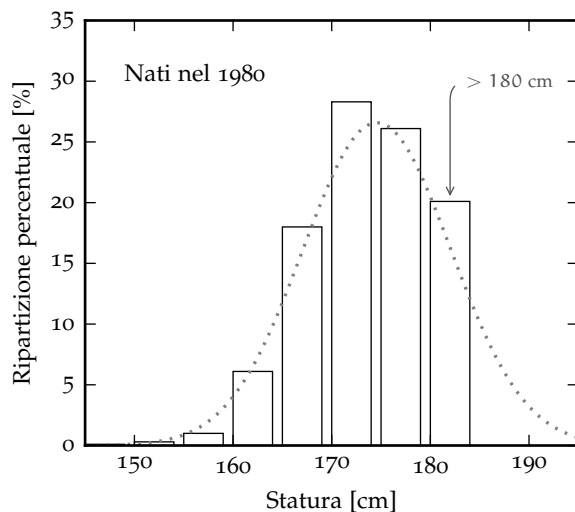


FIGURA 4.4. Ripartizione percentuale per classi di statura dei nati nel 1980, in base ai dati della tabella 12. La classe più a destra include tutti i coscritti più alti di 180 cm, il che rende difficile stimare la varianza del campione. La linea grigia tratteggiata rappresenta, a scopo puramente illustrativo, una possibile distribuzione generatrice.

fluttuazioni intorno al valor medio per la media di circa 400 000 misure. In effetti, a guardare bene la tabella, la media del campione cresce sistematicamente con il tempo—e di circa 1 mm l'anno, come mostrato in figura 4.5. Il fatto che questa crescita si veda così chiaramente implica che le fluttuazioni statistiche sul singolo punto debbono essere  $\ll 1$  mm, che è compatibile con la nostra stima grossolana di 0.2 mm.

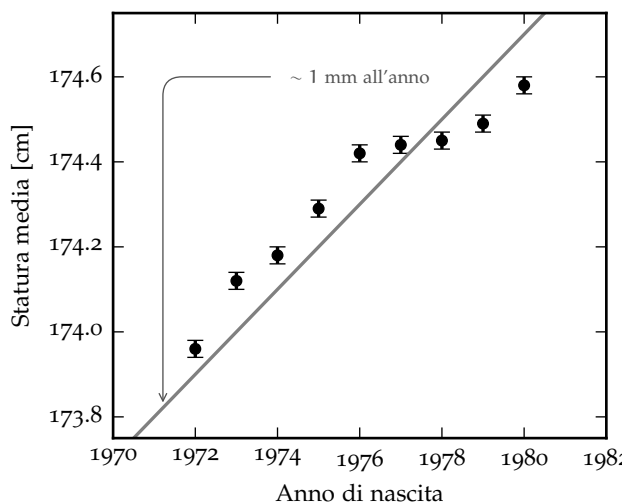


FIGURA 4.5. Andamento temporale della statura media dei coscritti in funzione dell'anno di nascita, in base ai dati della tabella 12. Le barre d'errore corrispondono alla deviazione standard della media per un campione di 400 000 unità, che abbiamo stimato rozzamente essere dell'ordine di 0.2 mm. Si vede chiaramente un aumento significativo della statura in funzione del tempo dell'ordine di  $\sim 1$  mm l'anno, il che implica che le fluttuazioni statistiche dei singoli punto debbono essere  $\ll 1$  mm.

#### 4.4 CALCOLO DELLA MEDIA E VARIANZA CAMPIONE IN PRATICA

È utile a questo punto vedere come si mette in pratica quanto abbiamo visto attraverso alcuni esempi concreti di implementazione (cioè attraverso frammenti di codice funzionanti). Il modo più semplice per calcolare la media e la varianza (o la deviazione standard) di un campione di dati è di fare due cicli separati sui dati stessi—il primo per calcolare la media campione  $m$  secondo la (127) ed il secondo per calcolare la varianza campione  $s_{n-1}^2$ , una volta nota  $m$ . Il frammento 4.1 mostra un esempio di implementazione di questa strategia.

Prima di procedere fermiamoci un secondo ad esaminare l'*output* del frammento 4.1. In questo caso la media e la varianza della distribuzione generatrice sono, per costruzione,  $\mu = 0$  e  $\sigma^2 = 1$  (per cui anche la deviazione standard è  $\sigma = 1$ ). Con un campione di  $n = 10000$  numeri il nostro algoritmo fornisce le stime  $m \approx 0.0098$  e  $\sigma \approx 0.9988$ —che sono vicini, ma ovviamente non identici, ai valori veri che abbiamo scelto in ingresso. La domanda è: questi valori sono abbastanza vicini oppure abbiamo modo di sospettare che qualcosa non vada? Per il valore della deviazione standard del campione non siamo ancora attrezzati per rispondere, ma per ciò che riguarda la media sappiamo che la sua deviazione

```

https://bitbucket.org/.../sample_stat1.py
1 def sample_statistics(sample):
2     """ Calculate the sample mean and (unbiased) standard
3     deviation in the dumbest possible way.
4     """
5     n = len(sample)
6     m = s = 0.
7     # First loop to calculate the mean.
8     for x in sample:
9         m += x
10    m /= n
11    # Second loop to calculate the variance.
12    for x in sample:
13        s += (x - m)**2.0
14    s /= (n - 1)
15    return m, s*0.5
16
17 # Quick test with normally-distributed random numbers.
18 import numpy
19 numpy.random.seed(1)
20 sample = numpy.random.normal(0., 1., size=10000)
21 print(sample_statistics(sample))
22
23 [Output]
24 (0.009772656699104992, 0.99883578672599571)

```

FRAMMENTO 4.1. Codice per il calcolo della media e della deviazione standard del campione secondo le (127) e (129) con due cicli separati. La funzione è applicata ad un campione di 10000 valori estratti da una funzione di distribuzione con media 0 e varianza 1.

standard è  $\sigma_m = \sigma/\sqrt{n} = \sigma/100 = 0.01$ . Ora, la nostra stima della media si discosta dalla media di circa una deviazione standard  $\sigma_m$ , che è proprio quello che ci aspettiamo. (Se ripetiamo l'esercizio con un campione diverso otterremo ovviamente una stima diversa, ma il punto fondamentale è che le fluttuazioni di  $m$  devono essere dello stesso ordine di grandezza di  $\sigma_m$ .)

#### 4.4.1 Un algoritmo alternativo con un ciclo singolo

Con qualche semplice manipolazione algebrica la formula per la varianza campione (129) può essere trasformata in

$$\begin{aligned}
 s_{n-1}^2 &= \frac{1}{(n-1)} \sum_{i=1}^n (x_i - m)^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i^2 + m^2 - 2mx_i) = \frac{1}{(n-1)} \left( \sum_{i=1}^n x_i^2 + nm^2 - 2m \sum_{i=1}^n x_i \right) = \\
 &= \frac{1}{(n-1)} \left( \sum_{i=1}^n x_i^2 + nm^2 - 2nm^2 \right) = \frac{1}{(n-1)} \left( \sum_{i=1}^n x_i^2 - nm^2 \right) = \frac{1}{(n-1)} \sum_{i=1}^n x_i^2 - \frac{n}{(n-1)} m^2,
 \end{aligned}$$

che è l'equivalente della (78) nel contesto della teoria dei campioni. Questa formulazione alternativa ci permette di implementare il calcolo della media e della deviazione standard del campione con un solo ciclo in cui accumuliamo in parallelo le somme parziali di  $x_i$  e  $x_i^2$ , come mostrato nel frammento 4.2. La questione non è semplicemente estetica o accademica, perché evitare il secondo ciclo ci permette di calcolare le variabili campione *in linea* senza bisogno di tenere traccia in memoria di tutti i valori  $x_i$  e di interrompere il processo in qualsiasi momento con tutta l'informazione rilevante nelle nostre mani (che è utile, ad esempio, quando non sappiamo in anticipo la dimensione del campione).

L'*output* di questo algoritmo alternativo è interessante per almeno due ragioni. Con lo stesso campione di test del frammento 4.1 (media  $\mu = 0$  e deviazione standard  $\sigma = 1$ ) otteniamo essenzialmente la stessa risposta—anche se, ad osservare attentamente, i valori delle deviazioni standard nei due casi differiscono tra di loro alla quindicesima cifra decimale. A questo livello si tratta semplicemente del fatto, imprescindibile, che la rappresentazione dei numeri in virgola mobile in un calcolatore è intrinsecamente inesatta. In questo caso particolare questo piccolo errore numerico è privo di conseguenze pratiche.

Ma una cosa ancora più interessante avviene quando sommiamo un numero grande ( $1 \times 10^9$  nel caso in questione) a tutti i valori nel campione iniziale. È ovvio che questa operazione è equivalente a campionare

```

https://bitbucket.org/.../sample_stat2.py
1 def sample_statistics(sample):
2     """ Calculate the sample mean and (unbiased) standard
3     deviation with one single loop.
4     """
5     n = len(sample)
6     # Calculate the mean and the variance with one loop.
7     m = s = 0.
8     for x in sample:
9         m += x
10        s += x**2
11    m /= n
12    s = (s - n*m*m)/(n - 1)
13    return m, s**0.5
14
15 # Quick test with normally-distributed random numbers.
16 import numpy
17 numpy.random.seed(1)
18 sample = numpy.random.normal(0., 1., size=10000)
19 print(sample_statistics(sample))
20
21 # And now add 1,000,000 to all the values in the sample.
22 sample += 1.e9
23 print(sample_statistics(sample))
24
25 [Output]
26 (0.009772656699104992, 0.99883578672599171)
27 (1000000000.0097739, 28.964542020709512)

```

FRAMMENTO 4.2. Calcolo della media e della deviazione standard del campione con un solo ciclo. Il codice è (marginalmente) meno complesso dell'equivalente frammento 4.1 ma, se la media del campione in ingresso è molto più grande della sua deviazione standard, in generale richiede la sottrazione di numeri potenzialmente molto grandi e vicini tra loro—per cui diventa numericamente instabile.

una distribuzione con media  $\mu = 1 \times 10^9$  e deviazione standard  $\sigma = 1$ —in particolare la stima della deviazione standard dovrebbe essere identica alla precedente. Eppure, se guardiamo la seconda riga di *output* del frammento 4.2 ci accorgiamo che l'algoritmo fornisce una stima del tutto insensata ( $\sim 30$ ) della deviazione standard stessa. Come è possibile? Il problema è una combinazione dei due fatti: (i) poiché gli  $x_i$  sono grandi, la somma parziale  $\sum x_i^2$  tende a crescere velocemente; e (ii) poiché  $\sigma \ll \mu$ ,  $\sum x_i^2$  e  $(\sum x_i)^2$  non sono molto diversi tra di loro—e quando sottraiamo due numeri molto grandi vicini tra loro, l'errore numerico può diventare arbitrariamente grande. Dal nostro punto di vista impariamo la lezione fondamentale che, a causa del fatto che l'aritmetica in virgola mobile non è esatta, due espressioni esattamente equivalenti se scritte sulla carta, possono fornire risultati diversi una volta implementati su un calcolatore.

#### 4.4.2 L'algoritmo di Welford

A questo punto possiamo chiederci quale sia il modo più robusto ed efficiente per implementare in pratica il calcolo della varianza campione. La risposta è in un algoritmo dovuto a Welford [15] e descritto in dettaglio in [11], di cui mostriamo una semplice implementazione nel frammento 4.3. (Si tratta di un'idea tanto ingegnosa quanto semplice, ed il lettore è senz'altro incoraggiato a consultare l'articolo originale.)

Rimandiamo il lettore alle referenze [15, 11] per una dimostrazione formale della correttezza dell'algoritmo e ci limitiamo a notare come esso funzioni (a parte l'usuale rumore numerico) anche nel caso in cui quello mostrato nel frammento 4.2 fallisce miseramente.

#### 4.4.3 E se io fossi veramente pigro?

Beh, c'è sempre la possibilità di usare i metodi nativi del modulo `numpy` di Python, come mostrato nel frammento 4.4—a patto ovviamente che il campione dei dati sia nella forma di un *array* di `numpy` (o di essere disposti a convertirlo in questa forma nel caso non lo fosse). Oltre ad essere sintatticamente più conciso, il codice sottostante è con ogni probabilità più efficiente e meglio testato di qualsiasi

```

https://bitbucket.org/.../sample_stat3.py
1 def sample_statistics(sample):
2     """ Calculate the sample mean and (unbiased) standard
3     deviation as in D. E. Knuth, TAOCP, Vol. 2, pag. 232.
4     """
5     n = m = s = 0.
6     for x in sample:
7         n += 1
8         delta = x - m
9         m += delta/n
10        s += delta*(x - m)
11    s /= (n - 1)
12    return m, s**0.5
13
14 # Quick test with normally-distributed random numbers.
15 import numpy
16 numpy.random.seed(1)
17 sample = numpy.random.normal(0., 1., size=10000)
18 print(sample_statistics(sample))
19
20 # And now add 1,000,000 to all the values in the sample.
21 sample += 1.e9
22 print(sample_statistics(sample))
23
24 [Output]
25 (0.0097726566991049712, 0.99883578672599072)
26 (1000000000.0097729, 0.99883579287765412)

```

FRAMMENTO 4.3. Calcolo della media e della deviazione standard campione secondo l'algoritmo di Welford [15], che permette di utilizzare un ciclo singolo e, allo stesso tempo, è numericamente stabile.

implementazione un programmatore casuale (come tipicamente è il Fisico medio) possa realizzare partendo da zero.

```

https://bitbucket.org/.../sample_stat4.py
1 import numpy
2
3 def sample_statistics(sample):
4     """ Calculate the sample mean and (unbiased) standard
5     deviation using numpy.
6     """
7     m = sample.mean()
8     s = sample.var(ddof=1)
9     return m, numpy.sqrt(s)
10
11 # Quick test with normally-distributed random numbers.
12 import numpy
13 numpy.random.seed(1)
14 sample = numpy.random.normal(0., 1., size=10000)
15 print(sample_statistics(sample))
16
17 # And now add 1,000,000 to all the values in the sample.
18 sample += 1.e9
19 print(sample_statistics(sample))
20
21 [Output]
22 (0.0097726566991049695, 0.99883578672599194)
23 (1000000000.0097727, 0.99883578624680525)

```

FRAMMENTO 4.4. Calcolo della media e della deviazione standard campione utilizzando i metodi nativi degli *array* del modulo *numpy* di Python. Si noti l'uso dell'argomento *ddof=1* per ottenere l'estimatore imparziale  $s_{n-1}$ , in assenza del quale la funzione calcolerebbe  $s_n$ .

A questo punto abbiamo esaminato quattro implementazioni diverse del calcolo della media e della deviazione standard del campione ed è probabilmente il momento di passare oltre.

## 4.5 COVARIANZA E CORRELAZIONE CAMPIONE

Supponiamo di avere una serie di campionamenti  $x_i$  ed  $y_i$ , con  $i = 1 \dots n$  di due variabili casuali  $x$  ed  $y$ , ovverosia una serie di  $n$  coppie ordinate  $(x_i, y_i)$ . Se diciamo  $m_x$  ed  $m_y$  le rispettive medie campione

$$m_x = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{e} \quad m_y = \frac{1}{n} \sum_{i=1}^n y_i$$

possiamo, analogamente a quanto abbiamo fatto in precedenza con la varianza, stimare la covarianza  $\text{Cov}(x, y)$  delle variabili di partenza attraverso la covarianza del campione

$$q_{xy} = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - m_x)(y_i - m_y). \quad (134)$$

Abbiamo sviscerato ampiamente la questione del  $(n-1)$  nella sezione 4.3.1 e non insisteremo oltre, così come non insisteremo oltre sull'efficienza e la stabilità degli algoritmi a disposizione per il calcolo della (134) in pratica.

La stima della correlazione  $r_{xy}$  (o semplicemente  $r$ ) del campione si scrive banalmente in termini delle stime  $s_x$  ed  $s_y$  delle deviazioni standard di  $x$  ed  $y$

$$s_x^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - m_x)^2 \quad \text{e} \quad s_y^2 = \frac{1}{(n-1)} \sum_{i=1}^n (y_i - m_y)^2$$

come

$$r_{xy} = \frac{q_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - m_x)(y_i - m_y)}{\sqrt{\sum_{i=1}^n (x_i - m_x)^2 \sum_{i=1}^n (y_i - m_y)^2}} \quad (135)$$

che è talvolta chiamato anche *coefficiente di correlazione lineare* o *indice di correlazione di Pearson* ed è molto utilizzato in pratica.

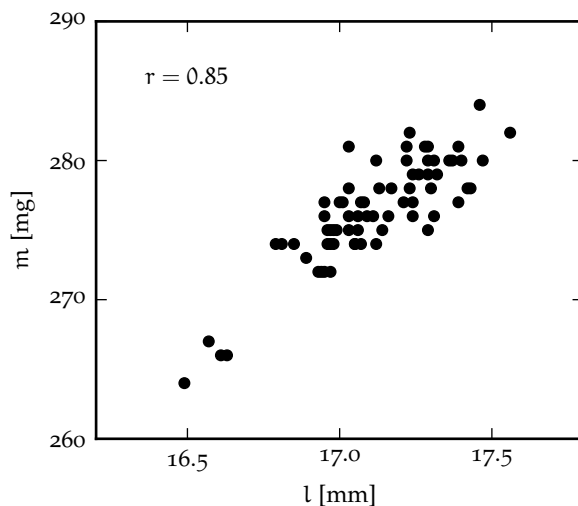


FIGURA 4.6. Grafico di dispersione della massa  $m$  in funzione della lunghezza  $l$  per un campione di 70 chiodini (le misure sono state realmente effettuate il 5 agosto 2005 da uno degli autori). Le misure di lunghezza sono state effettuate con un calibro Palmer (risoluzione 0.01 mm) e quelle di massa con una bilancia di precisione (risoluzione 1 mg), per cui le incertezze di misura sono molto più piccole della dispersione delle due quantità. È chiaro che a valori piccoli (grandi) della lunghezza tendono ad essere associati valori piccoli (grandi) della massa, per cui le due quantità sono correlate. L'indice di correlazione di Pearson, calcolato secondo la (135), è 0.85.

Uno dei modi più semplici per individuare una possibile correlazione tra due variabili in un campione è quello di realizzare un grafico di dispersione, come mostrato in figura 4.6. Se valori piccoli di una variabile tendono ad essere associati a valori piccoli (o grandi) dell'altra, e viceversa, questo è indice di una correlazione positiva (o negativa). Se le fluttuazioni attorno al valor medio delle due quantità appaiono indipendenti, cioè la distribuzione dei punti nel diagramma di dispersione non è elongata in una direzione privilegiata, allora le quantità sono *scorrelate*. L'indice di Pearson indica l'entità della correlazione ed è  $\pm 1$  se e solo se le due grandezze nel grafico di dispersione stanno esattamente su una retta.

► **ESEMPIO 4.6.** Si misurano la massa  $m$  (con una bilancia di precisione) e la lunghezza  $l$  (con un calibro Palmer) per un campione di 70 chiodini dello stesso materiale. Il grafico di dispersione delle due grandezze, mostrato in figura 4.6, indica una chiara correlazione, con un indice di Pearson di 0.85. (Si noti che in questo caso le incertezze di misura sono molto più piccole della dispersione intrinseca della due quantità.) In questo caso l'origine della correlazione è Fisica, e deriva dal fatto che i chiodini sono tutti fatti dello stesso materiale, per cui è naturale aspettarci che quelli più lunghi pesino di più (e viceversa). Se i chiodini fossero esattamente cilindrici, allora la lunghezza sarebbe proporzionale al volume e ci aspetteremmo che la correlazione fosse perfetta, poiché il rapporto tra massa e volume è la densità—ed è fissato. In pratica la presenza della testa e della punta (che saranno leggermente diverse da caso a caso) rendono la correlazione imperfetta.

#### 4.5.1 Correlazioni e rapporti di causa effetto

Per quanto la tentazione di fare il collegamento possa essere forte, non si può sottolineare abbastanza il fatto che *una correlazione tra due grandezze non implica necessariamente un rapporto di causa-effetto tra le grandezze stesse*. In altre parole, se  $x$  ed  $y$  sono correlate non è detto che  $x$  causi  $y$  o  $y$  causi  $x$ . Lo abbiamo visto in concreto nel caso dei chiodini dell'esempio 4.6, in cui non diremmo che la massa causa la lunghezza o viceversa, e la correlazione, più prosaicamente, è determinata da un fattore fisico (il fatto che i chiodi siano fatti tutti dello stesso materiale) che influenza entrambe le grandezze sotto studio. Negli studi clinici e sociologici, tanto per fare un esempio, spesso non è semplice fattorizzare l'effetto che determinate concause ambientali possono avere sulle grandezze oggetto di studio, e questo costituisce una delle difficoltà maggiori nel passaggio delicato dall'osservazione di una correlazione allo stabilire una relazione di causa ed effetto.

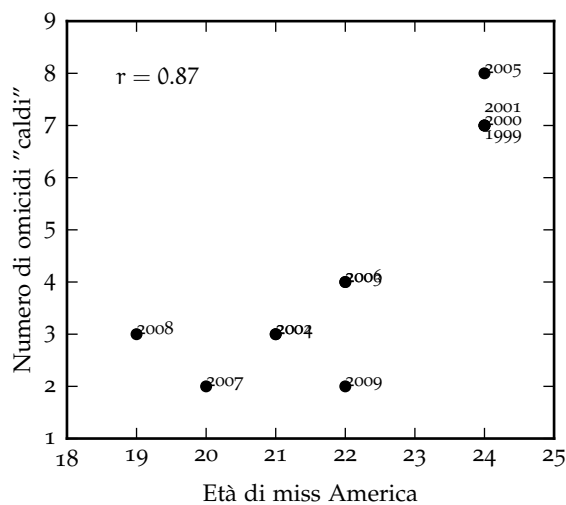


FIGURA 4.7. Grafico di dispersione del numero di omicidi per anno commessi negli Stati Uniti utilizzando vapore o oggetti caldi in funzione dell'età di Miss America (nello stesso anno)—per gli anni 1999–2009 (adattato da <http://tylervigen.com/spurious-correlations>). L'indice di correlazione di Pearson è 0.87 ma in questo caso è ovvio che la correlazione è puramente casuale e nessuno direbbe che l'età di Miss America sia legata causalmente con il numero di omicidi.

La materia è complessa e non abbiamo il tempo di sviscerarla in dettaglio. La figura 4.7 è un buon modo per convincersi del punto principale di questa sezione—correlazione e causalità non sono la stessa cosa—e che, a cercare bene è sorprendentemente facile trovare correlazioni tra cose che non hanno niente a che vedere l'una con l'altra. Se non sapevate che il numero di persone morte per affogamento in piscina per anno negli Stati Uniti è positivamente correlato con il numero di film (dello stesso anno) in cui Nicolas Cage appare, probabilmente è giunto il momento di dare un'occhiata a <http://tylervigen.com/spurious-correlations>.

## 4.6 MEDIA E VARIANZA DI UNA FUNZIONE DI VARIABILI CASUALI

Nella sezione 4.2 abbiamo imparato a *sommare* le distribuzioni. Prima di tornare di nuovo sulla propagazione degli errori (statistici) ci rimane ancora un ulteriore piccolo passo da fare. La domanda che ci

facciamo in questa sezione è la seguente: come possiamo stimare, anche approssimativamente, la media e la varianza di una funzione generica di variabili casuali?

#### 4.6.1 Il caso unidimensionale

Consideriamo dunque una variabile casuale  $x$ , di cui assumiamo di sapere la funzione di distribuzione—e sia  $\mu$  il valor medio di  $x$  e  $\sigma$  la sua deviazione standard. Come detto all'inizio della sezione, il nostro scopo è quello di stimare la media  $\mu_f$  e la varianza  $\sigma_f^2$  di una generica funzione  $f(x)$ . Poiché sappiamo che i valori di  $x$  tendono a *concentrarsi* attorno al valor medio, possiamo partire dallo sviluppo in serie di Taylor di  $f(x)$  attorno a  $\mu$  troncato al prim'ordine:

$$f(x) \approx f(\mu) + \frac{df(\mu)}{dx}(x - \mu) \quad \text{ovvero} \quad f(x) - f(\mu) \approx \frac{df(\mu)}{dx}(x - \mu).$$

La media si calcola formalmente come il valore di aspettazione  $E[f(x)]$ , che sviluppato al prim'ordine diviene

$$\mu_f = E[f(x)] \approx E\left[f(\mu) + \frac{df(\mu)}{dx}(x - \mu)\right] = E[f(\mu)] + \frac{df(\mu)}{dx}E[x - \mu].$$

(Ricordate:  $f(\mu)$  è una costante che può essere portata fuori dal valore di aspettazione, e così pure la derivata—sono entrambe calcolate in un punto e dunque non dipendono da  $x$ .) A questo punto, notando che  $E[x - \mu]$  si annulla banalmente per la linearità del valore di aspettazione, si ottiene

$$\mu_f \approx f(\mu). \quad (136)$$

Il valor medio di  $f(x)$ , cioè, può essere approssimato con il valore della funzione  $f(x)$  calcolata nel valor medio di  $x$ , a meno di termini del second'ordine in  $(x - \mu)$ . Che il termine al prim'ordine dello sviluppo si annulli è un fatto importante, che ci dice che la 136 vale *esattamente* se  $f(x)$  è una funzione lineare di  $x$ .

Per la varianza possiamo procedere analogamente, utilizzando il nostro sviluppo in serie di  $f(x)$  ed il risultato che abbiamo appena ottenuto per la media

$$\sigma_f^2 = E[(f(x) - \mu_f)^2] \approx E[(f(x) - f(\mu))^2] \approx E\left[\left(\frac{df(\mu)}{dx}(x - \mu)\right)^2\right] = \left(\frac{df(\mu)}{dx}\right)^2 E[(x - \mu)^2],$$

da cui

$$\sigma_f^2 \approx \left(\frac{df(\mu)}{dx}\right)^2 \sigma^2. \quad (137)$$

Notiamo, di passaggio, come la (137) assomigli, almeno ad un'occhiata superficiale, alla (16). Avremo modo di sottolineare nella prossima sezione come le due cose, sia pur connesse, siano fondamentalmente diverse.

#### 4.6.2 Il caso generale

Nel caso multidimensionale (quando, cioè, si ha una funzione  $f(x_1, x_2, x_3, \dots)$  di un certo numero di variabili casuali  $x_1, x_2, x_3, \dots$  con medie  $\mu_1, \mu_2, \mu_3, \dots$  e varianze  $\sigma_1^2, \sigma_2^2, \sigma_3^2, \dots$ ) il linguaggio rimane sostanzialmente lo stesso, ma la situazione diventa più interessante—e rientra in gioco la questione dell'indipendenza statistica tra le variabili che abbiamo già affrontato nella sezione 4.2 a proposito della somma.

Andiamo per gradi e cominciamo dal caso relativamente semplice di una funzione  $f(x_1, x_2)$  di due variabili casuali. Lo sviluppo in serie di Taylor al prim'ordine è in questo caso

$$f(x_1, x_2) \approx f(\mu_1, \mu_2) + \frac{\partial f(\mu_1, \mu_2)}{\partial x_1}(x_1 - \mu_1) + \frac{\partial f(\mu_1, \mu_2)}{\partial x_2}(x_2 - \mu_2).$$



La media di  $f$  (sviluppata al prim'ordine) sarà, esattamente come prima

$$\mu_f = E[f(x_1, x_2)] \approx f(\mu_1, \mu_2) + \frac{\partial f(\mu_1, \mu_2)}{\partial x_1} E[x_1 - \mu_1] + \frac{\partial f(\mu_1, \mu_2)}{\partial x_2} E[x_2 - \mu_2] = f(\mu_1, \mu_2). \quad (138)$$

Per la varianza il calcolo è leggermente più complicato. Partendo dalla definizione si ha, ricordando che come nel caso unidimensionale le derivate sono valutate nei valori medi, per cui sono *numeri* che passano al di fuori dei valori di aspettazione

$$\begin{aligned} \sigma_f^2 &= E[(f(x_1, x_2) - f(\mu_1, \mu_2))^2] \approx E\left[\left(\frac{\partial f(\mu_1, \mu_2)}{\partial x_1}(x_1 - \mu_1) + \frac{\partial f(\mu_1, \mu_2)}{\partial x_2}(x_2 - \mu_2)\right)^2\right] = \\ &= \left(\frac{\partial f(\mu_1, \mu_2)}{\partial x_1}\right)^2 \sigma_1^2 + \left(\frac{\partial f(\mu_1, \mu_2)}{\partial x_2}\right)^2 \sigma_2^2 + 2 \frac{\partial f(\mu_1, \mu_2)}{\partial x_1} \frac{\partial f(\mu_1, \mu_2)}{\partial x_2} \text{Cov}(x_1, x_2). \end{aligned}$$

L'espressione che abbiamo appena ricavato si può riscrivere più convenientemente utilizzando la definizione di coefficiente di correlazione lineare come

$$\sigma_f^2 \approx \left(\frac{\partial f(\mu_1, \mu_2)}{\partial x_1}\right)^2 \sigma_1^2 + \left(\frac{\partial f(\mu_1, \mu_2)}{\partial x_2}\right)^2 \sigma_2^2 + 2 \frac{\partial f(\mu_1, \mu_2)}{\partial x_1} \frac{\partial f(\mu_1, \mu_2)}{\partial x_2} \text{Corr}(x_1, x_2) \sigma_1 \sigma_2. \quad (139)$$

Le (138) e (139) si possono generalizzare al caso di un numero arbitrario di variabili nella forma compatta

$$\mu_f \approx f(\mu_1, \dots, \mu_n) \quad \text{e} \quad \sigma_f^2 \approx \sum_{i=1}^n \sum_{j=1}^n \frac{\partial f(\mu_1, \dots, \mu_n)}{\partial x_i} \frac{\partial f(\mu_1, \dots, \mu_n)}{\partial x_j} \text{Corr}(x_i, x_j) \sigma_i \sigma_j, \quad (140)$$

che costituisce la formula generale linearizzata per il calcolo della varianza di una funzione di variabili casuali.

La cosa interessante è che se le variabili casuali sono mutuamente indipendenti, tutte le correlazioni  $\text{Corr}(x_i, x_j)$  con  $i \neq j$  si annullano e la formula per la varianza si semplifica notevolmente: ricordando che  $\text{Corr}(x_i, x_i) = 1$

$$\sigma_f^2 \approx \sum_{i=1}^n \left(\frac{\partial f(\mu_1, \dots, \mu_n)}{\partial x_i}\right)^2 \sigma_i^2. \quad (141)$$

Vale la pena sottolineare come tutte le relazioni che abbiamo ricavato in questa sezione valgono approssimativamente quando i termini di ordine superiore al primo sono trascurabili ed esattamente nel caso lineare, come illustrato nell'esempio 4.7.

► **ESEMPIO 4.7.** Possiamo utilizzare il formalismo sviluppato in questa sezione per ricavare le formule per la media e la deviazione standard della somme di due variabili casuali indipendenti  $x_1$  ed  $x_2$ —che abbiamo calcolato direttamente nella sezione 4.2.1. Sia dunque  $f(x_1, x_2) = x_1 + x_2$ . La funzione è lineare in entrambe le variabili, nel senso che le derivate parziali di ordine superiore al primo rispetto ad entrambe le variabili si annullano—ragion per cui la (140) è in questo caso esatta. Abbiamo

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = \frac{\partial f(x_1, x_2)}{\partial x_2} = 1,$$

da cui segue direttamente

$$\mu_f = f(\mu_1, \mu_2) \quad \text{e} \quad \sigma_f = \sqrt{\sigma_1^2 + \sigma_2^2},$$

che sono gli stessi risultati della (120) e (122).

## 4.7 LA PROPAGAZIONE DELL'ERRORE STATISTICO

Siamo pronti per chiudere il cerchio—riprendiamo il filo della nostra discussione e riesaminiamo, nel linguaggio della probabilità e delle funzioni di distribuzione, la propagazione dell'errore che abbiamo discusso, in un contesto diverso, nella sezione 1.5. Abbiamo visto che, nella teoria dei campioni, l'incertezza di misura ha il significato di una stima della deviazione standard della distribuzione generatrice. Visto che sappiamo calcolare, almeno approssimativamente, la deviazione standard di una funzione arbitraria di variabili casuali, date le deviazioni standard delle variabili stesse, di fatto abbiamo tutti gli ingredienti per propagare gli errori in modo statisticamente corretto.

Supponiamo dunque di avere una serie di  $n$  grandezze indipendenti misurate  $x_i = \hat{x}_i \pm \sigma_i$  ed una generica funzione  $f(x_1, \dots, x_n)$ , e partiamo dal caso più semplice—quello cioè in cui le grandezze di partenza sono tutte indipendenti. L'equivalente della formula generale di propagazione dell'errore massimo (17) è

$$\sigma_f^2 \approx \sum_{i=1}^n \left( \frac{\partial f(\hat{x}_1, \dots, \hat{x}_n)}{\partial x_i} \right)^2 \sigma_i^2 \quad \text{ovvero} \quad \sigma_f \approx \sqrt{\sum_{i=1}^n \left( \frac{\partial f(\hat{x}_1, \dots, \hat{x}_n)}{\partial x_i} \right)^2 \sigma_i^2} \quad (142)$$

La 142 è formalmente identica alla (141) ma le due scritture sono logicamente diverse, perché in questo contesto i  $\sigma_i$  rappresentano in generale le nostre miglior stime delle deviazioni standard delle distribuzioni generatrici e non i valori calcolati a partire dalla forma analitica delle funzioni di distribuzioni stesse. La differenza è sottile ma importante. Il contenuto fisico fondamentale della 142 è che nel caso di grandezze indipendenti (che è tipico in laboratorio) *gli errori statistici non si sommano linearmente come gli errori massimi, ma in quadratura.*

► **ESEMPIO 4.8.** Torniamo per un attimo all'esempio 1.15 della misura dell'accelerazione di gravità al livello del suolo con un pendolo. La propagazione dell'errore statistico si legge in questo caso come

$$\sigma_g = \frac{4\pi^2}{T^2} \sqrt{\left( \sigma_l^2 + \frac{4l^2}{T^2} \sigma_T^2 \right)} \approx 0.11 \text{ m s}^{-2} \quad \text{anziché} \quad \Delta g = \frac{4\pi^2}{T^2} \left( \Delta l + 2 \frac{l}{T} \Delta T \right) \approx 0.15 \text{ m s}^{-2},$$

cioè l'incertezza statistica su  $g$  è significativamente più piccola di quella ottenuta attraverso la propagazione dell'errore massimo.

Possiamo andare oltre. Se le grandezze non sono indipendenti ed abbiamo delle stime  $r_{ij}$  dei coefficienti di correlazione  $\text{Corr}(x_i, x_j)$ , la 142 si generalizza come

$$\sigma_f^2 \approx \sum_{i=1}^n \sum_{j=1}^n \frac{\partial f(\hat{x}_1, \dots, \hat{x}_n)}{\partial x_i} \frac{\partial f(\hat{x}_1, \dots, \hat{x}_n)}{\partial x_j} r_{ij} \sigma_i \sigma_j. \quad (143)$$

► **ESEMPIO 4.9.** Si misura la massa  $m = \hat{m} \pm \sigma_m$  il volume  $V = \hat{V} \pm \sigma_V$  di un oggetto e si vuole stimare la densità  $\rho$  dell'oggetto stesso. La propagazione dell'errore nel caso generale si scrive come

$$\begin{aligned} \sigma_\rho^2 &= \left( \frac{\partial \rho(\hat{m}, \hat{V})}{\partial m} \right)^2 \sigma_m^2 + \left( \frac{\partial \rho(\hat{m}, \hat{V})}{\partial V} \right)^2 \sigma_V^2 + 2r_{mV} \frac{\partial \rho(\hat{m}, \hat{V})}{\partial m} \frac{\partial \rho(\hat{m}, \hat{V})}{\partial V} \sigma_m \sigma_V = \\ &= \frac{\hat{m}^2 \sigma_V^2 + \hat{V}^2 \sigma_m^2 - 2r_{mV} \hat{m} \hat{V} \sigma_m \sigma_V}{\hat{V}^4} \end{aligned}$$

Se le due misure sono indipendenti  $r_{mV}$  è nullo e la formula si semplifica in

$$\sigma_\rho^2 = \frac{\hat{m}^2 \sigma_V^2 + \hat{V}^2 \sigma_m^2}{\hat{V}^4}.$$

## 4.7.1 Un esempio di propagazione degli errori per variabili correlate

Concludiamo il capitolo con un esempio di propagazione degli errori nel caso di variabili non indipendenti—e a questo scopo ripartiamo dall'esempio 4.6 illustrato in figura 4.6. La figura 4.8 è una rappresentazione alternativa del grafico di dispersione in cui che comprende, oltre al grafico di dispersione stesso, le sue proiezioni sui due assi ortogonali—ovvero gli istogrammi unidimensionali delle due variabili in gioco. Questo tipo di rappresentazione riassume tutte le informazioni su una data coppia di quantità e si trova spesso utilizzata in pratica.

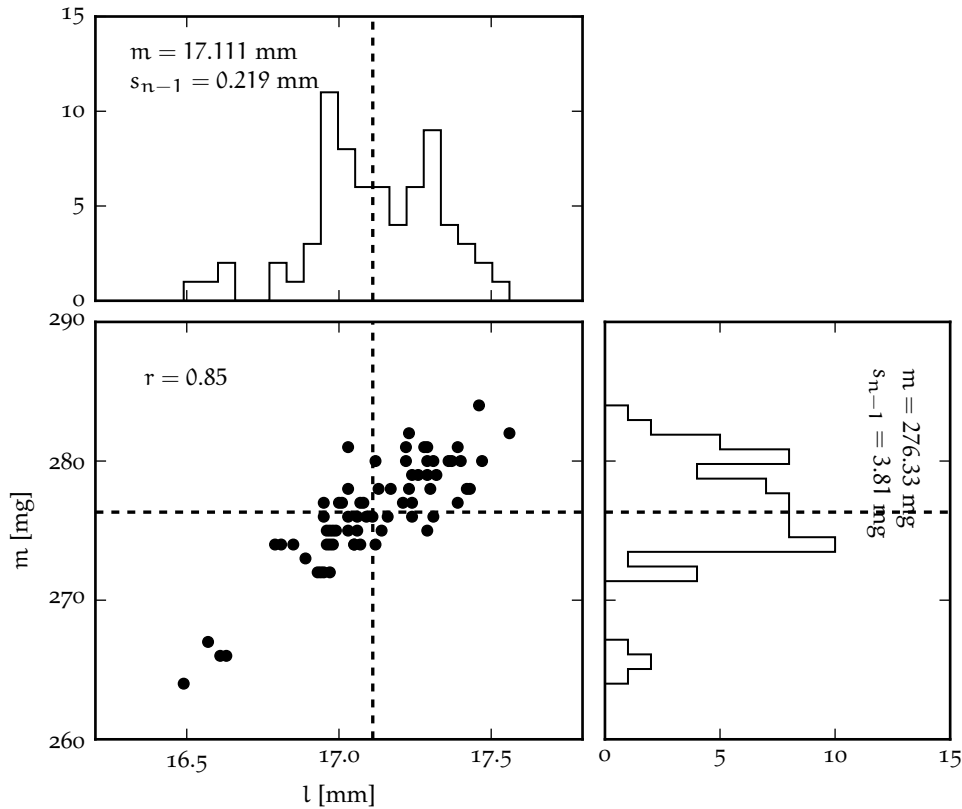


FIGURA 4.8. Grafico di dispersione dei dati mostrati in figura 4.6 con l'aggiunta dei due istogrammi unidimensionali delle due quantità in gioco. Le linee tratteggiate indicano le medie campione delle due variabili e, per completezza, i valori numerici delle medie stesse sono riportate nei due istogrammi insieme a quelli delle stime imparziali delle rispettive deviazioni standard.

Supponiamo adesso di voler stimare, a partire dai dati a nostra disposizione, la densità lineare dei nostri chiodini

$$\lambda = \frac{m}{l}.$$

(Se avessimo misurato il diametro dei chiodini, e assumendo i chiodini stessi perfettamente simmetrici, avremmo potuto anche stimare la densità del materiale propriamente detta, ma per i nostri fini si tratta di un dettaglio.) Al solito prenderemo la media campione e la stima della deviazione standard della media come valore centrale ed incertezza associata alla misura delle due grandezze di partenza e, in base alle informazioni in figura 4.8, scriveremo

$$m = \hat{m} \pm s_m = 276.33 \pm 0.45 \text{ mg} \quad \text{e} \quad l = \hat{l} \pm s_l = 17.111 \pm 0.026 \text{ mm}$$

(abbiamo diviso le stime delle deviazioni standard del campione per  $\sqrt{70}$  per passare alla deviazione

standard della media). Ora, se utilizzassimo acriticamente le formule per la propagazione degli errori per variabili indipendenti scriveremmo

$$\hat{\lambda} = \frac{\hat{m}}{\hat{l}} \quad \text{e} \quad \sigma_{\lambda}^2 = \frac{\hat{m}^2 \sigma_l^2 + \hat{l}^2 \sigma_m^2}{\hat{l}^4} \quad \text{ovverosia} \quad \lambda = 16.149 \pm 0.036 \text{ g m}^{-1}.$$

Se invece dividiamo, chiodino per chiodino, il valore della massa per il valore corrispondente della lunghezza, la deviazione standard della media per il campione risultante, che è rappresentato nell'istogramma in figura 4.9 è

$$s_{\lambda} = 0.014,$$

vale a dire circa due volte e mezzo più piccolo dell'errore propagato. Da dove viene, allora, la discrepanza?

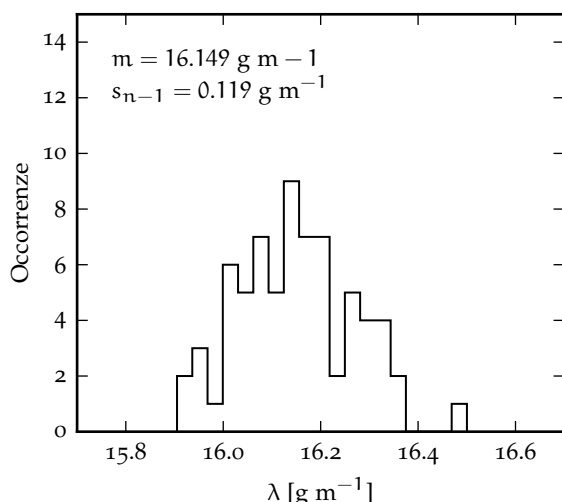


FIGURA 4.9. Istogramma dei valori di densità lineare  $\lambda_i = m_i/l_i$  ottenuti dividendo, chiodino per chiodino, i valori misurati di massa e lunghezza—per l'intero campione mostrato nel diagramma di dispersione in figura 4.8. Per completezza i valori della media campione e della stima imparziale della deviazione standard sono riportate esplicitamente.

La risposta è semplice:  $m$  e  $l$  hanno un coefficiente di correlazione diverso da zero, per cui non sono indipendenti, e la (142) non si applica. Dobbiamo invece utilizzare la formula completa (143)

$$\sigma_{\lambda}^2 = \frac{\hat{m}^2 \sigma_l^2 + \hat{l}^2 \sigma_m^2 - 2r_{ml} \hat{m} \hat{l} \sigma_m \sigma_l}{\hat{l}^4}$$

che, non per caso, fornisce un valore

$$\sigma_{\lambda} = 0.014$$

consistente con la stima  $s_{\lambda}$  fatta direttamente sul campione.

#### 4.8 IN BREVE...

► Possiamo modellizzare il processo di misura di una grandezza fisica in condizioni di ripetitività come il campionamento di una distribuzione di probabilità—che chiamiamo distribuzione generatrice. In questo schema il valore centrale e l'incertezza di misura coincidono con la nostra miglior stima della media e della deviazione standard della distribuzione generatrice. Ne segue che l'intervallo determinato dall'incertezza di misura non è garantito contenere il misurando, ma ha una probabilità data di contenerlo.

► Nella somma di variabili casuali indipendenti la media della somma è uguale alla somma delle medie e la varianza della somma è uguale alla somma delle varianze. Ne segue che le deviazioni standard si sommano in quadratura.

► Nel caso di  $n$  misure ripetute di una stessa grandezza in condizioni di dispersione non nulle la media e la varianza campione

$$\bar{m} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{e} \quad s_{n-1}^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{m})^2$$

sono le nostre migliori stime della media e della varianza della distribuzione generatrice.

- La deviazione standard della media di  $n$  campionamenti di una distribuzione generatrice data è  $\sqrt{n}$  volte più piccola della deviazione standard della distribuzione stessa. Questa è la legge fondamentale con cui scalano gli errori statistici in funzione della dimensione del campione.
- Gli errori statistici su grandezze indipendenti si sommano in quadratura—e non linearmente come gli errori massimi

$$s_f = \sqrt{\sum_{i=1}^n \left( \frac{\partial f(\hat{x}_1, \dots, \hat{x}_n)}{\partial x_i} \right)^2 s_i^2}.$$

Questo deriva essenzialmente dal fatto che se le grandezze sono indipendenti le fluttuazioni statistiche tendono in media a compensarsi ed è relativamente poco probabile che esse tendano a contribuire nello stesso verso per generare configurazioni estreme. Ne segue che l'errore massimo tende sistematicamente ad essere una sovrastima dell'incertezza di misura.



## DISTRIBUZIONI UNIVARIATE DI USO COMUNE

## 5.1 LA DISTRIBUZIONE BINOMIALE

Consideriamo un esperimento che abbia esattamente due esiti distinti possibili,  $E_1$  ed  $E_2$ , con probabilità  $P(E_1) = p$  e  $P(E_2) = 1 - p$ . Se ripetiamo  $n$  volte l'esperimento, assumendo che le realizzazioni siano indipendenti, qual è la probabilità di ottenere esattamente  $k$  volte (con  $0 \leq k \leq n$ ) l'esito  $E_1$ ?

Cominciamo dal calcolare la probabilità di ottenere  $k$  volte  $E_1$  (e, di conseguenza,  $n - k$  volte  $E_2$ ) *in un ordine particolare*—ad esempio  $E_1$  nelle prime  $k$  realizzazioni ed  $E_2$  nelle successive. Non è difficile convincersi, sfruttando la legge di moltiplicazione delle probabilità per eventi indipendenti (59), che

$$P\left(\underbrace{E_1 E_1 E_1 E_1 E_1 E_1 \dots}_{k \text{ volte}} \underbrace{E_2 E_2 E_2 E_2 E_2 E_2 \dots}_{n-k \text{ volte}}\right) = p^k (1-p)^{n-k}.$$

Ma questa è solo una delle possibili configurazioni in cui si può ottenere  $k$  volte l'esito  $E_1$ ; ce ne sono altre (ad esempio quella in cui otteniamo  $E_2$  nelle prime  $n - k$  realizzazioni ed  $E_1$  nelle successive e molte altre *miste*), tutte con la stessa probabilità  $p^k (1-p)^{n-k}$ . Il numero totale di combinazioni possibili coincide con il numero di sottoinsiemi di  $k$  elementi di un insieme di  $n$  elementi, ossia con il coefficiente binomiale  $n$  su  $k$ . Possiamo allora sfruttare il fatto che le probabilità per eventi disgiunti si sommano (e qui stiamo sommando più volte lo stesso numero) e scrivere la probabilità cercata come

$$\mathcal{B}(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k} \quad (144)$$

La (144) prende il nome di distribuzione binomiale, ed è mostrata per quattro combinazioni distinte di  $n$  e  $p$  in figura 5.1.

► **ESEMPIO 5.1.** Lanciamo una moneta per 4 volte. Le sequenze possibili sono complessivamente  $2^4 = 16$  e sono qui di seguito elencate: TTTT, CTTT, TCTT, CCTT, TTCT, CTCT, TCCT, CCCT, TTTC, CTTC, TCTC, CCTC, TTCC, CTCC, TCCC, CCCC. Se la moneta non è truccata ognuna di queste combinazioni ha la stessa probabilità  $P = 1/16$  di uscire. Le combinazioni in cui T (testa) figura esattamente due volte sono 6—per l'esattezza le combinazioni 4, 6, 7, 10, 11 e 13, numerate nell'ordine in cui esse sono riportate sopra. Notiamo esplicitamente che 6 è proprio il coefficiente binomiale 4 su 2

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{24}{2 \times 2} = 6.$$

La probabilità di avere esattamente due teste in quattro lanci è dunque  $6/16 = 3/8$  che, come è facile verificare, coincide con  $\mathcal{B}(2; 4, 1/2)$  definita in (144).

Visto che questo è il primo caso di funzione di distribuzione dipendente da parametri che esaminiamo, una questione veloce di notazione: nella (144)  $k$  è la variabile aleatoria, mentre  $n$  e  $p$  sono due parametri esterni completamente specificati dal problema che stiamo studiando. Nel seguito scriveremo sempre gli argomenti delle funzioni di distribuzione in quest'ordine (prima la variabile casuale e poi i parametri) ed avremo cura di separare la variabile da quest'ultimi utilizzando un punto e virgola.

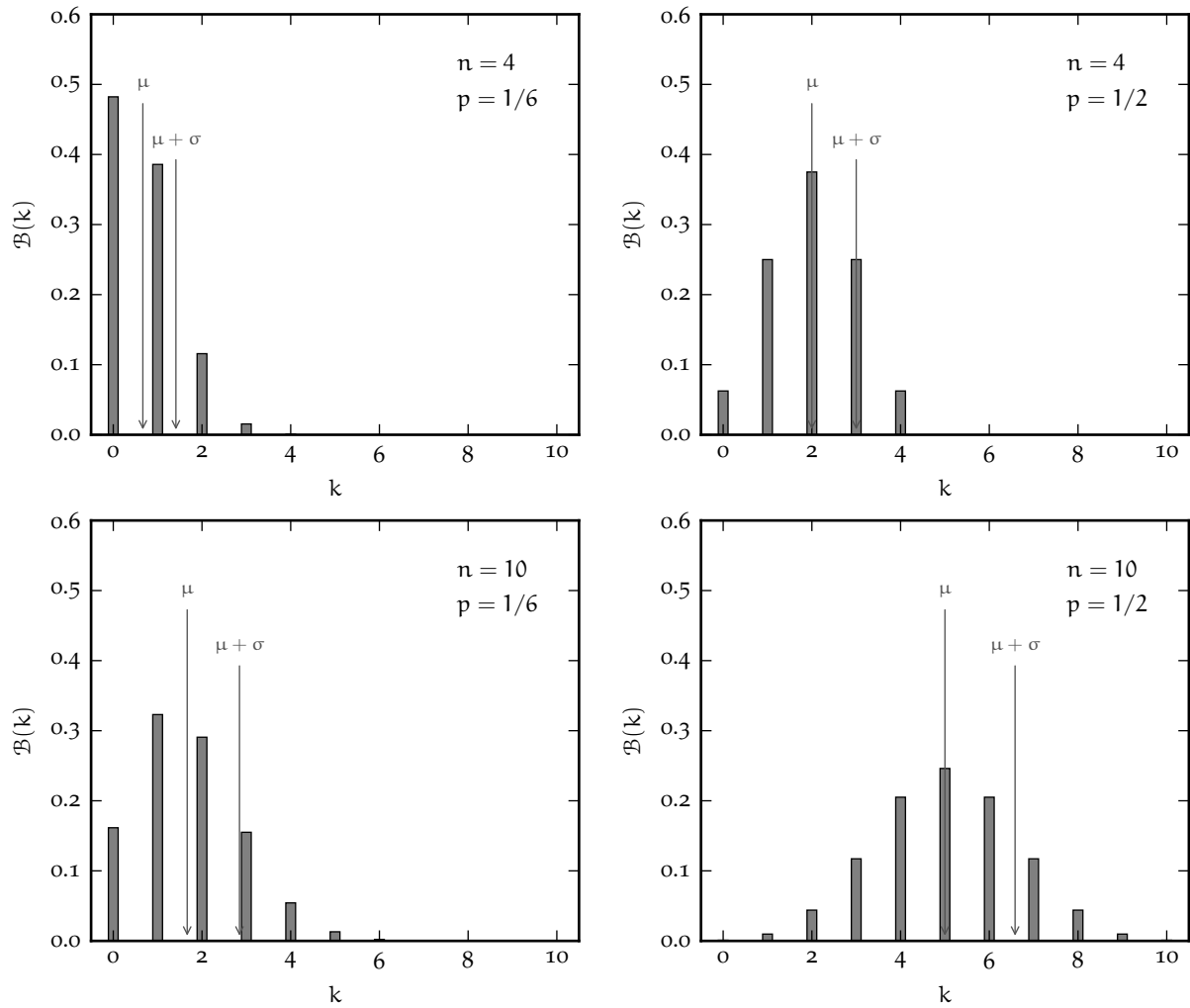


FIGURA 5.1. Esempi di distribuzione binomiale (144) per quattro diversi valori di  $n$  e  $k$ . Come discuteremo estensivamente nel seguito, all'aumentare della media  $\mu = np$ , la distribuzione tende progressivamente ad assumere una caratteristica forma a campana.

### 5.1.1 Normalizzazione, media e varianza

Sfruttando la formula per la potenza di un binomio si può verificare facilmente che, così come è scritta nella (144), la distribuzione binomiale è correttamente normalizzata:

$$\sum_{k=0}^n \mathcal{B}(k; n, p) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1.$$

La media si calcola secondo la definizione (72):

$$\mu = E[k] = \sum_{k=0}^n k \mathcal{B}(k; n, p) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

Il termine con  $k = 0$  non contribuisce alla somma a causa del  $k$  a moltiplicare, per cui possiamo riscrivere l'espressione precedente facendo partire la somma da 1:

$$\mu = \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=1}^n k \frac{np(n-1)!}{k(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k}.$$



A questo punto possiamo portare fuori dal segno di sommatoria il termine  $np$  a moltiplicare, dato che non dipende dall'indice  $k$  su cui si somma, e semplificare il  $k$  che è sia al numeratore che al denominatore:

$$\begin{aligned}\mu &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} = np \sum_{h=0}^m \frac{m!}{h!(m-h)!} p^h (1-p)^{m-h} = \\ &= np \sum_{h=0}^m \binom{m}{h} p^h (1-p)^{m-h}\end{aligned}$$

(abbiamo posto  $h = k - 1$  e  $m = n - 1$ ). Nell'ultimo passaggio riconosciamo la condizione di normalizzazione, per cui

$$\mu = np. \quad (145)$$

(La media è uguale al prodotto del numero di realizzazioni per la probabilità di successo in una realizzazione singola. Quello che uno si aspetterebbe intuitivamente, no?)

Ok, procediamo oltre. Per stimare la varianza utilizziamo la (78) ed il valore di aspettazione di  $k^2$  si calcola come:

$$E[k^2] = \sum_{k=0}^n k^2 \mathcal{B}(k; n, p) = \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}.$$

Esattamente come prima possiamo far partire la somma da 1, dato che il primo termine è nullo:

$$E[k^2] = \sum_{k=1}^n k^2 \frac{np(n-1)!}{k(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} = np \sum_{k=1}^n k \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k}$$

e ponendo, come prima,  $h = k - 1$  e  $m = n - 1$  si ottiene:

$$E[k^2] = np \sum_{h=0}^m (h+1) \mathcal{B}(h; m, p) = np \sum_{h=0}^m [h \mathcal{B}(h; m, p) + \mathcal{B}(h; m, p)] = np(mp + 1) = np(np - p + 1),$$

da cui è banale calcolare, utilizzando la (78)

$$\sigma^2 = np(1-p). \quad (146)$$

► **ESEMPIO 5.2.** Supponiamo di tirare un dado a sei facce per 100 volte e di osservare che la faccia numero 1 esce 3 volte. Possiamo concludere che il dado non è equo? Il numero  $k$  di uscite della faccia 1 in 100 lanci è una variabile aleatoria che segue una distribuzione binomiale con  $n = 100$  e  $p = 1/6$ . La media della distribuzione è  $np \approx 16.67$  e la deviazione standard è  $\sqrt{np(1-p)} \approx 3.73$ , per cui il nostro 3 dista  $\approx 3.7$  deviazioni standard dalla media. Il modo corretto di inquadrare il problema è calcolare la probabilità che  $k \leq 3$  nell'ipotesi che il dado sia equo. La disuguaglianza di Chebyshev fornisce un limite superiore (molto generoso) a  $1/3.7^2 \approx 7.3\%$ . Il calcolo diretto fornisce

$$P(k \leq 3) = P(k=0) + P(k=1) + P(k=2) + P(k=3) \approx 1.8 \times 10^{-5}.$$

Quindi è estremamente poco probabile che con 100 lanci di un dado equo la faccia 1 esca solo 3 volte (o meno).

## 5.1.2 Momenti di ordine superiore



Per calcolare i momenti di ordine superiore al secondo potremmo, in linea di principio, procedere come sopra, ma non ci vuole molto per accorgersi che il calcolo diviene tedioso molto velocemente. Fortunatamente non è difficile ricavare una formula per ricorrenza che rende il calcolo più agevole.

Notiamo infatti che

$$\begin{aligned} \frac{d}{dp} E[k^m] &= \frac{d}{dp} \sum_{k=0}^n k^m \mathcal{B}(k; n, p) = \sum_{k=0}^n k^m \binom{n}{k} \frac{d}{dp} [p^k (1-p)^{n-k}] = \\ &= \sum_{k=0}^n k^m \binom{n}{k} [kp^{k-1}(1-p)^{n-k} - (n-k)p^k(1-p)^{n-k-1}] = \\ &= \frac{1}{p} \sum_{k=0}^n k^{m+1} \binom{n}{k} p^k (1-p)^{n-k} - \frac{n}{1-p} \sum_{k=0}^n k^m \binom{n}{k} p^k (1-p)^{n-k} + \\ &+ \frac{1}{1-p} \sum_{k=0}^n k^{m+1} \binom{n}{k} p^k (1-p)^{n-k} = \left[ \frac{1}{p} + \frac{1}{1-p} \right] E[k^{m+1}] - \frac{n}{1-p} E[k^m], \end{aligned}$$

da cui

$$E[k^{m+1}] = p(1-p) \frac{d}{dp} E[k^m] + np E[k^m]. \quad (147)$$

Utilizzando la (147) e partendo dal momento di ordine 1 possiamo allora calcolare tutti i momenti di ordine superiore—ad esempio:

$$E[k^2] = p(1-p) \frac{d}{dp} E[k] + np E[k] = p(1-p) \frac{d(np)}{dp} + (np)^2 = np(1-p) + (np)^2 = np(np-p+1),$$

che sapevamo già, e

$$\begin{aligned} E[k^3] &= p(1-p) \frac{d}{dp} E[k^2] + np E[k^2] = p(1-p)(2n^2p - 2np + n) + n^2p^2(np-p+1) = \\ &= 3n^2p^2 - 3np^2 + np - 3n^2p^3 + 2np^3 + n^3p^3 = np(1-p)(1-2p) + 3n^2p^2(1-p) + n^3p^3. \end{aligned}$$

Il momento centrale di ordine 3 ed il coefficiente di asimmetria della distribuzione binomiale sono dunque rispettivamente

$$\begin{aligned} \mu_3 &= E[k^3] - 3\mu\sigma^2 - \mu^3 = np(1-p)(1-2p) + 3n^2p^2(1-p) + n^3p^3 - 3n^2p^2(1-p) - n^3p^3 = \\ &= np(1-p)(1-2p) \end{aligned}$$

e

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{(1-2p)}{\sqrt{np(1-p)}}. \quad (148)$$

## 5.2 DIGRESSIONE: DISTRIBUZIONE BINOMIALE E MISURE DI EFFICIENZA

Da un punto di vista pratico, la distribuzione binomiale è rilevante in Fisica perché regola la statistica delle misure di *efficienza*, che è un concetto fondamentale in molti contesti diversi. Facciamo un esempio per fissare le idee: un rivelatore di particelle è un dispositivo progettato per produrre un impulso elettrico quando è attraversato, appunto, da una particella (ad esempio un protone o un elettrone). In pratica non esiste un rivelatore *perfetto*, nel senso che, per quanto raro, si dà sempre il caso che una particella possa attraversare un rivelatore senza produrre un segnale. In questo caso l'efficienza  $\varepsilon$  è per definizione la probabilità che, dato il passaggio della particella, il segnale sia effettivamente emesso.

L'efficienza di un rivelatore si misura tipicamente facendo passare dal rivelatore un numero fissato (e noto)  $N$  di particelle e contando il numero di volte  $n \leq N$  in cui è effettivamente prodotto un

segnale elettrico in uscita. È chiaro che  $n$  è una variabile distribuita come una binomiale  $\mathcal{B}(n; N, \varepsilon)$  in cui l'efficienza ha il significato della probabilità elementare di successo  $p$ . La cosa interessante è che, tipicamente,  $\varepsilon$  è l'incognita che noi vogliamo *stimare* da un singolo campionamento della variabile  $n$  (dato  $N$ ). Si tratta del tipico problema di probabilità inversa come quelli che abbiamo discusso nella sezione 3.5.

### 5.2.1 La soluzione classica

Il problema si può attaccare nella forma più semplice, armati solo della nostra intuizione, dicendo che la nostra miglior stima  $\hat{\varepsilon}$  di  $\varepsilon$  sarà

$$\hat{\varepsilon} = \frac{n}{N}, \quad (149)$$

che ha la buona proprietà di valere  $\varepsilon$  in media—nel senso del valore di aspettazione—come si dimostra banalmente:

$$E[\hat{\varepsilon}] = E\left[\frac{n}{N}\right] = \frac{1}{N} E[n] = \frac{1}{N} N\varepsilon = \varepsilon.$$

Ma quale incertezza associamo a  $\hat{\varepsilon}$ ? Ciò che sappiamo dal problema di probabilità diretto è che la varianza di  $n$ , fissati  $\varepsilon$  ed  $N$ , è data da  $\text{Var}(n) = N\varepsilon(1 - \varepsilon)$ , che (non conoscendo  $\varepsilon$ ) possiamo stimare utilizzando  $\hat{\varepsilon} = n/N$ . Scriveremo allora

$$\text{Var}(\hat{\varepsilon}) = \text{Var}\left(\frac{n}{N}\right) = \frac{1}{N^2} \text{Var}(n) = \frac{\varepsilon(1 - \varepsilon)}{N} \approx \frac{\hat{\varepsilon}(1 - \hat{\varepsilon})}{N} = \frac{n(N - n)}{N^3},$$

da cui l'errore  $\sigma_{\hat{\varepsilon}}$  su  $\hat{\varepsilon}$

$$\sigma_{\hat{\varepsilon}} = \sqrt{\frac{n(N - n)}{N^3}}. \quad (150)$$

► **ESEMPIO 5.3.** Per misurare l'efficienza di un rivelatore si fanno passare  $N = 1000$  particelle dalla sua area attiva e si registra un segnale in  $n = 962$  casi. Utilizzando la (149) e la (150) scriveremo  $\varepsilon = 0.962 \pm 0.006$ .

Le (149) e (150) si trovano su tutti i libri di testo e sono ampiamente usate in pratica. Notiamo che la formula (150) per la stima dell'errore ha un problema patologico per  $n = 0$  e  $n = N$ —ove è identicamente nulla. Questo generalmente non è un problema, ma nella prossima sezione vedremo una formulazione più corretta che funziona anche nei casi limite.

### 5.2.2 Una soluzione basata sul teorema di Bayes



Il problema delle misure di efficienza è anche un'ottima palestra per approfondire il ruolo e l'utilizzo del teorema di Bayes nei problemi di probabilità inversa, che abbiamo cominciato a discutere nella sezione 3.5.2. Cominciamo da riscrivere la 63 nella forma appropriata per il nostro problema:

$$P(\varepsilon | n) = \frac{P(n | \varepsilon) P(\varepsilon)}{P(n)} \propto P(n | \varepsilon) P(\varepsilon). \quad (151)$$

(Breve riepilogo: la probabilità a posteriori è proporzionale al prodotto della verosimiglianza per la probabilità a priori, ed il denominatore  $P(n)$ , che non dipende da  $\varepsilon$  è una semplice costante di normalizzazione che, come vedremo, possiamo calcolare facilmente).  $P(n | \varepsilon)$ , come sappiamo, non è che  $\mathcal{B}(n; N, \varepsilon)$ , e l'unica cosa che rimane da fissare è la probabilità a priori  $P(\varepsilon)$ . Come abbiamo avuto occasione di dire, la scelta della *prior* è un punto delicato, ed argomento di discussione serrata tra le diverse scuole di statistica. Nel contesto di questo esercizio, assumeremo semplicemente che, non avendo nessuna informazione sull'efficienza  $\varepsilon$  che vogliamo misurare, possiamo assumere una probabilità a priori costante, per cui

$$P(\varepsilon | n) \propto \binom{N}{n} \varepsilon^n (1 - \varepsilon)^{N-n} \propto \varepsilon^n (1 - \varepsilon)^{N-n}$$

(L'ultimo passaggio è giustificato dal fatto che il coefficiente binomiale non dipende da  $\varepsilon$ .) Ok, fermiamoci un attimo a riprendere fiato. L'equazione che abbiamo scritto è notevole perché di fatto abbiamo trasformato una funzione di distribuzione propriamente normalizzata nella variabile casuale (discreta)  $n$  in una densità di probabilità (in generale non normalizzata) per la variabile casuale (continua)  $\varepsilon$ . La costante di normalizzazione  $c$  si calcola banalmente integrando per parti la verosimiglianza

$$\begin{aligned}\int_0^1 \varepsilon^n (1-\varepsilon)^{N-n} d\varepsilon &= \frac{1}{n+1} \left[ \varepsilon^{n+1} (1-\varepsilon)^{N-n} \Big|_0^1 + (N-n) \int_0^1 \varepsilon^{n+1} (1-\varepsilon)^{N-n-1} d\varepsilon \right] = \\ &= \frac{N-n}{n+1} \int_0^1 \varepsilon^{n+1} (1-\varepsilon)^{N-n-1} d\varepsilon\end{aligned}$$

e, iterando  $(N-n)$  volte

$$\int_0^1 \varepsilon^n (1-\varepsilon)^{N-n} d\varepsilon = \frac{(N-n)! n!}{N!} \int_0^1 \varepsilon^N d\varepsilon = \frac{(N-n)! n!}{(N+1)!}.$$

Si ha dunque in conclusione

$$P(\varepsilon | n) = \frac{(N+1)!}{(N-n)! n!} \varepsilon^n (1-\varepsilon)^{N-n}. \quad (152)$$

La (152) è molto di più di quello che abbiamo calcolato nella sezione precedente—essa è la densità di probabilità normalizzata per la nostra efficienza  $\varepsilon$ , data una misura  $n$  ed assumendo una *prior* non informativa. Con questo in mano possiamo calcolare un certo numero di cose, tra cui la media e la deviazione standard, che possiamo considerare come miglior stima di  $\varepsilon$  ed incertezza associata

$$\hat{\varepsilon} = E[\varepsilon] = \frac{n+1}{N+2} \quad (153)$$

$$\sigma_{\hat{\varepsilon}}^2 = \text{Var}(\varepsilon) = \frac{(n+1)(N-n+1)}{(N+3)(N+2)^2}, \quad (154)$$

che curano i casi patologici e si riducono alle espressioni ricavate in precedenza per  $n \gg 1$  e  $N \gg 1$ .

### 5.3 LA DISTRIBUZIONE MULTINOMIALE



La distribuzione multinomiale è una generalizzazione della distribuzione binomiale in cui lo schema è ancora quello delle  $n$  ripetizioni indipendenti di uno stesso esperimento, ma stavolta l'esperimento in questione può avere  $m$  (anziché 2) esiti distinti e mutuamente esclusivi  $E_1, \dots, E_m$  con probabilità associate  $p_1, \dots, p_m$ , soggette ai vincoli usuali

$$p_i \geq 0 \quad \text{e} \quad \sum_{i=1}^m p_k = 1. \quad (155)$$

In questa sezione non ricaveremo in modo deduttivo le proprietà della distribuzione multinomiale—ci limiteremo ad elencarle, sottolineando le analogie con la distribuzione binomiale. (Per molti versi il passaggio tra le due ha una corrispondenza uno ad uno con la generalizzazione dei coefficienti binomiali ai coefficienti multinomiali che abbiamo visto nella sezione 3.3.4.

► **ESEMPIO 5.4.** La ripetizione di  $n$  lanci di un dado a equo a sei facce è un esempio tipico in cui i numeri di volte  $k_1, \dots, k_6$  in cui escono le facce  $1, \dots, 6$  sono distribuite secondo una multinomiale con  $p_i = 1/6$ .

Cominciamo. La probabilità che in  $n$  ripetizioni del nostro esperimento si verifichi  $k_1$  volte l'evento  $E_1$  e  $k_2$  volte l'evento  $E_2 \dots e k_m$  volte l'evento  $E_m$  è data da

$$\mathcal{M}(k_1, \dots, k_m; n, p_1, \dots, p_m) = \binom{n}{k_1, \dots, k_m} p_1^{k_1} p_2^{k_2} \dots p_m^{k_m} = \frac{n!}{k_1! k_2! \dots k_m!} p_1^{k_1} p_2^{k_2} \dots p_m^{k_m}. \quad (156)$$

Notiamo esplicitamente che la (156) è la funzione di distribuzione congiunta delle  $m$  variabili casuali  $k_1, \dots, k_m$  ed è dipendente dagli  $m$  parametri  $p_1, \dots, p_m$ . Notiamo anche che la (156) vale solamente sotto l'assunzione implicita che  $\sum_{i=1}^m k_i = n$ , nel senso che sequenze di  $k_i$  la cui somma non corrisponda al numero totale di ripetizioni dell'esperimento non sono ovviamente ammesse.

► **ESEMPIO 5.5.** La probabilità che in 6 lanci di un dado equo a sei facce esca esattamente una volta ciascuna singola faccia (cioè esattamente un uno, esattamente un due e così via fino al sei) è data dal valore della distribuzione multinomiale per  $n = 6$ ,  $k_1 = \dots = k_6 = 1$  e  $p_1 = \dots = p_6 = 1/6$ , ovvero

$$P = \frac{6!}{6^6} \approx 0.0154, \quad (157)$$

ovverosia poco più dell'1%.

Non dimostreremo esplicitamente che la (156), così come è scritta, è correttamente normalizzata, e non calcoleremo esplicitamente la media e la varianza dei  $k_i$ —che, in perfetta analogia con i risultati che abbiamo ricavato per la distribuzione binomiale, sono

$$E[k_i] = np_i \quad \text{e} \quad \text{Var}(k_i) = np_i(1 - p_i). \quad (158)$$

Una cosa che possiamo chiederci a questo punto è se i  $k_i$ , visti come variabili indipendenti, siano tra loro indipendenti. Non sorprendentemente, la risposta è no, perché, come abbiamo detto prima, essi sono legati dalla relazione

$$\sum_{i=1}^m k_i = n.$$

Fisicamente questo corrisponde a dire che se lanciamo 100 volte un dado e si osserva, ad esempio, un numero di uscite della faccia 2 superiore alla media, questo dovrà essere bilanciato da un numero di uscite delle altre facce (alcune o tutte) *inferiore alla media*. (Per inciso, questo semplice argomento ci dice che la correlazione tra facce diverse deve essere negativa.) In effetti si può dimostrare che

$$\text{Cov}(k_i, k_j) = -np_i p_j \quad \text{da cui} \quad \text{Corr}(k_i, k_j) = -\sqrt{\frac{p_i p_j}{(1 - p_i)(1 - p_j)}}. \quad (159)$$

## 5.4 LA DISTRIBUZIONE DI POISSON

Consideriamo un processo *stazionario* nel dominio del tempo che produce il verificarsi di eventi indipendenti l'uno dall'altro—l'evento elementare potrebbe essere, ad esempio, il passaggio di una macchina da una strada trafficata, o l'arrivo di un cliente in un ufficio postale, o l'incontro di un conoscente in un luogo affollato. Le due domande con cui apriamo questa sezione sono: (i) cosa possiamo dire sulla probabilità che, in un dato intervallo di tempo, si verifichino  $k$  eventi quando in media se ne verificano  $\mu$ ? e (ii) tutto questo ha qualcosa a che vedere con la distribuzione binomiale che abbiamo appena analizzato in dettaglio? (Vedremo che la risposta alla seconda domanda è sì, ma dobbiamo precisare alcuni concetti prima di andare avanti.)

### 5.4.1 Processi Poissoniani

Procediamo con ordine e cominciamo con il chiarire la terminologia. Quando parliamo di *processo Poissoniano* (e, per fissare le idee, di processo Poissoniano nel dominio del tempo) intendiamo essenzialmente tre cose:

► (indipendenza) gli eventi elementari sono indipendenti, ovverosia il verificarsi di un evento ad un determinato istante di tempo non influenza la probabilità che un altro evento si verifichi (o non si verifichi) ad un istante successivo;

- (stazionarietà) la frequenza temporale media degli eventi è indipendente dal tempo, cioè il numero medio di eventi per unità di tempo  $\lambda$  è lo stesso in qualsiasi intervallo;
- (non simultaneità) non si possono verificare due o più eventi elementari nello stesso istante.

Fermiamoci per un secondo e cerchiamo di confrontare queste assunzioni (e specialmente le prime due) con gli esempi che abbiamo fatto all'inizio della sezione. Nel caso degli arrivi in ufficio postale possiamo veramente dire che l'ingresso di un cliente è indipendente da quelli dei clienti precedenti? Probabilmente no, perché se la fila nell'ufficio è troppo lunga potremmo essere scoraggiati e rinunciare ad entrare. E possiamo dire che la frequenza media degli ingressi è indipendente dal tempo? Sicuramente no, perché è ovvio che ci saranno momenti nella giornata (e nella settimana o nel mese) in cui l'afflusso medio sarà relativamente più alto (senza contare che l'ufficio, presumibilmente, sarà chiuso la notte). Se ci pensiamo bene, questo genere di dubbi varrà anche per gli altri due esempi che abbiamo fatto—e allora perché li abbiamo fatti? La risposta è semplice: mentre è molto difficile trovare una situazione reale in cui le nostre assunzioni siano verificate esattamente, vi sono molte situazioni in cui il modello Poissoniano fornisce una descrizione adeguata in intervalli di tempo abbastanza piccoli. Così la statistica dei conteggi per un rivelatore di raggi cosmici sarà con ottima approssimazione Poissoniana su tempi scala di minuti o di ore, anche se il flusso di raggi cosmici sulla superficie terrestre subisce variazioni su tempi scala più lunghi (e.g., per le variazioni della pressione atmosferica e per le variazioni delle caratteristiche dell'Eliosfera durante il ciclo Solare) per cui in effetti il processo non è strettamente Poissoniano. Vedremo altri esempi più in dettaglio nel seguito.

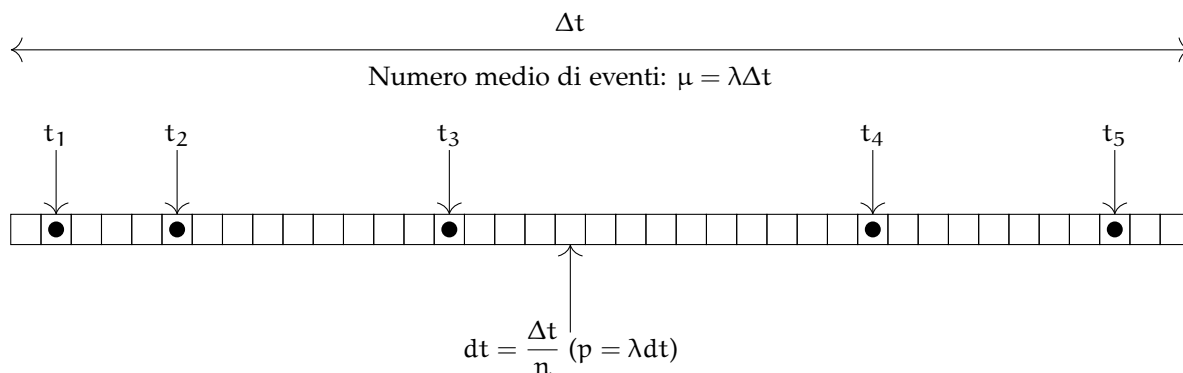


FIGURA 5.2. Rappresentazione schematica di un processo Poissoniano nel dominio del tempo con frequenza caratteristica  $\lambda$  (il numero medio di eventi in un intervallo di tempo  $\Delta t$  è dato da  $\mu = \lambda \Delta t$ ). Se dividiamo il nostro intervallo finito di lunghezza  $\Delta t$  in  $n$  intervallini di lunghezza  $dt = \Delta t/n$ , nel limite in cui  $n \rightarrow \infty$  (per cui possiamo assumere l'ipotesi di non simultaneità) la probabilità di registrare un evento in uno qualsiasi degli intervallini è  $p = \lambda dt$ .

Consideriamo dunque un intervallo di tempo  $\Delta t$ , che dividiamo in  $n$  intervalli più piccoli di lunghezza  $dt = \Delta t/n$  come mostrato in figura 5.2—with l'idea che alla fine faremo tendere  $n \rightarrow \infty$ . Il numero medio  $\lambda$  di eventi per unità di tempo (che si misura in  $s^{-1}$ ) è il parametro che determina il numero medio  $\mu = \lambda \Delta t$  di eventi nell'intero intervallo ed il numero medio di eventi  $\lambda dt = \mu/n$  in uno qualsiasi degli  $n$  sottointervalli—che per l'ipotesi di stazionarietà sono tutti equivalenti da questo punto di vista. Questo ci porta ad un punto interessante: se  $n$  è abbastanza grande (o, il che è lo stesso,  $dt$  è abbastanza piccolo) l'ipotesi di non simultaneità ci dice che in ogni sottointervallino si possono verificare esattamente 0 o 1 evento/i—ma non più di 1. Allora, se chiamiamo  $p$  la probabilità che in un generico sottointervallo si verifichi un evento, la probabilità che se ne verifichino 0 sarà  $(1 - p)$  ed il numero medio di eventi, che sappiamo essere  $\lambda dt$  si scrive come

$$\lambda dt = \frac{\mu}{n} = 0 \times (1 - p) + 1 \times p = p.$$

Cioè  $\lambda dt$  è non solo il numero medio di eventi per sottointervallo, ma anche la probabilità che si verifichi esattamente un evento in uno qualsiasi dei sottointervalli.

Questo ci offre la possibilità di calcolare immediatamente un certo numero di cose interessanti. Ad esempio: qual è la probabilità  $P(0)$  di osservare 0 eventi nell'intero intervallo  $t$  quando, come abbiamo visto, in media ne osserviamo  $\mu$ ? Beh, visto che per ipotesi gli eventi sono indipendenti, le probabilità corrispondenti si moltiplicano e nel limite  $n \rightarrow \infty$  la risposta è

$$P(0; \mu) = \lim_{n \rightarrow \infty} (1-p)^n = \lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right)^n = e^{-\mu}.$$

E se invece ci chiedessimo qual è la probabilità di osservare esattamente 1 evento nell'intero intervallo  $t$ ? Il ragionamento procede essenzialmente come prima, salvo il fatto che nel prodotto dobbiamo sostituire con  $p$  uno degli  $n$  termini  $(1-p)$  (stiamo richiedendo che non si verifichino eventi in  $n-1$  intervallini e che si verifichi un evento nell'intervallino rimanente) e dobbiamo moltiplicare per  $n$ , per tenere conto del fatto che siamo liberi di scegliere a caso tra gli  $n$  disponibili

$$P(1; \mu) = \lim_{n \rightarrow \infty} np(1-p)^{n-1} = \lim_{n \rightarrow \infty} \frac{n\mu}{n} \left(1 - \frac{\mu}{n}\right)^{n-1} = \mu e^{-\mu}.$$

Fermiamoci un attimo. Il lettore più accorto avrà capito che stiamo di fatto riderivando la distribuzione binomiale nel limite in cui la probabilità del singolo evento  $p = \lambda dt = \mu/n$  è piccola ed il numero di ripetizioni  $n$  è grande. Ma allora abbiamo tutti gli strumenti per calcolare la forma esplicita della funzione di distribuzione—che chiameremo distribuzione di Poisson. Lo faremo esplicitamente nella prossima sezione.

#### 5.4.2 La distribuzione di Poisson come limite della binomiale

Abbiamo visto che formalmente la distribuzione di Poisson si ottiene come limite della distribuzione binomiale quando  $n \rightarrow \infty$  e  $p \rightarrow 0$  in modo che la media  $\mu = np$  si mantenga costante. (Notiamo, per inciso, che se  $p \rightarrow 0$  allora  $(1-p) \rightarrow 1$  e la varianza della distribuzione limite tenderà a  $\sigma = np(1-p) \rightarrow np = \mu$ . Lo verificheremo direttamente.) Per il momento scriviamo la seconda condizione come  $p = \mu/n$ , per cui si ha

$$\lim_{n \rightarrow \infty} \mathcal{B}(k; n, p) = \lim_{n \rightarrow \infty} \frac{n!}{k!(n-k)!} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k} = \lim_{n \rightarrow \infty} \frac{n(n-1) \cdots (n-k+1)}{n^k} \frac{\mu^k}{k!} \left(1 - \frac{\mu}{n}\right)^{n-k}.$$

Fermiamoci per un attimo ad osservare questa espressione, perché ne troveremo di simili nel seguito. Non si tratta di un limite nel senso usuale del termine poiché  $k$  non è un numero, ma una variabile casuale—che, per di più, può assumere tutti i valori da 0 ad  $n$ . Non è ovvio, allora, che cosa voglia dire far tendere  $n \rightarrow \infty$  se non abbiamo una qualche sorta di prescrizione per  $k$ . Introduciamo la variabile (casuale) *ridotta*

$$\xi = \frac{k - np}{n}.$$

Il numeratore misura, in un qualche senso, le fluttuazioni della variabile casuale  $k$  attorno al suo valor medio, per cui sarà dell'ordine della deviazione standard della distribuzione  $\sigma = \sqrt{np(1-p)}$ . (Più precisamente, sarà poco probabile, e.g., per il teorema di Chebyshev, che  $\xi$  si discosti da 0 di una quantità molto più grande di  $\sigma$ .) Se questo è vero, allora  $\xi \rightarrow 0$  come  $1/\sqrt{n}$  per  $n \rightarrow \infty$ . Per la definizione di  $\xi$  si ha  $k = n(p + \xi)$ , il che ci permette di riscrivere il nostro limite come

$$\lim_{n \rightarrow \infty} \mathcal{B}(k; n, p) = \lim_{n \rightarrow \infty} \frac{n^k (1 - 1/n) \cdots (1 - p - \xi + 1/n)}{n^k} \frac{\mu^k}{k!} \left(1 - \frac{\mu}{n}\right)^{n(1-p-\xi)} = \lim_{n \rightarrow \infty} \frac{\mu^k}{k!} \left(1 - \frac{\mu}{n}\right)^n$$

(nel penultimo passaggio abbiamo sfruttato il fatto che  $p \rightarrow 0$  e  $\xi \rightarrow 0$ ). Mettendo tutto insieme otteniamo l'espressione chiusa per la distribuzione di Poisson, che è illustrata in figura 5.3 per diversi valori di  $\mu$ .

$$\mathcal{P}(k; \mu) = \frac{\mu^k}{k!} e^{-\mu}. \quad (160)$$

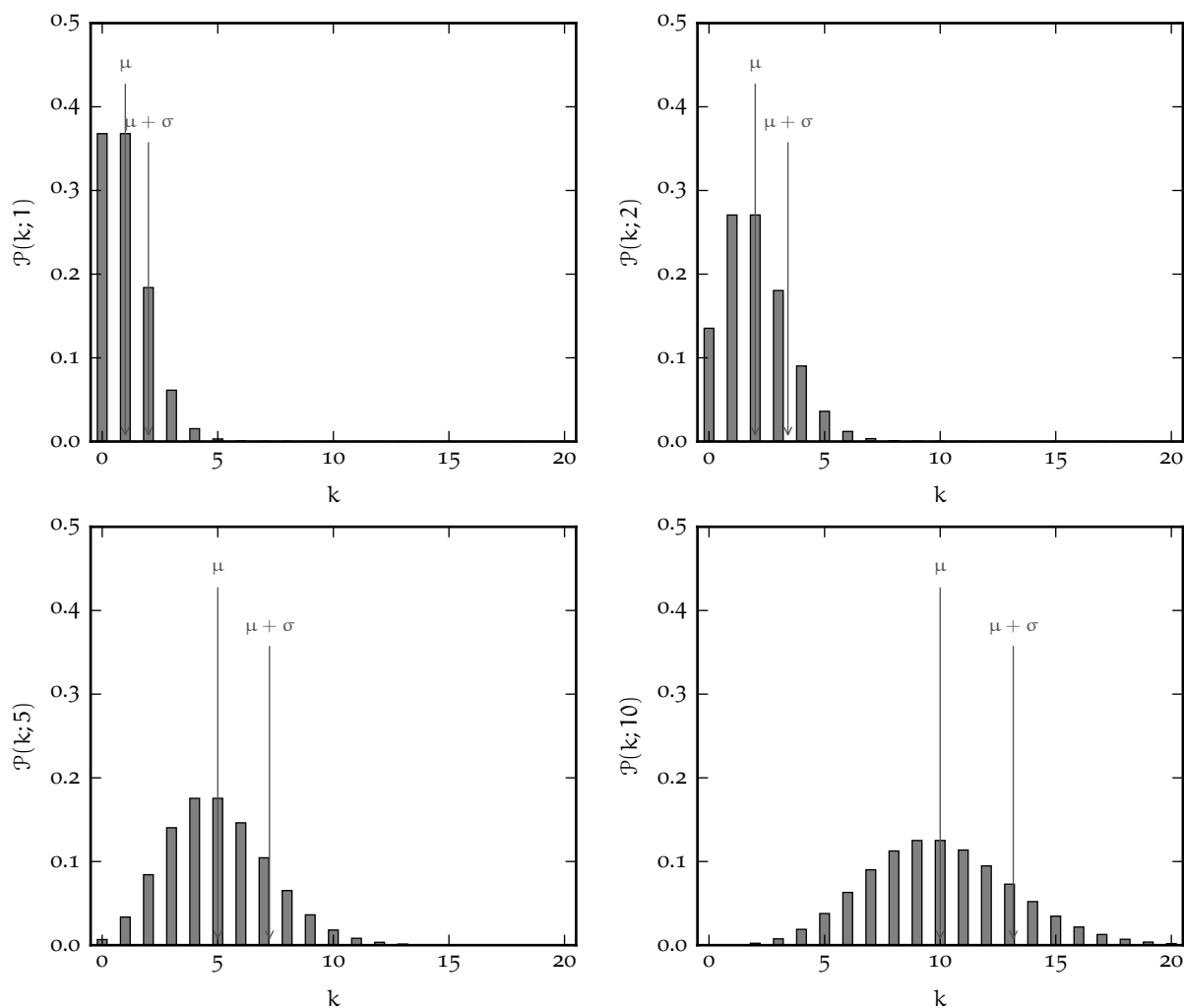


FIGURA 5.3. Esempi di distribuzione di Poisson (160) per diversi valori di  $\mu$ . Come discuteremo estensivamente nel seguito, all'aumentare di  $\mu$ , la distribuzione tende progressivamente ad assumere una caratteristica forma a campana.

Prima di andare avanti ci soffermiamo per un attimo sulle differenze tra la distribuzione binomiale e quella di Poisson—adesso che sappiamo che la seconda è un caso limite della prima. La distribuzione binomiale dipende da due parametri,  $n$  e  $p$ , che concorrono entrambi a determinare il valor medio e la varianza; la distribuzione di Poisson dipende da un solo parametro—la media, che incidentalmente coincide anche con la varianza.

Ma la cosa più importante è che, nello schema binomiale, la variabile casuale  $k$  è limitata superiormente ( $k \leq n$ ), mentre in quello di Poisson  $k$  può assumere qualsiasi valore intero da 0 a  $\infty$ —anche se in pratica la probabilità corrispondente decresce velocemente al crescere di  $k$ . Quando vi chiedete se un certo fenomeno segua la statistica binomiale o Poissoniana, questo è un buon indicatore: il valore massimo di occorrenze è fissato oppure no? (E, in pratica, vi sono molte situazioni in cui la vita non è né perfettamente binomiale, né perfettamente Poissoniana.)

#### 5.4.3 Due esempi di interesse storico ed alcune considerazioni

Prima di andare avanti ci soffermiamo brevemente su due esempi celebri di applicazione della distribuzione di Poisson, che dal nostro punto di vista sono utili per illustrare e sottolineare alcuni concetti fondamentali.

Apparentemente uno degli esempi classici (e quello che ha attratto l'attenzione iniziale sulla distribu-



► **ESEMPIO 5.6.** Supponiamo di lanciare tre dadi per 500 volte e di registrare, ad ogni lancio, la somma  $s$  delle uscite dei tre dadi in questione. Ci chiediamo la probabilità che il valore  $s = 4$  esca per 10 volte. Il modo formalmente corretto di impostare il problema è quello di utilizzare la statistica binomiale. Con tre dadi a sei facce si hanno  $6^3 = 216$  possibili combinazioni in totale; quelle che danno come somma 4 sono esattamente 3: (1, 1, 2), (1, 2, 1) e (2, 1, 1). La probabilità che la somma delle uscite in un lancio di tre dadi sia 4 è dunque  $p = 3/216$ . A questo punto è facile calcolare la probabilità che questo accada 10 volte in 500 lanci:

$$P = \mathcal{B}(10; 500, 3/216) = \binom{500}{10} \times (3/216)^{10} \times (213/216)^{490} \simeq 0.06933.$$

Alla luce di quanto detto in questo paragrafo possiamo anche applicare la statistica Poissoniana ( $n$  è grande e  $p$  è piccolo). Su 500 lanci  $s$  ammonterà a 4, in media, un numero di volte pari a:

$$\mu = 500 \times \frac{3}{216} = \frac{125}{18} \simeq 6.944$$

e la probabilità di avere 10 volte il valore 4 sarà dunque:

$$P = \mathcal{P}(10; 125/18) = \frac{6.944^{10}}{10!} \times e^{-6.944} \simeq 0.06928.$$

zione stessa) si deve a Vladislav Iosifovič von Bortkevič e risale al 1898. Esaminando i registri storici della cavalleria Prussiana von Bortkevič dimostrò in un celebre saggio [13] come il numero di soldati uccisi da un incidente a cavallo in un anno in un generico reparto della cavalleria stessa seguisse la distribuzione di Poisson.

$k$	$o_k$	$\mathcal{P}(k; 122/200)$	$e_k$
0	109	0.5434	108.67
1	65	0.3314	66.29
2	22	0.1011	20.22
3	3	0.0206	4.11
$\geq 4$	1	0.0036	0.71
	200	1.0000	200.00

**TABELLA 13.** Distribuzione del numero  $k$  di decessi per anno dovuti ad incidenti a cavallo per 10 reparti della cavalleria Prussiana [13]. I dati corrispondono ad un arco di tempo di 20 anni, per un totale di  $10 \times 20 = 200$  osservazioni e 122 decessi—con una media di  $122/200 = 0.61$  decessi l'anno. La tabella riporta le occorrenze osservate  $o_k$  e quelle attese  $e_k$  da una distribuzione di Poisson con media  $\mu = 0.61$ .

Il campione di dati in [13], riportato in tabella 13 si riferisce a 10 reparti osservati in un arco di tempo di 20 anni, per un totale di 200 osservazioni, con un numero complessivo di decessi pari a

$$\sum_{k=0}^4 k o_k = 0 \times 109 + 1 \times 65 + 2 \times 22 + 3 \times 3 + 1 \times 4 = 122$$

ed un numero medio di decessi per reparto per anno di  $122/200 \approx 0.61$ . Nella tabella 13  $o_k$  sono le occorrenze osservate per ciascun valore di  $k$  (e assommano a 200) ed  $e_k = 200 \times \mathcal{P}(k; 122/200)$  sono le occorrenze attese nel caso in cui il processo sia effettivamente Poissoniano, con la media  $\mu = 122/200 = 0.61$  stimata dai dati. Non abbiamo ancora gli strumenti per precisare il senso di questa affermazione, ma l'accordo tra  $o_k$  ed  $e_k$  è senza dubbio degno di nota.

Questo primo esempio è interessante perché apparentemente non ha molto a che vedere con lo schema binomiale. Dato l'oggetto della discussione (incidenti a cavallo) non abbiamo un vero e proprio processo elementare con una probabilità definita  $p$  di successo (se così vogliamo chiamarlo)—a meno che non pensiamo di registrare ogni volta che un soldato sale a cavallo e consideriamo un incidente come ad un evento che avviene in una certa frazione dei casi. Si tratterebbe di uno schema di difficile applicazione, ma soprattutto inutile: non abbiamo bisogno di inquadrare il problema in termini di tentativi e probabilità. La media del numero di incidenti mortali è sufficiente per avere un modello completo che fornisce una descrizione ragionevole della realtà. Allora la domanda che possiamo farci è: qual è l'implicazione fisica più rilevante del fatto che i nostri dati sono ben descritti da una distribuzione di Poisson? Evidentemente

il fatto che il fenomeno è sostanzialmente riproducibile su 10 divisioni distinte e stazionario su 20 anni—se ci pensiamo per un attimo non è una cosa che lascia indifferenti, no?

Il secondo esempio non è meno interessante. Per sottoporre a verifica il livello di precisione dell'artiglieria area Tedesca, in un articolo che ha tuttora un certo interesse storico, R. D. Clarke [2] ha diviso la zona sud di Londra in una griglia di  $24 \times 24 = 576$  regioni quadrate di eguale area ed ha contato il numero  $k$  di ordigni caduti entro ciascuna regione durante i bombardamenti Tedeschi di Londra nella seconda guerra mondiale. Le occorrenze di ciascun valore di  $k$  (per un totale di 537 ordigni nell'intervallo di tempo considerato) sono riportate in tabella 14.

$k$	$o_k$	$\mathcal{B}(k; 1/576, 537)$	$\mathcal{P}(k; 537/576)$	$e_k$
0	229	0.3933	0.3937	226.74
1	211	0.3673	0.3667	211.39
2	93	0.1712	0.1711	98.54
3	35	0.0531	0.0532	30.62
4	7	0.0123	0.0124	7.14
$\geq 5$	1	0.0027	0.0027	1.57
	576	1.0000	1.0000	576.00

TABELLA 14. Distribuzione delle occorrenze  $o_k$  della caduta di  $k$  ordigni su una griglia di 576 regioni di uguale area durante i bombardamenti Tedeschi di Londra della seconda guerra mondiale. Nell'ipotesi che la distribuzione delle bombe sia uniforme, il processo segue una legge binomiale nelle condizioni in cui vale l'approssimazione Poissoniana.

Nell'ipotesi in cui le bombe siano sganciate in modo uniforme nella regione considerata, ci aspettiamo che il numero  $k$  di bombe per regione sia descritto da una distribuzione binomiale con  $p = 1/576$  (la probabilità di scegliere a caso una delle 576 regioni) ed  $n = 537$  (il numero totale di ordigni). Ora, siamo proprio nel caso limite ( $p$  piccolo ed  $n$  grande) in cui la distribuzione binomiale tende a quella di Poisson, per cui, nelle nostre ipotesi, le occorrenze dovrebbero essere descritte altrettanto bene da una Poissoniana con media  $\mu = 537/576 \approx 0.932$ . In effetti entrambi i modelli forniscono una descrizione adeguata della realtà, come mostrato in tabella 14.

Questo esempio ha una ovvia interpretazione binomiale e costituisce una buona illustrazione della distribuzione di Poisson come limite della distribuzione binomiale. Si tratta anche di un esempio di processo Poissoniano che non avviene nel dominio del tempo—in questo caso la nostra variabile *dinamica* è una posizione nello spazio bidimensionale ed i nostri *intervalli* sono superfici. Da un punto di vista fisico il fatto che il processo segua una distribuzione di Poisson ci dice che esso è omogeneo nello spazio, che in questo contesto è l'equivalente della stazionarietà nel dominio del tempo.

#### 5.4.4 Normalizzazione, media e varianza

Sappiamo già che, come limite di una binomiale, la distribuzione di Poisson nella forma (160) ha media  $np \rightarrow \mu$  e varianza  $np(1-p) \rightarrow \mu$ , ma per completezza in questa sezione svolgiamo il calcolo esplicitamente. Per prima cosa si verifica banalmente che la distribuzione è correttamente normalizzata:

$$\sum_{k=0}^{\infty} \mathcal{P}(k; \mu) = \sum_{k=0}^{\infty} \frac{\mu^k}{k!} e^{-\mu} = e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = e^{-\mu} e^{\mu} = 1.$$

La media della distribuzione, come di consueto, si calcola formalmente secondo la (72)

$$E[k] = \sum_{k=0}^{\infty} k \mathcal{P}(k; \mu) = \sum_{k=0}^{\infty} k \frac{\mu^k}{k!} e^{-\mu} = e^{-\mu} \sum_{k=0}^{\infty} k \frac{\mu^k}{k!}.$$

Al solito il termine con  $k = 0$  non contribuisce alla somma, per cui possiamo far iniziare la somma stessa da  $k = 1$ :

$$E[k] = e^{-\mu} \sum_{k=1}^{\infty} k \frac{\mu^k}{k!} = e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^k}{(k-1)!}$$

e, ponendo  $h = k - 1$ ,

$$E[k] = e^{-\mu} \sum_{h=0}^{\infty} \frac{\mu^{h+1}}{h!} = \mu e^{-\mu} \sum_{h=0}^{\infty} \frac{\mu^h}{h!} = \mu e^{-\mu} e^{\mu} = \mu.$$

Come avevamo anticipato, il parametro  $\mu$  ha proprio il significato della media nel senso della definizione (72).

Per il calcolo della varianza partiamo, esattamente come abbiamo fatto nel caso della distribuzione binomiale, dal valore di aspettazione di  $k^2$ :

$$E[k^2] = \sum_{k=0}^{\infty} k^2 \mathcal{P}(k; \mu) = \sum_{k=0}^{\infty} k^2 \frac{\mu^k}{k!} e^{-\mu}.$$

Al solito eliminiamo il termine con  $k = 0$  e operiamo il cambiamento di variabile  $h = k - 1$ :

$$E[k^2] = \sum_{k=1}^{\infty} k^2 \frac{\mu^k}{k!} e^{-\mu} = \mu \sum_{h=0}^{\infty} (h+1) \frac{\mu^h}{h!} e^{-\mu} = \mu \left[ \sum_{h=0}^{\infty} h \frac{\mu^h}{h!} e^{-\mu} + \sum_{h=0}^{\infty} \frac{\mu^h}{h!} e^{-\mu} \right] = \mu^2 + \mu,$$

da cui, sfruttando la (78)

$$\sigma^2 = \mu. \quad (161)$$

Per la distribuzione di Poisson la varianza è uguale alla media e, di conseguenza, la deviazione standard è  $\sigma = \sqrt{\mu}$ . Facile da ricordare.

#### 5.4.5 Momenti di ordine superiore



Esattamente come abbiamo visto per la distribuzione binomiale, anche per la distribuzione di Poisson è possibile ricavare una relazione ricorsiva che consente di calcolare i momenti algebrici di ordine generico a partire da quelli di ordine più basso. Il ragionamento procede in modo analogo, partendo da

$$\begin{aligned} \frac{d}{d\mu} E[k^m] &= \frac{d}{d\mu} \sum_{k=0}^{\infty} k^m \frac{\mu^k}{k!} e^{-\mu} = \sum_{k=0}^{\infty} \frac{k^m}{k!} \frac{d}{d\mu} [\mu^k e^{-\mu}] = \sum_{k=0}^{\infty} \frac{k^m}{k!} [k\mu^{k-1} e^{-\mu} - \mu^k e^{-\mu}] = \\ &= \sum_{k=0}^{\infty} k^{m+1} \frac{\mu^{k-1}}{k!} e^{-\mu} - \sum_{k=0}^{\infty} k^m \frac{\mu^k}{k!} e^{-\mu} = \frac{1}{\mu} E[k^{m+1}] - E[k^m], \end{aligned}$$

da cui

$$E[k^{m+1}] = \mu \left( \frac{d}{d\mu} E[k^m] + E[k^m] \right). \quad (162)$$

Così, esattamente come prima

$$E[k^2] = \mu \left( \frac{d}{d\mu} E[k] + E[k] \right) = \mu(1 + \mu) = \mu^2 + \mu,$$

come già sapevamo, e

$$E[k^3] = \mu \left( \frac{d}{d\mu} E[k^2] + E[k^2] \right) = \mu(2\mu + 1 + \mu^2 + \mu) = \mu^3 + 3\mu^2 + \mu.$$

Il momento centrale di ordine 3 ed il coefficiente di asimmetria sono rispettivamente

$$\mu_3 = E[k^3] - 3\mu\sigma^2 - \mu^3 = \mu^3 + 3\mu^2 + \mu - 3\mu^2 - \mu^3 = \mu$$

e

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{1}{\sqrt{\mu}}. \quad (163)$$

(che si può ottenere direttamente come limite per  $p \rightarrow 0$  dell'espressione corrispondente per la binomiale ricordando che  $\mu = np$ .)

### 5.4.6 Distribuzione di Poisson e distanza tra eventi successivi

Torniamo per un attimo allo schema del nostro processo Poissoniano nel dominio del tempo mostrato in figura 5.2. Sappiamo che il numero di eventi  $k$  che si verificano in un intervallo di tempo fissato  $\Delta t$  segue la distribuzione (160). Cosa possiamo dire sulla distribuzione del tempo  $t$  che intercorre tra due eventi successivi?

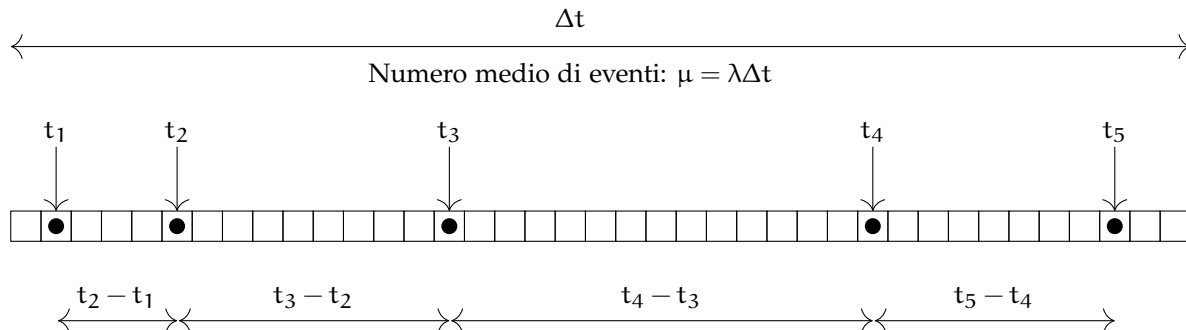


FIGURA 5.4. Rappresentazione schematica di un processo Poissoniano nel dominio del tempo con frequenza caratteristica  $\lambda$  (cfr. figura 5.2). La distanza temporale media tra due eventi successivi è  $\lambda^{-1}$  ma, data la natura aleatoria del processo, gli eventi, ovviamente, non sono equispaziati.

Per prima cosa  $t$  è una variabile aleatoria: è ovvio che gli eventi in un processo Poissoniano non sono temporalmente equispaziati—perché se così fosse il numero di eventi in un intervallo di tempo fissato non fluttuerebbe, ma sarebbe determinato a priori. Calcolare la media di  $t$  non è difficile: abbiamo in media  $\mu = \lambda \Delta t$  eventi in un intervallo di lunghezza  $\Delta t$  per cui il tempo medio tra due eventi successivi è

$$E[t] = \frac{\Delta t}{\mu} = \frac{\Delta t}{\lambda \Delta t} = \frac{1}{\lambda},$$

ovvero l'inverso della frequenza caratteristica del processo  $\lambda$ . (Notiamo che  $\lambda$  ha le dimensioni fisiche di  $[t]^{-1}$ , per cui la nostra equazione è, come deve essere, dimensionalmente corretta. Se  $\lambda$  è  $10 \text{ s}^{-1}$ , cioè 10 Hz, allora la distanza media tra due eventi successivi è 0.1 s.)

Ma possiamo fare molto di più. Possiamo calcolare l'espressione esplicita per la funzione di distribuzione di  $t$  notando che, dato il fatto che un evento si sia verificato ad un certo istante  $t_0$ , la probabilità (infinitesima) che l'evento successivo si verifichi entro un intervallino (infinitesimo) di durata  $dt$  ad una distanza temporale  $t$  è data dal prodotto

$$dP(t, dt) = e^{-\lambda t} \times \lambda dt,$$

in cui il primo termine rappresenta la probabilità che non si verifichi nessun evento tra i tempi  $t_0$  e  $t_0 + t$  ed il secondo la probabilità che si verifichi esattamente un evento tra i tempi  $t_0 + t$  e  $t_0 + t + dt$ . Dividendo entrambi i membri per  $dt$  possiamo calcolare la probabilità specifica per unità di tempo—vale a dire la nostra densità di probabilità

$$p(t; \lambda) = \lambda e^{-\lambda t}. \quad (164)$$

Le distanze tra eventi successivi in un processo Poissoniano sono distribuite esponenzialmente. Studieremo in dettaglio le proprietà della (164) nella sezione 5.6, ma anticipiamo che, come abbiamo già detto, la media di  $t$  è  $E[t] = \lambda^{-1}$ .

### 5.4.7 Somma di variabili Poissoniane

Consideriamo due variabili Poissoniane (indipendenti)  $l$  ed  $m$  con medie  $\mu_l$  e  $\mu_m$  rispettivamente. Ci proponiamo di capire come è distribuita la loro somma  $k = l + m$ . Formalmente dobbiamo sommare, per

un dato valore di  $k$ , su tutte le coppie di numeri positivi che danno come somma  $k$ ; ad esempio possiamo ottenere  $k = 2$  con  $l = 0$  e  $m = 2$  oppure  $l = 1$  e  $m = 1$  o ancora  $l = 2$  e  $m = 0$ . In generale:

$$P(k) = \sum_{l=0}^k \mathcal{P}(l; \mu_l) \mathcal{P}(k-l; \mu_m) \quad (165)$$

(abbiamo utilizzato la moltiplicazione delle probabilità per eventi indipendenti). Esplicitamente:

$$P(k) = \sum_{l=0}^k \frac{\mu_l^l}{l!} e^{-\mu_l} \cdot \frac{\mu_m^{k-l}}{(k-l)!} e^{-\mu_m} = e^{-(\mu_l + \mu_m)} \sum_{l=0}^k \frac{\mu_l^l \mu_m^{k-l}}{l!(k-l)!} = \frac{e^{-(\mu_l + \mu_m)}}{k!} \sum_{l=0}^k \binom{k}{l} \mu_l^l \mu_m^{k-l}.$$

Ma nella sommatoria finale riconosciamo la potenza di binomio (52), per cui:

$$P(k) = \frac{(\mu_l + \mu_m)^k}{k!} e^{-(\mu_l + \mu_m)} = \mathcal{P}(k; \mu_l + \mu_m), \quad (166)$$

cioè la somma di due variabili Poissoniane è ancora una variabile Poissoniana la cui media è la somma delle medie. (Va da sé che questo risultato si può estendere alla somma di un numero arbitrario di variabili Poissoniane.)

► **ESEMPIO 5.7 (MA IL POSTINO VIENE VERAMENTE TUTTI I GIORNI?).** Supponiamo di ricevere in media 75 lettere l'anno. Un giorno arriviamo a casa e troviamo nella cassetta della posta 5 lettere (nessuna delle quali era presente il giorno precedente). Possiamo concludere che il postino non consegna la posta tutti i giorni?

Per la discussione che segue assumeremo che la consegna della posta avvenga (di norma) anche la domenica e che il flusso di missive ad un certo indirizzo sia un fenomeno stazionario. La prima è un'assunzione *innocua* (che facciamo solo per evitare complicazioni inessenziali), mentre la seconda è un'assunzione *forte*—ma non del tutto irragionevole, o almeno non così irragionevole da rendere la discussione irrilevante. Sotto queste ipotesi il numero  $k$  di lettere che, giorno per giorno, troviamo nella cassetta della posta è una variabile Poissoniana, la cui media può essere stimata come  $m = 75/365 \approx 0.2$ . Possiamo allora riformulare in termini quantitativi la domanda iniziale chiedendoci quale sia la probabilità di trovare 5 o più lettere quando in media ce ne attendiamo 0.2:

$$P(k \geq 5) = \sum_{k=5}^{\infty} \mathcal{P}(k; m).$$

Si tratta del tipico caso in cui il teorema della probabilità totale permette di aggirare la serie infinita:

$$P(k \geq 5) = 1 - P(k < 5) = 1 - \sum_{k=0}^4 \mathcal{P}(k; m) = 1 - \sum_{k=0}^4 \frac{m^k}{k!} e^{-m} \approx 2.6 \times 10^{-6}.$$

La probabilità è abbastanza piccola (almeno sotto le ipotesi iniziali) da legittimarci ad un reclamo ufficiale.

Chiediamoci adesso quale sia la probabilità di trovare nella cassetta della posta 5 lettere nell'ipotesi in cui postino consegnino la corrispondenza solo una volta alla settimana. La somma di 7 variabili Poissoniane con media  $m$  è una variabile Poissoniana con media  $m_s = 7m = 525/365 \approx 1.44$ , e adesso:

$$P(k \geq 5) = 1 - P(k < 5) = 1 - \sum_{k=0}^4 \mathcal{P}(k; m_s) = \sum_{k=5}^{\infty} \frac{m_s^k}{k!} e^{-m_s} \approx 1.6\%,$$

che è piccola ma decisamente non piccola come prima.

## 5.5 LA DISTRIBUZIONE UNIFORME

La distribuzione uniforme è l'esempio più semplice di funzione di distribuzione di variabile casuale continua—la densità di probabilità è costante entro un intervallo finito e nulla fuori:

$$u(x; a, b) = \begin{cases} \frac{1}{(b-a)} & a \leq x \leq b \\ 0 & x < a; x > b \end{cases} \quad (167)$$

Un esempio particolare della densità di probabilità per una distribuzione uniforme (con  $a = 0$  e  $b = 1$ ) è mostrato in figura 5.5.

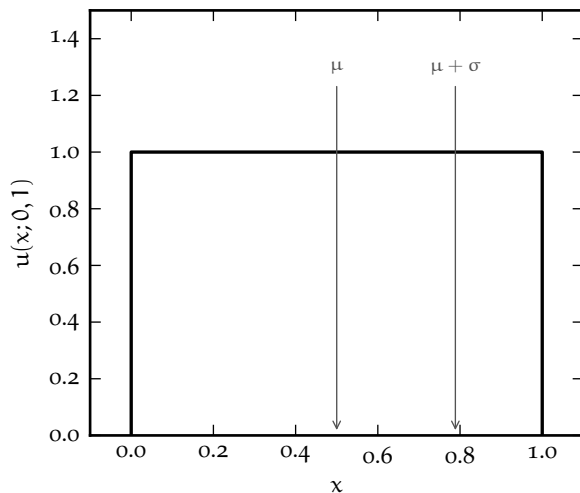


FIGURA 5.5. Esempio di distribuzione uniforme per  $a = 0$  e  $b = 1$ .

## 5.5.1 Normalizzazione, media, varianza e coefficiente di asimmetria

Si dimostra banalmente che la distribuzione, così come è scritta nella (167), è correttamente normalizzata:

$$\int_{-\infty}^{\infty} u(x; a, b) dx = \int_a^b \frac{1}{(b-a)} dx = \frac{1}{(b-a)} \int_a^b dx = \frac{1}{(b-a)} (b-a) = 1$$

(Abbiamo calcolato esplicitamente l'integrale, ma di fatto si tratta semplicemente dell'area di un rettangolo di base  $(b-a)$  e di altezza  $1/(b-a)$ —1, appunto.)

La media è data, per definizione, dal valore di aspettazione della variabile casuale

$$\mu = E[x] = \int_{-\infty}^{\infty} x u(x; a, b) dx = \frac{1}{(b-a)} \int_a^b x dx = \frac{1}{(b-a)} \left. \frac{x^2}{2} \right|_a^b = \frac{1}{(b-a)} \frac{(b^2 - a^2)}{2} = \frac{(b+a)}{2} \quad (168)$$

e coincide con il valor medio dell'intervallo su cui la densità di probabilità è non nulla—il che non dovrebbe sorprendere perché la funzione di distribuzione è simmetrica rispetto all'asse  $x = (b+a)/2$ . Per lo stesso motivo la mediana coincide con la media, mentre la moda non è definita poiché la densità di probabilità non ha un massimo.

Procediamo con il calcolo della varianza. Al solito partiamo dalla (78) e calcoliamo prima il valore di aspettazione di  $x^2$ :

$$\begin{aligned} E[x^2] &= \int_{-\infty}^{\infty} x^2 u(x; a, b) dx = \frac{1}{(b-a)} \int_a^b x^2 dx = \frac{1}{(b-a)} \left. \frac{x^3}{3} \right|_a^b = \frac{1}{(b-a)} \frac{(b^3 - a^3)}{3} = \\ &= \frac{1}{(b-a)} \frac{(b-a)(b^2 + ab + a^2)}{3} = \frac{(b^2 + ab + a^2)}{3} \end{aligned}$$

da cui

$$\begin{aligned}\sigma^2 &= E[x^2] - \mu^2 = \frac{(b^2 + ab + a^2)}{3} - \frac{(b+a)^2}{4} = \frac{(4b^2 + 4ab + 4a^2 - 3b^2 - 6ab - 3a^2)}{12} = \\ &= \frac{(b^2 - 2ab + a^2)}{12} = \frac{(b-a)^2}{12}\end{aligned}\quad (169)$$

e

$$\sigma = \frac{(b-a)}{\sqrt{12}}. \quad (170)$$

Questa  $\sqrt{12}$  al denominatore della deviazione standard è un numero importante da tenere a memoria, perché è quello che determina l'incertezza di misura (nel senso statistico) di tutti gli strumenti digitali—cioè di quegli strumenti che forniscono in uscita valori discreti (spaziati tra di loro di una quantità pari alla risoluzione strumentale) non affetti da fluttuazioni statistiche.

Per completezza la semilarghezza a metà altezza è pari alla semilarghezza dell'intervallo su cui la densità di probabilità è non nulla:

$$\text{HWHM} = \frac{(b-a)}{2} = \sqrt{3}\sigma \approx 1.73\sigma. \quad (171)$$

Poiché la distribuzione è simmetrica rispetto al valor medio, il coefficiente di asimmetria  $\gamma_1$ —e, più in generale, tutti i momenti centrali di ordine dispari—è banalmente nullo.

► **ESEMPIO 5.8.** Supponiamo di misurare la massa di un oggetto con una bilancia digitale con la risoluzione di un grammo; sia  $m = 58$  g il valore indicato dal display. Se possiamo escludere la presenza di effetti sistematici è ragionevole ammettere che il misurando sia compreso, con densità di probabilità uniforme, tra 57.5 e 58.5 g. La media della distribuzione sarà 58 g e la deviazione standard  $1/\sqrt{12} = 0.289$  g. Se vogliamo attribuire un errore statisticamente corretto alla nostra misura scriveremo, al livello di una deviazione standard:

$$m = 58.0 \pm 0.3 \text{ g}$$

### 5.5.2 Funzione cumulativa e quantili

La funzione cumulativa della distribuzione uniforme si calcola banalmente come

$$F(x) = \int_a^x \frac{1}{(b-a)} dt = \frac{(x-a)}{b-a} \quad a \leq x \leq b, \quad (172)$$

da cui la funzione di distribuzione inversa è

$$F^{-1}(q) = q(b-a) + a. \quad (173)$$

## 5.6 LA DISTRIBUZIONE ESPONENZIALE

Abbiamo già incontrato la distribuzione esponenziale nella sezione 5.4.6, a proposito della distanza tra due eventi successivi in un processo Poissoniano 164. In generale, dato un numero positivo  $\lambda > 0$  una distribuzione della forma

$$\varepsilon(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & 0 \leq x < \infty \\ 0 & x < 0 \end{cases} \quad (174)$$

si dice esponenziale con parametro  $\lambda$ . Vedremo tra un attimo che  $\lambda$  (che per ragioni dimensionali ha le dimensioni di  $x^{-1}$ ) ha il significato fisico dell'inverso della media della distribuzione, per cui a volte la (174) si trova anche scritta in funzione di  $\lambda^{-1}$ —che prende il nome di *vita media* nel caso di processi Poissoniani nel dominio del tempo e *cammino libero medio* nel caso di processi Poissoniani nel dominio dello spazio. Due esempi di distribuzione esponenziale, per due diversi valori di  $\lambda$ , sono mostrati in figura 5.6.

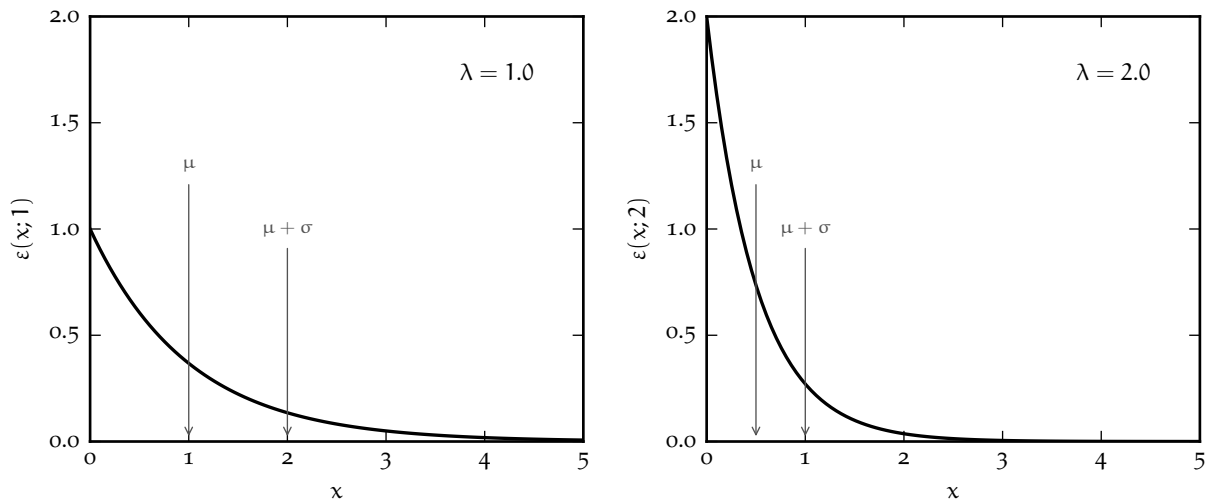


FIGURA 5.6. Esempi di distribuzione esponenziale (174) per due diversi valori di  $\lambda$  (1 e 2). Entrambe le distribuzioni sono correttamente normalizzate.

#### 5.6.1 Normalizzazione, media, varianza e coefficiente di asimmetria

La distribuzione esponenziale, così come scritta nella (174), è correttamente normalizzata:

$$\int_0^{\infty} \varepsilon(x; \lambda) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = -e^{-\lambda x} \Big|_0^{\infty} = 1.$$

La media della distribuzione si calcola al solito secondo la definizione 72 e vale

$$\begin{aligned} \mu = E[x] &= \int_{-\infty}^{\infty} x \varepsilon(x; \lambda) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^{\infty} r e^{-r} dr = \frac{1}{\lambda} \left( -r e^{-r} \Big|_0^{\infty} + \int_0^{\infty} e^{-r} dr \right) = \\ &= -\frac{1}{\lambda} e^{-r} \Big|_0^{\infty} = \frac{1}{\lambda}. \end{aligned} \quad (175)$$

(in cui abbiamo operato il cambiamento di variabile  $\lambda x = r$ , ed abbiamo calcolato l'integrale risultante per parti).

Per la varianza, al solito, si parte dal valore di aspettazione di  $x^2$ , che si calcola con lo stesso cambiamento di variabile  $\lambda x = r$  ed integrando per parti due volte

$$E[x^2] = \int_0^{\infty} x^2 \varepsilon(x; \lambda) dx = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2} \int_0^{\infty} r^2 e^{-r} dr = \frac{1}{\lambda^2} \left( -r^2 e^{-r} \Big|_0^{\infty} + 2 \int_0^{\infty} r e^{-r} dr \right) = \frac{2}{\lambda^2}.$$

Si ha dunque

$$\sigma^2 = E[x^2] - \mu^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \quad (176)$$

e

$$\sigma = \frac{1}{\lambda}. \quad (177)$$

La distribuzione esponenziale vale  $\lambda$  per  $x = 0$  e la retta orizzontale  $y = \lambda/2$  interseca la funzione densità di probabilità in  $x = 0$  ed in corrispondenza della radice dell'equazione

$$\frac{\lambda}{2} = \lambda e^{-\lambda x} \quad \text{ovvero} \quad x = \frac{\ln 2}{\lambda}.$$

Ne deriva immediatamente che la semilarghezza a metà altezza è data da

$$\text{HWHM} = \frac{\ln 2}{2\lambda} = \frac{\ln 2}{2} \sigma \approx 0.347 \sigma. \quad (178)$$



Integrando per parti come abbiamo fatto per il calcolo della media e della varianza è semplice ricavare per ricorrenza i momenti algebrici di ordine superiore

$$E[x^n] = \int_0^{\infty} x^n \lambda e^{-\lambda x} dx = -x^n e^{-\lambda x} \Big|_0^{\infty} + \frac{n}{\lambda} \int_0^{\infty} x^{n-1} \lambda e^{-\lambda x} dx = \frac{n}{\lambda} E[x^{n-1}].$$

Il momento algebrico di ordine 3, ad esempio, vale

$$E[x^3] = \frac{3}{\lambda} E[x^2] = \frac{6}{\lambda^3},$$

da cui

$$\mu_3 = E[x^3] - 3\mu\sigma^2 - \mu^3 = \frac{2}{\lambda^3} \quad \text{e} \quad \gamma_1 = 2. \quad (179)$$

(Che il coefficiente di asimmetria sia sempre positivo è cosa che non dovrebbe stupire perché la distribuzione ha una coda pronunciata per valori di  $x$  a destra della media.)

### 5.6.2 Funzione cumulativa e quantili

La funzione cumulativa della distribuzione esponenziale ha una espressione analitica semplice

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = -e^{-\lambda t} \Big|_0^x = 1 - e^{-\lambda x}, \quad (180)$$

che altrettanto semplicemente si può invertire per ricavare la funzione di distribuzione inversa

$$F^{-1}(q) = -\frac{\ln(1-q)}{\lambda}. \quad (181)$$

la funzione cumulativa e la funzione di distribuzione inversa sono mostrate in figura 5.7 per una distribuzione esponenziale con parametro  $\lambda = 1$ .

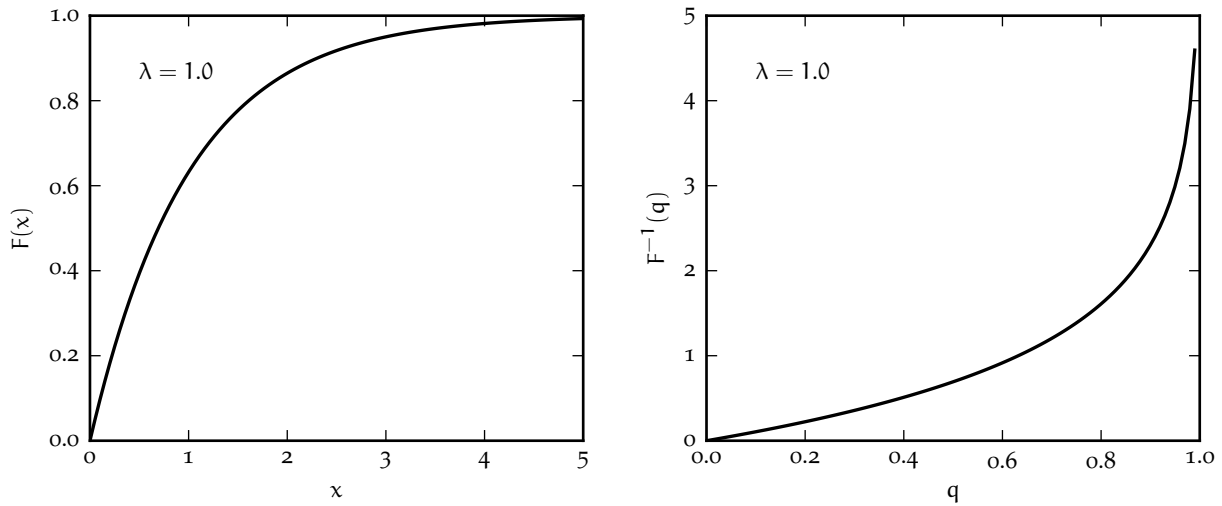


FIGURA 5.7. Funzione cumulativa e funzione di distribuzione inversa per una distribuzione esponenziale con parametro  $\lambda = 1$ .

### 5.6.3 Assenza di memoria

Una proprietà interessante della distribuzione esponenziale è costituita dal fatto che

$$P(x \geq x_1 + x_2) = \int_{x_1+x_2}^{\infty} \varepsilon(x; \lambda) dx = \lambda \int_{x_1+x_2}^{\infty} e^{-\lambda x} dx = e^{-\lambda(x_1+x_2)},$$

da cui segue banalmente che

$$P(x \geq x_1 + x_2) = P(x \geq x_1) P(x \geq x_2). \quad (182)$$

Ora, per definizione di probabilità condizionata, si ha anche che

$$P(x \geq x_1 + x_2) = P(x \geq x_1 + x_2 | x \geq x_1) P(x \geq x_1),$$

per cui la (182) si può anche scrivere come

$$P(x \geq x_1 + x_2 | x \geq x_1) = P(x \geq x_2). \quad (183)$$

Una variabile casuale che goda di questa proprietà si dice una variabile *senza memoria* (o anche *memory-less*).

► **ESEMPIO 5.9.** Supponiamo che il cammino libero medio  $l_0$  di una particella in un certo mezzo (omogeneo) sia 1 mm. Quando la particella penetra nel mezzo, in generale percorrerà una certa distanza  $x$  (che, a parità di condizioni iniziali, varia in modo casuale di volta in volta) prima di interagire; come detto prima, questa distanza è descritta da una densità di probabilità di tipo esponenziale con parametro  $\lambda = 1/l_0$ :

$$p(x) = \frac{1}{l_0} e^{-\frac{x}{l_0}}.$$

Il valore di aspettazione della variabile  $x$  è  $l_0$ . Il che significa, in altre parole, che la nostra particella percorre in media una distanza  $l_0$  prima di interagire (da cui il nome di cammino libero medio). Ci chiediamo quale sia la probabilità che la particella percorra una distanza  $l = 10l_0$  prima di interagire. La risposta è, banalmente

$$P(x \geq 10l_0) = e^{-10} \approx 4.54 \times 10^{-5},$$

cioè è estremamente poco probabile che una particella percorra una distanza maggiore a 10 volte il cammino libero medio prima di interagire.

► **ESEMPIO 5.10.** Supponiamo di *osservare*, a partire da un istante  $t_0$ , un nucleo radioattivo con vita media (inverso del parametro  $\lambda$ ) di 1 s. La probabilità che, dopo un secondo, il nucleo non sia decaduto è:

$$P(x \geq 1) = e^{-1} \approx 37\%$$

Supponiamo adesso che il nucleo non sia ancora decaduto dopo 10 s, il che è estremamente improbabile ma possibile (a proposito: quanto vale la probabilità di questo evento?). Ebbene: la probabilità che il nucleo non decada tra  $t_0 + 10$  s e  $t_0 + 11$  s è di nuovo il 37%.

Tanto per fissare le idee: se la stessa cosa valesse per un'automobile, ad ogni istante una vettura appena uscita dal concessionario ed una con 200000 km alle spalle avrebbero la stessa probabilità di rompersi entro il giorno successivo. Purtroppo, al contrario dei nuclei radioattivi le auto hanno memoria della propria storia!

## 5.7 LA DISTRIBUZIONE DI GAUSS

*"Everyone believes in it: experimentalists believing that it is a mathematical theorem, mathematicians believing that it is an empirical fact."* Questa osservazione, generalmente attribuita a Henri Poincaré, riassume efficacemente la rilevanza che la più *celebre* funzione di distribuzione—quella di Gauss, appunto—riveste nella teoria della probabilità e nella statistica. Come vedremo nel seguito, la distribuzione di Gauss può essere vista come il limite di una distribuzione di Poisson per  $\mu \rightarrow \infty$  o di una distribuzione binomiale nel limite  $n \rightarrow \infty$  (questa volta senza nessuna ipotesi aggiuntiva su  $p$ ). Inoltre, per il teorema del limite centrale, la distribuzione della media di un numero abbastanza grande di campionamenti di una variabile casuale, indipendentemente dalla sua distribuzione, è distribuita Gaussianamente. Questi sono solo alcuni dei motivi per cui questa distribuzione si trova così spesso in pratica.

## 5.7.1 La distribuzione di Gauss come limite della Poissoniana

Consideriamo una variabile casuale Poissoniana  $k$  e calcoliamo il logaritmo naturale della sua funzione di distribuzione

$$\ln \mathcal{P}(k; \mu) = \ln \left( \frac{\mu^k}{k!} e^{-\mu} \right) = k \ln \mu - \ln(k!) - \mu.$$

Il termine più problematico è chiaramente quello con il fattoriale, ma se  $\mu$  è molto grande (e quindi anche  $k$  è tendenzialmente molto grande, visto che  $\mu = E[k]$ ) possiamo utilizzare la formula di Stirling (47) e riscrivere la nostra espressione (questa volta in forma approssimata) come

$$\ln \mathcal{P}(k; \mu) \approx k \ln \mu - \frac{1}{2} \ln(2\pi k) - k \ln k + k - \mu.$$

L'idea è adesso quella di sviluppare in qualche modo questa espressione in serie di Taylor, ma il problema di fondo è lo stesso che abbiamo incontrato nel derivare la distribuzione di Poisson come limite della binomiale: possiamo fare il limite per  $\mu \rightarrow \infty$ , ma allo stesso tempo abbiamo bisogno di una prescrizione per  $k$ , che è una variabile casuale e, come tale, non è univocamente determinata da  $\mu$ . Introduciamo dunque la variabile casuale *ridotta*

$$\delta = \frac{k - \mu}{\mu} \quad \text{ovvero} \quad k = \mu(1 + \delta).$$

Le fluttuazioni di  $k$  attorno al valor medio (e quindi il valore del numeratore dell'espressione appena scritta) saranno dell'ordine di  $\sigma = \sqrt{\mu}$  e, per  $\mu \rightarrow \infty$ ,  $\delta \rightarrow 0$  come  $1/\sqrt{\mu}$ . Possiamo dunque scrivere

$$\ln \mathcal{P}(\delta; \mu) \approx \mu(1 + \delta) \ln \mu - \frac{1}{2} \ln(2\pi\mu(1 + \delta)) - \mu(1 + \delta) \ln(\mu(1 + \delta)) + \mu(1 + \delta) - \mu.$$

A questo punto possiamo sviluppare in serie attorno al valore  $\delta = 0$ —e vedremo tra un secondo che sviluppare al prim'ordine non basta ed avremo bisogno del termine con la derivata seconda. Il calcolo è un po' tedioso, ma le derivate rilevanti sono

$$\frac{d \ln \mathcal{P}(0; \mu)}{d\delta} = \left[ -\mu \ln(1 + \delta) - \frac{1}{2(1 + \delta)} \right]_{\delta=0} = -\frac{1}{2} \quad \text{e} \quad \frac{d^2 \ln \mathcal{P}(0; \mu)}{d\delta^2} = \left[ -\frac{\mu}{(1 + \delta)} + \frac{1}{2(1 + \delta)^2} \right]_{\delta=0} = -\mu,$$

da cui

$$\ln \mathcal{P}(\delta; \mu) \approx \ln \mathcal{P}(0; \mu) + \frac{d \ln \mathcal{P}(0; \mu)}{d\delta} \delta + \frac{1}{2} \frac{d^2 \ln \mathcal{P}(0; \mu)}{d\delta^2} \delta^2 = -\frac{1}{2} \ln(2\pi\mu) - \frac{1}{2} \delta - \frac{1}{2} \mu \delta^2.$$

Guardiamo più attentamente la relazione che abbiamo appena ottenuto. Per  $\delta \rightarrow 0$  il secondo termine si annulla (che è il motivo per cui non abbiamo troncato lo sviluppo al prim'ordine) ed il terzo è quello rilevante—nel valutare la forma indeterminata  $\mu \delta^2$  ricordiamo che  $\delta \rightarrow 0$  come  $1/\sqrt{\mu}$ . Possiamo adesso ripristinare il  $k$  che avevamo momentaneamente *nascosto*, col che otteniamo

$$\ln \mathcal{P}(k; \mu) \approx -\frac{1}{2} \ln(2\pi\mu) - \frac{1}{2} \frac{(k - \mu)^2}{\mu} \quad \text{ovvero} \quad \mathcal{P}(k; \mu) \approx \frac{1}{\sqrt{2\pi\mu}} e^{-\frac{1}{2} \frac{(k - \mu)^2}{\mu}}. \quad (184)$$

$k$	$\mathcal{P}(k; 100)$	Eq. (184)	Errore relativo
90	$2.504 \times 10^{-2}$	$2.420 \times 10^{-2}$	$3.36 \times 10^{-2}$
95	$3.601 \times 10^{-2}$	$3.521 \times 10^{-2}$	$2.24 \times 10^{-2}$
100	$3.986 \times 10^{-2}$	$3.989 \times 10^{-2}$	$8.34 \times 10^{-4}$
105	$3.440 \times 10^{-2}$	$3.521 \times 10^{-2}$	$2.34 \times 10^{-2}$
110	$2.342 \times 10^{-2}$	$2.420 \times 10^{-2}$	$3.31 \times 10^{-2}$
115	$1.272 \times 10^{-2}$	$1.295 \times 10^{-2}$	$1.84 \times 10^{-2}$

TABELLA 15. Illustrazione della validità della (184) come approssimazione delle distribuzioni di Poisson, per alcuni valori rappresentativi di  $k$  e  $\mu = 100$ . L'errore relativo è di qualche % al massimo, e l'approssimazione migliora al crescere di  $\mu$ .

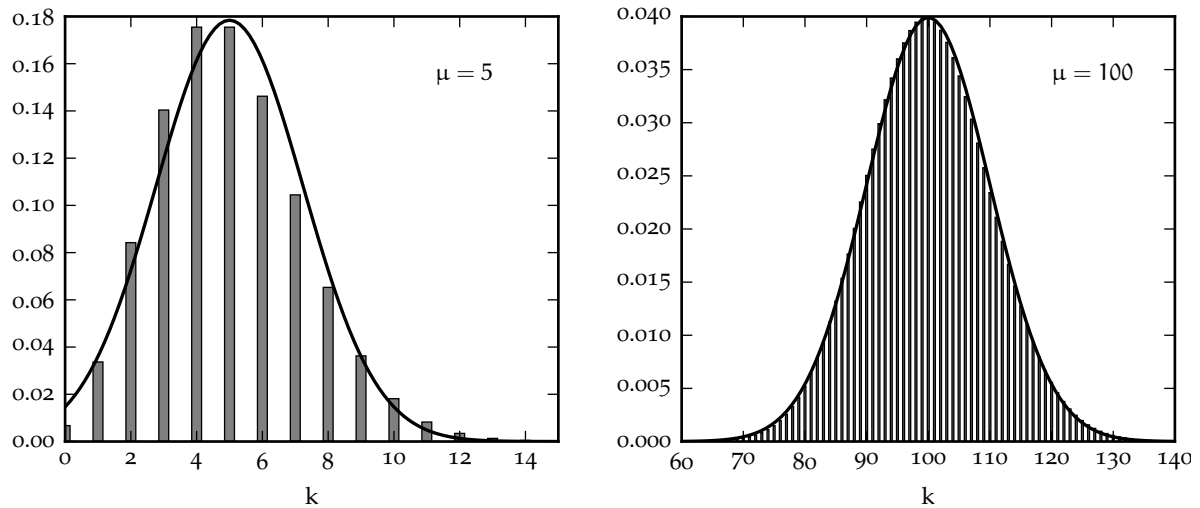


FIGURA 5.8. Approssimazione Gaussiana della distribuzione di Poisson per due diversi valori della media— $\mu = 5$  e  $\mu = 100$ .

A questo livello la (184) è un'espressione approssimata per la (160) che può essere utilizzata quando  $\mu$  è grande, come illustrato in figura 5.8 ed in tabella 15. Ma fermiamoci per un attimo ad osservare meglio la relazione che abbiamo appena ricavato. Per prima cosa la (184) è simmetrica rispetto al valor medio della distribuzione  $k = \mu$ . Questo non dovrebbe sorprendere, poiché sappiamo già che il coefficiente di asimmetria della distribuzione di Poisson vale  $\gamma_1 = 1/\sqrt{\mu}$ , che tende a 0 per  $\mu \rightarrow \infty$ . (In altre parole sapevamo già dalla (163) che la distribuzione di Poisson tende a diventare simmetrica per valori grandi della media.) La seconda osservazione è che, dato che varianza  $\sigma^2$  della distribuzione di Poisson è uguale al valor medio  $\mu$ , possiamo in effetti riscrivere la (184) nella forma alternativa

$$\mathcal{P}(k; \mu) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-\mu}{\sigma}\right)^2}.$$

Se adesso sostituiamo la variabile casuale discreta  $k$  con una variabile continua  $x$ , e lasciamo la deviazione standard  $\sigma$  libera di variare indipendentemente dalla media, abbiamo ottenuto la distribuzione di Gauss che studieremo in dettaglio nelle prossime sottosezioni. Torneremo più in dettaglio sul passaggio dal discreto al continuo nel seguito.

### 5.7.2 Normalizzazione, media e varianza

La funzione di distribuzione di Gauss (o distribuzione normale) si scrive dunque nella forma

$$N(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (185)$$

dove  $\mu$  e  $\sigma^2$ , come vedremo tra un attimo, sono proprio la media e la varianza della distribuzione nel senso delle (72) e (75)—e sono indipendenti tra di loro (cioè non sono più necessariamente coincidenti come nel caso Poissoniano). Una variabile casuale Gaussiana  $z$  con media 0 e varianza 1, che ha funzione di distribuzione (completamente fissata)

$$N(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad (186)$$

si dice variabile Gaussiana *in forma standard*. Una variabile Gaussiana generica  $x$  può essere trasformata nella corrispondente variabile in forma standard attraverso il cambio di variabile

$$z = \frac{x - \mu}{\sigma} \quad \text{ovvero} \quad x = \sigma z + \mu. \quad (187)$$

Si tratta di una trasformazione importante, che utilizzeremo spesso nel seguito, e per completezza, verifichiamo esplicitamente che tutto funzioni come deve:

$$E[z] = E\left[\frac{x - \mu}{\sigma}\right] = \frac{1}{\sigma} (E[x] - E[\mu]) = 0 \quad \text{e} \quad \text{Var}(z) = E[z^2] = E\left[\frac{(x - \mu)^2}{\sigma^2}\right] = \frac{1}{\sigma^2} E[(x - \mu)^2] = 1.$$

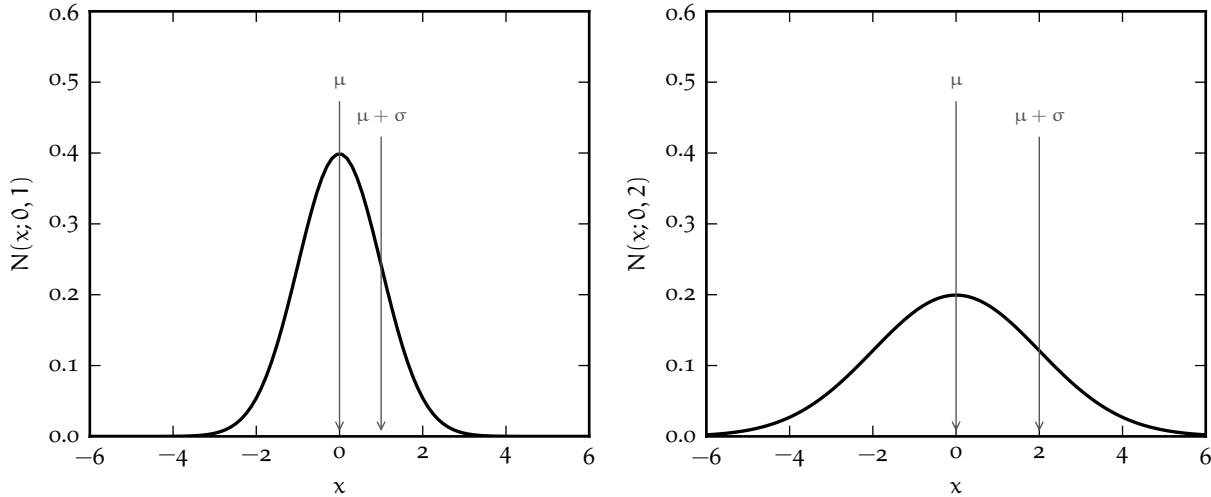


FIGURA 5.9. Esempi di distribuzione di Gauss (185) per  $\mu = 0$  e due diversi valori di  $\sigma$  (1 e 2). Si noti che l'altezza della distribuzione (cioè il valore in  $x = 0$ ) è inversamente proporzionale alla deviazione standard, per cui essa dimezza quando  $\sigma$  raddoppia.

La distribuzione di Gauss è mostrata in figura 5.9 per due valori diversi di  $\sigma$ . Notiamo esplicitamente che il valore massimo della funzione (che si ha per  $x = \mu$ ) è inversamente proporzionale a  $\sigma$ , per cui se raddoppiamo la deviazione standard della distribuzione, stiamo effettivamente raddoppiando la larghezza e dimezzando l'altezza. (L'integrale sotto la curva rimane costante per la condizione di normalizzazione.)

Verifichiamo che la distribuzione di Gauss, scritta nella forma (185), è correttamente normalizzata, e che—come accennato poco fa— $\mu$  e  $\sigma$  rappresentano la media e la deviazione standard della distribuzione. Scriviamo, al solito

$$\int_{-\infty}^{\infty} N(x; \mu, \sigma) dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz := \frac{1}{\sqrt{2\pi}} I_0.$$

Sfortunatamente questo integrale non ha espressione analitica, nel senso che non esiste una primitiva dell'integrando esprimibile in forma chiusa in termini di funzioni elementari<sup>1</sup>. Un modo semplice per calcolare l'integrale definito  $I_0$  è quello di scrivere il suo quadrato come prodotto di due integrali indipendenti, utilizzare il teorema di Fubini e passare in coordinate polari (facendo attenzione a come trasforma l'elemento di superficie  $dx dy \rightarrow r dr d\phi$ )

$$\begin{aligned} I_0^2 &= \left( \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx \right) \times \left( \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy \right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)} dx dy = \int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}r^2} r dr d\theta = \\ &= 2\pi \int_0^{\infty} e^{-\frac{1}{2}r^2} d\left(\frac{r^2}{2}\right) = -2\pi e^{-\frac{1}{2}r^2} \Big|_0^{\infty} = 2\pi \quad \text{da cui} \quad I_0 = \sqrt{2\pi}, \end{aligned}$$

che dimostra effettivamente che la distribuzione di partenza è correttamente normalizzata.

Prima di andare avanti vale la pena spendere un istante su una classe di integrali definiti (di cui il precedente è un caso particolare) che ci sarà utile nel seguito, ovvero

$$I_n = \int_{-\infty}^{\infty} x^n e^{-\frac{1}{2}x^2} dx = - \int_{-\infty}^{\infty} x^{n-1} d\left(e^{-\frac{1}{2}x^2}\right) = -x^{n-1} e^{-\frac{1}{2}x^2} \Big|_{-\infty}^{\infty} + (n-1) \int_{-\infty}^{\infty} x^{n-2} e^{-\frac{1}{2}x^2} dx.$$

<sup>1</sup> La questione è, almeno in parte, semantica, poiché l'integrale definito di  $e^{-\frac{1}{2}z^2}$  tra due estremi arbitrari può essere calcolato numericamente (e tabulato, come vedremo) con il grado di accuratezza desiderato.

Per ragioni di simmetria  $I_n$  è nullo per  $n$  dispari (l'integrando è dispari ed il dominio, sia pure infinito, è simmetrico rispetto allo zero). Per  $n$  pari, integrando per parti, si ottiene come abbiamo appena visto l'interessante relazione per ricorrenza

$$\begin{cases} I_0 = \sqrt{2\pi} \\ I_1 = 0 \\ I_n = (n-1) I_{n-2}, \end{cases} \quad (188)$$

da cui  $I_2 = \sqrt{2\pi}$ ,  $I_3 = 0$ ,  $I_4 = 3\sqrt{2\pi}$  e così via.

La media della distribuzione, formalmente, si può allora scrivere, operando lo stesso cambiamento di variabile di prima, come

$$E[x] = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + \mu) e^{-\frac{1}{2}z^2} dz = \frac{1}{\sqrt{2\pi}} (\sigma I_1 + \mu I_0) = \mu,$$

cosa che sapevamo già poiché la distribuzione è simmetrica rispetto al punto  $x = \mu$ . (Di più: questo implica anche che la mediana e la moda coincidono e valgono anch'esse  $\mu$ .) La varianza sarà, come anticipato

$$E[(x-\mu)^2] = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 z^2 e^{-\frac{1}{2}z^2} dz = \frac{1}{\sqrt{2\pi}} \sigma^2 I_2 = \sigma^2.$$

A questo punto la (188) ci permette di calcolare banalmente tutti i momenti centrali di ordine superiore, ma dal nostro punto di vista questo non è terribilmente interessante in quanto il coefficiente di asimmetria  $\gamma_1$  si annulla per ragioni di simmetria ed i momenti centrali di ordine pari non ci dicono molto altro sulla distribuzione.

È invece interessante, come abbiamo fatto con tutte le distribuzioni incontrate sino ad ora, calcolare il coefficiente di proporzionalità tra la semilarghezza a metà altezza e la deviazione standard—il che si fa risolvendo l'equazione

$$\frac{1}{2} \frac{1}{\sigma\sqrt{2\pi}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{da cui} \quad \frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2 = \ln 2 \quad \text{e} \quad x = \mu \pm \sqrt{2 \ln 2} \sigma.$$

Il valore della semilarghezza a metà altezza è dunque

$$\text{HWHM} = \sqrt{2 \ln 2} \sigma \approx 1.178\sigma. \quad (189)$$

### 5.7.3 L'integrale normale degli errori

Torniamo al problema dell'integrazione della distribuzione di Gauss. Nella sezione precedente abbiamo visto come calcolare un certo numero di integrali definiti (su tutta la retta reale), ma rimane il fatto che non esiste un'espressione analitica per la primitiva della densità di probabilità. Come facciamo, allora, a calcolare la probabilità che una generica variabile normale  $x$  sia contenuta in un dato intervallo finito? Il modo più naturale è passare attraverso la funzione cumulativa  $F(x)$  che abbiamo definito nel caso generale nella sezione 3.10:

$$P(x_1 \leq x \leq x_2) = F(x_2) - F(x_1).$$

Ora, data una variabile Gaussiana arbitraria, il calcolo del valore della funzione cumulativa in un punto generico si può sempre ricondurre all'integrale di una distribuzione di Gauss in forma standard mediante il cambiamento di variabile (187)

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz, \quad (190)$$

per cui in effetti l'unica cosa di cui abbiamo bisogno è una forma tabulata della funzione cumulativa per una variabile Gaussiana in forma standard—il cui grafico è mostrato per completezza in figura 5.10.

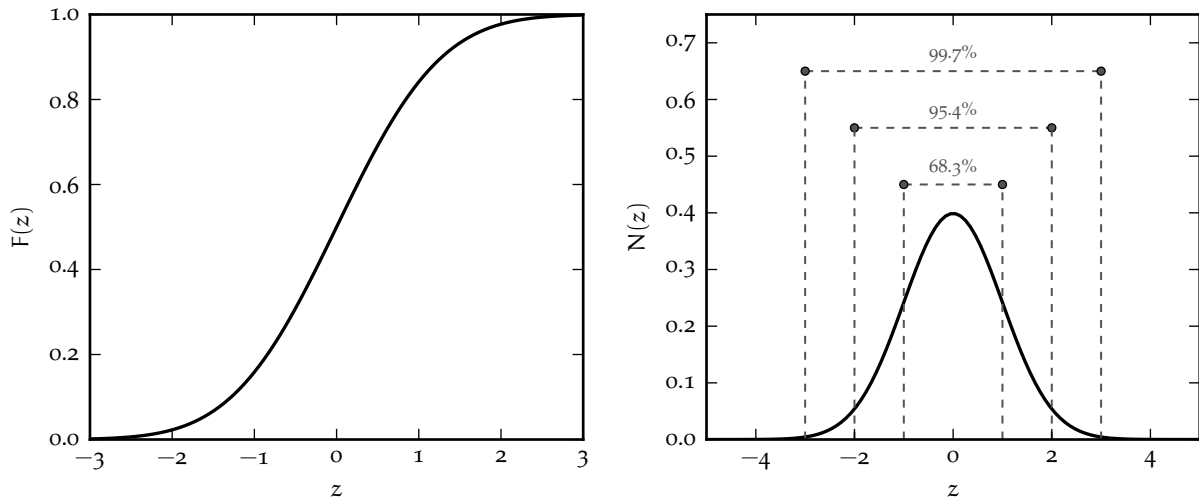


FIGURA 5.10. Grafico della funzione cumulativa per una variabile Gaussiana  $z$  in forma standard (cioè con media nulla e varianza unitaria) ed illustrazione della regola 68-95-99.7. La funzione cumulativa per una distribuzione di Gauss arbitraria si può ricondurre a questa tramite il cambiamento di variabile (187).

Anticipiamo che l'integrale di una distribuzione di Gauss entro 1, 2 e 3 deviazione standard dalla media vale rispettivamente

$$\begin{cases} P(\mu - 1\sigma \leq x \leq \mu + 1\sigma) = F(\mu + 1\sigma) - F(\mu - 1\sigma) \approx 0.6827 \\ P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = F(\mu + 2\sigma) - F(\mu - 2\sigma) \approx 0.9545 \\ P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) = F(\mu + 3\sigma) - F(\mu - 3\sigma) \approx 0.9973, \end{cases} \quad (191)$$

il che va talvolta sotto il nome di regola 68-95-99.7 (il 68% è un numero famoso ed indica appunto la probabilità che il valore di una variabile casuale Gaussiana disti meno di una deviazione standard dalla media.)

Sfortunatamente, per motivi storici, l'integrale che si trova più frequentemente tabulato (ed implementato numericamente nella maggior parte dei linguaggi di programmazione e dei programmi di analisi dati) non è la funzione cumulativa mostrata in figura 5.10, ma la *error function*, o *funzione degli errori* definita come

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \quad (192)$$

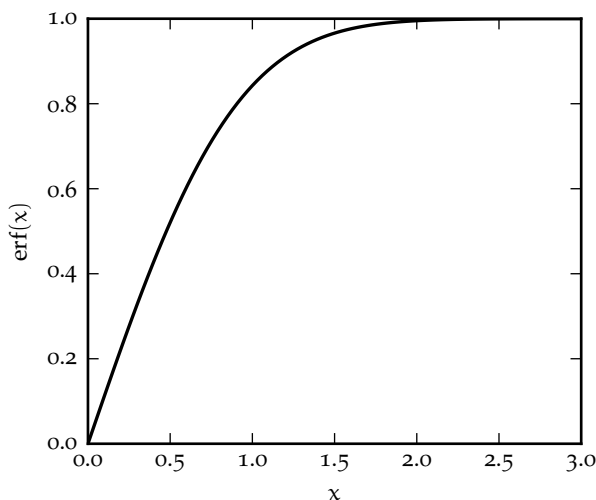


FIGURA 5.11. Grafico della *error function*  $\operatorname{erf}(x)$  definita dalla (192). La funzione  $\operatorname{erf}(x)$  si è diffusamente tabulata e la maggior parte dei linguaggi di programmazione e dei programmi di analisi dati ne offrono una implementazione numerica. La libreria standard di Python, ad esempio, la include nel modulo `math` come `math.erf`, ed il pacchetto SciPy ne offre una versione più sofisticata, che può essere applicata direttamente ad *array*, come `scipy.special.erf`. Alcune quantità legate alla *erf* sono tabulate nelle appendici A.1–A.5.

e mostrata in figura 5.11.

È chiaro che le due cose sono legate tra di loro, ma non coincidono esattamente. Fisicamente (e anche curiosamente, diremmo) la funzione degli errori rappresenta la probabilità che una variabile casuale con media 0 e varianza  $1/2$  cada nell'intervallo  $[-x, x]$ . Confrontando la (190) e la (192) si ottiene immediatamente

$$F(x) = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz + \int_0^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right), \quad (193)$$

che permette di passare dalla *error function* alla funzione cumulativa della gaussiana. Per una variabile Gaussiana in forma standard  $z$  la funzione cumulativa è così importante da meritare un nome tutto per sé (solitamente  $\Phi$ ) e la relazione precedente diviene

$$\Phi(z) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right). \quad (194)$$

<https://bitbucket.org/.../erf.py>

```

1 import math
2
3 def F(z):
4     """ Gaussian cumulative function.
5     """
6     return 0.5 + 0.5*math.erf(z/math.sqrt(2.))
7
8 # Integral of a gaussian within +/- 1, 2, and 3 sigma.
9 for z in [1, 2, 3]:
10     print(F(z) - F(-z))
11
12 [Output]
13 0.682689492137
14 0.954499736104
15 0.997300203937

```

FRAMMENTO 5.1. Frammento di codice per il calcolo della funzione cumulativa di una distribuzione di Gauss in forma standard, a partire dalla funzione erf del modulo math di Python. (Per completezza, il modulo scipy.special di SciPy offre un'implementazione alternativa della funzione degli errori che può operare direttamente su *array*.) Si confrontino i valori forniti in *output* dal programma con quelli riportati nella (191), oppure nella tabella in appendice A.4.

Il frammento di codice 5.1 illustra in modo sintetico una possibile applicazione pratica di quanto visto in questa sezione—il calcolo dei valori di probabilità corrispondenti alla regola 68-95-99.7 nella (191). Le appendici A.1–A.5 contengono i valori tabulati di varie quantità legate alla funzione cumulativa di una distribuzione di Gauss in forma standard, e possono essere utili quando non si ha a disposizione un calcolatore.

► ESEMPIO 5.11. Supponiamo di avere una variabile casuale  $x$  distribuita Gaussianamente con media  $\mu = 20$  e deviazione standard  $\sigma = 4$ . Qual è la probabilità che  $x$  sia compresa tra  $x_1 = 22$  e  $x_2 = 24$ ? La prima cosa da fare è trasformare gli estremi di integrazione nei valori corrispondenti per una variabile standard

$$z_1 = \frac{x_1 - \mu}{\sigma} = 0.5 \quad \text{e} \quad z_2 = \frac{x_2 - \mu}{\sigma} = 1.$$

A questo punto la probabilità cercata si scrive in termini della funzione cumulativa  $\Phi$  di una gaussiana in forma standard come

$$P(x_1 \leq x \leq x_2) = \Phi(1) - \Phi(0.5) = 0.8413 - 0.6915 = 0.1498.$$

(I valori si possono leggere sulle tavole in appendice A.2 o calcolare con un semplice programma come quello mostrato nel frammento 5.1.)



### 5.7.4 Alcuni commenti sul passaggio al continuo

Nella sezione 5.7.1 abbiamo derivato la forma analitica della distribuzione di Gauss come limite di una distribuzione di Poisson per  $\mu \rightarrow \infty$ , ma non abbiamo discusso in dettaglio le implicazioni del fatto che la distribuzione di Gauss è una distribuzione di variabile continua, mentre quella di Poisson una distribuzione di variabile discreta.

Torniamo per un attimo ad osservare con attenzione la figura 5.8. La prima cosa interessante da notare è che la bontà dell'approssimazione Gaussiana della distribuzione di Poisson migliora, per lo meno per quel che possiamo dire a livello qualitativo, al crescere della media; questo ce lo aspettavamo e non dovrebbe sorprenderci. L'altra, più sottile ma non meno importante, è che la *densità* delle barre che nel grafico rappresentano la distribuzione di Poisson tende pure ad aumentare al crescere della media. Ora, sappiamo che in generale, per lo meno per le distribuzioni unimodali che abbiamo incontrato fino a questo momento, i valori della probabilità (o della densità di probabilità nel caso continuo) sono significativamente diversi da zero solo entro alcune deviazioni standard dalla media. Per la distribuzione di Poisson, dato che  $\sigma = \sqrt{\mu}$ , questo significa che il numero di valori di  $k$  per cui la probabilità è apprezzabile cresce come  $\sqrt{\mu}$  al crescere di  $\mu$ , e che la differenza tra i valori di  $P(k)$  tra  $k$  contigui diviene via via più piccola. Nel nostro caso specifico questo significa che l'approssimazione Gaussiana può essere utilizzata non solo per stimare puntualmente il valore di  $P(k; \mu)$ , ma anche per calcolare la probabilità che una variabile Poissoniana sia compresa in un intervallo fissato, attraverso la relazione

$$P(k; \mu) \approx \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} N(x; \mu, \sqrt{\mu}) dx \quad \text{ovvero} \quad P(k_1 \leq k \leq k_2) = \sum_{k=k_1}^{k_2} P(k; \mu) \approx \int_{k_1-\frac{1}{2}}^{k_2+\frac{1}{2}} N(x; \mu, \sqrt{\mu}) dx. \quad (195)$$

Questa relazione può essere molto utile in pratica, perché permette di trasformare una somma (potenzialmente con un gran numero di addendi) in un integrale che è molto più agevole valutare numericamente.

► **ESEMPIO 5.12.** Sia data una variabile casuale  $k$  distribuita di Poissonianamente con media  $\mu = 1000$ . Qual è la probabilità  $P(950 \leq k \leq 1022)$ ? Il modo formalmente corretto di risolvere il problema è quello di calcolare

$$P(950 \leq k \leq 1022) = \sum_{k=950}^{1022} P(k; 1000) \approx 0.70822.$$

(Sono 73 termini. Senza un calcolatore sotto mano può essere tedioso.) L'alternativa è utilizzare l'approssimazione Gaussiana, trasformare gli estremi nei corrispondenti valori di una variabile in forma standard

$$z_1 = \frac{(950 - 1/2 - 1000)}{\sqrt{1000}} \approx -1.59695 \quad \text{e} \quad z_2 = \frac{(1022 + 1/2 - 1000)}{\sqrt{1000}} \approx 0.71151$$

e procedere utilizzando le tabelle della funzione cumulativa  $\Phi(z)$  o della *error function*

$$P(z_1 \leq z \leq z_2) = \Phi(z_2) - \Phi(z_1) \approx 0.70648.$$

La precisione dell'approssimazione Gaussiana è in questo caso al livello di 2 parti su 1000.

### 5.7.5 Una derivazione alternativa della distribuzione di Gauss

Prima di passare oltre ci soffermiamo un attimo su una derivazione alternativa della forma analitica della distribuzione di Gauss, particolarmente significativa per il suo carattere essenzialmente geometrico, dovuta ad Herschel [9].

Supponiamo di essere interessati alla distribuzione degli errori nella misure della posizione di una stella ed immaginiamo di fissare un riferimento cartesiano  $(x, y)$  centrato sulla posizione *vera* della stella

stessa. L'argomentazione di Herschel si basa su due semplici postulati che sembrano essere conseguenze banali dell'omogeneità dello spazio. Il primo è che gli errori su  $x$  e su  $y$  siano indipendenti e seguano la stessa funzione di distribuzione, cioè che la densità di probabilità della posizione *misurata* nello spazio bidimensionale si possa fattorizzare come

$$p(x, y) dx dy = \varphi(x)\varphi(y) dx dy. \quad (196)$$

Possiamo allora scrivere la stessa densità di probabilità in coordinate polari nel piano (ricordate che l'elemento infinitesimo di superficie  $dx dy$  diviene  $r dr d\phi$ )

$$p(x, y) dx dy = \psi(r, \phi) r dr d\phi$$

e, a questo punto, richiediamo il secondo postulato, ovverosia che la funzione di distribuzione sia indipendente dall'angolo  $\phi$

$$\varphi(x)\varphi(y) = \psi\left(\sqrt{x^2 + y^2}\right). \quad (197)$$

Ponendo  $y = 0$  nell'equazione precedente otteniamo  $\varphi(x)\varphi(0) = \psi(x)$ , che ci permette di eliminare  $\psi$  dal tavolo. Più precisamente possiamo riscrivere la (197) come

$$\varphi(x)\varphi(y) = \varphi\left(\sqrt{x^2 + y^2}\right) \varphi(0),$$

da cui, dividendo entrambi i membri per il quadrato di  $\varphi(0)$  e prendendo i logaritmi del risultato si ha

$$\ln\left(\frac{\varphi(x)}{\varphi(0)}\right) + \ln\left(\frac{\varphi(y)}{\varphi(0)}\right) = \ln\left(\frac{\varphi\left(\sqrt{x^2 + y^2}\right)}{\varphi(0)}\right). \quad (198)$$

Ora, è chiaro che questa relazione è soddisfatta (provate per sostituzione diretta) se e solo se

$$\ln\left(\frac{\varphi(x)}{\varphi(0)}\right) = cx^2,$$

ovvero

$$\varphi(x) = \varphi(0) e^{cx^2}, \quad (199)$$

ma perché la funzione di distribuzione sia normalizzabile  $c$  deve essere negativo, per cui di fatto l'espressione che abbiamo ricavato non è altro che una distribuzione di Gauss con media nulla.

Questa derivazione è particolarmente significativa per l'economia concettuale delle premesse: due condizioni puramente geometriche, in generale incompatibili tra di loro, diventano compatibili per una forma funzionale ben definita della distribuzione di probabilità di partenza—quella di Gauss.

## 5.8 IL TEOREMA CENTRALE DEL LIMITE

Abbiamo visto che la distribuzione di Gauss è il limite per  $\mu \rightarrow \infty$  della distribuzione di Poisson—e, di conseguenza, anche della distribuzione binomiale. Vedremo nel proseguo che altre distribuzioni (e.g., quella del  $\chi^2$  e la  $t$  di Student) tendono ad una Gaussiana nei limiti opportuni. Adesso è il momento di enunciare uno dei risultati più importanti del calcolo delle probabilità—il teorema centrale del limite.

**TEOREMA 5.1 (CENTRALE DEL LIMITE).** Siano date  $n$  variabili casuali indipendenti  $x_i$ , ciascuna con media  $\mu_i$  e varianza  $\sigma_i^2$  finite (non necessariamente uguali). Indipendentemente dalla forma particolare delle funzioni di distribuzione delle singole  $x_i$ , la somma  $S = \sum_{i=1}^n x_i$  è asintoticamente normale nel limite  $n \rightarrow \infty$ —nel senso che per  $n$  grande la funzione di distribuzione di  $S$  tende ad una distribuzione di Gauss come media e varianza date da

$$E[S] = \sum_{i=1}^n \mu_i \quad \text{e} \quad \text{Var}(S) = \sum_{i=1}^n \sigma_i^2.$$

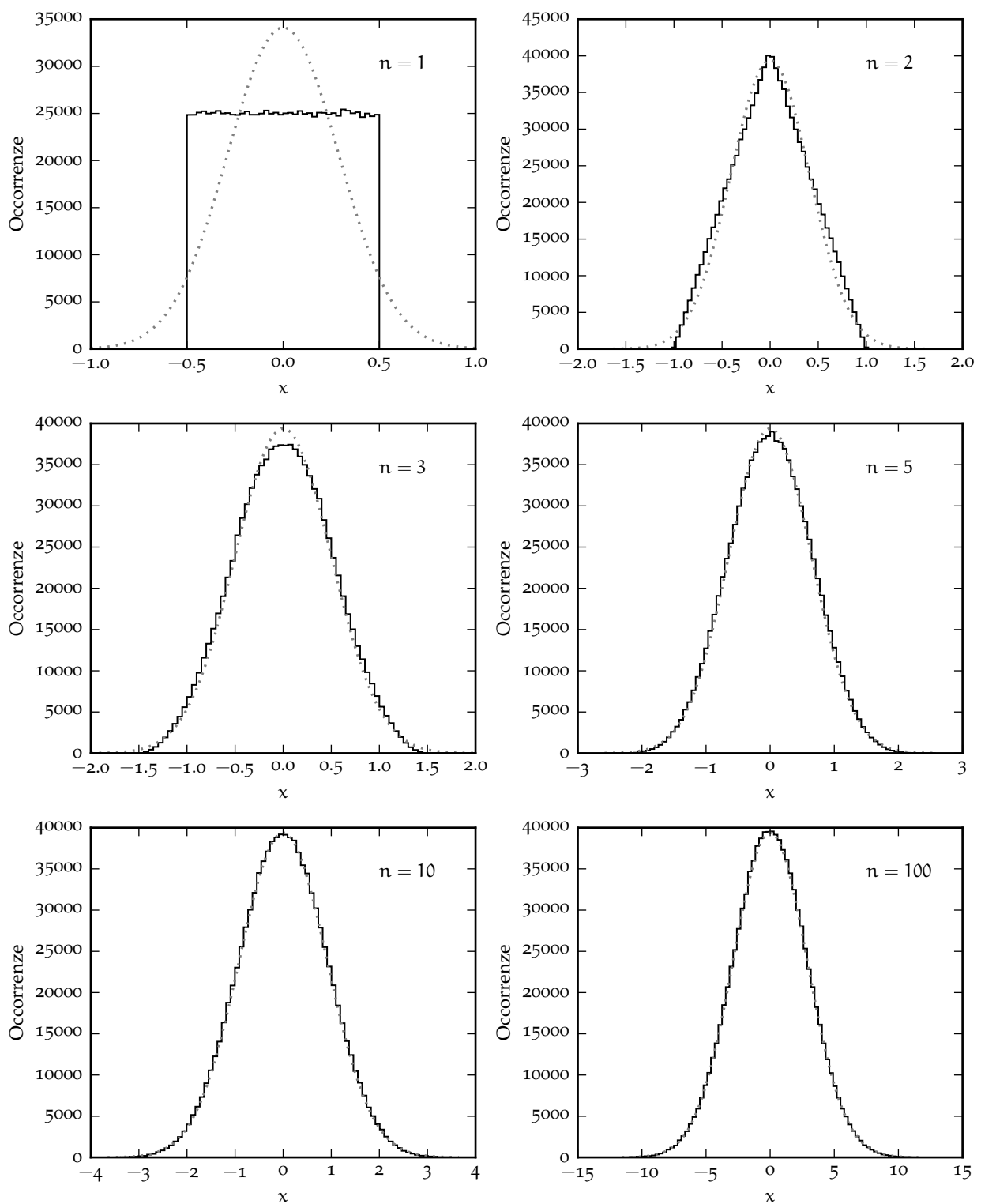


FIGURA 5.12. Illustrazione del teorema centrale del limite nel caso della somma di  $n$  variabili casuali distribuite uniformemente tra  $-1/2$  ed  $1/2$ . In ciascun grafico gli istogrammi si riferiscono ad un milione di campionamenti indipendenti della somma in questione, mentre linea grigia tratteggiata rappresenta una distribuzione di Gauss con media 0 e varianza  $n/12$ . La convergenza è relativamente veloce e per  $n = 10$  le due sono virtualmente indistinguibili.

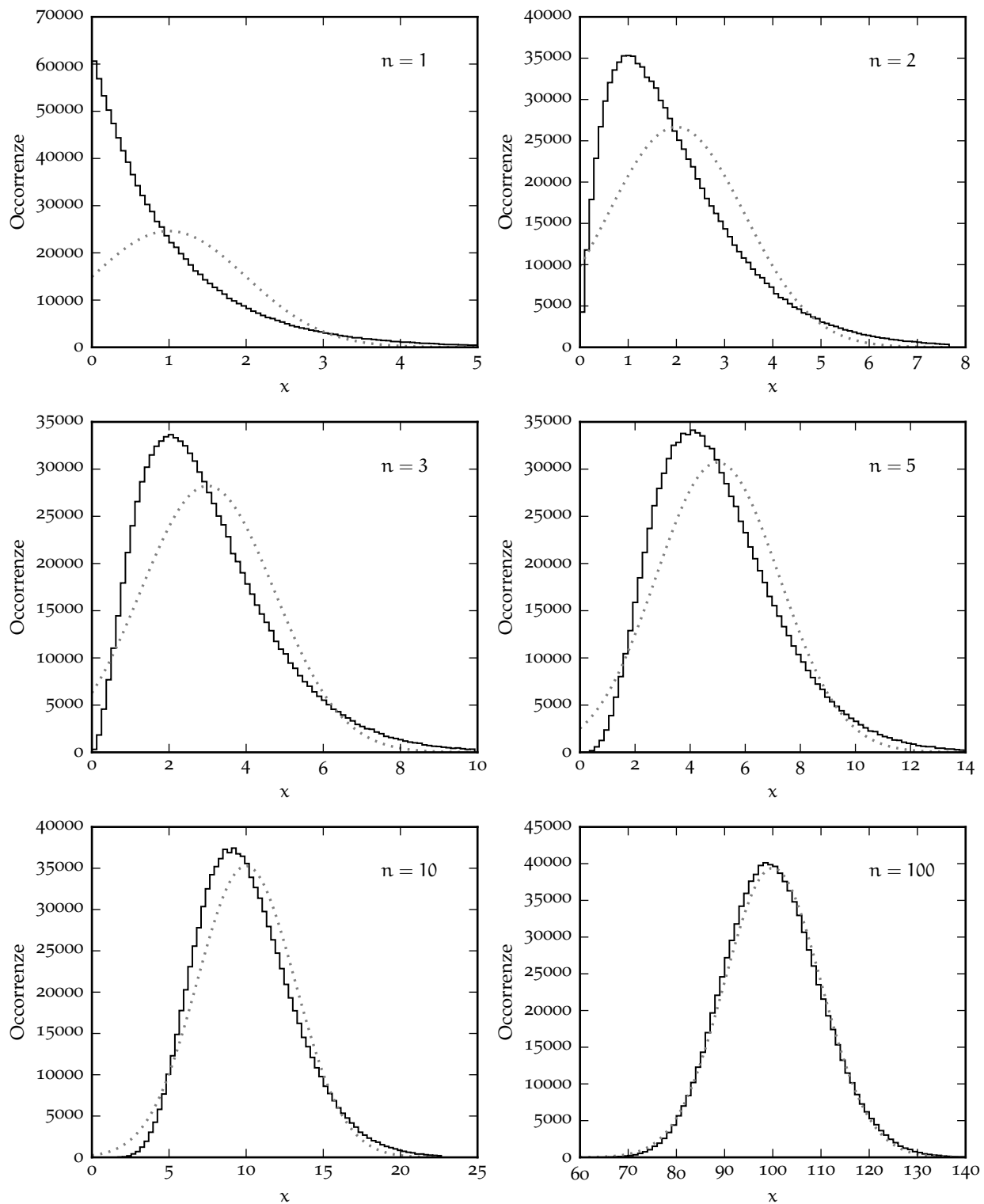


FIGURA 5.13. Illustrazione del teorema centrale del limite nel caso della somma di  $n$  variabili casuali distribuite esponenzialmente con parametro  $\lambda = 1$ . In ciascun grafico gli istogrammi si riferiscono ad un milione di campionamenti indipendenti della somma in questione, mentre linea grigia tratteggiata rappresenta una distribuzione di Gauss con media  $n$  e varianza  $n$ . La convergenza è molto più lenta che non nel caso della somma di variabili uniformi.

Ci limitiamo ad enunciare il teorema senza dimostrarlo. La legge di somma per i valor medi e le varianze la conosceamo già (vedi sezione 4.2.1); il risultato nuovo—ed è un risultato di un'importanza che non può essere sopravvalutata—è che se sommiamo abbastanza variabili la distribuzione risultante è approssimativamente Gaussiana indipendentemente da come sono distribuite le variabili di partenza.

Se ci pensiamo per un attimo, in realtà, di conseguenze del teorema centrale del limite ne abbiamo già incontrate alcune lungo la strada. Il fatto che la somma di un numero arbitrario di variabili binomiali (Poissoniane) sia distribuita ancora come una binomiale (Poissoniana), ci dice immediatamente, per il teorema centrale del limite, che sia la distribuzione binomiale che quella di Poisson debbono tendere ad una Gaussiana nel limite di media infinita.

### 5.8.1 Il teorema centrale del limite e la media campione

Il teorema centrale del limite ci garantisce anche che la media (che, a parte un fattore costante di normalizzazione  $1/n$  non è nient'altro che una somma) di un numero abbastanza grande di variabili casuali è distribuita approssimativamente come una Gaussiana. Questo implica che, quando scriviamo il risultato di una misura utilizzando la media e la deviazione standard della media, come nella sezione 4.3.2, di fatto il livello di confidenza è fissato al 68%

$$x = m \pm s_m \text{ [unità di misura]}. \quad (68\% \text{ CL}) \quad (200)$$

(a patto ovviamente che il numero di misure nel campione sia abbastanza grande).

## 5.9 IN BREVE...

Distribuzione	Espressione	Media	Varianza	Skewness
Binomiale	$\mathcal{B}(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$	$np$	$np(1-p)$	$\frac{1-2p}{\sqrt{np(1-p)}}$
Poissoniana	$\mathcal{P}(k; \mu) = \frac{\mu^k}{k!} e^{-\mu}$	$\mu$	$\mu$	$\frac{1}{\sqrt{\mu}}$
Uniforme	$u(x; a, b) = \begin{cases} \frac{1}{(b-a)} & a \leq x \leq b \\ 0 & x < a; x > b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	0
Esponenziale	$\varepsilon(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & 0 \leq x \leq \infty \\ 0 & x < 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	2
Gaussiana	$N(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$\mu$	$\sigma^2$	0
$\chi^2$	$\varrho(x; \nu) = \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} x^{\frac{\nu-2}{2}} e^{-\frac{x}{2}}$	$\nu$	$2\nu$	$\sqrt{\frac{8}{\nu}}$

TABELLA 16. Caratteristiche principali di alcune delle distribuzioni univariate di uso comune.

- In una situazione con sue soli esiti possibili, la distribuzione binomiale descrive la probabilità di avere  $k$  successi in  $n$  tentativi, data la probabilità  $p$  di successo nell'evento elementare.
- La distribuzione di Poisson è, formalmente, il limite della binomiale per  $p \rightarrow 0$  e  $n \rightarrow \infty$ , con  $\mu = np$  costante. Essa descrive, cosa più importante, una larga classe di fenomeni—i fenomeni stazionari, e permette di rispondere alla domanda: “qual è la probabilità che si verifichino  $k$  eventi quando in media se ne verificano  $\mu$ ?”

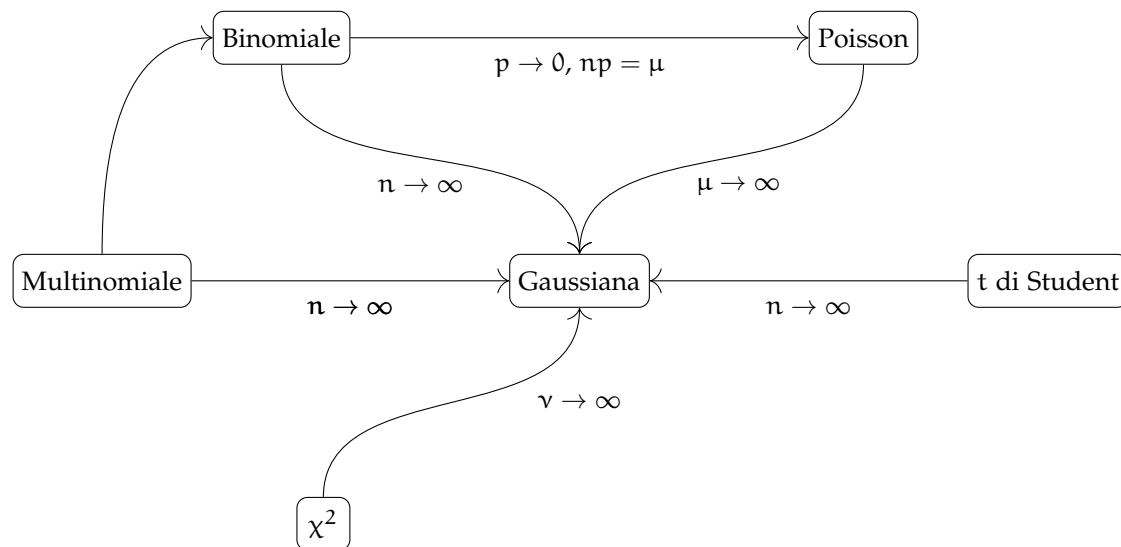


FIGURA 5.14. Rappresentazione grafica delle relazioni asintotiche tra le varie distribuzioni e dei limiti dei parametri in corrispondenza dei quali queste relazioni sono valide.

- La distribuzione costante è un buon modello per gli errori quando lo strumento di misura è digitale. (Tenete a mente il numero  $\sqrt{12}$  perché lo ri-incontrerete nella vostra carriera di Fisici.)
- La distribuzione esponenziale, tra le altre cose, descrive la distribuzione delle differenze di tempi tra eventi successivi in un processo stazionario (cioè quando la statistica dei conteggi è Poissoniana). Essa ha anche l'interessante proprietà dell'assenza di memoria.
- La distribuzione di Gauss può essere ricavata come limite della distribuzione di Poisson per  $\mu \rightarrow \infty$ , ma è in realtà un *attrattore* (cioè un limite sotto condizioni opportune) per molte distribuzioni, come mostrato in figura 5.14. L'integrale indefinito della Gaussiana non ha espressione analitica, per cui richiede di essere tabulato (vedi appendici A.1–A.5).
- La somma (e di conseguenza la media) di un numero abbastanza grande di variabili casuali tende, sotto ipotesi molto deboli, ad assumere una distribuzione Gaussiana indipendentemente dalla forma delle distribuzioni di partenza (teorema centrale del limite). Questo fissa il livello di confidenza (al 68%) quando scriviamo il risultato di una misura come  $m \pm s_m$ .

## INTRODUZIONE AI METODI MONTE CARLO

Avrete sentito tutti in vita vostra la parola *simulazione*—le simulazioni sono una parte rilevante del lavoro del fisico, ma cosa vuol dire esattamente eseguire una simulazione di un sistema fisico su un calcolatore? In questo capitolo cercheremo di rispondere, sia pur parzialmente, a questa domanda e di sviluppare i concetti di base rilevanti.

## 6.1 INTRODUZIONE: DUE SEMPLICI ESEMPI

Cominciamo con il primo esempio. Supponiamo di voler stimare numericamente il valore di  $\pi$  con un numero fissato di cifre significative. Storicamente i primi algoritmi per la stima di  $\pi$  erano di tipo puramente geometrico, basati sul calcolo del perimetro di poligoni, iscritti e circoscritti ad un cerchio di raggio unitario, con un numero sempre maggiore di lati—si deve ad Archimede la dimostrazione che  $223/71 \leq \pi \leq 22/7$ , completata utilizzando due poligoni regolari di 96 lati. Con la nascita del calcolo infinitesimale, il XVII e XVIII secolo hanno visto un fiorire di serie e prodotti infiniti che convergono a  $\pi$ , e che possono essere utilizzati in modo efficiente per calcolarne il valore con una precisione in linea di principio arbitraria. Noi utilizzeremo un terzo tipo di approccio, che è illustrato in figura 6.1 e che è interessante perché ha un equivalente *meccanico* che si può realizzare con un esperimento reale.

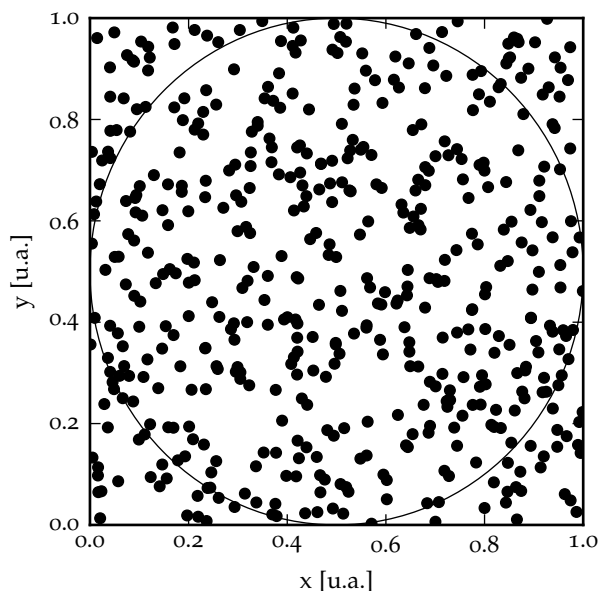


FIGURA 6.1. Rappresentazione grafica di  $N = 500$  punti casuali generati uniformemente entro un quadrato di lato unitario. In questa particolare realizzazione i punti che finiscono all'interno del cerchio inscritto sono esattamente  $n_c = 396$ , il che dà una stima di  $\pi \approx 4 \times 396/500 = 3.168$ . La precisione della nostra stima non è sconvolgente—essa è corretta fino alla prima cifra decimale, con un errore relativo di circa 8 parti su 1000—ma può essere incrementata a piacere (almeno in linea di principio) aumentando il numero di punti.

Consideriamo dunque un quadrato di lato  $\ell$  (che avrà area  $A_Q = \ell^2$ ) ed il cerchio in esso inscritto—che avrà area  $A_C = \pi\ell^2/4$ . Si ha ovviamente

$$\pi = \frac{4A_C}{A_Q},$$

che, data una stima di  $A_C$  ed  $A_Q$ , ci permette di stimare il valore di  $\pi$ . Procediamo dunque a disegnare il nostro quadrato sul pavimento e lanciamo in aria una manciata di coriandoli—facendo in modo che essi

lo ricoprono in modo uniforme. Il numero di coriandoli  $n_Q$  ( $n_C$ ) che cadono entro il quadrato (il cerchio) sarà in media proporzionale alla sua area  $A_Q$  ( $A_C$ ), per cui potremo scrivere

$$\pi \approx \frac{4n_C}{n_Q}.$$

L'approssimazione sarà tanto migliore quanto più alto è il numero di coriandoli che lanciamo e quanto più uniforme è la loro distribuzione.

Fermiamoci per un attimo a riflettere. Non sarebbe bello se potessimo fare tutto questo utilizzando un calcolatore, anziché imbrattare il pavimento? Non sarebbe, cioè, bello se il calcolatore permettesse di generare *numeri casuali* (o *random*, come si dice spesso)? Nel nostro caso un generatore di numeri casuali distribuiti uniformemente tra 0 ed 1 sarebbe tutto ciò di cui avremmo bisogno: con due estrazioni successive avremmo un punto  $(x, y)$  distribuito uniformemente nel quadrato unitario e procedendo così potremmo stimare  $\pi$  con precisione arbitraria, come mostrato in figura 6.1. La buona notizia è che tutto questo è possibile e qualsiasi linguaggio di programmazione fornisce strumenti per la generazione di numeri casuali—o meglio, come vedremo tra un attimo, pseudocasuali. Il frammento di codice 6.1 costituisce un'implementazione funzionante e riutilizzabile (sebbene non particolarmente efficiente) dell'algoritmo per la stima di  $\pi$  illustrato in figura 6.1.

```

1  import random
2  # Fix the seed to get always the same answer.
3  random.seed(1)
4
5  # Initialize the counters.
6  n_q = 1000000
7  n_c = 0
8  # Extract a number of random points...
9  for i in range(n_q):
10     x = random.uniform(0., 1.)
11     y = random.uniform(0., 1.)
12     # ...and check if they lie in the circle.
13     if (x*x + y*y <= 1.):
14         n_c += 1
15 # Estimate pi and print it.
16 pi = 4.*n_c/float(n_q)
17 print(pi)
18
19 [Output]
20 3.14138

```

FRAMMENTO 6.1. Codice per la stima di  $\pi$  utilizzando il generatore di numeri pseudocasuali della libreria *random* di Python, secondo l'algoritmo illustrato nel testo e nella figura 6.1. Sebbene la convergenza di questo algoritmo non sia stratosfericamente veloce, con 1000000 di punti si raggiunge un errore tipico di una parte su 1000 o meglio.

L'esempio della stima di  $\pi$  che abbiamo appena visto è importante e rappresentativo perché illustra l'idea che sta alla base dell'utilizzo di sequenze di numeri casuali come metodo per il calcolo di integrali—non dovrebbe essere troppo difficile, ad esempio, generalizzare il frammento 6.1 per calcolare numericamente l'integrale definito di una generica funzione di una variabile reale. Ma c'è un altro aspetto del problema, che è diverso ma altrettanto importante per il mestiere del Fisico, e che illustriamo con un altro esempio.

La macchina di Galton è un dispositivo ideato per semplici esperimenti di statistica, che permette tra le altre cose di illustrare in modo immediato la distribuzione binomiale ed il teorema centrale del limite—argomenti che tratteremo in dettaglio nei prossimi capitoli. Si tratta essenzialmente di una tavola su cui sono piantati dei chiodi in una matrice regolare come quella mostrata in figura 6.2. Quando la tavola è posta su un piano verticale, una pallina delle dimensioni opportune, lasciata cadere dalla sommità, può attraversarla fino al punto più basso urtando sui chiodi e deviando a destra o a sinistra (tipicamente con probabilità uguali) ad ogni livello. La posizione orizzontale di arrivo nella pallina tende ad assumere una caratteristica forma a campana quando il numero di livelli della macchina è abbastanza grande.

A questo punto vi starete chiedendo cosa c'entra tutto questo con i calcolatori e le sequenze di numeri casuali. Beh... si tratta di un altro esempio di un sistema che può essere simulato e studiato nel dettaglio



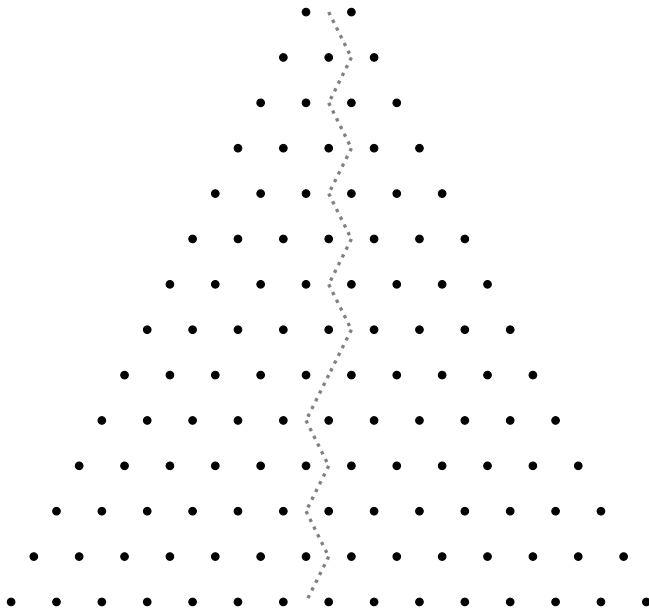


FIGURA 6.2. Rappresentazione schematica di una macchina di Galton con 14 livelli. La linea spezzata rappresenta una delle possibili traiettorie di una pallina lasciata cadere dalla sommità, sotto l'ipotesi che ad ogni livello la probabilità di deviare verso destra e quella di deviare verso sinistra siano entrambe  $1/2$ .

con un calcolatore attraverso l'ausilio di un generatore di numeri casuali. E, a pensarci meglio, la simulazione si può fare utilizzando lo stesso generatore casuale di numeri distribuiti uniformemente tra 0 ed 1 che abbiamo utilizzato prima: ad ogni livello possiamo estrarre un numero  $r$  tra 0 e 1 e dire che se  $r \leq 1/2$  ci spostiamo verso sinistra, mentre se  $r > 1/2$ , allora ci spostiamo a destra. La cosa interessante è che, nella nostra simulazione, possiamo seguire la pallina dall'inizio alla fine, proprio come faremmo nella realtà. Possiamo registrare, per ogni realizzazione del nostro esperimento, non solo il punto di arrivo, ma anche tutte le posizioni intermedie, che forniscono un quadro dettagliato del nostro *evento* elementare. In altre parole, ad una qualsiasi domanda che possiamo pensare di porci a proposito di una macchina di Galton reale, possiamo dare una risposta con un semplice programma al calcolatore. Il passo da questo semplice problema di carattere ricreativo alla modellizzazione del moto di una molecola in un gas o di una particella carica nella materia non è poi così lungo<sup>1</sup>.

[https://bitbucket.org/.../galton\\_box.py](https://bitbucket.org/.../galton_box.py)

```

1 import random
2 # Fix the seed to get always the same answer.
3 random.seed(1)
4
5 # Setup the simulation.
6 pos = 0.
7 step = 1.
8 path = [pos]
9 # Loop through the Galton box.
10 for i in range(10):
11     r = random.uniform(0., 1.)
12     if r <= 0.5:
13         pos -= step
14     else:
15         pos += step
16     path.append(pos)
17 # Print the actual path.
18 print(path)
19
20 [Output]
21 [0.0, -1.0, 0.0, 1.0, 0.0, -1.0, -2.0, -1.0, 0.0, -1.0, -2.0]
```

FRAMMENTO 6.2. Codice per la simulazione di una semplice macchina di Galton (in questo caso a 10 livelli) che utilizza internamente un generatore casuale di numeri reali distribuiti uniformemente tra 0 ed 1.

<sup>1</sup> Se volete un esempio più vicino alla vita di tutti i giorni pensate al vostro videogioco preferito. Come pensate che faccia il calcolatore a generare la varietà imprevedibile di situazioni che lo rende interessante?

## 6.2 CHE COSA È UN “NUMERO CASUALE”?

Abbiamo visto attraverso due esempi concreti che un generatore di numeri casuali, se usato opportunamente, può essere uno strumento utile per la modellizzazione e lo studio di sistemi reali in cui le fluttuazioni casuali giocano un qualche ruolo. (E, a questo punto, dovrebbe essere chiaro anche il collegamento tra il “Monte Carlo” nel titolo di questo capitolo e l’argomento del capitolo stesso.) Adesso si tratta di fare un passo indietro e chiedersi cosa sia un numero casuale e come si possano generare numeri casuali utilizzando un calcolatore.

È un numero casuale il 3? Ed il 5? È ovvio che entrambe queste domande sono prive di senso. Un numero è solo un numero e, di per sé, non è né casuale né non-casuale. Quello di casualità (o *randomicità*, per usare un orrendo neologismo che va di moda di questi tempi) è un concetto non banale da definire con precisione, ma la prima cosa importante da dire è che esso si applica solo alle *sequenze* di numeri—e non ai numeri singoli. E allora: cosa possiamo dire della sequenza di cifre decimali

$$S_1 = 6, 9, 0, 7, 6, 2, 1, 4, 6, 2, 2, 3, 1, 0, 2, 8, 8, 8, 2, 0, 8, 6, 4, 8, 5, 9, 9, 2, 8, 3?$$

È sufficientemente casuale oppure no? E ancora: vi è qualche differenza di carattere fondamentale (e se sì quale) tra  $S_1$  e le due sequenze alternative

$$S_2 = 0, 0$$

$$S_3 = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9?$$

Non abbiamo ancora risposto alla domanda con cui abbiamo aperto questa sezione, ma le sequenze  $S_2$  ed  $S_3$  non hanno proprio l’apparenza di essere casuali.  $S_2$  contiene solo al cifra 0 e  $S_3$  è la mera ripetizione delle 10 cifre di partenza, in ordine crescente.  $S_1$ , invece, non sembra avere niente che non vada—ed è chiaro che ci serve qualche strumento più sofisticato per prendere una decisione a proposito del suo livello di casualità. Eppure, se ci pensiamo meglio, non dovremmo assumere che, in una situazione veramente casuale, tutte le sequenze di cifre decimali di lunghezza 30 dovrebbero essere in un qualche senso equiprobabili? Se scegliamo una sequenza a caso, non dovremmo avere la stessa probabilità di estrarre  $S_1$ ,  $S_2$  ed  $S_3$ ? E allora qual è la differenza fondamentale tra  $S_1$  da una parte e  $S_2$  ed  $S_3$  dall’altra? Cosa è che ci fa percepire la prima come potenzialmente casuale e le altre due come decisamente non casuali? Se un generatore di numeri *random* ci fornisse  $S_3$  come prima sequenza, saremmo tentati di mettere in dubbio la sua bontà? E, se sì, perché?

Tutte queste domande sono perfettamente lecite, e mettono in luce alcune proprietà controintuitive del concetto di *randomicità*—che sono anche quelle che rendono difficoltosa la sua definizione. La cosa che rende  $S_2$  ed  $S_3$  diverse da  $S_1$  ai nostri occhi, è il fatto che esse sono caratterizzate da una struttura ben definita (un *pattern*) che è facile da identificare a prima vista. Possiamo descrivere  $S_2$  come “una sequenza di 30 zeri” e  $S_3$  come “tre ripetizioni della sequenza 0, 1, 2, 3, 4, 5, 6, 7, 8, 9”, mentre è difficile trovare una descrizione altrettanto concisa di  $S_1$ . E allora: è vero che, in una situazione realmente casuale  $S_1$ ,  $S_2$  ed  $S_3$  sono equiprobabili, ma è anche vero che ci sono relativamente poche sequenze con *pattern* così chiari come in  $S_2$  ed  $S_3$ , immerse in un oceano di sequenze *anonime* come  $S_1$ . È proprio questo che rende in un qualche senso  $S_2$  ed  $S_3$  speciali ai nostri occhi.

Il concetto di *pattern* è fondamentale in questo contesto, ma dobbiamo stare attenti a non enfatizzarlo eccessivamente come unico criterio per decidere se una data sequenza ha buone proprietà di *randomicità* oppure no. L’occhio umano è estremamente efficace nell’identificare strutture e ripetizioni nelle cose, e lasciando libero sfogo a questa nostra capacità di corre il rischio di eccedere. Guardiamo per un attimo  $S_1$  con attenzione, ad esempio. Ci sono due coppie di cifre che si ripetono (2, 2 e 9, 9) e addirittura una tripletta (8, 8, 8)! Questo vuol dire che in fondo nemmeno  $S_1$  è una sequenza casuale? Beh, se ci pensiamo un attimo la probabilità di avere una coppia di cifre decimali identiche in successione, assumendo che esse siano scelte a caso, è  $1/10$  e la probabilità di avere una tripletta è  $1/100$ , per cui trovare due coppie ed una tripletta in una sequenza di 30 cifre non deve sorprenderci. Anzi: se avessimo una sequenza di 1000 cifre senza nemmeno una ripetizione potremmo dire con buona probabilità che essa non è casuale!

Abbiamo insistito abbastanza sul tema ed è tempo di passare oltre. Abbiamo capito che ci sono alcune proprietà che una sequenza di numeri deve possedere per poter essere definita casuale. La più

ovvia è l'equidistribuzione (cioè il fatto che le diverse cifre debbono apparire in media con la stessa frequenza), ma chiaramente essa non è sufficiente— $S_3$  è equidistribuita ma chiaramente non casuale. Corrispondentemente esistono svariati test specifici ideati per verificare la casualità di una sequenza, basati ad esempio sulla distribuzione delle ripetizioni di cifre (e.g., coppie o triplette) o sulla distanza tra due occorrenze consecutive della stessa cifra. Si rimanda il lettore a [11] per una trattazione approfondita della materia. Per quel che ci riguarda adotteremo un approccio puramente empirico e diremo che una sequenza è casuale se, indipendentemente da come essa è stata ottenuta, ha le buone proprietà cui abbiamo sommariamente accennato (o, equivalentemente, supera tutti i test rilevanti).

### 6.2.1 Generatori casuali e pseudo-casuali

Esistono due approcci fondamentalmente diversi per la generazione di sequenze casuali. Il primo, che è per molti versi il più naturale (ed è anche il primo in senso storico) si basa sull'idea di utilizzare fenomeni naturali intrinsecamente casuali—come ad esempio il rumore termico su una resistenza elettrica o il decadimento di un campione radioattivo. Un generatore di questo tipo si dice *true random number generator* (TRNG), ed il servizio offerto da <https://www.random.org/> è un esempio moderno interessante in cui l'elemento di casualità è il rumore atmosferico. Vale sicuramente il tempo di un'occhiata veloce.

Dall'altra parte esistono generatori *pseudo-casuali*—che in Inglese si dicono *pseudo-random number generator* o PRNG—che utilizzano algoritmi basati su formule matematiche oppure tabelle precompilate per produrre sequenze di numeri completamente predeterminate che si comportano come se fossero casuali. Questo può sembrare sorprendente a prima vista: come è possibile che una sequenza predeterminata sia casuale? Beh, dobbiamo semplicemente ricordare cosa abbiamo detto alla fine della sezione precedente, e cioè che giudichiamo una sequenza dalla sue proprietà e non da come è stata generata. Se una sequenza supera i test di randomicità, allora è casuale per definizione. Vedremo alcuni esempi concreti nella prossima sezione.

Se dovessimo confrontare sommariamente le proprietà generali dei generatori casuali e pseudo-casuali diremmo essenzialmente tre cose:

- I generatori pseudo-casuali sono deterministici, mentre quelli casuali sono non deterministici. Nel primo caso la sequenza è univocamente determinata dal valore iniziale—e dunque prevedibile e *riproducibile*. In pratica la riproducibilità è una proprietà utile perché, come vedremo, ci permette di realizzare simulazioni in condizioni controllate.
- I generatori pseudo-casuali sono periodici, mentre quelli casuali sono aperiodici. La periodicità non è un problema in pratica a patto che il periodo sia abbastanza lungo. (Per completezza: il generatore pseudo-casuale della libreria standard di Python ha un periodo  $2^{19937} - 1 > 10^{6000}$  e ci sono meno di  $10^{27}$  nanosecondi nell'età dell'Universo, per cui la possibilità di esaurire la sequenza è abbastanza remota.)
- I generatori pseudo-casuali possono essere molto efficienti (in termini di numero di cifre prodotte nell'unità di tempo), mentre quelli casuali tendono ad essere comparativamente più lenti.

Non sorprenderà allora se i generatori pseudo-casuali sono quelli di gran lunga più utilizzati per applicazioni di modellizzazione e simulazione. Ma se avete bisogno di generare un numero casuale per l'estrazione di una lotteria e volete assicurarvi che nessuno possa barare, allora è consigliabile utilizzare un vero generatore casuale!

## 6.3 DIGRESSIONE: L'ARITMETICA MODULARE

L'operazione *modulo* non è altro che il resto della divisione intera tra due numeri—una cosa semplicissima ed apparentemente innocua. Ad esempio

$$29 \bmod 7 = 1 \quad \text{poiché} \quad 29 = 7 \times 4 + 1.$$

Eppure l'aritmetica modulare è materia estremamente interessante, e rilevante (tra le altre cose) per l'argomento di questo capitolo. Vediamo un esempio concreto: fissati due numeri interi positivi  $m$  e  $p$ , con  $p < m$ , e per qualsiasi numero intero  $k$  sappiamo calcolare banalmente la quantità

$$0 \leq g = k^p \bmod m \leq (m-1)$$

Ora, se proviamo a farlo, come mostrato nel frammento 6.3, vediamo subito che i risultati sono potenzialmente interessanti. Mentre  $k$  assume tutti i valori da 0 a  $m-1$  (in quest'ordine), la lista dei risultati esaurisce la lista degli  $m$  valori di  $k$ , ma in un ordine che pare aver poco a che vedere con quello di partenza. In un qualche senso abbiamo riordinato la lista dei primi  $m-1$  interi positivi in modo che appare casuale. Quello che accade è abbastanza chiaro: al crescere di  $k$  si tende ad avere  $k^p \gg m$  ed il risultato del modulo tende ad essere una funzione discontinua di  $k$ . L'equivalente meccanico del processo è, per certi aspetti la lancetta di un orologio che noi facciamo avanzare a scatti sempre più grandi. Quando uno scatto corrisponde a molti giri completi, allora la posizione in cui essa si ferma diviene essenzialmente indeterminata—un po' come in certi giochi da tavolo (avete presente il gioco della bottiglia?).

```

1  # https://bitbucket.org/.../modulo.py
2  # Define m and p, and initialize an empty list.
3  m = 11
4  p = 7
5  sequence = []
6  # Calculate k^p mod m for all positive k < m.
7  # Note ** is the power and % the modulo operator.
8  for k in range(m):
9      sequence.append(k**p % m)
10
11 print(sequence)
12
13 [Output]
14 [0, 1, 7, 9, 5, 3, 8, 6, 2, 4, 10]
```

FRAMMENTO 6.3. Esempio di calcolo di  $k^p \bmod m$  per  $m = 11$ ,  $p = 7$  e  $0 \leq k < m$ . Si sarebbe tentati di dire che l'effetto è quello di riordinare la lista di partenza  $[0 \dots m-1]$  in modo apparentemente casuale. (L'effetto qui non è eclatante, ma divertitevi a modificare il programma e ripetere l'esercizio con due numeri primi  $m$  e  $p$  ragionevolmente grandi).

### 6.3.1 Il logaritmo discreto

Il problema è anche più interessante di così. Adesso possiamo porci una domanda leggermente diversa, ovverosia: dato un intero positivo  $m$  e due interi  $0 \leq p, g < m$ , come possiamo calcolare il valore di  $k$  per cui

$$p^k \bmod m = g? \quad (201)$$

Il lettore non avrà mancato di notare l'analogia con il logaritmo ordinario nel campo dei numeri reali, ed in effetti la soluzione  $k$  della 201 si dice solitamente logaritmo discreto di  $g$  in base  $b$  (modulo  $m$ ).

Il problema del logaritmo discreto è importante in crittografia perché è considerato di difficile soluzione—almeno quando si tratta di numeri grandi. La figura 6.3 illustra quanto  $p^k \bmod m$  vari in modo irregolare al variare di  $k$ , per cui è estremamente difficile immaginare una strategia generale per calcolare il logaritmo discreto che non sia la forza bruta (cioè provare tutti i valori di  $k$  uno dopo l'altro fino a quando non troviamo la soluzione).

## 6.4 SEQUENZE PSEUDO-CASUALI

Una dei primi schemi per la generazione di sequenze pseudocasuali con un calcolatore è dovuta a John Von Neumann e risale alla metà degli anni '40 del '900. L'algoritmo di base è molto semplice: partiamo da un numero intero  $X_0$  di  $m$  cifre (che chiamiamo *seme* o *seed*), e definiamo una successione per

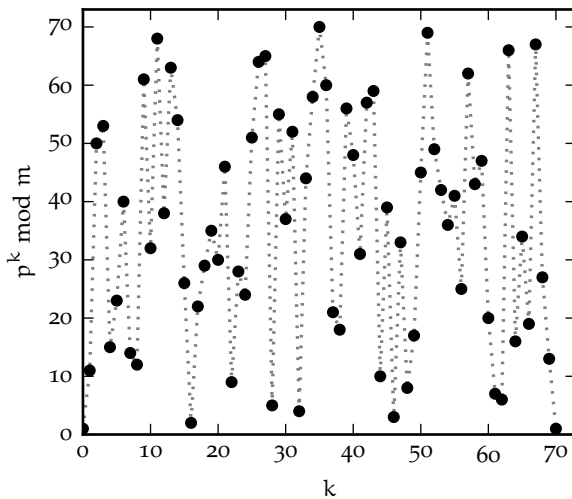


FIGURA 6.3. Illustrazione del problema del logaritmo discreto per  $m = 71$  e  $p = 11$ . Al variare di  $k = 0 \dots m - 1$  i valori di  $p^k \bmod m$  variano in modo estremamente irregolare, il che rende difficile calcolare la soluzione di  $p^k \bmod m = g$ , se non provando tutti i  $k$ .

ricorrenza con la prescrizione che ad ogni passo prendiamo le  $m$  cifre centrali del quadrato dell'elemento precedente—ovverosia, in rappresentazione decimale:

$$X_{n+1} = (X_n^2 / 10^{\frac{m}{2}}) \bmod 10^m. \quad (202)$$

Il metodo di Von Neumann è generalmente noto con il nome di *middle square*, ed è più complicato a spiegarsi descrittivamente che non con un esempio—cosa che facciamo prontamente nella tabella 17.

$n$	$X_n$	$X_n^2$	$X_n^2 / 10^{\frac{m}{2}}$	$X_{n+1}$
0	3333	11108889	111088	1088
1	1088	01183744	011837	1837
2	1837	03374569	033745	3745
3	3745	14025025	140250	0250
4	0250	00062500	000625	0625
5	0625	00390625	003906	3906
6	3906	15256836	152568	2568
7	2568	06594624	065946	5946
8	5946	35354916	353549	3549
9	3549	12595401	125954	5954

TABELLA 17. Illustrazione del metodo *middle square* per la generazione di sequenze pseudo-casuali con  $m = 4$  cifre decimali e seme iniziale  $X_0 = 3333$ . (Va da sé che in un calcolatore l'algoritmo sarebbe più efficiente se implementato in logica binaria anziché decimale.) Nella terza colonna sono sottolineate le 4 cifre centrali tra le 8 che costituiscono il quadrato dell' $n$ -simo numero nella sequenza, ma per completezza la quarta e la quinta colonna mostrano i singoli passi della (202) separatamente.

L'idea di base è che le cifre centrali del risultato dell'elevamento al quadrato dipendono in modo complicato dal numero di partenza e possono essere considerate come un numero di  $m$  cifre essenzialmente scorrelato da esso. Retrospectivamente possiamo dire che il metodo *middle square* non è soddisfacente secondo gli standard moderni, ed il suo interesse è ormai di carattere prevalentemente storico. Una delle limitazioni maggiori consiste nella tendenza dell'algoritmo di rimanere intrappolato in cicli di lunghezza molto breve. Potete verificare, ad esempio, che se scegliamo 2500 come seme per il nostro generatore a 4 cifre, la (202) trasforma  $X_0$  in se stesso indefinitamente. Purtuttavia forniamo un'implementazione utilizzabile del metodo nel frammento 6.4.

#### 6.4.1 Aspetti generali dei generatori pseudo-casuali

L'algoritmo *middle square* descritto nella sezione precedente, benché ormai di scarsa rilevanza pratica per i motivi che abbiamo visto, è importante perché esemplifica almeno due aspetti fondamentali di molti generatori pseudo-casuali utilizzati in pratica. Il primo è l'idea di base di generare ricorsivamente gli elementi (interi) della sequenza

$$X_{n+1} = f(X_n) \quad (203)$$

```

1  https://bitbucket.org/.../middle-square.py
2  # Middle-square random number generator with m = 4.
3  # Define the seed and print it out.
4  x = 3333
5  sequence = [x]
6
7  # Generate 10 numbers---note  $10^{\lfloor m/2 \rfloor} = 100$  and  $10^m = 10000$ .
8  for n in range(10):
9      x = (x**2 // 100) % 10000
10     sequence.append(x)
11
12 # Print the sequence.
13 print(sequence)
14
15 [Output]
16 [3333, 1088, 1837, 3745, 250, 625, 3906, 2568, 5946, 3549, 5954]

```

FRAMMENTO 6.4. Possibile implementazione dell'algoritmo *middle square* per la generazione della sequenza di 10 elementi illustrata in tabella 17.

a partire da un seme (o *seed*)  $X_0$  ed utilizzando una funzione  $f: [0, X_{\max}] \rightarrow [0, X_{\max}]$ . (Questa funzione può essere applicata all'elemento precedente della sequenza o, più in generale ad una  $n$ -tupla di elementi.) Il secondo è quello di sfruttare l'aritmetica modulare per far sì che la sequenza si comporti *come se fosse casuale*. Torneremo sull'argomento tra un attimo.

Prima di andare avanti soffermiamoci un momento sui due esempi con cui abbiamo aperto questo capitolo: il calcolo di  $\pi$  e la macchina di Galton—in entrambi i casi abbiamo utilizzato un generatore casuale di numeri reali compresi tra 0 ed 1. Qual è, allora la connessione con la (203)? Molto semplicemente, detto  $X_{\max}$  il valore massimo restituito da  $f(X)$ , che per esempio è  $10^m - 1$  nel generatore (202), possiamo semplicemente dividere  $X_n$  per  $X_{\max} + 1$

$$U_n = \frac{X_n}{X_{\max} + 1} \quad (204)$$

per ottenere una sequenza  $U_n$  di numeri reali tra 0 ed 1—o, più precisamente, nell'intervallo  $[0, 1)$ . Va da sé che la granularità non sarà infinita (tecnicamente avremo numeri razionali con precisione  $1/X_{\max}$ ), ma questo è irrilevante, perché l'aritmetica in virgola mobile su un calcolatore è comunque inesatta (come discusso sommariamente in appendice D).

Torniamo alla nostra definizione per ricorrenza. Una conseguenza immediata della (203) è che il *periodo* della sequenza sarà necessariamente limitato: avendo a disposizione un numero finito di interi, ad un certo punto la nostra funzione  $f$  trasformerà l'elemento  $n$ -esimo in un numero che abbiamo già ottenuto in precedenza e, da quel punto in poi, la sequenza si ripeterà in un ciclo infinito. Quando questo accade, la sequenza cessa di comportarsi come se fosse casuale, per cui il periodo costituisce un limite superiore al numero di elementi che si possono estrarre utilmente. Tutto ciò non è un problema in pratica se il periodo è abbastanza lungo, per cui è chiaro che questa è una delle considerazioni fondamentali nel disegno di un generatore.

#### 6.4.2 Lo schema lineare congruenziale

Lo schema lineare congruenziale (*linear congruential method*), proposto all fine degli anni '40 del secolo scorso, è ancora oggi abbastanza popolare, e lo discutiamo brevemente. La definizione di base è

$$X_{n+1} = (aX_n + c) \bmod m \quad (205)$$

in cui il moltiplicatore  $a$ , l'incremento  $c$  ed il modulo  $m$  sono tre numeri fissi opportunamente scelti. (Ovviamente abbiamo anche bisogno di un seme  $X_0$  per definire univocamente la sequenza—e diversi valori del seme danno origine a diverse sequenze.) Il fatto di prendere il risultato dell'operazione tra parentesi modulo  $m$  implica che il massimo numero ottenibile  $X_{\max} = m - 1$  e che il periodo massimo ottenibile è esattamente  $m$ . In generale non è garantito che, per una terna arbitraria di valori  $a$ ,  $c$  ed  $m$  la sequenza abbia periodo massimo, ma esistono condizioni ben definite perché questo accada [11].

Per completezza, il frammento 6.5 implementa un generatore *giocattolo* di numeri interi a due cifre decimali utilizzando lo schema lineare congruenziale. Notiamo, per inciso, che il suo periodo è il massimo possibile (100) per un generatore a due cifre decimali.

<https://bitbucket.org/.../lcg1.py>

```

1  # Toy two-digit (0--99) linear congruential generator.
2  a = 11
3  c = 3
4  m = 73
5  x = 81
6  sequence = []
7
8  # Generate 10 numbers.
9  for i in range(10):
10     x = (a*x + c) % m
11     sequence.append(x)
12
13 print(sequence)
14
15 [Output]
16 [18, 55, 24, 48, 20, 4, 47, 9, 29, 30]
```

FRAMMENTO 6.5. Esempio di implementazione di un generatore lineare congruenziale giocattolo di numeri interi a due cifre decimali con periodo massimo (100).

Il frammento 6.5, viceversa, illustra un generatore realistico, dotato di un periodo di  $2^{64} \approx 1.8 \times 10^{19}$ . Non è niente in confronto al generatore implementato nella libreria standard di Python (che, come abbiamo detto, ha un periodo di  $2^{1937} - 1$ ) ma si tratta pur sempre di qualcosa di utilizzabile in semplici applicazioni.

<https://bitbucket.org/.../lcg2.py>

```

1  # Real (i.e., usable) linear congruential generator.
2  a = 6364136223846793005
3  c = 1442695040888963407
4  m = 2**64
5  x = 1
6
7  # Generate 10 number between 0 and 1.
8  for i in range(10):
9     x = (a*x + c) % m
10     u = float(x)/(m - 1)
11     print(u)
12
13 [Output]
14 0.423209170873
15 0.509407442884
16 0.648359393963
17 0.382863390508
18 0.795447749254
19 0.500511282795
20 0.553935361313
21 0.0654193119742
22 0.839726109648
23 0.198440042789
```

FRAMMENTO 6.6. Esempio di implementazione di un generatore lineare congruenziale di numeri in virgola mobile (equidistribuiti tra 0 ed 1) utilizzabile in pratica, con periodo  $2^{64}$ .

## 6.5 BREVE DIGRESSIONE: BITMAP CASUALI

## 6.6 GENERAZIONE DI NUMERI PSEUDO-CASUALI CON DISTRIBUZIONE NON UNIFORME

A questo punto abbiamo un'idea abbastanza precisa di come si possano generare numeri pseudo-casuali distribuiti tra 0 ed 1. Il prossimo passo ovvio è chiedersi come si possano generare numeri pseudo-casuali con distribuzioni diverse.

## 6.7 IN BREVE...



---

FUNZIONI DI VARIABILI CASUALI

---

È giunto il momento di affrontare, almeno nel caso unidimensionale, un problema che abbiamo toccato in superficie nella sezione 4.6—ovverosia quello del calcolo esplicito della funzione di distribuzione di una *funzione di variabile casuale*  $y = f(x)$ , nota la funzione di distribuzione della variabile casuale  $x$ .

Può sembrare un argomento tecnico e, tutto sommato di scarso interesse, ma in questo capitolo vedremo che si tratta di un argomento fecondo di conseguenza, ed alcuni dei risultati ci serviranno per inquadrare quantitativamente i problemi di fit.

## 7.1 VARIABILI DISCRETE

Cominciamo con il caso (relativamente semplice) di una variabile casuale discreta. Uno potrebbe essere tentato di scrivere semplicemente

$$P(y = f(x_i)) = P(x = x_i) \quad (206)$$

ed in effetti questo è ciò che succede nel caso più elementare—quello cioè in cui la funzione  $f$  è biunivoca nell'intervallo di variabilità di  $x$  (vale a dire che per ogni valore  $y_i$  di  $y$  esiste uno ed un solo valore  $x_i$  di  $x$  tale che  $y_i$  è il trasformato di  $x_i$  attraverso la nostra funzione).

► **ESEMPIO 7.1.** Sia  $x$  l'uscita di un dado equo a sei facce. La funzione di distribuzione della variabile casuale  $y = x^2$  si può scrivere in forma tabellare come

$x_i$	1	4	9	16	25	36
$P(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6

La situazione è leggermente più complicata se lasciamo cadere l'ipotesi di biunivocità. È facile convincersi che in questo caso, preso un valore (ammesso) della variabile  $y$ :

$$P(y = f(x_i)) = \sum_{i=1}^m P(x = x_i) \quad (207)$$

dove la somma è estesa a tutti i valori (che qui supponiamo genericamente essere  $m$ ) di  $x$  per cui  $f(x_i) = \tilde{y}$ . È più difficile a dirsi che a farsi, e chiariamo il punto con l'esempio 7.2.

## 7.2 VARIABILI CONTINUE

Il caso di una variabile casuale continua è quello di gran lunga più interessante dal nostro punto di vista. Vedremo che le considerazioni che abbiamo fatto a proposito delle variabili discrete nella sezione precedente si applicano anche qui—e vedremo che, non a caso, calcolare la funzione di distribuzione di una funzione di variabile casuale nel caso continuo corrisponde essenzialmente a fare un cambiamento di variabile in un integrale.

Partiamo dunque, come prima, dal caso più semplice—quello in cui la funzione  $f(x)$  è biunivoca nell'intervallo di variabilità di  $x$ . (Per semplicità assumeremo anche che  $f(x)$  sia derivabile nell'intervallo

► **ESEMPIO 7.2.** Immaginiamo di avere un ipotetico dado simmetrico equo a sette facce—un oggetto cioè, la cui uscita  $x$  sia descritta dalla funzione di distribuzione

$x_i$	-3	-2	-1	0	1	2	3
$P(x_i)$	$1/7$	$1/7$	$1/7$	$1/7$	$1/7$	$1/7$	$1/7$

È chiaro che qui le cose non possono funzionare esattamente come prima. Per prima cosa la variabile casuale  $y = x^2$  potrà assumere solo i 4 valori 0, 1, 4, 9 poiché, ad esempio, il valore  $x_i = -2$  è trasformato in  $y = 4$  dalla funzione  $f(x)$  esattamente come  $x_i = 2$ . Allora avremo

$$P(y = 1) = P(x = -1) + P(x = 1) = \frac{1}{7} + \frac{1}{7} = \frac{2}{7},$$

e la funzione di distribuzione di  $y$  si scrive esplicitamente come:

$y_i$	0	1	4	9
$P(y_i)$	$1/7$	$2/7$	$2/7$	$2/7$

di interesse.) La biunivocità della corrispondenza tra le due variabili assicura che se  $x$  è compresa in un intervallo (infinitesimo) centrato su un generico valore  $x_0$ :

$$x \in [x_0 - dx, x_0 + dx]$$

allora (e solo allora)  $y$  è compresa in un intervallo (infinitesimo)

$$y \in [y_0 - dy, y_0 + dy] \quad \text{con} \quad y_0 = f(x_0) \quad \text{e} \quad dy = \left| \frac{df(x_0)}{dx} \right| dx.$$

(Se siete incerti su questo punto, tornate a leggere la sezione 1.5 ed osservate con attenzione la figura 1.3. Non abbiamo fatto niente di nuovo.) Se la nostra funzione  $f$  è biunivoca, possiamo anche invertirla per ottenere un'espressione per la vecchia variabile in funzione della nuova, sia puntualmente che in termini di differenziali

$$x = f^{-1}(y) \quad \text{e} \quad dx = \left| \frac{df^{-1}(y)}{dy} \right| dy.$$

Ma questo è tutto quello che ci serve. Se indichiamo con  $p_x(x)$  la densità di probabilità della nostra variabile originaria e con  $p_y(y)$  la densità di probabilità della variabile trasformata (che è esattamente quello che cerchiamo) allora l'unica cosa che dobbiamo garantire è che la probabilità associata in intervalli infinitesimi corrispondenti sia preservata

$$p_y(y)dy = p_x(x)dx = p_x(f^{-1}(y)) \left| \frac{df^{-1}(y)}{dy} \right| dy,$$

che deve valere per qualsiasi  $x$ . In definitiva, la nostra legge di trasformazione si legge

$$p_y(y) = p_x(f^{-1}(y)) \left| \frac{df^{-1}(y)}{dy} \right|. \quad (208)$$

La prescrizione per il cambio di variabile si può dunque esprimere a parole mediante il seguente processo a tre passi:

1. invertire la  $f(x)$  per ricavare un'espressione  $x = f^{-1}(y)$  per  $x$  in funzione di  $y$ ;
2. derivare rispetto a  $y$  l'espressione ottenuta al punto 1;
3. la funzione di distribuzione cercata è uguale al prodotto della funzione di distribuzione  $p_x(x)$  di  $x$ , espressa in funzione della nuova variabile  $y$ , moltiplicata per la derivata calcolata al punto 2.

Se adesso rilasciamo la richiesta che  $f(x)$  sia biunivoca, il ragionamento procede sostanzialmente invariato, con l'unica differenza che adesso dobbiamo fare una somma sulla molteplicità delle radici dell'equazione  $x = f^{-1}(y)$

$$p_y(y) = \sum_{i=1}^m p_x\left(f_i^{-1}(y)\right) \left| \frac{df_i^{-1}(y)}{dy} \right|. \quad (209)$$

### 7.3 UN ESEMPIO SIGNIFICATIVO: IL QUADRATO DI UNA VARIABILE CASUALE

Anziché procedere discutendo la materia in termini astratti, preferiamo illustrare quanto imparato fino a questo momento con due esempi concreti che ci saranno utili nel seguito: il calcolo della funzione di distribuzione del quadrato di una variabile casuale  $x$  con media  $\mu = 0$  e varianza  $\sigma^2 = 1$  distribuita uniformemente o Gaussianamente.

(Notiamo, per inciso, che si tratta di una situazione in cui le relazioni approssimate per la media e la varianza di una funzione di variabile casuale che abbiamo ricavato nella sezione 4.6 falliscono miseramente, poiché la derivata di  $f(x) = x^2$  si annulla nel valor medio della variabile casuale, ed il nostro sviluppo al prim'ordine è per definizione privo di significato.)

#### 7.3.1 Il quadrato di una distribuzione uniforme

Consideriamo dunque una variabile casuale  $x$  distribuita uniformemente tra  $-\sqrt{3}$  e  $\sqrt{3}$ —cioè con densità di probabilità

$$p_x(x) = \begin{cases} \frac{1}{2\sqrt{3}} & -\sqrt{3} \leq x \leq \sqrt{3} \\ 0 & x < -\sqrt{3}; x > \sqrt{3} \end{cases}$$

(se vi state chiedendo il motivo della scelta apparentemente curiosa dell'intervallo, è semplice: questa è la distribuzione uniforme con media  $\mu = 0$  e deviazione standard  $\sigma = 1$ , come si può verificare facilmente utilizzando i risultati che abbiamo ricavato nella sezione 5.5). Siamo interessati alla densità di probabilità  $p_y$  della variabile trasformata

$$y = f(x) = x^2.$$

Prima ancora di cominciare, notiamo che il valor medio di  $y$  è banale da calcolare grazie al fatto che la media di  $x$  è nulla

$$E[y] = E[x^2] = \sigma^2 + \mu^2 = \sigma^2 = 1.$$

La nuova variabile  $y$  è definita positiva, ed il supporto della densità di probabilità che stiamo cercando sarà  $[0, 3]$ . La distribuzione di partenza è simmetrica rispetto ad  $x = 0$ , per cui per ogni  $y > 0$  vi sono esattamente due valori  $x = \pm\sqrt{y}$  per cui  $f(x) = y$ —ed i valori della funzione di distribuzione e del modulo della derivata di  $f^{-1}$  sono identici in questi due punti. Ne segue che la nostra somma si semplifica in

$$p_y(y) = 2p_x\left(f^{-1}(y)\right) \left| \frac{df^{-1}(y)}{dy} \right|,$$

dove per  $f^{-1}(y)$  intendiamo la radice positiva. I due ingredienti che ci servono per calcolare  $p_y$  sono dunque

$$x = f^{-1}(y) = \sqrt{y} \quad \text{e} \quad \frac{df^{-1}(y)}{dy} = \frac{1}{2\sqrt{y}} \quad \text{da cui} \quad p_y(y) = \frac{1}{2\sqrt{3y}}. \quad (210)$$

La densità di probabilità appena calcolata è mostrata in figura 7.1. È facile verificare che la distribuzione, così scritta, è correttamente normalizzata.

$$\int_0^3 \frac{1}{2\sqrt{3y}} dy = \frac{1}{2\sqrt{3}} \int_0^3 \frac{1}{\sqrt{y}} dy = \left( \frac{y}{3} \right)^{\frac{1}{2}} \Big|_0^3 = 1.$$

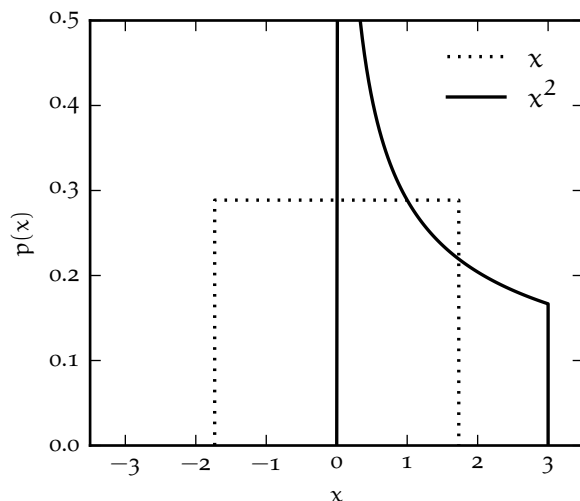


FIGURA 7.1. Grafico della funzione densità di probabilità del quadrato di una variabile casuale  $x$  distribuita uniformemente tra  $-\sqrt{3}$  e  $\sqrt{3}$ . (Per completezza la linea tratteggiata rappresenta la funzione di distribuzione della variabile di partenza.) È interessante notare come la densità di probabilità della variabile trasformata diverga in  $x = 0$ .

Per completezza calcoliamo esplicitamente media e varianza—anche se di fatto la media la conosciamo già:

$$E[y] = \int_0^3 y p_y(y) dy = \int_0^3 \frac{y}{2\sqrt{3y}} dy = \left( \frac{y}{3} \right)^{\frac{3}{2}} \Big|_0^3 = 1 \quad (211)$$

$$E[y^2] = \int_0^3 y^2 p_y(y) dy = \int_0^3 \frac{y^2}{2\sqrt{3y}} dy = \frac{y^{\frac{5}{2}}}{5\sqrt{3}} \Big|_0^3 = \frac{9}{5} \quad \text{da cui} \quad \text{Var}(y) = \frac{9}{5} - 1 = \frac{4}{5}. \quad (212)$$

### 7.3.2 Il quadrato di una Gaussiana

Passiamo ora ad un caso più interessante, ovvero il calcolo della funzione di distribuzione del quadrato di una variabile Gaussiana  $z$  in forma standard. Avremo modo di vedere nel capitolo 8 come la questione è collegata ai metodi di *fit*.

La quasi totalità delle cose che abbiamo detto nella sezione precedente sono ancora valide. È ancora vero che il valore di aspettazione di  $y$  è 1, ed è ancora vero che per ogni  $y > 0$  vi sono esattamente due valori  $x = \pm\sqrt{y}$  per cui  $f(x) = y$ , per cui si ha di nuovo

$$p_y(y) = 2p_x(f^{-1}(y)) \left| \frac{df^{-1}(y)}{dy} \right|.$$

(Una differenza notevole, se vogliamo, è che adesso il supporto della densità di probabilità di  $p_y(y)$  è l'intera semiretta  $y \geq 0$ .) In effetti abbiamo già tutti gli ingredienti che ci servono, e dobbiamo semplicemente sostituire  $y = z^2$  a  $z$  nella densità di probabilità della Gaussiana e moltiplicare per la derivata della funzione di trasformazione inversa

$$p_y(y) = \frac{2}{\sqrt{2\pi}} e^{-\frac{y}{2}} \frac{1}{2\sqrt{y}} = \frac{e^{-\frac{y}{2}}}{\sqrt{2\pi y}}. \quad (213)$$

Questa volta non (ri-)dimostriamo esplicitamente che l'espressione ottenuta è correttamente normalizzata e che la media della distribuzione è  $E[y] = 1$ , ma il calcolo della varianza è istruttivo e ci servirà nel seguito. Integrando per parti si ottiene

$$E[y^2] = \int_0^\infty \frac{y^2 e^{-\frac{y}{2}}}{\sqrt{2\pi y}} dy = \frac{1}{\sqrt{2\pi}} \int_0^\infty y^{\frac{3}{2}} e^{-\frac{y}{2}} dy = -\frac{2}{\sqrt{2\pi}} y^{\frac{3}{2}} e^{-\frac{y}{2}} \Big|_0^\infty + \frac{3}{\sqrt{2\pi}} \int_0^\infty y^{\frac{1}{2}} e^{-\frac{y}{2}} dy.$$

Ora, il primo termine si annulla, mentre nel secondo riconosciamo essenzialmente la definizione del valor medio, che conosciamo già, per cui

$$E[y^2] = 3E[y] = 3 \quad \text{e} \quad \text{Var}(y) = 3 - 1 = 2. \quad (214)$$

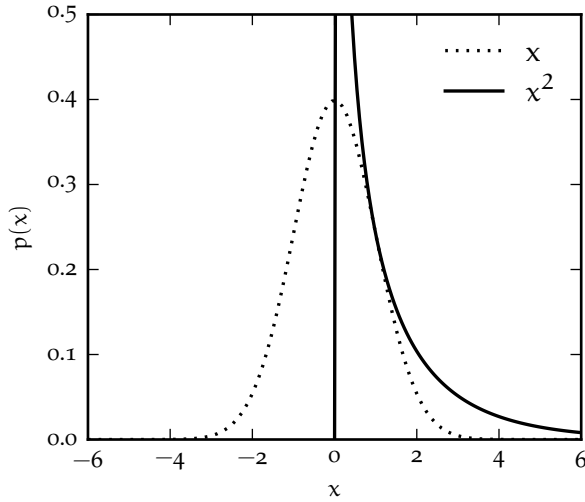


FIGURA 7.2. Grafico della funzione densità di probabilità del quadrato di una variabile casuale  $z$  Gaussiana in forma standard. (Per completezza la linea tratteggiata rappresenta la funzione di distribuzione della variabile di partenza.) Notiamo che anche qui la densità di probabilità della variabile trasformata diverge in  $x = 0$ .

Il quadrato di una variabile casuale Gaussiana in forma standard ha media 1 e varianza 2—teniamolo a mente perché ci sarà utile tra un attimo.

#### 7.4 LA DISTRIBUZIONE DEL $\chi^2$

Il risultato ottenuto nella sezione 7.3.2 ci offre lo spunto per introdurre una distribuzione che, pur non essendo strettamente legata al tema principale di questo capitolo, ci sarà utile nel seguito. Consideriamo la somma  $x$  di  $n$  variabili Gaussiane standard indipendenti *al quadrato*—che si indica solitamente, per ragioni che vedremo meglio in seguito, con il nome di  $\chi^2$  a  $n$  gradi di libertà

$$x = \sum_{i=1}^n z_i^2. \quad (215)$$

Come si scrive la funzione di distribuzione di  $x$ ? (Il problema non è proprio banale, ed abbiamo detto che la pertinenza con il filo conduttore di questo capitolo è limitata perché non si tratta di una funzione di una singola variabile casuale, ma della somma di variabili casuali diverse ed indipendenti.)

Prima ancora di esaminare il problema in dettaglio, grazie al fatto che sappiamo tutto sulla funzione di distribuzione di ciascuna delle  $z_i^2$ , possiamo dire un certo numero di cose su  $x$ . In particolare nella sezione 4.2.1 abbiamo ricavato le regole di somma delle medie e delle varianze per variabili casuali indipendenti, che ci permettono di scrivere banalmente

$$E[x] = \sum_{i=1}^n E[z_i^2] = \sum_{i=1}^n 1 = n \quad (216)$$

$$\text{Var}(x) = \sum_{i=1}^n \text{Var}(z_i^2) = \sum_{i=1}^n 2 = 2n. \quad (217)$$

Inoltre sappiamo che, per il teorema del limite centrale, la funzione di distribuzione di  $x$  tende ad una gaussiana—con media e varianza date dalle (216) e (217)—per  $n \rightarrow \infty$ .

Per completezza forniamo, senza dimostrazione, la soluzione completa del problema, nella forma della distribuzione del  $\chi^2$

$$p(x; n) = \frac{1}{2^{\frac{n}{2}} \Gamma(n/2)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}. \quad (218)$$

È facile verificare per sostituzione diretta che la funzione di distribuzione di una variabile Gaussiana in forma standard al quadrato che abbiamo ricavato nella sezione 7.3.2 non è altro che un  $\chi^2$  ad 1 grado di libertà. Per 2 gradi di libertà, viceversa, la distribuzione del  $\chi^2$  si riduce alla distribuzione esponenziale (174) con media  $\mu = 2$ . Vari esempi di distribuzione del  $\chi^2$ , per valori diversi del numero di

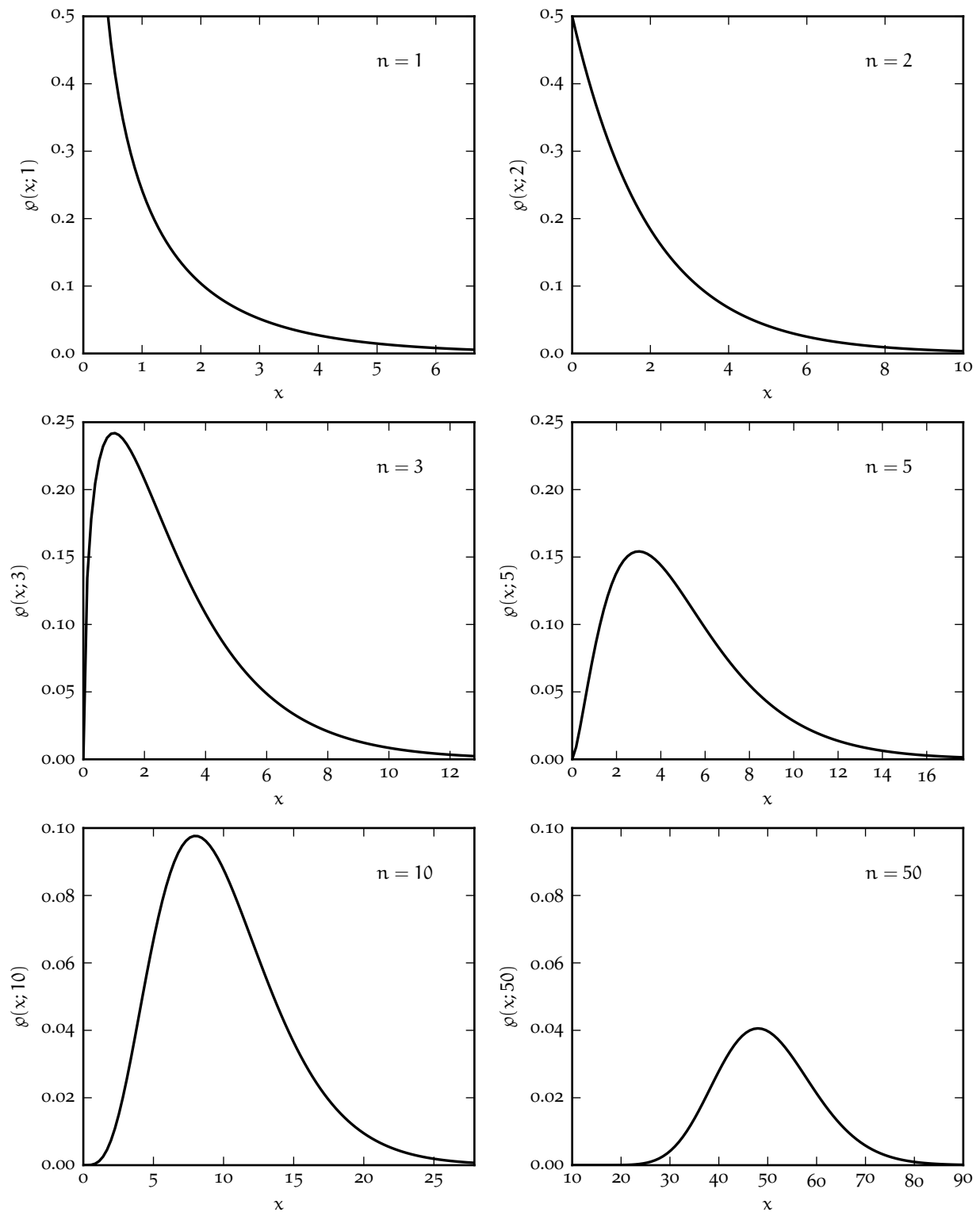


FIGURA 7.3. Esempi di distribuzione del  $\chi^2$ . Al variare del numero di gradi di libertà  $n$  la (218) genera una varietà notevole di forme differenti. Per  $n = 1$  essa coincide con la densità di probabilità di una variabile Gaussiana in forma standard al quadrato (213) e per  $n = 2$  con una distribuzione esponenziale (174) con media  $\mu = 2$ . Quando  $n$  è molto grande la distribuzione del  $\chi^2$  tende ad una Gaussiana per il teorema del limite centrale.

gradi di libertà, sono mostrati in figura 7.3. Anticipiamo che la funzione cumulativa per una variabile  $\chi^2$  è tabulata in appendice A.6.

### 7.5 UN CASO PECULIARE: LA DISTRIBUZIONE DI CAUCHY

Consideriamo una variabile casuale continua  $\phi$  distribuita uniformemente nell'intervallo  $[-\pi/2, \pi/2]$

$$p_{\phi}(\phi) = \begin{cases} \frac{1}{\pi} & -\frac{\pi}{2} \leq \phi \leq \frac{\pi}{2} \\ 0 & \phi < -\frac{\pi}{2}; \phi > \frac{\pi}{2} \end{cases}$$

Siamo interessati a calcolare la forma esplicita della densità di probabilità della variabile trasformata  $x$  definita da

$$x = f(\phi) = \gamma \tan \phi + x_0. \quad (219)$$

Prima di andare avanti con il calcolo, notiamo che questo problema ha un'interessante interpretazione geometrica, mostrata in figura 7.4: se abbiamo una sorgente luminosa posta a distanza  $\gamma$  da uno schermo infinito (in corrispondenza di un punto fissato  $x_0$ ) che emette uniformemente in tutte le direzioni, allora  $x$  rappresenta la distribuzione dei punti di intersezione dei raggi con lo schermo stesso.

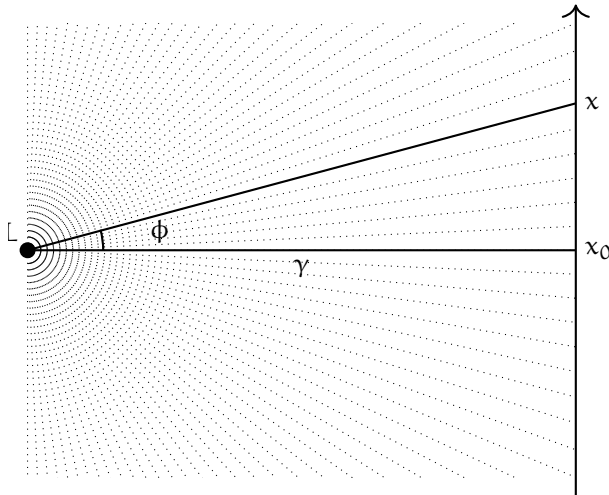


FIGURA 7.4. Interpretazione geometrica della trasformazione (219). Se la sorgente luminosa  $L$  è posta a distanza  $\gamma$  dallo schermo in corrispondenza del punto  $x_0$ ,  $x$  rappresenta l'intersezione con lo schermo stesso del generico raggio di luce emesso ad un angolo  $\phi$ .

Cominciamo, al solito, scrivendo la trasformazione inversa e derivando la relazione ottenuta rispetto alla variabile trasformata  $x$ :

$$\phi = f^{-1}(x) = \arctan\left(\frac{x - x_0}{\gamma}\right), \quad \text{da cui} \quad \frac{df^{-1}(x)}{dx} = \frac{\gamma}{(x - x_0)^2 + \gamma^2}.$$

A questo punto è banale trovare la densità di probabilità di  $x$ —si tratta essenzialmente di moltiplicare per  $1/\pi$ . La funzione di distribuzione di  $x$  si dice distribuzione di Cauchy (o, a seconda del contesto, distribuzione di Lorentz, Lorentziana o distribuzione di Breit-Wigner) e si scrive come

$$c(x; x_0, \gamma) = \frac{1}{\pi} \left[ \frac{\gamma}{(x - x_0)^2 + \gamma^2} \right]. \quad (220)$$

Notiamo esplicitamente che la distribuzione di Cauchy dipende da due parametri: il *parametro di posizione*  $x_0$  ed il *parametro di scala*  $\gamma$ . Nel caso particolare  $x_0 = 0$  e  $\gamma = 1$ , analogamente a quanto si fa per la distribuzione di Gauss, si parla di distribuzione di Cauchy in forma standard

$$c(x; 0, 1) = \frac{1}{\pi(1 + x^2)}. \quad (221)$$

Le densità di probabilità corrispondenti sono confrontate in figura 7.5.

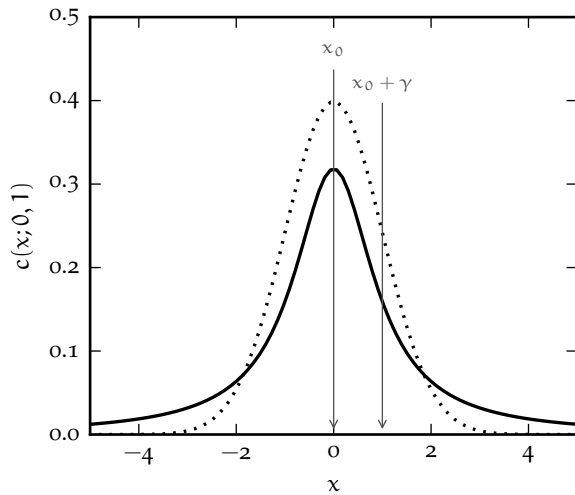


FIGURA 7.5. Grafico della distribuzione di Cauchy in forma standard (cioè con parametro di posizione  $x_0 = 0$  e parametro di scala  $\gamma = 1$ ). Per confronto, la linea tratteggiata rappresenta una distribuzione di Gauss in forma standard. La differenza più rilevante tra le due sta nel fatto che le code della distribuzione di Cauchy sono significativamente più pronunciate.

### 7.5.1 Alcune proprietà della distribuzione di Cauchy

La distribuzione di Cauchy, così come è scritta nella (220), è correttamente normalizzata—e l'integrale è banale in quanto, avendo ottenuto la densità di probabilità derivando la trasformazione inversa (modulo una costante), abbiamo di fatto l'espressione esplicita per la primitiva della densità di probabilità stessa

$$\int_{-\infty}^{\infty} c(x; x_0, \gamma) dx = \int_{-\infty}^{\infty} \frac{1}{\pi} \left[ \frac{\gamma}{(x - x_0)^2 + \gamma^2} \right] dx = \frac{1}{\pi} \arctan \left( \frac{x - x_0}{\gamma} \right) \Big|_{-\infty}^{\infty} = \frac{1}{\pi} \left( \frac{\pi}{2} + \frac{\pi}{2} \right) = 1.$$

Per inciso, questo ci permette immediatamente di calcolare la funzione cumulativa e la funzione di distribuzione inversa per la distribuzione in questione—ovvero

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan \left( \frac{x - x_0}{\gamma} \right) \quad \text{e} \quad F^{-1}(q) = x_0 + \gamma \tan \left( \pi q - \frac{\pi}{2} \right). \quad (222)$$

Veniamo alle cose interessanti, e concentriamoci un attimo per semplicità (ma senza perdere in generalità) sulla distribuzione in forma standard (221). La nostra solita espressione per la media

$$E[x] = \int_{-\infty}^{\infty} x c(x; 0, 1) dx = \int_{-\infty}^{\infty} \frac{x}{\pi(1 + x^2)} dx = \frac{1}{2\pi} \log(1 + x^2) \Big|_{-\infty}^{\infty}$$

porta in questo caso ad una forma indeterminata di tipo  $\infty - \infty^1$  per cui siamo costretti ad ammettere che la distribuzione di Cauchy non ha media definita. Questo è un problema—e non da poco—perché se la media non è definita non possiamo definire i momenti centrali di ordine superiore, primo tra tutti la varianza. Notiamo che se, in virtù del fatto che la funzione di distribuzione è pari, volessimo ignorare che la media non esiste e provassimo a calcolare una sorta di pseudo-varianza come momento centrale di ordine due, non andremmo molto lontano, poiché

$$E[x^2] = \int_{-\infty}^{\infty} x^2 c(x; 0, 1) dx = \int_{-\infty}^{\infty} \frac{x^2}{\pi(1 + x^2)} dx \rightarrow \infty.$$

(In questo caso, banalmente, l'integrando non tende a zero per  $x \rightarrow \pm\infty$ .) Il fatto che questo integrale diverga, fisicamente, è dovuto alle code della distribuzione, che sono molto più accentuate rispetto al caso, e.g., della Gaussiana. Più in generale possiamo dire che tutti i momenti algebrici di ordine dispari

<sup>1</sup> Non abbiamo gli strumenti per sviscerare l'argomento in tutti i suoi dettagli, ma vale la pena notare che la *parte* principale dell'integrale che esprime la media

$$\lim_{c \rightarrow \infty} \int_{-c}^c x c(x; 0, 1) dx$$

esiste ed è 0, come uno si aspetterebbe intuitivamente sulla base che la distribuzione è pari. Purtroppo prescrizioni diverse sul limite portano a valori diversi dell'integrale per cui, come abbiamo detto, l'integrale stesso non è definito.



della distribuzione di Cauchy sono indefiniti e tutti i momenti algebrici di ordine pari divergono; questo vale in generale—non solo per la distribuzione in forma standard.

Notiamo anche, per completezza, che la mediana e la moda della (221) sono entrambi pari al parametro di posizione  $x_0$ . La semilarghezza a metà altezza è data dalla soluzione (per la quantità  $x - x_0$ ) dell'equazione

$$\frac{1}{\pi} \left[ \frac{\gamma}{(x - x_0)^2 + \gamma^2} \right] = \frac{1}{2\pi\gamma} \quad \text{ovvero} \quad \text{HWHM} = \gamma,$$

cioè è pari al parametro di scala. In assenza di una media e di una deviazione standard definite, la mediana e la semilarghezza a metà altezza sono in questo caso le misure più utili di tendenza centrale e dispersione intorno alla media.

### 7.5.2 La distribuzione di Cauchy ed il teorema del limite centrale

Se vi ricordate una delle ipotesi necessarie per la validità del teorema centrale del limite è il fatto che le distribuzioni di partenza avessero media e varianza finite. Ebbene, la figura 7.6 illustra il fallimento spettacolare del teorema centrale del limite nel caso della distribuzione di Cauchy (confrontatela con le figure 5.12 e 5.13). Osserviamo la situazione in dettaglio. Il primo problema è che, a parte la scala sull'asse delle  $x$ , gli istogrammi in figura 7.6 sono essenzialmente identici l'uno all'altro, indipendentemente dal valore di  $n$ —cioè la loro forma non si avvicina, al crescere di  $n$ , ad una distribuzione di Gauss. Il secondo problema è che la larghezza relativa degli istogrammi cresce come  $n$ , anziché come  $\sqrt{n}$ , al crescere di  $n$ . (Va da sé che nessuno dei due è un reale problema: la distribuzione di Cauchy non ha media e varianza finite per cui semplicemente il teorema centrale del limite non vale per la somma di distribuzioni di Cauchy.)

Possiamo andare oltre. Si può dimostrare (cosa che non faremo) che la somma di due distribuzioni di Cauchy indipendenti è ancora una distribuzione di Cauchy con parametro di posizione uguale alla somma dei parametri di posizione e parametro di scala pari alla somma dei parametri di scala

$$x_1 \sim \text{Cauchy}(x_{01}, \gamma_1) \quad \text{e} \quad x_2 \sim \text{Cauchy}(x_{02}, \gamma_2) \rightarrow x_1 + x_2 \sim \text{Cauchy}(x_{01} + x_{02}, \gamma_1 + \gamma_2). \quad (223)$$

Può sembrare una proprietà banale, simile a quella che abbiamo già visto, ad esempio, per la distribuzione di Poisson, ma c'è una differenza fondamentale: dato che il parametro di scala coincide con la semilarghezza a metà altezza, stiamo dicendo che qui *le larghezze delle distribuzioni si sommano linearmente e non in quadratura*. Nel linguaggio delle distribuzioni ordinarie sarebbe come dire che nella somma di due variabili casuali indipendenti non si sommano le varianze, ma le deviazioni standard. È esattamente ciò che vediamo in figura 7.6: la somma di  $n$  distribuzioni di Cauchy in forma standard è una distribuzione di Cauchy con parametro di posizione  $x_0 = 0$  e parametro di scala  $\gamma = n$ —mentre la deviazione standard della distribuzione di Gauss che sarebbe predetta dal teorema centrale del limite cresce solo come  $\sqrt{n}$ .

Una conseguenza immediata (e per molti sorprendente) di questa proprietà è che la media di un numero arbitrario di campionamenti di una distribuzione di Cauchy è distribuita secondo una distribuzione di Cauchy identica a quella della variabile di partenza (e.g., la media di un numero arbitrario di campionamenti di una variabile di Cauchy in forma standard è distribuita ancora come una variabile di Cauchy in forma standard). Detto in altre parole la formula per la deviazione standard della media che abbiamo usato diffusamente fino a questo momento non vale (cosa che in fondo non sorprende poiché in questo contesto la deviazione standard non è definita) e se prendiamo il parametro di scala  $\gamma$  come incertezza di misura siamo costretti ad ammettere che l'errore sul singolo campionamento e l'errore sulla media sono la stessa cosa.

La distribuzione di Cauchy è spesso utilizzata come esempio di distribuzione patologica, ed ha un certo numero di altre proprietà interessanti, ma è arrivato il momento di andare avanti e cambiare argomento.

## 7.6 ANCORA SUL *random walk*

Nella sezione 4.2.2 abbiamo descritto sommariamente le peculiarità di base del problema del *random walk*. Ci torniamo sopra brevemente alla luce di ciò che abbiamo imparato nel frattempo: il teorema centrale del limite e i cambiamenti di variabile per variabili casuali.

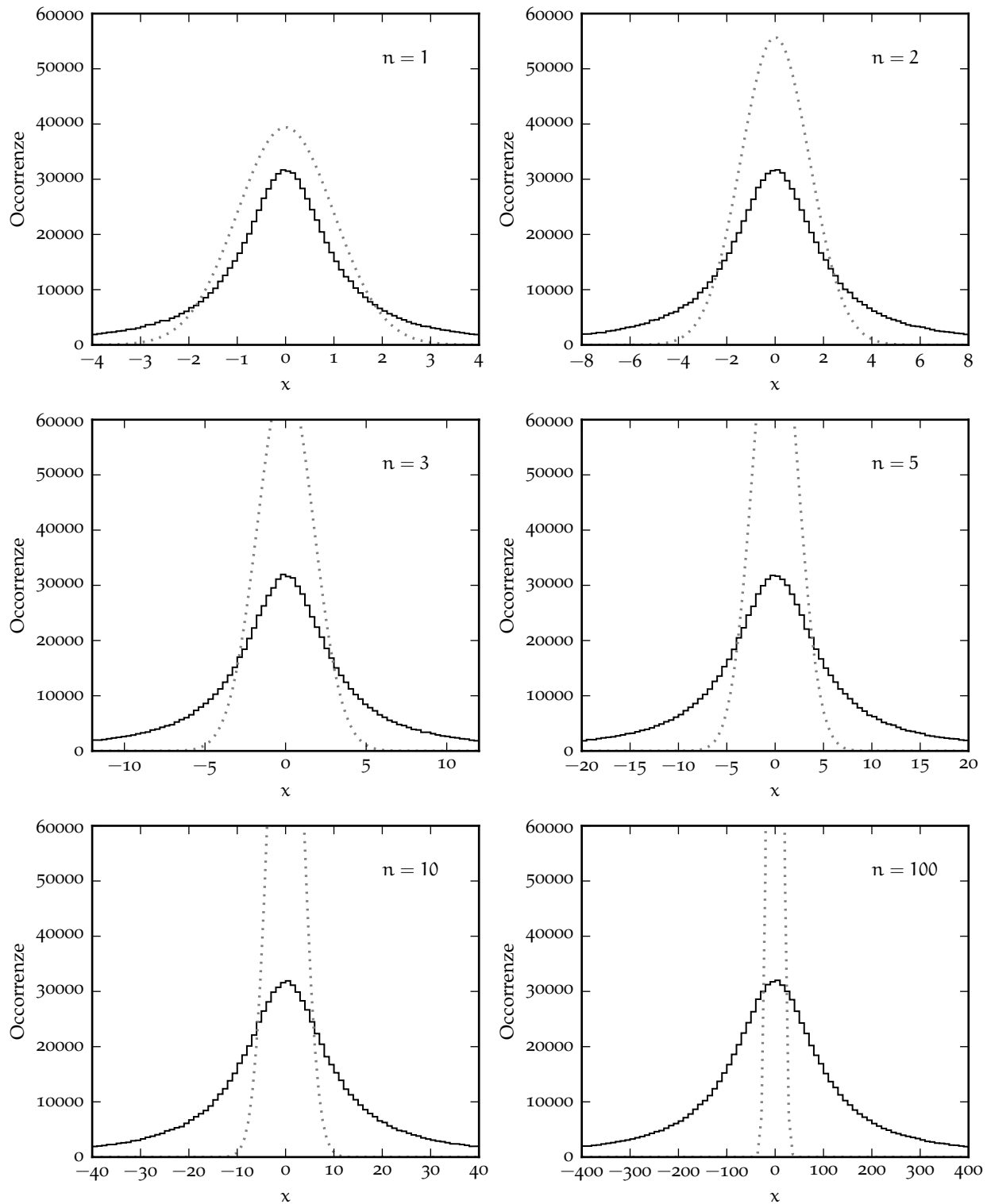


FIGURA 7.6. Illustrazione del fallimento del teorema centrale del limite nel caso della somma di  $n$  distribuzioni di Cauchy in forma standard. In ciascun grafico gli istogrammi si riferiscono ad un milione di campionamenti indipendenti della somma in questione, mentre linea grigia tratteggiata rappresenta una distribuzione di Gauss con media 0 e varianza  $n$ . Si noti come, a parte la scala sull'asse delle ascisse, gli istogrammi siano essenzialmente identici l'uno all'altro, indipendentemente dal valore di  $n$ , al contrario di ciò che accade nelle figure 5.12 e 5.13.

Nelle nostra ipotesi (ovverosia passi elementari di lunghezza  $\ell$  con probabilità  $1/2$  di muoversi verso destra o verso sinistra) avavamo ottenuto un'espressione esplicita per la media e la varianza della variabile casuale che rappresenta la coordinata  $x_n$  in cui si trova il nostro punto dopo  $n$  passi:

$$E[x_n] = 0 \quad \text{e} \quad \text{Var}(x_n) = n\ell^2.$$

Ora, il teorema centrale del limite ci garantisce che per  $n$  abbastanza grande la funzione di distribuzione di  $x_n$  è una Gaussiana con media e varianza date dalle espressioni che abbiamo appena scritto

$$p_{x_n}(x_n; \ell) = \frac{1}{\ell\sqrt{2\pi n}} e^{-\frac{x_n^2}{2n\ell^2}}.$$

ed il problema che dobbiamo affrontare è il calcolo esplicito della funzione di distribuzione della *distanza*

$$d_n = f(x_n) = |x_n| \quad (224)$$

del punto dall'origine dopo il passo  $n$ -esimo.

Procediamo con ordine. Esattamente come nel caso della trasformazione  $f(x) = x^2$  che abbiamo visto nella sezione 7.3 il supporto della variabile trasformata  $d_n$  sarà la semiretta reale positiva e per ogni valore di  $d_n$  ci saranno esattamente due valori di  $x_n$  (per la precisione  $-d_n$  e  $d_n$ ) che vengono trasformati in  $d_n$  dalla (224). Se ci restringiamo alla semiretta reale positiva abbiamo banalmente

$$x_n = f^{-1}(d_n) = d_n \quad \text{da cui} \quad \frac{df^{-1}(d_n)}{dd_n} = 1.$$

La densità di probabilità cercata sarà dunque—a meno di un fattore 2 dovuto a quanto detto sopra—uguale alla densità di probabilità della variabile di partenza, espressa in funzione della variabile trasformata e ristretta al dominio opportuno

$$p_{d_n}(d_n; \ell) = \sqrt{\frac{2}{\pi n\ell^2}} e^{-\frac{d_n^2}{2n\ell^2}} \quad \text{con} \quad d_n \geq 0. \quad (225)$$

A questo punto il valor medio di  $d_n$  si calcola con i nostri strumenti consueti, vale a dire

$$E[d_n] = \int_0^\infty d_n p_{d_n}(d_n; \ell) dd_n = \sqrt{\frac{2}{\pi n\ell^2}} \int_0^\infty d_n e^{-\frac{d_n^2}{2n\ell^2}} dd_n = \sqrt{\frac{2}{\pi n\ell^2}} \times n\ell^2 \int_0^\infty e^{-t} dt = \sqrt{\frac{2}{\pi}} \ell \sqrt{n} \quad (226)$$

che fissa finalmente il coefficiente di proporzionalità ( $\sqrt{2/\pi}$ ) che avevamo lasciato indefinito nella (126).



---

## METODI DI FIT

---

In questo capitolo riprendiamo, alla luce di quanto imparato fino a questo momento, il problema del fit di dati sperimentali, che abbiamo già incontrato nella sua variante grafica, e per modelli lineari o linearizzabili, nel capitolo 2.

### 8.1 DESCRIZIONE GENERALE DEL PROBLEMA

Come abbiamo già detto, dato un insieme di dati sperimentali, il problema di *fit* o *best fit* o *fitting* consiste nel cercare una funzione (o *modello*), possibilmente semplice, che descriva in modo soddisfacente i dati stessi. Vi sono due aspetti, diversi e ugualmente importanti, di questo problema. Il primo è—data una famiglia di modelli, ad esempio dipendente da uno o più parametri—trovare il modello specifico che meglio si adatta ai nostri dati secondo una qualche metrica ragionevole. Il secondo è valutare a posteriori il livello di accordo (o la bontà del fit) per capire se il nostro modello specifico è un buon modello oppure no. In questo capitolo cercheremo di sviscerare per quanto possibile entrambi questi aspetti.

Cominciamo con l'inquadrare in modo un po' più rigoroso gli ingredienti del problema. Data una serie di  $n$  coppie ordinate di misure di due generiche grandezze  $x$  ed  $y$ , con le rispettive incertezze

$$(x_i \pm \sigma_{x_i}, y_i \pm \sigma_{y_i}) \quad i = 1 \dots n$$

ed una famiglia di funzioni, dipendente da un numero  $m$  di parametri  $\theta_1 \dots \theta_m$ , che pensiamo possa rappresentare la dipendenza di  $y$  da  $x$

$$y = f(x; \theta_1 \dots \theta_m)$$

il nostro primo compito è dunque quello di trovare i valori  $\hat{\theta}_1 \dots \hat{\theta}_m$  dei parametri in modo da massimizzare (in un senso che preciseremo tra un attimo) l'accordo tra modello e dati.

► **ESEMPIO 8.1.** La famiglia di funzioni  $y = mx + q$  (che al variare di  $m$  e  $q$  genera tutte le rette del piano) è un modello a due parametri—che si dice *modello lineare*. È uno dei più semplici ad avere rilevanza pratica e ci torneremo in dettaglio nel seguito.

Vale la pena sottolineare subito che la scelta del modello (cioè della funzione di fit) è un passo essenziale del processo. In generale, dati  $n$  punti sperimentali, è banalmente possibile trovare un polinomio di grado  $n - 1$  che passi esattamente per tutti i punti, ma ciò non significa che questa sia una strategia di fit sensata. Anzi, esattamente il contrario. Per prima cosa, così facendo, ignoreremmo completamente le incertezze di misura, che sappiamo invece essere un ingrediente fondamentale di tutta la nostra costruzione. Inoltre il nostro fit ci restituirebbe un numero di valori pari a quello dei nostri punti sperimentali—senza riduzione di informazione e senza nessuna ovvia interpretazione fisica dei parametri. (In altre parole, è importante capire fin da subito che i problemi di fit sono concettualmente diversi dai problemi di interpolazione—qui stiamo cercando di fare qualcosa di leggermente più sofisticato che non il semplice far passare una curva da una serie di punti.) È chiaro allora che il modello va scelto in base a considerazioni di carattere fisico (ad esempio previsioni della teoria, o analogia con situazioni simili) che ci fanno preferire una famiglia di funzioni rispetto ad un'altra, e che per quanto possibile è bene che i parametri del modello siano riconducibili alla Fisica del sistema che stiamo studiando.

► **ESEMPIO 8.2.** Supponiamo di aver misurato la posizione di un punto materiale vincolato a muoversi su una retta in corrispondenza di un certo numero di istanti di tempo fissati e di essere interessati a determinare la legge oraria del punto stesso. Se abbiamo ragioni per pensare che il nostro oggetto si muova di moto uniforme possiamo usare un modello lineare. E se il modello lineare fornisce un buon fit, l'interpretazione fisica dei parametri di *best fit* è immediata: l'intercetta corrisponde alla posizione al tempo  $t = 0$  ed il coefficiente angolare alla velocità.

## 8.2 INTRODUZIONE INFORMALE AL FIT DEI MINIMI QUADRATI

Continuiamo la nostra discussione attaccando un problema leggermente più semplice di quello generale appena descritto, ma purtuttavia di grande rilevanza pratica. Faremo le tre ipotesi di lavoro seguenti:

1. le  $n$  misure sono tra loro indipendenti;
2. i valori  $y_i$  sono misurati in corrispondenza di  $x_i$  noti, ovverosia le incertezze di misura sugli  $x_i$  sono trascurabili;
3. gli errori di misura sulla variabile dipendente sono Gaussiani—cioè gli  $y_i$  tendono a fluttuare attorno al misurando con una distribuzione Gaussiana con deviazione standard  $\sigma_{y_i}$  nota a priori.

Prima ancora di cominciare la seconda delle assunzioni richiede un chiarimento. Che cosa vuol dire che gli errori sulla  $x$  sono trascurabili e come possiamo scrivere questa condizione? Evidentemente *non* come  $\sigma_{x_i} \ll \sigma_{y_i}$ —che in generale non è corretta nemmeno dimensionalmente, poiché  $x$  ed  $y$  non sono necessariamente grandezze omogenee. Ma nemmeno come  $\sigma_{x_i}/x_i \ll \sigma_{y_i}/y_i$  poiché una differenza relativa piccola sulla  $x$  può trasformarsi in una differenza sostanziale sulla  $y$ , se la derivata della funzione che lega tra loro la  $x$  e la  $y$  in un certo punto è abbastanza grande. Dobbiamo richiedere invece che per ciascuno dei punti misurati l'errore  $\sigma_{x_i}$ , propagato attraverso la funzione  $f(x)$  sulla variabile dipendente  $y$  sia molto più piccolo dell'errore di misura  $\sigma_{y_i}$  corrispondente o, in formule,

$$\left| \frac{df(x_i; \hat{\theta}_1 \dots \hat{\theta}_m)}{dx} \right| \sigma_{x_i} \ll \sigma_{y_i} \quad i = 1 \dots n. \quad (227)$$

(Se siete incerti su questo punto, tornate di nuovo a leggere la sezione 1.5 ed osservate con attenzione la figura 1.3.) Vedremo tra un attimo che questa condizione semplifica significativamente le cose, poiché nel confronto tra dati e modello ci permette di calcolare quest'ultimo in una serie punti ben definiti senza preoccuparci dell'errore sulla variabile indipendente.

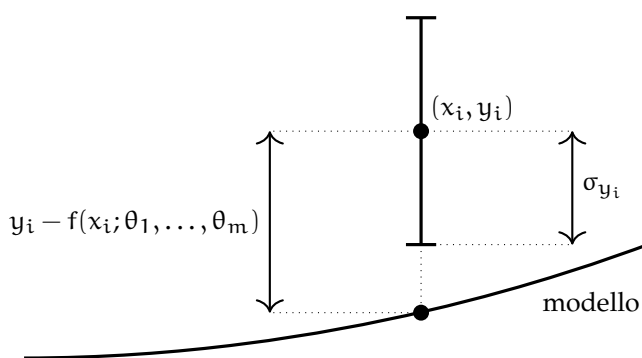


FIGURA 8.1. Significato geometrico dell'iesimo termine della somma (228): il quadrato della distanza tra dati e modello (per un insieme di valori dei parametri) misurate in barre d'errore.

Cominciamo dunque con il costruire la somma, che per motivi che saranno chiari tra un attimo chiameremo *chi quadro*

$$\chi^2(\theta_1, \dots, \theta_m) = \sum_{i=1}^n \left( \frac{y_i - f(x_i; \theta_1, \dots, \theta_m)}{\sigma_{y_i}} \right)^2. \quad (228)$$

A questo livello il  $\chi^2$  non è altro che la *somma dei quadrati delle differenze tra dati e modello, misurate in unità di barre d'errore*. (Non sarete sorpresi del fatto che abbiamo elevato al quadrato le distanze—al solito è per

evitare che fluttuazioni positive fluttuazioni negative si cancellino a vicenda.) È chiaro allora che questa quantità misura, in un qualche senso, la distanza complessiva tra una serie di dati ed un modello, in funzione del valore dei parametri da cui il modello dipende: più piccolo è il  $\chi^2$ , migliore l'accordo tra dati e modello—e viceversa.

Ora che abbiamo una misura oggettiva dell'accordo tra dati e modello, possiamo trovare la funzione di *best fit* semplicemente minimizzando il nostro  $\chi^2$  rispetto agli  $m$  parametri del modello:

$$\begin{cases} \frac{\partial \chi^2(\hat{\theta}_1, \dots, \hat{\theta}_m)}{\partial \theta_1} = 0 \\ \vdots \\ \frac{\partial \chi^2(\hat{\theta}_1, \dots, \hat{\theta}_m)}{\partial \theta_m} = 0. \end{cases} \quad (229)$$

La (229) fornisce un insieme di  $m$  equazioni che, almeno in linea di principio, ci permettono di ricavare gli  $m$  valori di *best fit*  $\hat{\theta}_1, \dots, \hat{\theta}_m$  per i parametri.

### 8.2.1 Fit dei minimi quadrati nel caso costante: la media pesata

L'esempio più semplice di fit ad un parametro (che chiameremo, per fissare le idee,  $q$ ) è quello con una funzione costante—cioè indipendente dalla variabile  $x$ :

$$f(x; q) = q \quad (\text{costante}).$$

Illustreremo questo primo caso pedissequamente in ogni passaggio. Il  $\chi^2$  si scrive come

$$\chi^2(q) = \sum_{i=1}^n \left( \frac{y_i - q}{\sigma_{y_i}} \right)^2.$$

(Notiamo esplicitamente che in questo caso le nostre  $n$  misure sperimentali non saranno coppie ordinate  $(x_i, y_i)$  ma singoli valori  $y_i$ , poiché nel modello la variabile indipendente non gioca alcun ruolo.) La condizione di minimo si scrive ponendo uguale a zero la derivata totale del  $\chi^2$  rispetto all'unico parametro

$$\frac{d\chi^2(q)}{dq} = -2 \sum_{i=1}^n \left( \frac{y_i - \hat{q}}{\sigma_{y_i}^2} \right) = 0. \quad (230)$$

Si tratta di un'equazione lineare in  $\hat{q}$  che si risolve banalmente, con poche manipolazioni algebriche, nella forma

$$\hat{q} = \frac{\sum_{i=1}^n \frac{y_i}{\sigma_{y_i}^2}}{\sum_{i=1}^n \frac{1}{\sigma_{y_i}^2}}. \quad (231)$$

L'espressione (231) si chiama *media pesata* delle  $y_i$  e costituisce la procedura statisticamente corretta con cui si combinano misure indipendenti della stessa quantità—ad esempio risultati di esperimenti diversi che misurano la stessa grandezza. Il motivo del nome sta nel fatto che se introduciamo le quantità (che chiameremo *pesi*)

$$w_i = \frac{1}{\sigma_{y_i}^2},$$

allora la (231) può essere riscritta nella forma equivalente, ma più compatta

$$\hat{q} = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}, \quad (232)$$

in cui la struttura della media pesata è più facilmente riconoscibile. Impariamo, come corollario, che i pesi statisticamente corretti da assegnare alle misure sono dunque gli inversi dei quadrati delle rispettive

incertezze—che, almeno a livello qualitativo, ha senso in quanto le misure con le incertezze più grandi debbono contare meno nella media.

A questo punto ci rimane da stimare l'incertezza da associare alla nostra stima  $\hat{q}$  di  $q$ . Formalmente dobbiamo, cioè, calcolare la varianza di  $\hat{q}$

$$\begin{aligned}\sigma_q^2 &= \text{Var} \left( \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i} \right) = \frac{\text{Var} \left( \sum_{i=1}^n w_i y_i \right)}{\left( \sum_{i=1}^n w_i \right)^2} = \frac{\sum_{i=1}^n \text{Var} (w_i y_i)}{\left( \sum_{i=1}^n w_i \right)^2} = \frac{\sum_{i=1}^n w_i^2 \text{Var} (y_i)}{\left( \sum_{i=1}^n w_i \right)^2} = \\ &= \frac{\sum_{i=1}^n w_i}{\left( \sum_{i=1}^n w_i \right)^2} = \frac{1}{\sum_{i=1}^n w_i} = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_{y_i}^2}}.\end{aligned}\quad (233)$$

Il calcolo non è esattamente banale, ma la cosa fondamentale da ricordare è che, in questo contesto, le  $y_i$  sono le variabili casuali mentre i  $\sigma_{y_i}$ , e di conseguenza i  $w_i$ , sono per ipotesi costanti note a priori, per cui passano fuori (elevate al quadrato, naturalmente) dall'operatore varianza. Detto questo motiviamo brevemente (ed in ordine) ciascun singolo passaggio:

1. abbiamo sfruttato il fatto che  $\text{Var}(cx) = c^2 \text{Var}(x)$ , cfr. (77);
2. la varianza della somma è pari alla somma delle varianze poiché le  $y_i$  sono indipendenti, cfr. (123);
3. abbiamo sfruttato di nuovo il fatto che  $\text{Var}(cx) = c^2 \text{Var}(x)$ , cfr. (77);
4. semplice sostituzione:  $\text{Var}(y_i) = \sigma_{y_i}^2 = 1/w_i$ .

In altre parole il quadrato dell'incertezza su  $\hat{q}$  è l'inverso della somma degli inversi delle incertezze (al quadrato) sulle misure di partenza. Non è difficile convincersi che  $\sigma_q$  è strettamente minore di ciascun  $\sigma_{y_i}$ —come è lecito aspettarsi sulla base del fatto che tutte le misure, più o meno a seconda dell'incertezza associata, contribuiscono informazione.

► **ESEMPIO 8.3.** Cinque gruppi sperimentali (che chiameremo A...E) misurano indipendentemente l'indice di rifrazione dell'acqua ottenendo i valori:  $n_A = 1.325 \pm 0.012$ ,  $n_B = 1.36 \pm 0.05$ ,  $n_C = 1.32 \pm 0.01$ ,  $n_D = 1.338 \pm 0.005$ ,  $n_E = 1.335 \pm 0.006$ . Se si vogliono combinare le misure per fornire una stima  $\hat{n}$  del misurando che includa in modo consistente tutta l'informazione disponibile il modo corretto di procedere è quello di eseguire una media pesata

$$\hat{n} = 1.3339 \pm 0.0034 \quad \text{oppure} \quad \hat{n} = 1.334 \pm 0.003.$$

Il calcolo è implementato nel frammento 8.1 ed il risultato è illustrato graficamente nella figura 8.2.

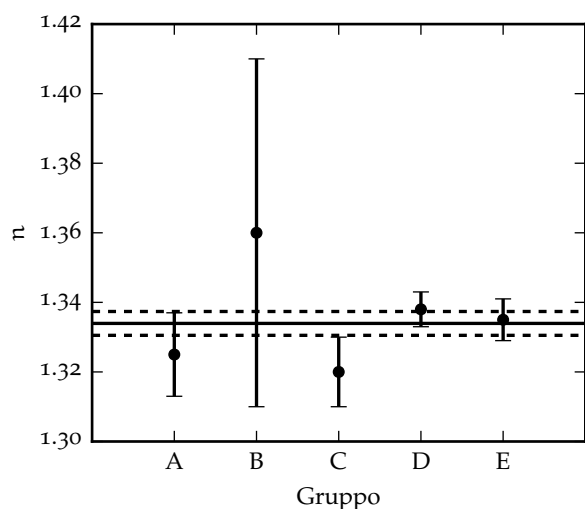


FIGURA 8.2. Illustrazione del calcolo della media pesata dei valori dell'indice di rifrazione nell'esempio 8.3. La linea continua rappresenta il valore della media pesata in questione e le due linee tratteggiate indicano l'incertezza (a  $1\sigma$ ) associata. Notiamo esplicitamente che la larghezza della banda di incertezza è più piccola della più piccola tra le incertezze di misura (quella del gruppo D).



```

1  https://bitbucket.org/.../weighted_average.py
2  def weighted_average(y, dy):
3      """Weighted average of a list of measurement.
4      """
5      S = Sy = 0.
6      for yi, dyi in zip(y, dy):
7          wi = (1./dyi)**2.
8          S += wi
9          Sy += wi*yi
10     q = Sy/S
11     dq = (1./S)**0.5
12     return q, dq
13
14 n = [1.325, 1.36, 1.32, 1.338, 1.335]
15 dn = [0.012, 0.05, 0.01, 0.005, 0.006]
16 print('q = %.4f +/- %.4f' % weighted_average(n, dn))
17
18 [Output]
19 q = 1.3339 +/- 0.0034

```

FRAMMENTO 8.1. Frammento di codice per il calcolo della media pesata delle misure dell'indice di rifrazione dell'acqua dell'esempio 8.3. La funzione prende in ingresso le liste dei valori delle misure e degli errori associati e restituisce la media pesata e l'incertezza associata. Il risultato è illustrato graficamente nella figura 8.2. Notiamo, per completezza, che l'incertezza  $\sigma_q$  sulla media pesata è più piccola del più piccolo tra gli errori di misura.

### 8.2.2 Fit dei minimi quadrati nel caso lineare

Il fit dei minimi quadrati nel caso lineare non è solo un esercizio di notevole importanza pratica; essendo uno degli esempi più semplici di fit a più di un parametro permette di discutere alcuni aspetti del problema generale della stima dei parametri che non emergono nella media pesata. Il nostro modello è in questo caso

$$f(x; q, m) = mx + q$$

da cui

$$\chi^2(q, m) = \sum_{i=1}^n \left( \frac{y_i - mx_i - q}{\sigma_{y_i}} \right)^2.$$

Come anticipato, questa volta abbiamo due condizioni di minimo per ricavare i valori dei due parametri

$$\begin{cases} \frac{\partial \chi^2(\hat{q}, \hat{m})}{\partial q} = -2 \sum_{i=1}^n \left( \frac{y_i - \hat{m}x_i - \hat{q}}{\sigma_{y_i}^2} \right) = 0 \\ \frac{\partial \chi^2(\hat{q}, \hat{m})}{\partial m} = -2 \sum_{i=1}^n \left( \frac{y_i - \hat{m}x_i - \hat{q}}{\sigma_{y_i}^2} \right) x_i = 0 \end{cases} \quad (234)$$

Allo scopo di semplificare la notazione per il calcolo esplicito della soluzione, introduciamo alcune quantità—che dipendono solamente dalle misure sperimentali e *non* dai valori dei parametri del modello, ed i cui valori numerici possono essere calcolati banalmente una volta che abbiamo a disposizione le misure stesse:

$$S_x^0 = \sum_{i=1}^n \frac{1}{\sigma_{y_i}^2} \quad S_x^1 = \sum_{i=1}^n \frac{x_i}{\sigma_{y_i}^2} \quad S_x^2 = \sum_{i=1}^n \frac{x_i^2}{\sigma_{y_i}^2} \quad S_{xy}^0 = \sum_{i=1}^n \frac{y_i}{\sigma_{y_i}^2} \quad S_{xy}^1 = \sum_{i=1}^n \frac{x_i y_i}{\sigma_{y_i}^2}, \quad (235)$$

col che possiamo allora riscrivere le condizioni di minimo come

$$\begin{cases} \hat{q} S_x^0 + \hat{m} S_x^1 = S_{xy}^0 \\ \hat{q} S_x^1 + \hat{m} S_x^2 = S_{xy}^1 \end{cases} \quad \text{ovvero, in forma matriciale} \quad \begin{bmatrix} S_x^0 & S_x^1 \\ S_x^1 & S_x^2 \end{bmatrix} \begin{bmatrix} \hat{q} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} S_{xy}^0 \\ S_{xy}^1 \end{bmatrix}.$$

L'algebra è tediosa, ma concettualmente il calcolo procede esattamente come per la media pesata nella sezione precedente per cui forniamo direttamente i risultati. Detta  $D$  la quantità

$$D = S_x^0 S_x^2 - (S_x^1)^2$$

(che non è altro che il determinante della matrice  $2 \times 2$  corrispondente al nostro sistema lineare) si ha:

$$\begin{cases} \hat{q} = \frac{S_{xy}^0 S_x^2 - S_{xy}^1 S_x^1}{D} \\ \hat{m} = \frac{S_{xy}^1 S_x^0 - S_{xy}^0 S_x^1}{D} \end{cases} \quad \text{con errori associati} \quad \begin{cases} \sigma_q^2 = \frac{S_x^2}{D} \\ \sigma_m^2 = \frac{S_x^0}{D} \end{cases} \quad (236)$$

Notiamo esplicitamente che, ai fini del calcolo delle incertezze, le uniche somme che dipendono dalle nostre variabili casuali  $y_i$  sono  $S_{xy}^0$  ed  $S_{xy}^1$ —le altre tre sono semplici *costanti* e possono essere trattate come tali nel calcolo delle varianze. Per completezza le quantità rilevanti sono

$$\text{Var}(S_{xy}^0) = \text{Var}\left(\sum_{i=1}^n \frac{y_i}{\sigma_{y_i}^2}\right) = \sum_{i=1}^n \text{Var}\left(\frac{y_i}{\sigma_{y_i}^2}\right) = \sum_{i=1}^n \frac{\text{Var}(y_i)}{\sigma_{y_i}^4} = \sum_{i=1}^n \frac{1}{\sigma_{y_i}^2} = S_x^0 \quad (237)$$

$$\text{Var}(S_{xy}^1) = \text{Var}\left(\sum_{i=1}^n \frac{x_i y_i}{\sigma_{y_i}^2}\right) = \sum_{i=1}^n \text{Var}\left(\frac{x_i y_i}{\sigma_{y_i}^2}\right) = \sum_{i=1}^n \frac{x_i^2 \text{Var}(y_i)}{\sigma_{y_i}^4} = \sum_{i=1}^n \frac{x_i^2}{\sigma_{y_i}^2} = S_x^2 \quad (238)$$

$$\text{Cov}(S_{xy}^0, S_{xy}^1) = \text{Cov}\left(\sum_{i=1}^n \frac{y_i}{\sigma_{y_i}^2}, \sum_{j=1}^n \frac{x_j y_j}{\sigma_{y_j}^2}\right) = \sum_{i,j=1}^n \frac{x_j}{\sigma_{y_i}^2 \sigma_{y_j}^2} \text{Cov}(y_i, y_j) = \sum_{i=1}^n \frac{x_i}{\sigma_{y_i}^2} = S_x^1, \quad (239)$$

(nell'ultima relazione abbiamo utilizzato il fatto che  $\text{Cov}(y_i, y_j)$  si annulla per  $i \neq j$  ed è uguale a  $\sigma_{y_i}^2$  per  $i = j$ ). Per completezza calcoliamo esplicitamente l'incertezza su  $\hat{q}$ , lasciando quella su  $\hat{m}$  per esercizio

$$\begin{aligned} \sigma_q^2 &= \frac{1}{D^2} \left[ (S_x^2)^2 \text{Var}(S_{xy}^0) + (S_x^1)^2 \text{Var}(S_{xy}^1) - 2 S_x^2 S_x^1 \text{Cov}(S_{xy}^0, S_{xy}^1) \right] = \\ &= \frac{1}{D^2} \left[ (S_x^2)^2 S_x^0 + (S_x^1)^2 S_x^2 - 2 S_x^2 (S_x^1)^2 \right] = \frac{1}{D^2} \left[ (S_x^2)^2 S_x^0 - (S_x^1)^2 S_x^2 \right] = \frac{S_x^2}{D^2} \left[ S_x^2 S_x^0 - (S_x^1)^2 \right] = \frac{S_x^2}{D}. \end{aligned}$$

(L'unica cosa da ricordare qui è che  $\text{Var}(x_1 - x_2) = \text{Var}(x_1) + \text{Var}(x_2) - 2\text{Cov}(x_1, x_2)$ .) Il frammento di codice 8.2 mostra una implementazione, elementare ma funzionante, dell'algoritmo.

```

https://bitbucket.org/.../linear_least_squares.py
1 def linear_least_squares(x, y, dy):
2     """Linear least squares fit.
3     """
4     Sx0 = Sx1 = Sx2 = Sxy0 = Sxy1 = 0.
5     for xi, yi, dyi in zip(x, y, dy):
6         wi = (1./dyi)**2.
7         Sx0 += wi
8         Sx1 += wi*xi
9         Sx2 += wi*xi*xi
10        Sxy0 += wi*yi
11        Sxy1 += wi*xi*yi
12    D = Sx0*Sx2 - Sx1**2.
13    q = (Sxy0*Sx2 - Sxy1*Sx1)/D
14    dq = (Sx2/D)**0.5
15    m = (Sxy1*Sx0 - Sxy0*Sx1)/D
16    dm = (Sx0/D)**0.5
17    return q, dq, m, dm
18
19 x = [1., 2., 3., 4., 5.]
20 y = [10.42, 10.96, 14.50, 16.58, 19.41]
21 dy = [0.50]*5
22 print('q = %.2f +/- %.2f, m = %.2f +/- %.2f' %\
23       linear_least_squares(x, y, dy))
24
25 [Output]
26 q = 7.29 +/- 0.52, m = 2.36 +/- 0.16

```

FRAMMENTO 8.2. Esempio di implementazione (completamente funzionale) di un fit dei minimi quadrati con un modello lineare. Per chiarezza, i nomi delle variabili utilizzate ricalcano da vicino quelli usati nel testo.

### 8.2.3 La covarianza tra i parametri nel fit lineare

A questo punto possiamo chiederci se le quantità  $\hat{q}$  e  $\hat{m}$  che abbiamo appena calcolato siano *indipendenti* o meno—e la risposta è no: anche se le misure  $y_i$  da cui partiamo sono, per ipotesi, indipendenti l'intercetta ed il coefficiente angolare di *best fit* sono (negativamente) correlati. (Non stupirà il fatto che questa non è una peculiarità del fit dei minimi quadrati lineare ma è una proprietà che vale generale, a meno che non si utilizzino tecniche specifiche per evitare la correlazione tra i parametri di *best fit*.)

In effetti abbiamo già in mano tutti gli ingredienti che ci servono per calcolare la covarianza cercata, ovvero sia

$$\begin{aligned} \text{Cov}(\hat{q}, \hat{m}) &= \text{Cov}\left(\frac{S_{xy}^0 S_x^2 - S_{xy}^1 S_x^1}{D}, \frac{S_{xy}^1 S_x^0 - S_{xy}^0 S_x^1}{D}\right) = \frac{1}{D^2} \left[ S_x^0 S_x^2 \text{Cov}(S_{xy}^0, S_{xy}^1) + \right. \\ &\quad \left. - S_x^1 S_x^2 \text{Cov}(S_{xy}^0, S_{xy}^0) - S_x^0 S_x^1 \text{Cov}(S_{xy}^1, S_{xy}^1) + (S_x^1)^2 \text{Cov}(S_{xy}^1, S_{xy}^0) \right] = \\ &= \frac{1}{D^2} \left[ S_x^0 S_x^1 S_x^2 - S_x^0 S_x^1 S_x^2 - S_x^0 S_x^1 S_x^2 + (S_x^1)^3 \right] = -\frac{S_x^1 (S_x^0 S_x^2 - (S_x^1)^2)}{D^2} = -\frac{S_x^1}{D}. \end{aligned} \quad (240)$$

Ne segue banalmente che le matrici di covarianza e di correlazione tra i parametri del fit si scrivono come

$$\Sigma = \frac{1}{D} \begin{bmatrix} S_x^2 & -S_x^1 \\ -S_x^1 & S_x^0 \end{bmatrix} \quad \text{e} \quad R = \begin{bmatrix} 1 & -\frac{S_x^1}{\sqrt{S_x^0 S_x^2}} \\ -\frac{S_x^1}{\sqrt{S_x^0 S_x^2}} & 1 \end{bmatrix} \quad (241)$$

Che la correlazione tra intercetta e coefficiente angolare in un fit sia diversa da zero (e di segno negativo) non dovrebbe stupire—lo abbiamo già visto implicitamente parlando della propagazione dell'errore massimo ed è illustrato nuovamente in figura 8.3. Il livello di correlazione tra i parametri aumenta all'aumentare del braccio di leva del fit e nel limite in cui la dispersione tra i valori  $x_i$  è molto più piccola (in modulo) dei valori numerici degli  $x_i$  stessi—cioè quando nella variabile indipendente i punti sono molto più distanti dall'origine di quanto non lo siano tra loro stessi—si ha la situazione interessante:

$$x_i \approx \tilde{x} \quad i = 1 \dots n \quad \text{da cui} \quad S_x^1 \approx \tilde{x} S_x^0, \quad S_x^2 \approx \tilde{x}^2 S_x^0 \quad \text{e infine} \quad \text{Corr}(\hat{q}, \hat{m}) = -\frac{S_x^1}{\sqrt{S_x^0 S_x^2}} \rightarrow -1. \quad (242)$$

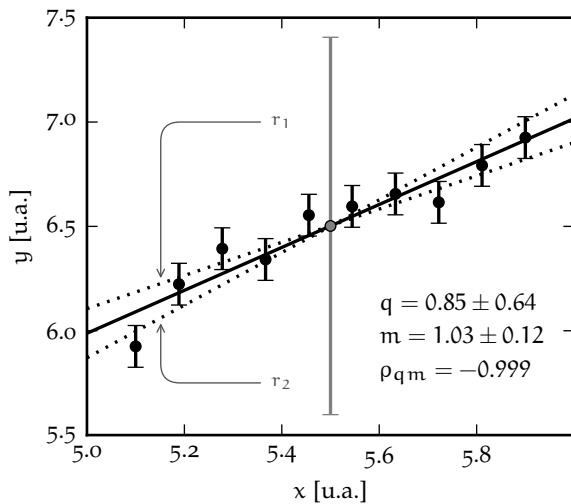


FIGURA 8.3. Illustrazione grafica della correlazione negativa tra intercetta e coefficiente angolare in un fit dei minimi quadrati con un modello lineare. La linea continua rappresenta la retta di *best fit*. Le linee tratteggiate  $r_1$  ed  $r_2$  mostrano che se incrementiamo (diminuiamo) il valore del coefficiente angolare, per far sì che la retta corrispondente continui ad essere un fit accettabile, siamo costretti a diminuire (incrementare) il valore del coefficiente angolare. La barra verticale al centro rappresenta l'incertezza (chiaramente assurda) sul valore del modello nel punto  $x = 5.5$  ottenuta propagando gli errori senza considerare la correlazione tra  $\hat{q}$  ed  $\hat{m}$ .

Dal punto di vista pratico la cosa è rilevante. Se vogliamo semplicemente riportare gli errori sui parametri le (236) sono sufficienti, ma se vogliamo propagare l'errore su una funzione dei parametri (ad esempio sull'interpolazione o l'estrapolazione della retta di *best fit*), allora il fatto che gli elementi fuori dalla diagonale della matrice di covarianza siano non nulli ci dice che la (142) non vale e siamo costretti ad utilizzare la (143), come illustrato nell'esempio 8.4.

► ESEMPIO 8.4. Con riferimento al fit lineare in figura 8.3 (che include i valori numerici dei parametri e delle incertezze associate) supponiamo di voler utilizzare il nostro modello di *best fit* per stimare l'incertezza di misura associata al valore della variabile indipendente  $f(x_0)$  in  $x_0 = 5.5$ . Se propagassimo banalmente gli errori come se  $\hat{q}$  ed  $\hat{m}$  fossero scorrelate avremmo

$$\sigma_f^2 = x_0^2 \sigma_m^2 + \sigma_q^2 \quad \text{ovvero} \quad \sigma_f = 0.90 \quad (243)$$

che è chiaramente assurdo, come illustrato dalla barra di errore grigia in corrispondenza di  $x_0$  nella figura 8.3. Il problema è che in questo caso il coefficiente di correlazione tra i parametri è prossimo a  $-1$  ed il risultato corretto è

$$\sigma_f^2 = x_0^2 \sigma_m^2 + \sigma_q^2 - 2x_0 \text{Corr}(\hat{q}, \hat{m}) \sigma_q \sigma_m \quad \text{ovvero} \quad \sigma_f = 0.030 \quad (244)$$

che è molto più ragionevole. La differenza tra i due numeri è un fattore 20—non stiamo parlando di una cosa da poco.

#### 8.2.4 Fit dei minimi quadrati con un polinomio di grado arbitrario



La discussione che abbiamo fatto nelle due sezioni precedenti può essere generalizzata senza enormi difficoltà (anche se i calcoli tendono a diventare complicati piuttosto rapidamente) al caso praticamente rilevante del fit dei minimi quadrati con un polinomio di grado  $k$  arbitrario—cioè del modello

$$f(x; a_0, a_1 \dots a_k) = a_0 + a_1 x + \dots + a_k x^k. \quad (245)$$

L'espressione per il  $\chi^2$  da minimizzare si scrive in questo caso come

$$\chi^2(a_0, a_1 \dots a_k) = \sum_{i=1}^n \left( \frac{y_i - (a_0 + a_1 x_i + \dots + a_k x_i^k)}{\sigma_{y_i}} \right)^2$$

e sua derivata rispetto al generico parametro  $a_j$  è banalmente

$$\frac{d\chi^2(a_0, a_1 \dots a_k)}{da_l} = -2 \sum_{i=1}^n \left( \frac{y_i - (a_0 + a_1 x_i + \dots + a_k x_i^k)}{\sigma_{y_i}^2} \right) x_i^l.$$

Le  $k+1$  condizioni di minimo (229) portano ad un sistema di  $k+1$  equazioni lineari nei parametri

$$\begin{cases} \hat{a}_0 \sum_{i=1}^n \frac{1}{\sigma_{y_i}^2} + \hat{a}_1 \sum_{i=1}^n \frac{x_i}{\sigma_{y_i}^2} + \dots + \hat{a}_k \sum_{i=1}^n \frac{x_i^k}{\sigma_{y_i}^2} = \sum_{i=1}^n \frac{y_i}{\sigma_{y_i}^2} \\ \hat{a}_0 \sum_{i=1}^n \frac{x_i}{\sigma_{y_i}^2} + \hat{a}_1 \sum_{i=1}^n \frac{x_i^2}{\sigma_{y_i}^2} + \dots + \hat{a}_k \sum_{i=1}^n \frac{x_i^{k+1}}{\sigma_{y_i}^2} = \sum_{i=1}^n \frac{x_i y_i}{\sigma_{y_i}^2} \\ \vdots \\ \hat{a}_0 \sum_{i=1}^n \frac{x_i^k}{\sigma_{y_i}^2} + \hat{a}_1 \sum_{i=1}^n \frac{x_i^{k+1}}{\sigma_{y_i}^2} + \dots + \hat{a}_k \sum_{i=1}^n \frac{x_i^{2k}}{\sigma_{y_i}^2} = \sum_{i=1}^n \frac{x_i^k y_i}{\sigma_{y_i}^2} \end{cases} \quad (246)$$

Definiamo allora, sul modello di ciò che abbiamo fatto nel caso del fit dei minimi quadrati lineare, le quantità

$$S_x^l = \sum_{i=1}^n \frac{x_i^l}{\sigma_{y_i}^2} \quad \text{e} \quad S_{xy}^l = \sum_{i=1}^n \frac{x_i^l y_i}{\sigma_{y_i}^2},$$

che, come prima, dipendono solamente dai dati e non dal modello—in altre parole, una volta che abbiamo le nostre misure, sono delle semplici somme il cui valore numerico può essere banalmente calcolato. Il nostro sistema di equazioni può allora essere riscritto in forma matriciale come

$$\begin{bmatrix} S_x^0 & S_x^1 & \dots & S_x^k \\ S_x^1 & S_x^2 & \dots & S_x^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ S_x^k & S_x^{k+1} & \dots & S_x^{2k} \end{bmatrix} \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_k \end{bmatrix} = \begin{bmatrix} S_{xy}^0 \\ S_{xy}^1 \\ \vdots \\ S_{xy}^k \end{bmatrix} \quad (247)$$

a la soluzione passa attraverso una inversione di matrice

$$\begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_k \end{bmatrix} = \begin{bmatrix} S_x^0 & S_x^1 & \dots & S_x^k \\ S_x^1 & S_x^2 & \dots & S_x^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ S_x^k & S_x^{k+1} & \dots & S_x^{2k} \end{bmatrix}^{-1} \begin{bmatrix} S_{xy}^0 \\ S_{xy}^1 \\ \vdots \\ S_{xy}^k \end{bmatrix} \quad (248)$$

che può essere eseguita utilizzando i metodi ordinari dell'algebra lineare. (Al di là della soluzione esplicita, la cosa essenziale da ricordare è che il fit dei minimi quadrati con un polinomio generico porta ad un insieme di equazioni lineari nei parametri che possono essere risolte in forma chiusa.)

### 8.2.5 Fit dei minimi quadrati non pesato

Quando le incertezze di misura  $\sigma_{y_i}$  non dipendono dall'indice  $i$  (cioè hanno lo stesso valore  $\sigma_y$  per tutti i punti sperimentali) possono passare fuori dal segno di sommatoria e si può utilizzare l'espressione

$$\sum_{i=1}^n (y_i - f(x_i; \theta_1, \dots, \theta_m))^2$$

come quantità da minimizzare nel fit. Va da sé che i  $\sigma_{y_i}$  sono ancora necessari per stimare le incertezze da associare ai parametri e per calcolare il  $\chi^2$  del fit, ma le nostre formule si semplificano sensibilmente. In questo caso si parla di fit dei minimi quadrati *non pesato*.

Così, per fissare le idee, i risultati che abbiamo ricavato per la media pesata nella sezione 8.2.1 divengono, nel caso non pesato

$$\hat{q} = \frac{1}{n} \sum_{i=1}^n y_i \quad \text{e} \quad \sigma_q = \frac{\sigma_y}{\sqrt{n}}, \quad (249)$$

cioè la miglior stima della nostra costante di fit  $q$  è data dalla media aritmetica delle misure  $y_i$  (il che non è sorprendente) e per quanto riguarda l'incertezza associata non abbiamo fatto altro che riscoprire per una via diversa la formula (131) per la deviazione standard della media.

## 8.3 IL TEST DEL $\chi^2$

Ora che abbiamo inquadrato il problema generale di *best fit*, e lo abbiamo risolto nei casi più semplici, è tempo di tornare alla seconda domanda con cui abbiamo aperto questo capitolo, ovvero: come si fa a decidere se un fit è soddisfacente o meno? È chiaro che, mentre il processo di minimizzazione del  $\chi^2$  ci fornisce i valori ottimali dei parametri, la risposta a questa altra domanda deve essere in qualche modo contenuta nel valore numerico del  $\chi^2$  stesso

$$\chi^2(\hat{\theta}_1, \dots, \hat{\theta}_m) = \sum_{i=1}^n \left( \frac{y_i - f(x_i; \hat{\theta}_1, \dots, \hat{\theta}_m)}{\sigma_{y_i}} \right)^2$$

che il fit restituisce in corrispondenza del minimo—ed in questa sezione cercheremo di capire come.

8.3.1 Il contenuto fisico del  $\chi^2$ : quanti gradi di libertà?

Se assumiamo per un attimo che il nostro modello costituisca una descrizione accurata della realtà e che le incertezze di misure siano stimate correttamente ogni singolo termine della somma (almeno nel caso in cui immaginiamo che esso sia calcolato in corrispondenza dei valori veri  $\bar{\theta}_1 \dots \bar{\theta}_m$  dei parametri)

$$\zeta_i = \frac{y_i - f(x_i; \bar{\theta}_1, \dots, \bar{\theta}_m)}{\sigma_{y_i}},$$

sarà distribuito per definizione come una variabile Gaussiana. Di più: come una Gaussiana in forma standard, nel senso che

$$\begin{aligned} E[\zeta_i] &= E\left[\frac{y_i - f(x_i; \bar{\theta}_1, \dots, \bar{\theta}_m)}{\sigma_{y_i}}\right] = \frac{1}{\sigma_{y_i}} E[y_i - f(x_i; \bar{\theta}_1 \dots \bar{\theta}_m)] = 0 \\ \text{Var}(\zeta_i) &= E[\zeta_i^2] = \frac{1}{\sigma_{y_i}^2} E[(y_i - f(x_i; \bar{\theta}_1 \dots \bar{\theta}_m))^2] = 1. \end{aligned}$$

Ora, dato che nella (228) ogni  $\zeta_i$  è elevato al quadrato, la nostra somma è, formalmente, una somma di  $n$  variabili Gaussiane in forma standard al quadrato, e noi abbiamo calcolato esplicitamente nella sezione 7.4 la sua funzione di distribuzione: una distribuzione del  $\chi^2$  ad  $n$  gradi di libertà. Sappiamo in particolare che

$$E[\chi^2(\bar{\theta}_1 \dots \bar{\theta}_m)] = n \quad \text{e} \quad \text{Var}(\chi^2(\bar{\theta}_1 \dots \bar{\theta}_m)) = 2n. \quad (250)$$

Possiamo chiederci se tutto questo valga anche se calcoliamo la nostra somma in corrispondenza dei valori di *best fit* dei parametri anziché nei valori (che peraltro non conosciamo) veri. Ebbene, sotto ipotesi relativamente deboli la quantità  $\chi^2(\hat{\theta}_1 \dots \hat{\theta}_n)$  è ancora distribuita come un  $\chi^2$  (cosa che non dimostreremo), ma il numero rilevante di gradi di libertà non è, in generale, pari al numero  $n$  di misure. Cerchiamo di capire perché.

Supponiamo di fittare la stessa serie di dati con un modello costante  $y = q$  ed un modello lineare  $y = mx + q$ . Ora, è ovvio che, poiché il modello lineare contiene al suo interno quello costante come caso particolare, esso è in grado di generare, tra le altre, tutte le funzioni costanti che il primo modello è in grado di generare. In altre parole il minimo  $\chi^2$  restituito da un fit lineare sarà per definizione più piccolo (o al massimo uguale) a quello restituito da un fit con una costante. Si tratta di una prima realizzazione importante: il numero di gradi di libertà del nostro problema non può dipendere solo dal numero di punti che fittiamo—in qualche modo deve dipendere anche dal numero di parametri liberi che utilizziamo nel fit. D'altra parte, se torniamo per un attimo al caso limite del fit con un polinomio di grado  $n - 1$ , che per definizione passa per tutti i punti, è chiaro che in quel caso il  $\chi^2$  del fit è identicamente nullo ed il nostro numero di gradi di libertà deve essere 0.

Cambiamo per un attimo prospettiva. Da un punto di vista dei principi primi, il numero di gradi di libertà del nostro problema è, per definizione, pari al numero di termini indipendenti nella somma (228). Allora è chiaro che per ogni parametro che fittiamo dai nostri dati introduciamo un vincolo—cioè una relazione tra i dati stessi—che porta alla diminuzione di un'unità nel numero di gradi di libertà<sup>1</sup>. (Se avete l'impressione che questo vi ricordi qualcosa che abbiamo già visto, è buon segno—rileggete con attenzione quello che abbiamo detto a proposito della stima imparziale della varianza campione nella sezione 4.3, perché il parallelo con il fatto che la sostituzione di  $\mu$  con  $m$  nel numeratore della formula per la varianza stessa richiedesse la sostituzione di  $n$  con  $n - 1$  al denominatore è perfettamente calzante). In definitiva *il numero di gradi di libertà è pari al numero dei punti sperimentali  $n$  meno il numero di parametri  $m$  che stimiamo dal fit*

$$\nu = n - m. \quad (251)$$

Vedremo nella sezione 8.3.4 che nel caso del test del  $\chi^2$  per una distribuzione ci sarà un altro piccolo colpo di scena, ma per il momento questo ci basta.

<sup>1</sup> Ad essere precisi, questo è tecnicamente vero solo nel caso in cui il vincolo sia di tipo lineare, ma si tratta di un dettaglio che è rilevante solo raramente, in pratica.

► **ESEMPIO 8.5.** Supponiamo di avere una serie di  $n$  misure  $y_i$ , con relative incertezze, e di voler verificare con un test del  $\chi^2$  l'ipotesi che esse siano compatibili con una costante  $q$ . Se questa costante è nota a priori (ad esempio dalla teoria), allora il numero di gradi di libertà del problema è  $\nu = m$ . Se, viceversa, abbiamo stimato  $q$  a partire dai dati, ad esempio mediante un fit dei minimi quadrati, allora la (231) ci permette di calcolare il valore di uno qualsiasi degli  $y_i$ , una volta noti i valori di tutti gli altri (oltre ovviamente a  $\hat{q}$ ). A questo punto gli  $n$  termini della somma (228) non sono più tutti indipendenti l'uno dall'altro, ed il numero di gradi di libertà è  $\nu = n - 1$ .

### 8.3.2 Il test del $\chi^2$ per una serie di dati

Siamo finalmente pronti per cominciare. In figura 8.4 è mostrato un fit ad una serie di 16 punti sperimentali con un modello quadratico  $f(x) = lx^2 + mx + q$ . Il modello ha 3 parametri per cui sappiamo che il numero di gradi di libertà del problema è  $\nu = 16 - 3 = 13$ .

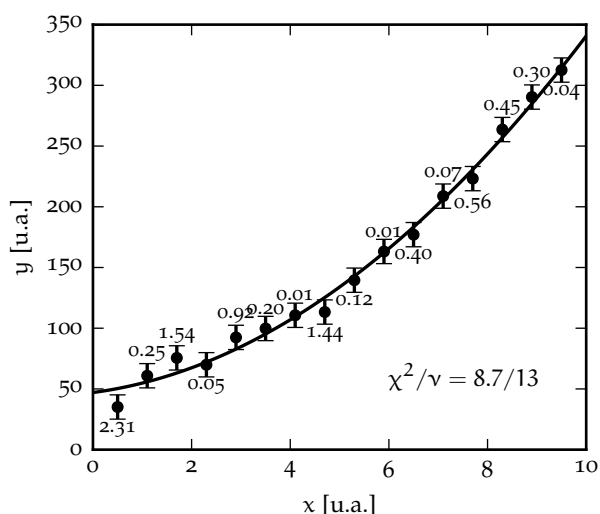


FIGURA 8.4. Esempio di fit ad una serie di dati con un modello quadratico  $f(x) = lx^2 + mx + q$ . Il modello ha 3 parametri e la serie 16 punti per cui il numero di gradi di libertà è  $\nu = 16 - 3 = 13$ . Il valore del  $\chi^2$  restituito dal fit è 8.7. Per completezza i contributi di ciascun punto alla somma (228) sono riportati sul grafico accanto ai punti stessi—e si tratta di un esercizio che è bene abituarsi a fare “ad occhio”, perché spesso è utile saper stimare approssimativamente il  $\chi^2$  di un fit senza un calcolatore a disposizione.

Se osserviamo attentamente la figura, non vi è niente che ci possa indurre a pensare che questo sia un cattivo fit: i punti sembrano oscillare attorno al modello di *best fit* con differenze dell'ordine delle barre d'errore; 7 punti stanno sotto al modello e 9 sopra, dove in media ce ne aspetteremmo 8 e 8, e non si notano regioni in cui il modello sovrastima o sottostima sistematicamente i dati; vi sono esattamente 3 punti che distano più di una barra d'errore dal modello, quando in media ce ne aspettiamo il 32% (ovverosia 1 meno il 68%), cioè circa 5. Insomma, niente che non vada. Ma cosa ci dice in questo caso il valore del  $\chi^2$  restituito dal fit—che, come indicato in figura, è 8.7?

Ricordiamo ancora una volta le regole del gioco: stiamo dicendo che se noi eseguiamo molte volte lo stesso esperimento in condizioni di ripetitività, i punti sperimentali sarebbero diversi ogni volta, così come i valori di *best fit* dei parametri del modello ed il valore del  $\chi^2$  riportato dal fit stesso. Se però (i) il modello è appropriato per il sistema fisico che stiamo studiando e (ii) le incertezze di misura sono stimate in modo statisticamente corretto (e sono Gaussiane), allora il  $\chi^2$  del fit è una variabile casuale distribuita come un chi quadro (appunto) con il numero appropriato di gradi di libertà.

E nel nostro caso specifico? Beh,  $\nu = 13$ , per cui ci aspettiamo che il  $\chi^2$  valga in media  $\mu = \nu = 13$ , con una deviazione standard di  $\sigma = \sqrt{2\nu} \approx 5.10$ . Il valore 8.7 che abbiamo ottenuto è entro una deviazione standard dalla media, per cui non abbiamo niente di cui preoccuparci. Possiamo fare di più: possiamo consultare le tavole in appendice A.6, da cui leggiamo che per un  $\chi^2$  a 13 gradi di libertà

$$P(\chi_{13}^2 \leq 8.7) \approx 20\%$$

(abbiamo interpolato più o meno ad occhio tra i valori 7.04 e 9.30 che si trovano in tabella): se ripetessimo molte volte l'esperimento nel 20% dei casi otterremo un  $\chi^2$  minore e nell'80% dei casi un  $\chi^2$  maggiore, ma il punto è che nessuno di questi due numeri è *troppo vicino* a 0, per cui possiamo concludere che il

nostro fit è un buon fit. Ora, questa è materia scivolosa che si presta facilmente a fraintendimenti per cui è bene chiarire le cose via via che si presentano. *Dire che un fit è un buon fit non equivale a dire che il modello è corretto.* (Per inciso, non si può mai dire che un modello è corretto perché in futuro misure più precise potrebbero dimostrare che non lo è. Può sembrare triste, ma l'unica cosa che si può veramente concludere in Fisica è che un modello è sbagliato.) Invece: *Dire che un fit è un buon fit equivale a dire che non abbiamo ragioni per rigettare il nostro modello.* Teniamolo ben in mente, perché non si tratta di semplice semantica, e ne parleremo più diffusamente nella sezione 8.3.6.

Abbiamo sviscerato in modo esaustivo il caso (relativamente semplice) in cui tutto va bene, ed è ora giunto il momento di esaminare in dettaglio l'ampia zoologia dei *fit cattivi*. Iniziamo con i due casi in cui il  $\chi^2$  è troppo alto mostrati in figura 8.5, che rappresentano un fit lineare agli stessi dati di figura 8.4 ed un fit quadratico ove però le barre d'errore sono state deliberatamente rimpicciolite.

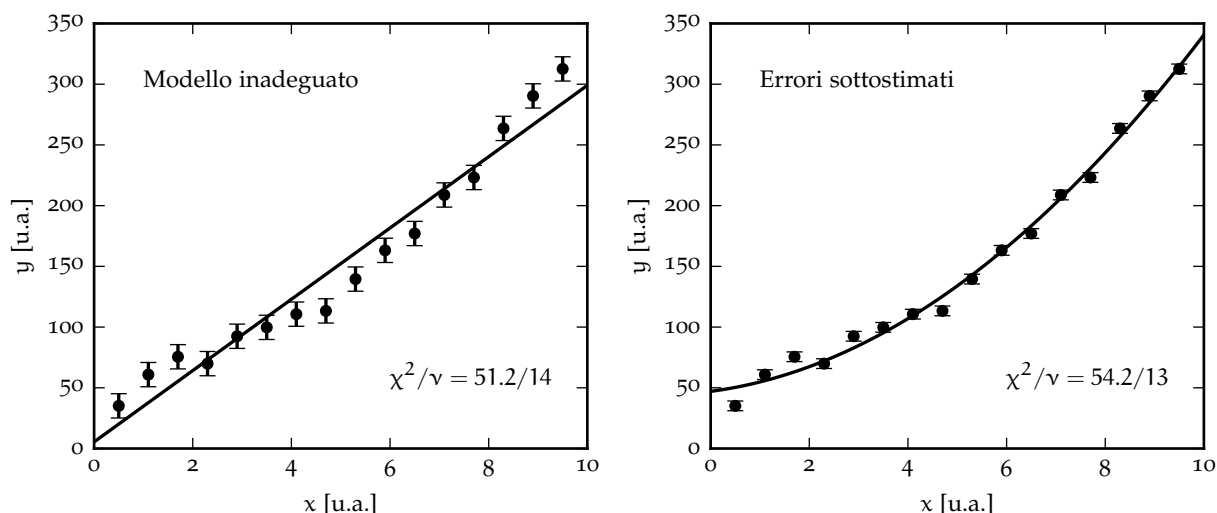


FIGURA 8.5. Esempi di fit lineare (a sinistra) e quadratico (a destra) degli stessi punti sperimentali di figura 8.4—nel secondo caso le barre d'errore sono state volutamente ridotte. In entrambi i casi il valore del  $\chi^2$  è troppo alto, ed indica un cattivo fit. (Notiamo esplicitamente che il fit lineare ha 2 parametri anziché 3, per cui il numero di gradi di libertà passa da 13 a 14.)

Per prima cosa: con 14 gradi di libertà ci aspettiamo un  $\chi^2$  di 14 con una deviazione standard di  $\sqrt{28} \approx 5.29$ , mentre con 13 gradi di libertà, come già sappiamo, ci aspettiamo un  $\chi^2$  di 13 con una deviazione standard di  $\sqrt{26} \approx 5.10$ . In entrambi i casi mostrati in figura 8.5 il valore del  $\chi^2$  restituito dal fit dista qualcosa dell'ordine di 7–8 deviazioni standard dalla media—cioè è troppo grande. (Ora, visto che abbiamo detto che il  $\chi^2$  è una sorta di misura della distanza complessiva del modello dai dati, il fatto che sia troppo grande deve essere male, non vi pare?) Possiamo di nuovo consultare le tavole in appendice A.6, che in questo caso ci dicono

$$P(\chi_{14}^2 \leq 51.2) > 99.99\% \quad \text{e} \quad P(\chi_{13}^2 \leq 54.2) > 99.99\%.$$

Fermiamoci un secondo: stiamo dicendo che se ripetessimo l'esperimento e se (i) il modello fosse corretto e (ii) gli errori stimati correttamente, allora la probabilità di ottenere un  $\chi^2$  minore di quello che abbiamo ottenuto sarebbe maggiore del 99.99%. Cioè che, nelle nostre due ipotesi, otterremo un  $\chi^2$  pari o maggiore a quello che abbiamo ottenuto in meno (e probabilmente molto meno) di un caso su 10000. A questo punto siamo costretti a scegliere tra due possibilità: o siamo stati incredibilmente sfortunati, oppure una delle nostre due assunzioni (o entrambe) non è verificata. Da Fisici, optiamo per la seconda—ed anticipiamo che la soglia per decidere in un verso o nell'altro è completamente arbitraria, ma ci torneremo più in dettaglio nella sezione 8.3.6.

Allora, se il  $\chi^2$  è troppo grande si hanno due casi, che sono legati uno ad uno alle nostre due ipotesi di lavoro: o il numeratore della somma (228) è troppo grande (cioè il nostro modello non è adeguato a fittare i dati) oppure il denominatore è troppo piccolo (cioè gli errori sono sottostimati). La differenza è, anche visualmente, ovvia nei due grafici in figura 8.5: in quello a sinistra si vede chiaramente che i



punti sperimentali hanno una certa concavità che il modello lineare non è in grado di catturare (nella parte centrale ci sono 8 punti consecutivi che stanno sotto al modello); in quello a destra gli errori sono semplicemente troppo piccoli. Nella realtà le due cose possono presentarsi insieme in modo difficilmente fattorizzabile, ma è ugualmente importante allenare l'occhio ad identificare questo tipo di situazioni.

Non abbiamo ancora finito. Ci rimane da trattare il caso in cui il  $\chi^2$  è troppo piccolo. (Se siete tentati di dire che un  $\chi^2$  piccolo non è un problema perché implica che il modello è molto vicino ai dati, non smettete di leggere perché vi sbagliate—e di grosso).

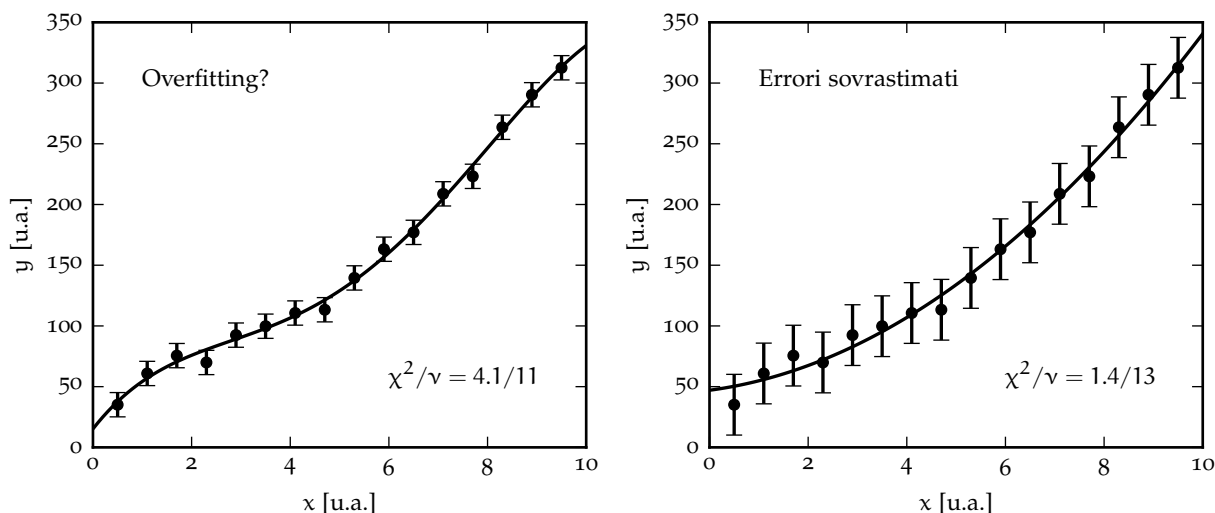


FIGURA 8.6. Esempi di fit con un polinomio di grado 4 (a sinistra) e quadratico (a destra) degli stessi punti sperimentali in figura 8.4—nel secondo caso le barre d'errore sono state volutamente allargate. In entrambi i casi il valore del  $\chi^2$  è troppo basso, ed indica un cattivo fit.

La figura 8.6 mostra un esempio di fit dei dati in figura 8.4 con un polinomio di grado 4 ed un fit quadratico in cui gli errori sono stati volutamente aumentati. Dalle tabelle in appendice A.6 leggiamo

$$P(\chi_{11}^2 \leq 4.1) \approx 2\% \quad \text{e} \quad P(\chi_{13}^2 \leq 1.4) < 0.5\%.$$

Cioè: se ripetessimo l'esperimento e se (i) il modello fosse corretto e (ii) gli errori stimati correttamente, la probabilità di ottenere un  $\chi^2$  pari a quello ottenuto o più piccolo sarebbe molto piccola—otterremmo quasi sempre un  $\chi^2$  più grande. Da Fisici, allora, dobbiamo accettare di nuovo che evidentemente una delle nostre ipotesi di lavoro (o entrambe) non è verificata. Il ragionamento procede esattamente come prima, e una delle possibilità è che abbiamo banalmente sovrastimato gli errori, come mostrato nel pannello di destra della figura 8.6. Ma ce n'è un'altra, e cioè che il nostro modello sia troppo *flessibile*; quando ciò accade, esso tende ad adattarsi alle fluttuazioni statistiche dei dati, ed il fit diviene artificialmente troppo buono, con un  $\chi^2$  corrispondentemente troppo piccolo—ma se ripetessimo l'esperimento su campione di dati diverso, il nostro modello di *best fit* sarebbe con ogni probabilità pessimo. Per completezza, questo fenomeno è noto come *over-fitting*, e l'esempio estremo è di nuovo l'interpolazione con un polinomio di grado  $n - 1$  che abbiamo tirato in ballo più volte dall'inizio del capitolo.

### 8.3.3 Il grafico dei residui

Una rappresentazione utile del fit che si trova spesso usata in pratica è il cosiddetto *grafico dei residui*, in cui si mostra, in funzione della variabile indipendente, la differenza tra i punti misurati ed il modello, calcolato in corrispondenza dei valori di *best fit* dei parametri (e dunque senza incertezze associate):

$$r_i = y_i - f(x_i; \hat{\theta}_1, \dots, \hat{\theta}_m) \quad \text{e} \quad \sigma_{r_i} = \sigma_{y_i}.$$

Il modello di *best fit* coincide dunque con l'asse  $y = 0$  e le differenze tra dati e modello sono visualizzate in modo più efficace che non in un grafico semplice in cui dati e modello sono banalmente sovrapposti.

Spesso il grafico dei residui è combinato con il grafico ordinario nella forma di due pannelli sovrapposti, come mostrato in figura 8.7.

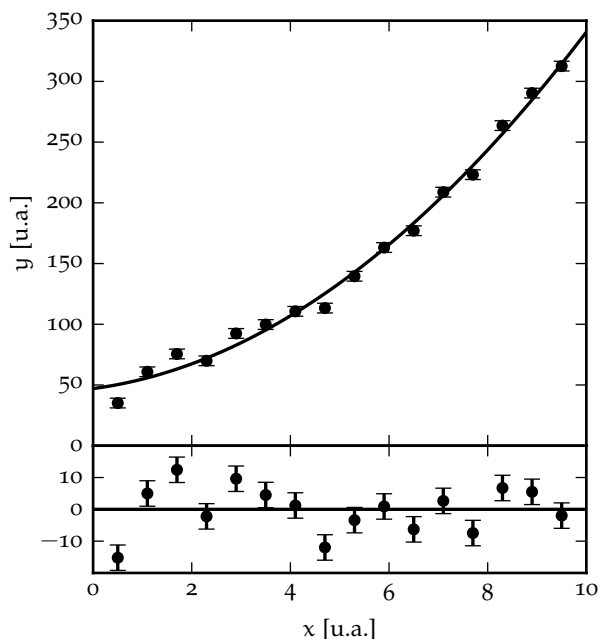


FIGURA 8.7. Esempio di grafico dei residui corrispondente al grafico di sinistra della figura 8.5. Il pannello inferiore (dei residui, appunto) mostra direttamente le deviazioni dei punti misurati rispetto al modello, e rende più facile valutare qualitativamente la bontà del fit. Notiamo esplicitamente che, poiché il modello è calcolato in un punto fissato, i valori numerici delle incertezze associate ai residui sono identiche alle incertezze di misura di partenza  $\sigma_{y_i}$ .

### 8.3.4 Fit dei minimi quadrati e test del $\chi^2$ per una distribuzione

Supponiamo adesso di voler fare un fit dei minimi quadrati o un test del  $\chi^2$  nel caso di una distribuzione—ad esempio, potremmo lanciare un dado a sei facce per 100 volte e vedere se l'ipotesi che il dado sia equo può essere rigettata. Oppure potremmo voler confrontare in senso statistico un istogramma con una distribuzione di variabile continua (e.g., una Gaussiana). È chiaro che questo problema è legato a quanto abbiamo detto fino a questo momento nella prima parte del capitolo. Ma come possiamo costruire una grandezza che sia l'equivalente della (228) in questo nuovo contesto?

La differenza fondamentale è che qui non abbiamo a che vedere con misure sperimentali, ma con *conteggi*. Cominciamo con l'indicare con  $o_i$  le occorrenze osservate in corrispondenza dell' $i$ -esimo valore della nostra variabile discreta o dell' $i$ -esimo canale del nostro istogramma. Il modello, d'altra parte, ci permetterà in generale (eventualmente in funzione dei suoi parametri) di calcolare le occorrenze attese  $e_i$  nella forma

$$e_i(\theta_1, \dots, \theta_m) = \begin{cases} \text{NP}(x_i; \theta_1, \dots, \theta_m) & \text{per una distribuzione di variabile discreta} \\ N \int_{x_i}^{x_{i+1}} p(x; \theta_1, \dots, \theta_m) dx & \text{per un istogramma} \end{cases} \quad (252)$$

dove l'ultimo integrale è inteso tra gli estremi dell' $i$ -esimo canale dell'istogramma ed  $N$  è il numero totale  $N = \sum_i o_i$  di occorrenze osservate.

► **ESEMPIO 8.6.** Nel caso dei nostri 100 lanci di un dado le occorrenze osservate  $o_i$  sono banalmente il numero di volte in cui ogni faccia è uscita. (Notiamo che si ha il vincolo  $\sum_{i=1}^6 o_i = 100$ .) Se ipotizziamo che il dado sia equo, il nostro modello prevede che le occorrenze attese siano identiche tra loro e pari a  $e_i = 100/6$ .

Con questi ingredienti possiamo scrivere quella che anticipiamo essere la risposta corretta alla nostra domanda iniziale, ovvero:

$$\chi^2 = \sum_{i=1}^n \frac{(o_i - e_i(\theta_1, \dots, \theta_m))^2}{e_i(\theta_1, \dots, \theta_m)} \quad \text{o più semplicemente} \quad \chi^2 = \sum_{i=1}^n \frac{(o_i - e_i)^2}{e_i}. \quad (253)$$

Fermiamoci per un secondo. Non vi è dubbio che la (253) ricordi, almeno superficialmente, la (228). Il numeratore  $(o_i - e_i(\theta_1, \dots, \theta_m))^2$  è banalmente la riscrittura di  $(y_i - f(x_i; \theta_1, \dots, \theta_m))^2$ , a patto di identificare  $o_i \rightarrow y_i$ —ciò che abbiamo misurato—ed  $e_i(\theta_1, \dots, \theta_m) \rightarrow f(x_i; \theta_1, \dots, \theta_m)$ —ovvero ciò che prevede il modello. Al denominatore abbiamo bisogno di una stima dell'errore di misura, che in questo caso coinciderà con la deviazione standard delle distribuzioni dei conteggi attesi  $e_i$ . Ma noi siamo abituati (lo abbiamo visto per le distribuzioni binomiale e Poissoniana) al fatto che gli errori sui conteggi scalano con la radice dei conteggi stessi  $\sigma_{e_i} \approx \sqrt{e_i}$  ovvero  $\sigma_{e_i}^2 \approx e_i$ . La (253) è dunque perfettamente ragionevole, sulla base di tutto quello che abbiamo imparato fino a questo momento.

C'è ancora un punto, sottile ma estremamente importante, da discutere. Nei casi in cui il numero totale  $N$  di eventi (o conteggi) è *fissato*—ad esempio perché abbiamo deciso a priori di lanciare la nostra moneta o il nostro dado 100 volte—gli  $n$  elementi della somma (253) non sono più tutti indipendenti tra di loro, perché non sono più indipendenti tra di loro le grandezze  $o_i$ . Se conosciamo il valore di  $n - 1$  occorrenze possiamo calcolare banalmente la rimanente attraverso

$$o_k = N - \sum_{i \neq k} o_i, \quad \text{ad esempio} \quad o_n = N - \sum_{i=0}^{n-1} o_i.$$

Ma si tratta di una cosa cui siamo ormai avvezzi: abbiamo un vincolo ulteriore che riduce di una unità il numero di gradi di libertà, per cui nel contesto del test del  $\chi^2$  per una distribuzione la (251) diviene

$$v = n - m - 1. \quad (254)$$

Ovvero: quando si ha a che fare con una variabile casuale discreta o un istogramma ed il numero totale di conteggi è fissato, *il numero di gradi di libertà è pari al numero dei dati sperimentali  $n$  meno il numero di parametri  $m$  che stimiamo dal fit meno 1.*

### 8.3.5 Poissoniana o binomiale? Un semplice esempio

È arrivato il momento di osservare più da vicino la (253). Molti di voi saranno saltati sulla sedia gridando: certo! La statistica dei conteggi è Poissoniana, e dato un numero di conteggi medio  $e_i$  la deviazione standard corrispondente è  $\sqrt{e_i}$ , per cui il denominatore della (253) rappresenta esattamente l'equivalente del quadrato dell'errore di misura. Di più: poiché se gli  $e_i$  sono abbastanza grandi la distribuzione di Poisson può essere approssimata con una Gaussiana, la (253) tende in questo limite alla somma di  $n$  variabili Gaussiane in forma standard indipendenti, per cui ha proprio il significato di un  $\chi^2$ .

Tutto questo è vero quando il numero totale di conteggi non è fissato, ma nel caso tipico in cui decidiamo a priori quante volte vogliamo ripetere un esperimento (e.g., i nostri soliti 100 lanci di un dado) le cose sono più complicate—anche se la (253), come vedremo tra un attimo, costituisce ancora la risposta corretta. In questo caso, se consideriamo l' $i$ -esimo valore della variabile casuale in questione (o l' $i$ -esimo canale dell'istogramma) *contro tutti gli altri*, non è difficile rendersi conto che il problema è in realtà binomiale. Per chiarire le idee con un esempio, se (con la consueta fantasia) lanciamo un dado a 6 facce per 100 volte, il numero di volte  $o_2$  in cui esce la faccia 2 sarà una variabile casuale distribuita secondo una binomiale  $\mathcal{B}(o_2; 100, 1/6)$ . In generale, detta  $p_i = P(x_i)$  la probabilità elementare che l'esito del nostro esperimento coincida con il valore  $i$ -esimo  $x_i$  della variabile casuale in questione o cada nel canale  $i$ -esimo del nostro istogramma, le occorrenze corrispondenti  $o_i$  saranno distribuite come una binomiale  $\mathcal{B}(o_i; N, p_i)$ , con

$$E[o_i] = Np_i = e_i \quad \text{e} \quad \text{Var}(o_i) = Np_i(1 - p_i) \neq e_i. \quad (255)$$

Il problema sta proprio nel fatto che in questo caso la varianza delle occorrenze attese è diversa dal valore di aspettazione delle occorrenze stesse, ed uno può legittimamente chiedersi dove sia finito il fattore  $(1 - p_i)$  che, ingenuamente, ci aspetteremmo di vedere al denominatore della (253). Lo vediamo in pratica con un esempio elementare.

Supponiamo dunque di lanciare una moneta  $N$  volte e di ottenere un numero di teste  $o_t$  ed un numero di croci  $o_c$ . A costo di essere pedanti ripetiamo che la somma  $o_t + o_c = N$  è fissata per cui  $o_c$  ed  $o_t$  non sono indipendenti—e.g., il numero di croci è fissato dal numero di teste nella forma di  $o_c = N - o_t$ .

Indichiamo con  $p$  la probabilità che l'uscita sia testa, e fissiamo  $p = 1/2$  a priori perché vogliamo verificare l'ipotesi che la moneta sia equa. In queste condizioni il modello, che è completamente specificato e non ha parametri liberi se non la normalizzazione, prevede

$$e_t = Np \quad e_c = N(1-p),$$

per cui il nostro conteggio del numero di gradi di libertà è  $\nu = n - m - 1 = 2 - 0 - 1 = 1$ . Proviamo a mettere tutto insieme e scrivere la (253) nel nostro caso specifico

$$\begin{aligned} \chi^2 &= \frac{(o_t - e_t)^2}{e_t} + \frac{(o_c - e_c)^2}{e_c} = \frac{(o_t - Np)^2}{Np} + \frac{(o_c - N(1-p))^2}{N(1-p)} = \\ &= \frac{(1-p)(o_t - Np)^2 + p(o_c - N(1-p))^2}{Np(1-p)} = \frac{(1-p)(o_t - Np)^2 + p(o_t - Np)^2}{Np(1-p)} = \frac{(o_t - Np)^2}{Np(1-p)} = \\ &= \frac{(o_t - e_t)^2}{e_t(1-p)}. \end{aligned}$$

(Nel terzultimo passaggio abbiamo eliminato  $o_c$  sfruttando la condizione di normalizzazione.) La nostra semplice manipolazione algebrica ha sortito due effetti distinti. Il primo è che, una volta scritta in funzione delle sole variabili indipendenti, la (253) ha un solo termine (e quindi, come avevamo anticipato) un solo grado di libertà anziché due. La seconda è che adesso al denominatore riconosciamo la varianza del numero atteso di teste  $e_t$  nell'ipotesi (corretta) che essa sia distribuita come una binomiale.

Il gioco funziona anche per  $n > 2$  e, con un leggero abuso di termini, potremmo dire che la (253) è corretta perché *maschera* da somma di  $n$  variabili Poissoniane indipendenti—in una forma elegante e simmetrica—quella che in realtà è la somma di  $n - 1$  variabili binomiali.

Ora, quando i valori medi attesi  $e_i$  sono abbastanza grandi<sup>2</sup> la distribuzione (binomiale) di ciascun termine della somma tende ad una Gaussiana in forma standard ed il cerchio si chiude, poiché la (253) acquista, sia pure solo approssimativamente, il significato di una variabile  $\chi^2$  e tutto quello che abbiamo discusso nella sezione precedente, *mutatis mutandis*, vale anche qui. (Per inciso, abbiamo discusso il caso di una distribuzione nel contesto del test del  $\chi^2$ , ma va da sé che il tutto si applica anche alla stima di parametri mediante fit dei minimi quadrati.)

► **ESEMPIO 8.7.** Si lancia un dado a sei facce per 100 volte, ottenendo per le facce 1...6 le seguenti occorrenze: 13, 12, 14, 20, 22, 19. Possiamo concludere che il dado non è equo? Si tratta di una semplice applicazione del test del  $\chi^2$ , in cui il numero atteso di occorrenze per ciascuna faccia è  $100/6$ . Possiamo usare la tabella qui di fianco segue per costruire la somma (253). Abbiamo  $n = 6$  punti e, poiché il modello è completamente specificato (non abbiamo stimato nessun parametro dai dati) il numero di gradi di libertà del problema è  $\nu = 6 - 0 - 1 = 5$ . Il valore del  $\chi^2$  che abbiamo ottenuto è 5.24, che è abbondantemente entro una deviazione standard dalla media attesa per 5 gradi di libertà, per cui non abbiamo ragione di pensare che il dado sia truccato.

$o_i$	$e_i$	$(o_i - e_i)^2 / e_i$
13	16.67	0.808
12	16.67	1.308
14	16.67	0.428
20	16.67	0.665
22	16.67	1.704
19	16.67	0.326
100	100	5.24

► **ESEMPIO 8.8.** Si lancia un dado a sei facce per 10000 volte, ottenendo per le facce 1...6 le seguenti occorrenze: 1300, 1200, 1400, 2000, 2200, 1900 (si tratta dell'esempio precedente in cui abbiamo moltiplicato tutti i numeri per un fattore 100). Cambiano le conclusioni? Anche senza costruire esplicitamente la tabella possiamo notare che in questo caso stiamo moltiplicando tutti i numeratori della somma per  $100^2$  ed i denominatori per 100, per cui l'effetto netto sarà che il  $\chi^2$  in questo caso è 100 volte più grande. Il numero di gradi di libertà è ancora 5 e con un  $\chi^2$  di 524 per 5 gradi di libertà possiamo affermare senza ombra di dubbio che il dado è truccato.

<sup>2</sup> Non esiste un limite netto a cosa significhi *abbastanza grande*, anche se in alcuni testi si trova  $e_i \geq 5$ . Se il numero di occorrenze per un determinato termine della somma è però inferiore ad 1 o 2 conviene sicuramente aggregare il termine stesso con uno (o più) dei vicini prima di calcolare il  $\chi^2$ .

## 8.3.6 Il concetto di p-value

Possiamo riassumere sommariamente quanto visto nelle sezioni precedenti dicendo che, quando facciamo un test del  $\chi^2$  per un fit con  $\nu$  gradi di libertà, ci aspettiamo il valore numerico del  $\chi^2$  stesso sia in media  $\nu$ , con una deviazione standard di  $\sqrt{2\nu}$ . Se il risultato è entro 2–3 deviazioni standard dalla media, non abbiamo ragione di rigettare il modello. Se il  $\chi^2$  è troppo grande, questo significa che il modello è inadeguato oppure gli errori sono sottostimati—o una combinazione delle due cose. Se il  $\chi^2$  è troppo piccolo, ciò significa che il modello ha troppi parametri oppure gli errori sono sovrastimati—o, di nuovo, una combinazione delle due. Ora, il test del  $\chi^2$  è il primo esempio di *verifica delle ipotesi* che incontriamo nel nostro percorso. Fino a questo momento ci siamo concentrati sulla sua applicazione al problema della bontà di un fit, ma più in generale, il  $\chi^2$  è una *statistica* (o *test statistics* in Inglese) che può essere utilizzato anche quando non è esplicitamente in gioco un fit.

In effetti gli esempi 8.7 e 8.8 sono rappresentativi della situazione in cui il modello che utilizziamo per il test è completamente specificato indipendentemente dalle osservazioni, e noi siamo semplicemente interessati a verificare la presenza o meno di un effetto fisico (“il dado è truccato?”). Possiamo anche riformulare il problema ribaltando la logica e chiederci se abbiamo elementi per rigettare l’ipotesi che l’effetto fisico che cerchiamo *non* sia all’opera (“il dado è equo?”). Quest’ultima si chiama tipicamente *ipotesi nulla* e si denota con  $H_0$ , e nel contesto della verifica delle ipotesi si definisce *p-value* la *probabilità di ottenere un risultato (e.g., un valore di  $\chi^2$ ) uguale o più estremo di quello che abbiamo ottenuto, assumendo che l’ipotesi nulla sia verificata*. Si tratta di una definizione piuttosto involuta, ma intuitivamente possiamo dire che il *p-value* misura il livello di incompatibilità di un particolare insieme di dati con un modello statistico—e più è piccolo il *p-value*, più è grande questa incompatibilità. Un valore troppo piccolo di *p-value* può allora essere interpretato come un’evidenza a sfavore dell’ipotesi nulla—sempre che tutte le assunzioni che vanno nel calcolo del *p-value* stesso siano verificate.

Quello di *p-value* è uno dei concetti più controversi nella statistica moderna [5, 14], tanto che il suo uso è stato recentemente bandito, non senza che questo destasse scalpore, da una nota rivista scientifica Statunitense: *Basic and Applied Social Psychology*. E il motivo di fondo della controversia è il fatto che la letteratura—sia quella specialistica che quella divulgativa—è costellata di cattivi esempi in cui il concetto di *p-value*, usato in modo fondamentalmente scorretto, ha generato confusione e disinformazione (per non parlare dei casi in cui il *p-value* è stato usato in modo deliberatamente fraudolento per supportare conclusioni non corroborate dai dati). Questo è dunque un buon momento per chiarire cosa *non* è il *p-value*.

- Il *p-value* non è la probabilità che l’ipotesi nulla sia corretta. La differenza può sembrare sottile, ma tornate per un attimo a rileggere la nostra definizione iniziale: essa è un’affermazione probabilistica a proposito dei dati (in relazione ad una certa ipotesi  $H_0$ ) e non a proposito dell’ipotesi. In altre parole, a questo punto dovremmo avere gli strumenti per capire al volo che le probabilità condizionate  $P(\text{dati} | H_0)$  e  $P(H_0 | \text{dati})$  non sono la stessa cosa. Così come non diremmo mai che la probabilità che un senatore sia donna è uguale alla probabilità che una donna sia senatore, allo stesso modo non dobbiamo farci ingannare da questo.
- A maggior ragione il *p-value* non è la probabilità che l’ipotesi alternativa ad  $H_0$  sia errata—ovverosia  $(1 - \text{p-value})$  non è la probabilità che l’ipotesi alternativa ad  $H_0$  sia corretta.
- Il *p-value* non è la probabilità che quello che abbiamo osservato sia semplicemente l’effetto di una fluttuazione statistica. Il calcolo del *p-value* è basato esattamente sull’assunzione che la nostra osservazione sia il prodotto del caso (sotto l’ipotesi nulla) per cui è chiaro che non può essere utilizzato per stimare la probabilità di questa assunzione.
- Il *p-value* non è la probabilità di rigettare per sbaglio l’ipotesi nulla. Di nuovo: la nostra definizione di *p-value* è un’affermazione a proposito dei dati, non dell’ipotesi nulla.

Parte della confusione nasce in pratica dal fatto che tutte le cose che il *p-value* non è che abbiamo appena elencato sono grandezze interessanti—per molti aspetti più del *p-value* stesso. Sarebbe bello avere una risposta generale alla domanda: “Qual è la probabilità che l’ipotesi nulla sia corretta?” ma il fatto è che, purtroppo, questa è una domanda intrinsecamente più complicata di quella (purtroppo più involuta)

► **ESEMPIO 8.9.** Supponiamo di aver lanciato un dado a sei facce per 100000 volte e di aver fatto un test del  $\chi^2$ , come negli esempi 8.7 e 8.8, che ha fornito questa volta un valore di 35.5. La nostra ipotesi nulla  $H_0$  è il fatto che il dado sia equo. Si tratta di un'ipotesi completamente specificata, che costituisce una descrizione completa del nostro problema e ci permette di calcolare le occorrenze attese da utilizzare nel test del  $\chi^2$ . Il *p-value* coincide con la probabilità che un  $\chi^2$  a 5 gradi di libertà sia maggiore o uguale a 35.5 e può essere calcolato facilmente con un *computer*

$$p\text{-value} = P(\chi_5^2 \geq 35.5) \approx 1.2 \times 10^{-6}.$$

Ora, quello che in sostanza il *p-value* ci dice in questo caso è che se il dado fosse equo e noi ripetessimo il nostro esperimento (cioè i nostri 100000 lanci) molte volte, allora otterremmo un  $\chi^2$  maggiore o uguale a quello che abbiamo ottenuto in questo primo tentativo (35.5) circa una volta su un milione. In questo caso il valore basso di *p-value* può essere interpretato come un'evidenza a sfavore dell'ipotesi nulla—cioè siamo costretti ad accettare che il dado non sia equo. Detto questo:

- $1.2 \times 10^{-6}$  non è la probabilità che il dado sia equo;
- $(1 - 1.2 \times 10^{-6})$  non è la probabilità che il dado non sia equo;
- $1.2 \times 10^{-6}$  non è la probabilità di aver osservato una fluttuazione statistica;
- $1.2 \times 10^{-6}$  non è la probabilità di commettere un errore dicendo che il dado non è equo.

► **ESEMPIO 8.10.** Un test del  $\chi^2$  sui dati relativi ai decessi per incidenti nella cavalleria Prussiana riportati nella tabella 13 fornisce un valore di 0.60 per  $5 - 1 - 1 = 3$  gradi di libertà, corrispondente ad un *p-value* di circa il 10%, per cui non abbiamo motivi per rigettare l'ipotesi che la distribuzione che regola il fenomeno sia Poissoniana.

► **ESEMPIO 8.11.** Un test del  $\chi^2$  sui dati relativi ai bombardamenti di Londra riportati nella tabella 14 fornisce un valore di 1.17 per  $6 - 1 - 1 = 4$  gradi di libertà, con un *p-value* del 12%, per cui, di nuovo, non abbiamo motivi per rigettare l'ipotesi che la distribuzione sottostante sia Poissoniana.

cui il *p-value* risponde. Da un punto di vista operativo la situazione è aggravata dall'usanza inveterata (propria soprattutto delle Scienze *soft*) di utilizzare una soglia fissata (tipicamente  $p = 0.05$ ) sul *p-value* per decidere se una determinata ipotesi può essere rigettata sulla base di una misura o un esperimento. Noi adotteremo il punto di vista (sobrio) che l'utilizzo appropriato del *p-value* può essere utile ricordando che esso non è, di per sé, una buona misura della correttezza di un modello o di un ipotesi.

Concludiamo questa sezione menzionando il fatto che in Fisica la soglia standard che si usa per annunciare una scoperta è molto più stringente del  $p = 0.05$  di cui sopra—si usa invece la regola delle  $5\sigma$  di cui forse avete sentito parlare. In questo contesto  $5\sigma$  si riferisce ad una distribuzione Gaussiana e si può tradurre facilmente nel *p-value* corrispondente leggendo la tabella in appendice A.5

$$p\text{-value} = P(z \geq 5) \approx 2.87 \times 10^{-7}.$$

Tanto per fissare le idee: la probabilità di ottenere 22 teste di seguito lanciando una moneta equa è  $p = 1/2^{22} \approx 2.38 \times 10^{-7}$ , simile al *p-value* corrispondente alle famose  $5\sigma$ —per cui in effetti stiamo dicendo che un Fisico ha bisogno di vedere con i suoi occhi 22 teste di seguito prima di dichiarare che una moneta non è equa.

### 8.3.7 Fit dei minimi quadrati e test del $\chi^2$ con errori non Gaussiani

Nonostante il formalismo che abbiamo sviluppato presupponga che le incertezze di misura siano Gaussiane, non tarderete ad accorgervi che in genere questo è un punto cui non viene prestata troppa attenzione, e sia il fit dei minimi quadrati che il test del  $\chi^2$  sono largamente utilizzati, in pratica, indipendentemente dal modello specifico di incertezze. Va da sé che il problema può essere risolto

completamente caso per caso—analiticamente o con metodi Monte Carlo—ma in questa sezione ci limitiamo ad alcune considerazioni di carattere generale, rivisitando passo per passo quello che abbiamo detto e cercando di capire cosa resta valido e cosa cambia se lasciamo cadere l'assunzione che gli errori siano Gaussiani.

Per prima cosa la (228) rimane una misura ragionevole della distanza complessiva tra dati e modello nella maggior parte dei casi, per cui i valori dei parametri che si ottengono dalla sua minimizzazione sono per lo più ragionevoli anche se le incertezze di misura non sono Gaussiane. Fanno potenzialmente eccezione i casi in cui gli errori di misura sono notevolmente asimmetrici—che è un caso che si presenta, ad esempio, quando si ha a che fare con una distribuzione con pochi conteggi.

Il calcolo delle incertezze di misura, sul modello delle (233) e (236), si basa essenzialmente sul fatto che per variabili casuali indipendenti la varianza della somma è uguale alla somma delle varianze—che è vero indifferentemente dalla distribuzione generatrice. Dunque se abbiamo stimato gli errori come la deviazione standard della distribuzione generatrice corrispondente, le incertezze sui valori di *best fit* dei parametri restituite dal fit saranno tendenzialmente ragionevoli anche nel caso non Gaussiano. (Avrete anche notato che, per lo meno nei casi più semplici, la stima dei parametri coinvolge delle somme quindi, per il teorema del limite centrale, è plausibile che la distribuzione a posteriori dei parametri sia Gaussiana anche se gli errori sulle grandezze misurate non lo sono.)

Il test del  $\chi^2$  è il punto in cui sorgono i veri problemi—nel senso che se gli errori non sono Gaussiani, allora per definizione la (228) non è distribuita come una variabile  $\chi^2$  e la maggior parte delle cose che abbiamo detto non è più vera. In pratica, se abbiamo stimato gli errori come la deviazione standard della distribuzione generatrice, allora il singolo termine della somma è ancora il quadrato di una variabile casuale  $\zeta$  con media 0 e varianza 1, per cui è ancora vero che

$$E[\zeta^2] = 1 \quad \text{e} \quad E[\chi^2(\hat{\theta}_1, \dots, \hat{\theta}_m)] = \nu$$

Se il numero di gradi di libertà del problema è abbastanza grande, il teorema del limite centrale ci viene di nuovo in aiuto garantendo che la distribuzione della nostra somma sia asintoticamente Gaussiana, ma la sua varianza non sarà più, in generale,  $2\nu$ .

Di fatto noi sappiamo la risposta esatta almeno per un caso praticamente rilevante—ovvero quello di errori uniformi (per esempio se utilizziamo uno strumento digitale ed assumiamo come incertezza la risoluzione strumentale diviso la  $\sqrt{12}$ ). Nella sezione 7.3.1 abbiamo visto che la varianza del quadrato di una distribuzione uniforme con varianza 1 è  $4/5$ , per cui in questa situazione avremo

$$\text{Var}(\chi^2(\hat{\theta}_1, \dots, \hat{\theta}_m)) = \frac{4}{5}\nu.$$

Meno della metà della varianza della corrispondente distribuzione del  $\chi^2$ : se convertiamo in probabilità il valore del  $\chi^2$  ottenuto con errori non Gaussiani utilizzando le tavole in appendice A.6 possiamo sbagliare di grosso. In questo caso se  $\nu$  è abbastanza grande possiamo utilizzare l'approssimazione Gaussiana (con la deviazione standard opportuna). Nel caso generale è necessario calcolare la densità di probabilità rilevante—analiticamente oppure attraverso una simulazione Monte Carlo.

#### 8.4 IL FIT IN PRATICA: METODI NUMERICI

Riprendiamo il filo della discussione e torniamo a parlare del problema generale dei fit. Nel contesto dei minimi quadrati il sistema di equazioni (229) ci fornisce una strategia generale per ricavare i valori di *best fit* dei parametri del modello e le incertezze di misura associate, e abbiamo visto che le (229) si possono risolvere in forma chiusa per lo meno nei casi più semplici. Possiamo ritenerci soddisfatti?

La risposta alla domanda è senza dubbio no, perché non è troppo difficile trovare un esempio di modello in cui le cose non vanno così lisce. Consideriamo ad esempio la famiglia di funzioni

$$f(x; \omega) = \sin(\omega x). \quad (256)$$

In questo caso se, come abbiamo fatto nel caso costante o lineare, scriviamo il  $\chi^2$  e minimizziamo rispetto ad  $\omega$  otteniamo l'equazione

$$\frac{d\chi^2(\hat{\omega})}{d\omega} = -2 \sum_{i=1}^n \left( \frac{y_i - \sin(\hat{\omega}x_i)}{\sigma_{y_i}^2} \right) x_i \cos(\hat{\omega}x_i) = 0 \quad (257)$$

che non c'è alcuna speranza di risolvere analiticamente in forma chiusa. Notiamo esplicitamente che si tratta di un problema ad un solo parametro, e nemmeno particolarmente esotico, per cui se non siamo in grado di risolvere questa tutta la teoria che abbiamo sviluppato non sarà di grande utilità pratica.

La differenza fondamentale tra la (257) la (230) e la (234) è che la prima equazione è non lineare nei parametri. Si parla in questo caso di fit dei minimi quadrati non lineare<sup>3</sup> e, come vedremo tra un attimo, la tecnica standard per affrontare il problema è quello dell'utilizzo di metodi numerici iterativi.

#### 8.4.1 Digressione: il metodo di Newton in una dimensione

Supponiamo di essere interessati a trovare il valore numerico della radice (o delle radici, nel caso ce ne fossero più di una) di una generica equazione  $f(x) = 0$  e supponiamo che l'equazione non si possa risolvere analiticamente in forma chiusa. Tanto per fissare le idee, il problema in questione potrebbe essere

$$\cos x - x = 0 \quad \text{ovvero} \quad \cos x = x$$

(in questo caso la radice è unica e si trova nell'intervallo  $[0, 1]$ ). Il metodo di Newton è uno dei metodi numerici più semplici per attaccare questo tipo di problema e lo discutiamo brevemente a scopo esemplificativo.

Supponiamo dunque di partire da un generico punto  $x_0$  e sviluppare la nostra funzione in serie di Taylor al prim'ordine attorno al punto stesso

$$f(x) \approx f(x_0) + \frac{df(x_0)}{dx}(x - x_0).$$

Ora, se vogliamo trovare i valori di  $x$  per cui  $f(x) = 0$ , che è il nostro problema di partenza, la condizione precedente si può riscrivere come

$$x \approx x_0 - \frac{f(x_0)}{\frac{df(x_0)}{dx}}.$$

Apparentemente non siamo andati molto avanti, perché: (i) la nostra equazione è approssimata e (ii) non è più semplice di quella da cui siamo partiti. Eppure la relazione che abbiamo ricavato ci fornisce l'idea per una strategia iterativa che potrebbe funzionare in pratica: data una stima  $x_n$  della soluzione, possiamo procedere per ricorrenza e ricavare una stima (sperabilmente) più accurata della nostra radice come

$$x_{n+1} = x_n - \frac{f(x_n)}{\frac{df(x_n)}{dx}}. \quad (258)$$

Geometricamente  $x_{n+1}$  non è nient'altro che la coordinata di intersezione della retta tangente ad  $f$  nel punto  $x_n$  con l'asse delle  $x$ , e quello che stiamo dicendo è che questa intersezione è, sotto opportune condizioni, una stima più accurata della radice dell'equazione di quanto non fosse il punto di partenza. Una discussione dettagliata delle limitazioni del metodo di Newton esula dai nostri scopi—è ovvio, ad esempio, che se ad una data iterazione incontriamo un massimo di  $f(x)$ , la (258) diverge poiché il denominatore della frazione si annulla. Purtroppo si tratta di un buon esempio per illustrare i tratti salienti di un tipico metodo numerico iterativo, in cui si parte da una stima (anche grossolana) del parametro che vogliamo stimare e questa stima viene aggiornata mediante un algoritmo che converge (se tutto va bene) al valore cercato.

<sup>3</sup> Vale la pena di sottolineare che il termine *lineare* in questo contesto, non si riferisce al modello, ma al sistema di equazioni (229) che permettono di ricavare i valori di *best fit* minimizzando il  $\chi^2$ .



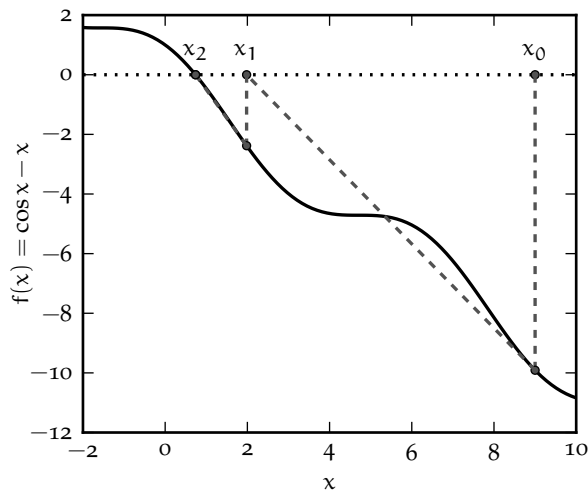


FIGURA 8.8. Illustrazione dei primi due passi del metodo di Newton per la stima delle radici dell'equazione  $\cos x - x = 0$  con la scelta del valore iniziale  $x_0 = 9$ . I valori numerici per i primi 6 passi dell'algoritmo sono elencati nella tabella 18.

Il procedimento è illustrato in figura 8.8 (per i primi due passi, ed una scelta del valore iniziale  $x_0 = 9$ ) nel caso  $f(x) = \cos x - x$  menzionato prima—in cui la nostra relazione per ricorrenza (258) si legge

$$x_{n+1} = x_n + \frac{(\cos x_n - x_n)}{(\sin x_n + 1)}.$$

I valori numerici forniti dai primi 6 passi dell'algoritmo sono elencati nella tabella 18 che, insieme alla figura, illustra chiaramente come il metodo di Newton possieda, almeno in questo caso, tutte le proprietà desiderabili di un algoritmo iterativo:

1. il valore fornito dal passo  $n$  è più vicino alla soluzione di quello fornito dal passo  $n - 1$ , cioè l'algoritmo converge alla radice cercata in modo monotono;
2. la convergenza è rapida, nel senso che sono necessari pochi passi per ottenere una stima accurata della radice cercata;
3. le differenze  $\delta = x_n - x_{n-1}$  tra due valori forniti da passi successivi dell'algoritmo tendono a decrescere rapidamente al crescere di  $n$ —tipicamente dopo pochi passi una nuova iterazione non modifica il valore restituito dal passo precedente.

L'ultima proprietà è particolarmente significativa, poiché, in condizioni ordinarie, il cambiamento relativo nel valore numerico della stima della radice è un buon indice di quanto siamo vicini alla soluzione corretta e può essere utilizzato come criterio per interrompere il processo iterativo—in altre parole: ci fermiamo quando continuare non cambia più le cose in modo apprezzabile.

$n$	$x_n$	$\delta = x_n - x_{n-1}$
0	9.0	—
1	1.98137488782	-7.01863
2	0.739511923363	-1.24186
3	0.739085173417	$-4.26750 \times 10^{-4}$
4	0.739085133215	$-4.02017 \times 10^{-8}$
5	0.739085133215	$-3.33067 \times 10^{-16}$
6	0.739085133215	0

TABELLA 18. Valori numerici forniti dai primi 6 passi del metodo di Newton applicato alla ricerca della radice dell'equazione  $\cos x - x = 0$ , partendo dal valore iniziale  $x_0 = 9$  (vedi figura 8.8). Si noti come già alla sesta iterazione la stima della radice non cambia nelle prime 12 cifre significative.

#### 8.4.2 Minimi quadrati non lineari e metodi iterativi

Torniamo al nostro problema di partenza, ovvero ai fit. Supponiamo di avere un punto materiale che si muove di moto circolare uniforme su di una circonferenza di raggio 1 m e di voler stimare la pulsazione angolare  $\omega$  del moto dal fit di una serie di misure della proiezione della posizione del punto sull'asse  $y$

ad istanti di tempo fissati—come mostrato in figura 8.9. Si tratta di un esempio puramente accademico che ci permette però di discutere in pratica il problema generale (non lineare) del fit dei minimi quadrati in uno dei casi più semplici tra quelli che non ammettono una soluzione analitica in forma chiusa, ovvero con il modello (256).

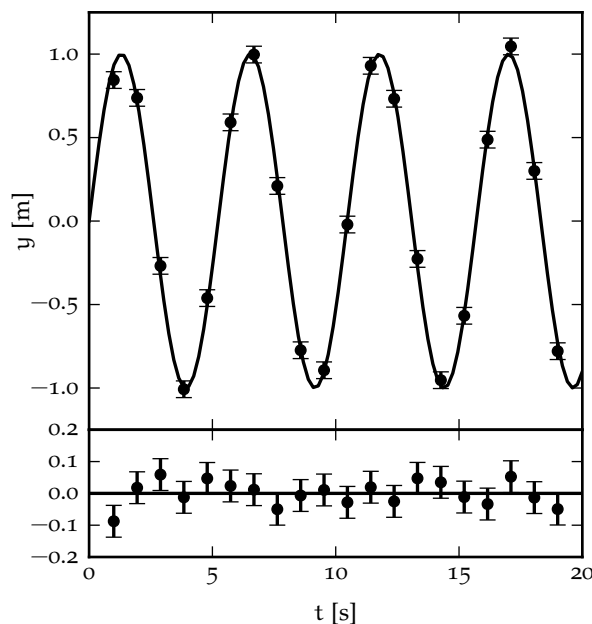


FIGURA 8.9. Esempio di fit numerico iterativo di una serie di dati con il modello (256), realizzato con la funzione `curve_fit` del modulo `scipy.optimize`. Il fit restituisce un valore  $\hat{\omega} = 1.2005 \pm 0.0011 \text{ rad s}^{-1}$ , con un  $\chi^2$  pari a 11.6 (per 19 gradi di libertà). Il grafico dei residui mostra chiaramente che, come indicato dal valore del  $\chi^2$ , il fit è un buon fit, poiché i punti misurati tendono ad oscillare attorno al modello in modo apparentemente casuale e l'entità caratteristica delle fluttuazioni è dello stesso ordine di grandezza delle incertezze di misura.

Qualsiasi programma di analisi dati che si rispetti fornisce almeno un *motore* di fit che permette di eseguire un fit dei minimi quadrati (lineare o non lineare) di una serie di dati, utilizzando come modello una funzione arbitraria dipendente da uno o più parametri. Questi motori di fit, va da sé, non trovano il minimo del  $\chi^2$  risolvendo il problema analiticamente come abbiamo fatto noi nel caso costante e lineare (e, d'altra parte, sappiamo che ciò non è in generale possibile). Lo fanno invece iterativamente, partendo da una stima iniziale dei parametri, anche grossolana (tipicamente fornita dall'utente), e provando valori diversi per i parametri stessi fino ad arrivare ragionevolmente vicino al minimo globale (se tutto va bene, ovviamente). Per fissare le idee potete immaginare che il calcolatore campioni per voi il valore del  $\chi^2$  in una regione rappresentativa dello spazio dei parametri e vi restituisca le coordinate del punto in cui il  $\chi^2$  stesso assume il valore minimo. Esistono strategie computazionalmente efficienti per trovare questo minimo—ed il metodo di Netwon che abbiamo descritto sommariamente è un esempio che funziona nei casi più semplici—ma l'argomento va ben al di là dello scopo di queste dispense, per cui non ci dilunghiamo oltre.

Il modulo `scipy.optimize` di Python offre un'interfaccia al fit dei minimi quadrati non lineare attraverso la funzione `curve_fit`, che utilizzeremo in questa sezione per illustrare i tratti generali dei metodi iterativi di fit, ma resta inteso che esiste una varietà sconfinata di programmi di analisi dati che offrono alternative altrettanto buone o addirittura migliori. Per i nostri scopi, la *segnatura* della funzione `curve_fit` è

$$popt, pcov = \text{scipy.optimize.curve\_fit}(f, xdata, ydata, p0=None, sigma=None) \quad (259)$$

(ad essere onesti la cosa è significativamente più complicata, ma i parametri che abbiamo elencati sono sufficienti per la discussione ed il lettore può trovare tutti i dettagli mancanti nella documentazione ufficiale disponibile su web <sup>4</sup>). Passiamo dunque brevemente in rassegna, in ordine di apparizione, i parametri che il nostro motore di fit richiede.

1. *f*: il nostro modello, ovvero la funzione di fit. È definita dall'utente nella forma di una funzione di Python che deve accettare la variabile indipendente come primo argomento, ed i parametri del modello come argomenti aggiuntivi (in ordine). Un punto importante da tenere a mente è che questa funzione

<sup>4</sup> [http://scipy.github.io/devdocs/generated/scipy.optimize.curve\\_fit.html](http://scipy.github.io/devdocs/generated/scipy.optimize.curve_fit.html)

deve essere capace di operare su *array* di *numpy*, per cui è buona regola utilizzare le funzioni native di *numpy* al suo interno.

2. *xdata*: il vettore delle coordinate  $x_i$  delle misure. Può essere una lista di Python o un *array* di *numpy*.
3. *ydata*: il vettore delle coordinate  $y_i$  delle misure. Può essere una lista di Python o un *array* di *numpy*, e la sua lunghezza deve essere uguale a quella di *xdata*.
4. *p0*: i valori iniziali dei parametri del modello, nella forma di una lista o tupla di Python o un *array* di *numpy*—in ogni caso di lunghezza pari al numero di parametri. (Se il modello ha un solo parametro, un numero in virgola mobile va bene.)
5. *sigma*: le incertezze di misura  $\sigma_{y_i}$  sui valori  $y_i$ . Operativamente può essere semplicemente un numero in virgola mobile—nel qual caso le incertezze sono assunte essere tutte uguali—oppure una lista di Python o un *array* di *numpy* della stessa lunghezza di *xdata* e *ydata*.

Avrete notato che i due argomenti finali *p0* e *sigma* sono opzionali (si vede dal fatto che hanno valori di *default* assegnati nella segnatura della funzione). Nel caso in cui essi non siano specificati esplicitamente dall'utente, il programma li fissa a sequenze di 1 della lunghezza opportuna—cioè i valori iniziali dei parametri sono fissati tutti ad 1 e le incertezze di misura sulla *y* sono pure fissate tutte ad 1.

Bene: sappiamo tutto sugli argomenti che possiamo passare al nostro motore di fit. Ci rimane da discutere è il significato della coppia di oggetti che la funzione ci restituisce.

1. *popt*: un *array* di *numpy* di lunghezza *m* contenente i valori ottimali dei parametri del modello. In altre parole queste sono proprio le stime dei parametri che cerchiamo.
2. *pcov*: la matrice di covarianza  $m \times m$  tra i parametri del modello. Ricordiamo che le incertezze sui parametri sono le radici quadrate degli elementi diagonali della matrice stessa.

```

https://bitbucket.org/.../fit_sin.py
1  import numpy
2  from scipy.optimize import curve_fit
3
4  def f(x, omega):
5      """Definition of the fit function.
6
7      Note that x goes first, then come the parameters.
8      """
9      return numpy.sin(omega*x)
10
11 # Definition of the input data set.
12 numpy.random.seed(100)
13 n = 20
14 dy = 0.05
15 omega = 1.2
16 x = numpy.linspace(1., 19., n)
17 y = f(x, omega) + numpy.random.normal(0, dy, size=n)
18 # Perform the fit and calculate the chisquare
19 popt, pcov = curve_fit(f, x, y, 1., dy)
20 chisq = (((y - f(x, *popt))/dy)**2).sum()
21 print('omega = %.4f +/- %.4f' %\
22       (popt, numpy.sqrt(pcov.diagonal())))
23 print('Chisquare = %.1f' % chisq)
24
25 [Output]
26 omega = 1.2005 +/- 0.0011
27 Chisquare = 11.6

```

FRAMMENTO 8.3. Frammento di codice per il fit della serie di dati in figura 8.9 con il modello (256). Per operare su *array* di *numpy*, nella definizione del modello alla linea 7 abbiamo utilizzato la funzione *sin* di *numpy*, piuttosto che quella del modulo *math* della libreria standard di Python. I dati sono generati aggiungendo al modello in ingresso *rumore* Gaussiano con media nulla e deviazione standard pari all'incertezza di misura su *y*. Il fit vero e proprio è eseguito alla linea 17. L'utilizzo di *\*popt* anziché *popt* nel calcolo del  $\chi^2$  alla riga 18 è un dettaglio tecnico interno a Python che consente di passare l'intero vettore dei valori di *best fit* dei parametri alla funzione—potete leggere qui tutti i dettagli. Notate che la stima finale di  $\omega$  è compatibile, entro l'incertezza, con il valore che abbiamo scelto in fase di generazione dei dati.

A questo punto siamo pronti per analizzare in dettaglio il frammento di codice 8.3. Guardatelo attentamente, leggete la didascalia, e cercate di collegare i singoli pezzi con quanto abbiamo detto fino ad ora.

## 8.4.3 Il problema dei valori iniziali

La peculiarità fondamentale dei minimi quadrati lineari è che, per definizione, tutte le derivate parziali del  $\chi^2$  rispetto ai parametri sono lineari nei parametri stessi. Questo vuol dire che, se guardiamo il  $\chi^2$  come una funzione dei parametri, la dipendenza sarà di tipo quadratico per ciascuno di essi. In un problema semplice ad un parametro come la media pesata il  $\chi^2$  in funzione del valore di *best fit* sarà una parabola e, più in generale, un paraboloide in dimensione  $m$  (dove  $m$  è, al solito il numero di parametri). Una conseguenza immediata di questa semplice osservazione è che in generale un problema di minimi quadrati lineare ha un solo minimo globale—che è il motivo per cui, nei casi di modelli costante e lineare, è stato possibile scrivere in forma chiusa la soluzione.

Ora, tutto questo non è vero in generale. In un fit dei minimi quadrati non lineare non vi è nessuna garanzia che il  $\chi^2$  abbia un solo minimo—tutt'altro: esso può avere un numero arbitrariamente grande di minimi locali che rendono non banale, come vedremo tra un attimo, il problema di trovare la soluzione iterativamente. E non si tratta, è importante sottolinearlo, di una questione puramente accademica—nel senso che non è necessario andare a cercare esempi *ad hoc* particolarmente esotici perché questo tipo di difficoltà si manifesti.

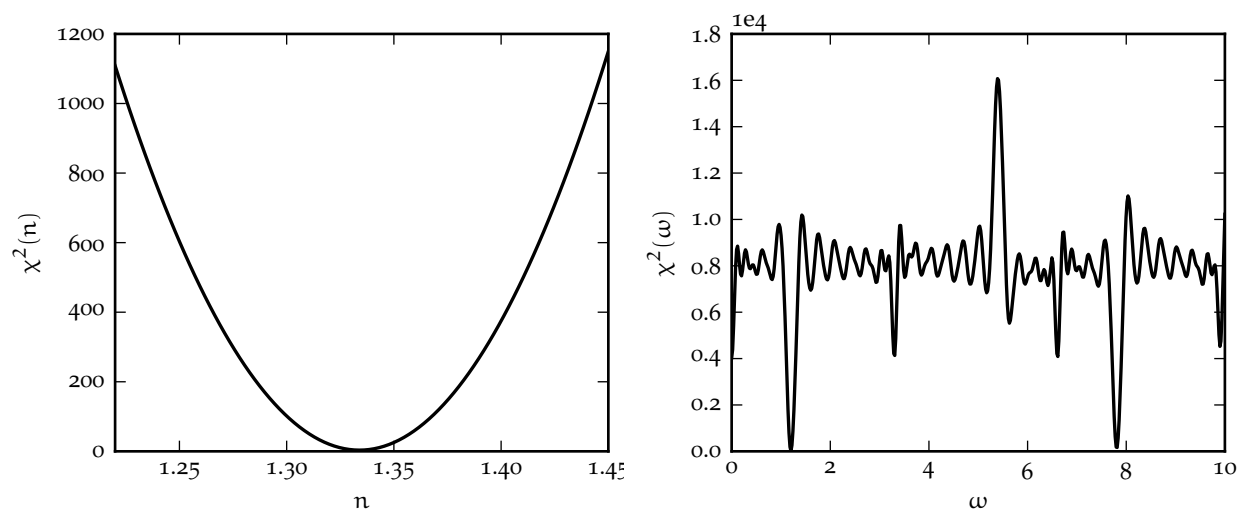


FIGURA 8.10. Esempi di andamenti del valore del  $\chi^2$  in funzione del parametro del modello di fit per la media pesata dell'esempio 8.3 (a sinistra) per il fit con il modello (256) dei dati in figura 8.9 (a destra). Il primo è un problema dei minimi quadrati lineare, il secondo non lineare.

Tutto questo è illustrato per due semplici problemi unidimensionali in figura 8.10 ove il  $\chi^2$  è mostrato in funzione del parametro di fit per la media pesata dell'esempio 8.3 e per il fit con il modello (256) dei dati in figura 8.9. I due grafici sono interessanti almeno per due motivi. Il primo è che essi ci permettono di visualizzare il processo di fit, almeno nel caso semplice di un solo parametro, come il moto di un punto materiale lungo un profilo rigido sotto l'azione delle forze di gravità. L'analogia non è esatta, ma il paragone è calzante. Nel nostro caso abbiamo costruito il profilo del  $\chi^2$  in funzione del valore del parametro con la forza bruta—campionandolo su una griglia regolare di punti. E la forza bruta è un possibile, anche se inefficiente, algoritmo di fit. Un buon metodo iterativo di fit non è altro che un algoritmo efficiente per far arrivare il nostro punto sul minimo del  $\chi^2$  il più velocemente possibile.

La seconda cosa interessante è il diverso ruolo che i valori iniziali dei parametri rivestono nel caso lineare ed in quello non lineare. Nel primo la convergenza è garantita dal fatto che il  $\chi^2$  ha un solo minimo, e partire da un punto molto lontano dal minimo stesso influisce solo sul numero di passi (e quindi sul tempo) necessari per arrivare alla convergenza. Nel caso non lineare la situazione è completamente differente, perché in generale il fit convergerà al minimo più vicino al punto corrispondente ai valori iniziali dei parametri (ricordate l'analogia con la pallina che rotola sul profilo del  $\chi^2$ ) e se questo è solamente un minimo locale, allora i valori di *best fit* che otteniamo sono semplicemente errati. La cosa è estremamente rilevante dal punto di vista pratico. Il fit mostrato in figura 8.9 corrisponde al minimo globale attorno a  $\omega \approx 1.2$  nel grafico di destra di figura 8.10—questo perché, come mostrato nel

frammento 8.3 abbiamo scelto  $\omega = 1$  come valore iniziale del nostro parametro. Se fossimo partiti da un valore iniziale diverso, il fit ci avrebbe restituito un valore diverso—ed errato—per il parametro, come mostrato in figura 8.11.

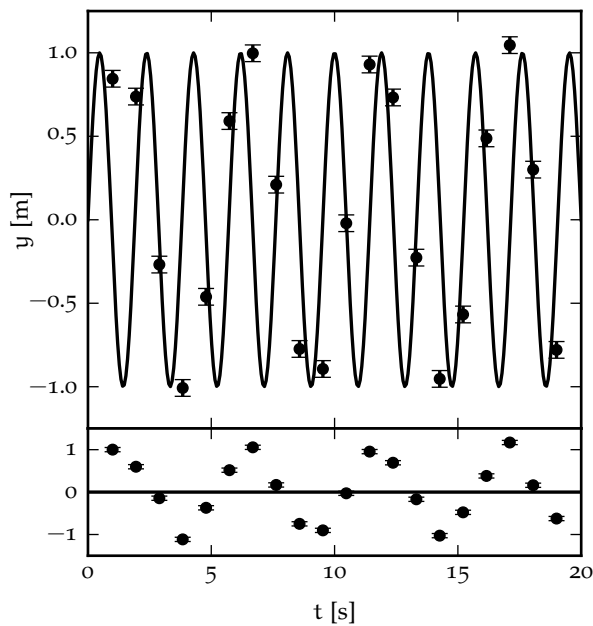


FIGURA 8.11. Esempio di mancata convergenza di un fit dei minimi quadrati non lineare in cui il valore iniziale del parametro non era abbastanza vicino al minimo globale. Il codice utilizzato per il fit è identico a quello mostrato nel frammento 8.3, con l'unica differenza che questa volta abbiamo scelto  $\omega = 3.2$  come valore iniziale. Il valore finale del parametro restituito dal fit è  $\hat{\omega} = 3.299 \pm 0.014 \text{ s}^{-1}$ , con un  $\chi^2$  (orribile) di 4068 per 19 gradi di libertà. (Il grafico dei residui è altrettanto orribile.) Notiamo, per completezza, che il valore a cui converge il fit in questo caso particolare corrisponde al minimo locale più prominente alla destra del minimo globale nel grafico di destra di figura 8.10.

Abbiamo capito che in pratica i fit di tipo numerico sono sensibili ai valori iniziali dei parametri che scegliamo. Lo abbiamo visto chiaramente in un caso relativamente semplice con un solo parametro ed è naturale aspettarsi che le cose non possano che peggiorare quando i parametri in gioco sono molti. A questo punto uno può chiedersi legittimamente: come facciamo ad accorgerci quando il fit non è andato a buon fine? La risposta è semplice: il valore del  $\chi^2$  ed il grafico dei residui sono due indicatori chiari che dovremmo essere perfettamente in grado di interpretare (vedi figura 8.11). La seconda domanda ovvia è: cosa facciamo nei casi in cui il fit non converge alla soluzione corretta? Anche qui la risposta è semplice: ci lasciamo guidare dai dati e cerchiamo per quanto possibile di assegnare valori iniziali ragionevoli (cioè vicini alla soluzione) ai parametri prima di ripetere il fit, come illustrato con un esempio concreto nella prossima sezione.

#### 8.4.4 Un esempio di fit complesso

In questa sezione ci soffermiamo brevemente, a scopo illustrativo, su un problema di fit a molti parametri che descrive un sistema fisico relativamente semplice—un oscillatore armonico smorzato—e lo facciamo non tanto perché il problema sia particolarmente instabile da un punto di vista numerico, quanto perché si tratta di una buona palestra per imparare a ragionare su come si scelgono in modo sensato i valori iniziali dei parametri.

Supponiamo dunque di aver campionato a tempi fissati la posizione  $y$  del nostro oscillatore in funzione del tempo. Il modello è in questo caso

$$f(t; A, \omega, \phi, \lambda, C) = Ae^{-\lambda t} \sin(\omega t + \phi) + C. \quad (260)$$

Abbiamo in totale 5 parametri:  $A$  rappresenta la semiampiezza iniziale di oscillazione,  $e^{-\lambda t}$  è il termine di smorzamento (che è esponenziale se assumiamo che la forza di attrito dipenda linearmente dalla velocità),  $\sin(\omega t + \phi)$  è il termine oscillatorio (in cui sia la pulsazione angolare che la fase sono incognite) e  $C$  rappresenta una costante di *offset* che tiene conto del fatto che l'oscillazione non è necessariamente centrata in 0.

Osserviamo dunque i punti in figura 8.12 e cerchiamo di capire come si possano stimare, uno per uno, i valori iniziali dei parametri.

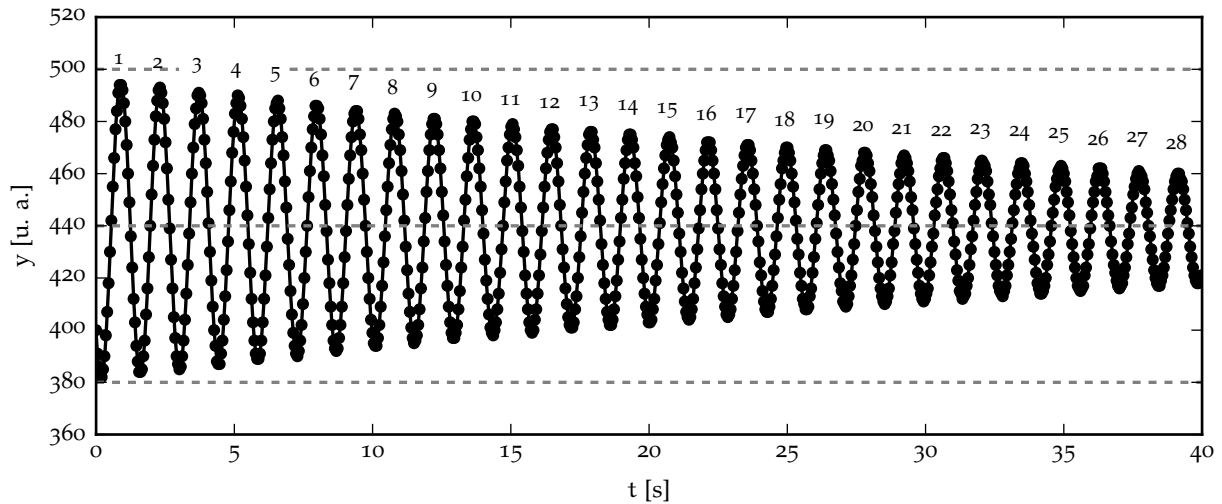


FIGURA 8.12. Esempio di fit dei minimi quadrati della posizione di un oscillatore smorzato in funzione del tempo con un modello a 5 parametri. Nei problemi non lineari a molti parametri la scelta dei valori iniziali per il fit è cruciale per garantire la convergenza—e le tre linee orizzontali in figura corrispondono ai riferimenti utilizzati per la stima dei parametri stessi.

1.  $A$ : come abbiamo detto  $A$  rappresenta la semiampiezza iniziale. Per  $t \approx 0$  i valori di  $y$  variano da circa 380 a circa 500 per cui una stima iniziale ragionevole è

$$A \approx \frac{(500 - 380)}{2} = 60.$$

2.  $\omega$ : se contiamo il numero di massimi (o di minimi) nei dati, possiamo dire che si hanno circa 28 oscillazioni complete in 40 s per cui il periodo di oscillazione e la pulsazione angolare saranno

$$T \approx \frac{40}{28} \approx 1.43 \text{ s} \quad \text{e} \quad \omega = \frac{2\pi}{T} \approx 4.39 \text{ s}^{-1}.$$

3.  $\phi$ : per  $t = 0$  l'oscillazione è prossima al suo punto più basso, per cui, all'ordine 0 possiamo assumere che  $\phi \approx -\pi/2$  o, equivalentemente,  $\phi \approx 3\pi/2$ .

4.  $\lambda$ : questo è interessante. Sappiamo che quando  $t = 1/\lambda$  la nostra ampiezza di oscillazione si è ridotta di un fattore  $1/e$ , ovvero sia a circa un terzo di quella. Possiamo rovesciare l'argomento e dire che, dato che per  $t \approx 40$  s l'ampiezza si è ridotta a circa un terzo di quella iniziale, possiamo stimare

$$\lambda \approx \frac{1}{40} = 0.025 \text{ s}^{-1}.$$

5.  $C$ : ragionando come al punto relativo ad  $A$  possiamo stimare il valore medio (ovverosia il centro) della nostra oscillazione come

$$C \approx \frac{(500 + 380)}{2} = 440.$$

Parametro	Valore iniziale	Valore di <i>best fit</i>
$A$	60	$57.78 \pm 0.06$
$\omega$	4.39	$4.43477 \pm 0.00006$
$\phi$	4.71	$3.895 \pm 0.001$
$\lambda$	0.025	$0.02469 \pm 0.00006$
$C$	440	$438.66 \pm 0.02$

TABELLA 19. Valori iniziali dei parametri (vedi il testo per i dettagli della scelta) e valori finali di *best fit* per il fit con il modello (260) ai dati mostrati in figura 8.12.

A questo punto abbiamo tutto ciò che ci serve per eseguire il fit vero e proprio—il cui risultato è mostrato in figura 8.12 insieme ai punti. La tabella 19 contiene un sommario dei valori iniziali dei parametri che abbiamo scelto e dei valori finali di *best fit*, con le incertezze associate. È degno di nota il fatto che 4 su 5 delle nostre stime iniziali sono entro il 10% dai valori finali.

## 8.5 FIT DI TIPO GENERALE

Ci rimane un ultimo punto importante da discutere: cosa succede quando l'assunzione di base (227) che abbiamo fatto all'inizio del capitolo, ovverosia che le incertezze di misura sulla variabile indipendente siano trascurabili, non è valida? La risposta è semplice: formalmente il metodo dei minimi quadrati nella forma in cui lo abbiamo discusso non può essere utilizzato—e non è un caso che la segnatura della funzione `curve_fit` di `scipy` non preveda un argomento per gli errori sulla  $x$ , poiché, in termini semplici, essi non sono un ingrediente del fit dei minimi quadrati. In pratica ignorare brutalmente gli errori sulla  $x$  ed eseguire ugualmente un fit dei minimi quadrati può avere in generale tre conseguenze distinte:

1. i valori dei parametri restituiti dal fit possono essere sistematicamente diversi dai valori corretti;
2. le incertezze sui parametri sono in generale sistematicamente sottostimate;
3. il valore del  $\chi^2$  che otteniamo è in generale sovrastimato e, dunque, non ha più il significato che gli abbiamo attribuito in tutta la discussione di questo capitolo.

Intendiamoci: gli effetti non sono sempre disastrosi, ma non è necessario andare a cercare esempi troppo esotici per far sí che essi siano apprezzabili. Al solito discuteremo l'argomento sulla base di un esempio concreto.

Supponiamo dunque di voler stimare il potere diottrico  $1/f$  di una lente convergente sfruttando la relazione per le lenti sottili tra la distanza oggetto-lente  $p$  e la distanza lente-immagine  $q$

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f}. \quad (261)$$

A questo scopo possiamo misurare una serie di  $n$  coppie di valori  $(p_i, q_i)$  con le rispettive incertezze di misura  $\sigma_{p_i}$  e  $\sigma_{q_i}$ . Con il cambiamento di variabile<sup>5</sup>

$$x_i = \frac{1}{p_i} \quad \sigma_{x_i} = \frac{\sigma_{p_i}}{p_i^2} \quad y_i = \frac{1}{q_i} \quad \sigma_{y_i} = \frac{\sigma_{q_i}}{q_i^2} \quad (262)$$

il nostro modello (261) diviene un modello lineare—e anzi un caso particolare di modello lineare poiché, se la teoria è corretta, il coefficiente angolare è previsto essere  $m = -1$ :

$$y = -x + \frac{1}{f}.$$

Possiamo dunque fare un fit dei nostri dati e stimare il potere diottrico cercato come l'intercetta della retta di *best fit*. Nella discussione che segue lasceremo il coefficiente della retta libero di variare nel fit sulla base del fatto che, in casi come questo, il valore del parametro restituito dal fit (o meglio, il suo livello di accordo con il valore atteso) costituisce un controllo utile di consistenza interna della misura.

Osserviamo con attenzione i punti sperimentali in figura 8.13. La prima domanda cui dobbiamo rispondere è: in questo caso specifico la condizione (227) è verificata oppure no? È chiaro che la risposta è no, poiché la derivata del modello rispetto alla variabile indipendente è pari al coefficiente angolare della retta di *best fit*, cioè dell'ordine di  $-1$ , per cui di fatto possiamo confrontare direttamente gli errori su  $x$  e quelli su  $y$ , e vediamo chiaramente che per i punti più a destra si ha  $\sigma_{x_i} \gg \sigma_{y_i}$ . Esattamente il contrario di quello che vorremmo.

Proviamo per un attimo a chiudere gli occhi e procedere ugualmente. La linea continua in figura 8.13 rappresenta il risultato di un fit dei minimi quadrati (ignorando banalmente gli errori sulla  $x$ ) con un

<sup>5</sup> C'è un'ulteriore piccola complicazione connessa a questo cambiamento di variabile: se assumiamo che gli errori su  $p$  e  $q$  siano Gaussiani, quelli sugli inversi in generale non lo saranno, ma a questo livello si tratta di un dettaglio tecnico che ignoriamo felicemente.

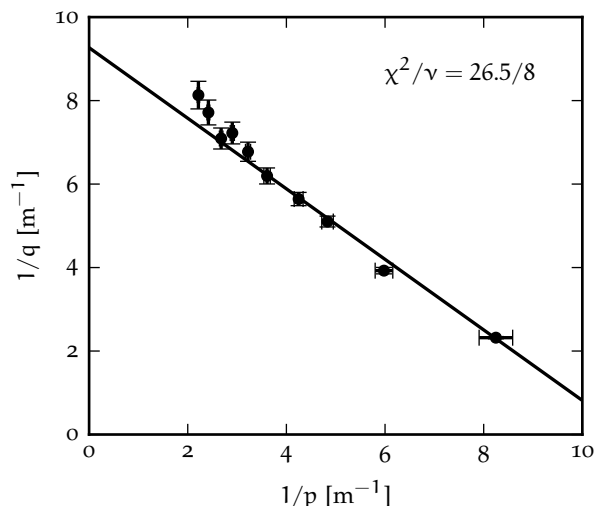


FIGURA 8.13. Esempio di fit dei minimi quadrati con un modello lineare eseguito ignorando brutalmente gli errori sulla variabile indipendente in una situazione in cui la (227) non è verificata. I valori dei parametri restituiti dal fit sono  $m = -0.845 \pm 0.030$  e  $1/r = 9.27 \pm 0.22$ . Notiamo esplicitamente che il coefficiente angolare è inconsistente con il valore atteso  $-1$ . Il  $\chi^2$  del fit è pari a 26.5, che per 8 gradi di libertà corrisponde ad un  $p$ -value al di sotto di  $10^{-3}$ .

modello lineare, in cui abbiamo lasciato liberi di variare, come abbiamo detto, sia l'intercetta che il coefficiente angolare. Il primo problema che ci troviamo a dover affrontare è che il coefficiente angolare restituito dal fit,  $m = -0.845 \pm 0.030$  non è compatibile con il valore atteso dalla teoria ( $-1$ ). Faremo l'ipotesi di lavoro (che verificheremo a posteriori) che le misure siano corrette e che il problema sia nel modo in cui abbiamo effettuato il fit. Ed in effetti non è difficile rendersi conto che qualcosa non va: la retta di *best fit*, per come abbiamo formulato il problema, è costretta a passare per il punto in basso a destra, che ha un errore sulla  $y$  relativamente piccolo. Eppure l'errore corrispondente sulla  $x$  (che, lo ricordiamo, abbiamo ignorato) è molto grande, ed è naturale pensare che se trovassimo il modo di metterlo in gioco, il risultato del fit potrebbe essere significativamente diverso.

Il secondo problema è che il  $\chi^2$  del fit è troppo alto: 26.5 per 8 gradi di libertà, che corrisponde ad un  $p$ -value inferiore a  $10^{-3}$ . Ma in fondo questo non dovrebbe stupire, visto che di fatto a abbiamo deliberatamente ignorato una parte consistente delle incertezze di misura.

#### 8.5.1 L'algoritmo di orthogonal distance regression (ODR)

Abbiamo capito che il fit dei minimi quadrati nel caso in cui gli errori sulla variabile indipendente non sono trascurabili è, come ci aspettavamo, problematico. La domanda ovvia, a questo punto, è: esistono algoritmi alternativi che ci permettono di attaccare il problema generale? E quali caratteristiche dovrebbe avere un algoritmo di questo tipo?

La risposta, chiaramente, è sì. E possiamo dire di più: intuitivamente ci aspettiamo di dover modificare la metrica che abbiamo usato fino a questo momento per valutare l'accordo fra dati e modello, vale a dire il  $\chi^2$  della (228), sotto due aspetti differenti:

1. al numeratore non possiamo più valutare semplicemente il modello nei punti misurati  $x_i$ , che non sono più fissati—il che equivale a dire che dobbiamo calcolare la distanza tra dati e modello non verticalmente, ma diagonalmente, come mostrato in figura 8.14;
2. al denominatore dobbiamo pesare opportunamente le due componenti ortogonali (sulla  $x$  e sulla  $y$ ) delle distanze di cui sopra, con le rispettive incertezze di misura.

È chiaro che questo problema è significativamente più complesso di quello dei minimi quadrati: non solo non vi è alcuna speranza di risolverlo analiticamente, ma anche implementare un algoritmo iterativo per la soluzione è cosa tutt'altro che banale.

Per fortuna *scipy* ci viene incontro con il pacchetto *odr* (per *orthogonal distance regression*) [1] che fa esattamente ciò che abbiamo appena detto. Il risultato, mostrato in figura 8.15 per gli stessi dati di figura 8.13 è più che incoraggiante. L'utilizzo delle incertezze di misura sulla variabile indipendente fa sì che la retta di *best fit* non sia più costretta a passare per l'ultimo punto (quello in basso a destra) e la nostra miglior stima per il coefficiente angolare è adesso  $m = -0.979 \pm 0.051$ —compatibile con il valore atteso  $-1$ . Il valore del parametro fisico che vogliamo stimare, ossia il potere diottrico della lente, è



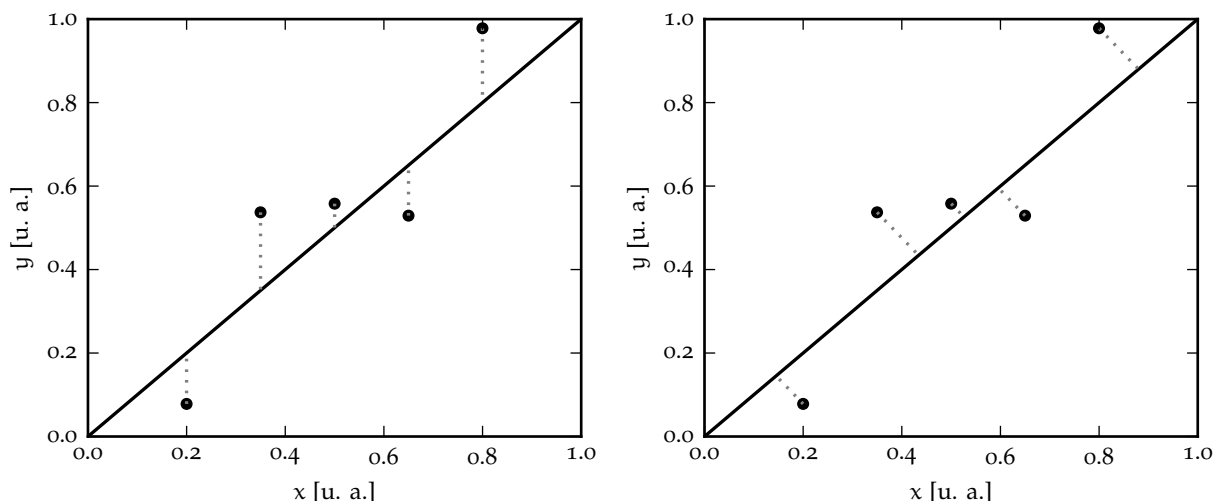


FIGURA 8.14. Rappresentazione grafica delle differenze che vanno al numeratore dei singoli termini del  $\chi^2$ , nel caso in cui gli errori sulla  $x$  siano trascurabili (sinistra) e nel caso dell'algoritmo di *orthogonal distance regression*. Per chiarezza di illustrazione le barre d'errore sui punti misurati non sono rappresentate.

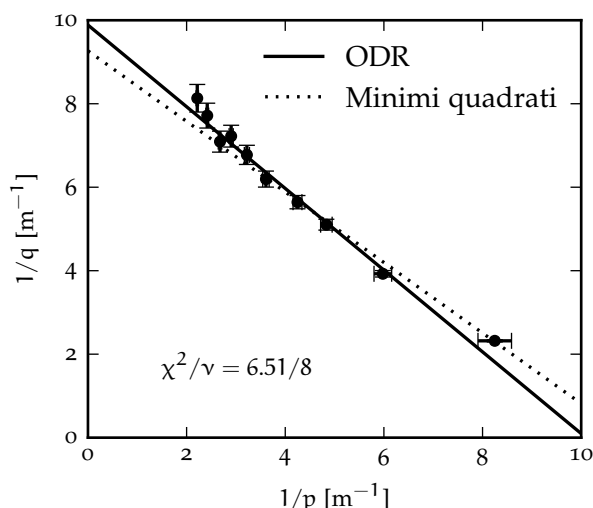


FIGURA 8.15. Esempio di fit ai dati di figura 8.13 con l'algoritmo di *orthogonal distance regression* della libreria *odr* di *scipy*. I valori dei parametri restituiti dal fit sono  $m = -0.979 \pm 0.051$  (questa volta compatibile con  $-1$ ) e  $1/\tau = 9.89 \pm 0.23$ . L'equivalente del  $\chi^2$  del fit è 6.52 per 8 gradi di libertà—cioè il fit è un buon fit. Per confronto la linea tratteggiata corrisponde al fit dei minimi quadrati mostrato in figura 8.13.

significativamente diverso dal caso precedente—e senza dubbio più corretto. E il  $\chi^2$  del fit è adesso 6.52 per 8 gradi di libertà, che indica che il fit è un buon fit.

L'unico notizia negativa, se così vogliamo dire, relativa a tutta la questione è che, per motivi storici, l'interfaccia di `scipy.odr` è completamente diversa da quella di `scipy.optimize.curve_fit`, per cui di fatto dobbiamo imparare una cosa nuova praticamente da zero. Persino l'ordine degli argomenti nella definizione del modello è invertita—prima i parametri e poi la variabile indipendente. Il codice utilizzato per eseguire il fit mostrato in figura 8.15 è riportato nel frammento 8.4 e commentato ampiamente nella didascalia corrispondente.

### 8.5.2 Una possibile modifica al metodo dei minimi quadrati

L'algoritmo di *orthogonal distance regression* che abbiamo visto nella sezione precedente costituisce una soluzione completa del problema generale di fit nel caso in cui gli errori sulla variabile indipendente non siano trascurabili, per cui possiamo chiederci se il metodo dei minimi quadrati in fondo non sia da buttare via.

La risposta è no. Una strategia alternativa è quella di considerare i punti  $x_i$  fissati e di propagare le incertezze corrispondenti  $\sigma_{x_i}$  sulla variabile indipendente, sommandole in quadratura (come si fa

```

https://bitbucket.org/.../odr.py
1 import numpy
2 from scipy.odr import odrpack
3
4 def f(pars, x):
5     """Definition of the (linear) fit function.
6
7     Note that the parameters go first, as a tuple,
8     then comes the actual array of x values.
9     """
10    return pars[0]*x + pars[1]
11
12 # Read the data from file.
13 x, dx, y, dy = numpy.loadtxt('data/lens.dat', unpack=True)
14 # Run the actual ODR.
15 model = odrpack.Model(f)
16 data = odrpack.RealData(x, y, sx=dx, sy=dy)
17 odr = odrpack.ODR(data, model, beta0=(1., 1.))
18 out = odr.run()
19 popt, pcov = out.beta, out.cov_beta
20 m, q = popt
21 dm, dq = numpy.sqrt(pcov.diagonal())
22 chisq = out.sum_square
23 # Print the fit output.
24 print('m = %.3f +/- %.3f, q = %.2f, %.2f' % (m, dm, q, dq))
25 print('Chisquare = %.1f' % chisq)
26
27 [Output]
28 m = -0.979 +/- 0.051, q = 9.89, 0.23
29 Chisquare = 6.5

```

FRAMMENTO 8.4. Frammento di codice per il fit dei dati in figura 8.15 con l'algoritmo di *orthogonal distance regression* implementato nel modulo `scipy.odr` di `numpy`. Tra le differenze in termini di interfacce rispetto alla funzione `scipy.optimize.curve_fit` notiamo che l'ordine degli argomenti della funzione che esprime il modello è invertito: prima i parametri (nella forma di una tupla e non separatamente) e poi la variabile indipendente. Il fit in sé, inoltre, non si risolve in una semplice chiamata di una funzione, ma richiede la preparazione di un certo numero di oggetti dedicati: il modello, i dati e l'algoritmo vero e proprio. L'argomento `beta0` del costruttore di `odrpack` rappresenta i valori iniziali dei parametri di fit.

per errori indipendenti) alle incertezze di partenza  $\sigma_{y_i}$ . In altre parole, possiamo definire delle nuove incertezze *efficaci*  $\sigma_i$  come

$$\sigma_i^2 = \sigma_{y_i}^2 + \left( \frac{df(x_i; \hat{\theta}_1 \dots \hat{\theta}_m)}{dx} \right)^2 \sigma_{x_i}^2 \quad (263)$$

ed utilizzare queste ultime in un fit dei minimi quadrati, minimizzando la quantità

$$\chi^2(\theta_1, \dots, \theta_m) = \sum_{i=1}^n \left( \frac{y_i - f(x_i; \theta_1, \dots, \theta_m)}{\sigma_i} \right)^2 = \sum_{i=1}^n \frac{(y_i - f(x_i; \theta_1, \dots, \theta_m))^2}{\sigma_{y_i}^2 + \left( \frac{df(x_i; \theta_1 \dots \theta_m)}{dx} \right)^2 \sigma_{x_i}^2} \quad (264)$$

Questo problema è, di nuovo, significativamente più complicato di quello di partenza, poiché il modello entra non solo al numeratore, ma anche, attraverso la sua derivata rispetto alla variabile indipendente, al denominatore. (Per inciso, non è difficile convincersi che, nelle ipotesi iniziali dei minimi quadrati (227), il secondo termine della (263) è per definizione trascurabile e torniamo esattamente nelle condizioni del minimo  $\chi^2$ .)

L'idea di base, a questo punto, è che possiamo procedere in tre passi:

1. eseguiamo un primo fit dei minimi quadrati ignorando brutalmente le incertezze sulla  $x$  (sappiamo che il risultato non sarà ottimale, ma sappiamo anche che, in condizioni ordinarie, non sarà nemmeno lontanissimo dalla soluzione);
2. utilizziamo i valori di *best fit* dei parametri che abbiamo ottenuto al passo 1 per stimare la derivata del modello in corrispondenza dei punti misurati

$$\frac{df(x_i; \hat{\theta}_1 \dots \hat{\theta}_m)}{dx}$$

e la utilizziamo per combinare le incertezze di misura sulle due coordinate, come indicato nella (263);

3. eseguiamo un nuovo fit dei minimi quadrati utilizzando le nuove stime delle incertezze.

In pratica può essere necessario ripetere i passi 2 e 3 un certo numero di volte per evitare che una stima non sufficientemente accurata dei parametri del modello al primo passo si rifletta in un valore inaccurato della derivata—e quindi dell'errore finale. Tipicamente, a meno di casi patologici, poche iterazioni sono sufficienti a raggiungere la convergenza.

```

https://bitbucket.org/.../least_squares_modified.py
1 import numpy
2 from scipy.optimize import curve_fit
3
4 def f(x, m, q):
5     """Definition of the (linear) fit function.
6     """
7     return m*x + q
8
9 # Read the data from file.
10 x, dx, y, dy = numpy.loadtxt('data/lens.dat', unpack=True)
11 # Run a first least-square fit (disregard dx).
12 popt, pcov = curve_fit(f, x, y, (1., 1.), dy)
13 # Iteratively update the errors and refit.
14 for i in range(3):
15     dxy = numpy.sqrt(dy**2. + (popt[0]*dx)**2.)
16     popt, pcov = curve_fit(f, x, y, popt, dxy)
17     chisq = (((y - f(x, *popt))/dxy)**2.).sum()
18     # Print the fit output.
19     print('Passo %d...' % i)
20     print(popt)
21     print(numpy.sqrt(pcov.diagonal()))
22     print(chisq)
23
24 [Output]
25 Passo 0...
26 [-0.9631888  9.81976774]
27 [ 0.04375748  0.19966969]
28 7.31388800403
29 Passo 1...
30 [-0.97092652  9.85277658]
31 [ 0.04469915  0.19821819]
32 6.57835391917
33 Passo 2...
34 [-0.97139716  9.8547628 ]
35 [ 0.04476226  0.19815346]
36 6.53672021395

```

FRAMMENTO 8.5. Frammento di codice per illustrare l'algoritmo dei minimi quadrati, modificato come spiegato nel testo per tenere in considerazione le incertezze di misura sulla variabile indipendente. Il fit iniziale è eseguito alla linea 13, e ci fornisce una prima stima del coefficiente angolare e dell'intercetta. I passi 2 e 3 descritti nel testo sono inclusi in un ciclo *for*, che ci permette di ripeterli per il numero di volte desiderato (nel nostro caso 3). Ad ogni passo la variabile *popt* contiene i valori di *best fit* dei due parametri, e l'elemento 0 dell'*array* è proprio la derivata di cui abbiamo bisogno per combinare gli errori secondo la (263)—il che avviene alla linea 16. (Il resto dovrebbe essere auto-esplicativo.) Il processo converge in pochi passi, e già alla prima iterazione il risultato è molto più vicino a quello corretto che non nel caso del fit dei minimi quadrati originale.

Il frammento di codice 8.5 illustra il concetto nel caso del fit lineare mostrato in figura 8.15. I modelli lineari si prestano particolarmente bene a questo tipo di approccio poiché il coefficiente angolare di *best fit* costituisce direttamente una stima della derivata necessaria per combinare gli errori sulla *x* e sulla *y*. Nel caso specifico il nostro fit dei minimi quadrati modificato converge in tre o quattro passi ad una soluzione relativamente vicina a quella fornita dal pacchetto *scipy.odr*—sia in termini di valori dei parametri, che in termini di incertezze sui parametri stessi e valori del  $\chi^2$ . Un confronto completo è riportato in tabella 20.

Parametro	MQ	ODR	MQ modificati
m	$-0.845 \pm 0.030$	$-0.978 \pm 0.051$	$-0.971 \pm 0.045$
q	$9.27 \pm 0.22$	$9.89 \pm 0.23$	$9.85 \pm 0.20$
$\chi^2$	26.5	6.51	6.54

TABELLA 20. Confronto tra i valori di *best fit* restituiti dai tre diversi algoritmi nel caso di un fit lineare ai dati mostrati nelle figure 8.13 e 8.15.

## 8.6 IN BREVE...

---

CENNI AL PRINCIPIO DI MASSIMA VEROSIMIGLIANZA

---

Torniamo per un secondo al secondo dei problemi che abbiamo introdotto a pagina 41: abbiamo un'urna contenente 5 palline—alcune rosse ed alcune blu (ma non conosciamo la proporzione tra i colori). Estraiamo una pallina, che risulta essere blu, e la mettiamo da parte. Estraiamo una seconda pallina, che stavolta è rossa. Qual è il numero di palline rosse nell'urna? Abbiamo già avuto modo di vedere come questa domanda non abbia una risposta univoca—e sicuramente non senza precisare meglio le condizioni al contorno del problema. Ma è giunto il momento di riesaminare il problema criticamente alla luce di ciò che abbiamo imparato lungo il percorso.

Supponiamo dunque per un attimo che l'urna contenga 2 palline rosse e 3 blu. In questo caso la probabilità che l'esito della nostra estrazione sia quello descritto nel problema—una pallina blu seguita da una rossa, è esattamente  $P = 3/5 \times 2/4 = 3/10$ . (L'uguaglianza è da leggere come: alla prima estrazione abbiamo 3 palline blu su 5, per cui la probabilità di estrarre una pallina blu è  $3/5$ ; alla seconda estrazione abbiamo 2 palline rosse su 4 rimanenti, per cui la probabilità di estrarre una pallina rossa è  $2/4$ ; poiché gli eventi sono indipendenti le probabilità si moltiplicano.) Ovviamente questo non risponde alla domanda di partenza, perché nessuno ci dice che nell'urna ci fossero effettivamente 2 palline rosse e 3 blu. Ci fornisce però una possibile chiave di lettura per affrontare il problema—per ogni configurazione iniziale dell'urna possiamo calcolare la probabilità di ottenere l'esito delle estrazioni di nostro interesse, vale a dire una pallina blu seguita da una rossa, come mostrato nella tabella 21.

$n_B$	$n_R$	$P$
0	5	$0/5 \times 4/4 = 0$
1	4	$1/5 \times 4/4 = 4/20$
2	3	$2/5 \times 3/4 = 6/20$
3	2	$3/5 \times 2/4 = 6/20$
4	1	$4/5 \times 1/4 = 4/20$
5	0	$5/5 \times 0/4 = 0$

TABELLA 21. Tabella dei valori della probabilità di estrarre una pallina blu seguita da una rossa per una qualsiasi configurazione iniziale di un'urna contenente  $n_B + n_R = 5$  palline.

A questo punto possiamo tentare di chiudere il cerchio dicendo che, se dovessimo scommettere sul numero di palline blu (o rosse) nell'urna, sarebbe sensato scegliere una delle due configurazioni ( $n_B = 2$  e  $n_R = 3$  oppure  $n_B = 3$  e  $n_R = 2$ ) che massimizzano la probabilità di ottenere la sequenza di estrazioni cercata. Come vedremo tra un attimo abbiamo appena enunciato il *principio di massima verosimiglianza* (o *maximum likelihood*).

### 9.1 DI NUOVO SUL CONCETTO DI VEROSIMIGLIANZA

Abbiamo incontrato il concetto di verosimiglianza per la prima volta nella sezione 3.5.2, e lo abbiamo utilizzato, anche se in modo implicito, per risolvere il problema generale della misura di efficienza nella sezione 5.2. Adesso è il momento di affrontare il problema in modo più organico.

Dato un modello probabilistico arbitrario, funzione di  $m$  parametri  $\theta_1 \dots \theta_m$ , la funzione di verosimiglianza non è altro che la probabilità di osservare un determinato esito  $d$  (che tecnicamente, lo ricordiamo, è una probabilità condizionata per la variabile casuale  $d$ ) *vista come funzione dei valori dei parametri del modello*

$$\mathcal{L}(\theta_1, \dots, \theta_m | d) = P(d | \theta_1, \dots, \theta_m). \quad (265)$$

Sembra poco più di un gioco di parole, ma come vedremo si tratta di un passaggio che ha implicazioni profonde. Lo illustriamo con un esempio.

### 9.1.1 Un esempio concreto: il campionamento singolo di una variabile Poissoniana

Supponiamo di osservare un processo Poissoniano con media  $\mu$ —per fissare le idee il numero  $k$  di decadimenti di un determinato campione radioattivo in 10 secondi. Se conosciamo *esattamente* la media  $\mu$  (per esempio  $\mu = 3$ ) possiamo scrivere, con ovvio significato dei termini,

$$\mathcal{P}(k; \mu = 3) = \frac{3^k}{k!} e^{-3}. \quad (266)$$

Questa espressione ci dice la probabilità che, fissata la media  $\mu$ , si misuri un generico valore di  $k$ ; si tratta, appunto, di una tipica proposizione di calcolo delle probabilità. Ma il fisico sperimentale si pone il problema inverso: da un valore *misurato*  $k$  (per esempio  $k = 3$ ) si vuole inferire il valore *incognito* di  $\mu$ . Allora possiamo riscrivere la (266) considerando  $k$  come costante e  $\mu$  come variabile indipendente:

$$\mathcal{L}(\mu; k = 3) = \frac{\mu^3}{3!} e^{-\mu}. \quad (267)$$

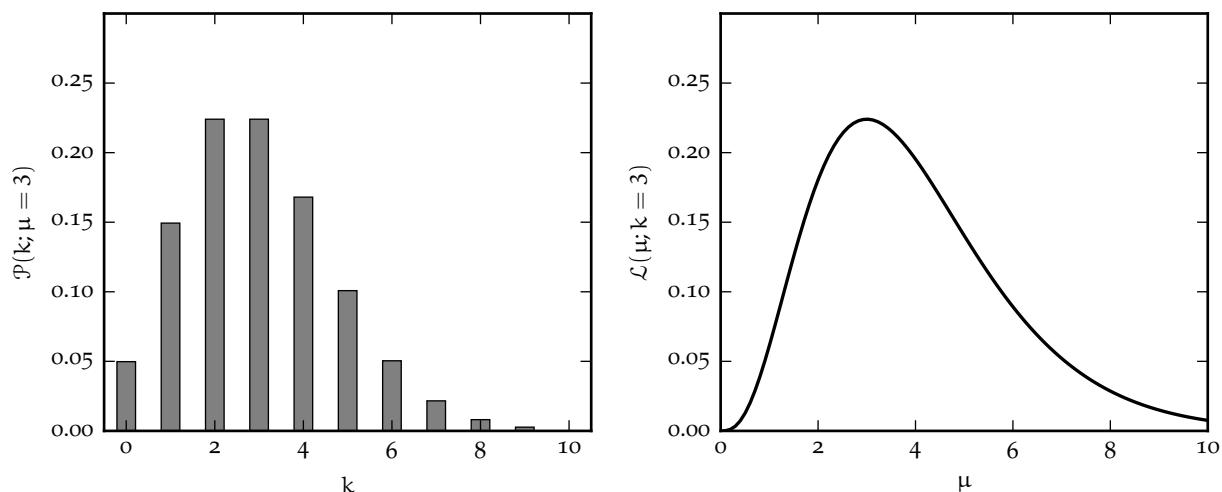


FIGURA 9.1. Distribuzione di Poisson  $\mathcal{P}(k; \mu = 3)$  e corrispondente funzione di verosimiglianza  $\mathcal{L}(\mu; k = 3)$ .

La distribuzione di Poisson  $\mathcal{P}(k; \mu = 3)$  e la corrispondente funzione di verosimiglianza  $\mathcal{L}(\mu; k = 3)$  sono mostrate in figura 9.1—a costo di essere pedanti ricordiamo che la prima è una funzione di distribuzione della variabile discreta  $k$ , la seconda è una funzione della variabile continua  $\mu$ . Il fatto che le due funzioni siano per certi aspetti simili nella *forma* non deve ingannare; è interessante notare, ad esempio, che la funzione di distribuzione è diversa da zero per  $k = 0$  mentre la funzione di verosimiglianza si annulla per  $\mu = 0$ —il che è semplicemente dovuto al fatto che, mentre da una Poissoniana con media  $\mu = 3$  si possono ottenere 0 conteggi, da una Poissoniana con media  $\mu = 0$  *non* si possono ottenere 3 conteggi. In sostanza non si tratta di una mera questione di notazione: come vedremo tra breve la distinzione tra la (266) e la (267) è profonda e gravida di conseguenze.

La funzione di verosimiglianza non è in generale una funzione di distribuzione, poiché non vi è nessuna ragione che essa sia correttamente normalizzata se integrata su tutti i valori possibili del parametro (o dei parametri) del modello. Il nostro esempio, incidentalmente, è una sorta di eccezione, nel senso che si può dimostrare per integrazione diretta che

$$\int_0^\infty \mathcal{L}(\mu; k = 3) d\mu = \int_0^\infty \frac{\mu^3}{3!} e^{-\mu} d\mu = 1.$$

La cosa non è in fondo così importante perché in questo contesto, come abbiamo visto, la cosa che ci interessa di più è il valore  $\hat{\mu}$  di  $\mu$  per cui la funzione di verosimiglianza assume il valore massimo. La derivata di  $\mathcal{L}$  è

$$\frac{d\mathcal{L}(\mu; k=3)}{d\mu} = \frac{\mu^2 e^{-\mu}(3-\mu)}{3!},$$

da cui  $\hat{\mu} = 3$ . In altre parole se misuriamo 3 conteggi, il principio di massima verosimiglianza ci dice che la nostra miglior stima della media del processo Poissoniano sottostante è 3. Ragionevole, no?

## 9.2 STIME DI MASSIMA VEROSIMIGLIANZA

Cerchiamo di formalizzare il problema in termini più generali. Sia dato un insieme di  $n$  misure sperimentali  $x_1, \dots, x_n$  ed un modello dipendente da  $m$  parametri  $\theta_1, \dots, \theta_m$ . Fissato un generico vettore di valori dei parametri il modello definisce univocamente la probabilità (o la densità di probabilità, nel caso i dati siano variabili continue) di ottenere proprio le nostre misure. Adesso noi scriviamo la funzione di verosimiglianza come

$$\mathcal{L}(\theta_1, \dots, \theta_m; x_1, \dots, x_n) = p(\theta_1, \dots, \theta_m | y_1, \dots, y_n). \quad (268)$$

Il *principio di massima verosimiglianza* si può enunciare dicendo che la miglior stima  $\hat{\theta}$  del vettore dei parametri  $\theta$  è quella che massimizza la funzione di verosimiglianza:

$$\begin{cases} \frac{\partial \mathcal{L}(\hat{\theta}_1, \dots, \hat{\theta}_m; x_1, \dots, x_n)}{\partial \theta_1} = 0 \\ \vdots \\ \frac{\partial \mathcal{L}(\hat{\theta}_1, \dots, \hat{\theta}_m; x_1, \dots, x_n)}{\partial \theta_m} = 0. \end{cases} \quad (269)$$

### 9.2.1 Stime di massima verosimiglianza nel caso Poissoniano

Supponiamo di aver campionato  $n$  volte una variabile Poissoniana  $k$  con media (incognita)  $\mu$  e di aver ottenuto i valori  $k_1, \dots, k_n$ . Come possiamo applicare il principio di massima verosimiglianza per stimare  $\mu$ ?

Per prima cosa scriviamo la probabilità  $P_i$  di ottenere il (singolo) valore  $k_i$  per un generico valore di  $\mu$ :

$$P_i = \mathcal{P}(k_i; \mu) = \frac{\mu^{k_i}}{k_i!} e^{-\mu}.$$

Ora, se i campionamenti sono indipendenti, la probabilità  $P$  di ottenere esattamente la nostra sequenza  $k_1, \dots, k_n$  non è altro che il prodotto delle probabilità singole

$$P = \prod_{i=1}^n P_i = \prod_{i=1}^n \frac{\mu^{k_i}}{k_i!} e^{-\mu},$$

e questa è anche la nostra verosimiglianza

$$\mathcal{L}(\mu; x_1, \dots, x_n) = \prod_{i=1}^n \frac{\mu^{k_i}}{k_i!} e^{-\mu}. \quad (270)$$

L'espressione 270 non è particolarmente comoda da derivare per trovare il massimo—essenzialmente perché la derivata del prodotto non è uguale al prodotto delle derivate. La funzione logaritmo ci viene però in aiuto per due motivi:

1. il logaritmo trasforma prodotti in somme (e la derivata della somma è uguale alle somma delle derivate);

2. il logaritmo è una funzione strettamente crescente, per cui il massimo di  $\ln \mathcal{L}$  corrisponde al massimo di like.

Scriviamo allora

$$\ln \mathcal{L}(\mu; k_1, \dots, k_n) = \sum_{i=1}^n \ln \left( \frac{\mu^{k_i}}{k_i!} e^{-\mu} \right) = \sum_{i=1}^n (k_i \ln \mu - \ln(k_i!) - \mu), \quad (271)$$

la cui derivata rispetto a  $\mu$  non è altro che

$$\frac{d \ln \mathcal{L}(\mu; k_1, \dots, k_n)}{d\mu} = \sum_{i=1}^n \left( \frac{k_i}{\mu} - 1 \right) = \frac{1}{\mu} \sum_{i=1}^n k_i - n,$$

che si annulla per

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n k_i. \quad (272)$$

Non sorprendentemente, il principio di massima verosimiglianza ci dice che la miglior stima della media del processo Poissoniano non è altro che la media aritmetica dei nostri campionamenti.

### 9.2.2 Stime di massima verosimiglianza nel caso Gaussiano

Ripetiamo adesso l'esercizio della sezione precedente nel caso di un campionamento ripetuto di una variabile Gaussiana con media  $\mu$  e deviazione standard  $\sigma$ —entrambe ignote.

Detti  $x_1, \dots, x_n$  gli  $n$  campionamenti della nostra variabile la funzione di verosimiglianza si scrive come

$$\mathcal{L}(\mu, \sigma; x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2}$$

ed il suo logaritmo è banalmente

$$\ln \mathcal{L}(\mu, \sigma; x_1, \dots, x_n) = \sum_{i=1}^n \ln \left( \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2} \right) = \sum_{i=1}^n \left[ -\ln \sigma - \frac{1}{2} \ln(2\pi) - \frac{1}{2} \left( \frac{x_i - \mu}{\sigma} \right)^2 \right]$$

Il principio di massima verosimiglianza si scrive in questo caso come

$$\begin{cases} \frac{\partial \ln \mathcal{L}(\hat{\mu}, \hat{\sigma}; x_1, \dots, x_n)}{\partial \mu} = \sum_{i=1}^n \frac{(x_i - \hat{\mu})}{\hat{\sigma}^2} = 0 \\ \frac{\partial \ln \mathcal{L}(\hat{\mu}, \hat{\sigma}; x_1, \dots, x_n)}{\partial \sigma} = \sum_{i=1}^n \left[ -\frac{1}{\hat{\sigma}} + \frac{(x_i - \hat{\mu})^2}{\hat{\sigma}^3} \right] = 0. \end{cases}$$

Ora, la prima delle due equazioni fornisce banalmente il risultato

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (273)$$

cioè, di nuovo, la miglior stima della media coincide con la media aritmetica dei nostri campionamenti. La seconda equazione si può riscrivere come

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2, \quad (274)$$

che corrisponde con la stima della varianza  $s_n^2$  definita nella (128). (I più acuti avranno notato che in questo caso la stima di massima verosimiglianza della varianza non è, per definizione, imparziale.)



### 9.2.3 Stime di massima verosimiglianza nel caso uniforme

A questo punto il lettore potrebbe essere indotto a pensare che il fatto che la stima di massima verosimiglianza della media di una distribuzione sia data dalla media aritmetica dei campionamenti sia un fatto generale—in fondo vale sia per il caso Poissoniano che per quello Gaussiano. Niente di più lontano dal vero, ed in effetti vedremo nella prossima sezione che nel caso di variabili continue la distribuzione di Gauss è l'unica a godere di questa (apparentemente banale) proprietà.

In questa sezione esaminiamo invece brevemente il caso interessante di un numero  $n$  campionamenti  $x_1, \dots, x_n$  di una variabile casuale  $x$  continua che sappiamo essere distribuita uniformemente—lo scopo è di stimare, utilizzando il principio di massima verosimiglianza, gli estremi  $a$  e  $b$  dell'intervallo di variabilità di  $x$ .

Procedendo come prima scriviamo la funzione di verosimiglianza

$$\mathcal{L}(a, b; x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{b-a} = \left( \frac{1}{b-a} \right)^n. \quad (275)$$

Notate niente di strano? Per prima cosa la funzione di verosimiglianza *non dipende dai campionamenti*  $x_i$ . Ma la cosa più grave è che il massimo di  $\mathcal{L}$  non si può trovare per derivazione per il semplice fatto che la (275) non ha massimo—e diverge per  $a = b$ . Ora, a meno che i nostri campionamenti non siano tutti identici noi sappiamo che questa condizione non può essere verificata, ed in effetti abbiamo i vincoli

$$a \leq \min_i \{x_i\} \quad \text{e} \quad b \geq \max_i \{x_i\}.$$

Allora banalmente, essendo il valore della verosimiglianza tanto più grande quanto più  $a$  e  $b$  sono vicini tra loro, siamo costretti ad ammettere che la soluzione del problema sia proprio

$$a = \min_i \{x_i\} \quad \text{e} \quad b = \max_i \{x_i\}. \quad (276)$$

Questo significa che la stima di massima verosimiglianza della media della distribuzione

$$\hat{\mu} = \frac{1}{2} \left( \min_i \{x_i\} + \max_i \{x_i\} \right) \quad (277)$$

coincide con la media aritmetica dei due campionamenti estremi e non dipende dagli altri  $n-2$ . Interessante, vero?

### 9.3 DIGRESSIONE: LA DISTRIBUZIONE DI GAUSS E GLI ERRORI DI MISURA

È interessante illustrare brevemente come, storicamente, Gauss sia arrivato a derivare la forma analitica della distribuzione che porta il suo nome.

Supponiamo di eseguire  $n$  campionamenti indipendenti  $x_i$  di una certa variabile aleatoria continua  $x$ , che assumiamo avere media  $\mu$  e funzione di distribuzione  $p(x; \mu)$ . Allora, come di consueto, possiamo scrivere la funzione di verosimiglianza come

$$\mathcal{L}(\mu; x_1, \dots, x_n) = \prod_{i=1}^n p(x_i; \mu) \quad \text{da cui} \quad \ln \mathcal{L}(\mu; x_1, \dots, x_n) = \sum_{i=1}^n \ln p(x_i; \mu),$$

e la stima di massima verosimiglianza di  $\mu$  è data dalla soluzione dell'equazione

$$\frac{d \ln \mathcal{L}(\hat{\mu}; x_1, \dots, x_n)}{d\mu} = \sum_{i=1}^n \frac{d \ln p(x_i; \hat{\mu})}{d\mu} = 0. \quad (278)$$

La domanda, apparentemente innocua, di Gauss è la seguente: esiste una forma funzionale di  $p(x; \mu)$  per cui la stima di massima verosimiglianza di  $\mu$  coincide con la media aritmetica dei campionamenti

$$\sum_{i=1}^n \frac{d \ln p(x_i; \hat{\mu})}{d\mu} = \frac{1}{n} \sum_{i=1}^n x_i?$$

È facile convincersi che l'eguaglianza che abbiamo appena scritto non è in generale vera. Se proviamo però a porre

$$\frac{d \ln p(x; \mu)}{d\mu} \propto (x - \mu)$$

essa è verificata in quanto si ha banalmente

$$\sum_{i=1}^n \frac{d \ln p(x_i; \hat{\mu})}{d\mu} \propto \sum_{i=1}^n (x_i - \hat{\mu}) = 0.$$

Tralasciando di dimostrare che questa condizione non è solo sufficiente, ma anche necessaria, ne segue che

$$\ln p(x; \mu) = c_1 (x - \mu)^2 + c_2$$

e infine

$$p(x; \mu) \propto e^{c_1 (x - \mu)^2}, \quad (279)$$

in cui, a parte la forma della normalizzazione, si riconosce l'espressione della distribuzione di Gauss (185) (notiamo esplicitamente che, perché  $p$  sia normalizzabile è necessario che  $c_1 < 0$ ). In questo senso la distribuzione normale si caratterizza come l'unica per cui il principio di verosimiglianza fornisce come miglior stima di una grandezza, sulla base di un numero finito di misure indipendenti, la media aritmetica delle misure stesse.

#### 9.4 INCERTEZZE NELLE STIME DI MASSIMA VEROSIMIGLIANZA

Ricapitoliamo brevemente. Abbiamo un principio generale, quello di massima verosimiglianza, che ci fornisce una prescrizione applicabile in generale per la stima di parametri—e ne abbiamo visto alcuni esempi di applicazione significativi. Fino ad ora, però, abbiamo evitato accuratamente la questione delle incertezze sui parametri, ed è giunto il momento di colmare la lacuna.

##### 9.4.1 Incertezze nel caso unidimensionale

La chiave per la stima delle incertezze sui parametri è una proprietà della funzione di verosimiglianza che enunciamo senza dimostrazione—e cioè il fatto che per grandi campioni (cioè, nel nostro linguaggio, per  $n$  grande) la funzione di verosimiglianza è asintoticamente normale. Nel caso unidimensionale si ha cioè

$$\mathcal{L}(\theta) = \frac{1}{\sigma_{\hat{\theta}} \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\theta - \hat{\theta}}{\sigma_{\hat{\theta}}} \right)^2}, \quad (280)$$

dove  $\sigma_{\hat{\theta}}$  è proprio l'incertezza sulla stima  $\hat{\theta}$  di  $\theta$  cui siamo interessati. Ora, se proviamo a calcolare la derivata seconda

$$\frac{d^2 \ln \mathcal{L}(\hat{\theta})}{d\theta^2} = \frac{d^2}{d\theta^2} \left[ -\ln \sigma_{\hat{\theta}} - \frac{1}{2} \ln(2\pi) - \frac{1}{2} \left( \frac{\theta - \hat{\theta}}{\sigma_{\hat{\theta}}} \right)^2 \right]_{\hat{\theta}} = -\frac{1}{\sigma_{\hat{\theta}}^2},$$

ovverossia

$$\sigma_{\hat{\theta}}^2 = -\frac{1}{\frac{d^2 \ln \mathcal{L}(\hat{\theta})}{d\theta^2}}. \quad (281)$$

Notiamo, per inciso, che se proviamo a sviluppare il logaritmo della funzione di verosimiglianza attorno a  $\hat{\theta}$ , sfruttando il fatto che la derivata prima si annulla in virtù del fatto che  $\hat{\theta}$  è un punto di massimo

$$\ln \mathcal{L}(\theta) \approx \ln \mathcal{L}(\hat{\theta}) + \frac{1}{2} \frac{d^2 \ln \mathcal{L}(\hat{\theta})}{d\theta^2} (\theta - \hat{\theta})^2 = \ln \mathcal{L}(\hat{\theta}) - \frac{1}{2} \frac{(\theta - \hat{\theta})^2}{\sigma_{\hat{\theta}}^2}$$

si ha l'interessante caratterizzazione dell'incertezza sulla stima del parametro  $\theta$  come il semintervallo determinato dai punti per cui la funzione di verosimiglianza si riduce di  $1/2$  rispetto al suo valore massimo

$$\ln \mathcal{L}(\hat{\theta} \pm \sigma_{\hat{\theta}}) \approx \ln \mathcal{L}(\hat{\theta}) - \frac{1}{2} = \ln \mathcal{L}_{\max} - \frac{1}{2}. \quad (282)$$

► **ESEMPIO 9.1.** Torniamo alla stima di massima verosimiglianza della media di una variabile Poissoniana che abbiamo descritto nella sezione ??

$$\ln \mathcal{L}(\mu; k_1, \dots, k_n) = \sum_{i=1}^n (k_i \ln \mu - \ln(k_i!) - \mu).$$

La derivata seconda del logaritmo della funzione di verosimiglianza si legge in questo caso come

$$\frac{d^2 \ln \mathcal{L}(\hat{\mu}; k_1, \dots, k_n)}{d\mu^2} = -\frac{1}{\hat{\mu}^2} \sum_{i=1}^n k_i = -\frac{n}{\hat{\mu}} \quad \text{da cui} \quad \sigma_{\hat{\mu}}^2 = \frac{\mu}{n},$$

in cui riconosciamo la varianza della media che abbiamo incontrato più volte lungo la strada.

#### 9.4.2 Incertezze nel caso generale

Nel caso generale, quando cioè si stima simultaneamente più di un parametro, la (281) si generalizza in un'espressione per l'inverso della matrice di covarianza

$$\Sigma_{ij}^{-1} = \frac{\partial^2 \ln \mathcal{L}(\hat{\theta}_1, \dots, \hat{\theta}_m)}{\partial \theta_i \partial \theta_j}, \quad (283)$$

che contiene sia l'informazione sull'incertezza sui parametri che quella sulla loro correlazione.

Nel limite di grandi campioni la funzione di verosimiglianza è ancora asintoticamente normale (multivariata) e di conseguenza  $\ln \mathcal{L}$  ha una forma iper-parabolica. Il contorno nello spazio dei parametri definito dall'equazione

$$\ln \mathcal{L}(\hat{\theta} \pm \sigma_{\hat{\theta}}) \approx \ln \mathcal{L}(\hat{\theta}) - \frac{1}{2} = \ln \mathcal{L}_{\max} - \frac{1}{2}. \quad (284)$$

ha piani tangenti situati a  $\pm 1$  deviazioni standard  $\sigma_{\hat{\theta}_1}, \dots, \sigma_{\hat{\theta}_m}$  dalle stime dei parametri  $\hat{\theta}_1, \dots, \hat{\theta}_m$ .

► **ESEMPIO 9.2.** Mettere esempio della Gaussiana.

## 9.5 IL $\chi^2$ COME APPLICAZIONE DEL PRINCIPIO DI MASSIMA VEROSIMIGLIANZA

### 9.6 IN BREVE...



---

## COMPLEMENTI DI STATISTICA

---

10.1 LA FUNZIONE GENERATRICE DEI MOMENTI

10.2 LA DISTRIBUZIONE  $t$  DI STUDENT

10.3 IN BREVE...



---

## BIBLIOGRAFIA

---

- [1] P. T. Boggs and J. E. Rogers. Orthogonal distance regression. *Statistical analysis of measurement error models and applications: proceedings of the AMS-IMS-SIAM joint summer research conference*, 112:186, 1990. Web link: [www.mechanicalkern.com/static/odr\\_ams.pdf](http://www.mechanicalkern.com/static/odr_ams.pdf).
- [2] R. D. Clarke. An application of the poisson distribution. *Journal of the Institute of Actuaries (London)*, 72:482, 1946. Web link: <https://www.actuaries.org.uk/documents/application-poisson-distribution>.
- [3] G. D'Agostini. Bayesian inference in processing experimental data: principles and basic applications. *Reports on Progress in Physics*, 66(9):1383, 2003. Web link: <http://arxiv.org/abs/physics/0304102>.
- [4] G. D'Agostini. On the so-called Boy or Girl Paradox. *ArXiv e-prints 1001.0708*, January 2010. Web link: <http://arxiv.org/abs/1001.0708>.
- [5] G. D'Agostini. Probably a discovery: bad mathematics means rough scientific communication. *ArXiv e-prints 1112.3620*, December 2011. Web link: <https://arxiv.org/abs/1112.3620>.
- [6] G. D'Agostini. Bertrand 'paradox' reloaded (with details on transformations of variables, an introduction to Monte Carlo simulation and an inferential variation of the problem). *ArXiv e-prints 1504.01361*, April 2015. Web link: <http://arxiv.org/abs/1504.01361>.
- [7] Richard Davis. The si unit of mass. *Metrologia*, 40(6):299, 2003. Web link: <http://stacks.iop.org/0026-1394/40/i=6/a=001>.
- [8] Comité Consultatif des Unités (CCU). A concise summary of the international system of units, the si. 2006. Web link: [http://www.bipm.org/utils/common/pdf/si\\_summary\\_en.pdf](http://www.bipm.org/utils/common/pdf/si_summary_en.pdf).
- [9] J. F. W. Herschel. Quetelet on probabilities. *Edinburgh Rev.*, 92:1–57, 1850.
- [10] Edwin Thompson Jaynes and G. Larry Bretthorst, editors. *Probability theory : the logic of science*. Cambridge University Press, Cambridge, UK, New York, 2003.
- [11] Donald E. Knuth. *The Art of Computer Programming, Volume 2 (3rd Ed.): Seminumerical Algorithms*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1997.
- [12] A. N. Kolmogorov. *Foundations of the theory of probability*. Chelsea Publishing Company, New York, 1956. Web link: <http://www.mathematik.com/Kolmogorov/index.html>.
- [13] V. von Bortkevič. *Das Gesetz der kleinen Zahlen [The law of small numbers]*. Leipzig, Germany, 1898.
- [14] R. L. Wasserstein and N. A. Lazar. The asa's statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2):129–133, 2016. Web link: <http://dx.doi.org/10.1080/00031305.2016.1154108>.
- [15] B. P. Welford. Note on a method for calculating corrected sums of squares and products. *Technometrics*, 4(3):419–420, 1962.





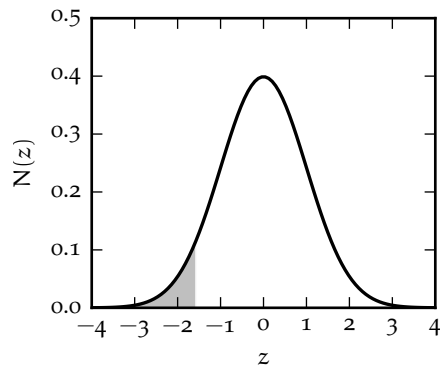
---

## TAVOLE NUMERICHE

---

In questa appendice riportiamo i valori tabulati di alcune funzioni statistiche rilevanti (e.g., l'integrale normale degli errori). Le tavole possono essere utili quando non si ha a disposizione un calcolatore.

## A.1 INTEGRALE NORMALE DEGLI ERRORI—I



Valori tabulati della funzione cumulativa di una variabile Gaussiana in forma standard

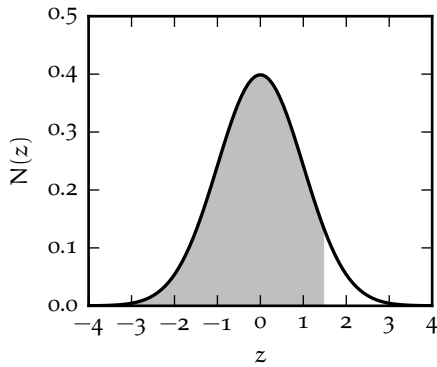
$$P = \Phi(z) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right).$$

Nella tabella la prima colonna identifica le prime due cifre di  $z$ , mentre la riga di intestazione identifica la seconda cifra decimale.

(Continua nella pagina successiva.)

$z$	9	8	7	6	5	4	3	2	1	0
-3.9	0.00003	0.00003	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00005	0.00005
-3.8	0.00005	0.00005	0.00005	0.00006	0.00006	0.00006	0.00006	0.00007	0.00007	0.00007
-3.7	0.00008	0.00008	0.00008	0.00008	0.00009	0.00009	0.00010	0.00010	0.00010	0.00011
-3.6	0.00011	0.00012	0.00012	0.00013	0.00013	0.00014	0.00014	0.00015	0.00015	0.00016
-3.5	0.00017	0.00017	0.00018	0.00019	0.00019	0.00020	0.00021	0.00022	0.00022	0.00023
-3.4	0.00024	0.00025	0.00026	0.00027	0.00028	0.00029	0.00030	0.00031	0.00032	0.00034
-3.3	0.00035	0.00036	0.00038	0.00039	0.00040	0.00042	0.00043	0.00045	0.00047	0.00048
-3.2	0.00050	0.00052	0.00054	0.00056	0.00058	0.00060	0.00062	0.00064	0.00066	0.00069
-3.1	0.00071	0.00074	0.00076	0.00079	0.00082	0.00084	0.00087	0.00090	0.00094	0.00097
-3.0	0.00100	0.00104	0.00107	0.00111	0.00114	0.00118	0.00122	0.00126	0.00131	0.00135
-2.9	0.00139	0.00144	0.00149	0.00154	0.00159	0.00164	0.00169	0.00175	0.00181	0.00187
-2.8	0.00193	0.00199	0.00205	0.00212	0.00219	0.00226	0.00233	0.00240	0.00248	0.00256
-2.7	0.00264	0.00272	0.00280	0.00289	0.00298	0.00307	0.00317	0.00326	0.00336	0.00347
-2.6	0.00357	0.00368	0.00379	0.00391	0.00402	0.00415	0.00427	0.00440	0.00453	0.00466
-2.5	0.00480	0.00494	0.00508	0.00523	0.00539	0.00554	0.00570	0.00587	0.00604	0.00621
-2.4	0.00639	0.00657	0.00676	0.00695	0.00714	0.00734	0.00755	0.00776	0.00798	0.00820
-2.3	0.00842	0.00866	0.00889	0.00914	0.00939	0.00964	0.00990	0.01017	0.01044	0.01072
-2.2	0.01101	0.01130	0.01160	0.01191	0.01222	0.01255	0.01287	0.01321	0.01355	0.01390
-2.1	0.01426	0.01463	0.01500	0.01539	0.01578	0.01618	0.01659	0.01700	0.01743	0.01786
-2.0	0.01831	0.01876	0.01923	0.01970	0.02018	0.02068	0.02118	0.02169	0.02222	0.02275
-1.9	0.02330	0.02385	0.02442	0.02500	0.02559	0.02619	0.02680	0.02743	0.02807	0.02872
-1.8	0.02938	0.03005	0.03074	0.03144	0.03216	0.03288	0.03362	0.03438	0.03515	0.03593
-1.7	0.03673	0.03754	0.03836	0.03920	0.04006	0.04093	0.04182	0.04272	0.04363	0.04457
-1.6	0.04551	0.04648	0.04746	0.04846	0.04947	0.05050	0.05155	0.05262	0.05370	0.05480
-1.5	0.05592	0.05705	0.05821	0.05938	0.06057	0.06178	0.06301	0.06426	0.06552	0.06681
-1.4	0.06811	0.06944	0.07078	0.07215	0.07353	0.07493	0.07636	0.07780	0.07927	0.08076
-1.3	0.08226	0.08379	0.08534	0.08691	0.08851	0.09012	0.09176	0.09342	0.09510	0.09680
-1.2	0.09853	0.10027	0.10204	0.10383	0.10565	0.10749	0.10935	0.11123	0.11314	0.11507
-1.1	0.11702	0.11900	0.12100	0.12302	0.12507	0.12714	0.12924	0.13136	0.13350	0.13567
-1.0	0.13786	0.14007	0.14231	0.14457	0.14686	0.14917	0.15151	0.15386	0.15625	0.15866
-0.9	0.16109	0.16354	0.16602	0.16853	0.17106	0.17361	0.17619	0.17879	0.18141	0.18406
-0.8	0.18673	0.18943	0.19215	0.19489	0.19766	0.20045	0.20327	0.20611	0.20897	0.21186
-0.7	0.21476	0.21770	0.22065	0.22363	0.22663	0.22965	0.23270	0.23576	0.23885	0.24196
-0.6	0.24510	0.24825	0.25143	0.25463	0.25785	0.26109	0.26435	0.26763	0.27093	0.27425
-0.5	0.27760	0.28096	0.28434	0.28774	0.29116	0.29460	0.29806	0.30153	0.30503	0.30854
-0.4	0.31207	0.31561	0.31918	0.32276	0.32636	0.32997	0.33360	0.33724	0.34090	0.34458
-0.3	0.34827	0.35197	0.35569	0.35942	0.36317	0.36693	0.37070	0.37448	0.37828	0.38209
-0.2	0.38591	0.38974	0.39358	0.39743	0.40129	0.40517	0.40905	0.41294	0.41683	0.42074
-0.1	0.42465	0.42858	0.43251	0.43644	0.44038	0.44433	0.44828	0.45224	0.45620	0.46017
-0.0	0.46414	0.46812	0.47210	0.47608	0.48006	0.48405	0.48803	0.49202	0.49601	0.50000

## A.2 INTEGRALE NORMALE DEGLI ERRORI—I



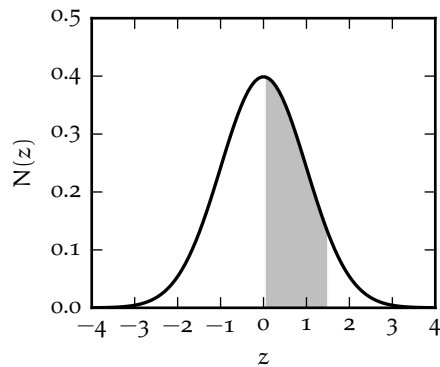
Valori tabulati della funzione cumulativa di una variabile Gaussiana in forma standard

$$P = \Phi(z) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right).$$

Nella tabella la prima colonna identifica le prime due cifre di  $z$ , mentre la riga di intestazione identifica la seconda cifra decimale.

$z$	0	1	2	3	4	5	6	7	8	9
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

## A.3 INTEGRALE NORMALE DEGLI ERRORI—II



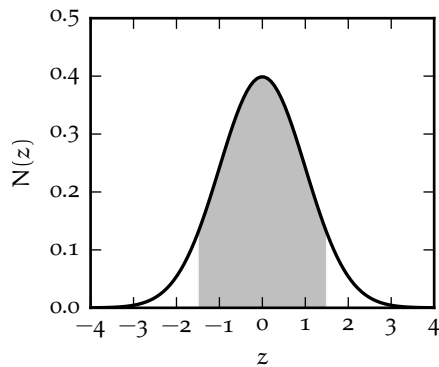
Valori tabulati per l'integrale di una distribuzione Gaussiana in forma standard corrispondente all'ombreggiatura in figura

$$P = \Phi(z) - \frac{1}{2} = \frac{1}{2} \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right).$$

Nella tabella la prima colonna identifica le prime due cifre di  $z$ , mentre la riga di intestazione identifica la seconda cifra decimale.

$z$	0	1	2	3	4	5	6	7	8	9
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41309	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2.0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900
3.1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929
3.2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950
3.3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965
3.4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976
3.5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983
3.6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989
3.7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49992	0.49992	0.49992	0.49992
3.8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995
3.9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997

## A.4 INTEGRALE NORMALE DEGLI ERRORI—III



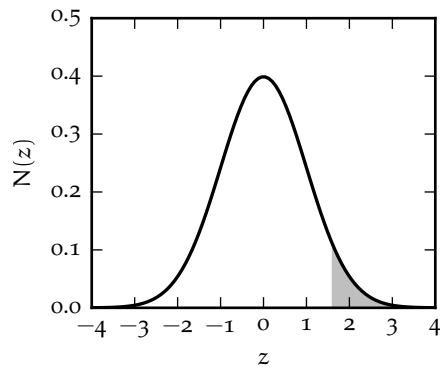
Valori tabulati per l'integrale di una distribuzione Gaussiana in forma standard corrispondente all'ombreggiatura in figura

$$P = \Phi(z) - \Phi(-z) = \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right).$$

Nella tabella la prima colonna identifica le prime due cifre di  $z$ , mentre la riga di intestazione identifica la seconda cifra decimale.

$z$	0	1	2	3	4	5	6	7	8	9
0.0	0.00000	0.00798	0.01596	0.02393	0.03191	0.03988	0.04784	0.05581	0.06376	0.07171
0.1	0.07966	0.08759	0.09552	0.10343	0.11134	0.11924	0.12712	0.13499	0.14285	0.15069
0.2	0.15852	0.16633	0.17413	0.18191	0.18967	0.19741	0.20514	0.21284	0.22052	0.22818
0.3	0.23582	0.24344	0.25103	0.25860	0.26614	0.27366	0.28115	0.28862	0.29605	0.30346
0.4	0.31084	0.31819	0.32551	0.33280	0.34006	0.34729	0.35448	0.36164	0.36877	0.37587
0.5	0.38292	0.38995	0.39694	0.40389	0.41080	0.41768	0.42452	0.43132	0.43809	0.44481
0.6	0.45149	0.45814	0.46474	0.47131	0.47783	0.48431	0.49075	0.49714	0.50350	0.50981
0.7	0.51607	0.52230	0.52848	0.53461	0.54070	0.54675	0.55275	0.55870	0.56461	0.57047
0.8	0.57629	0.58206	0.58778	0.59346	0.59909	0.60467	0.61021	0.61570	0.62114	0.62653
0.9	0.63188	0.63718	0.64243	0.64763	0.65278	0.65789	0.66294	0.66795	0.67291	0.67783
1.0	0.68269	0.68750	0.69227	0.69699	0.70166	0.70628	0.71086	0.71538	0.71986	0.72429
1.1	0.72867	0.73300	0.73729	0.74152	0.74571	0.74986	0.75395	0.75800	0.76200	0.76595
1.2	0.76986	0.77372	0.77754	0.78130	0.78502	0.78870	0.79233	0.79592	0.79945	0.80295
1.3	0.80640	0.80980	0.81316	0.81648	0.81975	0.82298	0.82617	0.82931	0.83241	0.83547
1.4	0.83849	0.84146	0.84439	0.84728	0.85013	0.85294	0.85571	0.85844	0.86113	0.86378
1.5	0.86639	0.86896	0.87149	0.87398	0.87644	0.87886	0.88124	0.88358	0.88589	0.88817
1.6	0.89040	0.89260	0.89477	0.89690	0.89899	0.90106	0.90309	0.90508	0.90704	0.90897
1.7	0.91087	0.91273	0.91457	0.91637	0.91814	0.91988	0.92159	0.92327	0.92492	0.92655
1.8	0.92814	0.92970	0.93124	0.93275	0.93423	0.93569	0.93711	0.93852	0.93989	0.94124
1.9	0.94257	0.94387	0.94514	0.94639	0.94762	0.94882	0.95000	0.95116	0.95230	0.95341
2.0	0.95450	0.95557	0.95662	0.95764	0.95865	0.95964	0.96060	0.96155	0.96247	0.96338
2.1	0.96427	0.96514	0.96599	0.96683	0.96765	0.96844	0.96923	0.96999	0.97074	0.97148
2.2	0.97219	0.97289	0.97358	0.97425	0.97491	0.97555	0.97618	0.97679	0.97739	0.97798
2.3	0.97855	0.97911	0.97966	0.98019	0.98072	0.98123	0.98173	0.98221	0.98269	0.98315
2.4	0.98360	0.98405	0.98448	0.98490	0.98531	0.98571	0.98611	0.98649	0.98686	0.98723
2.5	0.98758	0.98793	0.98826	0.98859	0.98891	0.98923	0.98953	0.98983	0.99012	0.99040
2.6	0.99068	0.99095	0.99121	0.99146	0.99171	0.99195	0.99219	0.99241	0.99264	0.99285
2.7	0.99307	0.99327	0.99347	0.99367	0.99386	0.99404	0.99422	0.99439	0.99456	0.99473
2.8	0.99489	0.99505	0.99520	0.99535	0.99549	0.99563	0.99576	0.99590	0.99602	0.99615
2.9	0.99627	0.99639	0.99650	0.99661	0.99672	0.99682	0.99692	0.99702	0.99712	0.99721
3.0	0.99730	0.99739	0.99747	0.99755	0.99763	0.99771	0.99779	0.99786	0.99793	0.99800
3.1	0.99806	0.99813	0.99819	0.99825	0.99831	0.99837	0.99842	0.99848	0.99853	0.99858
3.2	0.99863	0.99867	0.99872	0.99876	0.99880	0.99885	0.99889	0.99892	0.99896	0.99900
3.3	0.99903	0.99907	0.99910	0.99913	0.99916	0.99919	0.99922	0.99925	0.99928	0.99930
3.4	0.99933	0.99935	0.99937	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950	0.99952
3.5	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99963	0.99964	0.99966	0.99967
3.6	0.99968	0.99969	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976	0.99977	0.99978
3.7	0.99978	0.99979	0.99980	0.99981	0.99982	0.99982	0.99983	0.99984	0.99984	0.99985
3.8	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989	0.99989	0.99990	0.99990
3.9	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99993	0.99993	0.99993	0.99993

## A.5 INTEGRALE NORMALE DEGLI ERRORI—III

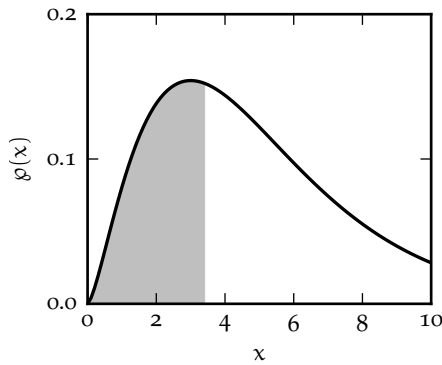


Valori tabulati per l'integrale di una distribuzione Gaussiana in forma standard corrispondente all'ombreggiatura in figura

$$P = 1 - \Phi(z) = \frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right).$$

Nella tabella la prima colonna identifica le prime due cifre di  $z$ , mentre la riga di intestazione identifica la seconda cifra decimale.

$z$	0	1	2	3	4	5	6	7	8	9
4.0	3.17e-05	3.04e-05	2.91e-05	2.79e-05	2.67e-05	2.56e-05	2.45e-05	2.35e-05	2.25e-05	2.16e-05
4.1	2.07e-05	1.98e-05	1.89e-05	1.81e-05	1.74e-05	1.66e-05	1.59e-05	1.52e-05	1.46e-05	1.39e-05
4.2	1.33e-05	1.28e-05	1.22e-05	1.17e-05	1.12e-05	1.07e-05	1.02e-05	9.77e-06	9.34e-06	8.93e-06
4.3	8.54e-06	8.16e-06	7.80e-06	7.46e-06	7.12e-06	6.81e-06	6.50e-06	6.21e-06	5.93e-06	5.67e-06
4.4	5.41e-06	5.17e-06	4.94e-06	4.71e-06	4.50e-06	4.29e-06	4.10e-06	3.91e-06	3.73e-06	3.56e-06
4.5	3.40e-06	3.24e-06	3.09e-06	2.95e-06	2.81e-06	2.68e-06	2.56e-06	2.44e-06	2.32e-06	2.22e-06
4.6	2.11e-06	2.01e-06	1.92e-06	1.83e-06	1.74e-06	1.66e-06	1.58e-06	1.51e-06	1.43e-06	1.37e-06
4.7	1.30e-06	1.24e-06	1.18e-06	1.12e-06	1.07e-06	1.02e-06	9.68e-07	9.21e-07	8.76e-07	8.34e-07
4.8	7.93e-07	7.55e-07	7.18e-07	6.83e-07	6.49e-07	6.17e-07	5.87e-07	5.58e-07	5.30e-07	5.04e-07
4.9	4.79e-07	4.55e-07	4.33e-07	4.11e-07	3.91e-07	3.71e-07	3.52e-07	3.35e-07	3.18e-07	3.02e-07
5.0	2.87e-07	2.72e-07	2.58e-07	2.45e-07	2.33e-07	2.21e-07	2.10e-07	1.99e-07	1.89e-07	1.79e-07
5.1	1.70e-07	1.61e-07	1.53e-07	1.45e-07	1.37e-07	1.30e-07	1.23e-07	1.17e-07	1.11e-07	1.05e-07
5.2	9.96e-08	9.44e-08	8.95e-08	8.48e-08	8.03e-08	7.60e-08	7.20e-08	6.82e-08	6.46e-08	6.12e-08
5.3	5.79e-08	5.48e-08	5.19e-08	4.91e-08	4.65e-08	4.40e-08	4.16e-08	3.94e-08	3.72e-08	3.52e-08
5.4	3.33e-08	3.15e-08	2.98e-08	2.82e-08	2.66e-08	2.52e-08	2.38e-08	2.25e-08	2.13e-08	2.01e-08
5.5	1.90e-08	1.79e-08	1.69e-08	1.60e-08	1.51e-08	1.43e-08	1.35e-08	1.27e-08	1.20e-08	1.14e-08
5.6	1.07e-08	1.01e-08	9.55e-09	9.01e-09	8.50e-09	8.02e-09	7.57e-09	7.14e-09	6.73e-09	6.35e-09
5.7	5.99e-09	5.65e-09	5.33e-09	5.02e-09	4.73e-09	4.46e-09	4.21e-09	3.96e-09	3.74e-09	3.52e-09
5.8	3.32e-09	3.12e-09	2.94e-09	2.77e-09	2.61e-09	2.46e-09	2.31e-09	2.18e-09	2.05e-09	1.93e-09
5.9	1.82e-09	1.71e-09	1.61e-09	1.51e-09	1.43e-09	1.34e-09	1.26e-09	1.19e-09	1.12e-09	1.05e-09
6.0	9.87e-10	9.28e-10	8.72e-10	8.20e-10	7.71e-10	7.24e-10	6.81e-10	6.40e-10	6.01e-10	5.65e-10
6.1	5.30e-10	4.98e-10	4.68e-10	4.39e-10	4.13e-10	3.87e-10	3.64e-10	3.41e-10	3.21e-10	3.01e-10
6.2	2.82e-10	2.65e-10	2.49e-10	2.33e-10	2.19e-10	2.05e-10	1.92e-10	1.81e-10	1.69e-10	1.59e-10
6.3	1.49e-10	1.40e-10	1.31e-10	1.23e-10	1.15e-10	1.08e-10	1.01e-10	9.45e-11	8.85e-11	8.29e-11
6.4	7.77e-11	7.28e-11	6.81e-11	6.38e-11	5.97e-11	5.59e-11	5.24e-11	4.90e-11	4.59e-11	4.29e-11
6.5	4.02e-11	3.76e-11	3.52e-11	3.29e-11	3.08e-11	2.88e-11	2.69e-11	2.52e-11	2.35e-11	2.20e-11
6.6	2.06e-11	1.92e-11	1.80e-11	1.68e-11	1.57e-11	1.47e-11	1.37e-11	1.28e-11	1.19e-11	1.12e-11
6.7	1.04e-11	9.73e-12	9.09e-12	8.48e-12	7.92e-12	7.39e-12	6.90e-12	6.44e-12	6.01e-12	5.61e-12
6.8	5.23e-12	4.88e-12	4.55e-12	4.25e-12	3.96e-12	3.69e-12	3.44e-12	3.21e-12	2.99e-12	2.79e-12
6.9	2.60e-12	2.42e-12	2.26e-12	2.10e-12	1.96e-12	1.83e-12	1.70e-12	1.58e-12	1.48e-12	1.37e-12
7.0	1.28e-12	1.19e-12	1.11e-12	1.03e-12	9.61e-13	8.95e-13	8.33e-13	7.75e-13	7.21e-13	6.71e-13
7.1	6.24e-13	5.80e-13	5.40e-13	5.02e-13	4.67e-13	4.34e-13	4.03e-13	3.75e-13	3.49e-13	3.24e-13
7.2	3.01e-13	2.80e-13	2.60e-13	2.41e-13	2.24e-13	2.08e-13	1.94e-13	1.80e-13	1.67e-13	1.55e-13
7.3	1.44e-13	1.34e-13	1.24e-13	1.15e-13	1.07e-13	9.91e-14	9.20e-14	8.53e-14	7.92e-14	7.34e-14
7.4	6.81e-14	6.32e-14	5.86e-14	5.43e-14	5.03e-14	4.67e-14	4.32e-14	4.01e-14	3.71e-14	3.44e-14
7.5	3.19e-14	2.96e-14	2.74e-14	2.54e-14	2.35e-14	2.18e-14	2.02e-14	1.87e-14	1.73e-14	1.60e-14
7.6	1.48e-14	1.37e-14	1.27e-14	1.17e-14	1.09e-14	1.00e-14	9.27e-15	8.60e-15	7.94e-15	7.38e-15
7.7	6.83e-15	6.27e-15	5.83e-15	5.38e-15	5.00e-15	4.61e-15	4.22e-15	3.94e-15	3.61e-15	3.33e-15
7.8	3.11e-15	2.89e-15	2.66e-15	2.44e-15	2.28e-15	2.05e-15	1.94e-15	1.78e-15	1.61e-15	1.50e-15
7.9	1.39e-15	1.28e-15	1.17e-15	1.11e-15	9.99e-16	9.44e-16	8.33e-16	7.77e-16	7.22e-16	6.66e-16

A.6 FUNZIONE CUMULATIVA DEL  $\chi^2$ 

Valori tabulati del valore  $\chi_p^2$  per cui la funzione cumulativa (cioè l'area ombreggiata in figura) per una variabile  $\chi^2$  a  $n$  gradi di libertà vale esattamente  $p$ .

Per un numero di gradi di libertà  $n > 40$  si può utilizzare un semplice programma al calcolatore oppure si può sfruttare l'approssimazione Gaussiana.

$z/p$	0.005	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99	0.995	0.999	0.9999
1	4e-5	2e-4	4e-3	2e-2	0.10	0.45	1.32	2.71	3.84	6.63	7.88	10.8	15.1
2	1e-2	2e-2	0.10	0.21	0.58	1.39	2.77	4.61	5.99	9.21	10.6	13.8	18.4
3	7e-2	0.11	0.35	0.58	1.21	2.37	4.11	6.25	7.81	11.3	12.8	16.3	21.1
4	0.21	0.30	0.71	1.06	1.92	3.36	5.39	7.78	9.49	13.3	14.9	18.5	23.5
5	0.41	0.55	1.15	1.61	2.67	4.35	6.63	9.24	11.1	15.1	16.7	20.5	25.7
6	0.68	0.87	1.64	2.20	3.45	5.35	7.84	10.6	12.6	16.8	18.5	22.5	27.9
7	0.99	1.24	2.17	2.83	4.25	6.35	9.04	12.0	14.1	18.5	20.3	24.3	29.9
8	1.34	1.65	2.73	3.49	5.07	7.34	10.2	13.4	15.5	20.1	22.0	26.1	31.8
9	1.73	2.09	3.33	4.17	5.90	8.34	11.4	14.7	16.9	21.7	23.6	27.9	33.7
10	2.16	2.56	3.94	4.87	6.74	9.34	12.5	16.0	18.3	23.2	25.2	29.6	35.6
11	2.60	3.05	4.57	5.58	7.58	10.3	13.7	17.3	19.7	24.7	26.8	31.3	37.4
12	3.07	3.57	5.23	6.30	8.44	11.3	14.8	18.5	21.0	26.2	28.3	32.9	39.1
13	3.57	4.11	5.89	7.04	9.30	12.3	16.0	19.8	22.4	27.7	29.8	34.5	40.9
14	4.07	4.66	6.57	7.79	10.2	13.3	17.1	21.1	23.7	29.1	31.3	36.1	42.6
15	4.60	5.23	7.26	8.55	11.0	14.3	18.2	22.3	25.0	30.6	32.8	37.7	44.3
16	5.14	5.81	7.96	9.31	11.9	15.3	19.4	23.5	26.3	32.0	34.3	39.3	45.9
17	5.70	6.41	8.67	10.1	12.8	16.3	20.5	24.8	27.6	33.4	35.7	40.8	47.6
18	6.26	7.01	9.39	10.9	13.7	17.3	21.6	26.0	28.9	34.8	37.2	42.3	49.2
19	6.84	7.63	10.1	11.7	14.6	18.3	22.7	27.2	30.1	36.2	38.6	43.8	50.8
20	7.43	8.26	10.9	12.4	15.5	19.3	23.8	28.4	31.4	37.6	40.0	45.3	52.4
21	8.03	8.90	11.6	13.2	16.3	20.3	24.9	29.6	32.7	38.9	41.4	46.8	54.0
22	8.64	9.54	12.3	14.0	17.2	21.3	26.0	30.8	33.9	40.3	42.8	48.3	55.5
23	9.26	10.2	13.1	14.8	18.1	22.3	27.1	32.0	35.2	41.6	44.2	49.7	57.1
24	9.89	10.9	13.8	15.7	19.0	23.3	28.2	33.2	36.4	43.0	45.6	51.2	58.6
25	10.5	11.5	14.6	16.5	19.9	24.3	29.3	34.4	37.7	44.3	46.9	52.6	60.1
26	11.2	12.2	15.4	17.3	20.8	25.3	30.4	35.6	38.9	45.6	48.3	54.1	61.7
27	11.8	12.9	16.2	18.1	21.7	26.3	31.5	36.7	40.1	47.0	49.6	55.5	63.2
28	12.5	13.6	16.9	18.9	22.7	27.3	32.6	37.9	41.3	48.3	51.0	56.9	64.7
29	13.1	14.3	17.7	19.8	23.6	28.3	33.7	39.1	42.6	49.6	52.3	58.3	66.2
30	13.8	15.0	18.5	20.6	24.5	29.3	34.8	40.3	43.8	50.9	53.7	59.7	67.6
31	14.5	15.7	19.3	21.4	25.4	30.3	35.9	41.4	45.0	52.2	55.0	61.1	69.1
32	15.1	16.4	20.1	22.3	26.3	31.3	37.0	42.6	46.2	53.5	56.3	62.5	70.6
33	15.8	17.1	20.9	23.1	27.2	32.3	38.1	43.7	47.4	54.8	57.6	63.9	72.0
34	16.5	17.8	21.7	24.0	28.1	33.3	39.1	44.9	48.6	56.1	59.0	65.2	73.5
35	17.2	18.5	22.5	24.8	29.1	34.3	40.2	46.1	49.8	57.3	60.3	66.6	74.9
36	17.9	19.2	23.3	25.6	30.0	35.3	41.3	47.2	51.0	58.6	61.6	68.0	76.4
37	18.6	20.0	24.1	26.5	30.9	36.3	42.4	48.4	52.2	59.9	62.9	69.3	77.8
38	19.3	20.7	24.9	27.3	31.8	37.3	43.5	49.5	53.4	61.2	64.2	70.7	79.2
39	20.0	21.4	25.7	28.2	32.7	38.3	44.5	50.7	54.6	62.4	65.5	72.1	80.6
40	20.7	22.2	26.5	29.1	33.7	39.3	45.6	51.8	55.8	63.7	66.8	73.4	82.1





---

## LA RELAZIONE DI LABORATORIO

---

Lo scopo della relazione di laboratorio è quella di descrivere, in modo più chiaro possibile, le procedure utilizzate e i risultati ottenuti durante l'esperienza. Comunicare il risultato del proprio lavoro è fondamentale nella ricerca scientifica, e le relazioni di laboratorio sono un primo passo verso l'acquisizione delle adeguate capacità di comunicazione. Chi non ha mai scritto una relazione (ad esempio alle scuole superiori), può incontrare alcune difficoltà iniziali. In questo capitolo sono quindi riportati alcuni suggerimenti e linee guida per la preparazione delle relazioni.

### B.1 CONSIDERAZIONI GENERALI

Nella stesura della relazione, sono fondamentali la *completezza*, la *concisione* e la *chiarezza*. Chi legge il testo, ad esempio un docente o un altro studente, deve capire cosa è stato fatto e come sono stati ottenuti i risultati. Uno dei punti chiave del metodo scientifico è la riproducibilità dei risultati. È quindi fondamentale che, leggendo la relazione, un altro sperimentatore abbia le informazioni necessarie per poter ripetere esattamente la misura, quindi è importante riportare le informazioni in modo completo, descrivendo ad esempio l'apparato sperimentale oppure le procedure seguite durante l'esecuzione dell'esperienza.

Bisogna tuttavia tenere presente la brevità e la concisione della relazione. La relazione infatti *non è un trattato di fisica*. E' giusto introdurre brevemente i concetti di fisica e le equazioni utilizzate nella relazione, ma senza dilungarsi troppo su tutti i dettagli del fenomeno che si vuole studiare. Inoltre, se la relazione è stata scritta in modo chiaro, sarà più facile rileggerla a distanza di tempo, ad esempio durante la preparazione all'esame di fine corso.

Tenendo presente questi aspetti, bisogna sottolineare tuttavia che la relazione *non è un diario delle attività svolte in laboratorio!*. Bisogna senza dubbio spiegare come sono state effettuate le misure e le difficoltà incontrate, ma evitando i dettagli non rilevanti. Ad esempio, è opportuno ridurre al minimo i commenti soggettivi e mantenere lo stile asciutto. Raccontare le scelte sperimentali adottate nel condurre una certa misura può essere utile (ad esempio gli accorgimenti utilizzati per misurare al meglio i periodi di oscillazione di un pendolo), ma solo se sono funzionali alla comprensione dell'esperienza e dei suoi risultati. E' molto più importante concentrarsi *sulle misure effettuate, sull'analisi dei dati e sui risultati ottenuti*.

### B.2 CARTA, PENNA E CALAMAIO, OPPURE IL COMPUTER?

Al momento di scrivere le prime relazioni, ci si trova in dubbio se prepararle al computer o scriverle a penna su un foglio. Esistono molte soluzioni per preparare la relazione in formato elettronico, a partire dai vari elaboratori di testo oggi disponibili per Linux o per Windows, oppure preparare un documento LaTeX. *Il punto chiave non è la forma in cui viene presentata la relazione, ma il suo contenuto*. Se non ci si sente sicuri nell'utilizzo del computer, si può scrivere la relazione a mano senza problemi, magari inserendo dei grafici realizzati al computer se necessario. È importante tenere presente che *scrivere la relazione al computer non corrisponde a una valutazione più alta*. L'unica accortezza che si richiede quando la relazione è redatta a mano, è che sia scritta con una grafia comprensibile!

## B.3 LA STRUTTURA DELLA RELAZIONE

Nella letteratura scientifica, gli articoli, o *paper*, sono basati su uno schema ben definito, che tende a seguire la cosiddetta struttura IMRaD (*Introduction, Methods, Results, and Discussion*). Questo schema organizzativo è il frutto di un'evoluzione di stile nel corso del tempo, mirata sempre di più a ottimizzare la comunicazione dei risultati scientifici. La produzione scientifica ha una crescita continua, e ogni giorno centinaia di nuovi articoli sono pubblicati in ogni disciplina. Avere uno schema predefinito aiuta gli scienziati a recuperare rapidamente le informazioni utili, senza dover passare troppo tempo a scovare i metodi o i risultati di un lavoro. Ad esempio sappiamo che nell'*Abstract* sono menzionati i punti essenziali di un lavoro, oppure che i dettagli circa l'apparato sperimentale sono discussi in una sezione chiamata *Methods* o simili.

In maniera simile, anche la relazione di laboratorio, che è una forma semplificata di *report* scientifico, può seguire uno schema definito, che può essere molto utile soprattutto nel corso delle prime esperienze di laboratorio. Ecco un possibile schema da seguire per la stesura della relazione:

1. Informazioni generali;
2. Scopo dell'esperienza;
3. Cenni teorici;
4. Apparato sperimentale;
5. Descrizione delle misure;
6. Analisi dei dati;
7. Conclusioni;

## B.3.1 Informazioni generali

All'inizio della relazione è ovviamente importante riportare il titolo dell'esperienza (ad esempio "Misura dell'accelerazione di gravità utilizzando un pendolo semplice"), la data e il nome dello studente o degli studenti del gruppo che ha condotto l'esperienza.

## B.3.2 Scopo dell'esperienza

Spiegare, in un paio di righe, lo scopo dell'esperienza e la metodologia che si intende utilizzare. Ad esempio misurare la densità di alcuni oggetti misurandole massa e volume.

## B.3.3 Cenni teorici

Descrivere, *brevemente*, i concetti fisici e il fenomeno che si sta studiando e mostrare la legge e le leggi che si intendono verificare. I cenni di teoria vanno citati in poche righe, e non si devono ricavare o dimostrare le equazioni utilizzate. È fondamentale introdurre le formule che si utilizzeranno nell'esperienza, scrivendole se necessario anche nella forma usata durante l'esperienza (ad esempio, passare da  $T$  a  $T^2$  per evidenziare andamenti costanti o lineari). Un paio di paragrafi sono necessari per questa sezione.

In questa sezione si deve, ove possibile, riportare anche un disegno che schematizzi le grandezze in gioco, ad esempio lo schema di un pendolo semplice, avendo cura di annotare con opportune lettere le grandezze descritte nelle equazioni (ad esempio lunghezze e angoli). Non è importante che il disegno sia raffinato, tridimensionale, o a colori, ma che descriva fedelmente il fenomeno. Quindi non serve perdere tempo a cercare su Internet un disegno molto bello: basta anche uno schema disegnato a mano, purché sia chiaro.

### B.3.4 Apparato sperimentale e strumenti

Si tratta di una sezione molto importante, in cui si richiede di descrivere prima di tutto gli strumenti utilizzati (calibri, bilancia, ecc), riportando per ciascuno la risoluzione. Bisogna poi riportare il materiale a disposizione (ad esempio molle, pendoli ecc).

### B.3.5 Descrizione delle misure

Descrivere le misure effettuate, le unità di misura e le incertezze associate a ciascuna grandezza. E' buona norma riportare i dati grezzi, quando non sono in numero eccessivo, in opportune tabelle. Nelle tabelle devono essere riportati in modo chiaro le misure con le opportune unità di misura e incertezze. Attenzione: bisogna dire cosa è stato fatto davvero (difficoltà e imprevisti compresi), non quello che sarebbe dovuto succedere in un mondo ideale!

### B.3.6 Analisi dei dati

Descrivere in questa sezione metodi di analisi utilizzati (ad esempio i *fit*), includendo le tabelle con i dati raccolti e gli eventuali grafici. E' molto importante descrivere in modo preciso i metodi di analisi utilizzati: ad esempio "metodo di fit basato sui minimi quadrati condotto utilizzando la funzione *curve\_fit* di Python", piuttosto che un generico "fit lineare" oppure "fit con Python". Quando si effettuano fit o test sulla bintà di un fit è sempre opportuno riportare il valore del  $\chi^2$ , i gradi di libertà e la probabilità associata ad esso. Bisogna sempre riportare come è stata stimata e propagata l'incertezza sulle varie quantità, e ovviamente anche per i risultati non bisogna mai dimenticarsi le unità di misura e le incertezze. In questa sezione naturalmente vanno inseriti i grafici costruiti a partire dai dati raccolti.

### B.3.7 Conclusioni

Si tratta di una delle sezioni più importanti e che al tempo stesso richiedono maggiore esperienza. In questa sezione si riportano infatti in modo chiaro i risultati finali dell'esperienza. Bisogna quindi discutere i risultati ottenuti dall'analisi dei dati, ad esempio la bontà di un risultato in termini di errore relativo e accordo con il modello teorico, oppure l'accordo o il disaccordo con una legge teorica utilizzando la probabilità associata al test del  $\chi^2$ .

Bisogna supportare le conclusioni con stime quantitative, ad esempio valutando la differenza dati-teoria alla luce delle incertezze di misura. È anche possibile discutere possibili idee per migliorare le misure o la stima degli errori.

## B.4 COSA NON DIMENTICARE, E COSA NON FARE

Una volta redatta la relazione seguendo lo schema appena descritto, prima di consegnarla conviene sempre controllare di non aver fatto alcuni errori o aver dimenticato delle informazioni importanti.

### B.4.1 Cose da non dimenticare

- ▶ Propagare sempre gli errori (possibilmente in modo corretto);
- ▶ Riportare sempre un disegno dell'apparato sperimentale;
- ▶ Descrivere sempre gli strumenti utilizzati;
- ▶ Includere sempre i grafici rilevanti e, se possibile, tutti i dati raccolti;
- ▶ Riportare sempre nelle tabelle, o sugli assi dei grafici, il nome della quantità misurata e la sua unità di misura;
- ▶ Riportare sempre con completezza i risultati di un fit, o di un test del  $\chi^2$ ;

- Evidenziare sempre opportunamente i risultati finali dell'esperienza.

#### B.4.2 Cose da non fare

- Non riportare mai la misura di una quantità senza riportare le unità di misura o l'incertezza associata;
- Non riportare mai in modo errato le cifre significative;
- Non riportare mai i valori delle misure direttamente sugli assi;
- Non discutere mai in maniera troppo qualitativa i risultati—non dire, ad esempio, che "l'accordo con la teoria è abbastanza buono...";
- Non usare mai vocaboli ed espressioni colorite di dubbia correttezza, o gergali, come "graficare", "plottare", "misure sospette", "fare senso", e simili.

#### B.5 IN CONCLUSIONE...

Scrivere una relazione di laboratorio fa parte del bagaglio culturale da acquisire quando si lavora nel mondo della ricerca e non solo. È importante ricordarsi i punti descritti in questo capitolo, e man mano applicarli man mano che diventeranno sempre più automatici e naturali. Ogni esperienza di laboratorio è un'occasione nuova per esercitarsi anche da questo punto di vista, in modo da imparare a descrivere il proprio lavoro in modo sempre più chiaro e rigoroso.

## ALCUNI STRUMENTI IN USO IN LABORATORIO

### C.1 IL CALIBRO VENTESIMALE

Il calibro (ed in particolare il calibro a corsoio—il tipo più diffuso e quello che utilizzeremo in laboratorio) è uno strumento di misura della lunghezza costituito da un regolo graduato su cui è libero di scorrere un *corsoio* e da un insieme di appendici che servono da battuta per la misura. Le parti essenziali del calibro, mostrate in figura C.1 sono:

1. i becchi esterni (dotati di coltelli), per misurare il diametro o le dimensioni esterne di un oggetto;
2. i becchi interni (anch'essi dotati di coltelli), per misurare il diametro interno di un oggetto;
3. l'asta di profondità, per misurare la profondità di una cavità o di un oggetto;
4. la scala principale graduata in mm;
5. la scala principale graduata in sedicesimi di pollice (che, lo ricordiamo, non fanno parte del SI);
6. il nonio graduato in ventesimi (o cinquantesimi) di mm;
7. il nonio graduato in frazioni di sedicesimi di pollice;
8. il sistema di bloccaggio del nonio.

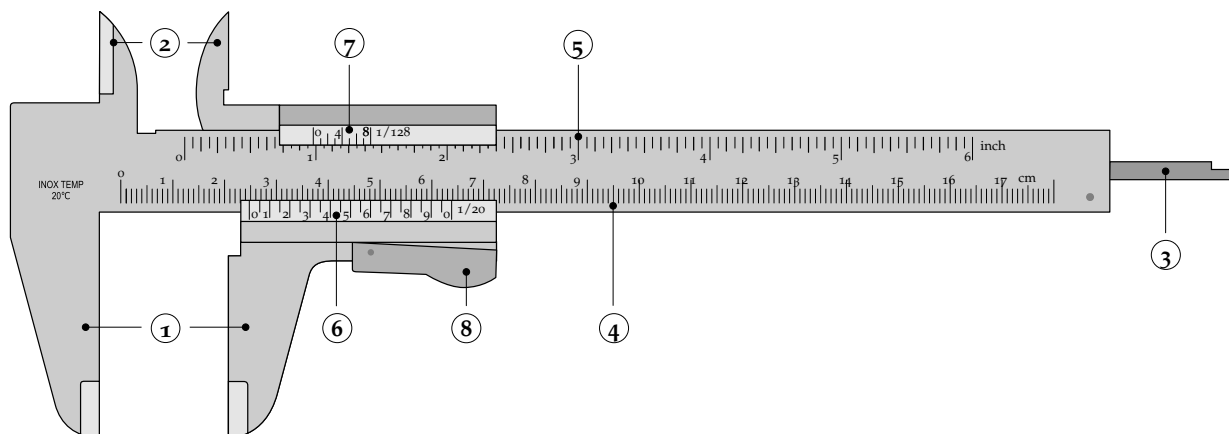


FIGURA C.1. Rappresentazione schematica di un calibro a corsoio ventesimale. Le parti dello strumento identificate dai numeri 1–8 sono descritte in dettaglio nel testo. (Figura adattata da [https://en.wikipedia.org/wiki/Calipers#/media/File:Vernier\\_caliper.svg](https://en.wikipedia.org/wiki/Calipers#/media/File:Vernier_caliper.svg).)

#### C.1.1 Il nonio

Il funzionamento del calibro a corsoio si basa essenzialmente su una delle sue componenti: il nonio—una scala graduata che, nei calibri che utilizzeremo in laboratorio, è lunga 39 mm e suddivisa in 20 parti

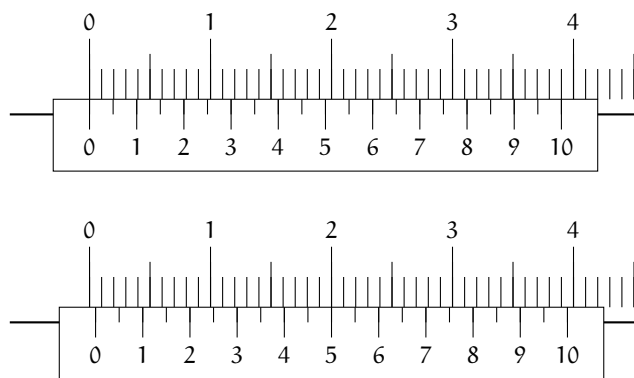


FIGURA C.2. Illustrazione del funzionamento del nonio. Quando lo zero del nonio coincide con lo zero della scala principale (figura in alto), tutte le altre tacche del nonio sono disallineate rispetto a quelle della scala principale. Se spostiamo il nonio di mezzo mm (figura in basso), l'unica tacca del nonio stessa a coincidere con una della scala principale è quella marcata con 5 (decimi di mm).

(quindi con 21 tacche equispaziate) da  $\frac{39}{20}$  mm ciascuna, e marcate per indicare i decimi di mm. Si tratta di un dispositivo tanto semplice quanto ingegnoso.

Quando il calibro è *chiuso* (cioè i becchi esterni sono a contatto tra loro), lo zero del nonio corrisponde con lo zero della scala principale; di conseguenza la ventunesima tacca del nonio (che è di nuovo marcata con uno zero) coincide con la tacca dei 39 mm della scala principale, mentre tutte le altre divisioni del nonio saranno per definizione disallineate con le divisioni della scala principale. Per inciso, questo è sempre vero: vi è al più (e sempre, entro la risoluzione limitata del nostro occhio) una tacca del nonio (o due, la prima e la ventunesima, nel caso dello zero) che coincide con una tacca della scala principale. Fino a qui tutto facile, ma adesso viene la parte interessante. Se spostiamo il nonio di  $\frac{1}{20}$  mm verso destra, la tacca dello zero sarà leggermente disallineata con lo zero della scala principale. La prima tacca del nonio a destra dello zero disterà  $\frac{1}{20} + \frac{39}{20} = 2$  mm dallo zero della scala principale e coinciderà quindi con la tacca dei 2 mm della scala stessa (tutte le altre tacche del nonio, come sappiamo, saranno disallineate con la scala principale). Se ci spostiamo ancora di  $\frac{1}{20}$  mm, la seconda tacca del nonio a destra dello zero disterà dallo zero stesso  $\frac{2}{20} + 2 \times \frac{39}{20} = 4$  mm e così via. In assenza di errori di lettura la risoluzione del calibro appena descritto è dunque  $\frac{1}{20}$  mm, da cui il nome di *calibro ventesimale*.

### C.1.2 Lettura del calibro

La tecnica di lettura del calibro si basa sul principio di funzionamento del nonio illustrato brevemente nella sezione precedente e, come illustrato in figura C.3, si può riassumere nelle due semplici regole che seguono:

1. la tacca della scala principale immediatamente inferiore allo zero del nonio indica la parte intera della lettura in mm;
2. l'unica tacca del nonio che coincide con una (qualsiasi) delle tacche della scala principale indica la parte decimale della lettura e si legge direttamente dal nonio stesso in decimi di mm.

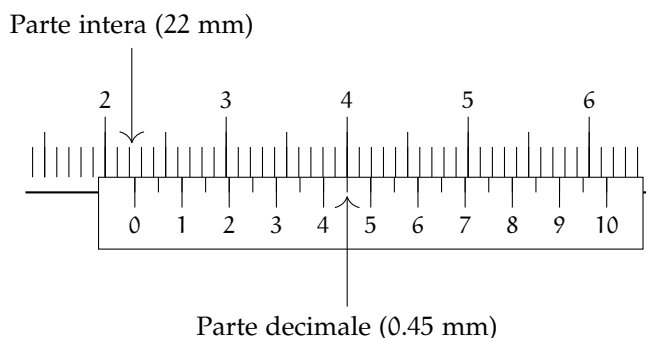


FIGURA C.3. Esempio di lettura del calibro ventesimale. La parte intera si legge sulla scala principale e corrisponde alla prima tacca a sinistra dello zero del nonio—in questo caso 22 mm. La parte decimale si legge sul nonio e corrisponde all'unica tacca che coincide con una delle tacche della scala principale—in questo caso  $\frac{4.5}{10}$  mm. La lettura completa è dunque  $22.45 \pm 0.05$  mm.

## C.2 IL MICROMETRO PALMER

Il micrometro Palmer è uno strumento di misura della lunghezza con risoluzione pari ad  $1/100$  mm. Il principio di funzionamento si basa sull'avanzamento di una vite micrometrica che spinge un cilindro mobile contro uno fisso—con l'oggetto da misurare posto tra i due come mostrato in figura C.4. L'idea di base è che se il passo della vite è abbastanza piccolo ed il suo diametro abbastanza grande, allora distanze relativamente piccole possono essere *amplificate* in rotazioni che possono essere facilmente lette su una scala.

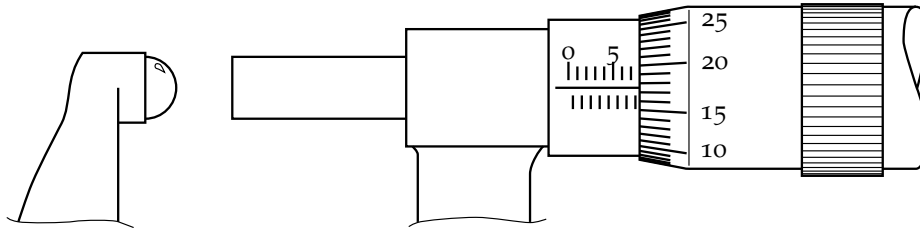


FIGURA C.4. Esempio di lettura di un micrometro Palmer. In questo caso particolare l'ultima tacca visibile sul cilindro fisso è quella che corrisponde a 7.5 mm—cui vanno aggiunti 0.17 mm corrispondenti alla rotazione della parte mobile. La lettura sarà dunque  $7.67 \pm 0.01$  mm. (Figura adattata da [https://commons.wikimedia.org/wiki/File:Lecture\\_micrometre.svg](https://commons.wikimedia.org/wiki/File:Lecture_micrometre.svg).)

## C.2.1 Lettura del calibro Palmer e consigli pratici

La lettura del calibro Palmer è semplice. Un giro completo del cilindro mobile corrisponde ad una escursione lineare di 0.5 mm e, corrispondentemente, il cilindro fisso è graduato con una scala di uguale passo. La lettura si fa dunque, al livello del mezzo millimetro, contando le tacche visibili sul cilindro fisso. A questa va aggiunta la frazione di mezzo millimetro corrispondente alla rotazione addizionale delle ghiera mobile, che è a sua volta suddivisa in 50 parti, come illustrato in figura C.4.

Data la risoluzione spinta (per un dispositivo puramente meccanico), è necessario fare attenzione al momento torcente applicato nel processo di misura—che, se troppo grande, potrebbe deformare l'oggetto da misurare, oltre a danneggiare il calibro. A questo scopo i calibri Palmer sono tipicamente dotati di un meccanismo di frizione, ed è *sempre dalla frizione (anziché dalla ghiera graduata) che bisogna avvitare il tamburo*.

È buona norma controllare sempre che, quando il tamburo è completamente avvitato, lo zero della ghiera corrisponda con l'asse graduato sul cilindro fisso. Se così non è, allora si ha un problema di zero che si può facilmente correggere per sottrazione alla fine della misura (o ricalibrando opportunamente il dispositivo).

È utile notare, infine, come in pratica sia semplice (per errore di lettura o imperfetta calibrazione dello strumento) sbagliare il conteggio delle tacche sul cilindro fisso—che corrisponde a sbagliare la lettura di 0.5 mm, o cinquanta volte la risoluzione del calibro. Per questo motivo è buona norma ricontrollare le misure fatte con il calibro Palmer con un calibro ventesimale che, pur avendo una risoluzione inferiore, può mettere in luce facilmente eventuali errori grossolani.





---

## SISTEMA BINARIO E RAPPRESENTAZIONE DEI NUMERI

---

Anche se in linea di principio questo non è l'unico approccio possibile, praticamente tutti i calcolatori moderni funzionano in logica binaria—cioè salvano e manipolano i dati al loro interno sotto forma di *bit*, o celle elementari di memoria che possono assumere due valori (convenzionalmente 0 o 1). Questo deriva essenzialmente dal fatto che i dispositivi di memorizzazione più diffusi si basano, a livello elementare, sul *flip-flop*, che è un circuito elettronico a due stati.

In questa appendice ci occupiamo brevemente della rappresentazione dei numeri in un calcolatore e dei problemi numerici che si incontrano tipicamente nell'aritmetica in virgola mobile.

### D.1 SISTEMI DI NUMERAZIONE IN BASE DIVERSA DA 10

Quando scriviamo il numero 127 (nella consueta notazione posizionale in base 10) quello che in effetti intendiamo è:

$$(127)_{10} = 1 \times 10^2 + 2 \times 10^1 + 7 \times 10^0.$$

(Notate che abbiamo indicato esplicitamente la base per evitare ambiguità, ma ove essa dovesse essere omessa intenderemo sempre che il numero è scritto in base 10). Più in generale possiamo *contare* in una base arbitraria  $n$ , con il comune intendimento che

$$(c_m \dots c_2 c_1 c_0)_n = c_m \times n^m + \dots + c_2 \times n^2 + c_1 \times n + c_0. \quad (285)$$

Tipicamente  $n$  è un numero intero più grande di 1, ma in generale questa restrizione non è necessaria—e i sistemi di numerazione in base negativa, o addirittura immaginaria o complessa, possono avere proprietà interessanti. Per  $n = 2, 8, 16$  e  $60$  si hanno i sistemi di numerazione binario, ottale, esadecimale e sessagesimale, che sono largamente usati in alcuni contesti. Così, ad esempio, nel sistema binario possiamo scrivere

$$(1101)_2 \triangleq 0b1101 = 1 \times 2^3 + 1 \times 2^2 + 1 \times 1 = 13,$$

ed in quello esadecimale

$$(24)_{16} \triangleq 0x24 = 2 \times 16 + 4 = 36.$$

L'autore di questa appendice, nel momento in cui scrive, ha quasi 0x28 anni, pesa circa 0x66 kg, ed è da lungo tempo convinto fautore dell'introduzione della bilancia pesapersona esadecimale. Per inciso, i sistemi binario ed esadecimale sono così utilizzati in pratica, che è comune trovarli indicati con i prefissi 0b e 0x, rispettivamente (come indicato negli esempi sovrastanti); nel seguito utilizzeremo anche noi questa notazione.

È facile convincersi che con  $m$  cifre si possono esprimere (in base  $n$ )  $n^m$  numeri diversi—ad esempio tutti quelli compresi tra 0 e  $n^m - 1$ . Così con 3 cifre decimali si possono esprimere i 1000 numeri tra 0 e 999, mentre con 8 cifre binarie si possono esprimere solo i  $2^8 = 256$  numeri tra 0 e 255.

#### D.1.1 Il sistema binario

Come abbiamo detto in avvio di sezione il sistema binario riveste un ruolo peculiare perché è il *linguaggio* parlato dai calcolatori. Convertire un numero intero dal sistema decimale a quello binario (e viceversa) è

[https://bitbucket.org/.../bin\\_dec.py](https://bitbucket.org/.../bin_dec.py)

```

1 # Decimal to binary...
2 print(bin(11))
3 print(bin(123))
4 print(bin(999))
5 # ...and binary to decimal.
6 print(0b1)
7 print(0b110011)
8 print(0b1111111)
9
10 [Output]
11 0b1011
12 0b1111011
13 0b1111100111
14 1
15 51
16 255

```

FRAMMENTO D.1. Esempi di conversione tra il sistema decimale e quello binario, e viceversa, utilizzando il linguaggio di programmazione Python.

banale, seguendo la (285). Qualsiasi linguaggio di programmazione degli di questo nome, inoltre, fornisce funzioni di conversione, come mostrato nel frammento D.1.

La situazione diventa più interessante non appena introduciamo nel *mix* il separatore decimale. Che cosa significa, ad esempio, la scrittura 2.25? Evidentemente

$$2.25 = 2 + 2 \times 10^{-1} + 5 \times 10^{-2},$$

il che vuol dire che, se non siamo interessati solamente ai numeri interi, dobbiamo generalizzare la nostra definizione (285) come

$$(c_m \dots c_2 c_1 c_0 . c_{-1} c_{-2} \dots)_n = c_m \times n^m + \dots + c_2 \times n^2 + c_1 \times n + c_0 + c_{-1} n^{-1} + c_{-2} n^{-2} \dots \quad (286)$$

In questo nuovo schema possiamo scrivere il nostro 2.25 nel sistema di numerazione binario semplicemente come

$$0b10.01 = 1 \times 2 + 0 + 0 \times 2^{-1} + 1 \times 2^{-2} = 2 + \frac{1}{4} = 2.25.$$

Fino a qui nessuna sorpresa, ma cosa succede se proviamo a fare la stessa cosa con il numero (apparentemente innocuo) 0.1? Per rispondere possiamo provare ad approssimare (da sopra e da sotto) con un numero fissato di cifre binarie dopo il punto fino a che non otteniamo il risultato esatto:

$$\begin{array}{lll}
\frac{1}{16} = 0b0.0001 & \leq 0.1 \leq & 0b0.0010 = \frac{1}{8} \\
\frac{3}{32} = 0b0.00011 & \leq 0.1 \leq & 0b0.00100 = \frac{1}{8} \\
\frac{3}{32} = 0b0.000110 & \leq 0.1 \leq & 0b0.000111 = \frac{7}{64} \\
\frac{3}{32} = 0b0.0001100 & \leq 0.1 \leq & 0b0.0001101 = \frac{13}{128} \\
\frac{25}{256} = 0b0.00011001 & \leq 0.1 \leq & 0b0.00011010 = \frac{13}{128} \\
\frac{51}{512} = 0b0.000110011 & \leq 0.1 \leq & 0b0.000110100 = \frac{13}{128}
\end{array}$$

Possiamo andare avanti quanto vogliamo, ma non è difficile convincersi che il numero decimale 0.1 *non* ha uno sviluppo binario finito—esattamente come la frazione  $1/3$ , che si può scrivere come  $(0.1)_3$  in base 3





---

## COLOPHON

---

Queste dispense sono state realizzate utilizzando esclusivamente *software* libero (se non sai cosa significa, è il momento di dare un’occhiata alla pagina web della Free Software Foundation).

Il testo è stato battuto interamente con emacs ed impaginato utilizzando  $\text{\LaTeX}$ —un programma per la realizzazione di documenti che utilizza  $\text{\TeX}$ , di Donald Knuth, come motore tipografico. La classe di stile utilizzata è una versione altamente personalizzata (che sono certo l’autore disapproverebbe, ma mira essenzialmente a risparmiare carta) del superbo pacchetto *classicthesis* di André Miede.

Le figure che non contengono grafici sono state realizzate direttamente in  $\text{\LaTeX}$  utilizzando il (niente meno che impressionante) pacchetto di grafica vettoriale *pgf*, insieme al suo *layer* sintattico di alto livello *tikz*, di Till Tantau. I grafici, invece, sono stati realizzati utilizzando il *backend* *pgf* di matplotlib, insieme ad alcune delle altre componenti dell’ecosistema di calcolo scientifico di Python—in particolare *numpy* e *scipy*.

I frammenti di codice sono eseguiti automaticamente durante la compilazione del documento da un semplice *script* in Python che appende l’*output* al frammento corrispondente. (In altre parole: sono garantiti funzionare, almeno sul computer dell’autore.) I colori sono aggiunti utilizzando l’ottimo evidenziatore di sintassi *Pygments*.

Le dispense sono state, per la maggior parte, sviluppate su macchine che utilizzano la distribuzione Fedora di GNU/Linux come sistema operativo. Tutte le componenti del progetto sono sotto controllo di configurazione ed il repository git è ospitato su bitbucket a

<https://bitbucket.org/lbaldini/statnotes>

“Shut up, we know you can play! Jesus...”  
—Steve Vai, Juice (1995)