
PREFAZIONE

INTRODUZIONE

INDICE

1	MISURE ED ERRORI: PRIMI PASSI	1
1.1	Concetti di base: errore massimo e cifre significative	1
1.1.1	Il concetto di errore massimo	1
1.1.2	Errore assoluto ed errore relativo	2
1.1.3	Cifre significative e convenzioni di scrittura	2
1.2	Stima dell'errore massimo in condizioni di ripetitività	3
1.2.1	Misure in regime di dispersione nulla: la risoluzione strumentale	3
1.2.2	Fluttuazioni casuali: stima a posteriori dell'errore massimo	3
1.3	Propagazione dell'errore massimo	4
1.3.1	Propagazione dell'errore massimo nelle somma e nella differenza	4
1.3.2	Propagazione dell'errore massimo nel prodotto e nel quoziente	5
1.3.3	Propagazione dell'errore massimo nel caso generale	6
1.4	Errori sistematici	9
1.4.1	Errori strumentali	10
1.4.2	Incertezze di modellizzazione	10
1.4.3	Identificazione degli effetti sistematici	11
1.4.4	Propagazione degli effetti sistematici	12
2	RAPPRESENTAZIONE GRAFICA E RIDUZIONE DEI DATI	13
2.1	Scatter plot	13
2.2	Fit grafici	13
2.3	Istogrammi	13
2.4	Scale non lineari	13
2.4.1	Leggi di potenza e scale bilogaritmiche	14
2.4.2	Fit grafici in scala bilogaritmica	16
2.4.3	Funzioni esponenziali e scale semilogaritmiche	17
2.4.4	Fit grafici in scala semilogaritmica	18
3	ELEMENTI DI TEORIA DELLA PROBABILITÀ	19
3.1	Definizione assiomatica della probabilità	19
3.2	Definizioni operative di probabilità	19
3.2.1	Definizione combinatoriale	20
3.2.2	Definizione frequentista	20
3.2.3	Definizione soggettiva	21
3.3	Leggi elementari della probabilità	21
3.4	Probabilità condizionata	22
3.4.1	Eventi indipendenti	22
3.5	Il teorema di Bayes: un semplice esempio	23
4	VARIABILI CASUALI E FUNZIONI DI DISTRIBUZIONE	25
4.1	Variabili casuali	25
4.2	Caratteristiche comuni alle distribuzioni	25
4.2.1	Valore di aspettazione	25
4.2.2	Misure di tendenza centrale	25
4.2.3	Misure di dispersione attorno alla media	26
4.2.4	La disuguaglianza di Chebyshev	27
4.2.5	Momenti di ordine superiore	28
4.2.6	La distribuzione cumulativa	28
5	DISTRIBUZIONI DI VARIABILE DISCRETA	29
5.1	La distribuzione binomiale	29
5.1.1	Normalizzazione, media e varianza	30
5.2	La distribuzione di Poisson	31
5.2.1	Stazionarietà e processi Poissoniani	32
5.2.2	Normalizzazione, media e varianza	35
5.2.3	Somma di variabili Poissoniane	35
5.2.4	La distribuzione di Poisson composta	36
6	DISTRIBUZIONE DI VARIABILE CONTINUA	37

INDICE

6.1	La distribuzione uniforme	37
6.1.1	Normalizzazione, media e varianza	37
6.2	La distribuzione esponenziale	38
6.2.1	Normalizzazione, media e varianza	39
6.3	La distribuzione di Gauss	40
6.3.1	Normalizzazione, media e varianza	40
6.3.2	L'integrale normale degli errori	41
6.3.3	La distribuzione di Gauss come limite della Poissoniana	41
6.3.4	La distribuzione di Gauss come limite della binomiale	42
6.3.5	La distribuzione di Gauss e gli errori di misura	42
6.3.6	Una derivazione alternativa della distribuzione di Gauss	43
7	VARIABILI CAMPIONE ED ESTIMATORI	45
7.1	Media e varianza campione	45
7.2	Deviazione standard della media	45
7.3	Il teorema del limite centrale	45
7.3.1	Un esempio interessante: il <i>random walk</i> in una dimensione	45
	Bibliografia	47

 MISURE ED ERRORI: PRIMI PASSI

1.1 CONCETTI DI BASE: ERRORE MASSIMO E CIFRE SIGNIFICATIVE

Il concetto di *errore* di misura gioca un ruolo centrale nelle scienze sperimentali. Non è concettualmente possibile *misurare* una grandezza fisica con infinita precisione, ed 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 degli errori di misura: se entriamo in un negozio di materiale elettrico per comperare 20 m di cavo non specifichiamo se ci aspettiamo che la tolleranza sulla lunghezza del cavo stesso debba essere di 1 mm, 1 cm, 10 cm o 1 m. Eppure il contesto ed il senso comune ci suggeriscono che in questa situazione specifica 1 cm o 10 cm sono valori più ragionevoli di 1 mm o 1 m. (Questo ha implicazioni pratiche: se, arrivati a casa, scopriremo che il cavo che abbiamo comprato è lungo 15 m ci sentiremo legittimati a tornare nel negozio e lamentarci, ma se la lunghezza fosse 19.95 m probabilmente no. D'altra parte se insistiamo con il negoziante che esigiamo una tolleranza di 1 mm, probabilmente ci sentiremo rispondere di rivolgersi ad un altro negozio.)

In Fisica, e nelle scienze sperimentali in genere, si cerca tipicamente di evitare ambiguità ed usare parole e frasi in 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. In questo caso è il bagaglio di conoscenza comune (che cominceremo a sviluppare in questo capitolo) a guidarci. Se, per esempio, leggiamo in un articolo scientifico che l'ampiezza di un certo segnale elettrico è 5 V, possiamo legittimamente aspettarci che l'incertezza associata sia 0.5 V o 1 V—ma non 10 V o 0.01 V. Per questo motivo, nello stesso contesto, le scritture 5.0 V o 5.00 V avrebbero un significato profondamente diverso.

1.1.1 Il concetto di errore massimo

È uso comune¹ scrivere il risultato della *misura* (sia essa diretta o indiretta) di una generica grandezza fisica x come:

$$x = \hat{x} \pm \Delta x. \quad (1)$$

I tre ingredienti della (1) hanno ciascuno un significato ben preciso che è bene fissare il prima possibile:

1. x è il valore (incognito) della grandezza che vogliamo misurare—che chiameremo anche *misurando* (cercheremo deliberatamente di evitare l'espressione *valore vero*, che pure si trova usata in letteratura);
2. \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;
3. Δx è l'*incertezza* (o errore) di misura—nel seguito useremo questi due termini essenzialmente senza alcuna distinzione. (Notiamo che per definizione Δx rappresenta la lunghezza di un intervallo ed è perciò una grandezza definita positiva.)

A questo livello Δx è da interpretarsi come *errore massimo*, e quando scriviamo la (1) stiamo dicendo che siamo *certi* che l'intervallo $[\hat{x} - \Delta x, \hat{x} + \Delta x]$ contenga il valore (incognito) x del misurando. (Più precisamente, \hat{x} e Δx definiscono il più piccolo intervallo che ci fornisca questa certezza.) Vedremo nel seguito che la nozione di errore massimo è patologicamente problematica (e la abbandoneremo il prima possibile) ma per il momento essa ha il vantaggio di permetterci di cominciare senza aver bisogno di strumenti particolarmente sofisticati.

¹ Per l'esattezza, allo scopo di evitare ambiguità, nel seguito cercheremo di seguire religiosamente la convenzione di indicare l'errore *massimo* con la lettera Δ e l'errore *statistico* (che introdurremo più avanti) con la lettera σ , per cui nel seguito scriveremo anche (ed, anzi, prevalentemente) $x = \hat{x} \pm \sigma_x$.

L'idea di base è che, almeno in linea di principio, migliorando il nostro apparato di misura possiamo rendere Δx piccolo a piacere e fare in modo che il valore centrale \hat{x} si avvicini arbitrariamente al misurando. Discuteremo brevemente nel seguito (e.g., nella sezione 1.4) perché questo, in pratica, non è generalmente possibile.

Notiamo anche che 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. Nel seguito cercheremo di usare questi due termini nell'accezione appena indicata.

1.1.2 Errore assoluto ed errore relativo

Δx si dice anche *errore assoluto* (per distinguerlo dall'errore relativo che definiremo tra un attimo), ed ha, per ovvie ragioni, le stesse dimensioni fisiche della misura a cui si riferisce. Definiamo *errore relativo* il rapporto tra l'incertezza della misura ed il suo valore centrale:

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

L'errore relativo è per definizione una quantità adimensionale (e positiva) ed è un concetto utile perché, ad esempio, un errore assoluto di 1 mm, è più critico sulla misura di una lunghezza di 1 cm piuttosto che su quella di 1 m: nel primo caso l'errore relativo è del 10%, mentre nel secondo è dello 0.1%.

1.1.3 Cifre significative e convenzioni di scrittura

La precisione di un risultato sperimentale è almeno parzialmente implicita nel modo in cui il risultato stesso è scritto. Abbiamo già avuto modo di osservare che in Fisica i numeri 5, 5.0 e 5.00 non hanno lo stesso significato—ed il motivo è che non hanno lo stesso numero di *cifre significative*.

La nozione di cifra significativa si definisce tramite le tre semplici regole qui di seguito:

1. la cifra più significativa è quella più a sinistra e diversa da zero;
2. la cifra meno significativa è quella più a destra;
3. tutte le cifre comprese tra la più significativa e la meno significativa sono significative.

► **ESEMPIO 1.1.** Negli esempi che seguono indichiamo con una sottolineatura la cifra meno significativa, con una soprallineatura quella più significativa e tra parentesi quadre il numero di cifre significative: $\overline{3}11\underline{5}$ [4], $\overline{3}212.\underline{5}$ [5], $0.\overline{03}2\underline{5}$ [3], $0.\overline{03}0\underline{0}$ [3], $\overline{3}00.\underline{00}$ [5].

► **ESEMPIO 1.2.** Nella nostra prescrizione il numero 300 ha 3 cifre significative, anche se, di per sé, non è ovvio che in questo caso i due zeri a destra siano veramente significativi (o non servano, piuttosto, solo ad indicare l'ordine di grandezza). In pratica si può risolvere l'ambiguità scrivendo esplicitamente 300. (con il punto decimale nel caso in cui le cifre significative siano 3) o 3×10^2 (nel caso di una sola cifra significativa).

La nozione di cifra significativa è strettamente connessa con quella di *arrotondamento* del valore centrale della misura. Il numero di cifre significative con cui si deve scrivere il risultato di una misura è determinato dall'incertezza associata: tipicamente si scrivono (arrotondando opportunamente) *tutte le cifre fino alla prima o alla seconda affetta da errore*—per il semplice motivo che le cifre successive sono essenzialmente mero *rumore* numerico, senza alcun contenuto quantitativo. In altre parole si scrive l'errore con una o due cifre significative, dopo di che si arrotonda opportunamente il valore della misura in modo da essere consistente, in termini di cifre decimali, con l'incertezza associata. Gli esempi che seguono dovrebbero chiarire il concetto.

► **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 ± 0.542	errore con troppe (tre) cifre significative	$[3.44 \pm 0.54]$
3.4 ± 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 ± 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 ± 100	ambigua per gli zeri meno significativi	$[(3.0 \pm 0.1) \times 10^3]$

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.

1.2 STIMA DELL'ERRORE MASSIMO IN CONDIZIONI DI RIPETITIVITÀ

Quando una misura è fatta più volte in condizioni di ripetitività si danno due casi tipici: (i) la *dispersione* delle misure è nulla, cioè otteniamo sempre lo stesso risultato, oppure (ii) il valore ottenuto *fluttua*, cioè due misure successive forniscono in generale risultati diversi tra loro.

1.2.1 Misure in regime di dispersione nulla: la risoluzione strumentale

La *risoluzione* di uno strumento è essenzialmente la più piccola *variazione* della quantità da misurare che è possibile apprezzare con lo strumento stesso². 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 come errore massimo di misura.

► **ESEMPIO 1.4.** 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.

► **ESEMPIO 1.5.** 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.

Per quanto apparentemente banali, i due esempi precedenti non devono trarre in inganno: la vita è spesso più complicata e in alcune situazioni identificare l'accuratezza di una misura (nel senso della distanza dal misurando) con la risoluzione dello strumento utilizzato può portare ad errori apprezzabili. Torneremo brevemente sulla questione nella sezione 1.4.

1.2.2 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. 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_i ($i = 1 \dots n$) della grandezza x cui siamo interessati, è ragionevole prendere la media aritmetica delle misure stesse come valore centrale

$$\hat{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

e la semidispersione come errore massimo

$$\Delta x = \frac{x_{\max} - x_{\min}}{2}. \quad (4)$$

² 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).

► **ESEMPIO 1.6.** Vogliamo misurare il periodo τ 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 del cronometro e dalla persona che fisicamente lo utilizza. Non è banale quantificare a priori l'incertezza di misura, ma possiamo sospettare che il tempo di reazione umano rappresenti 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 molto più piccola del tempo di reazione e di fatto non gioca alcun ruolo nella misura.) Seguendo la (3) e la (4) scriveremo $\tau = 2.15 \pm 0.06$ s.

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 da almeno due problemi insanabili. 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 in cui eseguire nuove misure (cioè aggiungere *informazione*) causa un incremento dell'incertezza. (Intuitivamente ci aspetteremmo che all'aumentare di n l'incertezza di misura diminuisca, ed in effetti vedremo nel seguito che, a patto di utilizzare una definizione più ragionevole di incertezza, questa intuizione è corretta.)

1.3 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 (che si incontra comunemente in laboratorio) di come l'errore massimo si propaghi nelle misure indirette.

Formalmente, dato un certo numero di grandezze indipendenti³ 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. Come è facile intuire, nel caso generale si tratta di un problema niente affatto banale. Fortunatamente, come vedremo tra un attimo, nella maggior parte dei casi di interesse pratico la soluzione del corrispondente problema *linearizzato* fornisce una risposta sufficientemente accurata.

1.3.1 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 con errore. I casi estremi sono quelli in cui gli errori 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 in una somma *gli errori massimi si sommano*

$$\Delta S = \frac{S_{\max} - S_{\min}}{2} = \Delta x + \Delta y. \quad (5)$$

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 $D = x - y$

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

³ 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.

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*:

$$\Delta D = \frac{D_{\max} - D_{\min}}{2} = \Delta x + \Delta y. \quad (6)$$

(Se siete veramente tentati di scrivere $\Delta D = \Delta x - \Delta y$ pensate a cosa succederebbe quando $\Delta x = \Delta y$. . . Triste a dirsi, ma quando le grandezze misurate sono indipendenti gli errori non si sottraggono mai.)

► **ESEMPIO 1.7.** Abbiamo misurato direttamente con un calibro ventesimale le due lunghezze $a = 2.50 \pm 0.05$ cm e $b = 2.35 \pm 0.05$ cm e vogliamo propagare l'errore sulla somma S e sulla differenza D . Applicando direttamente la (5) e la (6) si ha $S = 4.85 \pm 0.10$ cm e $D = 0.15 \pm 0.10$ cm.

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 (e di grande rilevanza pratica) quando si sottraggono misure vicine tra loro.

1.3.2 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

$$P_{\max} = (\hat{x} + \Delta x)(\hat{y} + \Delta y) = \hat{x}\hat{y} + (\hat{x}\Delta y + \hat{y}\Delta x + \Delta x\Delta y)$$

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

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 (niente a fatto inusuale, 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:

$$\Delta P = \frac{P_{\max} - P_{\min}}{2} = |\hat{x}| \Delta y + |\hat{y}| \Delta x \quad (7)$$

(abbiamo ripristinato i valori assoluti, che sono necessari nel caso generale in cui i valori centrali x o y sono negativi per garantire che ΔP sia positivo). Per coloro che si sentissero insoddisfatti da questa approssimazione apparentemente brutale, vale la pena ricordare che ΔP 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.

► **ESEMPIO 1.8.** Vogliamo propagare l'errore sul prodotto $P = ab$ delle due grandezze dell'esempio 1.7. I due termini della (7) sono $\hat{a}\Delta b \approx 0.12$ cm² e $\hat{b}\Delta a \approx 0.12$ cm²; per confronto, il termine che abbiamo trascurato è $\Delta a\Delta b \approx 0.0025$ cm², ossia circa 50 volte più piccolo. Possiamo dunque scrivere $P = 5.87 \pm 0.24$ cm² (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

$$\Delta Q = \frac{Q_{\max} - Q_{\min}}{2} = \frac{|\hat{x}| \Delta y + |\hat{y}| \Delta x}{\hat{y}^2}. \quad (8)$$

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 (8) si semplifica in $\Delta Q = \Delta x/\hat{y}$. Questo permette di fare misure (indirette) con un errore massimo inferiore alla risoluzione strumentale, come illustrato nell'esempio seguente.

► **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. Si ha $s = h/500$ —qui il denominatore è un numero non affetto da errore—per cui l'errore propagato sarà $\Delta s = \Delta h/500$ e la nostra stima dello spessore del singolo foglio sarà $s = 0.080 \pm 0.002$ mm, in cui l'errore massimo è (500 volte!) più piccolo della risoluzione del metro a nastro.

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

$$\frac{\Delta P}{|P|} = \frac{|\hat{x}| \Delta y + |\hat{y}| \Delta x}{|\hat{x}| |\hat{y}|} = \frac{\Delta x}{|\hat{x}|} + \frac{\Delta y}{|\hat{y}|} \quad (9)$$

$$\frac{\Delta Q}{|Q|} = \frac{|\hat{x}| \Delta y + |\hat{y}| \Delta x}{\hat{y}^2} \Big/ \frac{1}{|\hat{y}|} = \frac{|\hat{x}| \Delta y + |\hat{y}| \Delta x}{|\hat{x}| |\hat{y}|} = \frac{\Delta x}{|\hat{x}|} + \frac{\Delta y}{|\hat{y}|}, \quad (10)$$

ovvero *nelle somme e nei prodotti gli errori massimi relativi si sommano.*

1.3.3 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 Δx . Possiamo sviluppare f in serie di Taylor al prim'ordine attorno al punto \hat{x} :

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

A rigore al secondo membro dovremmo usare \pm o \mp a seconda che il valore della derivata sia positivo o negativo, ma per i nostri scopi possiamo prendere il modulo e scrivere

$$\Delta f = \frac{f_{\max} - f_{\min}}{2} \approx \left| \frac{df}{dx}(\hat{x}) \right| \Delta x. \quad (11)$$

Il simbolo \approx ci ricorda che la (11) 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, che la retta tangente ad f nel punto \hat{x} sia una buona approssimazione di f nell'intervallo $[\hat{x} - \Delta x, \hat{x} + \Delta x]$.

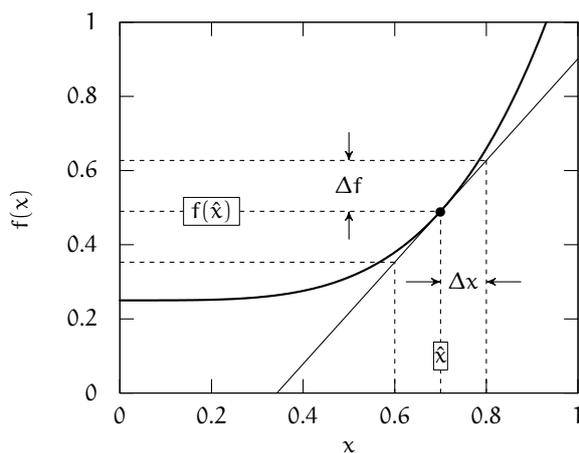


FIGURA 1. Rappresentazione grafica della linearizzazione della propagazione dell'errore su una generica funzione di una variabile (11). L'approssimazione è tanto più accurata quanto più la retta tangente ad f nel punto \hat{x} è vicina alla funzione nell'intervallo $[\hat{x} - \Delta x, \hat{x} + \Delta x]$.

Come detto in precedenza questa assunzione, che dipende sia dal valore di Δx che dalla forma della funzione $f(x)$ in un intorno di \hat{x} , va verificata caso per caso. (Un caso banale in cui il nostro sviluppo in serie troncato al prim'ordine non funziona è quello in cui $f'(\hat{x}) = 0$, come illustrato nell'esempio 1.11.)

► **ESEMPIO 1.10.** 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 (11), che nel caso specifico fornisce:

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

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 1.11.** Dato un angolo (misurato) $\theta = \hat{\theta} \pm \Delta\theta = (90 \pm 1)^\circ$ si vuole propagare l'errore sulla funzione $\cos \theta$. In questo caso la derivata prima che compare nella (11) si annulla nel punto $\hat{\theta}$

$$\left| \frac{d \cos \theta}{d\theta}(\hat{\theta}) \right| = |\sin \hat{\theta}| = 0,$$

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 oppure estendere lo sviluppo all'ordine successivo.

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

$$\Delta f \approx \left| \frac{\partial f}{\partial x}(\hat{x}, \hat{y}, \hat{z}, \dots) \right| \Delta x + \left| \frac{\partial f}{\partial y}(\hat{x}, \hat{y}, \hat{z}, \dots) \right| \Delta y + \left| \frac{\partial f}{\partial z}(\hat{x}, \hat{y}, \hat{z}, \dots) \right| \Delta z + \dots \quad (12)$$

Il lettore può verificare facilmente che la (12) permette di derivare le formule che abbiamo ricavato "a mano" per la propagazione dell'errore massimo nelle quattro operazioni elementari. Essa permette anche di ricavare una formula compatta (e utile in pratica) per la propagazione dell'errore massimo relativo su una grandezza del tipo $G(x, y, z, \dots) = x^\alpha y^\beta z^\gamma \dots$

$$\frac{\Delta G}{|G|} = \frac{|\alpha \hat{x}^{\alpha-1} \hat{y}^\beta \hat{z}^\gamma| \Delta x + |\hat{x}^\alpha \beta \hat{y}^{\beta-1} \hat{z}^\gamma| \Delta y + |\hat{x}^\alpha \hat{y}^\beta \gamma \hat{z}^{\gamma-1}| \Delta z \dots}{|\hat{x}^\alpha| |\hat{y}^\beta| |\hat{z}^\gamma| \dots} = |\alpha| \frac{\Delta x}{|\hat{x}|} + |\beta| \frac{\Delta y}{|\hat{y}|} + |\gamma| \frac{\Delta z}{|\hat{z}|} + \dots \quad (13)$$

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.*

► **ESEMPIO 1.12.** 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

$$T = 2\pi \sqrt{\frac{l}{g}}. \quad (14)$$

Misuriamo dunque direttamente $l = 0.984 \pm 0.003$ m e $T = 2.02 \pm 0.05$ s. La relazione scritta sopra si può invertire per ottenere un'espressione per g

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

da cui possiamo calcolare il valore centrale $\hat{g} = 9.52$ m s⁻². Non ci rimane che propagare l'errore su g utilizzando la 12. Per chiarire il significato di tutti i termini svolgeremo questo esempio pedissequamente:

$$\frac{\partial g}{\partial l} = \frac{4\pi^2}{T^2} \quad \text{e} \quad \frac{\partial g}{\partial T} = -\frac{8\pi^2 l}{T^3}, \quad \text{da cui} \quad \Delta g = \left| \frac{dg}{dl}(\hat{l}, \hat{T}) \right| \Delta l + \left| \frac{dg}{dT}(\hat{l}, \hat{T}) \right| \Delta T = \frac{4\pi^2}{\hat{T}^2} \left(\Delta l + \frac{2\hat{l}\Delta T}{\hat{T}} \right).$$

Scriveremo dunque il risultato come $g = 9.52 \pm 0.50$ m s⁻², che é compatibile con il valore *note* di 9.81 m s⁻². Notiamo esplicitamente che, dividendo per il valore centrale, otteniamo, in accordo con la 13

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

Per quanto (apparentemente) banale, l'esempio 1.12 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 (14)—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 θ_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 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 1.4.

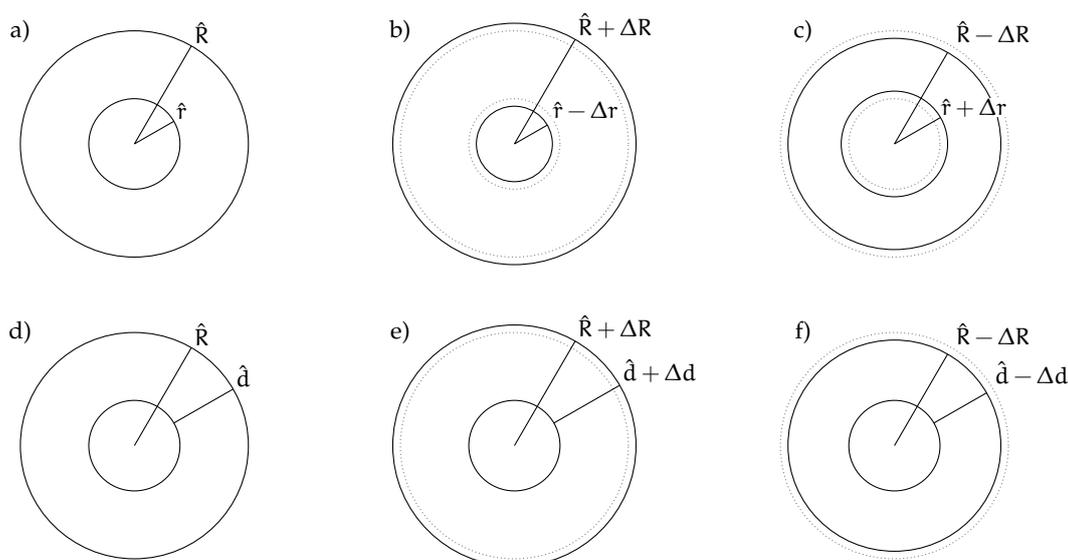


FIGURA 2. Illustrazione grafica della propagazione dell'errore massimo sulla misura della superficie di una corona circolare (esempi 1.13 e 1.14). Se misuriamo il raggio esterno R ed il raggio interno r , la configurazione di area massima si ha quando R è massimo e r è minimo (entro i rispettivi errori 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).

L'esempio che segue illustra, con gli strumenti limitati che abbiamo a disposizione a questo punto, il concetto di *indipendenza* delle misure nella propagazione dell'errore massimo (abbiamo già avuto modo di dire che, nel caso di grandezze non indipendenti la (12) semplicemente non vale). Esso 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.

► **ESEMPIO 1.13.** Propaghiamo l'errore massimo sulla superficie A di una corona circolare di cui si misurano il raggio esterno $R = \hat{R} \pm \Delta R = 15.0 \pm 0.1$ cm ed il raggio interno $r = \hat{r} \pm \Delta r = 6.0 \pm 0.1$ cm. Si ha

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

da cui possiamo calcolare il valore centrale della misura $\hat{A} = 594$ cm². L'errore si scrive come

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

per cui scriveremo la nostra misura come $A = 594 \pm 13$ cm². Si può verificare per sostituzione diretta che l'errore

massimo propagato coincide (approssimativamente) con la semidispersione delle superfici massima e minima mostrate in figura 2 (pannelli b e c):

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

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

► **ESEMPIO 1.14.** Ripetiamo l'esercizio precedente nell'ipotesi in cui misuriamo il raggio esterno $R = \hat{R} \pm \Delta R = 15.0 \pm 0.1 \text{ cm}$ e lo spessore della corona circolare $d = \hat{d} \pm \Delta d = 9.0 \pm 0.1 \text{ cm}$.

[Prima di andare avanti fermiamoci un attimo ad osservare la figura 2: i valori centrali di R e d corrispondono allo stesso oggetto fisico dell'esempio 1.13, ma le configurazioni di area massima e minima definite dagli involucri degli errori sono diverse nei due casi. Assicurati di aver capito questo prima di passare oltre.]

Al contrario dell'esempio precedente cominciamo dunque dal calcolo esplicito delle superfici massima e minima mostrate in figura 2 (pannelli e ed f):

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

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

La semidispersione delle due è $\approx 9 \text{ cm}^2$. A questo punto potremmo essere tentati di applicare il formalismo sviluppato in questa sezione ricavando il valore del raggio interno r a partire da R e d (con l'opportuna propagazione dell'errore massimo)

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

ed utilizzando poi la formula ricavata nell'esempio 1.13. È semplice verificare, però, che così facendo l'errore propagato ($\approx 17 \text{ cm}^2$) non coincide con la semidispersione delle superfici massima e minima che abbiamo appena calcolato. Il problema di fondo è che le quantità R ed r in questo caso non sono indipendenti poiché r è stato calcolato a partire da R . Un modo corretto di inquadrare il problema è invece quello di scrivere la superficie in funzione delle due grandezze (misurate) indipendenti

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

e propagare l'errore di conseguenza

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

L'errore è, in questo caso, più piccolo di quello calcolato nell'esempio 1.13: nel caso specifico è conveniente misurare raggio esterno e spessore piuttosto che raggio esterno e raggio interno.

1.4 ERRORI SISTEMATICI

Nella sezione 1.1.1 abbiamo introdotto la distinzione tra le nozioni di precisione ed accuratezza. Torniamo adesso su una domanda che abbiamo toccato brevemente più volte, ovvero: 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 dei cosiddetti *errori sistematici*⁴.

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.

⁴ Non avendo ancora introdotto gli *errori statistici* non abbiamo il bagaglio necessario per affrontare il nocciolo della questione, ma possiamo illustrare con alcuni semplici esempi le idee di base.

1.4.1 Errori strumentali

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

► **ESEMPIO 1.15.** Riesaminiamo brevemente l'esempio 1.5 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°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°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°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°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 1.16.** 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 $L = 2$ m aumenta di un fattore $\lambda L(T_2 - T_1) \approx 1$ mm—che è pari alla risoluzione del metro stesso. (Nel caso di un calibro cinquantessimale l'effetto può essere significativamente più vistoso.) 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?

1.4.2 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. Illustriamo con un esempio questo genere di situazione.

► **ESEMPIO 1.17.** Riesaminiamo l'esempio 1.12—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 massima θ_0 oltre la quale il nostro modello semplicistico (valido in regime di piccole oscillazioni) cessava di essere appropriato.

Partiamo dunque dallo sviluppo in serie di Taylor dell'espressione per il periodo in funzione di θ_0 :

$$T = T_0 \left(1 + \frac{1}{16}\theta_0^2 + \frac{11}{3072}\theta_0^4 + \dots \right), \quad \text{con} \quad T_0 = 2\pi\sqrt{\frac{l}{g}}. \quad (15)$$

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, almeno in parte, implicita nella domanda: la condizione è che ciò che abbiamo trascurato fosse molto più piccolo dell'errore di misura

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

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 performante), 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 1.12 sarà piccolo se Δl e ΔT sono piccoli, ma ad un certo punto entreremo nelle condizioni in cui non sarà più lecito utilizzare la (14) 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 θ_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 (15) 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 15, 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.

1.4.3 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 che utilizzato per interpretare i dati.

► **ESEMPIO 1.18.** Supponiamo di prendere i due termometri dell'esempio 1.15 e di immergerli in un (abbondante) misto di acqua e ghiaccio—cioè in un bagno termico a 0°C . Il primo termometro segna 0.2°C ed il secondo 1.5°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 ρ 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.

1.4.4 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 massimo⁵. La differenza di base è che gli errori sistematici *non sono indipendenti* tra misura e misura—se un termometro 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.

⁵ Come vedremo in seguito, non seguono nemmeno le regole di propagazione dell'errore statistico, che tutto sommato sono affini.

RAPPRESENTAZIONE GRAFICA E RIDUZIONE DEI DATI

In Fisica—e specialmente nella Fisica di frontiera—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 (qualsiasi cosa ciò voglia dire) 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 SCATTER PLOT

Uno *scatter plot* (o grafico di dispersione, anche se il nome italiano non è molto usato) è un diagramma in cui coppie ordinate di dati (ad esempio misure dirette) sono visualizzate come punti (eventualmente con barre d'errore) su un piano cartesiano. Si tratta probabilmente della più semplice rappresentazione possibile di una serie di dati—eppure è spesso utile, in pratica, per mettere in evidenza proprietà rilevanti dei dati stessi.

2.2 FIT GRAFICI

2.3 ISTOGRAMMI

2.4 SCALE NON LINEARI

Per l'occhio umano è relativamente semplice percepire deviazioni da un andamento lineare. Non è così semplice, ad esempio, distinguere *immediatamente* i grafici cartesiani di polinomi 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 del fit grafico.

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.

A questo proposito, vi è un aspetto ulteriore del problema che vale la pena illustrare brevemente—cioè il fatto che in natura esistono fenomeni che scalano in modo intrinsecamente logaritmico. La percezione del livello sonoro, ad esempio, scala approssimativamente con il logaritmo della pressione sonora¹ e si misura tipicamente in dB (decibel)

$$L = 10 \log_{10} \left(\frac{p^2}{p_{\text{ref}}^2} \right) = 20 \log_{10} \left(\frac{P}{P_{\text{ref}}} \right) \quad [\text{dB}] \quad (16)$$

(la pressione di riferimento è convenzionalmente il valore $P_{\text{ref}} = 20 \mu\text{Pa}$, che è considerata la soglia di udibilità dell'orecchio umano). Nello studio di questo tipo di fenomeni si ha spesso a che vedere con grandezze che possono variare di molti ordini di grandezza e per cui i *rapporti* relativi hanno molto più significato fisico delle *differenze relative*. Questo tipo di grandezze, come vedremo, può essere convenientemente rappresentato in scala logaritmica.

► **ESEMPIO 2.1.** Una normale conversazione tra due persone ad ≈ 1 m di distanza corrisponde ad una pressione sonora di circa 2–20 mPa, ovvero 40–60 dB, come si desume facilmente dalla (16). Si tratta di un valore tra 100 e

¹ Per essere precisi l'orecchio umano è sensibile solo a *variazioni* di pressione in un intervallo di frequenze compreso tra circa 20 Hz e 20 kHz (quello che si dice comunemente *spettro udibile*). Per *pressione sonora* qui intendiamo dunque lo scarto quadratico medio della componente variabile (a frequenze udibili) della pressione sonora. È sorprendente come la soglia di udibilità dell'orecchio umano ad 1 kHz sia circa $20 \mu\text{Pa}$ —cioè il nostro orecchio è in grado di cogliere variazioni di ampiezza un miliardo di volte più piccole rispetto alla pressione (costante) atmosferica, che è dell'ordine di 100 kPa.

1,000 volte più alto della soglia di udibilità—e, purtroppo, circa 10,000 volte più piccolo della soglia del dolore (130 dB).

In questo contesto la *decade*² (o, in contesti più specifici, *l'ottava*, il *tono* ed il *semitono*) è l'unità più conveniente per misurare quanto due valori differiscono tra loro. Formalmente la distanza in decadi tra due valori x_1 ed x_2 è data da

$$\text{Distanza in decadi}(x_1, x_2) = \log_{10} \left(\frac{x_2}{x_1} \right). \quad (17)$$

Analogamente, in musica, l'ottava è l'intervallo tra due note le cui frequenze siano una il doppio dell'altra; la distanza in ottave tra due frequenze ν_1 ed ν_2 si misura come

$$\text{Distanza in ottave}(\nu_1, \nu_2) = \log_2 \left(\frac{\nu_2}{\nu_1} \right). \quad (18)$$

Ancora: essendo un'ottava composta da 12 semitoni, la distanza in semitoni si scrive

$$\text{Distanza in semitoni}(\nu_1, \nu_2) = 12 \log_2 \left(\frac{\nu_2}{\nu_1} \right). \quad (19)$$

► ESEMPIO 2.2. Il numero di decadi tra 10^n e 10^{n+m} è esattamente

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

► ESEMPIO 2.3. La frequenza del LA dell'ottava centrale del pianoforte è 440 Hz, mentre quello dell'ottava successiva è 880 Hz. Le due frequenze differiscono di un fattore 2 per cui le note corrispondenti distano di un'ottava.

2.4.1 Leggi di potenza e scale bilogaritmiche

Si dice *legge di potenza* una funzione del tipo

$$y(x) = Cx^\beta, \quad (20)$$

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

► ESEMPIO 2.4. 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 $\beta = 2$.

► ESEMPIO 2.5. 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 $\beta = 1/2$.

► ESEMPIO 2.6. Una retta passante per l'origine $y(x) = Cx$ è un caso particolare di legge di potenza con $\beta = 1$.

² Mentre nel linguaggio comune il termine decade indica un intervallo di tempo di 10 anni, in Fisica esso si usa per indicare l'intervallo tra due valori (ad esempio 100 e 1,000, che differiscono tra loro di un fattore 10).

Calcolando il logaritmo in base 10 di entrambi i membri della (20), essa può essere riscritta come:

$$\log_{10}(y) = \log_{10}(Cx^\beta) = \log_{10}(C) + \beta \log_{10}(x).$$

Mediante il cambiamento di variabile

$$\begin{cases} x' = \log_{10}(x) \\ y' = \log_{10}(y) \end{cases}$$

e ponendo $C' = \log_{10}(C)$, essa diviene:

$$y'(x') = C' + \beta x'. \quad (21)$$

Nelle nuove variabili la (20) 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 $\log_{10}(y_i)$ in funzione di $\log_{10}(x_i)$ 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*. Assi, cioè, su cui la distanza tra due punti x_1 ed x_2 (misurata con il righello) sia proporzionale al numero di decadi $\log_{10}(x_2/x_1)$ che separano x_1 ed x_2 —piuttosto che alla loro differenza assoluta $(x_2 - x_1)$ come avviene su una normale scala lineare. Così in scala logaritmica la distanza fisica tra 1 e 10 è identica a quella tra 10 e 100 o 0.01 e 0.1³.

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* (figura 3).

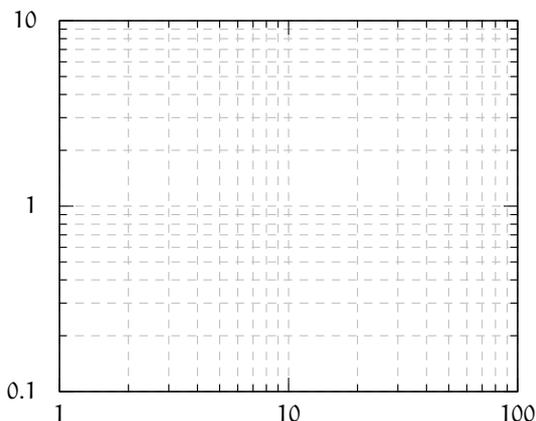


FIGURA 3. Esempio di carta bilogaritmica con due decadi sull'asse delle x e due decadi su quello delle y .

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.

► **ESEMPIO 2.7.** Si dispone di 5 sferette (di grandezza variabile) dello stesso materiale. Per ogni sferetta si misura il raggio r e la massa m . Fissata la densità ρ del materiale ci aspettiamo che le due grandezze siano legate dalla relazione:

$$m(r) = \frac{4}{3}\pi\rho r^3,$$

che è una legge di potenza con esponente 3. In figura 4 sono mostrati i grafici relativi in scala lineare e bilogaritmica (gli errori di misura sono troppo piccoli per essere rappresentati).

Nel primo caso (quello lineare) non è ovvio riconoscere una legge di potenza; inoltre è difficile leggere i valori di m per i due o tre punti più a sinistra, che tendono ad essere *schiacciati* nella parte inferiore del grafico.

Nel grafico in scala bilogaritmica, viceversa, il fatto che i punti si dispongano su una retta suggerisce immediatamente che ci troviamo di fronte ad una legge di potenza. La non linearità della scala delle y , inoltre,

³ Va da sé che lo 0 (ed a maggior ragione i valori negativi) non può essere rappresentato su scala logaritmica in quanto la distanza da esso un qualsiasi valore positivo divergerebbe. Questo è sostanzialmente una conseguenza del fatto che $\lim_{x \rightarrow 0^+} \log_{10}(x) = -\infty$.

rende molto più agevole la lettura dei dati, nonostante il fatto che il valore più grande di m si di tre ordini di grandezza più grande del più piccolo.

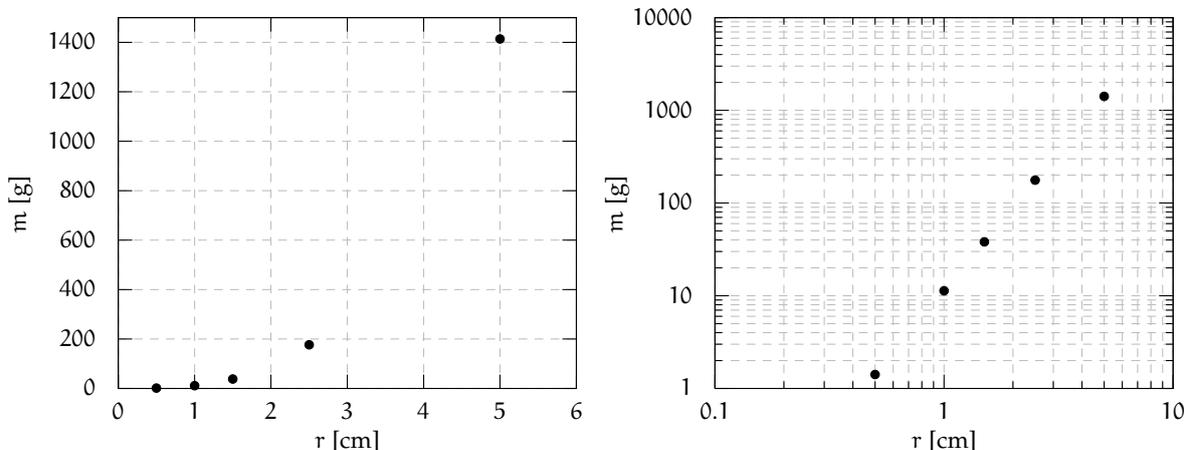


FIGURA 4. Grafici massa-volume relativi all'esempio 2.7, in scala lineare (a sinistra) e bilogaritmica (a destra).

2.4.2 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

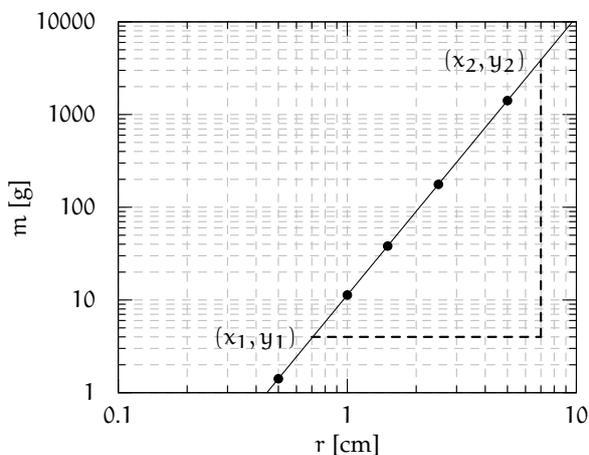


FIGURA 5. Fit grafico in carta bilogaritmica dei punti relativi all'esempio 2.7.

descritto per i grafici in scala lineare. Vi sono però almeno due differenza fondamentali che vale la pena descrivere in dettaglio.

Presi due punti sulla nostra retta di fit (figura 5), il coefficiente angolare si ricava dalla 21 come

$$\beta = \frac{y'_2 - y'_1}{x'_2 - x'_1} = \frac{\log_{10}(y_2) - \log_{10}(y_1)}{\log_{10}(x_2) - \log_{10}(x_1)} = \frac{\log_{10}(y_2/y_1)}{\log_{10}(x_2/x_1)}. \tag{22}$$

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, β si calcola usando la (22) (al solito propagando gli errori ove necessario). Se invece misuriamo *direttamente* i cateti del triangolo rettangolo in figura 5, dobbiamo usare la lunghezza di una decade (su ciascuno dei due assi separatamente) come unità di misura⁴.

⁴ Ne segue banalmente 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 (21), poniamo $x' = 0$, si ha $y' = C'$, ovvero $y = C$. Ma la condizione $x' = 0$ corrisponde, nelle variabili non trasformate, a $x = 1$

$$C = y(x' = 0) = y(x = 1), \quad (23)$$

per cui *in scala bilogaritmica 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 2.8.** Stimiamo i parametri del fit con una legge di potenza ai dati dell'esempio 2.7 (figura 5).

Fissiamo $x_2 = 7$ cm, $y_1 = 4$ g e leggiamo sul grafico le intersezioni con la funzione di fit e con l'asse $x = 1$: $x_1 = 0.70 \pm 0.05$ cm, $y_2 = 4000 \pm 500$ g, $y(1) = 15 \pm 5$ g (abbiamo preso come incertezza la metà del passo della griglia). Usando la (22) e la (23), e propagando gli errori, si ottiene

$$\beta = 3.00 \pm 0.15$$

$$C = 15 \pm 5 \text{ g}$$

Notiamo che β è compatibile con il valore atteso 3. Sapendo che $C = \frac{4}{3}\pi\rho$ possiamo ricavare banalmente una stima della densità: $\rho = 3.6 \pm 1.2$ g/cm³, che è compatibile con la densità dell'alluminio.

Gli errori relativi sono tutti piuttosto alti a causa della limitata risoluzione della griglia, ma misurando con un righello avremmo potuto essere significativamente più accurati.

2.4.3 Funzioni esponenziali e scale semilogaritmiche

Si dice *esponenziale* una funzione del tipo

$$y(x) = Ce^{-\lambda x}. \quad (24)$$

Nel caso (piuttosto frequente) in cui la variabile indipendente x rappresenti un tempo, il parametro λ 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

$$\frac{y(\tau)}{y(0)} = \frac{Ce^{-\lambda\tau}}{C} = \frac{Ce^{-\lambda/\lambda}}{C} = \frac{1}{e} \quad (25)$$

del 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

$$T_{1/2} = \tau \ln 2. \quad (26)$$

► **ESEMPIO 2.9.** Arnd Leike dell'Università di Monaco è stato insignito nel 2002 del premio IgNobel per la Fisica per aver dimostrato che il decadimento della schiuma della birra è esponenziale [5].

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

$$\log_{10}(y) = \log_{10}(Ce^{\lambda x}) = \log_{10}(C) + \lambda x \log_{10}(e).$$

Con il cambiamento di variabile

$$\begin{cases} x' = x \log_{10}(e) \\ y' = \log_{10}(y) \end{cases}$$

e ponendo, come prima, $C' = \log_{10}(C)$, la 24 diviene:

$$y' = C' + \lambda x' \quad (27)$$

cioè una retta.

A questo punto non stupirà il fatto che esistono grafici *semilogaritmici* e carte prestampate semilogaritmiche. La carta semilogaritmica è per una legge esponenziale cioè che la carta bilogaritmica è per una legge di potenza.

► **ESEMPIO 2.10.** Si immerge termometro in un misto di acqua e ghiaccio—per definizione a 0° . 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°) del bagno termico

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

(con λ costante caratteristica, determinata dai dettagli del sistema sotto studio). Si registra la temperatura segnata dal termometro ad intervalli di 0.5 s. I risultati sono mostrati, in scala lineare e bilogarithmica, in figura ??.

Nel primo caso (in scala lineare), non è ovvio riconoscere un esponenziale; inoltre è difficile leggere dal grafico i valori di T per tempi grandi, che di nuovo tendono ad essere schiacciati nella parte inferiore del grafico.

Nel grafico in carta semilogarithmica, viceversa, è immediato vedere che il decadimento delle temperatura è *approssimativamente* esponenziale. Si vede anche chiaramente una deviazione dall'andamento rettilineo, che indica che il nostro modello non è perfetto (spesso sono proprio le deviazioni dai modelli che cerchiamo di verificare ad essere interessanti in pratica).

2.4.4 Fit grafici in scala semilogarithmica

La discussione segue sostanzialmente quella della sezione 2.4.2. Presi due punti sulla funzione di fit, il coefficiente angolare si ricava dalla (24) come

$$\lambda = \frac{y'_2 - y'_1}{x'_2 - x'_1} = \frac{\log_{10}(y_2) - \log_{10}(y_1)}{(x_2 - x_1) \log_{10}(e)} = \frac{\log_{10}(y_2/y_1)}{(x_2 - x_1) \log_{10}(e)} \quad (28)$$

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

Un metodo alternativo per stimare graficamente λ è quello di utilizzare la (26): 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 (figura ??).

► **ESEMPIO 2.11.** Stimiamo λ per il fit relativo ai dati dell'esempio 2.10 mostrato in figura ??.
Supponiamo di avere una griglia abbastanza fine per leggere dal grafico $T_{1/2} = 6.0 \pm 0.5$ s. Dalla 26 ricaviamo

$$\lambda = \frac{\ln(2)}{T_{1/2}} = 0.116 \pm 0.010 \text{ s}^{-1}.$$

ELEMENTI DI TEORIA DELLA PROBABILITÀ

In questo capitolo sono passati brevemente in rassegna alcuni fatti fondamentali della teoria della probabilità.

3.1 DEFINIZIONE ASSIOMATICA DELLA PROBABILITÀ

Si deve a Kolmogorov [8] la prima costruzione rigorosa della teoria della probabilità, in una struttura che sostanzialmente sopravvive nei manuali di statistica 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 dallo *spazio campionario* Ω —cioè l'insieme (che per semplicità assumeremo qui numerabile) di tutte le possibili realizzazioni elementari di un dato fenomeno—e da una σ -algebra (chiusa rispetto alle operazioni di unione ed intersezione) \mathcal{F} di sottoinsiemi di Ω —tipicamente l'insieme di tutti i sottoinsiemi di Ω , o insieme di potenza di Ω —i cui elementi si dicono *eventi*.

► **ESEMPIO 3.1.** Nel lancio di un dado a sei facce lo spazio campionario Ω è dato dai sei valori possibili delle uscite $\{1\ 2\ 3\ 4\ 5\ 6\}$. L'insieme di potenza di Ω 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 ?), oltre ovviamente all'insieme vuoto \emptyset ed allo spazio campionario stesso Ω .

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

1. $0 \leq P(E_i) \leq 1 \ \forall E_i \in \mathcal{F}$;
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 (come del resto è ovvio dall'esempio 3.1) che la probabilità è definita su \mathcal{F} e non su Ω . Notiamo anche che il terzo assioma si estende banalmente all'unione numerabile di eventi che non abbiano elementi in comune.

► **ESEMPIO 3.2.** 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. Vedremo nel seguito che se il dado è equo, questa è non solo *una* probabilità, ma anche *la* probabilità.

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 3.2 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* (o *classica*) 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:

$$P(E) = \frac{n}{N} \quad (29)$$

► **ESEMPIO 3.3.** 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 3.4.** Supponiamo di avere un mazzo di 52 carte. La probabilità di estrarre un re è $P(\text{re}) = 4/52 = 1/13$; la probabilità di estrarre una carta di cuori è $P(\text{cuori}) = 13/52 = 1/4$.

► **ESEMPIO 3.5.** 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 $P(3) = 2/36 = 1/18$.

Anche senza entrare nella circolarità della definizione combinatoriale (non è chiaro come si possa richiedere che i casi siano equiprobabili quando ancora non si è definito la probabilità), è ovvio come il suo campo di applicabilità sia limitato ai casi 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 (non è affatto ovvio come si potrebbero definire i casi favorevoli ed i casi possibili). E, purtuttavia, si tratta di una nozione utile che utilizzeremo in seguito.

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 N di ripetizioni dell'esperimento è molto grande:

$$P(E) = \lim_{N \rightarrow \infty} \frac{n}{N} \quad (30)$$

La 30 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.

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

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

ma piuttosto in quello (più debole) della convergenza statistica:

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

Vale anche la pena sottolineare come, benché formalmente simili, la 29 e 30 siano completamente diverse dal punto di vista concettuale.

► **ESEMPIO 3.6.** 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).

3.2.3 Definizione soggettiva

Nella sua definizione *moderna* (o *soggettiva*) la probabilità di un evento E si identifica con la *misura del grado di fiducia che un individuo attribuisce al verificarsi di E* . Sebbene in questo schema il concetto di probabilità sembri perdere qualsiasi contenuto oggettivo, ad un esame più attento la definizione soggettiva di probabilità è tutt'altro che estranea all'esperienza quotidiana: quando scommettiamo (o accettiamo scommesse) con quote prefissate su un evento sportivo, tanto per fare un esempio, stiamo di fatto assegnando delle probabilità ben definite ai possibili esiti dell'evento stesso. (E si tratta di un problema che sarebbe difficile formalizzare nello schema combinatorio o in quello frequentista.)

3.3 LEGGI ELEMENTARI DELLA PROBABILITÀ

Come anticipato in precedenza, 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 Ω , si ha

$$P(\bar{E}) = 1 - P(E). \quad (31)$$

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

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

□

► **ESEMPIO 3.7.** Nel lancio di un dado equo a sei facce la probabilità che l'uscita non sia un 3 è data da

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

COROLLARIO 3.1. Per ogni evento E in \mathcal{F} , si ha $0 \leq P(E) \leq 1$.

COROLLARIO 3.2. $P(\emptyset) = 0$.

COROLLARIO 3.3. Se $E_1 \subset E_2$, allora $P(E_1) \leq P(E_2)$.

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

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2). \quad (32)$$

Dimostrazione. Possiamo scrivere E_2 come la somma di due insiemi disgiunti $E_2 = (E_1 \cap E_2) \cup (\bar{E}_1 \cap E_2)$ per cui

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

D'altra parte $E_1 \cup E_2 = \Omega \cap (E_1 \cup E_2) = (E_1 \cup \bar{E}_1) \cap (E_1 \cup E_2)$, che, in virtù della proprietà distributiva, può essere scritto come $E_1 \cup (\bar{E}_1 \cap E_2)$, da cui

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

Mettendo insieme le due equazioni sopra si ottiene la tesi. (La figura 6 illustra in modo intuitivo il contenuto della dimostrazione.) □

► **ESEMPIO 3.8.** Si consideri di nuovo il nostro mazzo di 52 carte. La probabilità di estrarre un re oppure una carta di cuori di cuori si può calcolare dal teorema di addizione delle probabilità come

$$P(\text{re o cuori}) = P(\text{re}) + P(\text{cuori}) - P(\text{re di cuori}) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}.$$

Se gli eventi E_1 e E_2 sono disgiunti (cioè $E_1 \cap E_2 = \emptyset$) allora la formula di addizione delle probabilità (32) si semplifica in

$$P(E_1 \cup E_2) = P(E_1) + P(E_2). \quad (33)$$

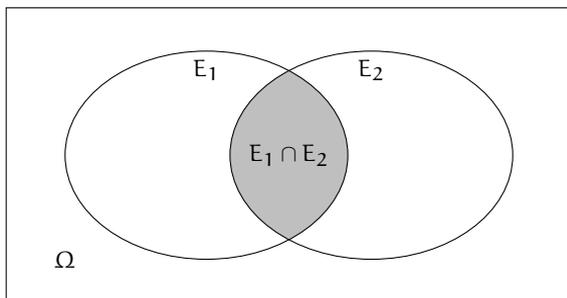


FIGURA 6. Diagramma di Venn per il teorema di addizione delle probabilità 3.2

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

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.

► ESEMPIO 3.9. Nel lancio di un dado equo a 6 facce la probabilità che esca il 2 oppure il 3 è data da

$$P(2 \text{ o } 3) = P(2) + P(3) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

3.4 PROBABILITÀ CONDIZIONATA

Dati due eventi E_1 ed E_2 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

$$P(E_1 | E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} \quad (34)$$

(va da sé che assumiamo che $P(E_2) \neq 0$). Intuitivamente, come illustrato in figura 6, 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 che abbiamo introdotto nella sezione 3.1. Di fatto ogni probabilità può essere vista come una probabilità condizionata notando che

$$P(E) = P(E | \Omega). \quad (35)$$

3.4.1 Eventi indipendenti

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

$$P(E_1 | E_2) = P(E_1). \quad (36)$$

Per la definizione di probabilità condizionata (34) la condizione di indipendenza tra eventi si può riscrivere come

$$P(E_1 \cap E_2) = P(E_1)P(E_2). \quad (37)$$

(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 Ω . In quel caso gli eventi si dicono *incompatibili*.)

► ESEMPIO 3.10. Consideriamo di nuovo il nostro mazzo di 52 carte. Come sappiamo la probabilità di estrarre un re è $P(\text{re}) = 1/13$ e la probabilità di estrarre una carta di cuori è $P(\text{cuori}) = 1/4$. 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 $P(\text{re di cuori}) = 1/52$. Il fatto che

$$P(\text{re})P(\text{cuori}) = \frac{1}{13} \cdot \frac{1}{4} = \frac{1}{52} = P(\text{re di cuori})$$

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

L'esempio precedente può sembrare così banale (in fondo non era ovvio sin dal principio che i due eventi fossero indipendenti?) da non meritare lo spazio che occupa sulla carta. In effetti non serve complicarlo di molto per capire che il concetto di indipendenza tra eventi è più sottile di quanto possa apparire a prima vista.

► **ESEMPIO 3.11.** Aggiungiamo un jolly al nostro mazzo di 52 carte (che diventa così di 53 carte). La probabilità di estrarre un re diviene $P(\text{re}) = 4/53$ mentre quella di estrarre una carta di cuori è adesso $P(\text{cuori}) = 13/53$. La probabilità di estrarre il re di cuori è banalmente $P(\text{re di cuori}) = 1/53$ per cui

$$P(\text{re}) P(\text{cuori}) = \frac{4}{53} \cdot \frac{13}{53} = \frac{52}{2809} \neq P(\text{re di cuori}),$$

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.

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 (34) si riscrive banalmente come

$$P(E_1 \cap E_2) = P(E_1 | E_2) P(E_2) = P(E_2 | E_1) P(E_1) \quad (38)$$

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

► **ESEMPIO 3.12.** Riesaminiamo brevemente l'esempio precedente alla luce di quanto appena detto. Tutto è perfettamente consistente:

$$P(\text{re di cuori}) = \begin{cases} P(\text{re} | \text{cuori}) P(\text{cuori}) = \frac{1}{13} \cdot \frac{13}{53} = \frac{1}{53} \\ P(\text{cuori} | \text{re}) P(\text{re}) = \frac{1}{4} \cdot \frac{4}{53} = \frac{1}{53} \end{cases}$$

3.5 IL TEOREMA DI BAYES: UN SEMPLICE ESEMPIO

La (38) ha come conseguenza immediata e banale il noto teorema di Bayes, che lega tra loro le probabilità condizionate $P(E_1 | E_2)$ e $P(E_2 | E_1)$

$$P(E_1 | E_2) = \frac{P(E_2 | E_1) P(E_1)}{P(E_2)}. \quad (39)$$

Nel seguito non faremo uso del teorema di Bayes, ma per completezza ne illustriamo qui il contenuto con un esempio.

Supponiamo di aver acquistato un rilevatore automatico di banconote false. Le specifiche fornite dalla casa produttrice dicono che l'oggetto ha un'efficienza nel segnalare banconote contraffatte del 99% ed un tasso di falsi positivi del 5%—cioè nel 5% dei casi (in media) segnala come contraffatta una banconota autentica. 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 99% o 95% (come uno potrebbe legittimamente 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.

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

ok: il dispositivo segnala la banconota come autentica

ko: il dispositivo segnala la banconota come contraffatta

Notiamo che questi eventi sono a due a due disgiunti e partizionano il nostro spazio Ω , che possiamo vedere come il prodotto cartesiano dei due insiemi di due elementi $\{A, C\}$ e $\{ok, ko\}$

$$\Omega = \{(A, ok), (A, ko), (C, ok), (C, ko)\},$$

in due modi differenti. Le specifiche del costruttore si possono allora riscrivere (e precisare) come

$$P(ko|C) = 0.99$$

$$P(ko|A) = 0.05$$

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).

Detto questo, siamo interessati alla probabilità che una banconota segnalata come contraffatta dal nostro rilevatore sia effettivamente contraffatta—che per la (39) possiamo scrivere come

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

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

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

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)

$$P(C|ko) = \frac{P(ko|C)P(C)}{P(ko|A)P(A) + P(ko|C)P(C)} = \frac{0.99 \times 0.001}{0.05 \times 0.999 + 0.99 \times 0.001} \approx 1.9\%$$

Sorprendentemente la probabilità cercata è dell'ordine dell'2% nonostante l'efficienza del nostro rilevatore sia prossima all'unità 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 $P(C|ko) = 0$ come anticipato.

(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?)

 VARIABILI CASUALI E FUNZIONI DI DISTRIBUZIONE

4.1 VARIABILI CASUALI

4.2 CARATTERISTICHE COMUNI ALLE DISTRIBUZIONI

4.2.1 Valore di aspettazione

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

$$E[f(x)] = \begin{cases} \sum f(x_k)P(x_k) & \text{per variabili discrete} \\ \int_{-\infty}^{\infty} f(x)p(x) dx & \text{per variabili continue,} \end{cases} \quad (40)$$

(in cui la sommatoria è estesa a tutti i valori x_k che la variabile x può assumere). Per la linearità di integrale e sommatoria, il valore di aspettazione è un *operatore lineare*, nel senso che

$$E[c_1 f(x) + c_2 g(x)] = c_1 E[f(x)] + c_2 E[g(x)], \quad (41)$$

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

$$E[c] = c. \quad (42)$$

4.2.2 Misure di tendenza centrale

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.

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

$$\mu = E[x] = \begin{cases} \sum x_k P(x_k) & \text{per variabili discrete} \\ \int_{-\infty}^{\infty} xp(x) dx & \text{per variabili continue.} \end{cases} \quad (43)$$

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

$$E[cx] = cE[x]. \quad (44)$$

Nel caso particolare di una variabile casuale che 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 (43) 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 (43) a seconda della probabilità corrispondente di modo che la media si sposti verso i valori più probabili.

► **ESEMPIO 4.1.** 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 4.2.** 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:

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

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

$$P(x \leq \mu_{1/2}) = P(x \geq \mu_{1/2}). \tag{45}$$

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.

► **ESEMPIO 4.3.** Nel caso del dado a sei facce qualsiasi numero nell'intervallo aperto $]3, 4[$ soddisfa la (45), per cui è necessario fissare una convenzione per definire la mediana (possiamo ad esempio prendere il valore centrale $7/2$ di tale intervallo).

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.

4.2.3 Misure di dispersione attorno alla media

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

$$\sigma^2 = E[(x - \mu)^2] = \begin{cases} \sum (x_k - \mu)^2 P(x_k) & \text{per variabili discrete} \\ \int_{-\infty}^{\infty} (x - \mu)^2 p(x) dx & \text{per variabili continue} \end{cases} \tag{46}$$

(nel seguito indicheremo la varianza di x con σ^2 , σ_x^2 o $\text{Var}(x)$ a seconda del contesto). Notiamo che se c è una costante

$$\text{Var}(cx) = E[(cx - E[cx])^2] = E[(cx - c\mu)^2] = E[c^2(x - \mu)^2] = c^2 E[(x - \mu)^2] = c^2 \text{Var}(x). \tag{47}$$

Definiamo inoltre la *deviazione standard* come radice quadrata della varianza

$$\sigma = \sqrt{\sigma^2}. \tag{48}$$

¹ 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.

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.

► **ESEMPIO 4.4.** Torniamo al nostro dado a sei facce. Utilizzando il valor medio calcolato nell'esempio 4.1 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 4.5.** Consideriamo la variabile casuale x distribuita uniformemente tra 0 ed 1 introdotta nell'esempio 4.2. La varianza si calcola come

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

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

Partendo dalla definizione di varianza (46), attraverso una semplice manipolazione algebrica

$$\sigma^2 = E[(x - \mu)^2] = E[x^2 - 2x\mu + \mu^2] = E[x^2] - E[2x\mu] + E[\mu^2] = E[x^2] - 2\mu E[x] + \mu^2 = E[x^2] - 2\mu^2 + \mu^2$$

si arriva relazione utile

$$\sigma^2 = E[x^2] - \mu^2. \quad (49)$$

Un concetto utile che si applica alle funzioni di distribuzione di variabile continua—ed in particolare a quelle unimodali—è quello di *semilarghezza a metà altezza*. Come mostrato in figura ?? 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

$$\text{FWHM} = |x_b - x_a|, \quad (50)$$

e la semilarghezza a metà altezza HWHM (*Half Width at Half Maximum* in Inglese) come

$$\text{HWHM} = \frac{|x_b - x_a|}{2}. \quad (51)$$

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

$$\text{HWHM} = c\sigma \quad (52)$$

con c dell'ordine dell'unità. Questo offre un'ulteriore interpretazione geometrica del concetto di deviazione standard come *larghezza* di una distribuzione.

4.2.4 La disuguaglianza di Chebyshev

Se misuriamo del deviazioni dal valor medio in unità di σ è *poco probabile* che esse siano molto grandi (da cui il nome di deviazione standard), come precisato dal teorema di seguito.

TEOREMA 4.1 (DI CHEBYSHEV). Sia x una variabile casuale tale che esistano finiti la media μ e la varianza σ^2 ; preso un numero positivo k si ha:

$$P(|x - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad (53)$$

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

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

da cui segue banalmente la tesi. □

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.

► **ESEMPIO 4.6.** 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 $P(|x - \mu| \geq 2\sigma) \leq 1/4$.

4.2.5 Momenti di ordine superiore

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

$$\mathcal{M}_n(x_0) = E[(x - x_0)^n] = \begin{cases} \sum (x_k - x_0)^n P(x_k) & \text{per variabili discrete} \\ \int_{-\infty}^{\infty} (x - x_0)^n p(x) dx & \text{per variabili continue.} \end{cases} \quad (54)$$

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

$$\lambda_n = \mathcal{M}_n(0) \quad (55)$$

ed i *momenti centrali* μ_n , ossia i momenti di ordine generico attorno al valor medio μ di x :

$$\mu_n = \mathcal{M}_n(\mu) \quad (56)$$

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).

La media e la varianza non sono le uniche quantità interessanti che si possono costruire con in momenti. A partire dal momento centrale di ordine 3 si definisce il *coefficiente di asimmetria* o *skewness* γ_1

$$\gamma_1 = \frac{\mu_3}{\sigma^2}. \quad (57)$$

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.

A partire dal momento centrale di ordine 4 si definisce la *curtosi* γ_2

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 \quad (58)$$

(vederemo che il -3 ha l'unico scopo di rendere la curtosi nulla per una distribuzione normale).

4.2.6 La distribuzione cumulativa

DISTRIBUZIONI DI VARIABILE DISCRETA

5.1 LA DISTRIBUZIONE BINOMIALE

Consideriamo un esperimento che abbia esattamente due esiti distinti possibili, E_1 ed E_2 , con probabilità

$$\begin{aligned} P(E_1) &= p \\ P(E_2) &= 1 - p. \end{aligned}$$

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 (37), che

$$P\left(\overbrace{E_1 E_1 E_1 E_1 E_1 \dots}^{k \text{ volte}} \overbrace{E_2 E_2 E_2 E_2 E_2 \dots}^{n-k \text{ volte}}\right) = p^k (1-p)^{n-k}.$$

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

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \quad (59)$$

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

$$\mathcal{B}(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k} \quad (60)$$

La (60) prende il nome di distribuzione binomiale. Notiamo esplicitamente che k è la variabile casuale, mentre n e p sono parametri fissati.

► **ESEMPIO 5.1.** Lanciamo una moneta per 4 volte. Le sequenze possibili sono complessivamente $2^4 = 16$ e sono qui di seguito elencate: TTTT, CT TT, 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

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{24}{2 \cdot 2} = 6.$$

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 (60).

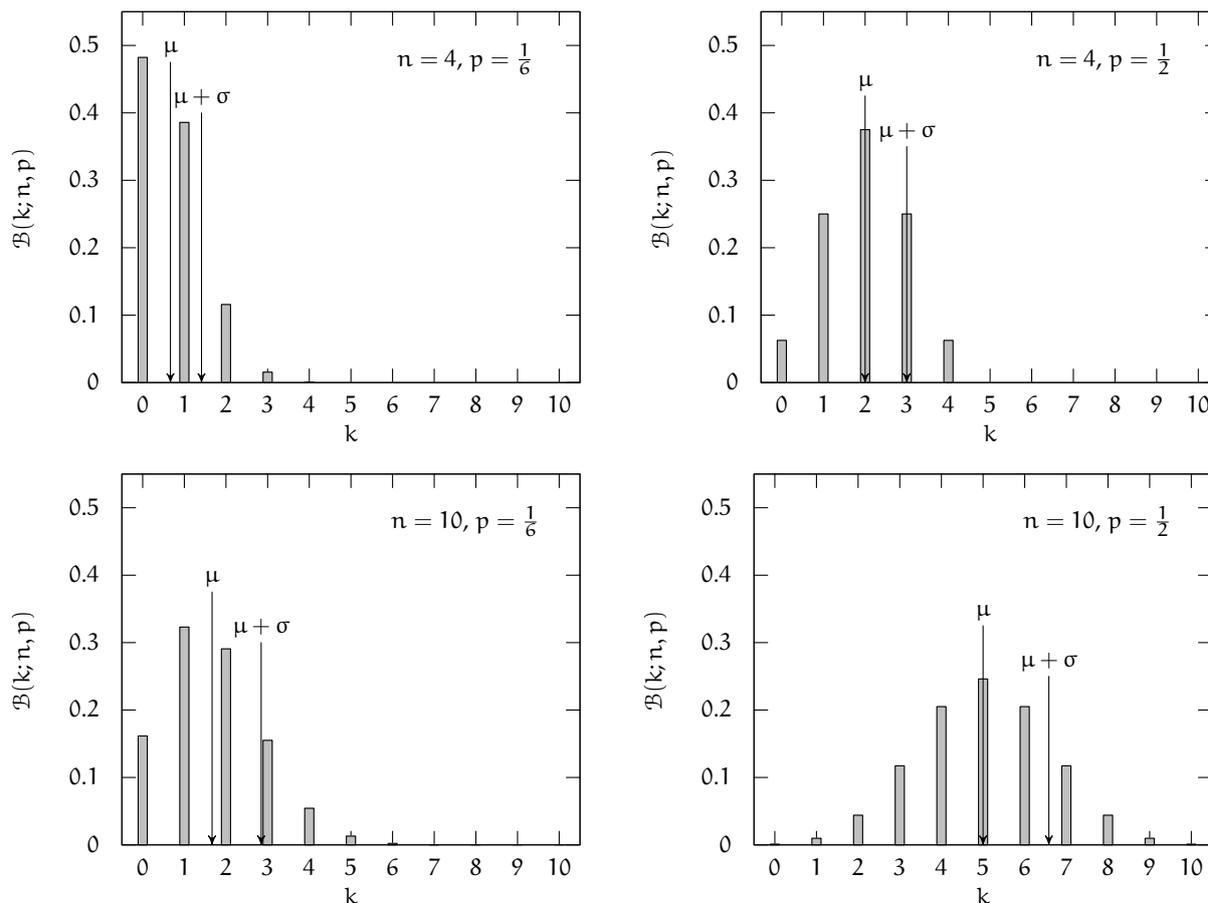


FIGURA 7. Esempi di distribuzione binomiale (60) per diversi valori di n e k . Come discuteremo estensivamente nel seguito, all'aumentare della media $\mu = np$, la distribuzione tende progressivamente ad assumere una caratteristica forma a campana.

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 (60), 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 (43):

$$\mu = E[k] = \sum_{k=0}^n k \mathcal{B}(k; n, p) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

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:

$$\mu = np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} = np \sum_{h=0}^m \frac{m!}{h!(m-h)!} p^k (1-p)^{m-h} = np \sum_{h=0}^m \binom{m}{h} p^k (1-p)^{m-h}$$

(abbiamo posto $h = k - 1$ e $m = n - 1$). Nell'ultimo passaggio riconosciamo la condizione di normalizzazione, per cui

$$\mu = np. \quad (61)$$

Per stimare la varianza utilizziamo la (49). 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:

$$E[k^2] = \sum_{k=1}^n k^2 \frac{np(n-1)!}{k(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} = np \sum_{k=1}^n k \frac{np(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k}$$

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 (49)

$$\sigma^2 = np(1-p). \quad (62)$$

5.2 LA DISTRIBUZIONE DI POISSON

La distribuzione di Poisson è una distribuzione di variabile discreta dipendente da un solo parametro μ —che, come vedremo tra un attimo, ha il significato della media nel senso della (43):

$$\mathcal{P}(k; \mu) = \frac{\mu^k}{k!} e^{-\mu}. \quad (63)$$

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 $p = \mu/n$, col che 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 ξ si discosti da 0 di una quantità molto più grande di σ .) Se questo è vero, allora $\xi \rightarrow 0$ come $1/\sqrt{n}$ per $n \rightarrow \infty$. Per la definizione di ξ si ha $k = n(p + \xi)$, il che ci permette di riscrivere il nostro limite come

$$\lim_{n \rightarrow \infty} \mathcal{B}(k; n, p) = \lim_{n \rightarrow \infty} \frac{n^k (1 - 1/n) \cdots (1 - p - \xi + 1/n)}{n^k} \frac{\mu^k}{k!} \left(1 - \frac{\mu}{n}\right)^{n(1-p-\xi)} = \lim_{n \rightarrow \infty} \frac{\mu^k}{k!} \left(1 - \frac{\mu}{n}\right)^n = \frac{\mu^k}{k!} e^{-\mu},$$

che è proprio la (63) (nel penultimo passaggio abbiamo sfruttato il fatto che $p \rightarrow 0$ e $\xi \rightarrow 0$).

► **ESEMPIO 5.2.** 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:

$$P = \mathcal{B}(10; 500, 3/216) = \binom{500}{10} \times (3/216)^{10} \times (213/216)^{490} \simeq 0.06933.$$

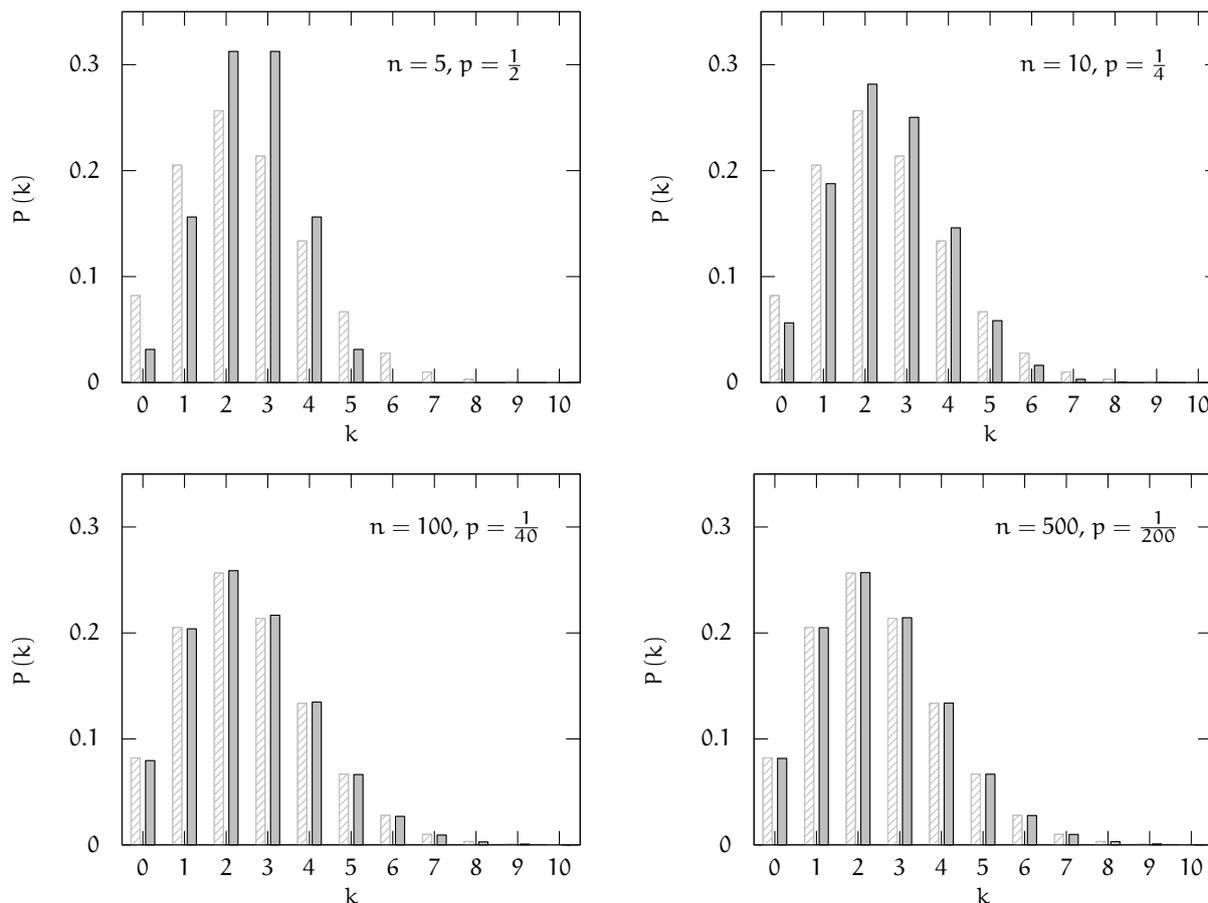


FIGURA 8. Esempi di distribuzione binomiale per diversi valori di n e p (con np costante). Il grafico ombreggiato in grigio (identico nei quattro pannelli) rappresenta una distribuzione di Poisson con media $\mu = 5/2$. Al crescere di n le due distribuzioni divengono via via più simili.

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:

$$P = \mathcal{P}(10; 125/18) = \frac{6.944^{10}}{10!} \times e^{-6.944} \simeq 0.06928.$$

5.2.1 Stazionarietà e processi Poissoniani

Fortunatamente l'utilità della distribuzione di Poisson non è limitata all'esempio appena illustrato. La seguente derivazione alternativa mette in luce in modo più trasparente il contenuto fisico ed il campo di applicabilità della statistica Poissoniana.

Supponiamo, per fissare le idee, di essere interessati ad un generico evento E caratterizzato da un tasso di probabilità per unità di tempo λ costante nel tempo. Questo significa essenzialmente che la probabilità (infinitesima) dP che l'evento avvenga in un intervallo (infinitesimo) dt è λdt —e questa probabilità non varia con il tempo. Notiamo, per inciso, che quest'ultima assunzione (cioè la *stazionarietà* del processo) è tutt'altro che banale e costituisce il prerequisito fondamentale di applicabilità di ciò che stiamo per discutere. Ci chiediamo quale sia la probabilità $P(k; t)$ che il nostro evento avvenga k volte in un intervallo di tempo finito t .

Per cominciare dividiamo il nostro intervallo finito t in n intervallini di durata $dt = t/n$ (faremo tendere $n \rightarrow \infty$ alla fine). In ogni intervallino la probabilità che l'evento *non* si verifichi è $1 - \lambda t/n$. Se assumiamo che ciò che avviene in un generico intervallino non influenzi ciò che avviene negli altri, possiamo utilizzare la formula di

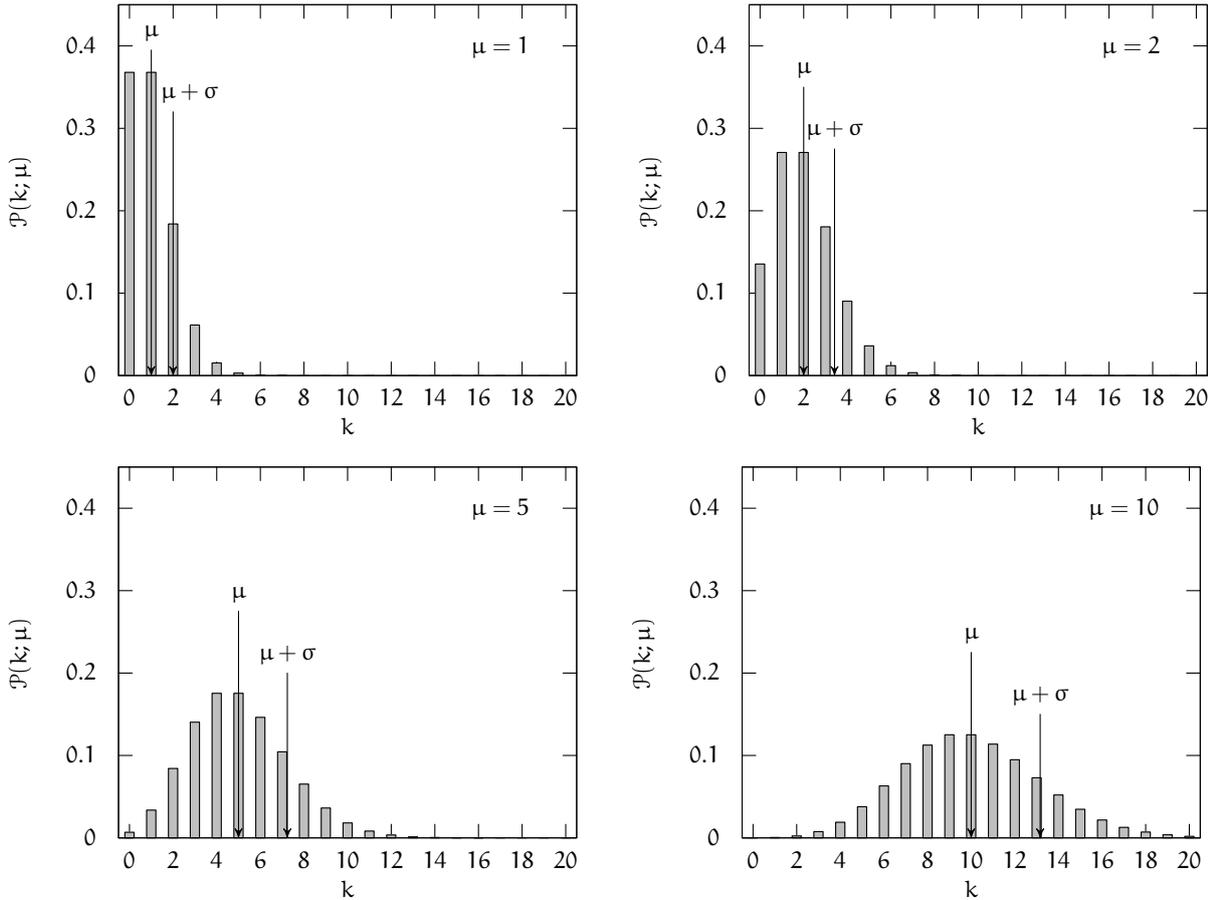


FIGURA 9. Esempi di distribuzione di Poisson (63) per diversi valori di μ . Come discuteremo estensivamente nel seguito, all'aumentare di μ , la distribuzione tende progressivamente ad assumere una caratteristica forma a campana.

moltiplicazione delle probabilità per eventi indipendenti e la probabilità che *non* si verifichino eventi in un tempo finito t è data da

$$P(0;t) = \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda t}{n}\right)^n = e^{-\lambda t}.$$

Se n è abbastanza grande (o, il che è lo stesso, dt abbastanza piccolo) la probabilità che si verifichino due o più eventi in un intervallino dt è trascurabile, e la probabilità che si verifichi esattamente un evento in un tempo finito t è data da

$$P(1;t) = \lim_{n \rightarrow \infty} n \lambda dt \left(1 - \frac{\lambda t}{n}\right)^{n-1} = \lambda t e^{-\lambda t}$$

(essenzialmente stiamo richiedendo che non si verifichino eventi in $n - 1$ intervallini e che si verifichi un evento nell'intervallino rimanente, che siamo liberi di scegliere a caso tra gli n disponibili). Il lettore più accorto avrà capito che stiamo di fatto di fatto riderivando la distribuzione binomiale nel limite in cui la probabilità del singolo evento (in questo caso λdt) è piccola ed il numero di eventi possibili (in questo caso n) è grande—cioè proprio nelle ipotesi in cui essa tende ad una Poissoniana. La probabilità cercata sarà dunque

$$P(k;t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad (64)$$

che coincide con la (63) modulo il cambiamento di variabile $\mu = \lambda t$.

Equivalentemente possiamo tentare di scrivere un'equazione differenziale (ricorsiva) per $P(k;t)$ notando che ci sono solo due modi (le cui probabilità si sommano) perché si verifichino k eventi in un intervallo di tempo $t + dt$: k eventi nella parte iniziale t e 0 nell'ultimo intervallo dt o $k - 1$ eventi nella parte iniziale t e 1 nell'ultimo intervallo dt . In formule

$$P(k;t + dt) = (1 - \lambda dt)P(k;t) + \lambda dt P(k - 1;t).$$

Possiamo riscrivere la relazione qui sopra come

$$\frac{P(k; t + dt) - P(k; t)}{dt} + \lambda P(k; t) - \lambda P(k - 1; t) = 0,$$

da cui, prendendo il limite $dt \rightarrow 0$ si ottiene infine

$$\frac{dP(k; t)}{dt} + \lambda P(k; t) - \lambda P(k - 1; t) = 0. \quad (65)$$

Non è difficile sostituire la (64) nella (65) e verificare che essa è effettivamente una soluzione.

Possiamo riassumere il contenuto di questa sezione dicendo che *nei processi stazionari la probabilità che si verifichino k eventi quando in media se ne verificano μ è data proprio da una distribuzione di Poisson con media μ* . È importante sottolineare che per stazionarietà qui non intendiamo solamente stazionarietà nel tempo; la statistica Poissoniana descrive altrettanto bene, ad esempio, processi stazionari in una dimensione spaziale, sulla superficie o nel volume. Esempi banali di variabili casuali che seguono, più o meno fedelmente, la statistica di Poisson sono: il numero di incidenti stradali in un intervallo di tempo fissato; il numero degli errori di battitura nelle pagine di un libro; il numero di gocce di cioccolata nei biscotti; il numero di decadimenti radiattivi per unità di tempo; il numero di interazioni che una particella elementare subisce attraversando uno strato di materiale. In breve: tutti i processi in cui si ha a che vedere con conteggi in regime stazionario.

► **ESEMPIO 5.3.** Apparentemente uno degli esempi classici di applicazione della distribuzione di Poisson (e quello che ha attratto l'attenzione iniziale sulla distribuzione 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 [1] come il numero di soldati uccisi da un incidente a cavallo in un anno in ciascuno dei reparti della cavalleria stessa obbedisse la distribuzione di Poisson. Il campione di dati includeva 10 reparti osservati in un arco di tempo di 20 anni, per un totale di 200 osservazioni (con un numero totale di morti pari a 122 ed un numero medio di morti per reparto per anno di $122/200 = 0.61$) e si può riassumere nella tabella di seguito,

Numero di morti (k)	o_k	$\mathcal{P}(k; 0.61)$	e_k
0	109	0.5434	108.7
1	65	0.3314	66.3
2	22	0.1011	20.2
3	3	0.0206	4.1
4	1	0.0063	1.3

in cui o_k sono le occorrenze osservate per ciascun valore di k (e assommano a 200) ed $e_k = 200 \times \mathcal{P}(k; 0.61)$ sono le occorrenze attese nel caso in cui il processo sia effettivamente Poissoniano. Non abbiamo ancora gli strumenti per precisare il senso di questa affermazione, ma l'accordo tra o_k ed e_k è sicuramente degno di nota.

► **ESEMPIO 5.4.** 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 *inoocua* (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:

$$P(k \geq 5) = 1 - P(k < 5) = 1 - \sum_{k=0}^4 \mathcal{P}(k; m) = \sum_{k=0}^4 \frac{m^k}{k!} e^{-m} \approx 2.6 \times 10^{-6}.$$

La probabilità è abbastanza piccola (almeno sotto le ipotesi iniziali) da legittimarci ad un reclamo ufficiale.

5.2.2 Normalizzazione, media e varianza

Si verifica banalmente che la distribuzione di Poisson, così come scritta nella (63) 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 si calcola formalmente secondo la (43)

$$E[k] = \sum_{k=0}^{\infty} k \mathcal{P}(k; \mu) = \sum_{k=0}^{\infty} k \frac{\mu^k}{k!} e^{-\mu} = e^{-\mu} \sum_{k=0}^{\infty} k \frac{\mu^k}{k!}.$$

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 μ ha proprio il significato della media nel senso della definizione (43).

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 (49)

$$\sigma^2 = \mu. \quad (66)$$

5.2.3 Somma di variabili Poissoniane

Consideriamo due variabili Poissoniane (indipendenti) l ed m con medie μ_l e μ_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:

$$P(k) = \sum_{l=0}^k \mathcal{P}(l; \mu_l) \mathcal{P}(k-l; \mu_m) \quad (67)$$

(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 (??), per cui:

$$P(k) = \frac{(\mu_l + \mu_m)^k}{k!} e^{-(\mu_l + \mu_m)} = \mathcal{P}(k; \mu_l + \mu_m), \quad (68)$$

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 5.5.** Torniamo adesso all'esempio 5.4 e chiediamoci qual è la probabilità di trovare nella cassetta della posta 5 lettere nell'ipotesi in cui postino consegnano 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:

$$P(k \geq 5) = 1 - P(k < 5) = 1 - \sum_{k=0}^4 \mathcal{P}(k; m_s) = \sum_{k=0}^4 \frac{m_s^k}{k!} e^{-m_s} \approx 1.6\%,$$

che è piccola ma decisamente non piccola come prima.

5.2.4 *La distribuzione di Poisson composta*

DISTRIBUZIONE DI VARIABILE CONTINUA

6.1 LA DISTRIBUZIONE UNIFORME

È il più semplice esempio di funzione di distribuzione di variabile casuale:

$$u(x; a, b) = \begin{cases} \frac{1}{(b-a)} & a \leq x \leq b \\ 0 & x < a; x > b \end{cases} \quad (69)$$

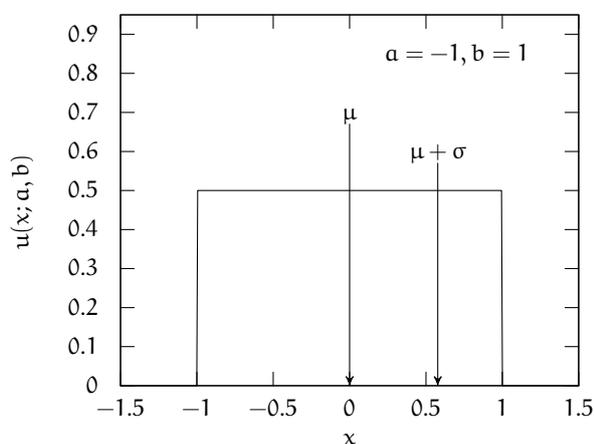


FIGURA 10. Esempio di distribuzione uniforme per $a = -1$ e $b = 1$.

6.1.1 Normalizzazione, media e varianza

La distribuzione, così come è scritta nella (69), è correttamente normalizzata:

$$\int_{-\infty}^{\infty} u(x; a, b) dx = \int_a^b \frac{1}{(b-a)} dx = \frac{1}{(b-a)} \int_a^b dx = \frac{1}{(b-a)} (b-a) = 1$$

La media è data da:

$$\mu = E[x] = \int_{-\infty}^{\infty} x u(x; a, b) dx = \frac{1}{(b-a)} \int_a^b x dx = \frac{1}{(b-a)} \frac{x^2}{2} \Big|_a^b = \frac{1}{(b-a)} \frac{(b^2 - a^2)}{2} = \frac{(b+a)}{2}$$

cioè coincide, come è naturale aspettarsi, con il valor medio dell'intervallo su cui la densità di probabilità è non nulla:

$$\mu = \frac{(b+a)}{2} \quad (70)$$

Il valore di aspettazione di x^2 è dato da:

$$\begin{aligned} E[x^2] &= \int_{-\infty}^{\infty} x^2 u(x; a, b) dx = \frac{1}{(b-a)} \int_a^b x^2 dx = \frac{1}{(b-a)} \frac{x^3}{3} \Big|_a^b = \frac{1}{(b-a)} \frac{(b^3 - a^3)}{3} = \\ &= \frac{1}{(b-a)} \frac{(b-a)(b^2 + ab + a^2)}{3} = \frac{(b^2 + ab + a^2)}{3} \end{aligned}$$

e la varianza, al solito:

$$\sigma^2 = E[x^2] - \mu^2 = \frac{(b^2 + ab + a^2)}{3} - \frac{(b + a)^2}{4} = \frac{(4b^2 + 4ab + 4a^2 - 3b^2 - 6ab - 3a^2)}{12} = \frac{(b^2 - 2ab + a^2)}{12},$$

cioè

$$\sigma^2 = \frac{(b - a)^2}{12}. \tag{71}$$

► **ESEMPIO 6.1.** 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:

$$m = 58.0 \pm 0.3 \text{ g}$$

6.2 LA DISTRIBUZIONE ESPONENZIALE

Una variabile casuale la cui funzione di distribuzione sia data da:

$$\varepsilon(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & 0 \leq x < \infty \\ 0 & x < 0 \end{cases} \tag{72}$$

(con $\lambda \geq 0$) viene chiamata variabile casuale esponenziale con parametro λ (oppure anche esponenzialmente distribuita con parametro λ). Per ragioni dimensionali, il parametro λ ha le dimensioni di x^{-1} .

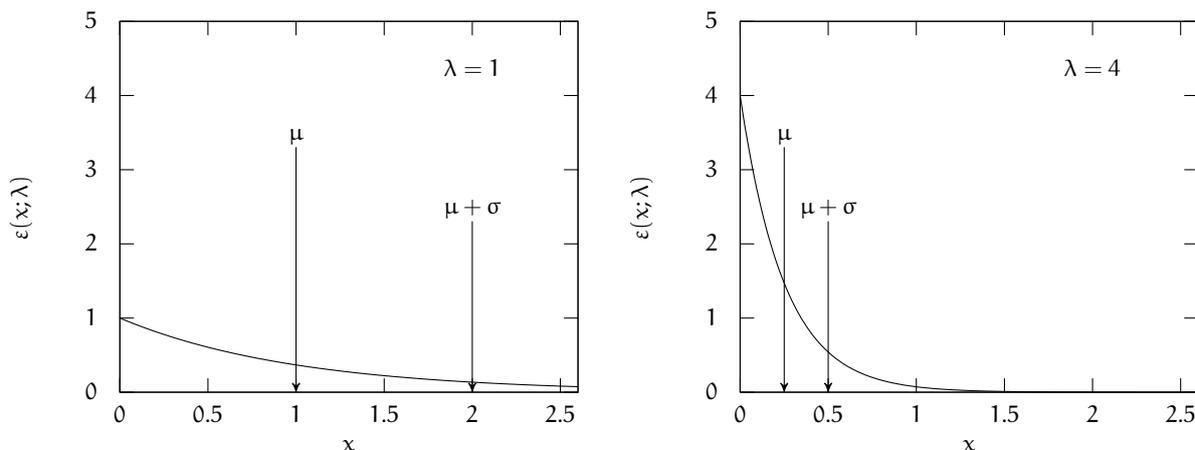


FIGURA 11. Esempi di distribuzione esponenziale (72) per due diversi valori di λ (1 e 4). Entrambe le distribuzioni sono correttamente normalizzate.

Un risultato particolarmente interessante (che non dimostreremo qui) è costituito dal fatto che la funzione di distribuzione esponenziale è la funzione di distribuzione dell'intervallo (di solito temporale) che intercorre tra due conteggi successivi in processi Poissoniani. Tipici esempi di variabili casuali distribuite esponenzialmente sono: il tempo dopo il quale un nucleo radioattivo decade (in questo caso l'inverso del parametro λ prende il nome di *vita media* del nucleo), il tempo che intercorre tra due successive chiamate di telefono sbagliate, ma anche la distanza che una particella percorre in un mezzo omogeneo prima di interagire con il mezzo stesso (in questo caso l'inverso del parametro λ prende il nome di *lunghezza di interazione* o *cammino libero medio*).

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

$$P(x \geq x_1 + x_2) = P(x \geq x_1) P(x \geq x_2) \tag{73}$$

Una variabile casuale che goda di questa proprietà si dice una variabile *senza memoria* (o anche *memory-less*). La condizione di assenza di memoria si può riscrivere banalmente come

$$P(x \geq x_1 + x_2 | x \geq x_1) = P(x \geq x_2). \quad (74)$$

► **ESEMPIO 6.2.** 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 6.3.** Supponiamo di *osservare*, a partire da un istante t_0 , un nucleo radioattivo con vita media (inverso del parametro λ) 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!

6.2.1 Normalizzazione, media e varianza

La distribuzione, al solito, è 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 vale

$$\mu = E[x] = \int_0^{\infty} x \varepsilon(x; \lambda) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx.$$

Con il cambiamento di variabile $\lambda x = r$, l'integrale può essere calcolato per parti:

$$\mu = \frac{1}{\lambda} \int_0^{\infty} r e^{-r} dr = \frac{1}{\lambda} \left(-r e^{-r} \Big|_0^{\infty} + \int_0^{\infty} e^{-r} dr \right) = \frac{1}{\lambda} - e^{-r} \Big|_0^{\infty} = \frac{1}{\lambda}. \quad (75)$$

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

$$\sigma^2 = E[x^2] - \mu^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \quad (76)$$

6.3 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, sotto opportune condizioni, della distribuzione binomiale e di quella di Poisson. 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.

La funzione di distribuzione di Gauss (o distribuzione normale) si scrive come:

$$N(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{77}$$

dove μ e σ , come vedremo tra un attimo, sono proprio la media e la deviazione standard della distribuzione nel senso delle (43) e (46). Una variabile casuale Gaussiana con media 0 e varianza 1

$$N(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{78}$$

si dice variabile Gaussiana in forma standard.

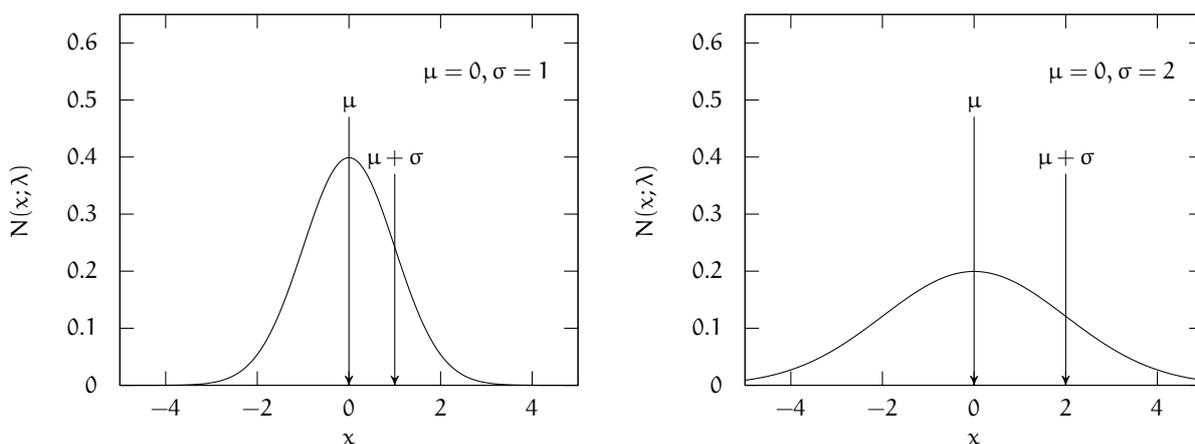


FIGURA 12. Esempi di distribuzione di Gauss (77) per $\mu = 0$ e due diversi valori di σ (1 e 2). Entrambe le distribuzioni sono correttamente normalizzate.

6.3.1 Normalizzazione, media e varianza

Verifichiamo che la distribuzione di Gauss, scritta nella forma (77), è correttamente normalizzata, e che—come accennato poco fa— μ e σ rappresentano la media e la deviazione standard della distribuzione. Scriviamo, al solito

$$\int_{-\infty}^{\infty} N(x; \mu, \sigma) dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz := \frac{1}{\sqrt{2\pi}} I_0$$

(abbiamo operato il cambio di variabile $z = (x - \mu)/\sigma$). Sfortunatamente questo integrale non ha espressione analitica, nel senso che non esiste una primitiva dell’integrando esprimibile in forma chiusa in termini di funzioni elementari¹. 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

$$\begin{aligned} I_0^2 &= \left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx \right) \times \left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy \right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)} dx dy = \int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}r^2} r dr d\theta = \\ &= 2\pi \int_0^{\infty} e^{-\frac{1}{2}r^2} d\left(\frac{r^2}{2}\right) = -2\pi e^{-\frac{1}{2}r^2} \Big|_0^{\infty} = 2\pi, \end{aligned}$$

¹ 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.

che dimostra che la distribuzione 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

$$I_n = \int_{-\infty}^{\infty} x^n e^{-\frac{1}{2}x^2} dx.$$

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 l'interessante relazione per ricorrenza

$$I_n = \int_{-\infty}^{\infty} x^n e^{-\frac{1}{2}x^2} dx = - \int_{-\infty}^{\infty} x^{n-1} d\left(e^{-\frac{1}{2}x^2}\right) = -x^{n-1} e^{-\frac{1}{2}x^2} \Big|_{-\infty}^{\infty} + (n-1) \int_{-\infty}^{\infty} x^{n-2} e^{-\frac{1}{2}x^2} dx,$$

ovvero

$$I_0 = \sqrt{2\pi} \quad (79)$$

$$I_n = (n-1) I_{n-2} \quad (80)$$

La media della distribuzione, formalmente, si può allora scrivere, operando lo stesso cambiamento di variabile di prima, come

$$E[x] = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + \mu) e^{-\frac{1}{2}z^2} dz = \frac{1}{\sqrt{2\pi}} (\sigma I_1 + \mu I_0) = \mu,$$

e la varianza come

$$E[(x-\mu)^2] = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 z^2 e^{-\frac{1}{2}z^2} dz = \frac{1}{\sqrt{2\pi}} \sigma^2 I_2 = \sigma^2.$$

Più in generale si ha che i momenti di ordine superiore (pari) attorno alla media sono dati da

$$E[(x-\mu)^n] = \frac{1}{\sqrt{2\pi}} \sigma^n I_n. \quad (81)$$

6.3.2 L'integrale normale degli errori

6.3.3 La distribuzione di Gauss come limite della Poissoniana

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). In entrambi i casi la dimostrazione non presenta difficoltà concettuali di rilievo—si tratta di utilizzare la formula di Stirling per lo sviluppo del fattoriale e calcolare il limite appropriato (e in entrambi i casi è comodo fare il limite nel logaritmo della quantità).

Consideriamo dunque una variabile casuale Poissoniana k e calcoliamo il logaritmo naturale della sua funzione di distribuzione

$$\ln \mathcal{P}(k; \mu) = \ln \left(\frac{\mu^k}{k!} e^{-\mu} \right) = k \ln \mu - \ln(k!) - \mu.$$

Su μ è molto grande (e quindi k è molto grande, visto che $\mu = E[k]$) possiamo utilizzare la formula di Stirling per il fattoriale e riscrivere la nostra espressione come

$$\ln \mathcal{P}(k; \mu) \approx k \ln \mu - \ln(\sqrt{2\pi k}) + k - k \ln k - \mu.$$

L'idea è adesso quella di sviluppare questa espressione in serie di Taylor intorno a μ . Per rendere l'argomento sufficientemente rigoroso introduciamo la variabile casuale *ridotta*

$$\delta = \frac{k - \mu}{\mu}$$

che, per inciso, può essere invertita per dare $k = \mu(1 + \delta)$. Le fluttuazioni di k attorno a 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 allora scrivere

$$\ln \mathcal{P}(k; \mu) \approx \mu(1 + \delta) \ln \mu - \ln \left(\sqrt{2\pi\mu(1 + \delta)} \right) + \mu(1 + \delta) - \mu(1 + \delta) \ln(\mu(1 + \delta)) - \mu.$$

A questo punto possiamo sviluppare in serie attorno al valore $\delta = 0$: il calcolo è un po' tedioso, ma le derivate rilevanti sono

$$\begin{aligned} \left. \frac{d \ln \mathcal{P}(\delta; \mu)}{d\delta} \right|_{\delta=0} &= -\mu \ln(1+\delta) - \frac{1}{2(1+\delta)} \Big|_{\delta=0} = -\frac{1}{2} \\ \left. \frac{d^2 \ln \mathcal{P}(\delta; \mu)}{d\delta^2} \right|_{\delta=0} &= -\frac{\mu}{(1+\delta)} + \frac{1}{2(1+\delta)^2} \Big|_{\delta=0} = -\mu, \end{aligned}$$

da cui

$$\ln \mathcal{P}(k; \mu) \approx \ln \mathcal{P}(0; \mu) + \left. \frac{d \ln \mathcal{P}(\delta; \mu)}{d\delta} \right|_{\delta=0} \frac{1}{2} + \left. \frac{d^2 \ln \mathcal{P}(\delta; \mu)}{d\delta^2} \right|_{\delta=0} \frac{\delta^2}{2} = -\ln(\sqrt{2\pi\mu}) - \frac{1}{2}\delta - \frac{1}{2}\mu\delta^2.$$

Per $\delta \rightarrow 0$ il secondo termine si annulla ed il terzo è quello rilevante—nel valutare la forma indeterminata $\mu\delta^2$ ricordiamo che $\delta \rightarrow 0$ come $1/\sqrt{\pi}$. Si ha allora

$$\ln \mathcal{P}(k; \mu) \approx -\ln(\sqrt{2\pi\mu}) - \frac{1}{2}\mu\delta^2 = -\ln(\sqrt{2\pi\mu}) - \frac{(k-\mu)^2}{2\mu}, \quad (82)$$

da cui infine

$$\mathcal{P}(k; \mu) \approx \frac{1}{\sqrt{2\pi\mu}} e^{-\frac{(k-\mu)^2}{2\mu}}. \quad (83)$$

che è esattamente l'espressione per una Gaussiana con media μ e deviazione standard $\sqrt{\mu}$. Dunque la distribuzione di Poisson tende ad una Gaussiana per $\mu \rightarrow \infty$. (Notiamo, per completezza, che per grandi valori di k possiamo approssimare la nostra variabile casuale, che è intrinsecamente discreta, come continua.)

► ESEMPIO 6.4. TBD.

6.3.4 La distribuzione di Gauss come limite della binomiale

Con un procedimento analogo a quello seguito nella sezione precedente—ma, se possibile, ancora più tedioso—si può dimostrare che la distribuzione di Gauss si ottiene come limite della distribuzione binomiale nel limite $np \rightarrow \infty$ e $nq \rightarrow \infty$.

Non daremo qui una dimostrazione rigorosa, ma notiamo che si tratta di un enunciato leggermente diverso da quello illustrato nella sezione 5.2. Notiamo però che non si tratta di un risultato inatteso, visto che una binomiale tende ad una distribuzione di Poisson per $n \rightarrow \infty$ e $np = \mu$ finita, e a sua volta una distribuzione di Poisson tende ad una distribuzione di Gauss per $\mu \rightarrow \infty$.

► ESEMPIO 6.5. TBD.

6.3.5 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 misurazioni indipendenti x_i di una certa grandezza θ . Fissato il valore *vero* (ed incognito) di θ , il generico x_i sarà una variabile casuale caratterizzata da una qualche funzione di distribuzione $\varphi(x_i; \theta)$ —e se le n misurazioni sono effettuate tutte nelle stesse condizioni possiamo assumere che la densità di probabilità φ sia la stessa per tutti gli x_i . Possiamo allora scrivere la densità di probabilità di ottenere una particolare serie di misure $x_1 \dots x_n$ come il prodotto

$$\mathcal{P}(x_1 \dots x_n | \theta) = \prod_{i=1}^n \varphi(x_i; \theta).$$

Se dovessimo, sulla base di questo, fornire una *stima* $\hat{\theta}$ di θ sarebbe ragionevole scegliere il valore che massimizza la probabilità a posteriori di ottenere la serie di misure in questione—imporre cioè che

$$\left. \frac{d\mathcal{P}(x_1 \dots x_n | \theta)}{d\theta} \right|_{\hat{\theta}} = 0,$$

o equivalentemente, dato che il logaritmo è una funzione monotona,

$$\left. \frac{d \ln \mathcal{P}(x_1 \dots x_n | \theta)}{d\theta} \right|_{\hat{\theta}} = \sum_{i=1}^n \left. \frac{d \ln \varphi(x_i; \theta)}{d\theta} \right|_{\hat{\theta}} = 0. \quad (84)$$

(Notiamo, per inciso, che il ragionamento appena descritto costituisce sostanzialmente il *principio di massima verosimiglianza* di cui parleremo molto più diffusamente in seguito. Passare al logaritmo in questo caso è comodo perché permette di trasformare un prodotto in una somma, e sfruttare poi il fatto che la derivata è un operatore lineare rispetto alla somma stessa.)

A prima vista la (84), così come è scritta, non appare particolarmente utile: il principio di massima verosimiglianza si usa tipicamente per calcolare la miglior stima della grandezza a cui siamo interessati, ma se non sappiamo la forma funzionale di $\varphi(x_i; \theta)$ non abbiamo alcun modo per ricavare $\hat{\theta}$. In questa semplice situazione, però, l'intuito ci dice che la miglior stima di θ dovrebbe essere la media aritmetica delle misure, dato che esse sono indipendenti e non c'è alcuna ragione di preferire in alcun modo una alle altre:

$$\hat{\theta} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (85)$$

La domanda originale di Gauss è, allora, leggermente diversa: assumendo che la (85) sia verificata, possiamo dedurne la forma di φ ? (Notiamo che la domanda è tutt'altro che banale, poiché la (84) e la (85) non sono, in generale, compatibili per una funzione di distribuzione φ generica.) Se proviamo a porre

$$\left. \frac{d \ln \varphi(x_i; \theta)}{d\theta} \right|_{\hat{\theta}} \propto (x_i - \theta)$$

la (84) e la (85) sono automaticamente soddisfatte, in quanto

$$\sum_{i=1}^n \left. \frac{d \ln \varphi(x_i; \theta)}{d\theta} \right|_{\bar{x}} \propto \sum_{i=1}^n (x_i - \bar{x}) = 0.$$

Tralasciando di dimostrare che questa condizione non è solo sufficiente, ma anche necessaria, ne segue che

$$\ln \varphi(x_i; \theta) = c_1 (x_i - \theta)^2 + c_2$$

e infine

$$\varphi(x_i; \theta) \propto e^{c_1 (x_i - \theta)^2}, \quad (86)$$

in cui, a parte la forma della normalizzazione, si riconosce l'espressione della distribuzione di Gauss (77) (notiamo esplicitamente che, perché φ 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.

6.3.6 Una derivazione alternativa della distribuzione di Gauss

Si deve ad Herschel [3] una derivazione della forma analitica della distribuzione di Gauss particolarmente significativa per il suo carattere essenzialmente geometrico (di fatto non si fa uso di alcun argomento di probabilità o statistica).

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. 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

$$\rho(x, y) dx dy = \varphi(x) dx \varphi(y) dy. \quad (87)$$

Possiamo scrivere la stessa densità di probabilità in coordinate polari nel piano

$$\rho(x, y) dx dy = \psi(r, \theta) r dr d\theta$$

e, a questo punto, richiediamo il secondo postulato, ovvero sia che la funzione di distribuzione sia indipendente dall'angolo θ

$$\varphi(x)\varphi(y) = \psi\left(\sqrt{x^2 + y^2}\right). \quad (88)$$

Ponendo $y = 0$ nell'equazione precedente otteniamo

$$\varphi(x)\varphi(0) = \psi(x),$$

col che possiamo riscrivere la (88) come

$$\ln\left(\frac{\varphi(x)}{\varphi(0)}\right) + \ln\left(\frac{\varphi(y)}{\varphi(0)}\right) = \ln\left(\frac{\varphi(\sqrt{x^2 + y^2})}{\varphi(0)}\right), \quad (89)$$

che è verificata se e solo se

$$\ln\left(\frac{\varphi(x)}{\varphi(0)}\right) = c_1 x^2,$$

ovvero

$$\varphi(x) = e^{c_1 x^2}, \quad (90)$$

che, come prima, è una distribuzione normale (il fatto che sia centrata in 0 deriva dalla nostra scelta del sistema di coordinate).

7

VARIABILI CAMPIONE ED ESTIMATORI

7.1 MEDIA E VARIANZA CAMPIONE

7.2 DEVIAZIONE STANDARD DELLA MEDIA

7.3 IL TEOREMA DEL LIMITE CENTRALE

7.3.1 *Un esempio interessante: il random walk in una dimensione*

BIBLIOGRAFIA

- [1] V. von Bortkevič, *Das Gesetz der kleinen Zahlen [The law of small numbers]*, Leipzig, Germany (1898).
- [2] A. De Moivre, *The doctrine of chances*, <https://archive.org/download/doctrineofchance00moiv/doctrineofchance00moiv.pdf>.
- [3] J. F. W. Herschel, *Quetelet on Probabilities*, Edinburgh Rev. Vol. 92 (1850), 1–57 (1850).
- [4] J. C. Maxwell, *On the Dynamical Theory of Gases*, Philosophical Transactions of the Royal Society of London, Vol. 157, 49–88 (1867), <http://www.jstor.org/stable/108968>
- [5] A. Leike, *Demonstration of the Exponential Decay Law Using Beer Froth*, European Journal of Physics, Vol. 23, 21–26 (2002).
- [6] T. Del Prete, *Methods of Statistical Data Analysis in High Energy Physics*, <http://www.pi.infn.it/atlas/documenti/note/statistica.ps.gz>.
- [7] M. Loreti, *Teoria degli errori e fondamenti di statistica (introduzione alla fisica sperimentale)*, <http://wwwcdf.pi.infn.it/labo/INDEX.html>.
- [8] A. N. Kolmogorov, *Foundations of the theory of probability*, Chelsea Publishing Company, New York, 1956, <http://www.mathematik.com/Kolmogorov/index.html>.
- [9] D. E. Knuth, *The Art of Computer Programming, Vol. 1, Fundamental Algorithms*, Addison Wesley, 2005, <http://www-cs-faculty.stanford.edu/~uno/taocp.html>.
- [10] D. E. Knuth, *The Art of Computer Programming, Vol. 2, Seminumerical Algorithms*, Addison Wesley, 2005, <http://www-cs-faculty.stanford.edu/~uno/taocp.html>.
- [11] D. E. Knuth, *The Art of Computer Programming, Vol. 3, Sorting and Searching*, Addison Wesley, 2005, <http://www-cs-faculty.stanford.edu/~uno/taocp.html>.
- [12] Donald E. Knuth, *The complexity of songs*, Communications of the ACM, Vol. 27 Issue 4, 344–346 (1984)

Bibliografia

COLOPHON

Queste dispense sono state redatte utilizzando esclusivamente *software* libero.

Il testo è stato composto con \LaTeX ed i grafici sono stati realizzati con gnuplot utilizzando il terminale tikz—il tutto su un sistema GNU/Linux (Fedora).