

# Probabilità continua

---

## 7.1 Variabili aleatorie

Il primo obiettivo di questa sezione è formalizzare il concetto di “probabilità che una certa misura dia un certo risultato con una certa approssimazione”; inoltre saremo in grado di dare una definizione precisa di cosa vuol dire che due misure sono indipendenti.

L’idea è di vedere una misura come una funzione a valori reali definita su uno spazio degli eventi che, a seconda dei casi, può essere l’insieme di tutti i possibili risultati di un esperimento, oppure l’insieme di tutti gli individui di una popolazione, o anche qualcosa di più generale. Infatti, “effettuare la misura” vuol dire associare un numero reale a ciascun possibile risultato dell’esperimento, o a ciascun individuo della popolazione.

Siccome i nostri strumenti di misura hanno una precisione finita, di solito il risultato della misura non sarà un numero reale preciso con tutte le sue infinite cifre decimali, ma un numero reale indicato con un certo errore; in altre parole, la misura ci può dire solo che il valore “vero” appartiene a un determinato intervallo di valori. Quindi siamo interessati alla probabilità che il valore della misura cada in un certo intervallo. L’unico caso in cui la misura può assumere un valore preciso è quando i possibili valori formano un insieme finito, o più in generale discreto, cioè un insieme i cui elementi sono ben separati fra di loro, per cui basta una precisione finita per distinguerli tutti (l’esempio tipico è dato da misure i cui possibili valori sono solo numeri interi).

**ESEMPIO 7.1** Supponiamo che la misura consista nel determinare la lunghezza della coda di una popolazione di cavie, usando un righello con la precisione di un millimetro (il tuo assistente non è riuscito a recuperare nulla di meglio). In tal caso lo spazio degli eventi è costituito dalla popolazione di cavie; la misura consiste nell’associare a ciascuna cavia la lunghezza della sua coda, misurata con il righello; e sei interessato alla probabilità che la misura ottenuta sia, per esempio,  $3 \pm 0.1$  cm, cioè che ricada nell’intervallo [2.9, 3.1].

ESEMPIO 7.2 Stavolta la misura consiste nell'effettuare la somma dei risultati del lancio di due dadi distinguibili. In questo caso lo spazio degli eventi consiste in tutti i possibili risultati del lancio di due dadi; la misura consiste nell'associare a ciascun risultato la somma dei valori; e sei interessato alla probabilità che la misura (la somma) ottenuta sia per esempio 8. Nota che in questo caso la misura può assumere solo un numero finito di valori, per cui ha senso chiedersi la probabilità che la misura assuma un valore preciso.

ESEMPIO 7.3 Supponiamo di irradiare una coltura batterica con una quantità data di radiazione, e che la misura consista nel determinare quanti batteri sono mutati a causa della radiazione. In questo caso lo spazio degli eventi consiste in tutti i possibili risultati dell'esperimento, e, siccome il numero di batteri della coltura è molto grande, la misura consiste nel determinare la percentuale di batteri mutati, per cui sei interessato alla probabilità che la percentuale sia per esempio  $10 \pm 0.1\%$ , cioè cada nell'intervallo  $[0.09, 0.11]$ .

ESEMPIO 7.4 Supponiamo di avere a disposizione una data quantità di materiale radioattivo, e che la misura consista nel contare la radiazione emessa in un'ora, che corrisponde al numero di atomi radioattivi che decadono nel materiale in un'ora. In questo caso lo spazio degli eventi consiste in tutti i possibili risultati dell'esperimento, la misura consiste nel contare il numero di decadimenti radioattivi in un'ora, e sei interessato alla probabilità che la misura ottenuta sia per esempio 27 218. Nota che in questo caso la misura può assumere solo valori interi (di fatto finiti, ma non sapendo a priori il numero di atomi presenti nel materiale non conviene mettere una limitazione a priori sui possibili valori della misura), per cui ha senso chiedersi la probabilità che la misura assuma un valore preciso.

Questo modo di pensare viene formalizzato dal concetto di variabile aleatoria. Sia  $\Omega$  uno spazio degli eventi, e indichiamo con  $\mathcal{A}$  la famiglia dei sottoinsiemi di  $\Omega$  di cui possiamo calcolare la probabilità (vedi la Sezione 2.5 e la Curiosità 2.1), e sia  $p: \mathcal{A} \rightarrow [0, 1]$  una distribuzione di probabilità. Allora una *variabile aleatoria* a valori reali è una funzione  $X: \Omega \rightarrow \mathbb{R}$  tale che gli insiemi  $X^{-1}(I) = \{p \in \Omega \mid X(p) \in I\}$  appartengano ad  $\mathcal{A}$  per ogni intervallo (aperto o chiuso) o semiretta<sup>1</sup> (aperta o chiusa)  $I$  di  $\mathbb{R}$ . In altre parole, stiamo dicendo che una variabile aleatoria è una funzione  $X: \Omega \rightarrow \mathbb{R}$  per cui siamo in grado di calcolare la probabilità che il valore di  $X$  cada in un certo intervallo (o semiretta)  $I$ .

*Osservazione 7.1* In questo contesto si utilizza spesso una notazione particolare per gli insiemi della forma  $X^{-1}(I)$ . Per l'esattezza si scrive

$$X^{-1}(I) = \{X \in I\},$$

---

<sup>1</sup> Non c'è bisogno di richiedere esplicitamente che  $X^{-1}(\mathbb{R})$  appartenga ad  $\mathcal{A}$ , in quanto  $X^{-1}(\mathbb{R}) = \Omega$  per ogni funzione  $X$ , e  $\Omega$  appartiene sempre ad  $\mathcal{A}$ .

e notazioni analoghe. Per esempio, si scrive

$$X^{-1}([2, 3]) = \{2 \leq X \leq 3\}, \quad X^{-1}((-\infty, -1)) = \{X < -1\},$$

e così via. Inoltre, la probabilità che la variabile aleatoria assuma un valore contenuto in  $I$  si scrive

$$p(X \in I)$$

e non  $p(\{X \in I\})$ , eliminando le parentesi graffe. Per esempio, si scrive

$$p(2 \leq X \leq 3), \quad p(X < -1).$$

*Osservazione 7.2* Se  $\mathcal{A}$  coincide con l'intera famiglia dei sottoinsiemi di  $\Omega$  (per esempio, se  $\Omega$  è finito) allora ogni funzione su  $\Omega$  a valori in  $\mathbb{R}$  è una variabile aleatoria.

*Osservazione 7.3* Vale la pena sottolineare che per il momento non ci stiamo occupando di *come* definire la distribuzione di probabilità su  $\Omega$ ; stiamo *assumendo* che ci sia. In particolare, non ci stiamo per ora preoccupando di come calcolare la probabilità che una variabile aleatoria assuma valore in un certo intervallo; stiamo soltanto formalizzando questi concetti, e cercando di identificarne alcune proprietà. La problematica di come determinare una distribuzione di probabilità su uno spazio degli eventi, o di come determinare la funzione di distribuzione (vedi oltre) di una variabile aleatoria, è stata discussa nel Capitolo 2, e ne ripareremo in alcuni casi specifici anche in questo capitolo.

Negli esempi precedenti abbiamo distinto misure (cioè variabili aleatorie) che potevano assumere a priori qualsiasi valore, e misure che potevano assumere soltanto un numero finito (o discreto) di valori. Vediamo di formalizzare anche questa distinzione.

Diremo che un insieme  $D \subset \mathbb{R}$  è *discreto* se esiste  $\varepsilon > 0$  tale che due elementi distinti di  $D$  distano sempre almeno  $\varepsilon$ .

**ESEMPIO 7.5** L'insieme  $\mathbb{N}$  dei numeri naturali è un insieme discreto, in quanto due numeri naturali diversi distano sempre di almeno  $\varepsilon = 1$ ; inoltre ogni insieme finito è chiaramente (perché?) discreto.

Diremo che una variabile aleatoria  $X: \Omega \rightarrow \mathbb{R}$  è *discreta* se la sua immagine  $X(\Omega) \subset \mathbb{R}$  è un insieme discreto (cioè se  $X$  può assumere solo un insieme discreto di valori).

*Osservazione 7.4* Nella Sezione 7.4 introdurremo il concetto di variabile aleatoria continua. Le variabili aleatorie continue non sono discrete, ma esistono variabili aleatorie particolarmente antipatiche che non sono né continue né discrete.

**ESEMPIO 7.6** Le misure degli Esempi 7.2 e 7.4 sono variabili aleatorie discrete, mentre le misure degli Esempi 7.1 e 7.3 sono variabili aleatorie continue.

*Osservazione 7.5* Se  $X$  è una variabile aleatoria discreta, e  $a \in \mathbb{R}$  è un suo possibile valore, allora ha perfettamente senso cercare di calcolare la probabilità dell'evento  $\{X = a\}$ , in quanto  $\{X = a\} = \{a - \varepsilon < X < a + \varepsilon\}$ , dove  $\varepsilon > 0$  è la distanza minima fra due valori distinti di  $X$ . Se  $X$  è una variabile aleatoria non discreta, invece, calcolare la probabilità dell'evento  $\{X = a\}$  potrebbe non avere molto senso, anche quando si può fare; ne ripareremo meglio nella Sezione 7.4.

Siamo ora in grado di definire precisamente il concetto di “misure indipendenti” che abbiamo citato più volte nel Capitolo 2. L’idea è che due misure sono indipendenti se il risultato di una non influenza il risultato dell’altra; o, in altre parole, se la probabilità dell’evento che la prima misura abbia un certo valore è indipendente dalla probabilità dell’evento che la seconda misura abbia un certo valore. Ricordando la definizione di eventi indipendenti introdotta nella Sezione 2.6 siamo portati alla seguente definizione: diremo che due variabili aleatorie  $X_1, X_2: \Omega \rightarrow \mathbb{R}$  definite sullo stesso spazio degli eventi sono *indipendenti* se

$$p(\{X_1 \in I_1\} \cap \{X_2 \in I_2\}) = p(X_1 \in I_1) \cdot p(X_2 \in I_2)$$

per ogni coppia di intervalli o semirette  $I_1, I_2 \subset \mathbb{R}$ . In altre parole,  $X_1$  e  $X_2$  sono indipendenti se gli eventi  $\{X_1 \in I_1\}$  e  $\{X_2 \in I_2\}$  sono indipendenti quali che siano gli intervalli o semirette  $I_1$  e  $I_2$ .

Più in generale, diremo che  $n$  variabili aleatorie  $X_1, \dots, X_n$  sono *indipendenti* se ogni volta che ne scegliamo  $1 \leq k \leq n$  diverse  $X_{i_1}, \dots, X_{i_k}$  abbiamo

$$p(\{X_{i_1} \in I_1\} \cap \dots \cap \{X_{i_k} \in I_k\}) = p(X_{i_1} \in I_1) \cdots p(X_{i_k} \in I_k)$$

per ogni  $k$ -upla di intervalli o semirette  $I_1, \dots, I_k \subset \mathbb{R}$ ; vedi la Curiosità 2.2.

**ESEMPIO 7.7** Sia  $\Omega = \{(i, j) \mid i, j \in \{1, 2, 3, 4, 5, 6\}\}$  lo spazio degli eventi del lancio di due dadi distinti, e mettiamo su  $\Omega$  la solita distribuzione di probabilità uniforme: ogni evento semplice ha probabilità  $1/36$ . Indichiamo poi con  $X_1: \Omega \rightarrow \mathbb{R}$  la variabile aleatoria che misura il lancio del primo dado, cioè  $X_1(i, j) = i$  per ogni  $(i, j) \in \Omega$ ; e con  $X_2: \Omega \rightarrow \mathbb{R}$  la variabile aleatoria che misura il lancio del secondo dado, cioè  $X_2(i, j) = j$  per ogni  $(i, j) \in \Omega$ . Chiaramente,  $X_1$  e  $X_2$  sono variabili aleatorie discrete, che possono assumere solo i valori  $1, \dots, 6$ . Vediamo che sono anche indipendenti. Se  $I_1 \subset \mathbb{R}$  è un intervallo (o una semiretta), l’insieme  $\{X_1 \in I_1\}$  è formato da quelle coppie di lanci di dadi in cui il risultato del primo dado appartiene a  $I_1$ . Se  $I_1 \cap \{1, \dots, 6\}$  contiene  $k_1$  elementi, allora  $\{X_1 \in I_1\}$  conterrà (perché?)  $6k_1$  eventi semplici, e quindi

$$p(X_1 \in I_1) = \frac{6k_1}{36} = \frac{k_1}{6},$$

dove abbiamo usato il fatto che su  $\Omega$  abbiamo messo la distribuzione di probabilità uniforme per cui la probabilità di qualsiasi evento è uguale al numero di elementi dell’evento diviso per 36.

Analogamente, se  $I_2 \subset \mathbb{R}$  è un altro intervallo (o semiretta) e  $I_2 \cap \{1, \dots, 6\}$  contiene  $k_2$  elementi, abbiamo

$$p(X_2 \in I_2) = \frac{k_2}{6} .$$

Infine, l'evento  $\{X_1 \in I_1\} \cap \{X_2 \in I_2\}$  contiene tutte e sole le coppie  $(i, j)$  con  $i \in I_1$  e  $j \in I_2$ ; quindi  $\{X_1 \in I_1\} \cap \{X_2 \in I_2\}$  contiene (perché?)  $k_1 k_2$  elementi. Dunque

$$p(\{X_1 \in I_1\} \cap \{X_2 \in I_2\}) = \frac{k_1 k_2}{36} = \frac{k_1}{6} \cdot \frac{k_2}{6} = p(X_1 \in I_1) \cdot p(X_2 \in I_2) ,$$

e abbiamo verificato che  $X_1$  e  $X_2$  sono indipendenti (cioè che il risultato del lancio del primo dado non influenza il risultato del lancio del secondo).

Per chiarezza, anche se non sarebbe strettamente necessario, ripeto questo argomento per due intervalli specifici. Poniamo  $I_1 = [1, 5/2]$ , e  $I_2 = [2, 4]$ . Allora  $I_1 \cap \{1, \dots, 6\} = \{1, 2\}$  e  $I_2 \cap \{1, \dots, 6\} = \{2, 3, 4\}$ , per cui

$$\begin{aligned} \{X_1 \in I_1\} &= \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), \\ &\quad (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6)\} \end{aligned}$$

ha  $2 \cdot 6 = 12$  elementi, e

$$\begin{aligned} \{X_2 \in I_2\} &= \{(1, 2), (2, 2), (3, 2), (4, 2), (5, 2), (6, 2), (1, 3), (2, 3), (3, 3), \\ &\quad (4, 3), (5, 3), (6, 3), (1, 4), (2, 4), (3, 4), (4, 4), (5, 4), (6, 4)\} \end{aligned}$$

ha  $6 \cdot 3 = 18$  elementi; in particolare  $p(X_1 \in I_1) = 12/36 = 1/3$ , e analogamente  $p(X_2 \in I_2) = 18/36 = 1/2$ . D'altra parte

$$\{X_1 \in I_1\} \cap \{X_2 \in I_2\} = \{(1, 2), (1, 3), (1, 4), (2, 2), (2, 3), (2, 4)\}$$

ha  $2 \cdot 3 = 6$  elementi, per cui ha probabilità  $6/36 = 1/6 = (1/3) \cdot (1/2)$  come previsto.

**ESEMPIO 7.8** Indicato con  $\Omega$  lo stesso spazio degli eventi dell'esempio precedente, sia  $X: \Omega \rightarrow \mathbb{R}$  la variabile aleatoria data dalla somma dei risultati dei due dadi:  $X(i, j) = i + j$ . Allora le variabili aleatorie  $X$  ed  $X_2$  (dove  $X_2$  è quella introdotta nell'esempio precedente) non sono indipendenti. Infatti, prendiamo  $I_2 = [2, 4]$  e  $I = [1, 2]$ . Allora  $\{X \in I\} = \{(1, 1)\}$ , per cui  $p(X \in I) = 1/36$ ; inoltre abbiamo visto che  $p(X_2 \in I_1) = 1/2$ . Ma  $\{X \in I\} \cap \{X_2 \in I_2\} = \emptyset$ ; quindi

$$p(\{X \in I\} \cap \{X_2 \in I_2\}) = p(\emptyset) = 0 \neq \frac{1}{72} = \frac{1}{36} \cdot \frac{1}{2} = p(X \in I) \cdot p(X_2 \in I_2) ,$$

e  $X$  e  $X_2$  non sono indipendenti (infatti, la somma dei due dadi dipende dal valore del secondo dado).

Concludo questa sezione introducendo una nozione che sarà più utile per le variabili aleatorie continue, ma ha senso anche quelle discrete. Sia  $X: \Omega \rightarrow \mathbb{R}$  una variabile aleatoria. La *funzione di distribuzione* (o *legge*) di  $X$  è la funzione  $F_X: \mathbb{R} \rightarrow [0, 1]$  definita da

$$F_X(t) = p(X \leq t).$$

In altre parole,  $F_X(t)$  è la probabilità che la variabile aleatoria  $X$  assuma un valore minore o uguale a  $t$ .

Conoscendo la funzione di distribuzione è possibile ricostruire moltissime caratteristiche della variabile aleatoria. Per esempio, da  $\Omega \setminus \{X \leq t\} = \{X > t\}$  ricaviamo

$$p(X > t) = 1 - F_X(t).$$

Poi se  $t_1 < t_2$  abbiamo  $\{t_1 < X \leq t_2\} = \{X \leq t_2\} \setminus \{X \leq t_1\}$ , per cui

$$p(t_1 < X \leq t_2) = F_X(t_2) - F_X(t_1).$$

*Osservazione 7.6* Se  $t_1 \leq t_2$  allora  $\{X \leq t_1\} \subseteq \{X \leq t_2\}$  e quindi  $F_X(t_1) \leq F_X(t_2)$ ; in altre parole, le funzioni di distribuzioni sono sempre crescenti. Inoltre è chiaro (perché?) che

$$\lim_{t \rightarrow -\infty} F_X(t) = 0 \quad \text{e} \quad \lim_{t \rightarrow +\infty} F_X(t) = 1.$$

*Osservazione 7.7* Se  $X$  è una variabile aleatoria discreta, allora  $F_X$  è una funzione costante a tratti. Infatti, se  $\varepsilon > 0$  è la distanza minima fra due valori di  $X$ , e  $x$  è un valore di  $X$ , allora l'intervallo aperto  $(x, x + \varepsilon)$  non contiene valori di  $X$ , per cui

$$\forall t \in (x, x + \varepsilon) \quad F_X(t) = p(X \leq t) = p(X \leq x) = F_X(x).$$

Inoltre, possiamo recuperare  $p(X = x)$  usando la funzione di distribuzione. Infatti,  $x$  è l'unico valore di  $X$  nell'intervallo  $(x - \varepsilon/2, x + \varepsilon/2]$ . Quindi abbiamo  $\{X = x\} = \{x - \varepsilon/2 < X \leq x + \varepsilon/2\}$  e

$$p(X = x) = F_X(x + \varepsilon/2) - F_X(x - \varepsilon/2).$$

Viceversa, conoscendo  $p(X = x)$  per ogni valore  $x$  di  $X$  possiamo ricostruire la funzione di distribuzione di  $X$ . Infatti se  $x_1 < x_2$  sono due valori consecutivi di  $X$  abbiamo  $F_X(t) = F_X(x_1)$  se  $x_1 \leq t < x_2$  e

$$F_X(x_2) = F_X(x_1) + p(X = x_2).$$

Quindi procedendo un valore alla volta determiniamo  $F_X$  su tutta la retta; vedi l'Esempio 7.9.

**CURIOSITÀ 7.1** Con un po' più di fatica possiamo usare la funzione di distribuzione per recuperare  $p(X = x)$  per ogni variabile aleatoria  $X$  e ogni  $x \in \mathbb{R}$ . Infatti, siccome

$$\{X < x\} = \bigcup_{t < x} \{X \leq t\} \quad \text{e} \quad \{X = x\} = \{X \leq x\} \setminus \{X < x\},$$

otteniamo

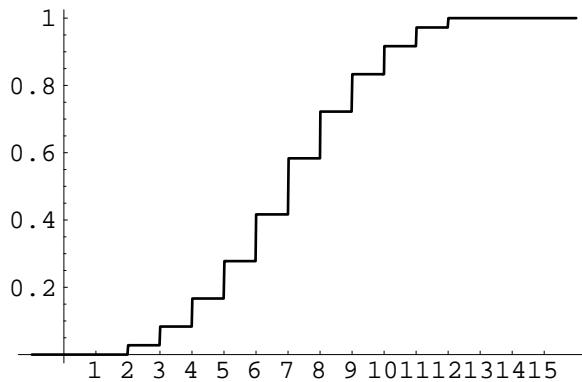
$$p(X < x) = \lim_{t \rightarrow x^-} F_X(t) \quad \text{e} \quad p(X = x) = F_X(x) - \lim_{t \rightarrow x^-} F_X(t);$$

nota che il limite esiste in quanto  $F$  è crescente e limitata. In particolare, se  $F_X$  è una funzione continua a sinistra allora  $p(X = x) = 0$  per ogni  $x \in \mathbb{R}$  — che equivale a dire che è impossibile effettuare la misura  $X$  con precisione infinita. Come vedremo nell'Osservazione 7.13, questo accade per le variabili aleatorie continue.

**ESEMPIO 7.9** Vediamo di determinare la funzione di distribuzione per la variabile aleatoria “somma del lancio di due dadi”. Sia  $\Omega$  lo spazio degli eventi dell'Esempio 7.7, e  $X: \Omega \rightarrow \mathbb{R}$  la variabile aleatoria somma data da  $X(i, j) = i + j$ . Allora  $X$  è una variabile aleatoria discreta, con valori possibili  $2, \dots, 12$ . Nell'Esempio 2.14 abbiamo calcolato la probabilità di  $\{X = k\}$  per  $k = 2, \dots, 12$ , ottenendo

$$\begin{aligned} p(X = 2) = p(X = 12) &= \frac{1}{36}, & p(X = 3) = p(X = 11) &= \frac{1}{18}, \\ p(X = 4) = p(X = 10) &= \frac{1}{12}, & p(X = 5) = p(X = 9) &= \frac{1}{9}, \\ p(X = 6) = p(X = 8) &= \frac{5}{36}, & p(X = 7) &= \frac{1}{6}. \end{aligned}$$

Quindi  $F_X(t) = 0$  finché  $t < 2$ ; poi  $F_X(2) = p(X \leq 2) = p(X = 2) = 1/36$ . La funzione di distribuzione  $F_X$  rimane costante in  $2 \leq t < 3$ , mentre in  $t = 3$  otteniamo  $F_X(3) = p(X \leq 3) = p(X = 2) + p(X = 3) = 3/36 = 1/12$ . Poi  $F_X$  rimane costante in  $3 \leq t < 4$ , e ha un altro scalino in  $F_X(4) = F_X(3) + p(X = 4) = 1/6$ . Procedendo in questo modo otteniamo la funzione di distribuzione il cui grafico è mostrato nella Fig. 7.1.



**Figura 7.1** La funzione di distribuzione della somma di due dadi.

## 7.2 Media e varianza di variabili aleatorie discrete

Nel Capitolo 3 abbiamo introdotto i concetti di media, varianza e deviazione standard di un insieme di misurazioni; in questa sezione vogliamo introdurre concetti analoghi per una misura — ovvero per una variabile aleatoria, almeno nel caso discreto (del caso continuo parleremo nella Sezione 7.4).

Per capire come procedere, supponiamo di avere uno spazio degli eventi finito  $\Omega = \{v_1, \dots, v_n\}$  con distribuzione di probabilità uniforme, e una variabile aleatoria (necessariamente discreta)  $X: \Omega \rightarrow \mathbb{R}$ . Nella Sezione 3.6 abbiamo definito la media  $\bar{x}$  delle misurazioni  $X(v_1), \dots, X(v_n)$  ponendo

$$\bar{x} = \frac{X(v_1) + \dots + X(v_n)}{n}.$$

Ora, alcune di queste misure potrebbero avere risultati uguali; potrebbe succedere che  $X(v_i) = X(v_j)$  per qualche  $i \neq j$ . Siano  $x_1, \dots, x_k$  (con  $k \leq n$ ) gli effettivi valori assunti da  $X$ , cioè  $X(\Omega) = \{x_1, \dots, x_k\}$ , e per  $j = 1, \dots, k$  indichiamo con  $f_j$  il numero di volte che  $X$  assume il valore  $x_j$ ; in altre parole,  $f_j$  è la cardinalità di  $\{X = x_j\}$ . Allora

$$X(v_1) + \dots + X(v_n) = f_1 x_1 + \dots + f_k x_k,$$

in quanto nella somma a destra abbiamo semplicemente raggruppato insieme gli eventuali addendi uguali della somma a sinistra. Quindi

$$\bar{x} = \frac{f_1}{n} x_1 + \dots + \frac{f_k}{n} x_k.$$

Ma su  $\Omega$  abbiamo messo la distribuzione uniforme di probabilità; quindi

$$\forall j = 1, \dots, k \quad p(X = x_j) = \frac{f_j}{n},$$

e dunque siamo arrivati alla formula

$$\bar{x} = \sum_{j=1}^k p_j x_j, \tag{7.1}$$

dove  $p_j = p(X = x_j)$ , che esprime la media come media ponderata (vedi anche la Curiosità 3.3).

La nostra idea intuitiva è che la media di una misura deve coincidere con la media di tutte le misurazioni possibili; quindi (7.1) dovrebbe rappresentare la media della variabile aleatoria  $X$  che abbiamo appena studiato.

È chiaro allora come definire la media di una variabile aleatoria discreta qualsiasi. Sia  $X: \Omega \rightarrow \mathbb{R}$  una variabile aleatoria discreta; la *media* (o *valor medio*, o *valore atteso*, o *speranza*) di  $X$  è il numero  $E(X)$  dato da

$$E(X) = \sum p_j x_j, \tag{7.2}$$

dove la somma è estesa a tutti i valori  $x_j$  di  $X$ , e  $p_j = p(X = x_j)$ . A volte scriveremo  $\mu_X$  invece di  $E(X)$ .

*Osservazione 7.8* “ $E$ ” sta per “expected”, che in inglese significa “atteso”. Si dice “valore atteso” perché, in un certo senso, la media rappresenta il valore che ci aspettiamo (o che speriamo...) la misura dia. S’intende che tutto ciò non è da prendere alla lettera, in quanto può capitare (vedi l’Esempio 7.10) che il valore atteso di una variabile aleatoria non sia fra i valori che la variabile può assumere!

*Osservazione 7.9* Se l’insieme dei valori della variabile aleatoria  $X$  è finito, allora (7.2) è una somma finita, e non ci sono problemi di interpretazione. Se invece l’insieme dei valori di  $X$  è infinito (ma discreto), allora (7.2) è una somma infinita, cioè una serie (vedi la Curiosità 5.9). Siccome gli addendi sono tutti numeri positivi o nulli, questa somma infinita si comporta come una somma normale, e il suo valore è semplicemente il limite delle somme parziali finite. Vedremo nella prossima sezione esempi di come manipolare somme infinite di questo genere.

La varianza di misurazioni era la media degli scarti quadratici, cioè delle differenze al quadrato  $(x_j - \bar{x})^2$ . La probabilità che compaia la differenza  $x_j - \bar{x}$  è chiaramente uguale alla probabilità che compaia  $x_j$  come valore della misura, per cui è uguale a  $p_j = p(X = x_j)$ . Scrivendo la media degli scarti quadratici come media ponderata possiamo allora definire la varianza  $\text{Var}(X)$  di una variabile aleatoria discreta  $X: \Omega \rightarrow \mathbb{R}$  ponendo

$$\text{Var}(X) = \sum p_j(x_j - \mu_X)^2,$$

dove anche stavolta la somma è estesa a tutti i possibili valori  $x_j$  di  $X$ . A volte scriveremo  $\sigma_X^2$  invece di  $\text{Var}(X)$ .

Infine, la *deviazione standard*  $\sigma_X$  di una variabile aleatoria discreta  $X: \Omega \rightarrow \mathbb{R}$  è la radice quadrata della varianza:

$$\sigma_X = \sqrt{\sigma_X^2}.$$

A volte scriveremo  $\text{DS}(X)$  invece di  $\sigma_X$ .

Vediamo ora qualche esempio di calcolo di valore atteso e varianza.

**ESEMPIO 7.10** Sia  $\Omega = \{1, \dots, 6\}$  lo spazio degli eventi del lancio di un dado non truccato, e indichiamo con  $X: \Omega \rightarrow \mathbb{R}$  la variabile aleatoria banale data dal valore del lancio del dado:  $X(j) = j$  per  $j = 1, \dots, 6$ . Siccome il dado non è truccato, su  $\Omega$  abbiamo la distribuzione uniforme di probabilità, per cui  $p_j = p(X = j) = \frac{1}{6}$  per  $j = 1, \dots, 6$ . Quindi

$$E(X) = p_1 \cdot 1 + p_2 \cdot 2 + \dots + p_6 \cdot 6 = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = 3.5.$$

In altre parole il valore atteso del lancio di un dado è 3.5, che è effettivamente nel mezzo di tutti i possibili valori, ma non può essere mai il risultato di un lancio!

In modo analogo calcoliamo la varianza del lancio di un dado non truccato:

$$\text{Var}(X) = p_1 \cdot (1 - 3.5)^2 + p_2 \cdot (2 - 3.5)^2 + \cdots + p_6 \cdot (6 - 3.5)^2 = \frac{17.5}{6} = 2.91\overline{6},$$

per cui  $\sigma_X = \sqrt{2.91\overline{6}} \simeq 1.71$ .

**ESEMPIO 7.11** Calcoliamo ora il valore atteso e la varianza della somma del lancio di due dadi non truccati. Usando le probabilità calcolate nell'Esempio 7.9 troviamo

$$\begin{aligned} E(X) &= \frac{1}{36}2 + \frac{1}{18}3 + \frac{1}{12}4 + \frac{1}{9}5 + \frac{5}{36}6 + \frac{1}{6}7 + \frac{5}{36}8 + \frac{1}{9}9 + \frac{1}{12}10 + \frac{1}{18}11 + \frac{1}{36}12 \\ &= \frac{252}{36} = 7, \end{aligned}$$

per cui il valore atteso del lancio di due dadi non truccati è 7, in accordo con la nostra intuizione. La varianza è data da

$$\begin{aligned} \text{Var}(X) &= \frac{1}{36}(2 - 7)^2 + \frac{1}{18}(3 - 7)^2 + \frac{1}{12}(4 - 7)^2 + \frac{1}{9}(5 - 7)^2 \\ &\quad + \frac{5}{36}(6 - 7)^2 + \frac{1}{6}(7 - 7)^2 + \frac{5}{36}(8 - 7)^2 + \frac{1}{9}(9 - 7)^2 \\ &\quad + \frac{1}{12}(10 - 7)^2 + \frac{1}{18}(11 - 7)^2 + \frac{1}{36}(12 - 7)^2 \\ &= \frac{210}{36} = 5.8\overline{3}, \end{aligned}$$

e la deviazione standard è  $\text{DS}(X) = \sqrt{35/6} \simeq 2.42$ .

La distribuzione binomiale studiata nella Sezione 2.10 può essere riformulata in termini di variabili aleatorie. La situazione è la seguente: lo spazio degli eventi  $\Omega$  consiste in tutti i possibili risultati di  $n$  esperimenti indipendenti (per esempio,  $n$  lanci di un dado, o le nascite di  $n$  figli). In ciascun esperimento un certo evento  $E$  (per esempio, un risultato pari, o la nascita di una femmina) può avvenire con probabilità  $p$ . Consideriamo allora la variabile aleatoria  $X: \Omega \rightarrow \mathbb{R}$  che conta il numero di volte che l'evento  $E$  è effettivamente accaduto negli  $n$  esperimenti (il numero di risultati pari in una data sequenza di  $n$  lanci, o il numero di figlie femmine in una data famiglia con  $n$  figli). Allora nella Sezione 2.10 abbiamo visto che  $X$  soddisfa la legge

$$p(X = k) = \binom{n}{k} p^k q^{n-k}, \quad (7.3)$$

per  $k = 0, \dots, n$ , dove  $q = 1 - p$ . Nota che, grazie all'Osservazione 7.6, a partire da (7.3) è possibile ricostruire completamente la funzione di distribuzione di  $X$ . Qualsiasi variabile aleatoria discreta che soddisfa (7.3) verrà detta *variabile aleatoria Bernoulliana* di tipo  $(n, p)$ .

Vogliamo calcolare valore atteso e varianza per una variabile aleatoria Bernoulliana di tipo  $(n, p)$ . Cominciamo con il valore atteso: siccome i possibili valori di  $X$

sono  $0, \dots, n$  otteniamo

$$\begin{aligned} E(X) &= \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} k = \sum_{k=1}^n \frac{n!}{(n-k)!k!} kp^k q^{n-k} = \sum_{k=1}^n \frac{n!}{(n-k)!(k-1)!} p^k q^{n-k} \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} q^{n-k} = np \sum_{h=0}^{n-1} \frac{(n-1)!}{(n-h-1)!h!} p^h q^{n-1-h} \\ &= np(p+q)^{n-1} = np \end{aligned}$$

(controlla di aver capito bene tutti i passaggi di questo conto). In altre parole, se un evento  $E$  in un esperimento può accadere con probabilità  $p$ , in  $n$  esperimenti ci aspettiamo accada  $np$  volte (in perfetto accordo con la nostra intuizione).

Il calcolo della varianza è simile ma un poco più complicato. L'idea è cercare di ricondursi a somme del tipo

$$\sum_{k=0}^{\ell} \binom{\ell}{k} p^k q^{\ell-k},$$

che valgono  $(p+q)^\ell = 1$ . Scrivendo

$$k - np = k - n(1-q) = nq - (n-k)$$

otteniamo

$$\begin{aligned} \text{Var}(X) &= \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} (k - np)^2 = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} (k - np)(nq - (n-k)) \\ &= n \sum_{k=0}^n \binom{n}{k} kp^k q^{n-k+1} - \sum_{k=0}^n \binom{n}{k} k(n-k)p^k q^{n-k} \\ &\quad - n^2 \sum_{k=0}^n \binom{n}{k} p^{k+1} q^{n-k+1} + n \sum_{k=0}^n \binom{n}{k} (n-k)p^{k+1} q^{n-k} \\ &= npq \left[ n \sum_{k=1}^n \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} q^{n-k} - \sum_{k=1}^{n-1} \frac{(n-1)!}{(n-k-1)!(k-1)!} p^{k-1} q^{n-k-1} \right. \\ &\quad \left. - n \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} + n \sum_{k=0}^{n-1} \frac{(n-1)!}{(n-k-1)!k!} p^k q^{n-k-1} \right] \\ &= npq \left[ n \sum_{h=0}^{n-1} \binom{n-1}{h} p^h q^{n-1-h} - (n-1) \sum_{h=0}^{n-2} \binom{n-2}{h} p^h q^{n-2-h} \right. \\ &\quad \left. - n \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} + n \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} \right] \\ &= npq [n(p+q)^{n-1} - (n-1)(p+q)^{n-2} - n(p+q)^n + n(p+q)^{n-1}] \\ &= npq. \end{aligned}$$

Quindi  $\sigma_X^2 = npq$ , e la deviazione standard è  $\sigma_X = \sqrt{npq}$ .

ESEMPIO 7.12 Riprendiamo l’Esempio 2.59: *supponendo che la probabilità che un paziente sottoposto a una data operazione muoia entro un mese sia del 12%, se sono state effettuate 4 operazioni, quanti pazienti ci aspettiamo siano ancora vivi dopo un mese?* Se la probabilità che un paziente muoia entro un mese dall’operazione è  $q = 12/100$ , la probabilità che sopravviva è  $p = 1 - q = 88/100$ . Vogliamo trovare il valore atteso della variabile Bernoulliana  $X$  di tipo  $(4, p)$  data dal numero di pazienti che sopravvivono dopo un mese. Il conto che abbiamo appena fatto ci dà

$$E(X) = 4p = 4 \cdot \frac{88}{100} = \frac{88}{25} = 3.52 ;$$

quindi ci aspettiamo che un mese dopo l’operazione siano ancora vivi fra 3 e 4 pazienti.

### 7.3 Distribuzione di Poisson

Il calcolo della distribuzione binomiale prevede di conoscere a priori il numero massimo  $n$  di eventi possibili; in molti casi, invece, non è detto che lo si sappia. Inoltre, la distribuzione binomiale necessita di sapere a priori la probabilità  $p$  che l’evento accada; in molti casi, invece, questa probabilità non è nota. L’osservazione del fenomeno può invece fornire il numero medio  $\mu$  (che per la distribuzione binomiale è uguale a  $np$ , come abbiamo visto nella sezione precedente) di eventi in un dato intervallo di tempo; e vogliamo vedere se riusciamo usando solo  $\mu$  a calcolare la probabilità del verificarsi di  $k$  eventi, con  $k \in \mathbb{N}$  qualsiasi, nel dato intervallo di tempo.

Supponiamo quindi di essere nella seguente situazione:

- abbiamo fissato un dato intervallo di tempo (o una data regione di spazio);
- in un qualsiasi istante di questo intervallo (o punto di questa regione) può accadere un dato evento, e noi non siamo in grado di predire quando (o dove);
- l’accadere o meno di un evento in un dato istante (o punto) è indipendente dall’accadere o meno dell’evento in un altro istante (o punto);
- sappiamo che in media accadranno  $\mu$  eventi nell’intervallo di tempo (o regione di spazio).

Indichiamo con  $X$  la variabile aleatoria che conta il numero di eventi accaduti nel fissato intervallo di tempo (o regione di spazio); il nostro obiettivo è calcolare la probabilità  $p(X = k)$  che accadano esattamente  $k$  eventi, quale che sia  $k \in \mathbb{N}$ . Chiaramente  $X$  è una variabile aleatoria discreta, ma l’insieme dei valori che può assumere non è finito. Il prossimo esempio contiene un elenco di situazioni rappresentabili con un modello di questo genere.

ESEMPIO 7.13 – Il numero di automobili che passano in un dato punto di una strada (sufficientemente lontano da semafori) in un determinato periodo di tempo;

- il numero di errori di battuta commessi scrivendo una pagina di testo;
- il numero di telefonate che un call center riceve in un minuto;
- il numero di mutazioni in una fissata sequenza di DNA sottoposto a una data radiazione;
- il numero di decadimenti radioattivi in una fissata quantità di sostanza radioattiva in un dato intervallo di tempo (molto più piccolo della vita media della sostanza radioattiva);
- il numero di alberi di leccio in una data regione di foresta mista;
- il numero di stelle in un dato volume di spazio;
- il numero di virus che possono infettare una data cellula in uno specificato intervallo di tempo in una coltura cellulare fissata;
- il numero di cavalleggeri prussiani uccisi dal calcio di un cavallo in un dato anno (questo esempio è dovuto a L.J. Bortkiewicz).

Per ricavare  $p(X = k)$  procediamo suddividendo il nostro intervallo di tempo (o regione di spazio) in  $n$  intervallini uguali, e indichiamo con  $Y_n$  la variabile aleatoria che conta il numero di intervallini in cui accade almeno un evento (o, come diremo, che vengono *colpiti*). Siccome l'accadere o meno di un evento in un intervallino è del tutto indipendente dall'accadere o meno di un evento in un altro intervallino, possiamo equiparare l'osservazione di un intervallino a un singolo esperimento, e quindi pensare di stare effettuando  $n$  esperimenti del tutto indipendenti l'uno dall'altro. Quindi  $Y_n$  è una variabile aleatoria Bernoulliana di tipo  $(n, p)$ , dove  $p$  è la probabilità che un dato intervallino sia colpito.

Ora, noi ci aspettiamo in media  $\mu$  eventi nell'intero intervallo; d'altra parte, ci aspettiamo che in media vengano colpiti  $np$  intervallini. Se  $n$  è molto più grande di  $\mu$ , in media ciascun intervallino sarà colpito al più una volta sola; quindi il numero medio di eventi deve coincidere col numero medio di intervallini colpiti, cioè  $\mu = np$ , ovvero  $p = \mu/n$ .

Riassumendo, se  $n$  è molto maggiore di  $\mu$  la probabilità che vengano colpiti  $k$  intervallini è

$$\begin{aligned}
p(Y_n = k) &= \binom{n}{k} p^k q^{n-k} = \binom{n}{k} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{n^k k!} \frac{\mu^k}{(1 - \frac{\mu}{n})^k} \left(1 - \frac{\mu}{n}\right)^n \\
&= \frac{n^k (1 - \frac{1}{n}) \cdots (1 - \frac{k-1}{n})}{n^k (1 - \frac{\mu}{n})^k} \frac{\mu^k}{k!} \left(1 - \frac{\mu}{n}\right)^n \\
&= \frac{\mu^k}{k!} \frac{(1 - \frac{1}{n}) \cdots (1 - \frac{k-1}{n})}{(1 - \frac{\mu}{n})^k} \left(1 - \frac{\mu}{n}\right)^n.
\end{aligned} \tag{7.4}$$

Benché molto improbabile, anche per  $n$  molto grande un intervallino potrebbe essere colpito più di una volta, per cui non abbiamo ancora trovato la probabilità che avvengano esattamente  $k$  eventi nell'intervallo. Per risolvere questo problema basta far tendere  $n$  all'infinito: in questo modo gli intervallini tendono a ridursi a

singoli punti, e quindi  $p(Y_n = k)$  tende alla probabilità  $p(X = k)$  che avvengano esattamente  $k$  eventi.

Nell'espressione finale di (7.4), la prima frazione non dipende da  $n$ , per cui non varia facendo tendere  $n$  all'infinito. Numeratore e denominatore della seconda frazione sono composti da un numero finito di fattori ( $k - 1$  nel numeratore,  $k$  nel denominatore) che tendono a 1 al tendere di  $n$  all'infinito; quindi la seconda frazione tende a 1 al tendere di  $n$  all'infinito. Ricordando il limite fondamentale (4.24) che abbiamo usato per definire la funzione esponenziale otteniamo quindi

$$p(X = k) = \frac{\mu^k}{k!} e^{-\mu}. \quad (7.5)$$

Ogni variabile aleatoria a valori numeri naturali che soddisfa (7.5) verrà detta *variabile aleatoria di Poisson* di media  $\mu$ .

**ESEMPIO 7.14** *Sapendo che in media in un anno 12 cavalleggeri prussiani vengono uccisi dal calcio di un cavallo, qual è la probabilità  $p$  che nel 1861 ne siano stati uccisi solo 7? Siccome il numero di cavalleggeri prussiani uccisi in un anno dal calcio di un cavallo è una variabile aleatoria di Poisson, possiamo applicare la (7.5) e trovare*

$$p = \frac{12^7}{7!} e^{-12} \simeq \frac{35\,831\,808}{5040} \cdot 6.14 \cdot 10^{-6} \simeq 0.044;$$

quindi la probabilità cercata è di circa il 4.4%.

**ESEMPIO 7.15** *Sapendo che una cellula in una data coltura cellulare viene infettata in media da 2 virus, qual è la probabilità che una data cellula venga infettata da almeno un virus? Indicando con  $X$  la variabile aleatoria di Poisson che conta il numero di virus che infettano una data cellula, abbiamo*

$$p(X \geq 1) = 1 - p(X = 0) = 1 - \frac{2^0}{0!} e^{-2} = 1 - e^{-2} \simeq 0.86.$$

Quindi c'è una probabilità di circa l'86% che una data cellula sia infettata.

**Osservazione 7.10** La probabilità che una variabile aleatoria di Poisson  $X$  assuma un qualche valore (0 compreso) dev'essere chiaramente 1; quindi dobbiamo avere

$$\sum_{k=0}^{+\infty} p(X = k) = 1. \quad (7.6)$$

La somma a sinistra è una somma infinita (una serie); stiamo sommando  $p(X = k)$  al variare di  $k$  in tutti i numeri naturali. Ricordando (7.5) otteniamo

$$\sum_{k=0}^{+\infty} p(X = k) = e^{-\mu} \sum_{k=0}^{+\infty} \frac{\mu^k}{k!}.$$

Ora, nella Curiosità 5.11 abbiamo fatto vedere che la serie a secondo membro ha come somma proprio  $e^\mu$ , cioè che

$$\sum_{k=0}^{+\infty} \frac{\mu^k}{k!} = e^\mu ; \quad (7.7)$$

quindi

$$\sum_{k=0}^{+\infty} p(X = k) = e^{-\mu} e^\mu = e^0 = 1 ,$$

e (7.6) è verificata. Se non lo fosse stata, avrebbe voluto dire che un qualche passaggio del nostro ragionamento era sbagliato; ma per fortuna non è così.

Un'altra verifica che dobbiamo fare è controllare che il valore atteso di una variabile aleatoria di Poisson di media  $\mu$  sia effettivamente  $\mu$ ; se così non fosse vorrebbe dire che la definizione che abbiamo dato di valore atteso non corrisponde all'idea intuitiva che volevamo formalizzare. Ma in realtà abbiamo

$$\begin{aligned} E(X) &= \sum_{k=0}^{+\infty} p(X = k) \cdot k = \sum_{k=0}^{+\infty} \frac{\mu^k}{k!} e^{-\mu} k = e^{-\mu} \sum_{k=1}^{+\infty} \frac{\mu^k}{(k-1)!} \\ &= \mu e^{-\mu} \sum_{k=1}^{+\infty} \frac{\mu^{k-1}}{(k-1)!} = \mu e^{-\mu} \sum_{h=0}^{+\infty} \frac{\mu^h}{h!} = \mu e^{-\mu} e^\mu \\ &= \mu , \end{aligned}$$

grazie a (7.7), come volevamo.

Calcoliamo ora anche la varianza di  $X$  (e vedrai di nuovo come la manipolazione di sommatorie infinite sia molto simile alla manipolazione delle sommatorie finite):

$$\begin{aligned} \text{Var}(X) &= \sum_{k=0}^{+\infty} p(X = k) \cdot (k - \mu)^2 = \sum_{k=0}^{+\infty} \frac{\mu^k}{k!} e^{-\mu} (k - \mu)^2 \\ &= e^{-\mu} \sum_{k=0}^{+\infty} \frac{\mu^k}{k!} k^2 - 2\mu e^{-\mu} \sum_{k=0}^{+\infty} \frac{\mu^k}{k!} k + \mu^2 e^{-\mu} \sum_{k=0}^{+\infty} \frac{\mu^k}{k!} \\ &= \mu e^{-\mu} \sum_{k=1}^{+\infty} \frac{\mu^{k-1}}{(k-1)!} k - 2\mu^2 + \mu^2 \\ &= \mu e^{-\mu} \sum_{h=0}^{+\infty} \frac{\mu^h}{h!} (h+1) - \mu^2 = \mu e^{-\mu} \sum_{h=0}^{+\infty} \frac{\mu^h}{h!} h + \mu e^{-\mu} \sum_{h=0}^{+\infty} \frac{\mu^h}{h!} - \mu^2 \\ &= \mu^2 + \mu - \mu^2 \\ &= \mu . \end{aligned}$$

In altre parole, *la varianza di una variabile aleatoria di Poisson è uguale alla sua media*. Questo risultato può essere usato per verificare se effettivamente gli eventi sono indipendenti.

**ESEMPIO 7.16** Vuoi studiare la distribuzione degli alberi di leccio in una foresta mista, e cercare di capire se la posizione di ciascun albero è indipendente da quella degli altri oppure se c'è una relazione. Dividi la foresta in molti riquadri di ugual area, e incarichi il tuo assistente di contare il numero di alberi di leccio presenti in ciascun riquadro. Se ciascun albero di leccio nasce in un punto indipendente dalla presenza di altri alberi di leccio, il conteggio del numero di alberi per riquadro dovrebbe seguire una distribuzione di Poisson; in particolare, la media del numero di alberi per riquadro dovrebbe essere circa uguale alla varianza del numero di alberi per riquadro. Se i dati raccolti dal tuo assistente soddisfano questo requisito, l'ipotesi di indipendenza viene rafforzata; ma se la varianza dei dati raccolti dal tuo assistente è sensibilmente diversa dalla media, vuol dire che l'ipotesi d'indipendenza non è coerente con i fatti.

Per la precisione, possono succedere due cose. Se la varianza è sensibilmente maggiore della media, vuol dire che il numero di alberi per riquadro varia molto: ci sono diversi riquadri con tanti alberi, e diversi riquadri con pochi alberi. In altre parole, gli alberi di leccio tendono a raggrupparsi in alcune zone per lasciarle libere altre. Ci possono essere vari motivi per questo fenomeno di raggruppamento (in inglese, *clustering*): i semi si discostano poco dall'albero genitore, il terreno nella foresta non è uniforme, o anche altri. La matematica ha segnalato la presenza di un fenomeno; compito del biologo è ora capire il perché di questo fenomeno, facendo delle ipotesi (che poi possono essere tradotte in un modello matematico e di nuovo verificate come abbiamo appena fatto per l'ipotesi d'indipendenza).

L'altra possibilità è che la varianza sia sensibilmente minore della media. Questo vuol dire che il numero di alberi per riquadro varia poco; in ogni riquadro c'è più o meno lo stesso numero di alberi. In altre parole, gli alberi di leccio sono disposti in modo più o meno uniforme in tutta la foresta. Di nuovo, ci possono essere diversi motivi per questo comportamento (tendenza a occupare tutto lo spazio possibile, competizione per la luce o per l'assorbimento di nutrimento dal terreno, oppure sono stati piantati da qualcuno...) ed è compito del biologo fare ipotesi sul perché; ma la matematica intanto ti assicura che una distribuzione uniforme è estremamente improbabile sia casuale, e richiede la presenza di un qualche tipo di relazione fra i vari alberi.

**ESEMPIO 7.17** Un'altra applicazione interessante della distribuzione di Poisson è una stima della carica elettrica di un elettrone. Sotto certe condizioni, il passaggio degli elettroni in un certo punto per un secondo segue una distribuzione di Poisson. Se in  $t$  secondi passano  $N$  elettroni, la corrente media in un secondo per quel punto è data da

$$I = e \frac{N}{t},$$

dove  $e$  è la carica dell'elettrone. D'altra parte, la deviazione standard della corrente passante per quel punto è  $\sigma_I = e\sigma_N$ , dove  $\sigma_N$  è la deviazione standard del numero di elettroni passanti per quel punto in un secondo. Se effettivamente il passaggio degli elettroni segue una distribuzione di Poisson, la varianza dev'essere uguale alla

media; quindi  $\sigma_N^2 = N/t$ , per cui

$$\sigma_I = e \sqrt{\frac{N}{t}} .$$

Dunque

$$e = \frac{\sigma_I^2}{I}$$

ci fornisce una stima per la carica dell'elettrone.

## 7.4 Variabili aleatorie continue

Supponiamo ora che  $X: \Omega \rightarrow \mathbb{R}$  sia una variabile aleatoria non discreta; come ne calcoliamo media e varianza?

Per giustificare la risposta, procediamo per approssimazioni. Per fissare le idee, supponiamo che  $X$  sia una misurazione, che effettuiamo con strumenti a poco a poco più sofisticati. Il primo strumento è veramente rozzo, e fornisce solo risultati arrotondati all'unità superiore. Per esempio, una misura di 3 significa che il valore vero della misura è in realtà compreso fra 2 e 3. Quindi, se indichiamo con  $X_1$  la misura fornita da questo strumento, abbiamo  $p(X_1 = 3) = p(2 < X \leq 3)$  e, in generale,

$$\forall j \in \mathbb{Z} \quad p(X_1 = j) = p(j - 1 < X \leq j) .$$

Possiamo rappresentare le probabilità associate a  $X_1$  costruendo un istogramma composto da colonne di base gli intervalli  $(j - 1, j]$  e altezza  $p_j = p(j - 1 < X \leq j)$ ; vedi la Fig. 7.2.

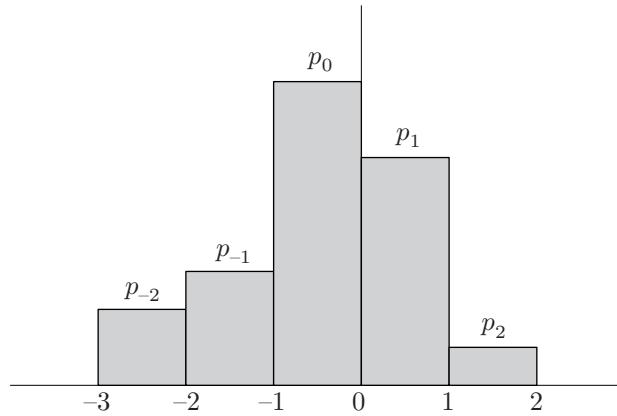


Figura 7.2 .

L'area della colonna  $j$ -esima è  $p_j = p_j \Delta j$  (dove  $\Delta j = j - (j - 1) = 1$ ) e rappresenta la probabilità che la misura  $X$  sia compresa fra  $j - 1$  e  $j$ . Inoltre, la media e la varianza di  $X_1$  sono date da

$$E(X_1) = \mu_1 = \sum_j p_j j \Delta j \quad \text{e} \quad \text{Var}(X_1) = \sum_j p_j (j - \mu_1)^2 \Delta j ,$$

dove abbiamo usato il fatto che  $\Delta j = 1$ , e possono essere considerate come una prima approssimazione della media e della varianza di  $X$ .

Abbiamo ricevuto un nuovo finanziamento, e ci possiamo permettere uno strumento migliore  $X_2$ , con una precisione di mezza unità; per esempio, una misura di  $1/2$  con  $X_2$  significa che il valore vero è compreso fra  $0$  e  $1/2$ . Quindi  $p(X_2 = 1/2) = p(0 < X \leq 1/2)$ ; in generale, se indichiamo con  $x_j = j/2$  un generico valore di  $X_2$ , abbiamo

$$\forall j \in \mathbb{Z} \quad p(X_2 = x_j) = p(x_{j-1} < X \leq x_j) .$$

Rappresentiamo di nuovo le probabilità associate a  $X_2$  con un istogramma in modo però che sia l'*area* delle colonne di base l'intervallo  $(x_{j-1}, x_j]$  a rappresentare la probabilità che la misura  $X$  sia compresa fra  $x_{j-1}$  e  $x_j$ . Se indichiamo con  $f_j$  l'altezza della colonna  $j$ -esima, dobbiamo avere

$$p_j = f_j \Delta x_j ,$$

dove  $p_j = p(X_2 = x_j)$  e  $\Delta x_j = x_j - x_{j-1} = 1/2$ ; quindi

$$f_j = \frac{p(x_{j-1} < X \leq x_j)}{\Delta x_j} = 2p_j .$$

In particolare, se  $a$  e  $b$  sono due possibili valori di  $X_2$ , la probabilità che la misura  $X$  cada fra  $a$  e  $b$  è data da

$$p(a < X \leq b) = \sum_h p_h = \sum_h f_h \Delta x_h ,$$

dove la somma è limitata a quegli  $h$  per cui  $a < x_h \leq b$ .

Infine, la media e la varianza di  $X_2$  sono date da

$$\begin{aligned} E(X_2) &= \mu_2 = \sum_j p_j x_j = \sum_j f_j x_j \Delta x_j \\ \text{Var}(X_2) &= \sum_j p_j (x_j - \mu_2)^2 = \sum_j f_j (x_j - \mu_2)^2 \Delta x_j , \end{aligned}$$

e possono essere considerate come una migliore approssimazione della media e della varianza di  $X$ .

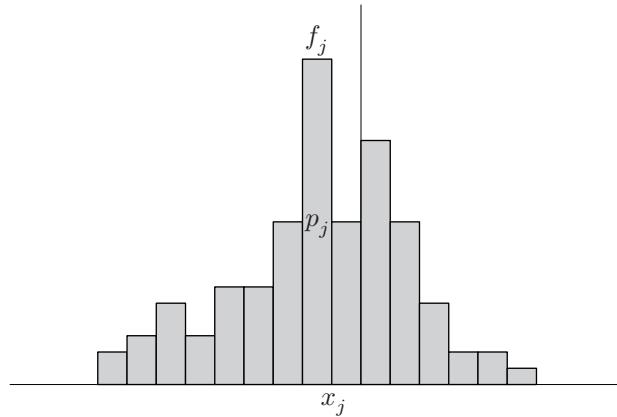
A questo punto è chiaro come si procede. Per ogni  $n \geq 1$  ci procuriamo uno strumento  $X_n$  con precisione  $1/n$ . Per  $x_j = j/n$ , poniamo

$$p_j = p(X_n = x_j) = p(x_{j-1} < X \leq x_j)$$

e

$$f_j = \frac{p(x_{j-1} < X \leq x_j)}{\Delta x_j} = np_j ,$$

e costruiamo l'istogramma con colonne di altezza  $f_j$  sopra  $(x_{j-1}, x_j]$  in modo che sia l'area delle colonne a rappresentare la probabilità che  $X$  cada fra  $x_{j-1}$  e  $x_j$ ; vedi la Fig. 7.3.



**Figura 7.3** .

Quindi se  $a$  e  $b$  sono due possibili valori di  $X_n$ , la probabilità che la misura  $X$  cada fra  $a$  e  $b$  è data dall'area dell'istogramma sull'intervallo  $(a, b]$ , cioè

$$p(a < X \leq b) = \sum_h p_h = \sum_h f_h \Delta x_h , \quad (7.8)$$

dove la somma è limitata a quegli  $h$  per cui  $a < x_h \leq b$ .

Inoltre, la media e la varianza di  $X_n$  sono date da

$$\begin{aligned} E(X_n) &= \mu_n = \sum_j p_j x_j = \sum_j x_j f_j \Delta x_j \\ \text{Var}(X_n) &= \sum_j p_j (x_j - \mu_n)^2 = \sum_j (x_j - \mu_n)^2 f_j \Delta x_j , \end{aligned} \quad (7.9)$$

e possono essere considerate come sempre migliori approssimazioni della media e della varianza di  $X$ ; nota che stavolta la somma è estesa a tutti i possibili valori di  $j$ , cioè con  $x_j$  che varia da  $-\infty$  a  $+\infty$ .

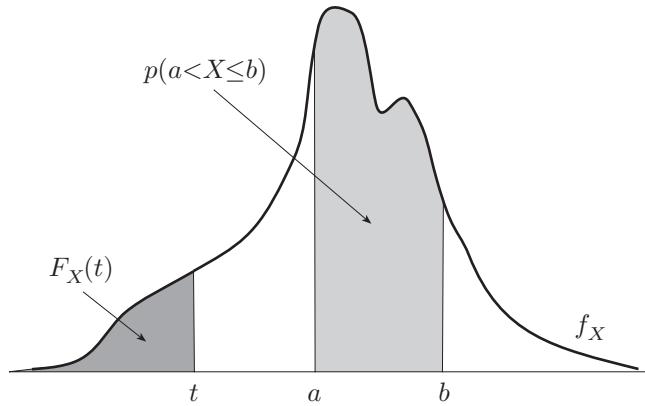
A questo punto non ci rimane che passare al limite per  $n$  che tende all'infinito. Diremo che la variabile aleatoria  $X$  è *continua* se per ogni  $x \in \mathbb{R}$  esiste il limite

$$f_X(x) = \lim_{\Delta x \rightarrow 0^+} \frac{p(x - \Delta x < X \leq x)}{\Delta x}; \quad (7.10)$$

la funzione  $f_X: \mathbb{R} \rightarrow \mathbb{R}$  è detta *densità di probabilità* della variabile aleatoria  $X$ . Ricordando la definizione di integrale vediamo subito che passando al limite (7.8) ci dice che la densità di probabilità è tale che

$$\forall a < b \quad p(a < X \leq b) = \int_a^b f_X(x) dx; \quad (7.11)$$

vedi la Fig. 7.4. In altre parole, *la probabilità che  $X$  cada fra  $a$  e  $b$  è uguale all'area del sottografo della densità di probabilità  $f_X$  sull'intervallo  $[a, b]$ .*



**Figura 7.4** Densità di probabilità.

In particolare, la funzione di distribuzione  $F_X$  di  $X$  è data da

$$F_X(t) = \int_{-\infty}^t f_X(x) dx;$$

di conseguenza, il primo teorema fondamentale del calcolo ci dice che *la densità di probabilità di una variabile aleatoria continua è la derivata della funzione di distribuzione*:

$$f_X = \frac{dF_X}{dx}.$$

**Osservazione 7.11** Più precisamente,  $f_X$  è la derivata sinistra della funzione di distribuzione, come si vede scrivendo (7.10) nella forma

$$f_X(x) = \lim_{\Delta x \rightarrow 0^+} \frac{F_X(x) - F_X(x - \Delta x)}{\Delta x} = \lim_{h \rightarrow 0^-} \frac{F_X(x + h) - F_X(x)}{h}.$$

*Osservazione 7.12* Dalla (7.10) si vede subito che  $f_X(x) \geq 0$  per ogni  $x \in \mathbb{R}$ ; inoltre (7.11) ci dice che

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

CURIOSITÀ 7.2 Viceversa, se abbiamo una funzione  $f: \mathbb{R} \rightarrow \mathbb{R}$  integrabile su ogni intervallo (o semiretta), con  $f(x) \geq 0$  sempre e tale che  $\int_{-\infty}^{+\infty} f(x) dx = 1$  allora possiamo definire una distribuzione di probabilità  $p: \mathcal{A} \rightarrow [0, 1]$  su  $\mathbb{R}$  ponendo

$$p(A) = \int_A f(x) dx,$$

dove  $\mathcal{A}$  è la famiglia dei sottoinsiemi di  $\mathbb{R}$  su cui  $f$  è integrabile (essenzialmente,  $\mathcal{A}$  è ottenuta prendendo tutte le possibili unioni e intersezioni finite o infinite numerabili di intervalli). Allora le proprietà dell'integrale permettono di verificare facilmente che  $p$  soddisfa gli assiomi (P1)–(P3) che abbiamo introdotto nella Sezione 2.5, per cui  $p$  è effettivamente una distribuzione di probabilità su  $\mathbb{R}$ . Per la precisione,  $p$  soddisfa anche l'assioma (P4) citato nella Curiosità 2.1; ma per vederlo occorre definire l'integrale in modo più generale di quanto abbiamo fatto noi, usando la *teoria dell'integrazione di Lebesgue*.

*Osservazione 7.13* Se  $X$  è una variabile aleatoria continua e  $a \in \mathbb{R}$ , abbiamo

$$\begin{aligned} p(X = a) &= F_X(a) - \lim_{t \rightarrow a^-} F_X(t) = \int_{-\infty}^a f_X(x) dx - \lim_{t \rightarrow a^-} \int_{-\infty}^t f_X(x) dx \\ &= \lim_{t \rightarrow a^-} \int_t^a f_X(x) dx = \int_a^a f_X(x) dx \\ &= 0. \end{aligned}$$

Questo non deve sorprenderti:  $p(X = a)$  è la probabilità che si riesca a determinare che la misura  $X$  vale esattamente  $a$  con precisione infinita, controllando tutte le infinite cifre decimali. Nella pratica sperimentale questo non succede mai (e nella pratica matematica ha probabilità zero di succedere, che è un modo più forbito di dire la stessa cosa); le misure sono sempre ottenute a meno di un errore, che equivale a dire che il valore “vero” appartiene a un determinato intervallo. Quindi nella pratica sperimentale siamo interessati alla probabilità che la misura cada in un dato intervallo, che è esattamente ciò che possiamo calcolare usando la densità di probabilità.

Infine, passando al limite su  $n$  che tende all'infinito in (7.9), partendo dalle medie e varianze delle approssimazioni  $X_n$  troviamo finalmente la definizione di media (o *valor medio*, *valore atteso*, o *speranza*) e varianza di una variabile aleatoria continua:

$$E(X) = \mu_X = \int_{-\infty}^{+\infty} x f_X(x) dx, \quad \text{Var}(X) = \sigma_X^2 = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx.$$

La deviazione standard  $\sigma_X$  di una variabile aleatoria continua è come al solito la radice quadrata della varianza:

$$\text{DS}(X) = \sigma_X = \sqrt{\sigma_X^2}.$$

*Osservazione 7.14* Il valor medio di una variabile aleatoria continua soddisfa proprietà analoghe a quelle della media di numeri. Per esempio, è l'unico  $\mu \in \mathbb{R}$  per cui l'integrale degli “errori”  $x - \mu$ , pesati in base alla densità di probabilità  $f_X$ , è nullo:

$$\int_{-\infty}^{+\infty} (x - \mu) f_X(x) dx = 0 \iff \mu = \mu \int_{-\infty}^{+\infty} f_X(x) dx = \int_{-\infty}^{+\infty} x f_X(x) dx = \mu_X.$$

Si può anche vedere che  $\mu_X$  minimizza l'integrale degli scarti quadratici  $(x - \mu)^2$ , pesati in base alla densità di probabilità  $f_X$ . Consideriamo la funzione  $s: \mathbb{R} \rightarrow \mathbb{R}$  data da

$$s(y) = \int_{-\infty}^{+\infty} (x - y)^2 f_X(x) dx.$$

Sviluppando il quadrato otteniamo

$$\begin{aligned} s(y) &= y^2 \int_{-\infty}^{+\infty} f_X(x) dx - 2y \int_{-\infty}^{+\infty} x f_X(x) dx + \int_{-\infty}^{+\infty} x^2 f_X(x) dx \\ &= y^2 - 2\mu_X y + \int_{-\infty}^{+\infty} x^2 f_X(x) dx. \end{aligned}$$

Quindi  $s$  è una funzione quadratica con coefficiente direttivo positivo, per cui ammette un unico minimo in  $y_0 = -b/2a = \mu_X$ , come voluto.

**CURIOSITÀ 7.3** Si può dimostrare che il valor medio della somma di due variabili aleatorie (entrambe discrete o continue)  $X_1$  e  $X_2$  è sempre uguale alla somma dei valori medi:

$$E(X_1 + X_2) = E(X_1) + E(X_2).$$

Una proprietà importante delle variabili aleatorie *indipendenti* è che anche la varianza della somma è uguale alla somma delle varianze: se  $X_1$  e  $X_2$  sono variabili aleatorie (entrambe discrete o continue) indipendenti allora

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2).$$

Nota che la funzione di distribuzione (o la densità di probabilità nel caso continua) della somma *non* è mai uguale alla somma delle funzioni di distribuzione (o delle densità di probabilità). Se lo fosse, ci sarebbero (perché?) degli eventi con probabilità maggiore di 1...

**CURIOSITÀ 7.4** Se  $X$  è una variabile aleatoria con funzione di distribuzione  $F_X$  continua, la *mediana*  $M_X$  di  $X$  è il punto di mezzo dell'intervallo  $F_X^{-1}(1/2)$ ; in altre parole, la probabilità che  $X$  assuma un valore minore o uguale a  $M_X$  è uguale alla probabilità che  $X$  assuma un valore maggiore di  $M_X$ . Se  $F_X$  non è continua, la mediana potrebbe non esistere; vedi l'Esempio 7.9. Anche se esiste, la mediana potrebbe essere diversa dal valor medio; vedi la Curiosità 7.5.

Le prossime (e ultime) tre sezioni di questo capitolo presentano i principali esempi di densità di probabilità di variabili aleatorie continue, calcolandone il valore atteso e la varianza.

## 7.5 Distribuzione uniforme

Nel caso discreto, una delle distribuzioni di probabilità più comuni era quella uniforme: tutti gli eventi semplici (o tutti i valori di una variabile aleatoria discreta) hanno la stessa probabilità di verificarsi. Nel caso continuo, dire che tutti i valori di una variabile aleatoria continua hanno la stessa probabilità di verificarsi non è particolarmente significativo, in quanto, come abbiamo visto nell’Osservazione 7.13, la probabilità che una variabile aleatoria continua  $X$  ammetta il valore esatto  $a \in \mathbb{R}$  è zero. Per formalizzare l’idea intuitiva dell’equiprobabilità di tutti i valori di  $X$  dobbiamo procedere in modo lievemente diverso: l’idea è che la probabilità che  $X$  assuma il valore  $a$  a meno di un errore assoluto pari a  $e$ , cioè  $p(a - e \leq X \leq a + e)$ , dipende solo dall’errore  $e$  (cioè dall’ampiezza dell’intervallo di valori considerato) e non da  $a$ . Che debba dipendere dall’errore è evidente: aumentando l’ampiezza dell’intervallo dei valori possibili la probabilità deve aumentare. L’uniformità è espressa dall’indipendenza da  $a$ .

**ESEMPIO 7.18** Siamo quasi alla fine delle dispense, per cui decidi che puoi fare a meno del tuo assistente, e lo depositi in mezzo al deserto in una giornata coperta, per cui non può neppure usare il sole per orientarsi. Il tuo assistente ha decisamente voglia di vivere (e magari vendicarsi), per cui invece di mettersi seduto a riflettere sui suoi errori sceglie una direzione a caso e s’incammina. Che probabilità ha di aver scelto l’unica (con l’approssimazione di un grado) direzione che lo porta verso l’oasi più vicina? Per semplicità, identifichiamo ciascuna direzione con l’angolo che forma con l’est; quindi se misuriamo gli angoli in gradi le possibili direzioni corrispondono all’intervallo  $(0, 360]$ . Se indichiamo con  $X$  la variabile aleatoria “direzione in cui si incammina l’assistente”, dobbiamo calcolare  $p(x_0 - 1 < X \leq x_0 + 1)$ , dove  $x_0 \in (0, 360]$  è l’unica direzione giusta, sapendo che l’assistente ha la stessa probabilità di scegliere qualsiasi direzione, in quanto non ha alcun punto di riferimento su cui basarsi. Dal discorso fatto prima, questa probabilità deve dipendere solo dal fatto che l’intervallo  $(x_0 - 1, x_0 + 1]$  è lungo 2, e dev’essere la stessa per tutti gli intervalli di lunghezza 2. Suddividiamo allora l’intervallo  $(0, 360]$  in 180 intervalli di lunghezza 2; gli assiomi della probabilità e l’ipotesi di uniformità ci dicono che

$$\begin{aligned} 1 &= p(X \in (0, 360]) = p(X \in (0, 2]) + p(X \in (2, 4]) + \cdots + p(X \in (358, 360]) \\ &= 180 p(x_0 - 1 < X \leq x_0 + 1), \end{aligned}$$

per cui la risposta è

$$p(x_0 - 1 < X \leq x_0 + 1) = \frac{1}{180} \simeq 0.6\%,$$

decisamente piccolina.

Vediamo di formalizzare il ragionamento fatto in questo esempio. Diremo che una variabile aleatoria  $X$  è *limitata* in  $(a, b]$  se i valori di  $X$  sono tutti contenuti nell'intervallo  $(a, b]$ . In particolare, abbiamo (perché?)

$$F_X(a) = 0 \quad \text{e} \quad F_X(b) = 1 ,$$

da cui segue (perché?) che  $F_X(t) = 0$  per ogni  $t \leq a$  e  $F_X(t) = 1$  per ogni  $t \geq b$ . Diremo poi che una variabile aleatoria  $X$  limitata in  $(a, b]$  ha una *distribuzione uniforme* sull'intervallo  $(a, b]$  se esiste una costante  $c > 0$  tale che

$$p(x_1 < X \leq x_2) = c(x_2 - x_1)$$

per ogni  $a \leq x_1 < x_2 \leq b$ . Questa formula chiaramente esprime l'idea di uniformità che abbiamo discusso prima. Inoltre, prendendo  $x_1 = a$  e  $x_2 = b$  è facile trovare quanto deve valere  $c$ :

$$1 = p(a < X \leq b) = c(b - a) \implies c = \frac{1}{b - a} .$$

Quindi  $X$  ha una distribuzione uniforme su  $(a, b]$  se e solo se

$$p(x_1 < X \leq x_2) = \frac{x_2 - x_1}{b - a}$$

per ogni  $a \leq x_1 < x_2 \leq b$ . In particolare, la funzione di distribuzione di  $X$  è data da (perché?)

$$F_X(t) = \begin{cases} 0 & \text{se } t \leq a , \\ \frac{t - a}{b - a} & \text{se } a \leq t \leq b , \\ 1 & \text{se } t \geq b . \end{cases}$$

La funzione di distribuzione è derivabile a sinistra, per cui  $X$  è una variabile aleatoria continua con densità di probabilità

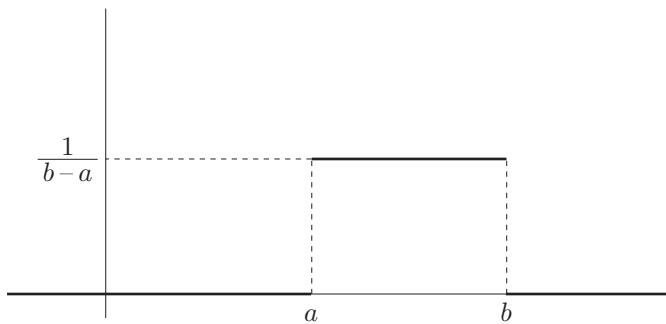
$$f_X(x) = \begin{cases} 0 & \text{se } x \leq a , \\ \frac{1}{b - a} & \text{se } a < x \leq b , \\ 0 & \text{se } t > b ; \end{cases}$$

vedi la Fig. 7.5. Nota che l'uniformità della distribuzione probabilità si traduce nel fatto che la densità di probabilità è costante nell'intervallo dei valori di  $X$ .

Come esercizio, calcoliamo il valor medio e la varianza di una variabile aleatoria con distribuzione uniforme sull'intervallo  $(a, b]$ . Il valor medio è

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_a^b \frac{x}{b - a} dx = \frac{1}{b - a} \frac{b^2 - a^2}{2} = \frac{a + b}{2} ;$$

(nota che siamo passati dall'integrale improprio su tutta la retta all'integrale definito su  $[a, b]$  perché la densità di probabilità è nulla al di fuori di quell'intervallo).



**Figura 7.5** Densità di probabilità uniforme.

Dunque il valor medio di  $X$  è il punto medio dell’intervallo  $(a, b]$ ; del resto, siccome tutti i valori di  $(a, b]$  sono equiprobabili, il valor medio non poteva essere altro che il punto di mezzo dell’intervallo.

Infine, la varianza:

$$\begin{aligned}\text{Var}(X) &= \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \left[ \int_a^b x^2 dx - (a+b) \int_a^b x dx + \frac{(a+b)^2}{4} \int_a^b dx \right] \\ &= \frac{1}{b-a} \left[ \frac{b^3 - a^3}{3} - \frac{(a+b)(b^2 - a^2)}{2} + \frac{(a+b)^2(b-a)}{4} \right] \\ &= \frac{b^2 + ab + a^2}{3} - \frac{b^2 + 2ab + a^2}{4} \\ &= \frac{(b-a)^2}{12},\end{aligned}$$

per cui la deviazione standard di una variabile aleatoria con distribuzione uniforme su  $(a, b]$  è

$$\sigma_X = \frac{b-a}{2\sqrt{3}}.$$

## 7.6 Distribuzione esponenziale

Un’altra distribuzione comune per variabili aleatorie continue si ricava studiando variabili aleatorie discrete di Poisson: per l’esattezza, vogliamo studiare il tempo che intercorre fra due eventi consecutivi in un fenomeno di Poisson.

Supponiamo quindi di voler esaminare un fenomeno che soddisfa le ipotesi della Sezione 7.3, e indichiamo con  $X$  la variabile aleatoria che misura il tempo fra due eventi successivi. Vogliamo trovare la funzione di distribuzione e la densità di probabilità (se esiste) di  $X$ .

**Osservazione 7.15** Siccome gli eventi in un fenomeno di Poisson sono del tutto indipendenti l'uno dall'altro, possiamo scegliere a piacimento l'istante in cui iniziamo a osservare il fenomeno. In particolare, possiamo iniziare a osservare il fenomeno quando avviene un evento, e misurare il tempo fra due eventi successivi diventa la stessa cosa di misurare quanto tempo passa senza eventi.

Indichiamo con  $\mu$  il valor medio di eventi in un intervallo di tempo unitario. Allora in un tempo  $t$  avremo in media  $\mu t$  eventi. Quindi se indichiamo con  $Y(t)$  la variabile aleatoria discreta che conta il numero di eventi nell'intervallo di tempo  $[0, t]$  i conti fatti nella Sezione 7.3 ci dicono che

$$\forall k \in \mathbb{N} \quad p(Y(t) = k) = \frac{(\mu t)^k}{k!} e^{-\mu t} .$$

Ora, dire che il primo evento avviene dopo un tempo  $t > 0$  è esattamente equivalente a dire che non ci sono eventi nell'intervallo  $[0, t]$ ; quindi

$$\forall t \geq 0 \quad p(X > t) = p(Y(t) = 0) = \frac{(\mu t)^0}{0!} e^{-\mu t} = e^{-\mu t} .$$

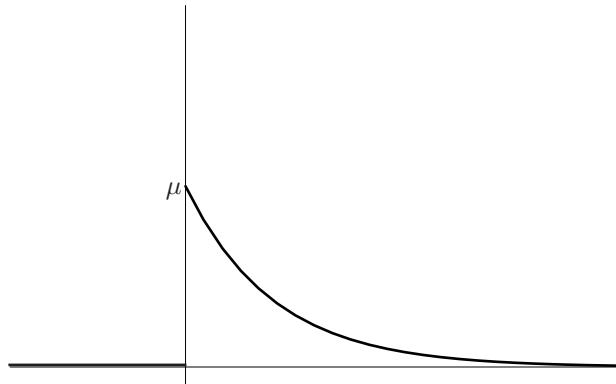
D'altra parte, il primo evento è sicuramente dopo il tempo 0, per cui  $p(X > t) = 1$  per ogni  $t < 0$ . Siccome  $F_X(t) = p(X \leq t) = 1 - p(X > t)$ , abbiamo trovato la funzione di distribuzione cercata:

$$F_X(t) = \begin{cases} 1 - e^{-\mu t} & \text{se } t > 0 , \\ 0 & \text{se } t \leq 0 . \end{cases}$$

Derivando otteniamo anche la densità di probabilità:

$$f_X(x) = \begin{cases} \mu e^{-\mu x} & \text{se } x > 0 , \\ 0 & \text{se } x \leq 0 ; \end{cases} \quad (7.12)$$

vedi la Fig. 7.6. Una variabile aleatoria continua con densità di probabilità data da (7.12) è detta *esponenziale di durata media  $\mu$* .



**Figura 7.6** Densità di probabilità esponenziale.

Non è difficile calcolare il valor medio di una variabile aleatoria esponenziale, integrando per parti:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} xf_X(x) dx = \int_0^{+\infty} x\mu e^{-\mu x} dx = -xe^{-\mu x} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-\mu x} dx \\ &= (0 - 0) - \frac{1}{\mu} e^{-\mu x} \Big|_0^{+\infty} = \frac{1}{\mu}. \end{aligned}$$

Non è un risultato sorprendente: se nell'unità di tempo ci sono in media  $\mu$  eventi, l'intervallo di tempo medio fra due eventi è ragionevole sia  $1/\mu$ .

Calcolando il valor medio abbiamo trovato che

$$\int_0^{+\infty} x\mu e^{-\mu x} dx = \frac{1}{\mu}.$$

Ricordando che inoltre  $\int_0^{+\infty} \mu e^{-\mu x} dx = 1$ , integrando di nuovo per parti possiamo calcolare la varianza:

$$\begin{aligned} \text{Var}(X) &= \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx = \int_0^{+\infty} \left(x - \frac{1}{\mu}\right)^2 \mu e^{-\mu x} dx \\ &= \int_0^{+\infty} x^2 \mu e^{-\mu x} dx - \frac{2}{\mu} \int_0^{+\infty} x \mu e^{-\mu x} dx + \frac{1}{\mu^2} \int_0^{+\infty} \mu e^{-\mu x} dx \\ &= -x^2 e^{-\mu x} \Big|_0^{+\infty} + 2 \int_0^{+\infty} x e^{-\mu x} dx - \frac{2}{\mu^2} + \frac{1}{\mu^2} \\ &= \frac{2}{\mu^2} - \frac{1}{\mu^2} = \frac{1}{\mu^2}. \end{aligned}$$

Quindi la deviazione standard di una variabile aleatoria esponenziale di durata media  $\mu$  è

$$\sigma_X = \frac{1}{\mu} = \mu_X.$$

**CURIOSITÀ 7.5** Se  $X$  è una variabile aleatoria esponenziale di durata media  $\mu$ , si ha  $F_X(t) = 1/2$  se e solo se  $e^{-\mu t} = 1/2$ , per cui la mediana (vedi la Curiosità 7.4) di  $X$  è

$$M_X = \frac{\log 2}{\mu} < \mu_X.$$

## 7.7 Distribuzione normale

In natura, moltissime variabili aleatorie continue hanno una densità di probabilità con una tipica forma a campana. Capita così frequentemente che in assenza di informazioni che suggeriscano il contrario, in statistica si assume spesso che le misure che si vogliono studiare abbiano questo tipo di densità di probabilità. In

questa sezione introdurremo queste funzioni a campana, e cercheremo di spiegare perché sono così frequenti.

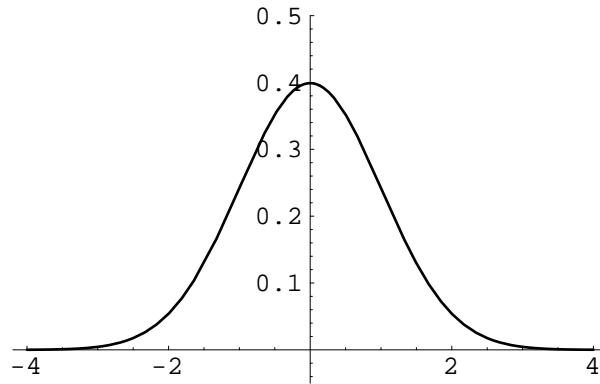
La funzione da cui nasce tutto è la *Gaussiana*  $G_{0,1}: \mathbb{R} \rightarrow \mathbb{R}$  data da

$$G_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} ;$$

la Fig. 7.7 ne contiene il grafico. Il fattore  $1/\sqrt{2\pi}$  serve ad assicurare (vedi la Curiosità 6.9) che

$$\int_{-\infty}^{+\infty} G_{0,1}(x) dx = 1 ; \quad (7.13)$$

quindi  $G_{0,1}$  può essere la densità di probabilità di una variabile aleatoria continua.



**Figura 7.7** La Gaussiana.

Supponiamo allora che  $X$  sia una variabile aleatoria continua di densità di probabilità  $G_{0,1}$ . Siccome

$$\frac{de^{-x^2/2}}{dx} = -xe^{-x^2/2} ,$$

il valore atteso di  $X$  è

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} xe^{-x^2/2} dx = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \Big|_{-\infty}^{+\infty} = 0 .$$

Un'integrazione per parti e (7.13) ci permettono di calcolare anche la varianza di  $X$ :

$$\begin{aligned} \text{Var}(X) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-x^2/2} dx = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \cdot (-xe^{-x^2/2}) dx \\ &= -\frac{1}{\sqrt{2\pi}} \left[ xe^{-x^2/2} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} e^{-x^2/2} dx \right] \\ &= 1 . \end{aligned}$$

Diremo quindi che una variabile aleatoria continua  $X$  è *normale* di media 0 e varianza 1 se la densità di probabilità di  $X$  è  $G_{0,1}$ .

*Osservazione 7.16* La funzione di distribuzione di una variabile aleatoria  $X$  normale di media 0 e varianza 1 è la funzione  $\Phi: \mathbb{R} \rightarrow [0, 1]$  data da

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-x^2/2} dx .$$

Questa funzione *non* è esprimibile tramite funzioni elementari (a parte che come integrale di  $e^{-x^2/2} \dots$ ). Un paio di osservazioni possiamo però farle lo stesso. Prima di tutto, la funzione  $G_{0,1}$  è una funzione pari, cioè con grafico simmetrico rispetto all'asse delle ascisse. Di conseguenza,

$$\forall t > 0 \quad \frac{1}{\sqrt{2\pi}} \int_{-t}^0 e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-x^2/2} dx .$$

Da questo segue in particolare (perché?) che

$$\Phi(0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 e^{-x^2/2} dx = \frac{1}{2} = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} e^{-x^2/2} dx .$$

Più in generale se  $t \geq 0$  abbiamo

$$\Phi(t) = 1 - \frac{1}{\sqrt{2\pi}} \int_t^{+\infty} e^{-x^2/2} dx = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-t} e^{-x^2/2} dx = 1 - \Phi(-t) ,$$

cioè

$$\forall t \in \mathbb{R} \quad \Phi(t) + \Phi(-t) = 1 . \tag{7.14}$$

Molto più di questo non si può fare; per questo motivo in molti testi di statistica si trovano delle tavole di valori approssimati di  $\Phi$  — e molti calcolatori sono ampiamente forniti di programmi per calcolarla numericamente. I valori che a noi interesseranno di più sono

$$\Phi(1) \simeq 0.84135 , \quad \Phi(2) \simeq 0.9772 , \quad \Phi(3) \simeq 0.9986 . \tag{7.15}$$

Da questo usando (7.14) si trova

$$p(-1 \leq X \leq 1) \simeq 68.27\% , \quad p(-2 \leq X \leq 2) \simeq 95.45\% , \quad p(-3 \leq X \leq 3) \simeq 99.73\% .$$

Si sa pure che  $p(-x \leq X \leq x) = 50\%$  per  $x \simeq 0.674$ , mentre  $p(-x \leq X \leq x) = 95\%$  per  $x \simeq 1.960$  e  $p(-x \leq X \leq x) = 99\%$  per  $x \simeq 2.576$ .

La varianza misura quanto è dispersa una variabile aleatoria rispetto alla media. Ora, la funzione  $x \mapsto G_{0,1}(x/\sigma)$  ha un grafico più allargato rispetto a quello di  $G_{0,1}$  se  $\sigma > 1$ , e più stretto se  $\sigma < 1$ ; quindi ci aspettiamo che una variabile aleatoria con densità di probabilità simile a  $G_{0,1}(x/\sigma)$  abbia varianza maggiore di 1 se  $\sigma > 1$ , e minore di 1 se  $\sigma < 1$ . Ora,  $G_{0,1}(x/\sigma)$  non può essere una densità di probabilità perché, effettuando la sostituzione  $y = x/\sigma$ , abbiamo

$$\int_{-\infty}^{+\infty} G_{0,1}(x/\sigma) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2\sigma^2} dx = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-y^2/2} dy = \sigma ;$$

dunque

$$G_{0,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$

ha integrale su  $\mathbb{R}$  uguale a 1, e quindi può essere una densità di probabilità. Siccome

$$\int_{-\infty}^{+\infty} xG_{0,\sigma}(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} xe^{-x^2/2\sigma^2} dx = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} ye^{-y^2/2} dy = 0 ,$$

e

$$\int_{-\infty}^{+\infty} x^2 G_{0,\sigma}(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-x^2/2\sigma^2} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 e^{-y^2/2} dy = \sigma^2 ,$$

una variabile aleatoria continua con densità di probabilità  $G_{0,\sigma}$  ha valor medio 0 e varianza  $\sigma^2$ .

Possiamo anche spostare la media. È molto ragionevole immaginare che per spostare la media basti traslare il grafico della densità di probabilità, cioè traslare le ascisse (operazione che chiaramente non cambia l'area del sottograffico). Quindi proviamo a porre

$$G_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} .$$

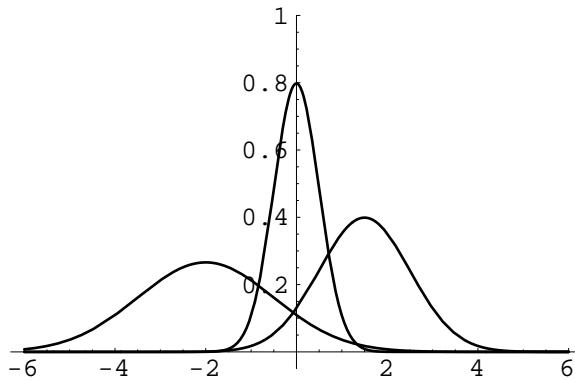
Usando la sostituzione  $y = (x - \mu)/\sigma$  si vede subito che

$$\begin{aligned} \int_{-\infty}^{+\infty} xG_{\mu,\sigma}(x) dx &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} xe^{-(x-\mu)^2/2\sigma^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma y + \mu) e^{-y^2/2} dy \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} ye^{-y^2/2} dy + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-y^2/2} dy \\ &= \mu , \end{aligned}$$

e

$$\begin{aligned} \int_{-\infty}^{+\infty} x^2 G_{\mu,\sigma}(x) dx &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x - \mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 e^{-y^2/2} dy = \sigma^2 ; \end{aligned}$$

quindi una variabile aleatoria continua con densità di probabilità  $G_{\mu,\sigma}$  ha valor medio  $\mu$  e varianza  $\sigma^2$ . Per questo motivo diremo che una variabile aleatoria continua  $X$  è *normale* di media  $\mu$  e varianza  $\sigma^2$  se la densità di probabilità di  $X$  è  $G_{\mu,\sigma}$ . La Fig. 7.8 mostra il grafico di  $G_{\mu,\sigma}$  per vari valori di  $\mu$  e  $\sigma$ .



**Figura 7.8** Varie Gaussiane.

*Osservazione 7.17* La funzione di distribuzione di una variabile aleatoria  $X$  normale di media  $\mu$  e varianza  $\sigma^2$  è data da

$$\begin{aligned} F_X(t) &= \int_{-\infty}^t G_{\mu,\sigma}(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^t e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(t-\mu)/\sigma} e^{-y^2/2} dy = \Phi\left(\frac{t-\mu}{\sigma}\right). \end{aligned}$$

Quindi (7.14) e (7.15) implicano che

$$\begin{aligned} F_X(\mu + \sigma) &= 1 - F_X(\mu - \sigma) = \Phi(1) \simeq 0.84135, \\ F_X(\mu + 2\sigma) &= 1 - F_X(\mu - 2\sigma) = \Phi(2) \simeq 0.9772, \\ F_X(\mu + 3\sigma) &= 1 - F_X(\mu - 3\sigma) = \Phi(3) \simeq 0.9986. \end{aligned}$$

Quindi

$$p(\mu - \sigma \leq X \leq \mu + \sigma) = F_X(\mu + \sigma) - F_X(\mu - \sigma) \simeq 68.27\%,$$

cioè *circa il 68.27% delle volte il valore di una variabile aleatoria normale cade nell'intervallo  $[\mu - \sigma, \mu + \sigma]$* . In modo analogo si vede che il 95.45% delle volte il valore di una variabile aleatoria normale cade nell'intervallo  $[\mu - 2\sigma, \mu + 2\sigma]$ , e che il 99.73% delle volte cade nell'intervallo  $[\mu - 3\sigma, \mu + 3\sigma]$ .

Il motivo per cui le variabili aleatorie normali sono così frequenti in natura è il *Teorema del limite centrale*, che abbiamo già citato nella Sezione 3.8. Esistono

varie versioni di questo risultato; qui voglio citarne quella forse più significativa per le applicazioni biologiche.

Supponiamo di avere una successione  $X_1, X_2, \dots, X_n, \dots$  di variabili aleatorie, ognuna con la propria media  $\mu_i$  e la propria varianza  $\sigma_i^2$ . Supponiamo che:

(C1) le  $X_i$  siano tutte limitate;

(C2) siano tutte *indipendenti*;

(C3) le varianze  $\sigma_i^2$  non siano troppo piccole, in modo da avere  $\sum_i \sigma_i^2 = +\infty$ ; per esempio, è sufficiente che la successione  $\{\sigma_i^2\}$  non tenda a zero (ma basta anche molto meno).

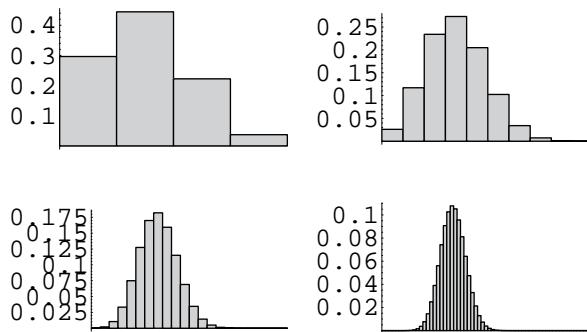
È importante notare che *non supponiamo nient'altro sulla funzione di distribuzione delle  $X_i$* . Inoltre, le  $X_i$  possono essere sia discrete sia continue. Indichiamo con  $S_n = X_1 + \dots + X_n$  la variabile aleatoria somma delle prime  $n$  variabili aleatorie, e con  $M_n = \frac{1}{n}S_n$  la media. Allora il *Teorema del limite centrale* ci dice che *sia  $S_n$  che  $M_n$  per  $n \rightarrow +\infty$  tendono* (in un senso che preciseremo fra un attimo) *a una variabile aleatoria normale*.

Perché questo risultato spiega l'abbondanza delle variabili aleatorie normali in natura? Il motivo è che molti fenomeni possono (almeno in prima approssimazione) essere rappresentati come il risultato dell'interazione di molte cause indipendenti che agiscono contemporaneamente. Se l'interazione è additiva (cioè gli effetti si sommano) il fenomeno che noi vediamo è dovuto alla somma di queste cause. Rappresentando le cause con variabili aleatorie, il Teorema del limite centrale ci dice che il fenomeno si comporta quindi come una variabile aleatoria molto vicina a una variabile aleatoria normale — *anche se le cause iniziali non avevano nulla a che fare con la distribuzione normale*. Questo non accade sempre (le ipotesi di indipendenza e di additività sono piuttosto forti), ma abbastanza spesso da giustificare la diffusione della distribuzione normale.

**ESEMPIO 7.19** Un esempio, semplificato ma non troppo, è quello della pigmentazione della pelliccia di alcuni animali. Supponiamo che il colore della pelliccia dipenda dalla presenza di  $n$  fattori,  $X_1, \dots, X_n$ . Ciascun fattore può essere presente con probabilità  $p$  (la stessa per ogni fattore), e la presenza di un fattore qualsiasi aumenta la pigmentazione di un'unità (opportuna; questa è l'additività). Supponendo che gli  $n$  fattori siano indipendenti, la variabile aleatoria  $S_n = X_1 + \dots + X_n$  che misura la pigmentazione è allora una variabile Bernouilliana di tipo  $(n, p)$ . La Fig. 7.9 mostra istogrammi associati a  $S_n$ , in cui la colonna  $i$ -esima ha base  $[i, i+1]$  e altezza  $p(S_n = i)$ , per  $p = 1/3$  e  $n = 2, 5$  e  $10$ , ed è evidente che la distribuzione sta tendendo a una distribuzione normale.

Il Teorema del limite centrale ci dice che otteniamo lo stesso risultato anche se ciascun fattore può apparire in più stati differenti, ogni stato con la propria probabilità e influenzando in modo diverso la pigmentazione; le uniche cose che contano sono il fatto che i vari fattori sono indipendenti, e che la loro influenza si somma.

**Osservazione 7.18** Un caso particolarmente importante del Teorema del limite centrale è quando le variabili aleatorie hanno tutte la stessa media  $\mu$  e la stessa varianza  $\sigma^2$ . Questo accade, per esempio, quando le  $X_i$  sono misure fatte su individui



**Figura 7.9** Convergenza della distribuzione binomiale.

diversi di una popolazione molto vasta; in questo caso  $\mu$  e  $\sigma^2$  sono la media e la varianza della popolazione. In questo caso il Teorema del limite centrale ci dice che per  $n$  abbastanza grande la distribuzione della media campionaria  $M_n$  approssima una distribuzione normale; ed è questo il fatto che avevamo utilizzato nella Sezione 3.8.

Vediamo ora di precisare in che senso le variabili aleatorie  $S_n$  e  $M_n$  tendono a una variabile aleatoria normale. Siccome si può dimostrare (vedi la Curiosità 7.3 e l’Osservazione 7.19) che

$$E(S_n) = \mu_1 + \cdots + \mu_n, \quad \text{Var}(S_n) = \sigma_1^2 + \cdots + \sigma_n^2,$$

e

$$E(M_n) = \frac{\mu_1 + \cdots + \mu_n}{n}, \quad \text{Var}(M_n) = \frac{\sigma_1^2 + \cdots + \sigma_n^2}{n^2},$$

la media e la varianza di  $S_n$  (e forse anche quelle di  $M_n$ ) potrebbero divergere al tendere di  $n$  all’infinito; in che senso allora  $S_n$  può tendere a una variabile aleatoria normale, che ha media e varianza finite?

L’idea è che dobbiamo normalizzare  $S_n$  e  $M_n$  in modo da portare la media a 0 e la varianza a 1, in modo da poterle confrontare meglio.

Sia  $X$  una variabile aleatoria qualsiasi, di media  $\mu_X$  e varianza  $\sigma_X^2$ . Allora la *centrata* di  $X$  è la variabile aleatoria  $X^{(c)}$  data da

$$X^{(c)} = X - \mu_X.$$

Invece la *standardizzata* di  $X$  è la variabile aleatoria  $X^{(s)}$  data da

$$X^{(s)} = \frac{1}{\sigma_X} X^{(c)} = \frac{X - \mu_X}{\sigma_X}.$$

Vediamo di calcolare media e varianza delle variabili aleatorie centralizzate e standardizzate. Prima di tutto abbiamo

$$F_{X^{(c)}}(t) = p(X^{(c)} \leq t) = p(X - \mu_X \leq t) = p(X \leq t + \mu_X) = F_X(t + \mu_X);$$

quindi derivando troviamo

$$f_{X^{(c)}}(x) = f_X(x + \mu_X),$$

e dunque ponendo  $y = x + \mu_X$  otteniamo

$$\begin{aligned} E(X^{(c)}) &= \int_{-\infty}^{+\infty} x f_{X^{(c)}}(x) dx = \int_{-\infty}^{+\infty} x f_X(x + \mu_X) dx = \int_{-\infty}^{+\infty} (y - \mu_X) f_X(y) dy \\ &= \int_{-\infty}^{+\infty} y f_X(y) dy - \mu_X \int_{-\infty}^{+\infty} f_X(y) dy = \mu_X - \mu_X = 0, \end{aligned}$$

e

$$\begin{aligned} \text{Var}(X^{(c)}) &= \int_{-\infty}^{+\infty} x^2 f_{X^{(c)}}(x) dx = \int_{-\infty}^{+\infty} x^2 f_X(x + \mu_X) dx \\ &= \int_{-\infty}^{+\infty} (y - \mu_X)^2 f_X(y) dy = \sigma_X^2. \end{aligned}$$

In altre parole, la variabile aleatoria centrata ha media 0 e varianza uguale a quella di  $X$ . Poi,

$$F_{X^{(s)}}(t) = p(X^{(s)} \leq t) = p(\sigma_X^{-1} X^{(c)} \leq t) = p(X^{(c)} \leq \sigma_X t) = F_{X^{(c)}}(\sigma_X t);$$

quindi derivando troviamo

$$f_{X^{(s)}}(x) = \sigma_X f_{X^{(c)}}(\sigma_X x),$$

e dunque ponendo  $y = \sigma_X x$  otteniamo

$$\begin{aligned} E(X^{(s)}) &= \int_{-\infty}^{+\infty} x f_{X^{(s)}}(x) dx = \int_{-\infty}^{+\infty} \sigma_X x f_{X^{(c)}}(\sigma_X x) dx \\ &= \frac{1}{\sigma_X} \int_{-\infty}^{+\infty} y f_{X^{(c)}}(y) dy = 0, \end{aligned}$$

e

$$\begin{aligned} \text{Var}(X^{(s)}) &= \int_{-\infty}^{+\infty} x^2 f_{X^{(s)}}(x) dx = \int_{-\infty}^{+\infty} \sigma_X x^2 f_{X^{(c)}}(\sigma_X x) dx \\ &= \frac{1}{\sigma_X^2} \int_{-\infty}^{+\infty} y^2 f_{X^{(c)}}(y) dy = 1. \end{aligned}$$

In altre parole, la variabile aleatoria standardizzata ha sempre media 0 e varianza 1.

*Osservazione 7.19* Quest'ultimo conto dice in realtà che

$$E(\sigma X) = \frac{E(X)}{\sigma} \quad \text{e} \quad \text{Var}(\sigma X) = \frac{\text{Var}(X)}{\sigma^2}$$

per ogni variabile aleatoria  $X$  e ogni  $\sigma > 0$ .

*Osservazione 7.20* Se  $X$  è normale di media  $\mu$  e varianza  $\sigma^2$ , si vede facilmente (controlla) che  $X^{(s)}$  è normale di media 0 e varianza 1.

Allora l'enunciato esatto del Teorema del limite centrale è: *se le ipotesi (C1)–(C3) sono soddisfatte allora le standardizzate  $S_n^{(s)}$  e  $M_n^{(s)}$  tendono per  $n \rightarrow +\infty$  a una variabile aleatoria normale di media 0 e varianza 1 nel senso che*

$$\forall t \in \mathbb{R} \quad \lim_{n \rightarrow +\infty} F_{S_n^{(s)}}(t) = \lim_{n \rightarrow +\infty} F_{M_n^{(s)}}(t) = \Phi(t).$$

*Osservazione 7.21* Quando le  $X_i$  hanno tutte la stessa media  $\mu$  e la stessa varianza  $\sigma^2$ ,abbiamo (vedi la Curiosità 7.3)

$$E(M_n) = \mu \quad \text{e} \quad \text{Var}(M_n) = \frac{\sigma^2}{n}.$$

Per questo motivo in un certo senso (e con un lieve abuso di linguaggio) si può dire che il teorema del limite centrale implica che  $M_n$  approssima sempre meglio una variabile aleatoria normale di media  $\mu$  e varianza  $\sigma^2/n$ . Mettendo insieme questo risultato con quanto detto nell'Osservazione 7.17 otteniamo le affermazioni sugli intervalli di confidenza anticipati nell'Osservazione 3.33 e nella Sezione 3.8.