

4

Bestiario

In questo capitolo presenteremo le principali classi di funzioni reali di variabile reale che si incontrano nella pratica scientifica. Vedremo anche alcune delle tecniche principali per trovare funzioni che approssimano meglio possibile dei dati sperimentali.

4.1 Funzioni lineari

Le funzioni reali di variabile reale più semplici (dopo le costanti...) sono le funzioni lineari. Rappresentano relazioni di proporzionalità: una funzione è lineare se il suo valore varia in modo proporzionale alla variazione dell'argomento. In altre parole, una funzione f è lineare se esiste un numero reale $m \in \mathbb{R}$ (di solito non nullo) tale che se la variabile indipendente x varia di una quantità q allora la variabile dipendente $f(x)$ varia di mq .

Vediamo come si deduce da questa definizione la formula che descrive una funzione lineare. Supponiamo di variare il valore della variabile indipendente da x_0 a x ; la variazione è quindi uguale a $q = x - x_0$, che tradizionalmente si indica con $\Delta x = x - x_0$. Se $f: \mathbb{R} \rightarrow \mathbb{R}$ è una funzione lineare, la variazione $\Delta f = f(x) - f(x_0)$ del suo valore deve soddisfare la relazione

$$\Delta f = m \Delta x .$$

Inserendo in questa formula le definizioni di Δf e Δx otteniamo

$$f(x) - f(x_0) = \Delta f = m \Delta x = m(x - x_0) ,$$

per cui

$$f(x) = mx + d , \tag{4.1}$$

con $d = f(x_0) - mx_0$.

Viceversa, supponiamo che la funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ sia data dalla formula (4.1). Allora

$$\Delta f = f(x) - f(x_0) = mx + d - (mx_0 + d) = m(x - x_0) = m\Delta x ,$$

cioè f rappresenta una relazione di proporzionalità.

Riassumendo, le *funzioni lineari* (a volte chiamate anche *funzioni lineari affini*, riservando il termine “lineare” alle funzioni di questo tipo con $d = 0$) sono tutte e sole le funzioni della forma (4.1) per opportuni $m, d \in \mathbb{R}$.

Nella pratica sperimentale, capita spesso di trovare dati che dipendono in maniera lineare da una variabile (almeno per certi intervalli della variabile; vedi l’Osservazione 4.1); si pone quindi il problema di come trovare la legge che esprime questa relazione a partire dai dati sperimentali. In altre parole, vogliamo recuperare i coefficienti m e d conoscendo alcuni punti del grafico della funzione. Vediamo un esempio molto semplice, ma già significativo, di questa situazione.

ESEMPIO 4.1 È noto che la percentuale di semi di una data pianta che germogliano dipende dalla temperatura ambiente. Per una determinata varietà di pomodoro, è stato verificato che alla temperatura di 12°C germoglia il 40% dei semi, mentre alla temperatura di 15°C germoglia il 70% dei semi. Trova la relazione fra la temperatura e la percentuale di semi germogliati, supponendo sia espressa da una funzione lineare. Indichiamo con $P(T)$ la percentuale di semi che germoglia alla temperatura di $T^\circ\text{C}$. Siccome abbiamo supposto che la funzione $P: \mathbb{R} \rightarrow \mathbb{R}$ che associa alla temperatura T la percentuale $P(T)$ sia lineare, possiamo scrivere

$$P(T) = mT + d$$

per opportuni $m, d \in \mathbb{R}$; il nostro obiettivo è usare i dati sperimentali per calcolare m e d . Noi sappiamo che $P(12) = 40$ e $P(15) = 70$; dunque

$$30 = 70 - 40 = P(15) - P(12) = \Delta P = m\Delta T = m(15 - 12) = 3m ,$$

(attenzione: in questo esempio la variabile indipendente si chiama T e la variabile dipendente P , per cui abbiamo scritto ΔT e ΔP invece di Δx e Δf) da cui deduciamo

$$m = \frac{30}{3} = 10 .$$

Per ricavare d basta notare che si deve avere

$$40 = P(12) = m \cdot 12 + d = 10 \cdot 12 + d = 120 + d ,$$

per cui

$$d = -80 .$$

Quindi l’unica funzione lineare che rappresenta correttamente i dati sperimentali è

$$P(T) = 10T - 80 . \tag{4.2}$$

Osservazione 4.1 È importante notare che per arrivare a questa soluzione abbiamo *supposto a priori* che la funzione da trovare fosse di tipo lineare; è un’ipotesi, e non una conseguenza. Del resto, da due sole coppie di dati è ben difficile immaginare, senza altre informazioni, quale possa essere l’andamento della funzione che

volevamo studiare. Quale sia l'ipotesi giusta da fare (se lineare, quadratica, esponenziale o di altro tipo) può venire suggerito da quanto sappiamo sul fenomeno biologico che stiamo studiando. Altrimenti, conviene fare molte più misure e cercare di capire quale sia la funzione che *meglio approssima* i dati che abbiamo ottenuti, tenendo presente che le misure sono sicuramente affette da errori sperimentali. Ne parleremo più in dettaglio nella Sezione 4.3.

Come già accennato nel capitolo precedente, lo scopo di ottenere una formula come la (4.2) è effettuare predizioni. Ci permette di dare risposte plausibili¹, senza bisogno di altre misure, a domande del tipo: quale percentuale di semi germoglierà alla temperatura di 14 °C? A quale temperatura germoglierà il 50% dei semi?

ESEMPIO 4.2 Supponiamo quindi che per questa varietà di pomodori la relazione fra la percentuale di semi che germogliano e la temperatura sia data dalla formula (4.2). Allora la percentuale di semi che germogliano a 14 °C è

$$P(14) = 10 \cdot 14 - 80 = 140 - 80 = 60\% .$$

Trovare la temperatura T a cui germoglia il 50% dei semi equivale invece a risolvere l'equazione $P(T) = 50$, cioè

$$50 = P(T) = 10T - 80 ;$$

quindi $10T = 130$, cioè $T = 13$ °C.

ESEMPIO 4.3 Per la stessa varietà di pomodori, vogliamo trovare quale percentuale di semi germoglierà alla temperatura di 10 °C, e a quale temperatura germoglierà il 90% dei semi. La risposta alla prima domanda è $P(10) = 10 \cdot 10 - 80 = 20\%$, mentre per rispondere alla seconda domanda risolviamo l'equazione $P(T) = 90$ ottenendo $T = 17$ °C.

Le predizioni dell'Esempio 4.2 sono frutto di una *interpolazione*. Infatti, abbiamo dati sperimentali sia per valori della variabile indipendente inferiori a quelli coinvolti in queste predizioni, sia per valori superiori: sappiamo cosa succede a 12 e 15 °C, e deduciamo cosa accade a 13 e 14 °C. Invece, le predizioni dell'Esempio 4.3 sono frutto di una *estrapolazione*: i valori della variabile indipendente coinvolti nelle predizioni (10 e 17 °C) sono esterni all'intervallo dei valori della variabile indipendente per cui abbiamo dati sperimentali. Le estrapolazioni sono sempre molto più rischiose delle interpolazioni, in quanto l'ipotesi iniziale (che la relazione sia di tipo lineare) potrebbe valere solo all'interno di un determinato intervallo di valori.

ESEMPIO 4.4 Usando la (4.2) “prediciamo” che alla temperatura di 19 °C germoglierà il $P(19) = 10 \cdot 19 - 80 = 110\%$ dei semi, cosa piuttosto improbabile a meno

¹ Plausibili, e non certe: vedi l'Osservazione 4.2.

di generazione spontanea di nuovi semi dal nulla... Analogamente, la “predizione” che alla temperatura di 5 °C germogli il $P(5) = 10 \cdot 5 - 80 = -30\%$ dei semi ha ben poco senso.

Dunque quando si ipotizza un certo andamento per dei dati sperimentali è importante indicare con chiarezza l’intervallo dei valori per cui si ritiene valida l’ipotesi; al di fuori di quei valori l’estrapolazione potrebbe non avere senso anche se la funzione che rappresenta l’andamento dei dati è ancora definita.

Osservazione 4.2 Determinare l’intervallo dei valori in cui la formula ottenuta può essere valida è spesso un esercizio di buon senso: nel nostro caso, una percentuale negativa o maggiore del 100% non ha senso, per cui dobbiamo escludere i valori che danno risultati del genere. Nella pratica sperimentale, c’è però un ulteriore passo importante da fare: confrontare le predizioni sensate (le interpolazioni) ottenute con nuovi risultati sperimentali. Infatti, le nostre predizioni sono basate su un’ipotesi (che la funzione fosse di tipo lineare), ipotesi che dobbiamo verificare nei fatti. Se le nostre predizioni sono in buon accordo con le nuove misure (tenendo presente gli inevitabili errori sperimentali) allora possiamo dirci soddisfatti della nostra ipotesi; se invece non lo sono, dobbiamo cambiare ipotesi (vedi gli Esempi 4.7, 4.9, 4.18 e 4.27, e l’Esercizio 4.2 della Sezione 4.4).

Lasciamo ora crescere in pace i nostri pomodori, e vediamo come si affrontano in generale i problemi che abbiamo risolto in questo caso particolare. Supponiamo di avere due coppie $P_0 = (x_0, y_0)$ e $P_1 = (x_1, y_1)$ di dati; vogliamo trovare una funzione lineare $f(x) = mx + d$ tale che P_0 e P_1 appartengano al grafico di f , cioè tale che $f(x_0) = y_0$ e $f(x_1) = y_1$. Imitando il procedimento usato nell’Esempio 4.1 troviamo

$$y_1 - y_0 = f(x_1) - f(x_0) = \Delta f = m\Delta x = m(x_1 - x_0) ,$$

per cui

$$m = \frac{y_1 - y_0}{x_1 - x_0} = \frac{\Delta f}{\Delta x} .$$

Osservazione 4.3 Ovviamente stiamo supponendo che $x_1 \neq x_0$, in quanto altrimenti P_0 e P_1 non potrebbero (perché?) essere due punti del grafico di una sola funzione (a meno che non siano uguali, nel qual caso devi strigliare il tuo assistente e imporgli di misurare due coppie di dati diverse, se vuole sperare di ottenere un qualche risultato).

Una volta trovato m , è facile recuperare anche d : infatti

$$d = f(x_0) - mx_0 = y_0 - mx_0 .$$

Nota che

$$\begin{aligned} f(x_1) - mx_1 &= f(x_0) + (f(x_1) - f(x_0)) - mx_0 - m(x_1 - x_0) \\ &= f(x_0) - mx_0 + \Delta f - m\Delta x = f(x_0) - mx_0 , \end{aligned}$$

per cui si ottiene lo stesso valore di d sia usando P_0 sia usando P_1 .

Osservazione 4.4 Dalla formula $\Delta f = m\Delta x$ possiamo dedurre che

$$f(x) = f(x_0 + \Delta x) = f(x_0) + m(x - x_0) = f(x_0) + \frac{\Delta f}{\Delta x} \cdot (x - x_0),$$

formula che ritroveremo in un contesto diverso nel prossimo capitolo.

Dunque ci basta conoscere due punti del grafico di una funzione lineare per ricavare l'espressione della funzione. Viceversa, data la funzione è facile tracciarne il grafico. Infatti, sappiamo già che il grafico di $f(x) = mx + d$ dev'essere una retta; quindi ci basta trovarne due punti. Per esempio, un punto può essere l'intersezione con l'asse delle ordinate: ponendo $x = 0$ troviamo il punto² $(0, d)$. Una volta ottenuto un punto $P_0 = (x_0, y_0)$ del grafico, ogni altro punto $P_1 = (x_1, y_1)$ si ottiene con la formula

$$(x_1, y_1) = P_0 + (\Delta x, m\Delta x),$$

dove $\Delta x = x_1 - x_0$.

Osservazione 4.5 I grafici delle funzioni lineari sono tutte e sole le rette non parallele all'asse delle ordinate. Per avere tutte le rette dobbiamo considerare insiemi di equazione $ax + by + c = 0$. Quando $b \neq 0$ ricaviamo $y = -(a/b)x - (c/b)$, cioè il grafico della funzione lineare $f(x) = mx + d$ con $m = -a/b$ e $d = -c/b$. Se invece $b = 0$ (e $a \neq 0$) otteniamo $x = -c/a$, cioè la retta parallela all'asse delle ordinate passante per il punto $(-c/a, 0)$. Analogamente, se $a = 0$ e $b \neq 0$ otteniamo $y = -c/b$, che è la retta parallela all'asse delle ascisse passante per il punto $(0, -c/b)$, ovvero il grafico della funzione costante $f(x) = -c/b$.

Nell'Esempio 4.2, oltre a ricavare l'ordinata conoscendo l'ascissa (la percentuale conoscendo la temperatura), abbiamo risolto il problema inverso di trovare l'ascissa conoscendo l'ordinata (la temperatura conoscendo la percentuale). In altre parole, dato il valore y_0 abbiamo risolto l'equazione lineare $f(x) = y_0$. Siccome $f(x) = mx + d$, vediamo subito che:

- se $m \neq 0$ l'equazione $f(x) = y_0$ ha come unica soluzione $x = (y_0 - d)/m$;
- se $m = 0$ e $d \neq y_0$ l'equazione $f(x) = y_0$ non ha soluzioni;
- se $m = 0$ e $d = y_0$ l'equazione $f(x) = y_0$ ha infinite soluzioni (ogni valore di x va bene).

Osservazione 4.6 Se $m \neq 0$, dire che l'equazione $f(x) = y_0$ ha un'unica soluzione quale che sia $y_0 \in \mathbb{R}$ equivale a dire che la funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ data da $f(x) = mx + d$ è invertibile. L'inversa è la funzione $f^{-1}: \mathbb{R} \rightarrow \mathbb{R}$ che fornisce la soluzione dell'equazione: $f^{-1}(y) = (y - d)/m$.

² Per questo motivo d è a volte chiamato *intercetta delle ordinate*. Invece, il coefficiente m è spesso chiamato *coefficiente angolare*, per motivi che vedremo nella Sezione 4.10.

La relazione $\Delta f = m\Delta x$ permette di determinare facilmente quando una funzione lineare è crescente o decrescente. In generale, una funzione è *crescente* se aumentando il valore dell'argomento aumenta anche il valore della funzione; ed è *decrescente* se invece aumentando il valore dell'argomento il valore della funzione diminuisce. In altre parole, f è crescente se $x_0 \leq x_1$ implica $f(x_0) \leq f(x_1)$, mentre è decrescente se $x_0 \leq x_1$ implica $f(x_0) \geq f(x_1)$. In altre parole ancora, il grafico di una funzione crescente sale andando verso destra; quello di una funzione decrescente invece scende.

Osservazione 4.7 Una funzione è invece *strettamente crescente* se $x_0 < x_1$ implica $f(x_0) < f(x_1)$, escludendo la possibilità che si abbia $f(x_0) = f(x_1)$; ed è *strettamente decrescente* se $x_0 < x_1$ implica $f(x_0) > f(x_1)$. Infine una funzione crescente o decrescente si dice *monotona* (e non monòtona, anche se l'idea è la stessa: è una funzione che non cambia mai modo di crescere).

Ora, dire che $x_0 < x_1$ equivale a dire che $\Delta x = x_1 - x_0 > 0$; analogamente, dire che $f(x_0) < f(x_1)$ equivale a dire che $\Delta f = f(x_1) - f(x_0) > 0$ (rispettivamente, $f(x_0) > f(x_1)$ equivale a $\Delta f < 0$). Se f è una funzione lineare, sappiamo che $\Delta f = m\Delta x$; quindi

- se $m > 0$ allora $\Delta x > 0$ implica $\Delta f > 0$, cioè f è strettamente crescente;
- se $m < 0$ allora $\Delta x > 0$ implica $\Delta f < 0$, cioè f è strettamente decrescente;
- se $m = 0$ allora $\Delta f \equiv 0$, cioè f è costante.

Conoscere la crescenza o la decrescenza di una funzione aiuta a trovarne i punti di massimo e di minimo. Diremo che un punto x_0 è un *punto di massimo* (rispettivamente, *punto di minimo*) per una funzione f su un intervallo $[a, b]$ se $f(x_0) \geq f(x)$ per ogni $x \in [a, b]$ (rispettivamente, $f(x_0) \leq f(x)$ per ogni $x \in [a, b]$). In altre parole, x_0 è un punto di massimo (minimo) se $(x_0, f(x_0))$ è un punto del grafico di f sopra l'intervallo $[a, b]$ con l'ordinata più alta (bassa). Il valore assunto dalla funzione (l'ordinata del grafico) in un punto di minimo (rispettivamente, di massimo) sull'intervallo $[a, b]$ viene detto (*valore*) *minimo* (rispettivamente *massimo*) di f sull'intervallo, e viene indicato con $\min f$ (rispettivamente, $\max f$), o con $\min_{x \in [a,b]} f(x)$ (rispettivamente, $\max_{x \in [a,b]} f(x)$) in caso sia importante ricordare l'intervallo che si sta considerando.

Osservazione 4.8 Trovare i punti di massimo o di minimo è fondamentale per le applicazioni della matematica. Infatti, in natura vale spesso un principio del minimo sforzo: la configurazione che si realizza (fra le infinite possibili) è quella che minimizza una qualche quantità. Per esempio, la luce segue il cammino più breve, i semi dei fiori cercano di disporsi in modo da minimizzare lo spreco di spazio, e così via.

Vogliamo trovare minimo e massimo di una funzione f monotona sull'intervallo $[a, b]$. Per definizione di intervallo, abbiamo $a \leq x \leq b$ per ogni $x \in [a, b]$. Se f è crescente, questo implica che $f(a) \leq f(x) \leq f(b)$; quindi se f è crescente sull'intervallo $[a, b]$ un punto di minimo è a , con valore minimo $f(a)$, e un punto

di massimo è b , con valore massimo $f(b)$. Un ragionamento analogo (controlla) ci dice che *se f è decrescente sull'intervallo $[a, b]$ un punto di minimo è b , con valore minimo $f(b)$, e un punto di massimo è a , con valore massimo $f(a)$.*

Osservazione 4.9 Se una funzione è strettamente crescente o decrescente (cioè strettamente monotona) su un intervallo chiuso allora ha un unico punto di minimo e un unico punto di massimo (perché?). Se invece non è strettamente monotona, potrebbe averne anche più di uno; per esempio, se f è costante allora tutti i punti sono contemporaneamente sia di massimo sia di minimo. Invece, il valore minimo e il valore massimo su un dato intervallo sono sempre unici (perché?).

CURIOSITÀ 4.1 Attenzione: dimostrare che un punto di minimo o un punto di massimo esiste può essere a volte anche molto complicato. E certe volte potrebbe anche non esistere. Per esempio, la funzione $f(x) = x$ non ha né minimo né massimo sull'intera retta reale \mathbb{R} : per quanto grande o quanto piccolo tu scelga $M \in \mathbb{R}$ esistono sempre $x_1, x_2 \in \mathbb{R}$ tali che $f(x_1) < M < f(x_2)$. Lo stesso problema si verifica su intervalli limitati ma non chiusi: la stessa funzione f non ha né minimo né massimo sull'intervallo aperto $(0, 1)$ (perché? Ricordati che 0 e 1 non appartengono all'intervallo considerato...). Per fortuna, il *Teorema di Weierstrass* assicura che tutte le funzioni continue (che sono la quasi totalità delle funzioni che considereremo in questo corso, e che definiremo nella Curiosità 4.9) hanno sempre almeno un punto di massimo e almeno un punto di minimo su qualsiasi intervallo *chiuso* della retta reale. Un esempio di funzione *non continua* che non ammette né massimo né minimo su un intervallo chiuso è la funzione $f: [-1, 1] \rightarrow \mathbb{R}$ definita da

$$f(x) = \begin{cases} |x| & \text{se } x \neq -1, 0, 1, \\ 1/2 & \text{se } x = -1, 0, 1. \end{cases}$$

Prova a tracciarne il grafico.

Abbiamo osservato che le funzioni lineari con coefficiente angolare non nullo sono sempre strettamente monotone; quindi quanto visto ci permette di trovarne massimi e minimi su intervalli chiusi. Per l'esattezza, se $f(x) = mx + d$ si ha

- se $m > 0$ il punto di minimo di f sull'intervallo $[a, b]$ è a , mentre il punto di massimo è b ;
- se $m < 0$ il punto di minimo di f sull'intervallo $[a, b]$ è b , mentre il punto di massimo è a .

Informazioni su crescenza e decrescenza aiutano anche a risolvere le disequazioni. Supponiamo di voler risolvere la disequazione $f(x) \geq y_0$ su un intervallo $[a, b]$ in cui la funzione f sia crescente. Ci sono tre casi possibili:

- se $y_0 \leq \min f = f(a)$, allora $f(x) \geq y_0$ per ogni $x \in [a, b]$, cioè tutti gli $x \in [a, b]$ sono soluzione della disequazione;
- se $y_0 > \max f = f(b)$, allora $f(x) < y_0$ per ogni $x \in [a, b]$, cioè la disequazione non ha soluzione in $[a, b]$;
- se $f(a) = \min f < y_0 \leq \max f = f(b)$ allora $f(x) \geq y_0$ per ogni $x \in [x_0, b]$, dove x_0 è la più piccola soluzione (quando esiste; vedi la Curiosità 4.2) dell'equazione $f(x) = y_0$ in $[a, b]$.

CURIOSITÀ 4.2 Un'altra proprietà non completamente banale delle funzioni continue definite su intervalli chiusi è che per ogni $y_0 \in [\min f, \max f]$ l'equazione $f(x) = y_0$ ammette sempre una

soluzione più piccola e una soluzione più grande (*teorema dei valori intermedi*). Sfortunatamente, le funzioni monotone non sono necessariamente continue; ma c'è un modo per aggirare il problema. Infatti, se f è crescente sull'intervallo $[a, b]$ allora per ogni $y_0 \in [\min f, \max f]$ esiste un più piccolo $x_0 \in [a, b]$ tale che $f(x) \geq y_0$ per ogni $x > x_0$, e $f(x) < y_0$ per ogni $x < x_0$. Se f è continua allora necessariamente $f(x_0) = y_0$, come prima; se f non è continua potrebbe succedere che $f(x_0)$ sia strettamente minore di y_0 . Quindi se f è crescente le soluzioni della disequazione $f(x) \geq y_0$ sono gli elementi dell'intervallo chiuso $[x_0, b]$ se $f(x_0) = y_0$, e gli elementi dell'intervallo semiaperto $(x_0, b]$ se $f(x_0) < y_0$. Un esempio di funzione crescente non continua è la $f: [-1, 1] \rightarrow \mathbb{R}$ data da:

$$f(x) = \begin{cases} x & \text{se } -1 \leq x \leq 0, \\ x + 1 & \text{se } 0 < x \leq 1. \end{cases}$$

Prova a tracciarne il grafico, e verifica che la disequazione $f(x) \geq 1$ ha come soluzione gli elementi dell'intervallo semiaperto $(0, 1]$.

Se vogliamo risolvere la disequazione $f(x) \leq y_0$ su un intervallo $[a, b]$ in cui la funzione f sia crescente, ci sono di nuovo tre casi possibili:

- se $y_0 < \min f = f(a)$, allora $f(x) > y_0$ per ogni $x \in [a, b]$, cioè la disequazione non ha soluzione in $[a, b]$;
- se $y_0 \geq \max f = f(b)$, allora $f(x) \leq y_0$ per ogni $x \in [a, b]$, cioè tutti gli $x \in [a, b]$ sono soluzione della disequazione;
- se $f(a) = \min f \leq y_0 < \max f = f(b)$ allora $f(x) \leq y_0$ per ogni $x \in [a, x_0]$, dove x_0 è la più grande soluzione (quando esiste; vedi la Curiosità 4.2) dell'equazione $f(x) = y_0$ in $[a, b]$.

Osservazione 4.10 Ragionamenti analoghi si applicano al caso di intervalli non chiusi, di intervalli illimitati, alle disequazioni strette (cioè con $>$ o $<$ invece di \geq e \leq), e alle funzioni decrescenti; lasciamo il compito di scrivere esplicitamente cosa si ottiene nei vari casi a te e al tuo assistente. *Attenzione:* il tuo obiettivo non dev'essere imparare a memoria tutti i casi possibili, ma capire come si ottengono, in modo da poter ripetere il ragionamento quando ti serve solo nei casi che ti servono (con notevole risparmio di tempo e di memoria).

Vediamo cosa questi ragionamenti ci dicono nel caso delle funzioni lineari. Vogliamo risolvere la disequazione $mx + d \geq y_0$; troviamo

- Se $m > 0$ (cioè $f(x) = mx + d$ è crescente) allora le soluzioni sono gli elementi della semiretta $[x_0, +\infty)$, dove $x_0 = (y_0 - d)/m = f^{-1}(y_0)$ è l'unica soluzione dell'equazione $mx + d = y_0$.
- Se $m < 0$ (cioè $f(x) = mx + d$ è decrescente) allora le soluzioni sono gli elementi della semiretta $(-\infty, x_0]$, dove $x_0 = (y_0 - d)/m = f^{-1}(y_0)$ è l'unica soluzione dell'equazione $mx + d = y_0$.

In maniera analoga (esercizio per te) si risolve la disequazione $mx + d \leq y_0$.

In particolare, se $m > 0$ la disequazione $mx + d > y_0$ ha soluzione una semiretta della forma $(x_0, +\infty)$ quale che sia $y_0 \in \mathbb{R}$. Questo vuol dire che se $m > 0$ possiamo rendere $f(x) = mx + d$ arbitrariamente grande a patto di scegliere x sufficientemente grande: per quanto grande sia $M > 0$ possiamo sempre trovare $x_0 > 0$

(sufficientemente grande) tale che $f(x) > M$ non appena $x > x_0$. In simboli,

$$\forall M > 0 \quad \exists x_0 > 0 : x > x_0 \implies f(x) > M .$$

Quando questo accade, si dice che $f(x)$ ha *limite $+\infty$* per x che *tende a $+\infty$* , e si scrive

$$\lim_{x \rightarrow +\infty} f(x) = +\infty .$$

Sempre supponendo $m > 0$, hai anche visto che la disequazione $mx + d < y_0$ ha soluzione una semiretta della forma $(-\infty, x_0)$ quale che sia $y_0 \in \mathbb{R}$. Questo vuol dire che se $m > 0$ *possiamo rendere $f(x) = mx + d$ arbitrariamente negativa a patto di scegliere x sufficientemente negativo*: per quanto grande sia $M > 0$ possiamo sempre trovare $x_0 < 0$ (sufficientemente negativo) tale che $f(x) < -M$ non appena $x < x_0$. In simboli,

$$\forall M > 0 \quad \exists x_0 < 0 : x < x_0 \implies f(x) < -M .$$

Stavolta si dice che $f(x)$ ha *limite $-\infty$* per x che *tende a $-\infty$* , e si scrive

$$\lim_{x \rightarrow -\infty} f(x) = -\infty .$$

Se $m < 0$ la situazione si inverte. In questo caso la disequazione $mx + d > y_0$ ha soluzione una semiretta della forma $(-\infty, x_0)$ quale che sia $y_0 \in \mathbb{R}$. Questo vuol dire che se $m < 0$ *possiamo rendere $f(x) = mx + d$ arbitrariamente grande a patto di scegliere x sufficientemente negativo*: per quanto grande sia $M > 0$ possiamo sempre trovare $x_0 < 0$ (sufficientemente negativo) tale che $f(x) > M$ non appena $x < x_0$. In simboli,

$$\forall M > 0 \quad \exists x_0 < 0 : x < x_0 \implies f(x) > M .$$

Si dice che $f(x)$ ha *limite $+\infty$* per x che *tende a $-\infty$* , e si scrive

$$\lim_{x \rightarrow -\infty} f(x) = +\infty .$$

Infine, se $m > 0$ la disequazione $mx + d < y_0$ ha soluzione una semiretta della forma $(x_0, +\infty)$ quale che sia $y_0 \in \mathbb{R}$. Questo vuol dire che se $m > 0$ *possiamo rendere $f(x) = mx + d$ arbitrariamente negativa a patto di scegliere x sufficientemente grande*: per quanto grande sia $M > 0$ possiamo sempre trovare $x_0 > 0$ (sufficientemente grande) tale che $f(x) < -M$ non appena $x > x_0$. In simboli,

$$\forall M > 0 \quad \exists x_0 > 0 : x > x_0 \implies f(x) < -M .$$

Stavolta si dice che $f(x)$ ha *limite $-\infty$* per x che *tende a $+\infty$* , e si scrive

$$\lim_{x \rightarrow +\infty} f(x) = -\infty .$$

A volte, quando una funzione f ha limite $\pm\infty$, si dice che f *diverge* a $\pm\infty$.

Riassumendo, il comportamento di una funzione lineare $f(x) = mx + d$ quando x è sufficientemente grande o sufficientemente negativo è dato da:

- se $m > 0$ allora $\lim_{x \rightarrow \pm\infty} f(x) = \pm\infty$;
- se $m < 0$ allora $\lim_{x \rightarrow \pm\infty} f(x) = \mp\infty$.

Con questo abbiamo concluso lo studio delle funzioni lineari. Nelle prossime sezioni cercheremo (per quanto possibile) di studiare in modo analogo funzioni più complesse.

4.2 Funzioni quadratiche

Le funzioni lineari sono tutte monotone: sempre crescenti o decrescenti. Non tutti i fenomeni naturali sono rappresentabili con funzioni monotone; spesso servono funzioni che un po' crescono e un po' decrescono.

ESEMPIO 4.5 Hai provato a far saltare il tuo assistente, sperando che la sua altitudine fosse descritta da una funzione monotona. Invece, sfortunatamente, è salito solo per poco e poi è tornato giù. La sua altitudine è stata inizialmente crescente, ha raggiunto un massimo, e poi è diventata decrescente.

Inoltre, anche le funzioni monotone non è detto che siano lineari, cioè che rappresentino relazioni di proporzionalità.

ESEMPIO 4.6 La superficie esterna di una cellula sferica dipende dal quadrato del raggio della cellula, per cui non aumenta in modo proporzionale al raggio.

Il tipo più semplice di funzioni non monotone (e quindi non lineari) è dato dalle *funzioni quadratiche*: funzioni $f: \mathbb{R} \rightarrow \mathbb{R}$ della forma

$$f(x) = ax^2 + bx + c .$$

Il grafico di una funzione quadratica è una curva chiamata *parabola*.

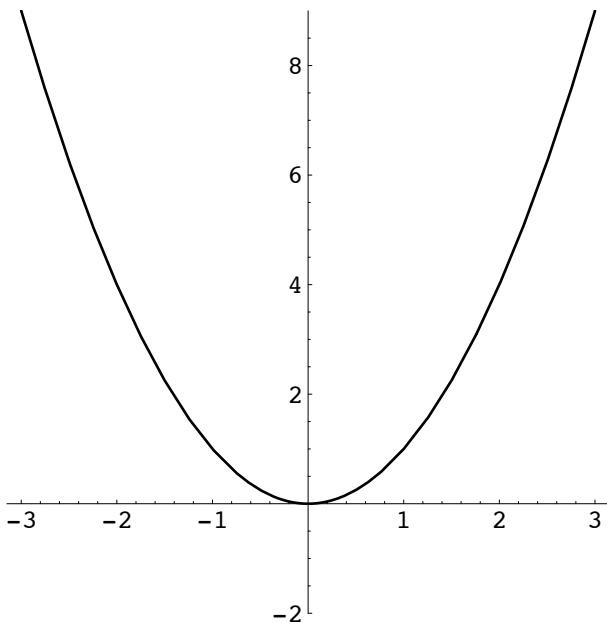
CURIOSITÀ 4.3 Più in generale, una *parabola* nel piano è il luogo dei punti la cui distanza da una retta data (detta *direttrice* della parabola) è uguale alla distanza da un punto dato (detto *fuoco* della parabola). Si può dimostrare che una parabola qualsiasi si ottiene sempre ruotando e traslando il grafico di una funzione quadratica.

Il primo obiettivo di questa sezione è trovare come collegare le proprietà geometriche (l'aspetto) del grafico di una funzione quadratica ai suoi coefficienti. Cominciamo studiando la funzione quadratica più semplice di tutte:

$$f(x) = x^2 ,$$

il cui grafico è rappresentato nella Figura 4.1.

La prima osservazione evidente è che $f(x) \geq 0$ sempre, e che $f(x) = 0$ se e solo se $x = 0$. In particolare,

**Figura 4.1** $f(x) = x^2$.

- (a) $f(x) = x^2$ ha un solo punto di minimo $\bar{x} = 0$, con valore minimo $\bar{y} = f(\bar{x}) = 0$; il punto $(\bar{x}, \bar{y}) = (0, 0)$ è detto vertice della parabola grafico di f .
- (b) la parabola grafico di $f(x) = x^2$ interseca l'asse delle ordinate nel punto di ordinata $c = f(0) = 0$.

Chiaramente, $(-x)^2 = x^2$, cioè $f(-x) = f(x)$ per qualsiasi x . In altri termini,

- (c) il grafico di $f(x) = x^2$ è simmetrico rispetto alla retta $x = \bar{x} = 0$ (l'asse delle ordinate), che è detta asse della parabola grafico di f .

Osservazione 4.11 Una funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ tale che $f(-x) = f(x)$ per ogni $x \in \mathbb{R}$ si dice *funzione pari*; se invece $f(-x) = -f(x)$ per ogni $x \in \mathbb{R}$ si dice che f è una *funzione dispari*. Un esempio di funzione dispari è $f(x) = 2x$.

CURIOSITÀ 4.4 Ogni funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ si può scrivere (in modo unico) come somma di una funzione pari e una funzione dispari. Infatti, ponendo $f_+(x) = \frac{1}{2}(f(x) + f(-x))$ e $f_-(x) = \frac{1}{2}(f(x) - f(-x))$ si vede subito che f_+ è pari, f_- è dispari, e $f = f_+ + f_-$.

Ora, se $0 \leq x_0 < x_1$ abbiamo $f(x_0) = x_0^2 < x_1^2 = f(x_1)$; invece se $x_0 < x_1 \leq 0$ abbiamo $0 \leq -x_1 < -x_0$ e $f(x_1) = f(-x_1) = (-x_1)^2 < (-x_0)^2 = f(-x_0) = f(x_0)$. Quindi

- (d) $f(x) = x^2$ è strettamente decrescente nella semiretta $(-\infty, \bar{x}]$ e strettamente crescente nella semiretta $[\bar{x}, +\infty)$, dove $\bar{x} = 0$. In questo caso, si dice anche che la parabola ha la *concavità rivolta verso l'alto*.

Inoltre, per ogni $y_0 > 0$ la diseguaglianza $f(x) > y_0$ ha come soluzione le semirette $(\sqrt{y_0}, +\infty)$ e $(-\infty, -\sqrt{y_0})$. Quindi possiamo rendere $f(x)$ arbitrariamente grande a patto di prendere x sufficientemente grande o sufficientemente negativo; usando la simbologia dei limiti introdotta nella sezione precedente possiamo dire che

- (e) se $f(x) = x^2$ allora $\lim_{x \rightarrow \pm\infty} f(x) = +\infty$.

Per concludere la descrizione geometrica della parabola grafico di $f(x) = x^2$ dobbiamo misurarne in qualche modo la larghezza. Un modo per farlo è vedere come cresce l'ordinata allontanandosi dal vertice: se cresce molto la parabola è stretta, se cresce poco la parabola è larga. Nel nostro caso si ha

- (f) se $f(x) = x^2$ allora $f(x) - f(\bar{x}) = 1 \cdot (x - \bar{x})^2$, dove $\bar{x} = 0$. In particolare, allontanandosi di un'unità dal vertice l'ordinata varia di $a = f(\bar{x} + 1) - f(\bar{x}) = 1$.

Possiamo effettuare un'analisi analoga sul grafico della funzione $f(x) = -x^2$. In questo caso si ottiene (vedi la Figura 4.2):

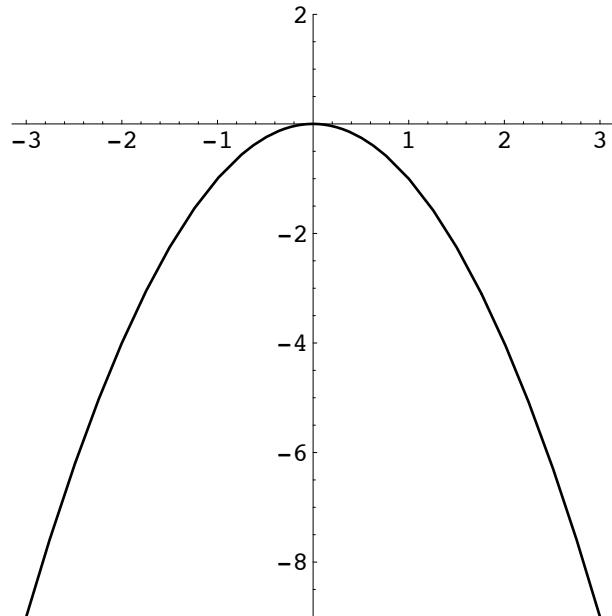


Figura 4.2 $f(x) = -x^2$.

- (a) $f(x) = -x^2$ ha un solo punto di massimo $\bar{x} = 0$, e il valore massimo è $\bar{y} = f(\bar{x}) = 0$; il punto $(\bar{x}, \bar{y}) = (0, 0)$ è sempre detto *vertice* della parabola grafico di f .
- (b) il grafico di $f(x) = -x^2$ interseca l'asse delle ordinate nel punto di ordinata $c = f(0) = 0$.
- (c) il grafico di $f(x) = -x^2$ è simmetrico rispetto alla retta $x = \bar{x} = 0$ (l'asse delle ordinate), che è ancora detta *asse della parabola* grafico di f .

- (d) $f(x) = -x^2$ è strettamente crescente nella semiretta $(-\infty, \bar{x}]$ e strettamente decrescente nella semiretta $[\bar{x}, +\infty)$, dove $\bar{x} = 0$. In questo caso, si dice che la parabola ha la *concavità rivolta verso il basso*.
- (e) se $f(x) = -x^2$ allora $\lim_{x \rightarrow \pm\infty} f(x) = -\infty$.
- (f) se $f(x) = -x^2$ allora $f(x) - f(\bar{x}) = -1 \cdot (x - \bar{x})^2$, con $\bar{x} = 0$. In particolare, allontanandosi di un'unità dal vertice l'ordinata varia di $a = f(\bar{x} + 1) - f(\bar{x}) = -1$.

Vogliamo far vedere che ogni parabola grafico di funzione quadratica soddisfa opportune variazioni delle proprietà (a)–(f), ed è completamente determinata dalle coordinate (\bar{x}, \bar{y}) del vertice e dalla larghezza $a = f(\bar{x} + 1) - f(\bar{x})$. Per farlo, vediamo come possiamo spostare il vertice e cambiare la larghezza, e che effetto ha sulla funzione quadratica.

Come primo passo, proviamo a variare la larghezza della parabola. Abbiamo visto che la parabola grafico di x^2 sale di 1 unità se ci spostiamo dal vertice di 1 unità. Se invece salisse di $a > 1$ unità spostandosi orizzontalmente dal vertice di 1 unità la parabola sarebbe più stretta (in quanto raggiunge l'ordinata 1 prima di $x = 1$); se salisse di $0 < a < 1$ unità spostandosi orizzontalmente dal vertice di 1 unità la parabola sarebbe più larga (in quanto raggiunge l'ordinata 1 oltre $x = 1$). Per ottenere questo effetto è sufficiente moltiplicare la funzione per a , cioè passare dalla funzione x^2 alla funzione ax^2 .

Osservazione 4.12 Questo procedimento si può applicare anche per a negativi. In questo caso $a = -|a| < 0$, per cui moltiplicare per a equivale a moltiplicare prima per $|a| > 0$ (modificando la larghezza della parabola) e poi per -1 . Quest'ultima operazione effettua una simmetria rispetto all'asse delle ascisse, ribaltando il grafico; vedi la Figura 4.3.

Osservazione 4.13 Moltiplicare per a le ordinate corrisponde a cambiare l'unità di misura (e l'orientazione, se $a < 0$) sull'asse delle ordinate: si ottiene lo stesso effetto dividendo per $|a|$ l'unità di misura (e invertendo l'orientazione se $a < 0$). Infatti, la vecchia unità di misura, che aveva ordinata 1 nelle vecchie coordinate, ora ha ordinata a , per cui la nuova unità di misura (che ha coordinata 1 nelle nuove coordinate) è $1/|a|$ volte la vecchia (con orientazione opposta se $a < 0$).

Le proprietà della funzione $f(x) = ax^2$ si ottengono subito da quelle di x^2 (tenendo presente il segno di a):

- (a) $f(x) = ax^2$ ha un solo punto di minimo (se $a > 0$; di massimo se $a < 0$) $\bar{x} = 0$, e il valore minimo (o massimo) è $\bar{y} = f(\bar{x}) = 0$; il vertice della parabola ha ancora coordinate $(\bar{x}, \bar{y}) = (0, 0)$.
- (b) Il grafico di $f(x) = ax^2$ interseca l'asse delle ordinate nel punto di ordinata $c = f(0) = 0$.
- (c) Il grafico di $f(x) = ax^2$ è simmetrico rispetto all'asse $x = \bar{x} = 0$.
- (d) Il grafico di $f(x) = ax^2$ ha la concavità rivolta verso l'alto se $a > 0$, e rivolta verso il basso se $a < 0$.
- (e) Se $f(x) = ax^2$ allora $\lim_{x \rightarrow \pm\infty} f(x) = +\infty$ se $a > 0$, e $\lim_{x \rightarrow \pm\infty} f(x) = -\infty$ se $a < 0$.

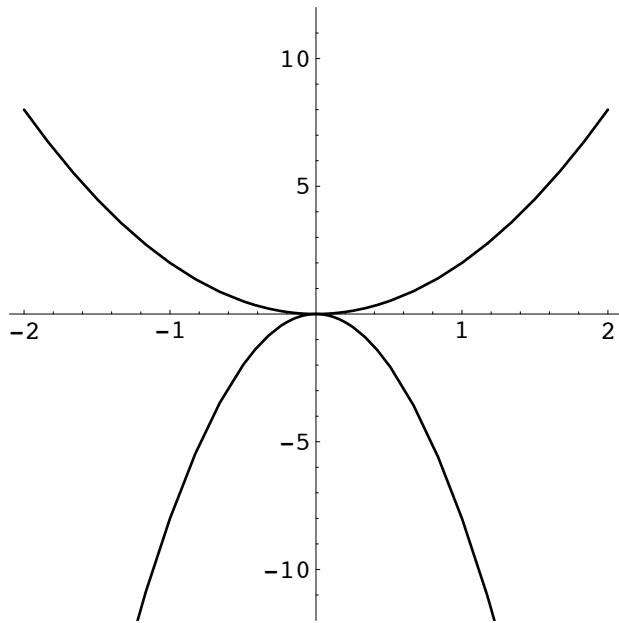


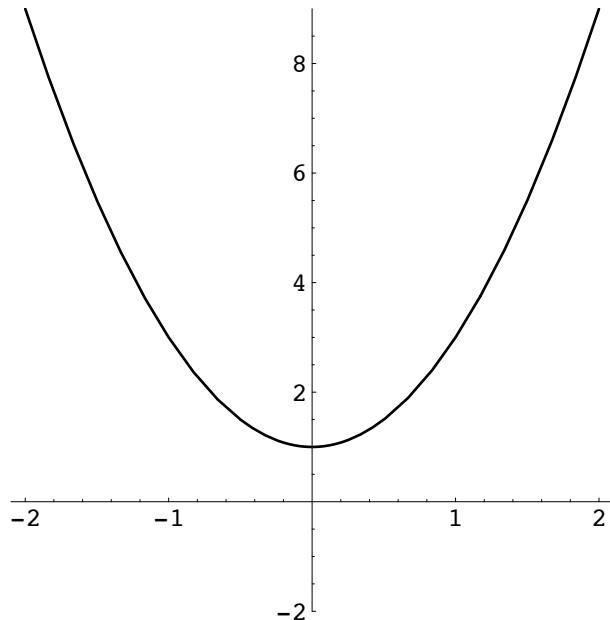
Figura 4.3 $f(x) = 2x^2$ e $f(x) = -8x^2$.

- (f) Se $f(x) = ax^2$ allora $f(x) - f(\bar{x}) = a \cdot (x - \bar{x})^2$, dove $\bar{x} = 0$. In particolare, allontanandosi di un'unità dal vertice l'ordinata varia di $f(\bar{x} + 1) - f(\bar{x}) = a$. Proviamo ora a spostare in direzione verticale il vertice della parabola grafico di ax^2 . Per portare il vertice nel punto $(0, \gamma)$ è sufficiente traslare l'intero grafico di una distanza pari a γ nella direzione verticale. In altre parole, dobbiamo sommare γ alle ordinate del grafico, cioè passare dalla funzione ax^2 alla funzione $ax^2 + \gamma$; vedi la Figura 4.4.

Osservazione 4.14 In generale, traslando verticalmente di una quantità γ il grafico di una funzione f si ottiene il grafico della funzione $f + \gamma$. Inoltre, invece di traslare in direzione verticale il grafico di una quantità γ avremmo potuto traslare in direzione verticale gli assi della quantità $-\gamma$ (cioè sottrarre γ alle ordinate) ottenendo lo stesso risultato. In altre parole, *traslare il piano in direzione verticale di una quantità γ equivale a sottrarre γ alle ordinate*.

Le proprietà della funzione $f(x) = ax^2 + \gamma$ si ottengono subito da quelle di ax^2 :

- $f(x) = ax^2 + \gamma$ ha un solo punto di minimo (se $a > 0$; di massimo se $a < 0$) $\bar{x} = 0$, e il valore minimo (o massimo) è $\bar{y} = f(\bar{x}) = \gamma$; il vertice della parabola ha ora coordinate $(\bar{x}, \bar{y}) = (0, \gamma)$.
- Il grafico di $f(x) = ax^2 + \gamma$ interseca l'asse delle ordinate nel punto di ordinata $c = f(0) = \gamma$.
- Il grafico di $f(x) = ax^2 + \gamma$ è simmetrico rispetto all'asse $x = \bar{x} = 0$.
- Il grafico di $f(x) = ax^2 + \gamma$ ha la concavità rivolta verso l'alto se $a > 0$, e rivolta verso il basso se $a < 0$.

**Figura 4.4** $f(x) = ax^2 + \gamma$.

- (e) Se $f(x) = ax^2 + \gamma$ allora $\lim_{x \rightarrow \pm\infty} f(x) = +\infty$ se $a > 0$, e $\lim_{x \rightarrow \pm\infty} f(x) = -\infty$ se $a < 0$.
- (f) Se $f(x) = ax^2 + \gamma$ allora $f(x) - f(\bar{x}) = a \cdot (x - \bar{x})^2$, dove $\bar{x} = 0$. In particolare, allontanandosi di un'unità dal vertice l'ordinata varia di $f(\bar{x} + 1) - f(\bar{x}) = a$.

Il passo successivo consiste nel traslare orizzontalmente il vertice (e quindi l'asse) della parabola. Abbiamo visto che traslare verso l'alto di una quantità γ il grafico è equivalente a sottrarre γ alle ordinate (cioè a spostare gli assi verso il basso di una quantità γ). Per lo stesso motivo, *traslare in direzione orizzontale di una quantità β il grafico equivale a traslare in direzione orizzontale gli assi della quantità $-\beta$, cioè a sottrarre β alle ascisse*. In altre parole, il grafico della funzione $f(x) = a(x - \beta)^2 + \gamma$, ottenuta sostituendo $x - \beta$ a x in $ax^2 + \gamma$, è ottenuto traslando orizzontalmente di una quantità β il grafico di $ax^2 + \gamma$. Ne segue che (vedi la Figura 4.5)

- (a) $f(x) = a(x - \beta)^2 + \gamma$ ha un solo punto di minimo (se $a > 0$; di massimo se $a < 0$) $\bar{x} = \beta$, e il valore minimo (o massimo) è $\bar{y} = f(\bar{x}) = \gamma$; il vertice della parabola ha quindi coordinate $(\bar{x}, \bar{y}) = (\beta, \gamma)$.
- (b) Il grafico di $f(x) = a(x - \beta)^2 + \gamma$ interseca l'asse delle ordinate nel punto di ordinata $c = f(0) = a\beta^2 + \gamma$.
- (c) Il grafico di $f(x) = a(x - \beta)^2 + \gamma$ è simmetrico rispetto all'asse $x = \bar{x} = \beta$.
- (d) Il grafico di $f(x) = a(x - \beta)^2 + \gamma$ ha la concavità rivolta verso l'alto se $a > 0$, e rivolta verso il basso se $a < 0$.
- (e) Se $f(x) = a(x - \beta)^2 + \gamma$ allora $\lim_{x \rightarrow \pm\infty} f(x) = +\infty$ se $a > 0$, e $\lim_{x \rightarrow \pm\infty} f(x) = -\infty$ se $a < 0$.

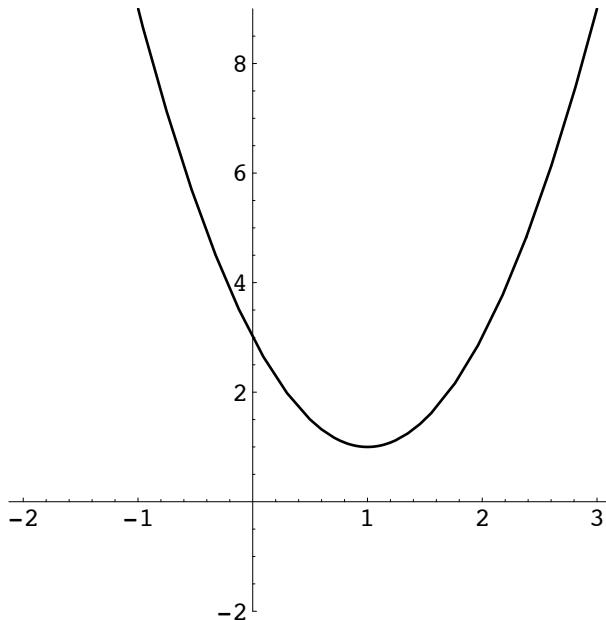


Figura 4.5 $f(x) = a(x - \beta)^2 + \gamma$.

- (f) Se $f(x) = a(x - \beta)^2 + \gamma$ allora $f(x) - f(\bar{x}) = a \cdot (x - \bar{x})^2$, con $\bar{x} = \beta$. Quindi allontanandosi di un'unità dal vertice l'ordinata varia di $f(\bar{x} + 1) - f(\bar{x}) = a$.

Osservazione 4.15 Vale la pena notare che c'è un'altra operazione ancora che potremmo a priori fare: moltiplicare le ascisse per un valore $\alpha \neq 0$, che equivale (come nel caso delle ordinate) a dividere per $|\alpha|$ l'unità di misura sull'asse delle ascisse (e a invertire l'orientazione se $\alpha < 0$). In questo modo arriveremmo a una funzione della forma $a(\alpha x - \beta)^2 + \gamma$. Fra un attimo vedremo però che, per le funzioni quadratiche, questa operazione non è necessaria: abbiamo già ottenuto tutte le funzioni quadratiche possibili senza bisogno di ulteriori operazioni. Invece è una trasformazione utile per studiare funzioni più complicate (vedi l'Osservazione 4.42).

Ora, $f(x) = a(x - \beta)^2 + \gamma$ è chiaramente una funzione quadratica: infatti svolgendo il quadrato troviamo

$$a(x - \beta)^2 + \gamma = ax^2 - 2a\beta x + a\beta^2 + \gamma = ax^2 + bx + c ,$$

dove

$$b = -2a\beta , \quad c = a\beta^2 + \gamma . \quad (4.3)$$

La cosa interessante è che vale anche il viceversa: ogni funzione quadratica si può scrivere nella forma $a(x - \beta)^2 + \gamma$. Infatti, ricavando β e γ dalle (4.3) in funzione di a , b e c (supponendo ovviamente $a \neq 0$), troviamo

$$\beta = -\frac{b}{2a} , \quad \gamma = c - a\beta^2 = c - \frac{b^2}{4a} = \frac{4ac - b^2}{4a} , \quad (4.4)$$

e quindi

$$ax^2 + bx + c = a \left(x + \frac{b}{2a} \right)^2 + c - \frac{b^2}{4a} = a(x - \beta)^2 + \gamma. \quad (4.5)$$

I risultati che abbiamo ottenuto sul grafico delle funzioni della forma $a(x - \beta)^2 + \gamma$ si possono quindi tradurre nel caso di funzioni quadratiche qualsiasi:

- (a) $f(x) = ax^2 + bx + c$ ha un solo punto di minimo (se $a > 0$; di massimo se $a < 0$) $\bar{x} = -b/2a$, e il valore minimo (o massimo) è $\bar{y} = f(\bar{x}) = c - b^2/4a$; il vertice della parabola ha quindi coordinate $(\bar{x}, \bar{y}) = (-b/2a, c - b^2/4a)$.
- (b) Il grafico di $f(x) = ax^2 + bx + c$ interseca l'asse delle ordinate nel punto di ordinata $f(0) = c$.
- (c) Il grafico di $f(x) = ax^2 + bx + c$ è simmetrico rispetto all'asse $x = \bar{x} = -b/2a$.
- (d) Il grafico di $f(x) = ax^2 + bx + c$ ha la concavità rivolta verso l'alto se $a > 0$, e rivolta verso il basso se $a < 0$. In altre parole, se $a > 0$ la funzione è strettamente decrescente nella semiretta $(-\infty, -b/2a]$ e strettamente crescente nella semiretta $[-b/2a, +\infty)$, mentre se $a < 0$ la funzione è strettamente crescente nella semiretta $(-\infty, -b/2a]$ e strettamente decrescente nella semiretta $[-b/2a, +\infty)$.
- (e) Se $f(x) = ax^2 + bx + c$ allora $\lim_{x \rightarrow \pm\infty} f(x) = +\infty$ se $a > 0$, e $\lim_{x \rightarrow \pm\infty} f(x) = -\infty$ se $a < 0$.
- (f) Se $f(x) = ax^2 + bx + c$ allora $f(x) - f(\bar{x}) = a \cdot (x - \bar{x})^2$, dove $\bar{x} = -b/2a$. In particolare, $f(\bar{x} + 1) - f(\bar{x}) = a$.

Dunque data la formula ora siamo in grado di tracciare il grafico. Vediamo ora come risolvere il problema inverso: dato il grafico (o, almeno, alcuni punti del grafico) ricavare la formula.

Un primo caso è quando abbiamo le coordinate (\bar{x}, \bar{y}) del vertice e (supponendo che il vertice non sia sull'asse delle ordinate, cioè che $\bar{x} \neq 0$) il punto $(0, c)$ di intersezione del grafico con l'asse delle ordinate. Allora i conti precedenti, e in particolare le (4.3), ci dicono che la funzione dev'essere $f(x) = ax^2 + bx + c$ con a e b dati da

$$b = -2a\bar{x}, \quad a = \frac{c - \bar{y}}{\bar{x}^2}.$$

Se invece il vertice è sull'asse delle ordinate, cioè $\bar{x} = 0$, allora sappiamo soltanto che $f(x) = ax^2 + c$ con $c = \bar{y}$; per trovare a servono altre informazioni (quali, per esempio, la larghezza della parabola).

Spesso, invece, conosciamo alcuni punti del grafico, senza però sapere quale sia il vertice. Per determinare la funzione, servono tre punti; vediamo come in un esempio.

ESEMPIO 4.7 *Torniamo a studiare i semi di pomodoro dell'Esempio 4.1. Sai già che alla temperatura di 12 °C germoglia il 40% dei semi, mentre alla temperatura di 15 °C germoglia il 70% dei semi. Un'ulteriore misurazione ha rivelato che alla temperatura di 9 °C germoglia il 20% dei semi. Dimostra che allora la relazione fra la temperatura e la percentuale di semi che germogliano non può essere lineare. Supponendo che sia quadratica, determinala.* Indichiamo nuovamente con $P(T)$ la percentuale di semi che germogliano alla temperatura T . Noi sappiamo che

$P(9) = 20$, $P(12) = 40$ e $P(15) = 70$. Se P fosse una funzione lineare, $\Delta P/\Delta T$ dovrebbe essere costante; invece

$$\frac{P(15) - P(12)}{15 - 12} = \frac{70 - 40}{3} = 10 \neq \frac{20}{3} = \frac{40 - 20}{3} = \frac{P(12) - P(9)}{12 - 9}.$$

Supponiamo allora che $P: \mathbb{R} \rightarrow \mathbb{R}$ sia una funzione quadratica, cioè che si abbia $P(T) = aT^2 + bT + c$; dobbiamo trovare a , b , $c \in \mathbb{R}$ in modo da avere $P(9) = 20$, $P(12) = 40$ e $P(15) = 70$. In altre parole, a , b e c devono soddisfare il seguente sistema di equazioni lineari:

$$\begin{cases} 81a + 9b + c = P(9) = 20, \\ 144a + 12b + c = P(12) = 40, \\ 225a + 15b + c = P(15) = 70. \end{cases} \quad (4.6)$$

Per risolvere questo sistema, sottraiamo la prima equazione dalla seconda, e la seconda dalla terza; otteniamo

$$\begin{cases} 63a + 3b = 20, \\ 81a + 3b = 30. \end{cases} \quad (4.7)$$

Sottraendo di nuovo la prima equazione dalla seconda otteniamo

$$18a = 10, \quad \text{cioè} \quad a = \frac{10}{18} = \frac{5}{9}.$$

Sostituendo questo valore nella seconda equazione in (4.7) troviamo

$$81 \cdot \frac{5}{9} + 3b = 20, \quad \text{cioè} \quad b = -5;$$

e sostituendo i valori di a e b trovati nella prima equazione in (4.6) recuperiamo infine

$$81 \cdot \frac{5}{9} - 9 \cdot 5 + c = 20, \quad \text{cioè} \quad c = 20.$$

Quindi la formula cercata è

$$P(T) = \frac{5}{9}T^2 - 5T + 20.$$

Questa formula ha qualche vantaggio su quella lineare. Per esempio, non è mai negativa; infatti, ha minimo per $T = -(-5)/2(5/9) = 9/2$, con valore minimo $P(9/2) = 20 - (-5)^2/4(5/9) = 35/4 > 0$. Ma anche lei può essere valida solo in un determinato intervallo di temperature. Infatti, $P(T)$ ricomincia ad aumentare quando la temperatura scende sotto $9/2$ °C, comportamento biologicamente alquanto improbabile; e $P(T) > 100$ se T è troppo grande (o sufficientemente negativo). Per esempio, $P(18) = 110$.

Il procedimento usato nel precedente esempio può essere applicato a qualsiasi funzione quadratica. Supponiamo di voler trovare la funzione quadratica $f(x) = ax^2 + bx + c$ il cui grafico passi per i punti (x_1, y_1) , (x_2, y_2) e (x_3, y_3) , con x_1 , x_2 ed x_3 tutti distinti. Vogliamo quindi trovare a , b e c in modo che $f(x_j) = y_j$ per $j = 1, 2, 3$. In altre parole, a , b e c devono risolvere il sistema

$$\begin{cases} x_1^2 a + x_1 b + c = y_1 , \\ x_2^2 a + x_2 b + c = y_2 , \\ x_3^2 a + x_3 b + c = y_3 . \end{cases}$$

Sottraiamo la prima equazione dalla seconda, e la seconda dalla terza; otteniamo

$$\begin{cases} (x_2^2 - x_1^2)a + (x_2 - x_1)b = y_2 - y_1 , \\ (x_3^2 - x_2^2)a + (x_3 - x_2)b = y_3 - y_2 . \end{cases} \quad (4.8)$$

Siccome $x_2^2 - x_1^2 = (x_2 - x_1)(x_2 + x_1)$ e $x_2 - x_1 \neq 0$, possiamo dividere la prima equazione per $x_2 - x_1$. Analogamente possiamo dividere la seconda equazione per $x_3 - x_2$, e otteniamo

$$\begin{cases} (x_2 + x_1)a + b = \frac{y_2 - y_1}{x_2 - x_1} , \\ (x_3 + x_2)a + b = \frac{y_3 - y_2}{x_3 - x_2} . \end{cases}$$

Sottraendo di nuovo la prima equazione dalla seconda otteniamo

$$(x_3 - x_1)a = \frac{y_3 - y_2}{x_3 - x_2} - \frac{y_2 - y_1}{x_2 - x_1} .$$

Siccome $x_3 - x_1 \neq 0$, da questa equazione possiamo ricavare a ; sostituendo il valore trovato nel sistema precedente otteniamo b , e sostituendo nel sistema iniziale troviamo anche c .

Osservazione 4.16 La generica funzione lineare dipendeva da due parametri (m e d); per determinarla avevamo bisogno di conoscere due punti del grafico. La generica funzione quadratica dipende da tre parametri (a , b e c); per determinarla abbiamo bisogno di conoscere tre punti del grafico. Tutto ciò non è un caso: si può dimostrare che se abbiamo una famiglia di funzioni dipendenti da k parametri, per determinare univocamente una funzione della famiglia servono k condizioni (indipendenti in un senso opportuno), quali per esempio richiedere il passaggio del grafico per k punti dati. Vedremo un altro esempio di questo fenomeno nella Sezione 4.4.

La (4.5) è molto utile anche per risolvere le equazioni di secondo grado. L'idea è che l'equazione $ax^2 + bx + c = y_0$ ha soluzione se e solo se la retta $y = y_0$ interseca il grafico di $f(x) = ax^2 + bx + c$. Questo accade solo se o $a > 0$ e y_0 è maggiore del minimo di f , oppure $a < 0$ e y_0 è minore del massimo di f . Quindi bisogna

confrontare y_0 con l'ordinata $\bar{y} = \gamma$ del vertice, che si legge facilmente da (4.5). Per la precisione, l'equazione $ax^2 + bx + c = y_0$ diventa

$$a(x - \beta)^2 + \gamma = y_0 , \quad \text{cioè} \quad (x - \beta)^2 = \frac{y_0 - \gamma}{a} .$$

Quindi ha soluzione reale se e solo se $(y_0 - \gamma)/a \geq 0$, e in tal caso le soluzioni sono

$$x_{\pm} = \beta \pm \sqrt{\frac{y_0 - \gamma}{a}} = \frac{-b \pm \sqrt{b^2 - 4a(c - y_0)}}{2a} . \quad (4.9)$$

Il caso $y_0 = 0$ è particolarmente interessante (e tutti gli altri possono esservi ricondotti sostituendo $c - y_0$ al posto di c). La quantità

$$D = b^2 - 4ac = -4a\bar{y}$$

è detta *discriminante* della funzione $f(x) = ax^2 + bx + c$. Il discriminante è positivo se e solo se a e l'ordinata \bar{y} del vertice hanno segno opposto, e si annulla se e solo se l'ordinata del vertice si annulla, cioè se e solo se il vertice è sull'asse delle ascisse. Insomma, o usando la posizione del vertice e la concavità del grafico, oppure usando il segno del discriminante in (4.9), giungiamo alla conclusione che l'equazione $ax^2 + bx + c = 0$ ha

- due soluzioni reali distinte se $D > 0$, cioè se a e \bar{y} hanno segno opposto;
- un'unica soluzione reale se $D = 0$, cioè se $\bar{y} = 0$, e in tal caso la soluzione è \bar{x} ;
- nessuna soluzione reale se $D < 0$, cioè se a e \bar{y} hanno lo stesso segno.

Come esercizio, riotteni questo risultato tracciando il grafico di $ax^2 + bx + c$ nei vari casi.

Usando le informazioni che abbiamo sulla crescenza e decrescenza delle funzioni quadratiche, possiamo facilmente trovare i massimi e i minimi in intervalli chiusi. Ci sono due casi da considerare:

- se l'intervallo $[a_0, a_1]$ non contiene l'ascissa del vertice di $f(x) = ax^2 + bx + c$, allora f è monotona in quell'intervallo, per cui (come abbiamo visto nella precedente sezione) i punti di minimo e massimo di f in $[a_0, a_1]$ sono gli estremi a_0 e a_1 .
- se l'intervallo $[a_0, a_1]$ contiene l'ascissa \bar{x} del vertice, sappiamo già che il punto di minimo (se $a > 0$) o di massimo (se $a < 0$) è \bar{x} . Siccome f è monotona nei due intervalli $[a_0, \bar{x}]$ e $[\bar{x}, a_1]$, si vede subito (perché?) che il punto di massimo (se $a > 0$) o di minimo (se $a < 0$) è quello³ fra i due estremi a_0 e a_1 su cui f assume il valore più grande (se $a > 0$) o più piccolo (se $a < 0$).

Queste tecniche ci permettono anche di risolvere facilmente le disequazioni di secondo grado. Perché la disequazione $ax^2 + bx + c \geq y_0$ possa avere soluzione occorre che il grafico di $f(x) = ax^2 + bx + c$ sia in qualche punto al di sopra

³ O entrambi se $f(a_0) = f(a_1)$.

della retta $y = y_0$. Mettendo insieme (4.9) con ciò che sappiamo sulla crescenza e decrescenza di f troviamo

- se $a > 0$ e
 - $y_0 \leq \bar{y}$, la disequazione $ax^2 + bx + c \geq y_0$ è soddisfatta per tutti i valori di $x \in \mathbb{R}$;
 - $y_0 > \bar{y}$, la disequazione $ax^2 + bx + c \geq y_0$ è soddisfatta per $x \in (-\infty, x_-]$ e per $x \in [x_+, +\infty)$, dove x_{\pm} sono dati da (4.9);
- se $a < 0$ e
 - $y_0 > \bar{y}$, la disequazione $ax^2 + bx + c \geq y_0$ non è mai soddisfatta;
 - $y_0 \leq \bar{y}$, la disequazione $ax^2 + bx + c \geq y_0$ è soddisfatta per $x \in [x_+, x_-]$, dove x_{\pm} sono dati da (4.9).

Analoghi risultati (esercizio: se sei confuso, aiutati tracciando il grafico nei vari casi) si trovano per la disequazione $ax^2 + bx + c \leq y_0$.

Osservazione 4.17 Quando $y_0 = 0$, il segno di \bar{y} è legato al segno di a tramite il segno del discriminante. In particolare, se $a > 0$ abbiamo $\bar{y} \geq 0$ se e solo se $D \leq 0$, mentre se $a < 0$ abbiamo $0 > \bar{y}$ se e solo se $D < 0$. Quindi quando $y = 0$ possiamo riformulare (esercizio per te) i risultati precedenti usando a e D invece di a e \bar{y} .

Osservazione 4.18 Come fatto nella scorsa sezione, i risultati sulle disequazioni ci permettono anche di studiare l'andamento all'infinito delle funzioni quadratiche. Per esempio, abbiamo appena visto che se $a > 0$ allora per ogni $y_0 \in \mathbb{R}$ possiamo trovare $x_{\pm} \in \mathbb{R}$ tali che $f(x) \geq y_0$ non appena $x \geq x_+$ oppure $x \leq x_-$. In altre parole, possiamo rendere $f(x)$ arbitrariamente grande a patto di scegliere x sufficientemente grande o sufficientemente negativo. Usando la terminologia già introdotta, abbiamo quindi dimostrato che

$$a > 0 \quad \implies \quad \lim_{x \rightarrow \pm\infty} ax^2 + bx + c = +\infty ,$$

in accordo con quanto avevamo già visto. In maniera analoga si dimostra che

$$a < 0 \quad \implies \quad \lim_{x \rightarrow \pm\infty} ax^2 + bx + c = -\infty .$$

4.3 Il metodo dei minimi quadrati

In questa sezione presenteremo due applicazioni dello studio delle funzioni quadratiche che abbiamo appena completato.

La prima applicazione consiste nel mantenere una promessa fatta nell'Osservazione 3.27. Siano $x_1, \dots, x_n \in \mathbb{R}$ dei numeri reali (dei dati); vogliamo trovare il punto di minimo della funzione

$$f(x) = \sum_{i=1}^n (x - x_i)^2 .$$

Se sviluppiamo i quadrati, vediamo subito che f è una funzione quadratica:

$$f(x) = \sum_{i=1}^n (x^2 - 2x_i x + x_i^2) = nx^2 - 2 \left(\sum_{i=1}^n x_i \right) x + \sum_{i=1}^n x_i^2,$$

per cui $f(x) = ax^2 + bx + c$ con

$$a = n, \quad b = -2 \sum_{i=1}^n x_i, \quad c = \sum_{i=1}^n x_i^2.$$

Quindi il punto di minimo di f è

$$\bar{x} = -\frac{b}{2a} = \frac{1}{n} \sum_{i=1}^n x_i,$$

cioè la media aritmetica dei dati, come promesso.

La seconda applicazione consiste in un metodo (detto *metodo dei minimi quadrati*) per trovare la retta che meglio approssima un dato insieme di dati.

Supponiamo di avere n coppie di dati $(x_1, y_1), \dots, (x_n, y_n)$, e di sospettare che le ordinate dipendano in modo lineare dalle ascisse. Anche se la nostra ipotesi è corretta, è molto improbabile che le n coppie di dati giacciono esattamente su una retta, in quanto non possiamo evitare gli errori sperimentali; abbiamo quindi bisogno di una tecnica che ci fornisca la “migliore” (in un senso da specificare) approssimazione lineare di questi dati, e al contempo una misura della bontà di questa approssimazione — in quanto, se la “migliore” approssimazione fosse cattiva vorrebbe dire che la nostra ipotesi di dipendenza lineare non è compatibile con i dati, e quindi dev’essere scartata.

Cominciamo con definire quanto il grafico di una funzione lineare $f(x) = mx + d$ approssima l’insieme $\{(x_1, y_1), \dots, (x_n, y_n)\}$. La coppia (x_i, y_i) appartiene al grafico di f se e solo se $y_i = mx_i + d$; quindi l’errore $\delta_i = mx_i + d - y_i$ misura la distanza che c’è fra il dato sperimentale (x_i, y_i) e il dato teorico $(x_i, f(x_i))$ che si avrebbe se la funzione f rappresentasse esattamente il fenomeno che stiamo studiando. Abbiamo quindi n errori, $\delta_1, \dots, \delta_n$; tenendo presente che a noi non importa il segno dell’errore ma solo la sua grandezza, e ricordando quanto fatto studiando la varianza, una misura di quanto la funzione $f(x) = mx + d$ approssima i dati è data dalla media dei quadrati degli errori:

$$S(m, d) = \frac{1}{n} \sum_{i=1}^n \delta_i^2 = \frac{1}{n} \sum_{i=1}^n (mx_i + d - y_i)^2.$$

Nota che la funzione S dipende dai due parametri m e d che determinano la funzione lineare f ; quindi S è una funzione di due variabili reali.

Il nostro obiettivo è trovare il punto di minimo di S , cioè i valori \bar{m} e \bar{d} di m e d che rendono $S(m, d)$ più piccola possibile⁴. La corrispondente funzione lineare $f(x) = \bar{m}x + \bar{d}$ sarà allora quella che meglio approssima i dati da cui siamo partiti; e ci rimarrà solo da trovare una misura di quanto buona sia questa approssimazione.

Per trovare questo punto di minimo procederemo in questo modo. Prima di tutto faremo vedere che, per ogni $m \in \mathbb{R}$ fissato, la funzione $d \mapsto S(m, d)$ è una funzione quadratica di d con coefficiente del termine quadrato positivo; quindi ammette un unico punto di minimo, che indicheremo con $d_0(m)$; vedremo anche come $d_0(m)$ dipende da m . Poi dimostreremo che anche la funzione $S(m, d_0(m))$ è una funzione quadratica (di m , stavolta) con coefficiente del termine quadrato positivo; quindi anche lei ha un unico punto di minimo \bar{m} , a cui corrisponde il valore $\bar{d} = d_0(\bar{m})$. Allora (\bar{m}, \bar{d}) è il punto di minimo cercato. Infatti, per ogni $(m, d) \in \mathbb{R}^2$ abbiamo

$$S(m, d) \geq S(m, d_0(m)) \geq S(\bar{m}, d_0(\bar{m})) = S(\bar{m}, \bar{d})$$

come voluto (studia bene la precedente catena di diseguaglianze fin quando non sei certo d'aver capito perché è vera e perché è proprio quello che ci serve).

Ok, cominciamo. Sviluppando i quadrati nella definizione di $S(m, d)$ otteniamo

$$\begin{aligned} S(m, d) &= \frac{1}{n} \sum_{i=1}^n (x_i^2 m^2 + d^2 + y_i^2 + 2x_i m d - 2x_i y_i m - 2y_i d) \\ &= d^2 + 2(m\bar{x} - \bar{y})d + m^2 \frac{1}{n} \sum_{i=1}^n x_i^2 - 2m \frac{1}{n} \sum_{i=1}^n x_i y_i + \frac{1}{n} \sum_{i=1}^n y_i^2 , \end{aligned}$$

dove \bar{x} è la media aritmetica di x_1, \dots, x_n e \bar{y} è la media aritmetica di y_1, \dots, y_n . Quindi per ogni m fissato $d \mapsto S(m, D)$ è effettivamente una funzione quadratica di d , in quanto possiamo scrivere $S(m, d) = ad^2 + bd + c$ con

$$a = 1 , \quad b = 2(m\bar{x} - \bar{y}) , \quad c = m^2 \frac{1}{n} \sum_{i=1}^n x_i^2 - 2m \frac{1}{n} \sum_{i=1}^n x_i y_i + \frac{1}{n} \sum_{i=1}^n y_i^2 .$$

In particolare, $a = 1 > 0$ e il punto di minimo $d_0(m)$ di questa funzione è

$$d_0(m) = -\frac{b}{2a} = \bar{y} - m\bar{x} .$$

Per calcolare $S(m, d_0(m))$ ricordiamoci che il valore minimo di una funzione

⁴ Ed è questo il motivo per cui questo metodo si chiama dei *minimi quadrati*.

quadratica (con $a > 0$) è $c - b^2/4a$, per cui

$$\begin{aligned} S(m, d_0(m)) &= m^2 \frac{1}{n} \sum_{i=1}^n x_i^2 - 2m \frac{1}{n} \sum_{i=1}^n x_i y_i + \frac{1}{n} \sum_{i=1}^n y_i^2 - \frac{1}{4} 4(m\bar{x} - \bar{y})^2 \\ &= \left[\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right] m^2 - 2 \left[\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y} \right] m + \frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2 \\ &= \tilde{a} m^2 + \tilde{b} m + \tilde{c}. \end{aligned}$$

con

$$\tilde{a} = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2, \quad \tilde{b} = -2 \left[\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y} \right], \quad \tilde{c} = \frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2.$$

Dunque abbiamo ottenuto, come promesso, una funzione quadratica di m . Inoltre, il coefficiente \tilde{a} del termine quadrato è positivo; infatti, l'Osservazione 3.32 ci dice che

$$\tilde{a} = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \text{Media}(x_i^2) - \text{Media}(x_i)^2 = \text{Var}(x_i) > 0,$$

come voluto⁵.

Quindi la funzione $S(m, d_0(m))$ ammette un unico punto di minimo

$$\bar{m} = -\frac{\tilde{b}}{2\tilde{a}} = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y}}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}, \quad (4.10)$$

e la funzione $S(m, d)$ ammette un unico punto di minimo (\bar{m}, \bar{d}) con

$$\bar{d} = \bar{y} - \bar{m}\bar{x}. \quad (4.11)$$

La retta grafico della funzione $f(x) = \bar{m}x + \bar{d}$ che meglio approssima i dati si chiama *retta di regressione (lineare)*.

Osservazione 4.19 Mentre il modo migliore di calcolare \bar{d} è usare la formula (4.11), ci sono altre formule per il calcolo di \bar{m} oltre a (4.10). Prima di tutto notiamo che possiamo scrivere (4.10) così:

$$\bar{m} = \frac{\text{Media}(x_i y_i) - \text{Media}(x_i)\text{Media}(y_i)}{\text{Media}(x_i^2) - \text{Media}(x_i)^2}, \quad (4.12)$$

⁵ La varianza non può essere nulla, a meno che tutti gli x_i siano uguali; ma in tal caso sapremmo già che i dati giacciono su una retta (verticale), e quindi non avremmo neppure cominciato questi conti.

o anche come

$$\overline{m} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2}, \quad (4.13)$$

dove \overline{xy} è la media aritmetica dei prodotti x_1y_1, \dots, x_ny_n , e $\overline{x^2}$ è la media aritmetica dei quadrati x_1^2, \dots, x_n^2 , per cui \overline{m} è la differenza fra la media dei prodotti e il prodotto delle medie divisa per la differenza fra la media dei quadrati e il quadrato della media.

Un'altra formula si ottiene notando che, come abbiamo già visto, il denominatore di (4.10) è uguale a $\text{Var}(x_i) = \frac{1}{n} \sum_{i=1}^n (\overline{x} - x_i)^2$. Una formula analoga vale anche per il numeratore: infatti

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n x_i y_i - \overline{x} \cdot \overline{y} &= \frac{1}{n} \sum_{i=1}^n x_i y_i - \overline{x} \frac{1}{n} \sum_{i=1}^n y_i - \overline{y} \frac{1}{n} \sum_{i=1}^n x_i + \overline{x} \cdot \overline{y} \\ &= \frac{1}{n} \sum_{i=1}^n (\overline{x} - x_i)(\overline{y} - y_i), \end{aligned}$$

per cui possiamo scrivere

$$\overline{m} = \frac{\sum_{i=1}^n (\overline{x} - x_i)(\overline{y} - y_i)}{\sum_{i=1}^n (\overline{x} - x_i)^2}. \quad (4.14)$$

Quale formula usare dipende dalle situazioni, da quali altri calcoli hai già effettuato o devi effettuare, da quale ti ricordi meglio e anche dai tuoi gusti personali.

Rimane da stabilire quanto bene la retta di regressione approssima i dati. Una prima informazione ci è data dal valore minimo che abbiamo trovato,

$$\begin{aligned} S(\overline{m}, \overline{d}) &= \tilde{c} - \frac{\tilde{b}^2}{4\tilde{a}} = \tilde{c} - \tilde{a} \overline{m}^2 \\ &= \frac{1}{n} \sum_{i=1}^n y_i^2 - \overline{y}^2 - \frac{\left[\frac{1}{n} \sum_{i=1}^n x_i y_i - \overline{x} \cdot \overline{y} \right]^2}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \overline{x}^2} \\ &= \frac{(\overline{y^2} - \overline{y}^2)(\overline{x^2} - \overline{x}^2) - (\overline{xy} - \overline{x} \cdot \overline{y})^2}{\overline{x^2} - \overline{x}^2}, \end{aligned}$$

dove $\overline{y^2}$ è la media aritmetica di y_1^2, \dots, y_n^2 , come al solito. Ora, le coppie di dati stanno tutte sulla retta di regressione se e solo se $S(\overline{m}, \overline{d}) = 0$. Però il valore di $S(\overline{m}, \overline{d})$ da solo non è una buona misura della qualità della retta di regressione, in quanto ha il solito problema degli errori assoluti: se i dati sono grandi allora l'errore è grande in valore assoluto, anche quando è piccolo rispetto ai valori assoluti dei dati. Ci serve invece un errore relativo. Siccome $S(\overline{m}, \overline{d})$ misura la media degli errori quadratici nelle ordinate, la quantità giusta a cui confrontarla è lo scarto

quadratico medio delle ordinate, cioè la varianza⁶ $\text{Var}(y_i) = \bar{y}^2 - \bar{y}^2$. Quindi siamo condotti a considerare la quantità

$$\frac{S(\bar{m}, \bar{d})}{\text{Var}(y_i)} = 1 - \frac{(\bar{xy} - \bar{x} \cdot \bar{y})^2}{(\bar{x}^2 - \bar{x}^2)(\bar{y}^2 - \bar{y}^2)},$$

che è sempre maggiore o uguale di zero, e si annulla se e solo se tutti i dati sono sulla retta di regressione. Inoltre, è anche sempre minore o uguale di 1 (perché?); quindi anche se i dati sono molto grandi rimane con valore assoluto limitato.

Abbiamo quasi finito: rimane da fare un passaggio analogo a quello che porta dalla varianza alla deviazione standard. La quantità $S(\bar{m}, \bar{d})/\text{Var}(y_i)$ si annulla se e solo se $(\bar{xy} - \bar{x} \cdot \bar{y})^2 / (\bar{x}^2 - \bar{x}^2)(\bar{y}^2 - \bar{y}^2) = 1$, cioè (estraendo la radice quadrata) se e solo se $(\bar{xy} - \bar{x} \cdot \bar{y}) / \sqrt{(\bar{x}^2 - \bar{x}^2)(\bar{y}^2 - \bar{y}^2)} = \pm 1$. Introduciamo allora il *coefficiente di correlazione di Pearson*

$$\text{CP} = \frac{\bar{xy} - \bar{x} \cdot \bar{y}}{\sqrt{(\bar{x}^2 - \bar{x}^2)(\bar{y}^2 - \bar{y}^2)}} \in [-1, 1].$$

Per quanto abbiamo detto, il coefficiente di correlazione di Pearson misura la bontà dell'approssimazione fornita dalla retta di regressione: se è sufficientemente vicino a 1 o a -1, allora l'approssimazione è buona; se invece è vicino a 0, vuol dire che i dati non seguono affatto un andamento lineare.

Osservazione 4.20 “Sufficientemente vicino a ± 1 ” di solito vuol dire almeno 0.9 in valore assoluto; almeno 0.95 è anche meglio.

Osservazione 4.21 Il segno del coefficiente di Pearson è lo stesso di \bar{m} .

Osservazione 4.22 I conti fatti nell'Osservazione 4.19 ci forniscono un'altra formula per il coefficiente di Pearson:

$$\text{CP} = \frac{\bar{xy} - \bar{x} \cdot \bar{y}}{\text{DS}(x_i)\text{DS}(y_i)} = \frac{\sum_{i=1}^n (\bar{x} - x_i)(\bar{y} - y_i)}{\sqrt{\sum_{i=1}^n (\bar{x} - x_i)^2 \sum_{j=1}^n (\bar{y} - y_j)^2}}.$$

Concludiamo questa sezione con un esempio di calcolo della retta di regressione e del coefficiente di Pearson; altri esempi li vedremo nella Sezione 4.9.

ESEMPIO 4.8 Riprendiamo il nostro gruppo di 15 cavie; vogliamo vedere se c'è una relazione lineare fra il loro peso (in decigrammi) e la loro età (in giorni; sono cavie molto giovani). Per procedere prepariamo una tabella (Tabella 4.1) con cinque colonne: l'età (la nostra x), il peso (la nostra y), i prodotti xy , i quadrati x^2 , e i

⁶ Che è nulla se e solo se tutte le y_i sono uguali; ma in tal caso i dati sono chiaramente su una retta (orizzontale), e di nuovo non ci saremmo imbarcati in questi calcoli.

quadrati y^2 . Poi calcoliamo la media aritmetica dei dati di ciascuna colonna; con questi dati possiamo trovare i coefficienti della retta di regressione e il coefficiente di Pearson.

| Cavia | Età (x) | Peso (y) | xy | x^2 | y^2 |
|--------------|---------|----------|--------|--------|-------|
| 1 | 61 | 28 | 1708 | 3721 | 784 |
| 2 | 76 | 32 | 2432 | 5776 | 1024 |
| 3 | 80 | 37 | 2960 | 6400 | 1369 |
| 4 | 66 | 29 | 1914 | 4356 | 841 |
| 5 | 71 | 31 | 2201 | 5041 | 961 |
| 6 | 68 | 30 | 2040 | 4624 | 900 |
| 7 | 78 | 32 | 2496 | 6084 | 1024 |
| 8 | 55 | 26 | 1430 | 3025 | 676 |
| 9 | 74 | 32 | 2368 | 5476 | 1024 |
| 10 | 60 | 27 | 1620 | 3600 | 729 |
| 11 | 65 | 29 | 1885 | 4225 | 841 |
| 12 | 70 | 30 | 2100 | 4900 | 900 |
| 13 | 64 | 28 | 1792 | 4096 | 784 |
| 14 | 73 | 31 | 2263 | 5329 | 961 |
| 15 | 68 | 31 | 2108 | 4624 | 961 |
| <i>Media</i> | 68.6 | 30.2 | 2087.8 | 4751.8 | 918.6 |

TABELLA 4.1

Usando per esempio la formula (4.13) otteniamo

$$\begin{aligned}\overline{m} &= \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{2087.8 - 68.6 \cdot 30.2}{4751.8 - 68.6^2} \simeq 0.351 , \\ \overline{d} &= \overline{y} - \overline{m} \cdot \overline{x} \simeq 30.2 - 0.351 \cdot 67 \simeq 6.136 ,\end{aligned}$$

per cui la retta di regressione è

$$f(x) = 0.351x + 6.136 .$$

Il coefficiente di Pearson è

$$\text{CP} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{(\overline{x^2} - \overline{x}^2)(\overline{y^2} - \overline{y}^2)}} = \frac{2087.8 - 68.6 \cdot 30.2}{\sqrt{(4751.8 - 68.6^2)(918.6 - 30.2^2)}} \simeq 0.927 ,$$

per cui la retta di regressione approssima piuttosto bene i dati, come si può vedere dalla Figura 4.6, che contiene sia i dati sia la retta di regressione.

Esercizio 4.1 Calcola la retta di regressione e il coefficiente di Pearson partendo dai dati della Tabella 4.1 ma supponendo che la cavia 3 abbia 55 giorni d'età.

Osservazione 4.23 Come hai visto (vero?) risolvendo il precedente esercizio, la presenza anche di un solo dato spurio può falsare di molto la retta di regressione,

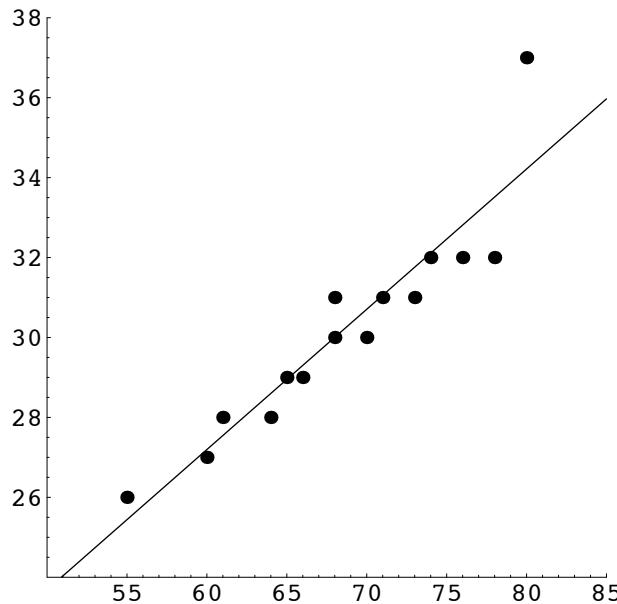


Figura 4.6 Retta di regressione.

e dare un coefficiente di Pearson molto basso. Per questo motivo nella pratica sperimentale conviene sempre esaminare i dati raccolti per eliminare dati evidentemente spuri, e (possibilmente) investigare i motivi che hanno portato alla presenza dei dati spuri (semplici errori o un fenomeno nuovo?).

4.4 Funzioni polinomiali

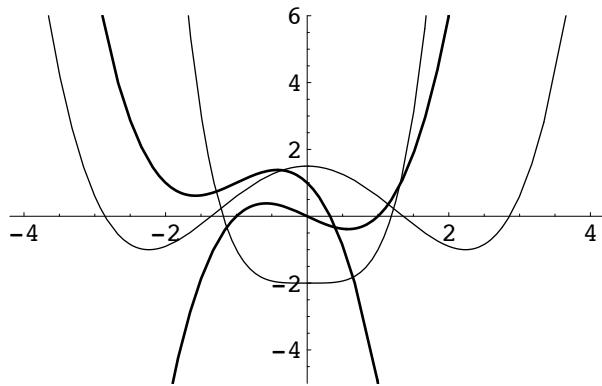
Dopo le funzioni quadratiche, si possono considerare funzioni di terzo grado, o di quarto grado, o più in generale *funzioni polinomiali*, cioè funzioni $f: \mathbb{R} \rightarrow \mathbb{R}$ espresse da un polinomio:

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 , \quad (4.15)$$

dove $n \in \mathbb{N}$ è il *grado* della funzione polinomiale (o del polinomio), e $a_0, \dots, a_n \in \mathbb{R}$ sono i *coefficienti*; si suppone sempre che $a_n \neq 0$. Non abbiamo ancora gli strumenti necessari⁷ per effettuare uno studio dettagliato di queste funzioni; mi limiterò quindi a citare alcuni fatti, in parte analoghi a quanto abbiamo già visto, che possono essere utili nel loro studio. La Figura 4.7 comunque contiene i grafici di alcune funzioni polinomiali, giusto per darti un'idea di che faccia possano avere.

La prima osservazione è che per x molto grande in valore assoluto l'addendo $a_n x^n$ in (4.15) è molto più grande degli altri, per cui il comportamento della fun-

⁷ Ne introdurremo molti nel prossimo capitolo.

**Figura 4.7** Funzioni polinomiali.

zione f per x molto grande in valore assoluto è dettato dal comportamento di $a_n x^n$. In particolare:

- se $a_n > 0$ e n è pari allora

$$\lim_{x \rightarrow \pm\infty} a_n x^n + \cdots + a_0 = \lim_{x \rightarrow \pm\infty} a_n x^n = +\infty ;$$

- se $a_n > 0$ e n è dispari allora

$$\lim_{x \rightarrow \pm\infty} a_n x^n + \cdots + a_0 = \lim_{x \rightarrow \pm\infty} a_n x^n = \pm\infty ;$$

- se $a_n < 0$ e n è pari allora

$$\lim_{x \rightarrow \pm\infty} a_n x^n + \cdots + a_0 = \lim_{x \rightarrow \pm\infty} a_n x^n = -\infty ;$$

- se $a_n < 0$ e n è dispari allora

$$\lim_{x \rightarrow \pm\infty} a_n x^n + \cdots + a_0 = \lim_{x \rightarrow \pm\infty} a_n x^n = \mp\infty .$$

CURIOSITÀ 4.5 Vediamo come dimostrare correttamente queste affermazioni. L'idea è scrivere

$$a_n x^n + \cdots + a_0 = a_n x^n \left(1 + \sum_{j=1}^n \frac{a_j}{a_n} \frac{1}{x^j} \right) , \quad (4.16)$$

e far vedere che per $|x|$ grande il termine fra parentesi non è troppo distante da 1. Poniamo

$$R = \frac{2n \max\{|a_0|, \dots, |a_n|\}}{|a_n|} \geq 2n > 1 .$$

Se $|x| > R$ e $0 \leq j \leq n-1$ abbiamo $|x|^{n-j} > |x| > R > 2n|a_j|/|a_n|$, cioè

$$\frac{|a_j|}{|a_n|} \frac{1}{|x|^{n-j}} < \frac{1}{2n} ,$$

e quindi

$$\sum_{j=0}^{n-1} \frac{|a_j|}{|a_n|} \frac{1}{|x|^{n-j}} < n \cdot \frac{1}{2n} = \frac{1}{2}.$$

Ricordando le seguenti fondamentali proprietà del valore assoluto

$$|a| - |b| \leq |a - b| \leq |a + b| \leq |a| + |b|, \quad |a \cdot b| = |a| \cdot |b|,$$

per $|x| > R$ otteniamo

$$1 + \sum_{j=0}^{n-1} \frac{a_j}{a_n} \frac{1}{x^{n-j}} \geq 1 - \left| \sum_{j=0}^{n-1} \frac{a_j}{a_n} \frac{1}{x^{n-j}} \right| \geq 1 - \sum_{j=0}^{n-1} \frac{|a_j|}{|a_n|} \frac{1}{|x|^{n-j}} \geq 1 - \frac{1}{2} = \frac{1}{2}.$$

Inoltre,

$$1 + \sum_{j=0}^{n-1} \frac{a_j}{a_n} \frac{1}{x^{n-j}} \leq 1 + \left| \sum_{j=0}^{n-1} \frac{a_j}{a_n} \frac{1}{x^{n-j}} \right| \leq 1 + \sum_{j=0}^{n-1} \frac{|a_j|}{|a_n|} \frac{1}{|x|^{n-j}} \leq 1 + \frac{1}{2} = \frac{3}{2},$$

sempre per $|x| > R$. Ricordando (4.16), non appena $a_n x^n > 0$ e $|x| > R$ otteniamo

$$\frac{3}{2} a_n x^n \geq a_n x^n + \cdots + a_0 \geq \frac{1}{2} a_n x^n,$$

mentre se $a_n x^n < 0$ e $|x| > R$ otteniamo

$$\frac{3}{2} a_n x^n \leq a_n x^n + \cdots + a_0 \leq \frac{1}{2} a_n x^n.$$

Queste due stime implicano immediatamente le affermazioni volute.

La seconda osservazione riguarda il numero di condizioni necessarie per determinare i coefficienti di un polinomio di grado n . Abbiamo visto che il grafico di una funzione lineare era completamente determinato dal passaggio per due punti, e che il grafico di una funzione quadratica era completamente determinato dal passaggio per tre punti. Analogamente, il grafico di un polinomio di grado n è completamente determinato dal passaggio per $n+1$ punti.

Il metodo per trovare il polinomio dati $n+1$ punti è analogo a quello visto per le funzioni quadratiche. Supponiamo di avere $n+1$ punti $(x_0, y_0), \dots, (x_n, y_n)$ con ascisse x_0, \dots, x_n tutte distinte. Trovare un polinomio $f(x) = a_n x^n + \cdots + a_0$ il cui grafico passi per questi punti, cioè tale che $f(x_j) = y_j$ per $j = 0, \dots, n$, equivale a risolvere il sistema lineare

$$\begin{cases} x_0^n a_n + x_0^{n-1} a_{n-1} + \cdots + a_0 = y_0, \\ \vdots \\ x_n^n a_n + x_n^{n-1} a_{n-1} + \cdots + a_0 = y_n, \end{cases}$$

di $n+1$ equazioni nelle $n+1$ incognite a_0, \dots, a_n . Per risolvere questo sistema si sottrae ogni equazione dalla successiva; se necessario, si dividono, come in (4.8), i

coefficienti del sistema ottenuto per un opportuno fattore comune, utilizzando le formule

$$x^{k+1} - y^{k+1} = (x - y) \sum_{i=0}^k x^i y^{k-i}; \quad (4.17)$$

e si ripete il procedimento col nuovo sistema. Dopo n passaggi si arriva a una sola equazione lineare con a_n come unica incognita; ricavata a_n si sostituisce il valore trovato nei sistemi precedenti, ricavando a_{n-1} e poi a_{n-2} e così via fino ad a_0 . Questa tecnica per determinare i coefficienti dei polinomi è detta *metodo delle differenze*.

CURIOSITÀ 4.6 La formula (4.17) si dimostra col seguente conto:

$$\begin{aligned} (x - y) \sum_{i=0}^k x^i y^{k-i} &= \sum_{i=0}^k x^{i+1} y^{k-i} - \sum_{i=0}^k x^i y^{k-(i-1)} \\ &= x^{k+1} + \sum_{i=0}^{k-1} x^{i+1} y^{k-i} - \sum_{h=0}^{k-1} x^{h+1} y^{k-h} - y^{k+1} \\ &= x^{k+1} - y^{k+1}, \end{aligned}$$

dove abbiamo posto $h = i - 1$ nell'ultima sommatoria.

ESEMPIO 4.9 Studiamo nuovamente i semi di pomodoro degli Esempi 4.1 e 4.7. Sai già che alla temperatura di 12°C germoglia il 40% dei semi, alla temperatura di 15°C germoglia il 70% dei semi, e che alla temperatura di 9°C germoglia il 20% dei semi. Non contento, il tuo assistente effettua un'ulteriore misura, scoprendo che alla temperatura di 18°C germoglia l'85% dei semi. Trova un polinomio di terzo grado che rappresenti questi dati. Dobbiamo trovare $a_0, a_1, a_2, a_3 \in \mathbb{R}$ che soddisfano il sistema

$$\begin{cases} 729a_3 + 81a_2 + 9a_1 + a_0 = 20, \\ 1728a_3 + 144a_2 + 12a_1 + a_0 = 40, \\ 3375a_3 + 225a_2 + 15a_1 + a_0 = 70, \\ 5832a_3 + 324a_2 + 18a_1 + a_0 = 85. \end{cases}$$

La prima serie di sottrazioni ci dà

$$\begin{cases} 999a_3 + 63a_2 + 3a_1 = 20, \\ 1647a_3 + 81a_2 + 3a_1 = 30, \\ 2457a_3 + 99a_2 + 3a_1 = 15. \end{cases}$$

Siccome i coefficienti di a_1 sono uguali nelle tre equazioni, non abbiamo bisogno di effettuare divisioni e possiamo procedere direttamente con la seconda serie di sottrazioni:

$$\begin{cases} 648a_3 + 18a_2 = 10, \\ 810a_3 + 18a_2 = -15. \end{cases}$$

L'ultima sottrazione ci dà $162a_3 = -25$, cioè $a_3 = -25/162$. Mettendo questo valore nelle equazioni precedenti e risalendo troviamo $a_2 = 55/9$, $a_1 = -1265/18$,

e $a_0 = 270$, per cui il polinomio cercato è

$$P(T) = -\frac{25}{162}T^3 + \frac{55}{9}T^2 - \frac{1265}{18}T + 270.$$

La Figura 4.8 contiene sia i dati sia i grafici delle funzioni che abbiamo ottenuto nei vari esempi.

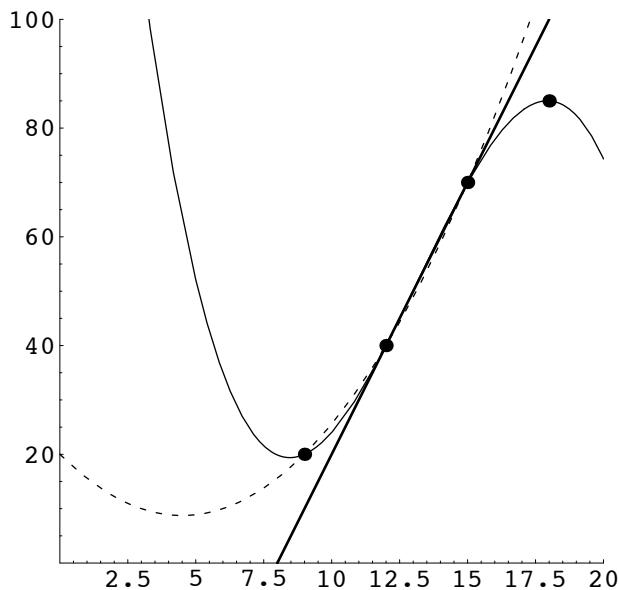


Figura 4.8

Osservazione 4.24 Gli Esempi 4.1, 4.7 e 4.9 mostrano che i conti nel metodo delle differenze sono più semplici se le ascisse x_0, \dots, x_n sono equispaziate, cioè se $x_1 - x_0 = x_2 - x_1 = \dots = x_{n-1} - x_n$.

Esercizio 4.2 Usa il metodo dei minimi quadrati per determinare la retta di regressione per i dati dell'Esempio 4.9, calcola il coefficiente di Pearson, e confronta pregi e difetti delle varie formule (retta di regressione inclusa) che abbiamo trovato per rappresentare la relazione fra temperatura e percentuale di semi germinati.

L'ultima osservazione che ci servirà riguarda le radici di un polinomio. Una radice di un polinomio $f(x)$ è un numero reale $x_0 \in \mathbb{R}$ tale che $f(x_0) = 0$. La regola di Ruffini dice che $x_0 \in \mathbb{R}$ è radice del polinomio f se e solo se esiste un polinomio q tale che $f(x) = (x - x_0)q(x)$, dove q ha grado di uno minore rispetto al grado di f . Ora, se x_0 è radice anche di q , deve esistere un terzo polinomio q_1 , di grado minore di quello di q , tale che $q(x) = (x - x_0)q_1(x)$, per cui $f(x) = (x - x_0)^2q_1(x)$. Ripetendo questo procedimento, prima o poi troveremo un numero naturale $r > 0$

(e minore o uguale del grado di f) e un polinomio q_r tali che

$$f(x) = (x - x_0)^r q_r(x) \quad \text{con} \quad q_r(x_0) \neq 0;$$

il numero r è detto *molteplicità* di x_0 come radice di f .

CURIOSITÀ 4.7 Se q_r ha una radice x_1 di molteplicità $s > 0$, possiamo ripetere questa costruzione con q_r , trovando un polinomio p_s tale che $f(x) = (x - x_0)^r (x - x_1)^s p_s(x)$. Procedendo in questo modo si riesce a dimostrare che ogni polinomio f si può scrivere in modo unico come prodotto

$$f(x) = a p_1(x)^{r_1} \cdots p_k(x)^{r_k}, \quad (4.18)$$

dove $a \in \mathbb{R}$ è il coefficiente direttore di f , r_1, \dots, r_k sono numeri interi maggiori di zero, e p_1, \dots, p_r sono polinomi *monici* (cioè con coefficiente direttore uguale a 1) *irriducibili* (cioè non si possono scrivere come prodotto di due altri polinomi monici). I polinomi irriducibili sono l'equivalente per i polinomi dei numeri primi per i numeri naturali; e (4.18) è l'equivalente della decomposizione in fattori primi di un numero naturale. Infine, si può anche dimostrare che i polinomi monici irriducibili a coefficienti reali sono o polinomi lineari della forma $x - x_0$ oppure polinomi quadratici senza radici reali (cioè con discriminante negativo).

4.5 Funzioni potenza

Un'altra famiglia importante di funzioni è costituito dalle *funzioni potenza*, che sono funzioni della forma

$$f(x) = ax^p,$$

dove $a \neq 0$ è un numero reale e p è un numero razionale (ma vedi anche l'Osservazione 4.26), detto *esponente* della funzione potenza.

Osservazione 4.25 Se p è un numero naturale, $p \in \mathbb{N} = \{0, 1, 2, \dots\}$, la funzione $f(x) = ax^p$ è una particolare funzione polinomiale, e quindi è definita su tutta la retta reale: $f: \mathbb{R} \rightarrow \mathbb{R}$. Se p è un numero intero negativo, $p \in \mathbb{Z}^- = \{-1, -2, \dots\}$, la funzione f è una particolare funzione razionale (vedi la prossima sezione) ed è definita per $x \neq 0$, cioè $f: \mathbb{R}^* \rightarrow \mathbb{R}$. Infine, se p è un numero razionale non intero, $p \in \mathbb{Q} \setminus \mathbb{Z}$, allora f è definita solo per $x \geq 0$, cioè $f: \mathbb{R}^+ \rightarrow \mathbb{R}$.

Osservazione 4.26 Come accenneremo nelle Sezioni 4.7 e 4.8 e vedremo nel prossimo capitolo, è possibile dare un senso anche alle potenze *irrazionali* di un numero non negativo, per cui potremo considerare funzioni potenza con esponente qualsiasi (ma solo con argomento reale non negativo).

CURIOSITÀ 4.8 Supponiamo che $p \in \mathbb{R} \setminus \mathbb{Q}$ sia un numero irrazionale, e $x \in \mathbb{R}^+$ un numero non negativo. Siccome i numeri razionali possono approssimare bene quanto vogliamo qualsiasi numero reale, un modo per calcolare la potenza irrazionale x^p si basa sul fatto che esiste un numero reale y tale che la potenza *razionale* x^q è arbitrariamente vicina a y non appena q è un numero razionale sufficientemente vicino a p ; allora si pone $x^p = y$. Quindi le potenze razionali forniscono approssimazioni arbitrariamente buone delle potenze irrazionali. Un altro modo per esprimere questo concetto è dire che per ogni $\varepsilon > 0$ (arbitrariamente piccolo) esiste un $\delta > 0$ (sufficientemente piccolo) tale che se q è un numero razionale che dista da p meno di δ (cioè $|q - p| < \delta$) allora x^q dista da y meno di ε (cioè $|x^q - y| < \varepsilon$).

Osservazione 4.27 Nella Sezione 4.8 vedremo anche che il grafico di una funzione potenza $f(x) = ax^p$ è completamente determinato dal passaggio per due punti.

Le funzioni potenza ax^p con $p \in \mathbb{N}$ hanno un comportamento molto simile a quello di ax se p è dispari, e a quello di ax^2 se p è pari. Infatti, con le tecniche introdotte nelle Sezioni 4.1 e 4.2 non è difficile vedere (esercizio per te e il tuo assistente) che

- se p è dispari, la funzione $f(x) = ax^p$ è monotona (crescente se $a > 0$, decrescente se $a < 0$), e

$$\lim_{x \rightarrow \pm\infty} ax^p = \begin{cases} \pm\infty & \text{se } a > 0, \\ \mp\infty & \text{se } a < 0; \end{cases}$$

- se p è pari (e non nullo), la funzione $f(x) = ax^p$ ha un punto di minimo (se $a > 0$) o un punto di massimo (se $a < 0$) in $x = 0$, è monotona (crescente o decrescente a seconda del segno di a), nelle semirette $(-\infty, 0]$ e $[0, +\infty)$, e

$$\lim_{x \rightarrow \pm\infty} ax^p = \begin{cases} +\infty & \text{se } a > 0, \\ -\infty & \text{se } a < 0. \end{cases}$$

Osservazione 4.28 Nota che il tipo di monotonia su $\mathbb{R}^+ = [0, +\infty)$ e il limite a $+\infty$ di ax^p dipende solo dal segno di a e non dalla parità di p .

Osservazione 4.29 Il confronto fra funzioni potenza con diverso esponente può fornire interessanti conseguenze biologiche legate a questioni di scala. Prendiamo un particolare individuo di una data specie animale (o vegetale), e scegliamo una sua lunghezza caratteristica ℓ : per esempio, possiamo indicare con ℓ il diametro della testa. Tutte le altre lunghezze di questo particolare individuo sono banalmente proporzionali a ℓ : i coefficienti di proporzionalità si ottengono semplicemente dividendo per ℓ la lunghezza che si vuole considerare. Se ora prendiamo un altro individuo della stessa specie, i coefficienti di proporzionalità delle sue lunghezze rispetto a ℓ saranno un po' diversi, ma non *tropo* diversi: per esempio, il secondo individuo sarà un po' più alto o un po' più basso, ma facendo parte della stessa specie è molto improbabile che sia alto più del doppio o meno della metà. Quindi possiamo considerare questa lunghezza ℓ come *rappresentativa* della specie; tutte le altre lunghezze in tutti gli altri individui della stessa specie saranno proporzionali a ℓ con coefficienti di proporzionalità approssimativamente costanti. Di conseguenza, tutte le *superfici* di individui della stessa specie saranno proporzionali a ℓ^2 , con coefficienti di proporzionalità approssimativamente costanti; e tutti i *volumi* saranno proporzionali a ℓ^3 , con coefficienti di proporzionalità approssimativamente costanti. Ora, i fenomeni di scambio con l'esterno (assorbimento di ossigeno, emissione di calore, eccetera) di un individuo avvengono usualmente attraverso la superficie, e quindi avranno andamenti proporzionali a ℓ^2 ; invece, i fenomeni metabolici (consumo di ossigeno, produzione di calore, eccetera) sono di solito proporzionali al volume (al numero di cellule coinvolte), e quindi proporzionali a ℓ^3 . La conseguenza di tutto ciò è che (come sarà chiarito dai prossimi esempi) *non è possibile variare eccessivamente le dimensioni di una data specie animale senza danneggiare l'equilibrio fra il metabolismo interno e l'ambiente esterno che le permette di vivere*; in un certo

senso, i rapporti fra il metabolismo e la forma di una specie ne determinano le dimensioni ideali.

ESEMPIO 4.10 Con buona pace dei film dell'orrore, un ragno gigante, ottenuto ingrandendo 100 volte un ragno usuale, ha poche possibilità di sopravvivere. Il consumo di ossigeno è proporzionale al volume del ragno, volume che è passato da un multiplo di ℓ^3 a un uguale multiplo di $(10\ell)^3 = 1000\ell^3$. D'altra parte, l'assorbimento di ossigeno è proporzionale alla superficie interna dei polmoni, superficie che è passata da un multiplo di ℓ^2 a un uguale multiplo di $(10\ell)^2 = 100\ell^2$. Il consumo di ossigeno è quindi aumentato di 1000 volte, mentre l'assorbimento di ossigeno solo di 100 volte; il ragno gigante riceve solo 1/10 dell'ossigeno che gli servirebbe, e quindi muore soffocato sotto gli occhi increduli dell'eroe del film.

In maniera analoga, con buona pace dei film di fantascienza, un uomo rimpicciolito di 10 volte si sentirebbe piuttosto male. Infatti, la perdita di calore attraverso l'epidermide è proporzionale alla superficie del corpo, cioè a ℓ^2 , e quindi si è ridotta di 1/100. Ma il calore prodotto dal corpo umano, necessario alla nostra sopravvivenza visto che siamo animali a sangue caldo, è proporzionale al volume del corpo, cioè a ℓ^3 , e quindi si è ridotto di 1/1000. Quindi un uomo rimpicciolito 10 volte perderebbe attraverso l'epidermide 10 volte più calore di quello che produce, e quindi probabilmente morirebbe di freddo.

Osservazione 4.30 La superficie di assorbimento dell'ossigeno nei polmoni in realtà ha una struttura frastagliata, di tipo frattale, e di conseguenza l'assorbimento di ossigeno è più efficiente diventando proporzionale a ℓ^p con $p > 2$; ma in ogni caso non raggiunge ℓ^3 .

ESEMPIO 4.11 È noto che la forza muscolare di un muscolo delle gambe è approssimativamente proporzionale al numero di fibre muscolari che lo compongono, e quindi è proporzionale alla superficie trasversa del muscolo, cioè a ℓ^2 . Inoltre, l'energia prodotta dal muscolo è proporzionale alla forza per la lunghezza, e quindi è uguale a $c_1\ell^3$ per un'opportuna costante $c_1 > 0$. D'altra parte, l'energia necessaria per un salto di altezza h è proporzionale al prodotto dell'altezza per il peso del corpo, e quindi è uguale a $c_2h\ell^3$ per un'opportuna costante $c_2 > 0$. Ne segue che la massima altezza possibile h di un salto deve soddisfare $c_1\ell^3 = c_2h\ell^3$, cioè $h = c_1/c_2$; in particolare, h non dipende da ℓ .

Ora, una pulce comune è in grado di saltare a un'altezza pari a 200 volte la propria altezza; quanto sarà in grado di saltare una pulce 10 volte più grande? E una pulce 100 volte più grande? La pulce comune è in grado di saltare a un'altezza $h = 200c_3\ell$ per una costante $c_3 > 0$ opportuna. Una pulce 10 volte più grande avrà lunghezza caratteristica 10ℓ , ma uguali costanti di proporzionalità c_1 , c_2 e c_3 . Siccome la massima altezza possibile per un salto dipende solo dalle costanti di proporzionalità, la pulce 10 volte più grande può saltare solo alla stessa altezza della pulce comune, e quindi al massimo a 20 volte la propria altezza. Analogamente, una pulce 100 volte più grande riuscirà a saltare solo il doppio della propria altezza, e una pulce mastodontica 1000 volte più grande della pulce comune riuscirà a saltare solo un quinto della propria altezza. Per intenderci, se una pulce comune

alta circa 1 mm riesce a saltare 20 cm, una pulce alta 1 metro riuscirebbe a saltare sempre soltanto 20 cm...

Le funzioni $f(x) = ax^p$ con esponente $p \in \mathbb{Q}^+ \setminus \mathbb{N}$ razionale positivo non intero sono definite solo su \mathbb{R}^+ , e hanno comportamento analogo a quelle con esponente naturale: sono crescenti se $a > 0$, decrescenti se $a < 0$, e il loro limite all'infinito è uguale a $\pm\infty$ a seconda del segno di a . La Figura 4.9 contiene il grafico di alcune di queste (con $a = 1$). Nota che quelle con esponente maggiore di 1 hanno la concavità rivolta verso l'alto, mentre quelle con esponente minore di 1 hanno la concavità rivolta verso il basso. Nel prossimo capitolo vedremo come verificare rigorosamente questa affermazione.

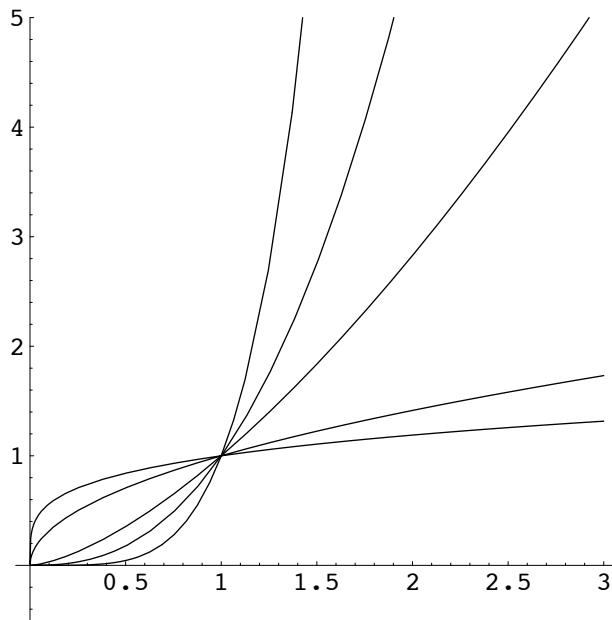


Figura 4.9 Funzioni potenza.

Più interessanti sono le funzioni $f(x) = ax^p$ con $p \in \mathbb{Q}^-$; siccome (almeno quando $p \in \mathbb{Z}^-$) sono funzioni razionali, le discutiamo nella prossima sezione (Osservazione 4.32).

4.6 Funzioni razionali

Una *funzione razionale* è un quoziente di polinomi:

$$f(x) = \frac{a_m x^m + a_{m-1} x^{m-1} + \cdots + a_0}{b_n x^n + b_{n-1} x^{n-1} + \cdots + b_0},$$

con $m, n \in \mathbb{N}$, $a_0, \dots, a_m, b_0, \dots, b_n \in \mathbb{R}$ e $a_m, b_n \neq 0$ (e di solito si assume anche $n \geq 1$, perché altrimenti f sarebbe un polinomio); il numero $d = \max\{m, n\}$ è

detto *grado* della funzione razionale. Ovviamente, lo studio dell'andamento delle funzioni razionali generiche, come per i polinomi, richiede strumenti che ancora non abbiamo; a parte alcune osservazioni finali, ci concentreremo quindi sulle funzioni razionali di grado 1, note anche come *funzioni lineari fratte*.

L'esempio più semplice di funzione lineare fratta è la funzione (potenza)

$$f(x) = \frac{a}{x} = ax^{-1},$$

con $a \neq 0$. Rappresenta le relazioni di proporzionalità inversa: infatti, un punto (x, y) appartiene al grafico di f se e solo se

$$xy = a,$$

per cui il prodotto fra l'argomento e il valore della funzione è costante su tutto il dominio della funzione.

Osservazione 4.31 In particolare, basta conoscere un punto (x_0, y_0) del grafico per determinare la funzione a/x , in quanto $a = x_0 y_0$.

La prima osservazione importante è che la funzione $f(x) = a/x$ non è definita su tutto l'asse reale: il suo dominio non è \mathbb{R} . Infatti, il quoziente a/x non è definito per $x = 0$, per cui la funzione f è definita solo su $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$. Si dice anche che 0 è una *singolarità* per la funzione f .

Quando una funzione ha una singolarità (un punto in cui non è definita), è importante cercare di capire come si comporta vicino alla singolarità. Cominciamo supponendo $a > 0$ e $x > 0$. Quando $x > 0$ diventa piccolo (per esempio, minore di $1/n$ per n arbitrariamente grande) allora $1/x$ diventa grande (per esempio, maggiore di n) e quindi anche a/x diventa grande (per esempio, maggiore di an). In altre parole, *possiamo rendere $f(x) = a/x$ arbitrariamente grande a patto di scegliere x sufficientemente piccolo e positivo*.

Abbiamo già visto come tradurre in simboli i concetti di “arbitrariamente grande” e “sufficientemente grande”; una procedura analoga si usa per il concetto di “sufficientemente piccolo”. La frase precedente diventa: per ogni $M > 0$ (arbitrariamente grande) esiste $\delta > 0$ (sufficientemente piccolo) tale che se $0 < x < \delta$ allora $f(x) > M$. In simboli,

$$\forall M > 0 \quad \exists \delta > 0: \quad 0 < x < \delta \implies f(x) > M.$$

Usando la terminologia dei limiti, diremo che *il limite di $f(x)$ per x che tende a 0 da destra (o da sopra, o che tende a 0^+) è $+\infty$* :

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

Quando $x < 0$ negativo diventa piccolo (in valore assoluto), $1/x$ diventa grande in valore assoluto ma rimane negativo, cioè diventa molto negativo. Quindi, sempre assumendo $a > 0$, *possiamo rendere $f(x) = a/x$ arbitrariamente negativa a*

patto di scegliere x sufficientemente piccolo e negativo. In altre parole, per ogni $M > 0$ (arbitrariamente grande) esiste $\delta > 0$ (sufficientemente piccolo) tale che se $-\delta < x < 0$ allora $f(x) < -M$. In simboli,

$$\forall M > 0 \quad \exists \delta > 0: \quad -\delta < x < 0 \implies f(x) < -M .$$

Usando la terminologia dei limiti, diremo che *il limite di $f(x)$ per x che tende a 0 da sinistra (o da sotto, o che tende a 0^-) è $-\infty$* :

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

Una conseguenza di questo comportamento è che il grafico di f si avvicina sempre più all'asse delle ordinate quando x tende a zero (si avvicina in alto se $x > 0$, in basso se $x < 0$). Si dice che l'asse delle ordinate è un *asintoto verticale* della funzione f .

Ovviamente, se $a < 0$ i segni si invertono; lascio a te il compito di dimostrare che

$$\text{se } a < 0 \text{ allora } \lim_{x \rightarrow 0^\pm} f(x) = \mp\infty ,$$

e di scrivere a parole e in simboli il significato di questa formula.

Vediamo ora cosa succede allontanandoci da zero. Cominciamo come al solito supponendo $a > 0$. Se $0 < x_0 < x_1$ allora $0 < 1/x_1 < 1/x_0$ e $0 < a/x_1 < a/x_0$; quindi $0 < x_0 < x_1$ implica $f(x_0) > f(x_1) > 0$, per cui f è strettamente decrescente nella semiretta $(0, +\infty)$. In modo analogo si dimostra che f è strettamente decrescente (ma negativa) nella semiretta $(-\infty, 0)$. Se invece $a < 0$ s'inverte tutto; riassumendo,

- se $a > 0$, la funzione $f(x) = a/x$ è strettamente decrescente e negativa in $(-\infty, 0)$, mentre è strettamente decrescente e positiva in $(0, +\infty)$;
- se $a < 0$, la funzione $f(x) = a/x$ è strettamente crescente e positiva in $(-\infty, 0)$, mentre è strettamente crescente e negativa in $(0, +\infty)$.

In particolare, quando x diventa grande (e $a > 0$) la funzione $f(x) = a/x$ decresce rimanendo positiva e diventando arbitrariamente piccola. Una cosa simile accade quando x diventa molto negativo (e quando $a < 0$): cambia il segno, cambia la crescenza, ma in ogni caso $f(x)$ diventa arbitrariamente piccolo in valore assoluto a patto di scegliere x sufficientemente grande o sufficientemente negativo. Ormai avrai capito il trucco per tradurre espressioni quali “arbitrariamente piccolo”: la frase precedente diventa “per ogni $\varepsilon > 0$ (arbitrariamente piccolo) esiste $M > 0$ (sufficientemente grande) tale che se $x > M$ o $x < -M$ allora $|f(x)| < \varepsilon$ ”. In simboli,

$$\forall \varepsilon > 0 \quad \exists M > 0: \quad x > M \text{ o } x < -M \implies |f(x)| < \varepsilon ,$$

o anche

$$\lim_{x \rightarrow \pm\infty} f(x) = 0 ,$$

che si legge “il limite di $f(x)$ per x che tende a $\pm\infty$ è 0”, o anche “ $f(x)$ tende a 0 per x che tende a $\pm\infty$ ”. Una conseguenza di questo comportamento è che il

grafico di f si avvicina sempre più all'asse delle ascisse quando x tende a $\pm\infty$ (si avvicina da sopra a $+\infty$ se $a > 0$ e a $-\infty$ se $a < 0$, e da sotto a $-\infty$ se $a > 0$ e a $+\infty$ se $a < 0$). Si dice che l'asse delle ascisse è un *asintoto orizzontale* della funzione f . La Figura 4.10 contiene il grafico di $f(x) = a/x$ con $a > 0$, che riassume visivamente tutte le proprietà che abbiamo discusso, asintoti compresi. Questo grafico è un esempio di *iperbole equilatera*⁸.

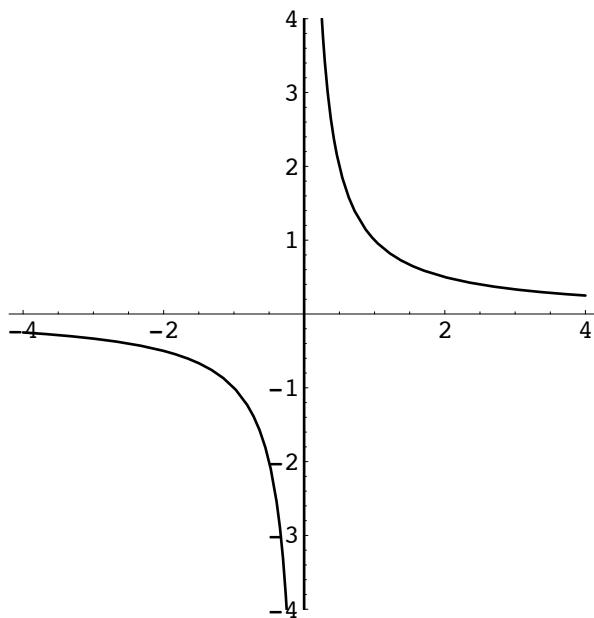


Figura 4.10 $f(x) = 1/x$.

Osservazione 4.32 Le funzioni potenza $f(x) = ax^p$ con p razionale negativo possono venire studiate con tecniche analoghe. In particolare, su $(0, +\infty)$ sono tutte positive e strettamente decrescenti (se $a > 0$, o negative e crescenti se $a < 0$), hanno limite 0 a $+\infty$ con la retta delle ascisse come asintoto orizzontale, e limite $\pm\infty$ (a seconda del segno di a) per x che tende a 0^+ , con la retta delle ordinate come asintoto verticale. Su $(-\infty, 0)$ sono definite solo quando $p \in \mathbb{Z}^-$, e in tal caso hanno un andamento analogo a quello di a/x in $(-\infty, 0)$ se p è dispari, e a quello di $|a/x|$ se p è pari. In particolare, se p è pari abbiamo che $f(x)$ diventa arbitrariamente grande a patto di prendere x sufficientemente piccolo in valore assoluto, cioè per ogni $M > 0$ (arbitrariamente grande) esiste $\delta > 0$ (sufficientemente piccolo)

⁸ Il termine “equilatera” serve a indicare che i due asintoti sono ortogonali. In generale, un’iperbole è il luogo dei punti del piano per cui il valore assoluto delle differenze delle distanze da due punti dati (detti *fuochi*) è costante. Un’iperbole ha sempre due asintoti, ma non necessariamente ortogonali.

tale che $0 < |x| < \delta$ implica $f(x) > M$, ovvero

$$\lim_{x \rightarrow 0} f(x) = +\infty$$

senza bisogno di distinguere se x tende a 0 da destra o da sinistra.

Lo studio di qualsiasi funzione razionale fratta può venire ricondotto a quello di a/x , proprio come avevamo ricondotto lo studio di qualsiasi funzione quadratica a quello di x^2 . Infatti, se $c \neq 0$ si ha

$$\frac{ax+b}{cx+d} = \frac{(a/c)x + b/c}{x + d/c} = \frac{\frac{a}{c}(x + d/c) + \frac{b}{c} - \frac{ad}{c^2}}{x + d/c} = \frac{a}{c} + \frac{(bc-ad)/c^2}{x + d/c}. \quad (4.19)$$

Ricordando la Sezione 4.2 vediamo che il grafico di $f(x) = (ax+b)/(cx+d)$ si ottiene a partire da quello di $1/x$ con le seguenti operazioni:

- moltiplichiamo le ordinate per $(bc-ad)/c^2$;
- sottraiamo a/c alle ordinate, traslando il grafico in direzione verticale della quantità a/c ;
- sommiamo d/c alle ascisse, traslando il grafico in direzione orizzontale della quantità $-d/c$.

Quindi la funzione $f(x) = (ax+b)/(cx+d)$ soddisfa le seguenti proprietà:

- ha una singolarità in $x_0 = -d/c$;
- il suo grafico è un'iperbole equilatera con asintoto orizzontale la retta $y = a/c$ e asintoto verticale la retta $x = -d/c$;
- nelle semirette $(-\infty, -d/c)$ e $(-d/c, +\infty)$ è strettamente decrescente se si ha $bc-ad > 0$, e strettamente crescente se $bc-ad < 0$ (ed è costante se $bc-ad=0$).

La presenza dell'asintoto orizzontale $y = a/c$ vuol dire che $f(x)$ è arbitrariamente vicina al valore a/c non appena x è sufficientemente grande o sufficientemente negativo. Ora, $f(x)$ è arbitrariamente vicina ad a/c se e solo se la differenza $f(x) - a/c$ è arbitrariamente piccola in valore assoluto. Quindi dire che $y = a/c$ è un asintoto orizzontale per f equivale a dire che *per ogni $\varepsilon > 0$ (arbitrariamente piccolo) esiste $M > 0$ (sufficientemente grande) tale che se $x > M$ o $x < -M$ allora $|f(x) - a/c| < \varepsilon$* . In simboli,

$$\forall \varepsilon > 0 \quad \exists M > 0: \quad x > M \text{ o } x < -M \implies |f(x) - a/c| < \varepsilon,$$

o anche

$$\lim_{x \rightarrow \pm\infty} f(x) = a/c.$$

Analogamente, la presenza dell'asintoto verticale $x = -d/c$ vuol dire che $f(x)$ è arbitrariamente grande (o arbitrariamente negativa) non appena x è sufficientemente vicino a $x_0 = -d/c$. Ora, x è sufficientemente vicina a x_0 se e solo se la differenza $x-x_0$ è arbitrariamente piccola. Quindi dire che $x = x_0$ è un asintoto verticale per f equivale (almeno quando $bc-ad > 0$) a dire che *per ogni $M > 0$ (arbitrariamente grande) esiste $\varepsilon > 0$ (sufficientemente piccolo) tale che se $0 < |x - x_0| < \varepsilon$ allora $|f(x)| > M$* .

allora $f(x) > M$, e se $-\varepsilon < x - x_0 < 0$ allora $f(x) < -M$. Usando il simbolo di limite questo si scrive

$$\lim_{x \rightarrow x_0^+} f(x) = +\infty \quad \text{e} \quad \lim_{x \rightarrow x_0^-} f(x) = -\infty.$$

Lascio a te il compito di scrivere cosa succede se $bc - ad < 0$. La Figura 4.11 contiene il grafico di una funzione lineare fratta con rappresentati anche gli asintoti.

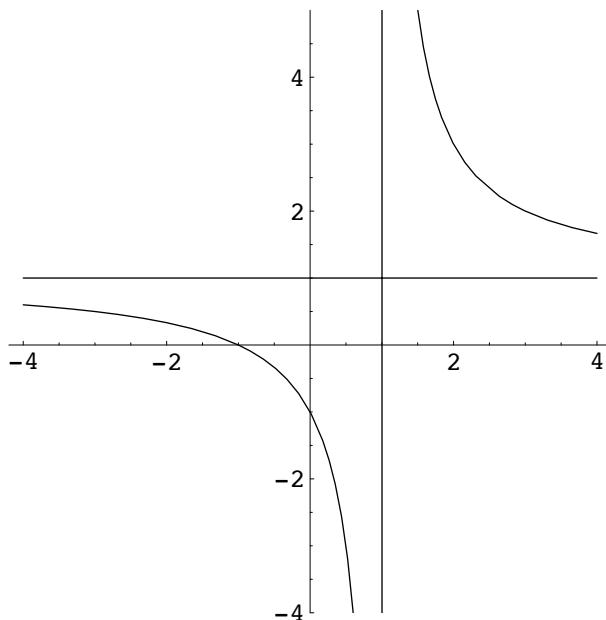


Figura 4.11 $f(x) = (x + 1)/(x - 1)$.

Osservazione 4.33 Una conseguenza immediata di (4.19) è che i punti (x, y) del grafico di $f(x) = (ax + b)/(cx + d)$ sono tutti i punti del piano che soddisfano la condizione

$$(x - \alpha)(y - \beta) = k$$

con $\alpha = -d/c$, $\beta = a/c$ e $k = (bc - ad)/c^2$.

Osservazione 4.34 Abbiamo visto cosa vuol dire che una funzione ha limite infinito quando x tende all'infinito; cosa vuol dire che ha limite un valore finito quando x tende all'infinito; e cosa vuol dire che ha limite infinito quando x tende a un valore finito (da destra, da sinistra o da entrambi i lati). Rimane da dire cosa vuol dire che ha limite un valore finito quando x tende a un valore finito. La definizione non dovrebbe stupirti: diremo che la funzione f ha *limite $\ell \in \mathbb{R}$ quando x tende a x_0* , e scriveremo

$$\lim_{x \rightarrow x_0} f(x) = \ell,$$

se $f(x)$ si avvicina arbitrariamente a ℓ a patto di prendere x sufficientemente vicino a x_0 , o, in altre parole, se per ogni $\varepsilon > 0$ (arbitrariamente piccolo) esiste $\delta > 0$ (sufficientemente piccolo) tale che $0 < |x - x_0| < \delta$ implica $|f(x) - \ell| < \varepsilon$.

Osservazione 4.35 La condizione $0 < |x - x_0| < \delta$ è equivalente a richiedere $0 < x - x_0 < \delta$ oppure $-\delta < x - x_0 < 0$. Utilizzando solo una di queste due condizioni otteniamo (come visto prima) il concetto di limite per x che tende a x_0 da sopra (cioè $x \rightarrow x_0^+$) oppure da sotto (cioè $x \rightarrow x_0^-$).

CURIOSITÀ 4.9 Possiamo usare il concetto di limite per definire la continuità di una funzione. Per l'esattezza, una funzione $f: I \rightarrow \mathbb{R}$, dove $I \subseteq \mathbb{R}$ è un intervallo, è continua in un punto $x_0 \in I$ se

$$\lim_{x \rightarrow x_0} f(x) = f(x_0),$$

cioè se il suo valore in x_0 coincide con il suo limite (sia da sopra che da sotto) in x_0 , o, ancora, se $f(x)$ diventa arbitrariamente vicino a $f(x_0)$ a patto di prendere x sufficientemente vicino a x_0 . La funzione $f: I \rightarrow \mathbb{R}$ è poi detta continua se lo è in ogni punto del suo dominio I .

Vediamo ora un'applicazione biologica delle funzioni lineari fratte.

ESEMPIO 4.12 In un esperimento si trova⁹ che la velocità v (in cm/sec) con cui un muscolo sartorio della coscia di una rana si estende per sollevare un peso p (in grammi) soddisfa la relazione

$$v(p) = 0.95 \left(\frac{70 - p}{p + 12} \right).$$

In particolare, questa funzione ha una singolarità in $p = -12$; ma siccome chiaramente ci interessa solo per $p \geq 0$, la presenza della singolarità non è un problema. È una funzione lineare fratta della forma $(ax + b)/(cx + d)$ con $a = -0.95$, $b = 66.5$, $c = 1$ e $d = 12$. In particolare, $(bc - ad)/c^2 = 77.9 > 0$ per cui v è strettamente decrescente per $p > -12$ (in altre parole, maggiore il peso più lentamente si estende il muscolo, osservazione piuttosto ragionevole). Di conseguenza, la massima velocità di estensione si ha per $p = 0$, cioè in assenza di carico, e vale $v(0) \approx 5.54$ cm/sec. Il limite all'infinito $a/c = -0.95$ è negativo; dunque il grafico deve intersecare l'asse delle ascisse. Infatti $v(70) = 0$, che vuol dire che se $p = 70$ g la gamba della rana non riesce a estendersi (velocità zero!), cioè la rana non riesce a sollevare un peso di 70 g (o maggiore). In particolare, questa formula può essere valida solo per valori di p nell'intervallo $[0, 70]$.

Vediamo ora cosa possiamo dire sul comportamento di una funzione razionale qualsiasi

$$f(x) = \frac{p(x)}{q(x)} = \frac{a_m x^m + a_{m-1} x^{m-1} + \cdots + a_0}{b_n x^n + b_{n-1} x^{n-1} + \cdots + b_0}. \quad (4.20)$$

⁹ Nella Sezione 4.9 vedremo come il metodo dei minimi quadrati può essere usato anche per interpolare funzioni lineari fratte.

Cominciamo col capire dove è definita. Gli unici punti in cui potrebbero esserci dei problemi sono le radici del denominatore q . Se x_0 è una radice di q e il numeratore non si annulla in x_0 , cioè $p(x_0) \neq 0$, allora per calcolare $f(x_0)$ dovremmo dividere per zero, che non è possibile; quindi *le radici del denominatore che non sono radici del numeratore sono sicuramente singolarità per f* . Se invece x_0 è radice anche del numeratore, dobbiamo confrontare le molteplicità. Nella Sezione 4.4 abbiamo visto che se x_0 è radice sia di p che di q possiamo scrivere $p(x) = (x - x_0)^r p_1(x)$ e $q(x) = (x - x_0)^s q_1(x)$ con $p_1(x_0), q_1(x_0) \neq 0$. Quindi

$$f(x) = \frac{(x - x_0)^r p_1(x)}{(x - x_0)^s q_1(x)} = (x - x_0)^{r-s} \frac{p_1(x)}{q_1(x)}. \quad (4.21)$$

Ma allora se $r \geq s$ il punto x_0 non è una singolarità di f , in quanto $q_1(x_0) \neq 0$; invece x_0 rimane una singolarità di f se $r < s$, perché in tal caso è $(x - x_0)^{r-s}$ ad avere una singolarità in x_0 (l'esponente è negativo). Inoltre (4.21) implica che, semplificando un'opportuna potenza di $x - x_0$, possiamo esprimere f come quoziente di polinomi tali che x_0 non sia una radice comune di numeratore e denominatore. Ripetendo questo procedimento per tutte le radici del denominatore troviamo che *ogni funzione razionale si può esprimere come rapporto di polinomi privi di radici comuni, e in tal caso le singolarità coincidono con le radici del denominatore*.

Supponiamo allora che $f(x) = p(x)/q(x)$ sia una funzione razionale tale che p e q non abbiano radici comuni, e sia x_0 una radice del denominatore di molteplicità $r > 0$. Per quanto visto possiamo scrivere

$$f(x) = \frac{1}{(x - x_0)^r} \frac{p(x)}{q_1(x)} \quad \text{con} \quad p(x_0), q_1(x_0) \neq 0.$$

In particolare, $p(x)/q_1(x)$ ammette limite finito non nullo $\ell = p(x_0)/q_1(x_0)$ per x che tende a x_0 , per cui il comportamento di $f(x)$ per x vicino a x_0 sarà analogo a quello di $\ell/(x - x_0)^r$. In particolare, il limite di $f(x)$ per x che tende a x_0 (da sopra o da sotto) sarà uguale a quello di $\ell/(x - x_0)^r$, e quindi varrà $\pm\infty$ a seconda del segno di ℓ e della parità di r (e a seconda se x tende a x_0 da sopra o da sotto). In ogni caso, la retta $x = x_0$ è un asintoto verticale.

Osservazione 4.36 Nel ragionamento precedente abbiamo implicitamente usato alcune proprietà algebriche dei limiti. Per l'esattezza, le seguenti formule valgono *quasi* sempre:

$$\begin{aligned} \lim_{x \rightarrow x_0} [f(x) \pm g(x)] &= \lim_{x \rightarrow x_0} f(x) \pm \lim_{x \rightarrow x_0} g(x), \\ \lim_{x \rightarrow x_0} [f(x)g(x)] &= \lim_{x \rightarrow x_0} f(x) \cdot \lim_{x \rightarrow x_0} g(x), \\ \lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} &= \frac{\lim_{x \rightarrow x_0} f(x)}{\lim_{x \rightarrow x_0} g(x)} \quad \text{se } \lim_{x \rightarrow x_0} g(x) \neq 0. \end{aligned} \quad (4.22)$$

Il “quasi” si riferisce al fatto che possono sorgere dei problemi se qualcuno di questi limiti è infinito. In particolare, se a secondo membro otteniamo una delle seguenti

forme indeterminate

$$+\infty - \infty, \quad \pm\infty \cdot 0, \quad \frac{\pm\infty}{\pm\infty},$$

allora per scoprire quanto fa il limite a primo membro (ammesso che esista!) abbiamo bisogno di maggiori informazioni (e di metodi che vedremo nel prossimo capitolo). Altre apparizioni di limiti infiniti non creano grossi problemi, invece: se poniamo

$$\begin{aligned} +\infty + \infty &= +\infty, \quad -\infty - \infty = -\infty, \\ \ell \pm \infty &= \pm\infty \quad \text{e} \quad \frac{\ell}{\pm\infty} = 0 \quad \text{per } \ell \in \mathbb{R}, \\ \ell \cdot \pm\infty &= \frac{\pm\infty}{\ell} = \pm\infty \quad \text{per } \ell > 0, \quad \ell \cdot \pm\infty = \frac{\pm\infty}{\ell} = \mp\infty \quad \text{per } \ell < 0, \end{aligned}$$

allora le formule (4.22) rimangono valide (come pure rimangono valide se al posto di x_0 mettiamo $\pm\infty$). Infine, anche lo studio del limite di un quoziente quando il denominatore tende a zero richiede maggiori informazioni. Se il limite del numeratore è non nullo (o infinito), il limite del *valore assoluto* del quoziente è $+\infty$, ma il limite del quoziente potrebbe essere $+\infty$, $-\infty$ o non esistere affatto. Infine, se anche il limite del numeratore è nullo siamo in presenza della forma indeterminata $0/0$, che studieremo nel prossimo capitolo.

Infine, vediamo cosa possiamo dire sul comportamento di una funzione razionale f , scritta nella forma (4.20), quando x tende all'infinito. Abbiamo visto (nella Sezione 4.4) che per $|x|$ abbastanza grande, $p(x)$ si comporta come $a_m x^m$ e $q(x)$ si comporta come $b_n x^n$; quindi $f(x)$ si comporta come la funzione potenza

$$\frac{a_m x^m}{b_n x^n} = \frac{a_m}{b_n} x^{m-n}$$

quando $|x|$ è abbastanza grande.

CURIOSITÀ 4.10 Per l'esattezza, quanto visto nella Curiosità 4.5 ci dice che esiste un $R > 0$ sufficientemente grande tale che se $|x| > R$ e $a_m x^m, b_n x^n > 0$ allora $\frac{1}{2} a_m x^m \leq p(x) \leq \frac{3}{2} a_m x^m$ e $\frac{1}{2} b_n x^n \leq q(x) \leq \frac{3}{2} b_n x^n$, per cui

$$\frac{1}{3} \frac{a_m}{b_n} x^{m-n} \leq \frac{p(x)}{q(x)} = f(x) \leq 3 \frac{a_m}{b_n} x^{m-n}.$$

Stime analoghe si ottengono anche per gli altri possibili segni di $a_m x^m$ e $b_n x^n$.

Di conseguenza,

$$\lim_{x \rightarrow \pm\infty} f(x) = \lim_{x \rightarrow \pm\infty} \frac{a_m}{b_n} x^{m-n} = \begin{cases} 0 & \text{se } n > m, \\ \frac{a_m}{b_n} & \text{se } n = m, \\ \pm\infty & \text{se } n < m, \end{cases}$$

dove il segno nell'ultimo caso dipende dal segno di a_m/b_n , dalla parità di $m - n$, e da dove si sta calcolando il limite (se a $+\infty$ o a $-\infty$).

ESEMPIO 4.13 La legge che descrive il comportamento delle lenti convesse sottili è

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}, \quad (4.23)$$

dove u è la distanza fra l'oggetto e il centro della lente, v è la distanza fra l'immagine e il centro della lente, e f è la lunghezza focale della lente. Vogliamo studiare la dipendenza della distanza $s = u + v$ fra oggetto e immagine dalla distanza u fra l'oggetto e il centro di una lente di lunghezza focale $f = 10$ cm. Siccome $v = s - u$, otteniamo

$$\frac{1}{10} = \frac{1}{u} + \frac{1}{s-u} = \frac{s}{u(s-u)} \implies u(s-u) = 10s \implies s = \frac{u^2}{u-10}.$$

Quindi la dipendenza è data da una funzione razionale di grado 2. Notiamo prima di tutto che s ha una singolarità in $u = 10$: questo vuol dire che quando l'oggetto si avvicina al fuoco della lente allora l'immagine scappa all'infinito (in quanto $s(u)$ tende a $+\infty$ per u che tende a 10^+). Se $u < 10$ allora s diventa negativa, cosa fisicamente insensata; ma infatti la legge (4.23) vale solo per $u, v \geq f$.

Siccome il grado del numeratore è maggiore del grado del denominatore, e i coefficienti dei termini di grado massimo del numeratore e del denominatore sono positivi, abbiamo che s tende a $+\infty$ quando u tende all'infinito. Detta così è un'affermazione fisicamente ovvia: se l'oggetto scappa all'infinito, la distanza dall'immagine (che è dall'altro lato della lente) tende all'infinito. Possiamo renderla però più interessante notando che

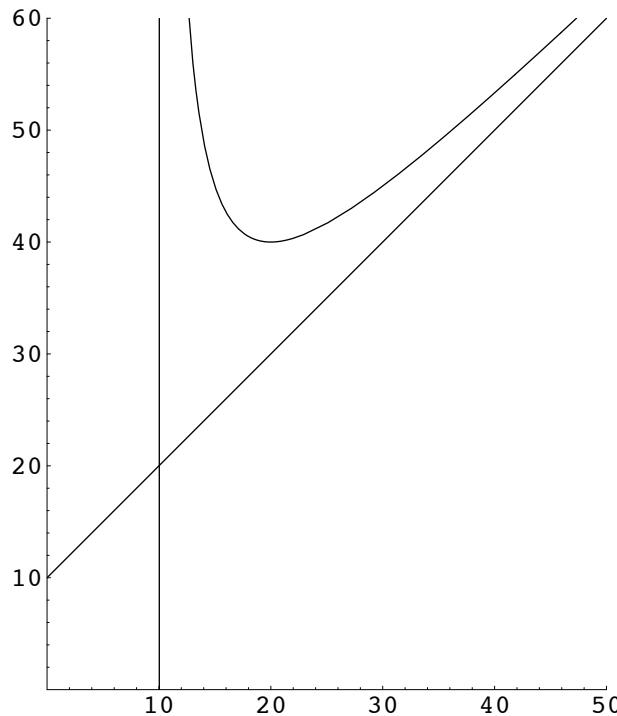
$$\frac{u^2}{u-10} = \frac{u(u-10) + 10u}{u-10} = u + \frac{10(u-10) + 100}{u-10} = u + 10 + \frac{100}{u-10}.$$

Siccome $100/(u-10)$ tende a zero quando u tende a $+\infty$, vediamo che s si comporta come $u + 10$ quando u diventa grande¹⁰. Ma $s = u + v$; quindi $v = s - u$ tende a 10 quando u tende all'infinito — che fisicamente vuol dire che quando l'oggetto scappa all'infinito l'immagine si avvicina quanto vogliamo al fuoco della lente.

Dunque s tende all'infinito sia quando u tende a $+\infty$ sia quando u tende a 10^+ . Ma allora la funzione s non può essere monotona nella semiretta $(10, +\infty)$; inoltre, siccome s scappa all'infinito in entrambi gli estremi di $(10, +\infty)$, sembra molto ragionevole supporre che s abbia almeno un punto di minimo all'interno di questa semiretta. Questo punto di minimo è chiaramente interessante, in quanto permette di trovare la distanza minima fra l'oggetto e l'immagine; vedremo nel prossimo capitolo come fare per (dimostrare che esiste e) trovarlo. La Figura 4.12 contiene il grafico della funzione s assieme all'asintoto verticale e all'asintoto obliquo¹¹.

¹⁰ E infatti il grafico di s si avvicina sempre più alla retta grafico di $u + 10$; si dice che questa retta è un *asintoto obliqua* per il grafico di s .

¹¹ Sì, il grafico di s è un'iperbole non equilatera.

Figura 4.12 $s(u) = u^2/(u - 10)$.

4.7 Funzioni esponenziali

Le funzioni esponenziali e logaritmiche (di cui parleremo nella prossima sezione) sono fra le più importanti e utili nella modellizzazione di fenomeni naturali, probabilmente persino più importanti delle funzioni polinomiali. Prima di presentare una definizione ufficiale, voglio discutere tre esempi in cui appaiono in maniera naturale funzioni di tipo esponenziale.

ESEMPIO 4.14 Supponiamo di avere una popolazione di cellule in ambiente ideale: calore e luce proprio adatti, e cibo in abbondanza per tutte. In queste condizioni, ogni cellula si riproduce sdoppiandosi dopo un tempo t_0 misurato in secondi (e che, per semplicità, supporremo uguale per ogni cellula). Se all'inizio la popolazione di cellule è composta da p_0 individui, da quanti individui sarà composta al tempo nt_0 ? L'idea è che ogni t_0 secondi tutte le cellule si riproducono, per cui la popolazione raddoppia. Quindi se al tempo 0 la popolazione è composta da $p(0) = p_0$ cellule, al tempo t_0 sarà composta da $p(t_0) = 2p_0$ cellule, al tempo $2t_0$ sarà composta da $p(2t_0) = 2(2p_0) = 4t_0$ cellule, e in generale al tempo nt_0 sarà composta da

$$p(nt_0) = 2^n p_0$$

cellule.

ESEMPIO 4.15 *Interesse semplice e interesse composto.* Questo esempio, anche se non di interesse strettamente biologico, è sicuramente interessante per i biologi, che come ogni altro lavoratore nella società contemporanea deve necessariamente contrattare con le banche. Supponi di avere a disposizione un capitale di c_0 euro, e di volerlo depositare in banca per ricavarne un qualche interesse. Il funzionario bancario, con un sorriso da un orecchio all'altro, ti propone il seguente schema, detto di *interesse semplice*, secondo lui molto vantaggioso: a intervalli regolari di tempo (per esempio, una volta all'anno) la banca ti concede un interesse pari al $p\%$ del capitale iniziale. Non devi fare nulla: basta che non tocchi i tuoi soldi in banca e questi crescono da soli. Se l'interesse viene attribuito a intervalli di tempo t_0 , e se al tempo 0 depositi il capitale c_0 , allora al tempo t_0 avrai un capitale uguale a $\text{IS}(t_0) = c_0 + \frac{p}{100}c_0$, al tempo $2t_0$ avrai un capitale uguale a $\text{IS}(2t_0) = \text{IS}(t_0) + \frac{p}{100}c_0 = c_0 + \frac{2p}{100}c_0$, e in generale al tempo nt_0 avrai un capitale uguale a

$$\text{IS}(nt_0) = c_0 + n \frac{p}{100} c_0 .$$

Tu guardi il funzionario di banca, e decidi che sorride troppo. Rifletti un attimo, e realizzi che in questo modo l'interesse viene calcolato sempre e soltanto sul capitale iniziale; il fatto che il capitale aumenti nel tempo non viene tenuto in considerazione. E questo non è giusto: se al tempo t_0 hai in banca un capitale pari a $\text{IS}(t_0)$, allora il nuovo interesse dev'essere calcolato su questa cifra, e non sul capitale iniziale! Il sorriso del funzionario di banca si restringe mentre ammette che effettivamente è previsto anche questo schema, detto di *interesse composto*. Con questo schema, al tempo t_0 hai un capitale $\text{IC}(t_0)$ uguale a quello ottenuto con l'interesse semplice, ma al tempo $2t_0$ hai un capitale $\text{IC}(2t_0) = (1 + \frac{p}{100})\text{IC}(t_0) = (1 + \frac{p}{100})^2 c_0$, e in generale al tempo nt_0 avrai un capitale pari a

$$\text{IC}(nt_0) = \left(1 + \frac{p}{100}\right)^n c_0 .$$

Ricordandoti la formula dello sviluppo del binomio ottieni

$$\left(1 + \frac{p}{100}\right)^n = 1 + n \frac{p}{100} + \binom{n}{2} \left(\frac{p}{100}\right)^2 + \dots > 1 + n \frac{p}{100}$$

non appena $n > 1$, per cui l'interesse composto ti fornisce più soldi dell'interesse semplice. Richiedi quindi che ti venga applicato l'interesse composto, e il funzionario, pur mugugnando, accetta.

Leggendo le clausole scritte in piccolo nel contratto, noti però un fatto curioso. Mentre l'interesse composto al tuo capitale viene attribuito a intervalli di t_0 , in caso tu andassi in rosso e chiedessi un prestito alla banca, la banca calcolerebbe l'interesse sul tuo prestito a intervalli di $t_0/2$. Insospettito, chiedi lumi al funzionario; lui (sorriso ridotto al lumicino) ti rassicura dicendo che sì, è vero, calcolano l'interesse a intervalli di $t_0/2$, ma ogni volta applicano un interesse dimezzato: solo di $p/2\%$. Subodorando qualcosa, provi a fare il conto: ponendo (per semplicità di notazione) $r = p/100$, partendo da un capitale t_0 il metodo della banca dopo un

tempo t_0 porta a un capitale pari a

$$\left(1 + \frac{r}{2}\right) \left(1 + \frac{r}{2}\right) c_0 = \left(1 + r + \frac{r^2}{4}\right) c_0 > (1+r)c_0 = \text{IC}(t_0) ;$$

il metodo adottato dalla banca fa aumentare il capitale (o il debito) più velocemente dell'interesse composto che ti avevano concesso! Minacci di portare il tuo capitale a un'altra banca; borbottando “Questi matematici rompiscatole...” e col sorriso completamente scomparso, il funzionario cede e accetta di usare lo stesso metodo di calcolo degli interessi sia per i crediti che per i debiti del tuo conto.

La storia non finisce qui. Il funzionario di banca, rimasto solo, riflette. Calcolando un interesse della metà il doppio delle volte, il capitale che si ottiene è maggiore. Che succede calcolando un interesse di un terzo il triplo delle volte? Un veloce conto mostra che il capitale che si ottiene dopo un tempo t_0 calcolando un interesse di $r/3$ dopo $t_0/3$ è

$$\left(1 + \frac{r}{3}\right)^3 c_0 = \left(1 + 3\frac{r}{3} + 3\frac{r^2}{9} + \frac{r^3}{27}\right) c_0 = \left(1 + r + \frac{r^2}{3} + \frac{r^3}{27}\right) c_0 > \left(1 + \frac{r}{2}\right)^2 c_0 ;$$

è più alto di quello che si otteneva applicando la metà dell'interesse il doppio delle volte! Al funzionario brillano gli occhi. Allora applicando un quarto dell'interesse il quadruplo delle volte sarà ancora più alto, e un quinto cinque volte sarà ancora più alto... e applicando un interesse minuscolo un numero sufficientemente alto di volte potremo ottenere un capitale sempre più alto, astronomico! Il funzionario si precipita dal suo capo, sicuro che una scoperta del genere gli assicurerà una promozione, se non addirittura un posto di dirigente! Il capo, che un poco di matematica dell'università ancora se la ricorda, ringrazia il funzionario ma lo rimanda al suo posto suggerendogli di iscriversi a Biologia e studiarsi un po' di matematica e statistica. Infatti, il funzionario ha ragione a dire che la procedura che ha proposto produce interessi crescenti, cioè che

$$1 + r < \left(1 + \frac{r}{2}\right)^2 < \left(1 + \frac{r}{3}\right)^3 < \cdots < \left(1 + \frac{r}{n}\right)^n < \left(1 + \frac{r}{n+1}\right)^{n+1} < \cdots ;$$

ma il funzionario ha torto a dire che in questo modo si possono ottenere interessi arbitrariamente grandi. Infatti, si può dimostrare (vedi la Curiosità 4.11) che esiste un numero reale, il *numero di Nepero*, usualmente indicato con e , che vale

$$e = 2.718281828459 \dots ,$$

tale che $(1+r/n)^n$ sia arbitrariamente vicino a e^r a patto di prendere n abbastanza grande:

$$\lim_{n \rightarrow +\infty} \left(1 + \frac{r}{n}\right)^n = e^r . \quad (4.24)$$

Quindi applicando lo schema proposto dal funzionario dopo un tempo t_0 si ottiene al massimo un capitale pari a $e^r c_0$, che è più di quanto si ottenesse prima, ma non

è arbitrariamente grande essendo strettamente minore di $3^r c_0$ (in quanto $e < 3$ e la funzione potenza è strettamente crescente). Per esempio, se $r = 1/100$ si ha

$$\text{IS}(t_0) = \text{IC}(t_0) = 1.01 c_0 < e^r c_0 \simeq 1.01005 c_0 .$$

Osservazione 4.37 Il limite in (4.24) è il limite di una *successione*, cioè di una sequenza di numeri $x_0, x_1, x_2, \dots, x_n, \dots$. Abbiamo già notato che una successione è semplicemente una funzione con dominio l'insieme \mathbb{N} dei numeri naturali; essendo $\mathbb{N} \subset \mathbb{R}$, la definizione di limite di una successione è analoga a quella di limite di una funzione per n che tende a $+\infty$. Diremo che una successione $\{x_1, x_2, \dots\}$ (o, come spesso scriveremo, $\{x_n\}$) ha *limite* $\ell \in \mathbb{R}$ se x_n è arbitrariamente vicino a ℓ a patto di prendere n sufficientemente grande, cioè se per ogni $\varepsilon > 0$ (arbitrariamente piccolo) esiste $N > 0$ (sufficientemente grande) tale che $n > N$ implica $|x_n - \ell| < \varepsilon$. Analogamente diremo che la successione $\{x_n\}$ ha *limite* $+\infty$ ($-\infty$) se per ogni $M > 0$ (arbitrariamente grande) esiste $N > 0$ (sufficientemente grande) tale che $n > N$ implica $x_n > M$ (rispettivamente, $x_n < -M$).

Osservazione 4.38 Nell'Esempio 4.15 abbiamo supposto $r > 0$. In realtà, il limite (4.24) vale anche per $r = 0$ (ovvio: $1^n = 1 = e^0$ quale che sia $n \in \mathbb{N}$, per cui la successione è costante) e per $r < 0$; quindi vale per ogni $r \in \mathbb{R}$.

Osservazione 4.39 In realtà, modulo sapere come si calcolano potenze con esponente irrazionale qualsiasi, il limite in (4.24) può essere visto come il limite di una funzione: infatti si ha

$$\forall r \in \mathbb{R} \quad \lim_{t \rightarrow +\infty} \left(1 + \frac{r}{t}\right)^t = e^r . \quad (4.25)$$

Ora, t tende a $+\infty$ se e solo se $x = 1/t$ tende a 0^+ ; quindi (4.25) si può scrivere anche come

$$\forall r \in \mathbb{R} \quad \lim_{x \rightarrow 0^+} (1 + rx)^{1/x} = e^r . \quad (4.26)$$

In particolare, prendendo $r = 1$ vediamo che $(1 + x)^{1/x}$ è molto vicino a e se $x > 0$ è sufficientemente piccolo; o, in altre parole, e^x è molto vicino a $1 + x$ se $x > 0$ è sufficientemente piccolo.

CURIOSITÀ 4.11 Vediamo come si dimostra (4.24) per $r > 0$. Poniamo

$$x_n(r) = \left(1 + \frac{r}{n}\right)^n ;$$

cominciamo dimostrando che $\{x_n(r)\}$ è una successione *strettamente crescente*, cioè tale che $x_n(r) < x_{n+1}(r)$ per ogni $n \in \mathbb{N}$. Usando lo sviluppo del binomio possiamo scrivere

$$\begin{aligned} x_n(r) &= \sum_{k=0}^n \binom{n}{k} \frac{r^k}{n^k} = \sum_{k=0}^n \frac{r^k}{k!} \frac{n(n-1)\cdots(n-k+1)}{n \cdot n \cdots n} \\ &= \sum_{k=0}^n \frac{r^k}{k!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) . \end{aligned}$$

Siccome

$$\forall j = 1, \dots, n-1 \quad 1 - \frac{j}{n+1} > 1 - \frac{j}{n},$$

otteniamo

$$\begin{aligned} x_{n+1}(r) &= \sum_{k=0}^{n+1} \frac{r^k}{k!} \left(1 - \frac{1}{n+1}\right) \cdots \left(1 - \frac{k-1}{n+1}\right) \\ &> \sum_{k=0}^n \frac{r^k}{k!} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) + \frac{r^{n+1}}{(n+1)!} \left(1 - \frac{1}{n+1}\right) \cdots \left(1 - \frac{n}{n+1}\right) \\ &> x_n(r). \end{aligned}$$

Quindi la successione $\{x_n(r)\}$ è strettamente crescente. Adesso dimostriamo che (contrariamente a quanto pensava il funzionario di banca) è una successione *limitata*. Diremo che un sottoinsieme $A \subset \mathbb{R}$ è *limitato superiormente* se esiste $M \in \mathbb{R}$ tale che $x \leq M$ per ogni $x \in A$; che è *limitato inferiormente* se esiste $M_1 \in \mathbb{R}$ tale che $M_1 \leq x$ per ogni $x \in A$; e che è *limitato* se è limitato sia superiormente sia inferiormente. Siccome $x_n(r) > x_1(r) = 1 + r$ per ogni $n \in \mathbb{N}$, la nostra successione è limitata inferiormente; facciamo vedere che è anche limitata superiormente. Sceglieremo un numero naturale n_0 maggiore di r . Notando che $1 - j/n < 1$ per ogni $n, j \in \mathbb{N}^*$ troviamo che per $n > n_0$ si ha

$$\begin{aligned} x_n(r) &< \sum_{k=0}^n \frac{r^k}{k!} = \sum_{k=0}^{n_0-1} \frac{r^k}{k!} + \sum_{k=n_0}^n \frac{r^k}{k!} = \sum_{k=0}^{n_0-1} \frac{r^k}{k!} + \frac{r^{n_0}}{n_0!} \sum_{k=n_0}^n \frac{r^{k-n_0}}{(n_0+1) \cdots k} \\ &< \sum_{k=0}^{n_0-1} \frac{r^k}{k!} + \frac{r^{n_0}}{n_0!} \sum_{k=n_0}^n \left(\frac{r}{n_0}\right)^{k-n_0} = \sum_{k=0}^{n_0-1} \frac{r^k}{k!} + \frac{r^{n_0}}{n_0!} \sum_{k=0}^{n-n_0} \left(\frac{r}{n_0}\right)^k. \end{aligned}$$

Usando la (4.17) con $x = r/n_0$ e $y = 1$ deduciamo che

$$x_n(r) < \sum_{k=0}^{n_0-1} \frac{r^k}{k!} + \frac{r^{n_0}}{n_0!} \frac{1 - (r/n_0)^{n-n_0+1}}{1 - r/n_0} < \sum_{k=0}^{n_0-1} \frac{r^k}{k!} + \frac{r^{n_0}}{n_0!} \frac{1}{1 - r/n_0} = M. \quad (4.27)$$

Abbiamo quindi ottenuto una stima $x_n(r) < M$ con M indipendente da n . Quindi la successione $\{x_n(r)\}$ continua a crescere senza mai poter superare la barriera imposta da M ; è naturale pensare che vada ad accumularsi da qualche parte. Ora, una delle proprietà più importanti dell'insieme dei numeri reali è l'esistenza degli estremi superiori. Sia $A \subset \mathbb{R}$ un sottoinsieme limitato superiormente; l'estremo superiore di A è il *minimo* numero $M \in \mathbb{R}$ tale che $x \leq M$ per ogni $x \in A$ ¹². I numeri reali sono costruiti in modo tale da assicurare l'esistenza degli estremi superiori (e inferiori; vedi la Nota 12) di insiemi limitati. Indichiamo con e_r l'estremo superiore della nostra successione $\{x_n(r)\}$; rimane da far vedere che e_r è il limite della nostra successione. Infatti, sceglieremo $\varepsilon > 0$ arbitrariamente piccolo. Sicuramente $e_r - x_n(r) \geq 0$ per ogni $n \in \mathbb{N}$, in quanto $e_r \geq x_n(r)$ per definizione. Ora, se avessimo

¹² In maniera analoga si definisce l'estremo inferiore di un sottoinsieme A limitato inferiormente come il massimo $M_1 \in \mathbb{R}$ tale che $M_1 \leq x$ per ogni $x \in A$. A volte si dice che un insieme A non limitato superiormente (inferiormente) ha estremo superiore (inferiore) $+\infty$ ($-\infty$).

$e_r - x_n(r) \geq \varepsilon$ per ogni $n \in \mathbb{N}$, avremmo $e_r - \varepsilon \geq x_n(r)$ per ogni n , e quindi $e_r - \varepsilon$ sarebbe un numero *minore* di e_r maggiore o uguale di tutti gli elementi della successione, per cui e_r non sarebbe l'estremo superiore. Siccome e_r è l'estremo superiore, questo non può succedere; quindi deve esistere $N \in \mathbb{N}$ tale che $e_r - x_N(r) < \varepsilon$, cioè $e_r - \varepsilon < x_N(r)$. Ma $\{x_n(r)\}$ è una successione crescente; quindi $e_r - \varepsilon < x_n(r)$ per ogni $n > N$, cioè (ricordando che $e_r - x_n(r)$ è sempre positivo) $|e_r - x_n(r)| < \varepsilon$ per ogni $n > N$, come voluto.

Abbiamo quindi dimostrato che la successione $\{x_n(r)\}$ ammette limite e_r . In particolare, la successione $\{x_n(1)\}$ ha limite e_1 ; per concludere, dobbiamo far vedere che $e_r = (e_1)^r$, almeno per $r = p/q$ razionale (che è la situazione in cui sappiamo esattamente cosa vuol dire elevare un numero alla r). Cominciamo notando che per ogni $p > 0$ naturale la successione $\{x_p(r), x_{2p}(r), \dots, x_{mp}(r), \dots\}$ è contenuta nella successione $\{x_n(r)\}$, per cui è crescente; inoltre siccome per ogni $n \in \mathbb{N}$ esiste $m \in \mathbb{N}$ tale che $mp > n$, la successione $\{x_{mp}(r)\}$ ha anche lo stesso estremo superiore di $\{x_n(r)\}$ — e quindi lo stesso limite. Ma allora

$$\begin{aligned} e_r &= \lim_{m \rightarrow +\infty} \left(1 + \frac{r}{mp}\right)^{mp} = \lim_{m \rightarrow +\infty} \left[\left(1 + \frac{r/p}{m}\right)^m\right]^p = \left[\lim_{m \rightarrow +\infty} \left(1 + \frac{r/p}{m}\right)^m\right]^p \\ &= (e_{r/p})^p, \end{aligned}$$

dove abbiamo usato il fatto che il limite del prodotto è uguale al prodotto dei limiti (Osservazione 4.36). In particolare, prendendo $r = 1$ otteniamo $e_1 = (e_{1/p})^p$, cioè $e_{1/p} = (e_1)^{1/p}$. Quindi se $r = p/q \in \mathbb{Q}^+$ è razionale positivo otteniamo

$$e_{p/q} = (e_{1/q})^p = (e_1)^{p/q},$$

come voluto. Il numero di Nepero e è quindi il nostro e_1 ; nota infine che la stima (4.27) con $r = 1$ e $n_0 = 2$ ci dà

$$e = e_1 < 1 + 1 + \frac{1}{2!} \cdot \frac{1}{1 - 1/2} = 3.$$

Lasciamo gli ambienti bancari e vediamo un altro esempio.

ESEMPIO 4.16 Decadimento radioattivo. Un atomo radioattivo, quando decade, libera particelle alfa, che possono indurre il decadimento di altri atomi. Quindi più atomi ci sono, più atomi decadono; il numero di atomi che decadono nell'unità di tempo è proporzionale al numero di atomi radioattivi presenti. Se indichiamo con $N(t)$ il numero di atomi radioattivi presenti al tempo t in un campione di isotopo radioattivo, esiste quindi una costante $\lambda > 0$, detta *costante di decadimento*, tale che

$$\Delta N = N(t_1) - N(t_0) = -\lambda(t_1 - t_0)N(t_0) = -\lambda\Delta t N(t_0),$$

o anche

$$\frac{\Delta N}{\Delta t} = -\lambda N,$$

almeno per Δt piccolo (e il segno meno è dovuto al fatto che il numero di atomi radioattivi diminuisce col passare del tempo). Vediamo come recuperare la forma della funzione N . Supponiamo che il nostro campione contenga N_0 atomi radioattivi al tempo $t_0 = 0$. Allora al tempo t_1 (piccolo) ne conterrà

$$N(t_1) = N_0 - \lambda t_0 N_0 = (1 - \lambda t_1)N_0.$$

Analogamente, al tempo $2t_1$ ne conterrà $N(2t_1) = (1 - \lambda t_1)N(t_1) = (1 - \lambda t_1)^2 N_0$, e in generale

$$N(kt_1) = (1 - \lambda t_1)^k N_0 .$$

Ponendo $t = kt_1$ questa formula diventa

$$N(t) = \left(1 - \frac{\lambda t}{k}\right)^k N_0 .$$

Ora, questa formula vale solo per t multiplo di t_1 . Inoltre, di fatto abbiamo supposto che tutti i decadimenti avvengano solo agli istanti multipli di t_1 ; in realtà, avvengono in continuazione. Quindi per calcolare $N(t)$ per $t > 0$ fissato ma qualsiasi (e per modellizzare meglio il fenomeno reale) dobbiamo rendere t_1 arbitrariamente piccolo, facendolo tendere a 0^+ . Siccome $k = t/t_1$, far tendere t_1 a zero equivale a far tendere k a $+\infty$, per cui otteniamo il risultato cercato:

$$N(t) = \lim_{k \rightarrow +\infty} \left(1 - \frac{\lambda t}{k}\right)^k N_0 = e^{-\lambda t} N_0 .$$

Questa è la formula del *decadimento esponenziale* di costante λ . A volte, per descrivere il decadimento radioattivo di un dato isotopo, invece della costante λ si usa il *tempo di dimezzamento* (o *emivita*) $t_{1/2}$, definito come il tempo necessario perché il numero di atomi radioattivi si dimezzi. In altre parole, $t_{1/2}$ deve soddisfare la seguente condizione:

$$e^{-\lambda t_{1/2}} = \frac{1}{2} .$$

Nella prossima sezione (Esempio 4.19) vedremo come calcolare $t_{1/2}$ partendo da λ .

Siamo ora pronti per definire cos'è una *funzione esponenziale*: è una funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ della forma

$$f(x) = aq^x ,$$

per opportuni $q > 0$ (detto *base* della funzione esponenziale) e $a \neq 0$.

Osservazione 4.40 Se $q = 1$ abbiamo $f(x) = a$ per ogni $x \in \mathbb{R}$, cioè ogni funzione esponenziale di base 1 è costante. Per questo motivo la base 1 non si usa praticamente mai.

Osservazione 4.41 Le proprietà delle potenze ci dicono che

$$aq^{-x} = a \left(\frac{1}{q}\right)^x \quad \text{e} \quad aq^{cx} = a(q^c)^x ;$$

quindi ogni funzione della forma $f(x) = aq^{cx}$ con $q > 0$ e $a, c \neq 0$ si può scrivere come funzione esponenziale, a patto di scegliere una base opportuna.

Osservazione 4.42 La funzione esponenziale per antonomasia è quella di base e e con coefficiente 1. È la più importante, tanto da avere meritato un simbolo specifico:

$$\exp(x) = e^x.$$

Nel prossimo capitolo vedremo perché questa funzione esponenziale è particolarmente amata dai matematici; nella prossima sezione vedremo inoltre che ogni funzione esponenziale può essere scritta nella forma $f(x) = a \exp(cx)$ per opportuni $a, c \neq 0$; vedi (4.33).

Osservazione 4.43 Attenzione a non confondere le funzioni esponenziali con le funzioni potenza. Nelle funzioni esponenziali, la base è fissata e la variabile è a esponente; nelle funzioni potenza, è l'esponente a essere fissato, mentre la variabile è la base.

Vediamo di studiare le funzioni esponenziali seguendo lo schema delle sezioni precedenti. Prima di tutto, la generica funzione esponenziale dipende da due parametri (q e a); quindi ci aspettiamo che per determinarla bastino due punti del suo grafico. Infatti, se $f(x) = aq^x$ e $x_0, x_1 \in \mathbb{R}$ abbiamo

$$f(x_1) = aq^{x_1} = aq^{x_0} \cdot q^{x_1 - x_0} = q^{x_1 - x_0} f(x_0);$$

quindi se (x_0, y_0) e (x_1, y_1) sono due punti con ascisse distinte del grafico di f , cioè $y_0 = f(x_0)$ e $y_1 = f(x_1)$ con $x_0 \neq x_1$, otteniamo

$$q = \left(\frac{y_1}{y_0} \right)^{1/(x_1 - x_0)} \quad \text{e} \quad a = \frac{y_0}{q^{x_0}} = \left(\frac{y_0^{x_1}}{y_1^{x_0}} \right)^{1/(x_1 - x_0)}.$$

Un altro modo per trovare a è notare che $a = f(0)$.

Ricaviamo ora le caratteristiche del grafico di $f(x) = aq^x$. Prima di tutto,

- (a) una funzione esponenziale non si annulla mai. Più precisamente, $f(x) > 0$ per ogni $x \in \mathbb{R}$ se $a > 0$, e $f(x) < 0$ per ogni $x \in \mathbb{R}$ se $a < 0$.

Poi, un numero maggiore di 1 diventa sempre più grande se viene elevato a esponenti più grandi, cioè $x_0 < x_1$ implica $q^{x_0} < q^{x_1}$. Invece, un numero positivo minore di uno diventa sempre più piccolo se elevato a esponenti più grandi; infatti, se $0 < q < 1$ allora $1/q > 1$, per cui $x_0 < x_1$ implica

$$1/q^{x_0} = (1/q)^{x_0} < (1/q)^{x_1} = 1/q^{x_1},$$

e quindi $q^{x_0} > q^{x_1}$. Tenendo presente anche il segno di a otteniamo le seguenti informazioni sulla monotonia delle funzioni esponenziali:

- (b) se $f(x) = aq^x$ allora f è strettamente crescente se $a > 0$ e $q > 1$, oppure se $a < 0$ e $0 < q < 1$; ed è strettamente decrescente se $a > 0$ e $0 < q < 1$, oppure se $a < 0$ e $q > 1$.

Nell'Esempio 4.15 abbiamo visto che la successione $(1+x/k)^k$ tende crescendo a e^x quando $x > 0$; in particolare, prendendo $k = 1$ otteniamo

$$\forall x > 0 \quad e^x > 1 + x .$$

In particolare, siccome $1 + x$ tende a $+\infty$ per x che tende a $+\infty$, anche e^x , essendo più grande, non può fare diversamente:

$$\lim_{x \rightarrow +\infty} e^x = +\infty .$$

Siccome $e^{-x} = 1/e^x$ ponendo $y = -x$ otteniamo

$$\lim_{x \rightarrow -\infty} e^x = \lim_{y \rightarrow +\infty} e^{-y} = \lim_{y \rightarrow +\infty} \frac{1}{e^y} = \frac{1}{\lim_{y \rightarrow +\infty} e^y} = 0 ,$$

dove abbiamo usato le proprietà dei limiti descritte nell'Osservazione 4.36.

Osservazione 4.44 Abbiamo usato anche un'altra proprietà importante dei limiti: se $f(x) \leq g(x)$ per tutti gli x abbastanza vicini a $x_0 \in \mathbb{R}$, allora

$$\lim_{x \rightarrow x_0} f(x) \leq \lim_{x \rightarrow x_0} g(x) ;$$

funzioni più grandi hanno limiti più grandi. Questo vale anche per $x_0 = \pm\infty$, dove per “abbastanza vicino” a $+\infty$ (rispettivamente, a $-\infty$) intendiamo x abbastanza grande (rispettivamente, abbastanza negativo), cioè $x > M$ (rispettivamente, $x < -M$) per qualche $M > 0$. In particolare, abbiamo il risultato noto come Teorema dei due carabinieri: se $f(x) \leq g(x) \leq h(x)$ per ogni x vicino a x_0 , e $\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} h(x) = L$ (che può essere finito o infinito), allora anche $\lim_{x \rightarrow x_0} g(x) = L$. Infine, un'avvertenza: la diseguaglianza stretta delle funzioni $f(x) < g(x)$ non implica la diseguaglianza stretta dei limiti. Per esempio, $1/x < 2/x$ per ogni $x > 0$ ma

$$\lim_{x \rightarrow +\infty} \frac{1}{x} = 0 = \lim_{x \rightarrow +\infty} \frac{2}{x} .$$

Come detto nell'Osservazione 4.42 e come dimostreremo nella prossima sezione, ogni funzione esponenziale si può scrivere nella forma $a \exp(cx)$, con $c > 0$ se la base è maggiore di 1 e $c < 0$ se la base è minore di 1. Quindi tenendo presente il segno di a otteniamo

(c) se $f(x) = aq^x$ allora

$$\lim_{x \rightarrow +\infty} f(x) = \begin{cases} +\infty & \text{se } a > 0 \text{ e } q > 1 , \\ -\infty & \text{se } a < 0 \text{ e } q > 1 , \\ 0 & \text{se } 0 < q < 1 ; \end{cases}$$

e

$$\lim_{x \rightarrow -\infty} f(x) = \begin{cases} 0 & \text{se } q > 1, \\ +\infty & \text{se } a > 0 \text{ e } 0 < q < 1, \\ -\infty & \text{se } a < 0 \text{ e } 0 < q < 1. \end{cases}$$

Dunque se $a > 0$ la funzione $f(x) = aq^x$ può assumere sia valori arbitrariamente grandi sia valori arbitrariamente piccoli positivi, mentre se $a < 0$ può assumere sia valori arbitrariamente negativi sia valori arbitrariamente piccoli negativi; e per passare dagli uni agli altri deve assumere anche tutti i valori intermedi (vedi la Curiosità 4.2). Quindi

(d) l'immagine di $f(x) = aq^x$ è tutto \mathbb{R}^+ se $a > 0$, o tutto \mathbb{R}^- se $a < 0$.

Osservazione 4.45 Una funzione strettamente monotona è sempre iniettiva: se $x_0 \neq x_1$ si deve avere $f(x_0) < f(x_1)$ o $f(x_0) > f(x_1)$, e non può mai succedere che $f(x_0) = f(x_1)$. Quindi una funzione esponenziale $f(x) = aq^x$ è una funzione bigettiva dal suo dominio \mathbb{R} alla sua immagine \mathbb{R}^+ (o \mathbb{R}^- a seconda del segno di a); in particolare, è una funzione *invertibile*. Come vedrai, questo fatto sarà essenziale nella prossima sezione.

La Figura 4.13 contiene il grafico di un paio di funzioni esponenziali.

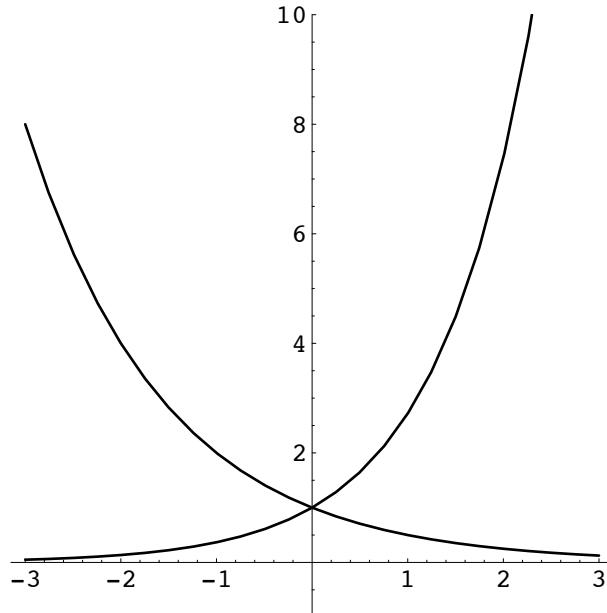


Figura 4.13 Grafici di e^x e $2^{-x} = (1/2)^x$.

Le funzioni esponenziali possono venire usate per costruire altre funzioni con comportamenti interessanti.

ESEMPIO 4.17 I neuroni comunicano gli uni con gli altri scambiando cariche elettriche. La carica elettrica si accumula su un dato neurone fino al momento in cui il potenziale supera una certa soglia (dipendente dalle condizioni fisico-chimiche del neurone, dei neuroni a cui è collegato, e delle altre cellule circostanti), quando scarica su un altro neurone. Sperimentalmente si è trovato che prima della scarica il potenziale elettrico $V(t)$ (misurato in milliVolt) accumulato da un neurone al tempo t segue un andamento della forma

$$V(t) = VQ_0(1 - e^{-kt}),$$

per opportuni $k, V_0 > 0$, dove l'origine dei tempi è fissata a un istante in cui il neurone è completamente scarico, cioè $V(0) = 0$. Vogliamo tracciare il grafico di $V(t)$, e capire il significato di V_0 . La funzione $V(t)$ è ottenuta a partire da una funzione esponenziale di base e^{-k} con una serie di operazioni; seguiamole e vediamo cosa succede al grafico. Siccome $k > 0$, la base e^{-k} è minore di 1, per cui la funzione esponenziale $f_1(t) = e^{-kt}$ è positiva, strettamente decrescente, con limite $+\infty$ per t che tende a $-\infty$, e con limite 0 per t che tende a $+\infty$. La funzione $f_2(t) = -e^{-kt}$ è ottenuta da f_1 moltiplicando per -1 ; quindi il grafico di f_2 è il simmetrico rispetto all'asse delle ascisse del grafico di f_1 . Quindi f_2 è negativa, strettamente crescente, con limite $-\infty$ per t che tende a $-\infty$, e con limite 0 per t che tende a $+\infty$. La funzione $f_3(t) = 1 - e^{-kt}$ è ottenuta da f_2 sommando 1, operazione che abbiamo visto corrispondere a traslare verso l'alto di 1 unità il grafico di f_2 . In particolare, $f_3(t) < 0$ se e solo se $f_2(t) < -1$ se e solo se $f_1(t) > 1$, e quindi se e solo se $t < 0$ (in quanto $f_1(0) = 1$ e f_1 è strettamente decrescente). Quindi f_3 è negativa sulla semiretta $(-\infty, 0)$, si annulla in 0, è positiva sulla semiretta $(0, +\infty)$, è strettamente crescente, ha limite $-\infty$ per t che tende a $-\infty$, e ha limite 1 per t che tende a $+\infty$.

Infine, $V(t) = V_0 f_3(t)$ è ottenuta da f_3 moltiplicando le ordinate per V_0 . Di conseguenza, V è negativa sulla semiretta $(-\infty, 0)$, si annulla in 0, è positiva sulla semiretta $(0, +\infty)$, è strettamente crescente, ha limite $-\infty$ per t che tende a $-\infty$, e ha limite V_0 per t che tende a $+\infty$. In particolare, il potenziale accumulato dal neurone (in assenza di scariche) aumenta col tempo tendendo ad appiattirsi verso il valore limite V_0 , che rappresenta quindi il potenziale massimo che può accumularsi sul neurone lasciato evolvere per conto suo senza influenze esterne. La Figura 4.14 contiene il grafico di $V(t)$ con $V_0 = k = 2$.

Le funzioni $f: \mathbb{R} \rightarrow \mathbb{R}$ della forma

$$f(x) = a \left(1 - e^{-k(x-x_0)} \right) + b \quad (4.28)$$

con $a, k > 0$ e $x_0, b \in \mathbb{R}$ sono utili ogni qualvolta si voglia rappresentare una quantità che assume un valore specificato b in un punto specificato x_0 , inizia a crescere in maniera quasi lineare per $x > x_0$, e poi (per l'intervento di fenomeni di saturazione) tende ad appiattirsi verso il valore limite $a + b$.

Osservazione 4.46 La frase “cresce in maniera quasi lineare” in un intervallo vuol dire che esiste una funzione lineare $g: \mathbb{R} \rightarrow \mathbb{R}$ tale che $|f(x) - g(x)|$ sia piccolo quando

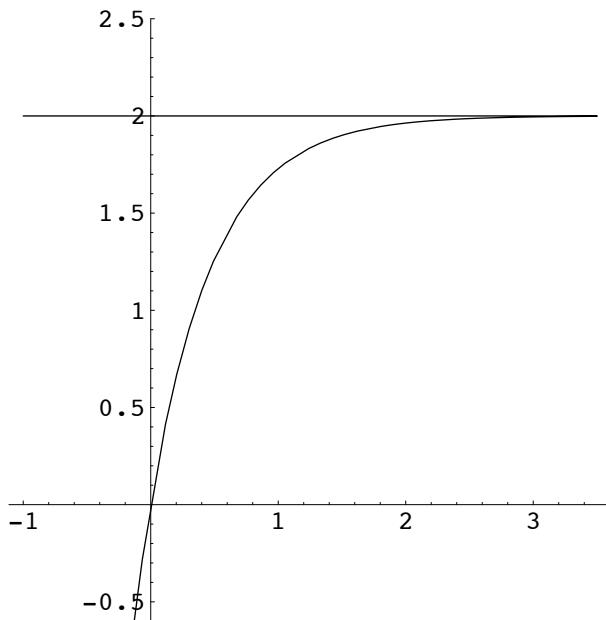


Figura 4.14 Grafico di $V(t) = 2(1 - e^{-2t})$.

x appartiene a quell'intervallo. Per esempio, nel prossimo capitolo vedremo che se $f(x) = a(1 - e^{-k(x-x_0)}) + b$ e $g(x) = b + ka(x - x_0)$ allora

$$\forall x > x_0 \quad 0 \leq |f(x) - g(x)| \leq a \frac{k^2(x - x_0)^2}{2}. \quad (4.29)$$

Ora, se indichiamo con $x_1 > x_0$ l'unico punto tale che $f(x_1) = b + a/3$, allora un conto che vedremo nella prossima sezione (Esempio 4.20) ci dirà che

$$\forall x \in [x_0, x_1] \quad |f(x) - g(x)| < 0.09a, \quad (4.30)$$

per cui nell'intervallo $[x_0, x_1]$ la differenza fra f e una funzione lineare non supera un decimo della variazione totale di f .

Le funzioni della forma (4.28) sono di solito usate solo per seguire lo sviluppo di un fenomeno che presenta saturazione (cioè appiattimento verso un valore limite) partendo da un punto preciso x_0 con un valore preciso $f(x_0) = b$, e proseguendo con $x \geq x_0$. Alcune situazioni, invece, non hanno un punto di partenza preciso individuato a priori; sappiamo solo che la quantità $f(x)$ che vogliamo studiare può variare da un valore limite minimo a_- a un valore limite massimo a_+ , con fenomeni di saturazione per x grande e di saturazione inversa (cioè partenza piatta da un valore limite) per x molto negativo. Le funzioni esponenziali sono utili anche in questo caso.

Dati $a_- < a_+$, vogliamo trovare una funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ strettamente crescente e tale che

$$\lim_{x \rightarrow -\infty} f(x) = a_- , \quad \lim_{x \rightarrow +\infty} f(x) = a_+ .$$

Prima di tutto, traslando verticalmente di a_- il grafico vediamo che basta trovare una funzione $f_1: \mathbb{R} \rightarrow \mathbb{R}$ strettamente crescente tale che

$$\lim_{x \rightarrow -\infty} f_1(x) = 0 , \quad \lim_{x \rightarrow +\infty} f_1(x) = a > 0 ,$$

dove $a = a_+ - a_-$; infatti, se abbiamo f_1 allora $f(x) = f_1(x) + a_-$ è come voluto. Poi, ponendo $f_1(x) = af_2(x)$ vediamo che basta trovare una funzione $f_2: \mathbb{R} \rightarrow \mathbb{R}^+$ strettamente crescente tale che

$$\lim_{x \rightarrow -\infty} f_2(x) = 0 , \quad \lim_{x \rightarrow +\infty} f_2(x) = 1 .$$

Dovendo essere strettamente crescente, f_2 è sempre positiva; quindi la funzione $f_3(x) = 1/f_2(x)$ è sempre positiva, è strettamente decrescente, ed è tale che

$$\lim_{x \rightarrow -\infty} f_3(x) = +\infty , \quad \lim_{x \rightarrow +\infty} f_3(x) = 1 .$$

Allora la funzione $f_4(x) = f_3(x) - 1$ è anch'essa positiva, strettamente decrescente, e soddisfa

$$\lim_{x \rightarrow -\infty} f_4(x) = +\infty , \quad \lim_{x \rightarrow +\infty} f_4(x) = 0 .$$

Ora, noi conosciamo funzioni con questo comportamento; per esempio,

$$f_4(x) = e^{-k(x-x_0)} ,$$

con $k > 0$ e $x_0 \in \mathbb{R}$. Risalendo otteniamo

$$\begin{aligned} f(x) &= f_1(x) + a_- = (a_+ - a_-)f_2(x) + a_- = \frac{a_+ - a_-}{f_3(x)} + a_- = \frac{a_+ - a_-}{1 + f_4(x)} + a_- \\ &= \frac{a_+ - a_-}{1 + e^{-k(x-x_0)}} + a_- . \end{aligned}$$

Le funzioni della forma

$$f(x) = \frac{a}{1 + e^{-k(x-x_0)}} + b , \tag{4.31}$$

con $a, k > 0$ e $x_0, b \in \mathbb{R}$, sono talvolta dette *funzioni logistiche*, e hanno esattamente l'andamento cercato: sono positive, strettamente crescenti, hanno come immagine l'intervallo $(b, a+b)$, e

$$\lim_{x \rightarrow -\infty} f(x) = b , \quad \lim_{x \rightarrow +\infty} f(x) = a + b .$$

Inoltre, $f(x_0) = b + a/2$, cioè x_0 è il punto in cui f raggiunge il punto di mezzo della sua immagine.

Osservazione 4.47 Le funzioni logistiche hanno un andamento quasi lineare vicino al punto x_0 . Infatti, tecniche che vedremo nel prossimo capitolo mostrano che

$$\forall x \in [x_{-1}, x_1] \quad \left| f(x) - \left(b + \frac{a}{2} + \frac{ak}{4}(x - x_0) \right) \right| < 0.09a,$$

dove $x_{\pm 1} \in \mathbb{R}$ sono gli unici punti per cui $f(x_{\pm 1}) = b + (a/2) \pm (a/4)$. Quindi nell'intervallo $[x_{-1}, x_1]$ la funzione logistica f differisce da una funzione lineare meno di un decimo della variazione totale a .

ESEMPIO 4.18 Proviamo a usare funzioni logistiche per studiare la dipendenza fra la temperatura e la percentuale di germinazione dei soliti semi di pomodoro. Siamo in effetti proprio in una situazione che sembra fatta apposta per questo tipo di approccio: abbiamo due valori limite, uno superiore (100) e uno inferiore (0); è molto ragionevole assumere che la percentuale $P(T)$ cresca con la temperatura; e che ci siano fenomeni di saturazione e di saturazione inversa. Siccome il limite inferiore è $b = 0$ e il limite inferiore è $a + b = 100$, vogliamo usare una funzione del tipo

$$P(T) = \frac{100}{1 + e^{-k(T-T_0)}}.$$

Dobbiamo determinare k e T_0 . Nella Sezione 4.9 vedremo come usare il metodo dei minimi quadrati per farlo al meglio; ma possiamo ottenere dei buoni risultati anche procedendo un po' a occhio.

Prima di tutto, T_0 dev'essere scelto in modo tale che $P(T_0)$ valga $b + a/2 = 50$. Siccome le misure sperimentali danno $P(12) = 40$ e $P(15) = 70$, dobbiamo avere $12 < T_0 < 15$. Ora, $(a/2) - (a/4) = 25$ e $(a/2) + (a/4) = 75$; quindi l'intervallo $[12, 15]$ è contenuto nell'intervallo $[x_{-1}, x_1]$ in cui la funzione logistica è quasi lineare (Osservazione 4.47). La interpolazione lineare che abbiamo visto nell'Esempio 4.2 vale 50 in 13; quindi una possibile stima per T_0 è $T_0 = 13$.

Rimane da trovare k o, equivalentemente, e^{-k} ; infatti, se poniamo $q = e^{-k}$ allora possiamo riscrivere f nella forma

$$P(T) = \frac{100}{1 + q^{T-13}}.$$

Siccome conosciamo $P(12) = 40$, e $12 - 13 = -1$, viene naturale imporre la condizione

$$40 = P(12) = \frac{100}{1 + q^{-1}} \quad \Rightarrow \quad q = \frac{2}{3}.$$

Quindi otteniamo

$$P(T) = \frac{100}{1 + (2/3)^{T-13}};$$

la Figura 4.15 ne riporta il grafico, dove si vede che non solo abbiamo una buona approssimazione dei dati sperimentali, ma anche un andamento che sembra compatibile con estrazioni credibili. Nell'Esempio 4.27 vedremo che il metodo dei

minimi quadrati fornirà la funzione

$$P(T) = \frac{100}{1 + (0.702)^{T-12.94}} ,$$

che non è troppo distante da quella ottenuta qui a mano.

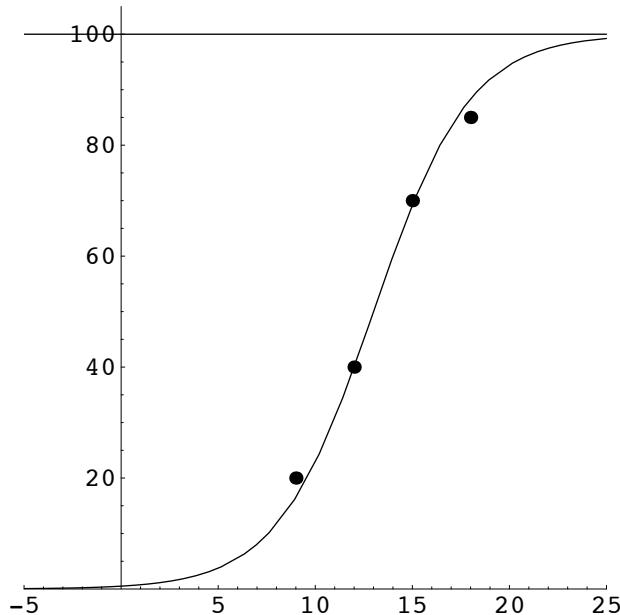


Figura 4.15 Una funzione logistica.

Infine, un'altra funzione importante che si può ottenere usando gli esponenziali è la *funzione Gaussiana*

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) ,$$

di cui puoi vedere il grafico nella Figura 4.16. Usando le tecniche viste in questa sezione non dovrebbe esserti difficile dimostrare (esercizio) che f è sempre positiva, è strettamente crescente in $(-\infty, 0]$, è strettamente decrescente in $[0, +\infty)$, ha un punto di massimo in $x = 0$, e ha limite 0 per x che tende a $\pm\infty$.

Osservazione 4.48 Vedremo nel Capitolo 6 che la distribuzione normale a cui abbiamo accennato nel capitolo precedente si esprime proprio usando la funzione Gaussiana.

CURIOSITÀ 4.12 Un'altra funzione che i matematici amano particolarmente è

$$f(x) = \begin{cases} e^{-1/x^2} & \text{se } x \geq 0 , \\ 0 & \text{se } x \leq 0 , \end{cases}$$

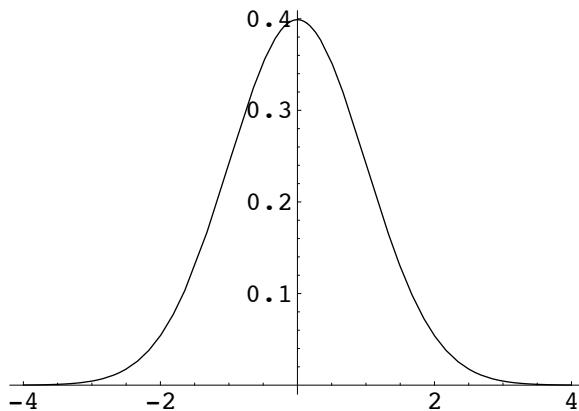


Figura 4.16 La funzione Gaussiana.

il cui grafico è riportato nella Figura 4.17. La particolarità di questa funzione è di essere (in un senso che discuteremo nel prossimo capitolo) infinitamente piatta nell'origine.

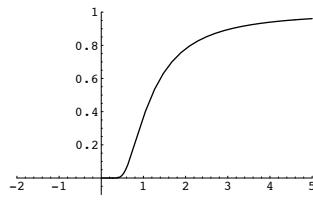


Figura 4.17 $f(x) = e^{-1/x^2}$.

4.8 Funzioni logaritmiche

Nell'Osservazione 4.45 abbiamo notato che la funzione $f(x) = q^x$ è *invertibile* come funzione da \mathbb{R} a \mathbb{R}^+ , non appena $q > 0$, $q \neq 1$. La funzione inversa si chiama *logaritmo in base $q \neq 1$* , e si indica con $\log_q: \mathbb{R}^+ \rightarrow \mathbb{R}$. In particolare, i logaritmi sono definiti solo sui numeri (strettamente) positivi.

Essendo la funzione inversa della funzione q^x , il logaritmo in base q soddisfa le identità seguenti:

$$\log_q(q^x) = x, \quad q^{\log_q x} = x.$$

In particolare,

$$\log_q 1 = \log_q(q^0) = 0, \quad \text{e} \quad \log_q q = \log_q(q^1) = 1.$$

Siccome il grafico dell'inversa di un funzione f si ottiene scambiando ascisse e ordinate, cioè prendendo il simmetrico del grafico di f rispetto alla retta $y = x$, dalle proprietà delle funzioni esponenziali deduciamo le seguenti proprietà delle funzioni logaritmiche:

- se $q > 1$ allora \log_q è strettamente crescente, negativa in $(0, 1)$, positiva in $(1, +\infty)$, e

$$\lim_{x \rightarrow +\infty} \log_q x = +\infty, \quad \lim_{x \rightarrow 0^+} \log_q x = -\infty;$$

se invece $q < 1$ allora \log_q è strettamente decrescente, positiva in $(0, 1)$, negativa in $(1, +\infty)$, e

$$\lim_{x \rightarrow +\infty} \log_q x = -\infty, \quad \lim_{x \rightarrow 0^+} \log_q x = +\infty.$$

La Figura 4.18 contiene i grafici di $\log_q x$ per $q > 1$ e per $q < 1$.

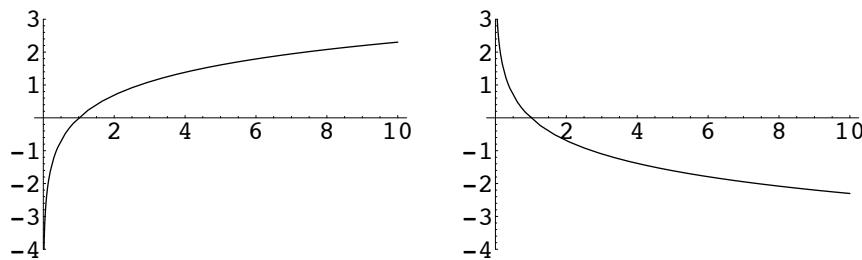


Figura 4.18 Logaritmi.

Le funzioni logaritmo godono di una proprietà estremamente importante. Applicando \log_q all'identità $q^x q^y = q^{x+y}$ otteniamo

$$\log_q(q^x q^y) = x + y.$$

Ora, se poniamo $a = q^x$ e $b = q^y$, ricaviamo $x = \log_q a$ e $y = \log_q b$, per cui

$$\forall a, b > 0 \quad \log_q(a \cdot b) = \log_q a + \log_q b; \quad (4.32)$$

il logaritmo del prodotto è uguale alla somma dei logaritmi. In particolare, se $a = b$ otteniamo $\log_q(a^2) = 2\log_q a$; con $b = a^2$ otteniamo $\log_q(a^3) = 3\log_q a$, e, più in generale,

$$\forall a > 0 \forall n \in \mathbb{N} \quad \log_q(a^n) = n \log_q a.$$

Inoltre

$$0 = \log_q 1 = \log_q \left(a \cdot \frac{1}{a} \right) = \log_q a + \log_q \frac{1}{a},$$

per cui

$$\forall a > 0 \quad \log_q \frac{1}{a} = -\log_q a.$$

Infine,

$$\forall a, b > 0 \quad \log_q \frac{a}{b} = \log_q \left(a \cdot \frac{1}{b} \right) = \log_q a - \log_q b .$$

Nell'Osservazione 4.42 avevamo detto che ogni funzione esponenziale può essere scritta nella forma $a e^{cx}$ per opportuni $a, c \in \mathbb{R}$. Adesso possiamo verificarlo, usando anche basi diverse da e . Dati $p, q > 0$, per definizione di logaritmo abbiamo $p = q^{\log_q p}$; quindi

$$\forall a \in \mathbb{R} \quad \forall p, q > 0 \quad ap^x = aq^{(\log_q p)x} . \quad (4.33)$$

In particolare, applicando \log_q a entrambi i membri (con $a = 1$) otteniamo

$$\forall x \in \mathbb{R} \quad \forall p, q > 0 \quad \log_q(p^x) = x \log_q p .$$

La formula (4.33), sempre con $a = 1$, ci permette anche di trovare come cambia il logaritmo cambiando base. Infatti, mettendo $\log_q x / \log_q p$ al posto di x otteniamo

$$p^{\log_q x / \log_q p} = q^{\log_q x} = x ,$$

e applicando \log_p a entrambi i membri deduciamo l'importante formula

$$\log_p x = \frac{\log_q x}{\log_q p} . \quad (4.34)$$

In particolare, ponendo $x = q$ ricaviamo

$$\log_p q = \frac{1}{\log_q p} .$$

Inoltre, prendendo $p = 1/q$ otteniamo

$$\log_{1/q} x = \frac{\log_q x}{\log_q(1/q)} = -\log_q x ,$$

per cui possiamo limitarci a lavorare con basi $q > 1$.

Osservazione 4.49 La formula (4.34) ci dice che tutti i logaritmi sono multipli della stessa funzione; quindi non c'è bisogno di lavorare con logaritmi in base qualsiasi ma è sufficiente fissare una base e lavorare solo con quella. La base più comune¹³ in ambito matematico è il numero di Nepero e ; il logaritmo in base e è detto *logaritmo naturale*, e viene indicato con \log (con la base sottintesa) o con \ln . In ambito

¹³ Per motivi che vedremo nel prossimo capitolo.

applicativo¹⁴ la base più comune è invece 10. Il logaritmo in base 10 è talvolta detto *logaritmo decimale*, e si indica con Log, o con log quando non si corre il rischio di confonderlo con il logaritmo naturale. Infine, in ambito informatico talvolta si usa anche il logaritmo in base 2, indicato (senza molta fantasia) con \log_2 .

Osservazione 4.50 Il logaritmo decimale di un numero positivo è strettamente legato alla sua rappresentazione decimale. Per esempio, abbiamo $n \leq \text{Log } x < n+1$ se e solo se $10^n \leq x < 10^{n+1}$, per cui la parte intera di $\text{Log } x$ è il numero di cifre (prima della virgola) nella rappresentazione decimale di x . Inoltre, Log funziona bene in combinazione con la notazione scientifica. Infatti, se scriviamo $x = a \cdot 10^b$ con $1 \leq a < 10$ e $b \in \mathbb{N}$, allora $\text{Log } x = b + \text{Log } a$, con $0 \leq \text{Log } a < 1$. Si dice che $\text{Log } a$ è la *mantissa* di $\text{Log } x$, mentre b è la *caratteristica* di $\text{Log } x$.

Osservazione 4.51 Nell’Osservazione 4.39 abbiamo visto che e^x è molto vicino a $1+x$ quando $|x|$ è piccolo.¹⁵ Da questo segue che $\log(1+x)$ è molto vicino a x quando $|x|$ è piccolo; scriveremo

$$\log(1+x) \approx x \quad \text{per } x \approx 0. \quad (4.35)$$

Questa formula è uno dei motivi per cui log è chiamato “logaritmo naturale”; la formula analoga per logaritmi in altre basi è meno elegante. Infatti, la (4.34) implica

$$\log_q(1+x) \approx \frac{x}{\log_e q} \quad \text{per } x \approx 0,$$

formula più brutta della (4.35).

Osservazione 4.52 La definizione formale del simbolo \approx è la seguente: $f(x) \approx g(x)$ per $x \approx x_0$ (dove x_0 può essere un numero reale oppure $\pm\infty$) se

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = 1.$$

Se questo accade, a volte si dice che f e g hanno *lo stesso ordine* in x_0 .

CURIOSITÀ 4.13 Supponiamo che la funzione g sia *limitata* vicino a x_0 , cioè che esista $M > 0$ tale che $|g(x)| < M$ per ogni x sufficientemente vicino a x_0 (dove “sufficientemente vicino a $+\infty$ ” significa “abbastanza grande”, e “sufficientemente vicino a $-\infty$ ” significa “abbastanza negativo”). Siccome si può scrivere

$$f(x) - g(x) = \left(\frac{f(x)}{g(x)} - 1 \right) g(x),$$

se $f(x) \approx g(x)$ per $x \approx x_0$ e $g(x)$ è limitata vicino a x_0 , allora $f(x) - g(x)$ tende a zero per x che tende a x_0 .

¹⁴ Per i motivi che vedrai nella prossima osservazione.

¹⁵ Per l’esattezza, nel prossimo capitolo vedremo che $|e^x - 1 - x| \leq e^{|x|} x^2 / 2$.

Osservazione 4.53 Un problema che non affronteremo ma che è importante menzionare è come calcolare i logaritmi in pratica. Esistono degli algoritmi che permettono di approssimare il logaritmo di un numero dato con la precisione voluta usando solo le quattro operazioni; e calcolatori e calcolatrici usano questi algoritmi per calcolare i logaritmi. Certo, è molto meglio adesso di quando (fino a non molti anni fa) si dovevano usare le tavole; ma tieni comunque presente che i logaritmi forniti dai calcolatori sono approssimazioni, e in alcuni casi è importante poter controllare la precisione di queste approssimazioni.

Vediamo ora alcune applicazioni biologiche dei logaritmi, cominciando con alcune questioni lasciate in sospeso nelle sezioni precedenti.

Osservazione 4.54 Il grafico di una funzione potenza $f(x) = ax^p$ è completamente determinato dal passaggio per due punti (x_0, y_0) e (x_1, y_1) , con $x_0 \neq x_1$ e $x_1/x_0, y_1/y_0 > 0$. Infatti da $f(x_j) = y_j$ per $j = 0, 1$ otteniamo $(x_1/x_0)^p = y_1/y_0$, e quindi

$$p = \frac{\log(y_1/y_0)}{\log(x_1/x_0)} \quad \text{e} \quad a = \frac{y_0}{x_0^p} .$$

ESEMPIO 4.19 Nell'Esempio 4.16 avevamo visto che la costante di decadimento λ e il tempo di dimezzamento $t_{1/2}$ di un materiale radioattivo erano legati da

$$e^{-\lambda t_{1/2}} = \frac{1}{2} .$$

Calcolando il logaritmo naturale di entrambi i membri otteniamo

$$-\lambda t_{1/2} = \log \frac{1}{2} = -\log 2 ,$$

per cui il tempo di dimezzamento è dato da

$$t_{1/2} = \frac{\log 2}{\lambda} .$$

In particolare, $e^{-\lambda t} = e^{-(\log 2)t/t_{1/2}} = 2^{-t/t_{1/2}}$, per cui la formula per il numero $N(t)$ di atomi radioattivi presenti al tempo t diventa

$$N(t) = N_0 2^{-t/t_{1/2}} .$$

ESEMPIO 4.20 Nell'Osservazione 4.46 avevamo bisogno di risolvere l'equazione

$$a(1 - e^{-k(x_1 - x_0)}) + b = b + \frac{a}{3} .$$

Sottraendo b a entrambi i membri, e dividendo il risultato per a , ci riconduciamo a

$$1 - e^{-k(x_1 - x_0)} = \frac{1}{3} , \quad \text{ovvero} \quad e^{-k(x_1 - x_0)} = \frac{2}{3} .$$

Calcolando il logaritmo naturale di entrambi i membri otteniamo

$$k(x_1 - x_0) = \log(3/2) ,$$

per cui la soluzione dell'equazione è $x_1 = x_0 + k^{-1} \log(3/2)$. Nella formula (4.29), però, non compariva direttamente x_1 , ma piuttosto $k^2(x - x_0)^2/2$ per $x_0 \leq x \leq x_1$. Quindi otteniamo

$$\forall x \in [x_0, x_1] \quad \frac{k^2(x - x_0)^2}{2} \leq \frac{k^2(x_1 - x_0)^2}{2} = \frac{[\log(3/2)]^2}{2} \simeq 0.0822 ,$$

da cui segue subito (4.30).

ESEMPIO 4.21 Si è verificato sperimentalmente in numerose situazioni che le variazioni *percepite* nell'intensità di uno stimolo non sono proporzionali alla variazione assoluta dell'intensità, ma alla variazione relativa. Per esempio, mentre percepiamo bene la differenza di peso fra 10 g e 20 g, abbiamo difficoltà a percepire la differenza di peso fra 1000 g e 1010 g, anche se la variazione assoluta di peso è la stessa. Aumentare la dose di un farmaco da 5 mg a 10 mg potrebbe causare una risposta significativamente diversa; aumentarla da 1000 mg a 1005 mg probabilmente no. La frequenza del La centrale è 440 Hz, quella del La un'ottava sopra è 880 Hz, e quella del La due ottave sopra è 1760 Hz: l'aumento di frequenza necessario per passare da un'ottava percepita alla successiva non è costante, ma raddoppia ogni volta.

In tutti questi casi (e altri) vale la *legge di Weber: la variazione assoluta dell'intensità percepita è proporzionale alla variazione relativa dello stimolo*. In simboli, se indichiamo con P l'intensità percepita, con s lo stimolo, e con ΔP e Δs le relative variazioni assolute, abbiamo la formula

$$\Delta P = \lambda \frac{\Delta s}{s} . \quad (4.36)$$

Nel prossimo capitolo vedremo che questo implica che P deve dipendere dal logaritmo di s , e più precisamente che si deve avere

$$P(s) = \lambda \log(s/s_0) , \quad (4.37)$$

dove s_0 rappresenta la soglia della percezione: si percepisce un segnale solo se lo stimolo ha un'intensità almeno pari a s_0 .

Diamo un'idea del perché la funzione (4.37) può soddisfare una legge tipo (4.36), almeno per variazioni relative $\Delta s/s$ piccole. Se P è data da (4.37), otteniamo

$$\begin{aligned} \Delta P &= P(s + \Delta s) - P(s) = \lambda \log \frac{s + \Delta s}{s_0} - \lambda \log \frac{s}{s_0} = \lambda \log \frac{s + \Delta s}{s} \\ &= \lambda \log \left(1 + \frac{\Delta s}{s} \right) \approx \lambda \frac{\Delta s}{s} \end{aligned}$$

per $\Delta s/s \approx 0$, grazie all'Osservazione 4.51.

ESEMPIO 4.22 Un tipico caso di applicazione della legge di Weber consiste nella misura dell'intensità percepita dei suoni. L'intensità assoluta dei suoni è una potenza per unità d'area, e quindi è misurata in Watt/m². La soglia della percezione uditiva per un suono di frequenza 1000 Hz è $I_0 = 10^{-12}$ W/m²: l'orecchio umano di solito non è in grado di sentire suoni di intensità inferiore (e con quella frequenza). Allora l'intensità percepita $P(I)$ di un suono di intensità assoluta I è misurata in *decibel*, abbreviati in dB, con la formula

$$P(I) = 10 \operatorname{Log} \left(\frac{I}{I_0} \right) \text{ dB}. \quad (4.38)$$

Il prossimo esempio contiene un'applicazione di queste formule.

ESEMPIO 4.23 Una centrifuga di laboratorio causa un rumore di 64 dB a una distanza di 1 m. Qual è l'intensità assoluta del suono? Se le centrifughe sono due, come cambia il valore in decibel del suono? Se ricopriamo entrambe le centrifughe con un isolante acustico che a 1 m di distanza riduce il suono di 45 dB, di quanto abbiamo ridotto l'intensità assoluta del suono? La (4.38) ci dice che per rispondere alla prima domanda dobbiamo risolvere l'equazione

$$10 \operatorname{Log} \left(\frac{I}{I_0} \right) = 64.$$

Dividendo per 10 e applicando la funzione esponenziale 10^x a entrambi i membri otteniamo

$$I = 10^{6.4} I_0 = 10^{6.4-12} = 10^{-5.6} \simeq 2.5 \cdot 10^{-6} \text{ W/m}^2.$$

Avere due centrifughe equivale a raddoppiare l'intensità assoluta del suono emesso. Siccome

$$10 \operatorname{Log} \left(\frac{2I}{I_0} \right) = 10 \operatorname{Log} 2 + 10 \operatorname{Log} \left(\frac{I}{I_0} \right),$$

raddoppiando l'intensità il rumore in decibel aumenta di $10 \operatorname{Log} 2 \simeq 3.01$ dB, per cui il rumore totale è circa

$$64 + 3.01 = 67.01 \text{ dB}.$$

L'isolante acustico riduce il rumore di 45 dB, portandolo a $67.01 - 45 = 22.01$ dB. Quindi l'intensità assoluta I_{rid} del suono ridotto deve soddisfare l'equazione

$$10 \operatorname{Log} \left(\frac{I_{\text{rid}}}{I_0} \right) = 22.01,$$

e quindi vale

$$I_{\text{rid}} = 10^{2.201} I_0 = 10^{-9.779} \simeq 1.589 \cdot 10^{-10} \text{ W/m}^2.$$

4.9 Tecniche di interpolazione

I logaritmi permettono di usare il metodo dei minimi quadrati anche per interpolare funzioni esponenziali, potenza, e (almeno in parte) logistiche; vedremo anche una tecnica per interpolare funzioni razionali fratte.

In tutti questi casi, l'idea è di applicare opportune trasformazioni ad ascisse e ordinate in modo da trasformare la funzione che si vuole interpolare in una funzione lineare, su cui poi applicare il metodo dei minimi quadrati.

L'esempio più semplice è dato dalle funzioni esponenziali. Calcolando il logaritmo di entrambi i membri di $f(x) = aq^x$ con $a > 0$ otteniamo

$$\text{Log } f(x) = \text{Log } a + (\text{Log } q)x ;$$

in altre parole, se f è una funzione esponenziale di x allora $\text{Log } f$ è una funzione lineare di x . Dunque per interpolare una funzione esponenziale $f(x) = aq^x$ nei dati $(x_1, y_1), \dots, (x_n, y_n)$, con ascisse tutte distinte e ordinate tutte positive, si procede come segue:

- si calcolano le coppie $(x_1, \text{Log } y_1), \dots, (x_n, \text{Log } y_n)$;
- si applica il metodo dei minimi quadrati per trovare la retta di regressione $y = \bar{m}x + \bar{d}$ dei dati $(x_1, \text{Log } y_1), \dots, (x_n, \text{Log } y_n)$;
- si ricava $q = 10^{\bar{m}}$ e $a = 10^{\bar{d}}$;
- si calcola il coefficiente di Pearson per controllare la qualità dell'interpolazione.

Osservazione 4.55 Ovviamente nulla vieta di usare il logaritmo naturale \log (o un logaritmo in qualsiasi altra base) al posto del logaritmo decimale Log .

Vediamo un esempio di applicazione di questo metodo.

ESEMPIO 4.24 Una colonia di batteri cresce in condizioni ideali di luce, calore e abbondanza di cibo. Il tuo assistente conta (uno per uno...) il numero di batteri $N(t)$ presenti al tempo t , a partire da $t = 0$ fino a $t = 120$ minuti, ottenendo i risultati riportati nelle prime due colonne della Tabella 4.2. Visti i dati (e le abitudini riproduttive dei batteri), tu sospetti una crescita esponenziale della forma $N(t) = aq^t$; vogliamo usare il metodo dei minimi quadrati per trovare a e q , e calcolare il coefficiente di correlazione di Pearson per verificare la plausibilità del tuo sospetto.

La prima cosa da fare è creare la Tabella 4.2, analoga alla Tabella 4.1 ma con $y = \text{Log } N$. *Attenzione:* per semplicità, nella Tabella 4.2 ho riportato i valori arrotondati alla seconda cifra decimale, ma i conti sono stati fatti con un numero ben maggiore di cifre decimali, in modo da avere una precisione di almeno due cifre decimali anche nel risultato finale (ricorda quanto abbiamo visto nel Capitolo 1 sui calcoli approssimati).

Partendo da questi dati otteniamo, con le tecniche viste nella Sezione 4.3, i parametri della retta di regressione:

$$\bar{m} \simeq 0.013 , \quad \bar{d} \simeq 2.67 .$$

| Tempo ($x = t$) | Numero (N) | $y = \log N$ | xy | x^2 | y^2 |
|-------------------|----------------|--------------|--------|--------|----------|
| 0 | 470 | 2.67 | 0 | 0 | 7.14 |
| 10 | 653 | 2.81 | 28.15 | 100 | 7.92 |
| 25 | 1019 | 3.01 | 75.20 | 625 | 9.05 |
| 45 | 1720 | 3.24 | 145.60 | 2025 | 10.47 |
| 60 | 3055 | 3.49 | 209.10 | 3600 | 12.15 |
| 75 | 4812 | 3.68 | 276.17 | 5625 | 13.56 |
| 100 | 10101 | 4.00 | 400.44 | 10000 | 16.03 |
| 120 | 18646 | 4.27 | 512.47 | 14400 | 18.24 |
| Media | 54.375 | 5060 | 3.40 | 205.89 | 4546.875 |
| | | | | | 11.82 |

TABELLA 4.2

Quindi

$$a = 10^{\bar{d}} \simeq 469.59, \quad q = 10^{\bar{m}} \simeq 1.03,$$

per cui otteniamo $N(t) = 469.59 \cdot 1.03^t$. Il coefficiente di Pearson è $CP \simeq 0.9996$, che vuol dire che la funzione esponenziale che abbiamo ottenuto approssima molto bene i dati.

Sarebbe naturale adesso rappresentare in un grafico cartesiano sia i dati sperimentali sia la funzione esponenziale interpolata, per poter valutare anche visivamente la qualità dell'approssimazione e poter individuare la presenza di eventuali dati spuri. Però c'è un problema: i valori sulle ordinate presentano un intervallo di variazione molto ampio (variano da 470 a 18646), rendendo difficile la loro rappresentazione in un grafico cartesiano usuale di dimensioni umane.

Questo problema si risolve usando un *grafico* (o *diagramma*) *semilogaritmico*, che ha sull'asse delle ascisse la scala lineare usuale, ma sull'asse delle ordinate una *scala logaritmica*.

La scala logaritmica su un asse si ottiene scegliendo, come al solito, un'origine e un'unità di misura (e un'orientazione) sull'asse, ma poi assegnando al numero reale positivo $x > 0$ il punto sull'asse distante $\log x$ unità dall'origine. Per esempio, al numero 1 viene associata l'origine, al numero 10 il punto posto 1 unità a destra (o sopra) l'origine, al numero $100 = 10^2$ il punto posto 2 unità a destra (o sopra) l'origine, al numero $1/10 = 10^{-1}$ il punto posto 1 unità a sinistra (o sotto) l'origine, al numero 4 il punto posto $\log 4 \simeq 0.6$ unità a destra (o sopra) l'origine, e così via. Il risultato è mostrato in Figura 4.19.

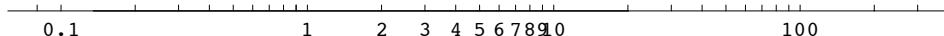


Figura 4.19 La scala logaritmica.

Rappresentare il grafico di una funzione f in un diagramma semilogaritmico equivale quindi a mettere le ordinate in scala logaritmica, che equivale a disegnare in un diagramma cartesiano usuale il grafico di $\log f$. In particolare, rappresentare il grafico di una funzione esponenziale $f(x) = aq^x$ in un diagramma semilogaritmico equivale a disegnare il grafico di $\log f = \log a + (\log q)x$, che è una retta: *il grafico di una funzione esponenziale in un diagramma semilogaritmico è rappresentato da una retta.* Nella Figura 4.20 puoi vedere i dati e la funzione esponenziale interpolata dell'Esempio 4.24, rappresentati a sinistra in un diagramma cartesiano usuale, e a destra in un diagramma semilogaritmico. In particolare, la funzione interpolata diventa la retta di regressione.

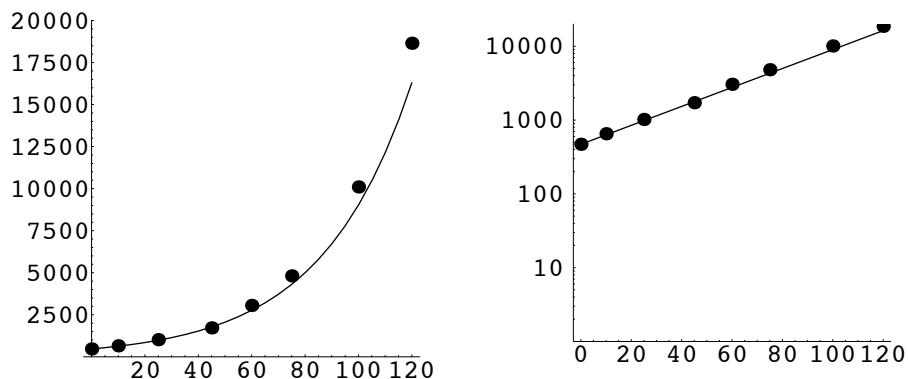


Figura 4.20 Un diagramma semilogaritmico.

Passiamo alle funzioni potenza. Calcolando il logaritmo di entrambi i membri di $f(x) = ax^p$ con $a > 0$ otteniamo

$$\log f(x) = \log a + p \log x ;$$

in altre parole, *se f è una funzione potenza di x allora $\log f$ è una funzione lineare di $\log x$.* Dunque per interpolare una funzione potenza $f(x) = ax^p$ nei dati $(x_1, y_1), \dots, (x_n, y_n)$, con ascisse tutte distinte e ordinate tutte positive, si procede come segue:

- si calcolano le coppie $(\log x_1, \log y_1), \dots, (\log x_n, \log y_n)$;
- si applica il metodo dei minimi quadrati per trovare la retta di regressione $y = \bar{m}x + \bar{d}$ dei dati $(\log x_1, \log y_1), \dots, (\log x_n, \log y_n)$;
- si ricava $p = \bar{m}$ e $a = 10^{\bar{d}}$;
- si calcola il coefficiente di Pearson.

Vediamo un esempio di applicazione di questo metodo.

ESEMPIO 4.25 Stavolta il tuo assistente ha misurato il numero medio $M(D)$ di uova deposte al giorno da femmine di gruppi di moscerini della frutta allevati

a densità D (misurata in numero di moscerini per centimetro quadro) differenti. Visti i dati (e le abitudini riproduttive dei moscerini), tu stavolta sospetti una crescita della forma $M(D) = aD^p$; vogliamo usare il metodo dei minimi quadrati per trovare a e p , e calcolare il coefficiente di correlazione di Pearson per verificare la plausibilità del tuo sospetto.

La prima cosa da fare è creare la Tabella 4.3, che contiene i dati e una colonna in più rispetto al solito, contenente $x = \log D$.

| <i>Densità</i> (D) | $x = \log D$ | <i>Uova</i> (M) | $y = \log M$ | xy | x^2 | y^2 |
|------------------------|--------------|---------------------|--------------|------|-------|-------|
| 5 | 0.70 | 28.7 | 1.46 | 1.02 | 0.49 | 2.12 |
| 10 | 1.00 | 22.6 | 1.35 | 1.35 | 1 | 1.83 |
| 20 | 1.30 | 19.5 | 1.29 | 1.68 | 1.69 | 1.66 |
| 40 | 1.60 | 16.2 | 1.21 | 1.94 | 2.57 | 1.46 |
| 80 | 1.90 | 14.2 | 1.15 | 2.19 | 3.62 | 1.33 |
| <i>Media</i> | 31 | 20.24 | 1.29 | 1.64 | 1.87 | 1.68 |

TABELLA 4.3

Partendo da questi dati otteniamo i parametri della retta di regressione:

$$\bar{m} \simeq -0.25, \quad \bar{d} \simeq 1.62.$$

Quindi

$$a = 10^{\bar{d}} \simeq 41.63, \quad p = \bar{m} \simeq -0.25,$$

per cui otteniamo $N(D) = 41.63 \cdot D^{-0.25}$. Il coefficiente di Pearson è $CP \simeq -0.995$, negativo perché la retta di regressione è decrescente, e anche stavolta la funzione che abbiamo ottenuto approssima molto bene i dati sperimentali.

Possiamo rappresentare le funzioni potenza in un *diagramma bilogaritmico*, in cui usiamo una scala logaritmica sia sull'asse delle ascisse sia sull'asse delle ordinate. Rappresentare il grafico di una funzione f in un diagramma bilogaritmico equivale a mettere ascisse e ordinate in scala logaritmica, che equivale a disegnare in un diagramma cartesiano usuale il grafico di $g(\tilde{x}) = \log f(10^{\tilde{x}})$. Infatti, alla coppia $(x, f(x))$ del grafico di f viene associato il punto $(\log x, \log f(x))$ nel diagramma bilogaritmico; ponendo $\tilde{x} = \log x$ abbiamo $x = 10^{\tilde{x}}$, per cui $(\log x, \log f(x)) = (\tilde{x}, \log f(10^{\tilde{x}}))$, e quindi $(\log x, \log f(x))$ è un punto del grafico di $\log f(10^{\tilde{x}})$.

In particolare, rappresentare il grafico di una funzione potenza $f(x) = ax^p$ in un diagramma bilogaritmico equivale a disegnare il grafico di

$$\log f(10^{\tilde{x}}) = \log a + p \log(10^{\tilde{x}}) = \log a + p\tilde{x},$$

che è una retta: *il grafico di una funzione potenza è rappresentato in un diagramma bilogaritmico da una retta*. Nella Figura 4.21 puoi vedere i dati e la funzione potenza

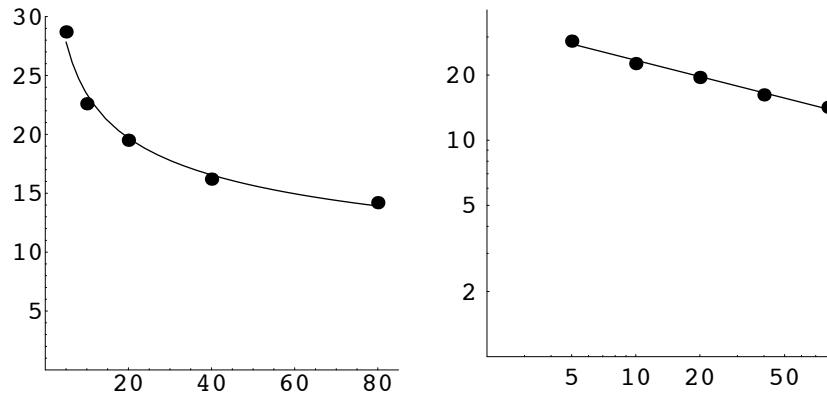


Figura 4.21 Diagramma bilogaritmico.

interpolata dell’Esempio 4.25, rappresentati a sinistra in un diagramma cartesiano usuale e a destra in un diagramma bilogaritmico.

Con un po’ di fantasia e qualche conto possiamo usare queste tecniche per interpolare anche altri tipi di funzioni. Per esempio, se $f(x) = a(1 - e^{-k(x-x_0)}) + b$ allora

$$1 - \frac{f(x) - b}{a} = e^{-k(x-x_0)} \iff \log\left(1 - \frac{f(x) - b}{a}\right) = -kx + kx_0 ,$$

cioè $\log(1 - (f(x) - b)/a)$ è una funzione lineare di x . Se, per altre vie, abbiamo delle informazioni su a (il valore limite) e b (il valore assunto nel punto iniziale x_0), allora possiamo applicare il metodo dei minimi quadrati per ricavare k e x_0 . Oppure, se abbiamo delle informazioni su b e x_0 , possiamo usare il metodo dei minimi quadrati per diversi valori di a cercando di ricavare la migliore approssimazione di x_0 . Oppure ancora, possiamo variare a e b cercando di portare il coefficiente di Pearson più vicino possibile a ± 1 .

ESEMPIO 4.26 Hai dato ordine al tuo assistente di misurare il potenziale elettrico di un neurone isolato di ratto per un’ora; i dati ottenuti sono riportati nella Tabella 4.4 (tempo in minuti, potenziale in milliVolt). Da quanto abbiamo visto, ti aspetti un andamento della forma $V(t) = V_0(1 - e^{-kt})$, e vuoi ricavare i valori di V_0 e k dai dati. La tecnica sopra illustrata suggerisce di calcolare la retta di regressione di $\log(1 - V(t)/V_0)$ per vari valori di V_0 (ricorda che stai supponendo $b = 0$) fino a ottenere una retta che passa per l’origine (in accordo con l’ipotesi $x_0 = 0$) e, probabilmente, un buon coefficiente di Pearson. La Tabella 4.4 riporta, oltre ai dati sperimentali, i valori di $-k = \bar{m}$, $kx_0 = \bar{d}$ e del coefficiente di Pearson ottenuti per vari valori di V_0 . Nota che, siccome V_0 è il valore limite, testiamo valori di V_0 superiori ai valori di potenziale misurati.

Dalla Tabella 4.4 si vede che il valore di V_0 che fornisce il valore di \bar{d} più vicino a 0 è $V_0 = 70$ mV, che è anche quello col miglior coefficiente di Pearson. Quindi puoi sentirti sufficientemente sicuro affermando che il potenziale di questo neurone

| Tempo ($x = t$) | Potenziale (V) | V_0 | $\bar{m} = -k$ | $\bar{d} = kx_0$ | CP |
|-------------------|--------------------|-------|----------------|------------------|----------|
| 0 | 0 | 66 | -0.071 | 0.304 | -0.974 |
| 10 | 25.41 | 68 | -0.053 | 0.078 | -0.998 |
| 20 | 41.51 | 70 | -0.045 | 0.001 | -0.99998 |
| 30 | 51.87 | 72 | -0.040 | -0.043 | -0.9990 |
| 40 | 58.58 | 74 | -0.036 | -0.069 | -0.9972 |
| 50 | 62.59 | | | | |
| 60 | 65.30 | | | | |

TABELLA 4.4

di ratto ha un andamento ben descritto dalla funzione $V(t) = 70(1 - e^{-0.045t})$. Dati e funzione interpolata sono mostrati nella prima immagine in Figura 4.22.

Argomenti analoghi si possono applicare anche alle funzioni logistiche. Infatti da $f(x) = b + a/(1 + e^{-k(x-x_0)})$ deduciamo

$$\begin{aligned} \frac{f(x) - b}{a} = \frac{1}{1 + e^{-k(x-x_0)}} &\iff \frac{a}{f(x) - b} - 1 = e^{-k(x-x_0)} \\ &\iff \log\left(\frac{a}{f(x) - b} - 1\right) = -kx + kx_0 . \end{aligned}$$

Quindi avendo informazioni sui valori limite a e b si può usare il metodo dei minimi quadrati per determinare $-k$ e kx_0 ; oppure, di nuovo, possiamo variare i parametri a e/o b cercando di ottenere il valore migliore per il coefficiente di Pearson.

ESEMPIO 4.27 Vogliamo interpolare una funzione logistica ai dati relativi alle percentuali di germinazione dei semi di pomodoro in funzione della temperatura. In questo caso abbiamo una scelta praticamente obbligata per i valori limite: $a = 100$ e $b = 0$. La Tabella 4.5 contiene nelle prime due colonne i dati sperimentali; nella terza colonna i valori delle ordinate trasformati secondo la legge $y = \log((100/P) - 1)$ suggerita dal ragionamento precedente; e le colonne successive i valori necessari per l'applicazione del metodo dei minimi quadrati.

| Temp. ($x = T$) | Perc. (P) | $y = \log\left(\frac{100}{P} - 1\right)$ | xy | x^2 | y^2 |
|-------------------|---------------|--|--------|-------|-------|
| 0 | 470 | 2.67 | 0 | 0 | 7.14 |
| 9 | 20 | 1.39 | 12.48 | 81 | 1.92 |
| 12 | 40 | 0.41 | 4.87 | 144 | 0.16 |
| 15 | 70 | -0.85 | -12.70 | 225 | 0.72 |
| 18 | 85 | -1.73 | -31.20 | 324 | 3.01 |
| Media | 13.5 | 53.75 | -0.20 | -6.65 | 193.5 |
| | | | | | 1.45 |

TABELLA 4.5

Dai valori della Tabella 4.5 si ottiene $-k = \bar{m} \simeq -0.354$ e $kx_0 = \bar{d} \simeq 4.58$; quindi $e^{-k} \simeq 0.702$ e $x_0 = \bar{d}/k \simeq 12.94$ per cui la funzione logistica interpolata è

$$P(T) = \frac{100}{1 + 0.702^{T-12.94}},$$

come annunciato nell'Esempio 4.18. Il coefficiente di Pearson è $CP \simeq -0.998$, per cui l'interpolazione è decisamente buona.

ESEMPIO 4.28 Una colonia di batteri cresce in condizioni non ideali, che possono sostenere la vita di solo un quantitativo limitato di individui. Anche stavolta il tuo assistente conta il numero $N(t)$ di batteri presenti al tempo t a partire da $t = 0$ fino a $t = 120$, ottenendo i risultati riportati nelle prime due colonne della Tabella 4.6. Questa volta tu sospetti che $N(t)$ sia rappresentato da una funzione logistica. È naturale ipotizzare che il valore limite a $-\infty$ valga $b = 0$; invece non ritieni di fare ipotesi a priori sul valore di x_0 , il punto di mezzo della funzione logistica, o su a , il valore limite a $+\infty$. Quindi applichi il metodo dei minimi quadrati per diversi valori di a , valori ovviamente scelti più grandi dei valori misurati, ottenendo le ultime quattro colonne della Tabella 4.6.

| Tempo ($x = t$) | Numero (N) | a | $\bar{m} = -k$ | $\bar{d} = kx_0$ | CP |
|-------------------|----------------|------|----------------|------------------|---------|
| 0 | 470 | 2116 | -0.063 | 1.58 | -0.9956 |
| 10 | 653 | 2118 | -0.061 | 1.51 | -0.9972 |
| 25 | 1007 | 2120 | -0.059 | 1.47 | -0.9977 |
| 45 | 1492 | 2122 | -0.057 | 1.43 | -0.9978 |
| 60 | 1866 | 2124 | -0.056 | 1.40 | -0.9977 |
| 75 | 2012 | 2126 | -0.055 | 1.37 | -0.9974 |
| 100 | 2099 | 2128 | -0.054 | 1.35 | -0.9971 |
| 120 | 2112 | 2130 | -0.053 | 1.33 | -0.9967 |

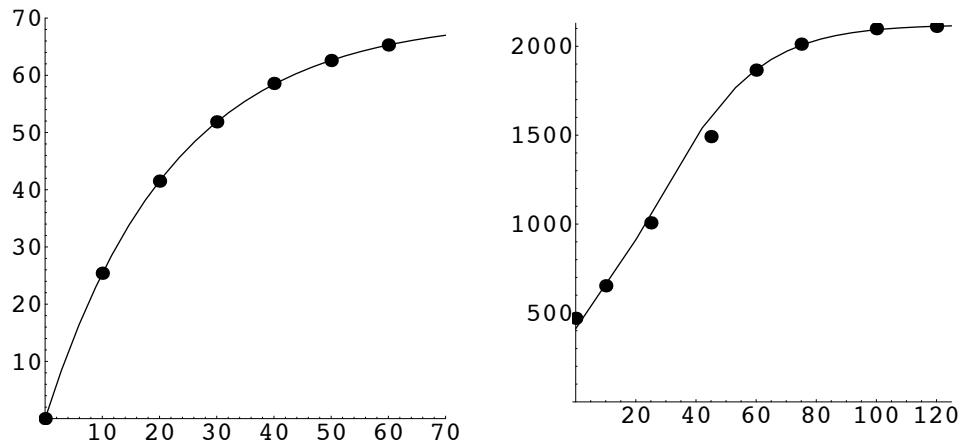
TABELLA 4.6

Dai risultati ottenuti si vede che il valore di a che fornisce il miglior coefficiente di Pearson è $a = 2122$, per il quale otteniamo $k \simeq 0.057$ e $x_0 = \bar{d}/k \simeq 24.97$. Quindi questo metodo ci fornisce la seguente formula

$$N(t) = \frac{2122}{1 + e^{-0.057(t-24.97)}} \simeq \frac{2122}{1 + 0.94^{t-24.97}}.$$

Dati e funzione interpolata sono mostrati nella seconda immagine in Figura 4.22.

Osservazione 4.56 Il metodo dei minimi quadrati consiste nel trovare il punto di minimo di una funzione quadratica di due variabili (m e d). Le tecniche di interpolazione che stiamo discutendo negli ultimi esempi richiedono di trovare il punto di minimo di una funzione non necessariamente quadratica e di più di due variabili (per esempio, a , b , k e x_0). Ci sono tecniche analitiche per affrontare

**Figura 4.22** Tecniche di interpolazione.

problemi di questo genere, ma sfortunatamente sono al di là di quanto possiamo trattare in questo corso.

Come ultimo esempio di questa tecnica vediamo le funzioni lineari fratte. La (4.19) ci dice che ogni funzione lineare fratta (non costante) può essere scritta nella forma

$$f(x) = \alpha + \frac{1}{\beta x + \gamma},$$

dove $\alpha = \lim_{x \rightarrow \pm\infty} f(x)$ è il valore limite di f all'infinito. Quindi otteniamo

$$f(x) - \alpha = \frac{1}{\beta x + \gamma} \quad \iff \quad \frac{1}{f(x) - \alpha} = \beta x + \gamma.$$

Dunque $1/(f(x) - \alpha)$ è funzione lineare di x , e possiamo applicare le solite tecniche di interpolazione.

ESEMPIO 4.29 Vediamo come si è ottenuta la formula dell'Esempio 4.12. Il tuo assistente ha misurato la velocità di espansione del muscolo sartorio della coscia di una rana che sollevava un determinato peso, aumentandolo di 10 grammi in 10 grammi finché la povera rana non ce l'ha fatta più. Le misure ottenute sono riportate nelle prime due colonne della Tabella 4.7. Visti i dati (e conoscendo le abitudini riproduttive delle rane, che sembrano non c'entrare niente ma non si sa mai) hai ipotizzato un andamento lineare fratto, e hai applicato il metodo dei minimi quadrati per diversi valori di α ; l'unica limitazione che ti sei posto è che α dev'essere negativo (in quanto dev'essere minore dei valori misurati). Le ultime quattro colonne della Tabella 4.7 riportano i valori di β e γ e il coefficiente di Pearson ottenuti al variare di α .

| Peso ($x = p$) | Velocità (v) | α | $\bar{m} = \beta$ | $\bar{d} = \gamma$ | CP |
|------------------|------------------|----------|-------------------|--------------------|--------|
| 0 | 5.542 | -0.35 | 0.037 | -0.103 | 0.9672 |
| 10 | 2.591 | -0.65 | 0.019 | 0.101 | 0.9955 |
| 20 | 1.484 | -0.95 | 0.013 | 0.154 | 0.9999 |
| 30 | 0.905 | -1.25 | 0.009 | 0.170 | 0.9979 |
| 40 | 0.548 | -1.55 | 0.007 | 0.172 | 0.9934 |
| 50 | 0.306 | | | | |
| 60 | 0.132 | | | | |
| 70 | 0 | | | | |

TABELLA 4.7

Il valore migliore di α è quindi $\alpha = -0.95$, col quale otteniamo

$$v(p) = -0.95 + \frac{1}{0.013p + 0.154} \simeq 0.95 \left(\frac{70 - p}{p + 12} \right),$$

che è la formula usata nell'Esempio 4.12. Dati e funzione sono riportati nella Figura 4.23.

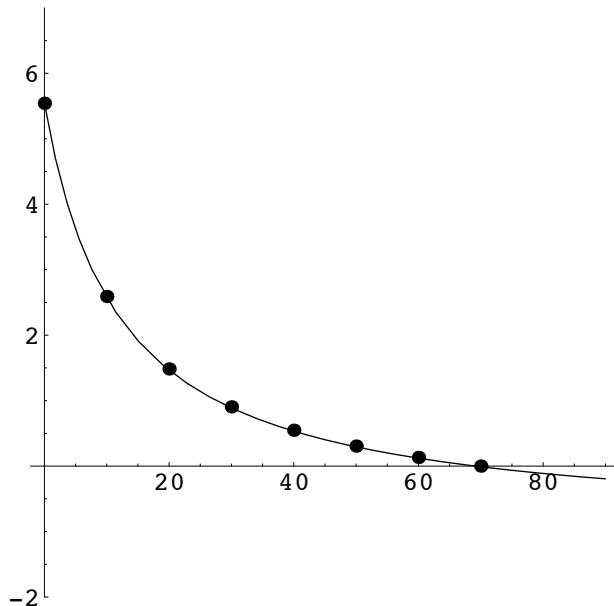


Figura 4.23 Interpolazione di funzioni lineari fratte.

4.10 Funzioni trigonometriche

Le funzioni che abbiamo visto finora sono decisamente utili e hanno una vasta applicabilità, ma non sono adatte a rappresentare un tipo di fenomeni piuttosto comune in natura: i fenomeni periodici. Un fenomeno è *periodico* se si ripete (più o meno) uguale a intervalli prestabiliti di tempo. Analogamente, una funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ è *periodica di periodo $T > 0$* se

$$\forall x \in \mathbb{R} \quad f(x + T) = f(x).$$

In altre parole, una funzione f è periodica di periodo T se il suo grafico non cambia traslandolo orizzontalmente (a destra o a sinistra) di T unità; in altre parole ancora, l'intero grafico si ottiene giustappponendo copie del grafico di f ristretta a un qualsiasi intervallo chiuso di lunghezza T .

Nessuna delle funzioni che abbiamo visto finora era periodica (a parte le costanti). Scopo di questa sezione è introdurre le funzioni periodiche di base; in un certo senso, tutte le altre funzioni periodiche si possono ottenere a partire da queste (vedi la fine del capitolo).

Nella matematica elementare esistono degli oggetti che presentano una natura periodica: gli angoli. Sai bene che un angolo di 90° e un angolo di $90 + 360 = 450^\circ$ sono in realtà lo stesso angolo: aggiungendo o togliendo 360° da un angolo la sua ampiezza non muta. Il nostro obiettivo sarà usare questa periodicità naturale per costruire funzioni periodiche a valori reali.

Abbiamo appena parlato di angoli misurandoli in gradi, ma il grado non è l'unità di misura migliore per gli angoli. Il principale problema è che si tratta di una unità di misura del tutto slegata dall'unità di misura delle lunghezze, e quindi priva di significato geometrico. Un'unità di misura molto più naturale, e strettamente legata all'unità di misura delle lunghezze è invece il radiante (abbreviato, se necessario, con “rad”). Per definizione, l'angolo di 1 radiante è l'angolo che sottende un arco di lunghezza 1 in una circonferenza di raggio 1. Di conseguenza, *la misura in radianti di un angolo coincide con la lunghezza dell'arco sotteso in una circonferenza di raggio unitario*.

Per esempio, l'angolo giro (360°) sottende l'intera circonferenza, che è lunga 2π ; quindi l'angolo giro misura 2π radianti. L'angolo piatto (180°) sottende mezza circonferenza, che è lunga π ; quindi l'angolo piatto misura π radianti. Analogamente, l'angolo retto (90°) sottende un quarto di circonferenza, per cui misura $\pi/2$ radianti. In generale, un angolo di d° sottende $d/360$ di circonferenza, per cui misura $2\pi d/360$ radianti:

$$d^\circ = \frac{d}{180} \pi \text{ rad}.$$

In particolare, la misura in radianti di angoli acuti comuni è

$$1^\circ = \frac{\pi}{180} \text{ rad}, \quad 30^\circ = \frac{\pi}{6} \text{ rad}, \quad 45^\circ = \frac{\pi}{4} \text{ rad}, \quad 60^\circ = \frac{\pi}{3} \text{ rad}, \quad 90^\circ = \frac{\pi}{2} \text{ rad}.$$

Viceversa, avendo la misura in radianti si recupera la misura in gradi moltiplicando

per $180/\pi$:

$$r \text{ rad} = \left(\frac{180r}{\pi} \right)^\circ .$$

In particolare,

$$1 \text{ rad} = \left(\frac{180}{\pi} \right)^\circ \simeq 57.29^\circ ,$$

dove 57.29° significa “57 gradi e 29 centesimi di grado”; nella pratica scientifica i minuti e i secondi di grado non si usano mai.

Gli angoli hanno un verso: possiamo percorrerli in senso orario oppure in senso antiorario. Di conseguenza, la misura degli angoli ha un segno: esistono angoli positivi e angoli negativi. Per convenzione, *si considera positivo un angolo percorso in senso antiorario*, e negativo un angolo percorso in senso orario. Per esempio, l'angolo dalla semiretta positiva delle ascisse alla semiretta positiva delle ordinate è $+\pi/2$ (in radianti), mentre l'angolo dalla semiretta positiva delle ascisse alla semiretta negativa delle ordinate è $-\pi/2$.

Come accennato sopra, la misura degli angoli contiene una periodicità intrinseca. Siccome ruotare di un angolo giro (360° ovvero 2π radianti) ci riporta al punto di partenza, un angolo di r radianti (o d°) e un angolo di $r + 2\pi$ radianti (o $(d + 360)^\circ$) identificano lo stesso spicchio di piano. Analogamente, un angolo di r radianti (o d°) e un angolo di $r - 2\pi$ radianti (o $(d - 360)^\circ$) identificano lo stesso spicchio di piano; in generale, un angolo di r radianti (o d°) e un angolo di $r + 2k\pi$ radianti (o $(d + 360k)^\circ$) identificano lo stesso spicchio di piano quale che sia $k \in \mathbb{Z}$. In questo senso, la misura degli angoli è periodica di periodo 2π radianti (o 360°).

Possiamo ora sfruttare questa periodicità degli angoli per definire le due principali funzioni periodiche. Sia $C \subset \mathbb{R}^2$ la circonferenza nel piano cartesiano di centro l'origine e raggio unitario. Per ogni $\phi \in \mathbb{R}$ l'angolo di ϕ radianti (misurato a partire dalla semiretta positiva delle ascisse) identifica in modo unico un punto $P(\phi)$ della circonferenza C , ottenuto partendo dal punto $(1, 0)$ e ruotando di ϕ , cioè seguendo la circonferenza per una lunghezza pari a $|\phi|$ procedendo in senso antiorario se $\phi \geq 0$, e in senso orario se $\phi \leq 0$. Il coseno $\cos \phi$ dell'angolo ϕ è allora l'ascissa del punto $P(\phi)$, mentre il seno¹⁶ $\sin \phi$ di ϕ è l'ordinata di $P(\phi)$; vedi la Figura 4.24.

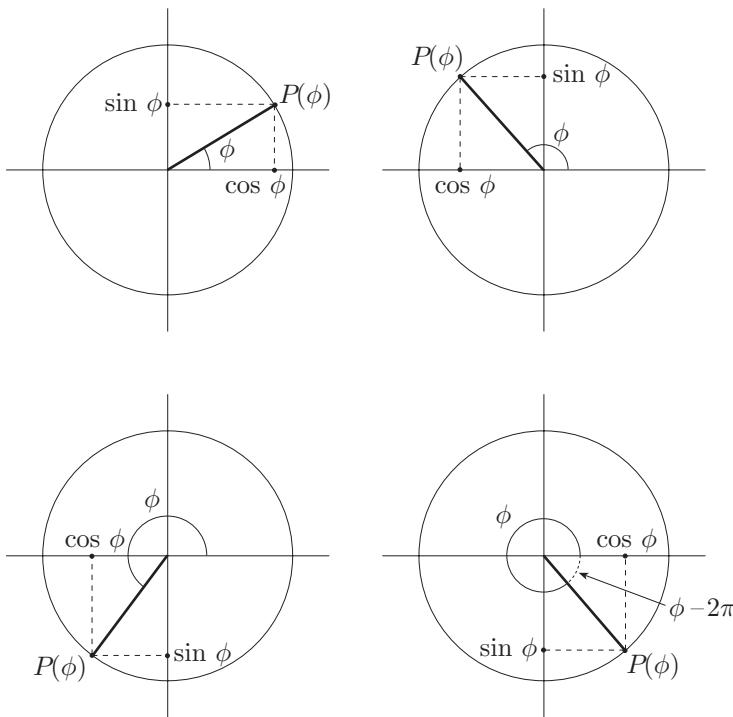
In questo modo abbiamo definito due funzioni $\cos, \sin: \mathbb{R} \rightarrow \mathbb{R}$ chiaramente periodiche di periodo 2π :

$$\forall k \in \mathbb{Z} \quad \cos(\phi + 2k\pi) = \cos \phi \quad \text{e} \quad \sin(\phi + 2k\pi) = \sin \phi .$$

Siccome l'ascissa e l'ordinata dei punti della circonferenza di centro l'origine e raggio unitario variano fra 1 e -1 , lo stesso fanno seno e coseno:

$$\forall \phi \in \mathbb{R} \quad -1 \leq \cos \phi \leq 1 , \quad -1 \leq \sin \phi \leq 1 ,$$

¹⁶ In questo corso indicheremo sempre il seno con il simbolo internazionale \sin , e mai con sen .

**Figura 4.24** Seno e coseno.

e, più precisamente, le funzioni seno e coseno hanno come immagine esattamente l'intervallo chiuso $[-1, 1]$.

Seguendo l'andamento del coseno per angoli crescenti, vediamo che parte da un punto di massimo $\cos 0 = 1$ e poi decresce, passando per $\cos(\pi/2) = 0$ fino a raggiungere il punto di minimo $\cos \pi = -1$. A quel punto ricomincia a crescere, ripassando per $\cos(3\pi/2) = 0$ fino a ritornare a un punto di massimo $\cos(2\pi) = 1$. Per $\phi > 2\pi$ (o $\phi < 0$) la periodicità ci assicura che l'andamento si ripete. In particolare,

- $\cos \phi = 1$ nel punto di massimo $\phi = 0$ e, più in generale, nei punti di massimo $\phi = 2k\pi$, con $k \in \mathbb{Z}$;
- $\cos \phi = -1$ nel punto di minimo $\phi = \pi$ e, più in generale, nei punti di minimo $\phi = \pi + 2k\pi$, con $k \in \mathbb{Z}$;
- $\cos \phi$ si annulla in $\phi = \pi/2, 3\pi/2$ e, più in generale, negli angoli $\phi = \pi/2 + k\pi$ con $k \in \mathbb{Z}$;
- $\cos \phi$ è positivo nell'intervallo $(-\pi/2, \pi/2)$ e, più in generale, negli intervalli $(2k\pi - \pi/2, 2k\pi + \pi/2)$ al variare di $k \in \mathbb{Z}$;
- $\cos \phi$ è strettamente decrescente nell'intervallo $[0, \pi]$ e, più in generale, negli intervalli $[2k\pi, (2k+1)\pi]$ al variare di $k \in \mathbb{Z}$;
- $\cos \phi$ è strettamente crescente nell'intervallo $[\pi, 2\pi]$ e, più in generale, negli intervalli $[(2k-1)\pi, 2k\pi]$ al variare di $k \in \mathbb{Z}$.

La Figura 4.25 mostra il grafico di $\cos \phi$ (linea continua).

Seguendo invece il seno per ϕ crescenti, vediamo che parte da $\sin 0 = 0$ e poi cresce, raggiungendo un punto di massimo $\sin(\pi/2) = 1$. A quel punto comincia a decrescere, annullandosi di nuovo in $\sin \pi = 0$, fino a raggiungere un punto di minimo $\cos(3\pi/2) = -1$. Poi risale, fino a tornare a $\sin(2\pi) = 0$. Per $\phi > 2\pi$ (o $\phi < 0$) la periodicità ci assicura che l'andamento si ripete. In particolare,

- $\sin \phi = 1$ nel punto di massimo $\phi = \pi/2$ e, più in generale, nei punti di massimo $\phi = 2k\pi + \pi/2$, con $k \in \mathbb{Z}$;
- $\sin \phi = -1$ nel punto di minimo $\phi = -\pi/2$ e, più in generale, nei punti di minimo $\phi = 2k\pi - \pi/2$, con $k \in \mathbb{Z}$;
- $\sin \phi$ si annulla in $\phi = 0, \pi$ e, più in generale, negli angoli $\phi = k\pi$ con $k \in \mathbb{Z}$;
- $\sin \phi$ è positivo nell'intervallo $(0, \pi)$ e, più in generale, in $(2k\pi, (2k+1)\pi)$ al variare di $k \in \mathbb{Z}$;
- $\sin \phi$ è strettamente crescente nell'intervallo $[-\pi/2, \pi/2]$ e, più in generale, negli intervalli $[2k\pi - \pi/2, 2k\pi + \pi/2]$ al variare di $k \in \mathbb{Z}$;
- $\cos \phi$ è strettamente decrescente nell'intervallo $[\pi/2, 3\pi/2]$ e, più in generale, negli intervalli $[2k\pi + \pi/2, 2k\pi + 3\pi/2]$ al variare di $k \in \mathbb{Z}$.

La Figura 4.25 mostra il grafico di $\sin \phi$ (linea tratteggiata).

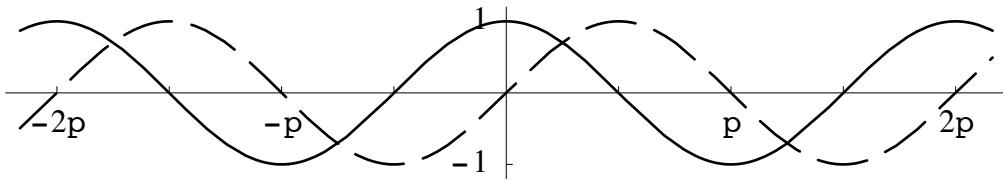


Figura 4.25 Coseno e seno.

Osservazione 4.57 Dalla figura è evidente (e lo verificheremo rigorosamente fra poco) che il grafico di $\cos \phi$ si ottiene traslando verso sinistra di $\pi/2$ il grafico di $\sin \phi$. Il grafico di $\sin \phi$ è detto *sinusoide*; qualunque altra curva che si ottiene da una sinusoide tramite traslazioni o moltiplicazioni di ascisse e/o ordinate si chiama *curva sinusoidale*, ed è il grafico di una *funzione sinusoidale*.

Osservazione 4.58 Una funzione sinusoidale $f: \mathbb{R} \rightarrow \mathbb{R}$ è determinata da quattro grandezze:

- il *periodo* (che per seno e coseno è 2π);
- l'*ampiezza*, data da $(M - m)/2$, dove M è il valore massimo e m il valore minimo di f , ed è quindi metà dell'intervallo di variazione dei valori di f (e per seno e coseno vale 1);
- il *valor medio*, dato da $(M + m)/2$, ed è quindi il punto centrale dell'intervallo di variazione di f (e per seno e coseno vale 0);
- la *fase*, che è il primo punto non negativo di massimo (per cui il coseno ha fase 0 mentre il seno ha fase $\pi/2$).

Vedremo fra poco come ottenere funzioni sinusoidali che abbiano periodo, ampiezza, valor medio e fase prestabiliti.

Osservazione 4.59 Seno e coseno ci forniscono un esempio di funzioni che *non* hanno limite per x che tende a $+\infty$ o $-\infty$. Infatti, al crescere di x il valore di $\sin x$ e di $\cos x$ continua a oscillare fra -1 e 1 , senza mai tendere ad alcun limite, finito o infinito; e la stessa cosa accade quando x diventa arbitrariamente negativo. Sempre usando seni e coseni è possibile costruire funzioni che *non* ammettono limite anche per x che tende a un valore $x_0 \in \mathbb{R}$ finito. Per esempio, la funzione $f(x) = \sin(1/x)$, definita per ogni $x \neq 0$, non ammette limite per x che tende a 0 , né da sopra né da sotto: infatti, quando x si avvicina a zero $1/x$ diventa arbitrariamente grande (o arbitrariamente negativo), per cui $\sin(1/x)$ oscilla freneticamente fra -1 e 1 senza mai tendere a un valore limite, come si può vedere nella Figura 4.26.

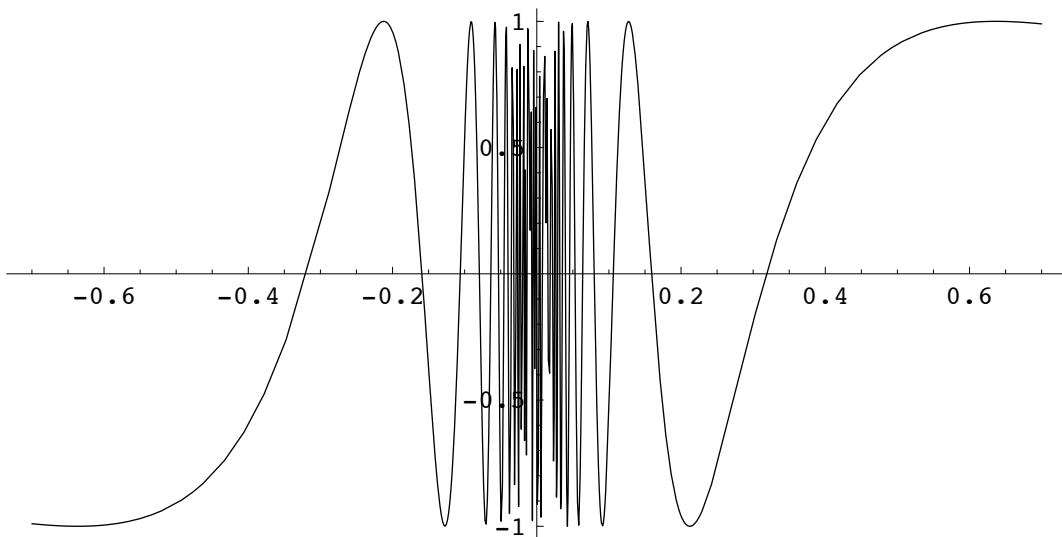


Figura 4.26 $f(x) = \sin(1/x)$.

Osservazione 4.60 Una conseguenza dell'osservazione precedente è che non sempre i limiti esistono. In diversi casi capire se un dato limite esiste o meno e, se esiste, calcolarlo, può essere un'impresa tutt'altro che banale. Un primo esempio di limite di questo genere l'abbiamo incontrato nella Sezione 4.7 (Osservazione 4.39 e Curiosità 4.11); un altro esempio è legato alla funzione seno. Noi sappiamo che $\sin 0 = 0$. Inoltre, $\sin \phi$ è l'ordinata del punto $P(\phi)$, che si ottiene partendo da $P(0) = (1, 0)$ e seguendo la circonferenza per una lunghezza ϕ . Ora, per ϕ piccolo l'arco da $P(0)$ a $P(\phi)$ pare distare sempre meno dalla corda da $P(0)$ a $P(\phi)$, e quest'ultima tende a confondersi con il segmento verticale di altezza $\sin \phi$ uscente da $P(\phi)$. Quindi sembra plausibile aspettarsi che qualcosa del genere valga:

$$\lim_{\phi \rightarrow 0} \frac{\sin \phi}{\phi} = 1 , \quad (4.39)$$

cioè $\sin x \approx x$ per $x \approx 0$. Ti deve essere ben chiaro che il discorso precedente *non* dimostra la validità di (4.39), ma semplicemente rende plausibile una simile

affermazione — e quindi spinge a cercare di verificare se è vera o non lo è. La Curiosità 4.17 conterrà una dimostrazione corretta di (4.39); la Curiosità 4.14 contiene invece un esempio in cui un ragionamento di questo genere porta a una conclusione sbagliata. La morale è che fidarsi è bene ma non fidarsi è meglio: quando si tratta di ragionamenti che coinvolgono limiti o, più in generale, quantità che diventano arbitrariamente piccole o arbitrariamente grandi, bisogna procedere con i piedi di piombo e controllare (o far controllare a un matematico) ogni passaggio.

CURIOSITÀ 4.14 Indichiamo con ℓ_1 la lunghezza di una semicirconferenza che ha come diametro l'intervallo $[0, 1]$. Poi indichiamo con ℓ_2 la somma delle lunghezze delle semicirconferenze di diametro $[0, 1/2]$ e $[1/2, 1]$. Più in generale, indichiamo con ℓ_n la somma delle lunghezze delle n semicirconferenze di diametro gli intervalli $[0, 1/n]$, $[1/n, 2/n], \dots, [(n-1)/n, 1]$. Tutte queste semicirconferenze si avvicinino sempre di più all'intervallo $[0, 1]$; sembrerebbe quindi naturale aspettarsi che la somma delle loro lunghezze si avvicini sempre più alla lunghezza dell'intervallo $[0, 1]$, cioè che si abbia

$$\lim_{n \rightarrow +\infty} \ell_n = 1 .$$

Invece, questo è falso. Per vederlo, basta calcolare ℓ_n . Infatti, ℓ_1 è la lunghezza di una semicirconferenza di diametro 1, cioè raggio $1/2$, per cui $\ell_1 = \pi$. Al passo n , ciascuna semicirconferenzina ha diametro $1/n$, per cui ha lunghezza π/n . Ma ce ne sono n : quindi $\ell_n = n \cdot (\pi/n) = \pi$. In altre parole, $\ell_n = \pi$ per ogni n , e il risultato esatto è quindi

$$\lim_{n \rightarrow +\infty} \ell_n = \pi .$$

La spiegazione di questo apparente paradosso è che le circonferenzine pur avvicinandosi al segmento diventano sempre più curve; e questo aumento della curvatura controbilancia il rimpicciolimento del diametro lasciando costante la lunghezza totale.

Vediamo ora di raccogliere le principali proprietà delle funzioni seno e coseno. Prima di tutto, cambiare di segno a un angolo significa ruotare in senso opposto; quindi i punti $P(\phi)$ e $P(-\phi)$ sono simmetrici rispetto all'asse delle ascisse, e dunque

$$\cos(-\phi) = \cos \phi \quad \text{e} \quad \sin(-\phi) = -\sin \phi . \quad (4.40)$$

In altre parole, il coseno è una funzione pari mentre il seno è una funzione dispari.

Ruotando di $\pm 180^\circ$ (cioè aggiungendo $\pm \pi$ agli angoli) cambiamo di segno ascisse e ordinate dei punti; quindi

$$\cos(\phi \pm \pi) = -\cos \phi , \quad \sin(\phi \pm \pi) = -\sin \phi . \quad (4.41)$$

Ruotando in senso antiorario di 90° (cioè aggiungendo $\pi/2$ agli angoli) trasformiamo le ascisse in ordinate, e le ordinate nelle ascisse cambiate di segno; quindi

$$\cos\left(\phi + \frac{\pi}{2}\right) = -\sin \phi , \quad \sin\left(\phi + \frac{\pi}{2}\right) = \cos \phi , \quad (4.42)$$

come anticipato nell'Osservazione 4.57. Queste proprietà sono visualizzate nella Figura 4.27.

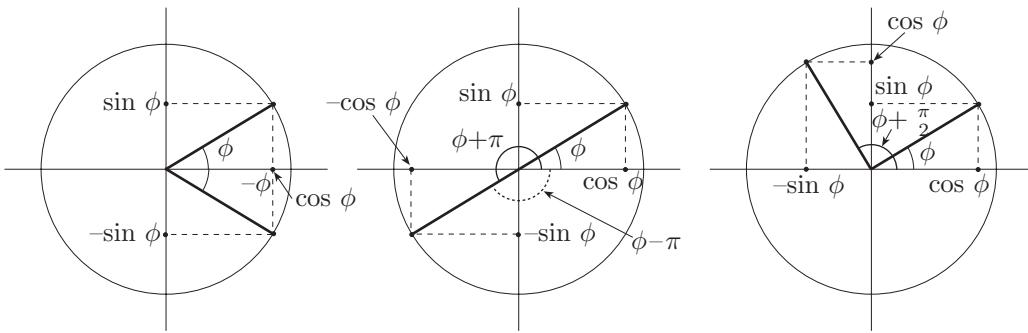


Figura 4.27 .

E se ruotiamo in senso orario di $\pi/2$ radianti? Cosa succede lo deduciamo subito da quanto ottenuto:

$$\cos\left(\phi - \frac{\pi}{2}\right) = \cos\left[\left(\phi + \frac{\pi}{2}\right) - \pi\right] = -\cos\left(\phi + \frac{\pi}{2}\right) = \sin\phi , \quad (4.43)$$

e analogamente (esercizio)

$$\sin\left(\phi - \frac{\pi}{2}\right) = -\cos\phi . \quad (4.44)$$

Essendo $P(\phi)$ un punto della circonferenza di centro l'origine e raggio unitario, la distanza fra $P(\phi)$ e l'origine è esattamente 1. Scrivendo la distanza dall'origine in coordinate otteniamo la relazione fondamentale che lega seni e coseni:

$$\sin^2\phi + \cos^2\phi \equiv 1 , \quad (4.45)$$

dove $\sin^2\phi$ è il quadrato di $\sin\phi$, e analogamente $\cos^2\phi$ è il quadrato di $\cos\phi$.

Osservazione 4.61 In particolare, otteniamo

$$\sin\phi = \pm\sqrt{1 - \cos^2\phi} , \quad \cos\phi = \pm\sqrt{1 - \sin^2\phi} .$$

Attenzione, però: il segno davanti alle radici quadrate dipende da ϕ , in quanto deve coincidere con il segno di $\sin\phi$ (o $\cos\phi$).

Osservazione 4.62 La relazione (4.45), assieme a (4.39) e alle proprietà dei limiti, ci permette di calcolare due altri limiti importanti:

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{1 - \cos x}{x} &= \lim_{x \rightarrow 0} \left(\frac{1 - \cos^2 x}{x} \cdot \frac{1}{1 + \cos x} \right) = \left(\lim_{x \rightarrow 0} \frac{\sin^2 x}{x} \right) \cdot \left(\lim_{x \rightarrow 0} \frac{1}{1 + \cos x} \right) \\ &= \left(\lim_{x \rightarrow 0} \sin x \right) \cdot \left(\lim_{x \rightarrow 0} \frac{\sin x}{x} \right) \cdot \left(\lim_{x \rightarrow 0} \frac{1}{1 + \cos x} \right) = 0 \cdot 1 \cdot \frac{1}{2} = 0 , \end{aligned}$$

e

$$\begin{aligned}\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} &= \lim_{x \rightarrow 0} \left(\frac{1 - \cos^2 x}{x^2} \cdot \frac{1}{1 + \cos x} \right) = \left(\lim_{x \rightarrow 0} \frac{\sin^2 x}{x^2} \right) \cdot \left(\lim_{x \rightarrow 0} \frac{1}{1 + \cos x} \right) \\ &= \left(\lim_{x \rightarrow 0} \frac{\sin x}{x} \right)^2 \cdot \left(\lim_{x \rightarrow 0} \frac{1}{1 + \cos x} \right) = 1^2 \cdot \frac{1}{2} = \frac{1}{2},\end{aligned}$$

cioè $1 - \cos x \approx \frac{1}{2}x^2$ per $x \approx 0$.

Osservazione 4.63 Usando (4.45) possiamo ricavare il valore di seno e coseno di altri angoli speciali. Per esempio, il punto $P(\pi/4)$ appartiene alla bisettrice del primo quadrante; quindi $\sin(\pi/4) = \cos(\pi/4) > 0$. Ma allora (4.45) implica $2\sin^2(\pi/4) = 1$, per cui

$$\sin \frac{\pi}{4} = \cos \frac{\pi}{4} = \frac{\sqrt{2}}{2}.$$

L'angolo di $\pi/3$ radianti (cioè 60°) è l'angolo interno di un triangolo equilatero. Quindi l'origine, $P(\pi/3)$ e $(1, 0)$ formano un triangolo equilatero di lato 1; siccome nei triangoli equilateri l'altezza coincide con la mediana deduciamo

$$\cos \frac{\pi}{3} = \frac{1}{2}, \quad \sin \frac{\pi}{3} = \sqrt{1 - \frac{1}{2^2}} = \frac{\sqrt{3}}{2}.$$

Usando (4.40) e (4.42) otteniamo anche

$$\cos \frac{\pi}{6} = \cos \left(-\frac{\pi}{6} \right) = \sin \left(-\frac{\pi}{6} + \frac{\pi}{2} \right) = \sin \frac{\pi}{3} = \frac{\sqrt{3}}{2}, \quad \sin \frac{\pi}{6} = \frac{1}{2}.$$

CURIOSITÀ 4.15 Un altro modo per dedurre che $\sin \pi/4 = \cos \pi/4$ è il seguente:

$$\sin \frac{\pi}{4} = -\cos \left(\frac{\pi}{4} + \frac{\pi}{2} \right) = -\cos \frac{3\pi}{4} = -\cos \left(-\frac{\pi}{4} + \pi \right) = \cos \left(-\frac{\pi}{4} \right) = \cos \frac{\pi}{4}.$$

Un'altra formula molto importante è quella che permette di calcolare il coseno della differenza di due angoli:

$$\cos(\phi - \psi) = \cos \phi \cos \psi + \sin \phi \sin \psi. \quad (4.46)$$

La Curiosità 4.16 mostra come ricavare questa formula; ma prima vediamo come usarla per calcolare il coseno della somma di due angoli, e il seno della somma e della differenza di due angoli. Prima di tutto usando (4.40) otteniamo

$$\begin{aligned}\cos(\phi + \psi) &= \cos(\phi - (-\psi)) = \cos \phi \cos(-\psi) + \sin \phi \sin(-\psi) \\ &= \cos \phi \cos \psi - \sin \phi \sin \psi.\end{aligned} \quad (4.47)$$

Poi (4.43) e (4.44) ci danno

$$\begin{aligned}\sin(\phi + \psi) &= \cos \left(\phi + \psi - \frac{\pi}{2} \right) = \cos \phi \cos \left(\psi - \frac{\pi}{2} \right) - \sin \phi \sin \left(\psi - \frac{\pi}{2} \right) \\ &= \cos \phi \sin \psi + \sin \phi \cos \psi.\end{aligned} \quad (4.48)$$

Analogamente (esercizio),

$$\sin(\phi - \psi) = \sin \phi \cos \psi - \cos \phi \sin \psi .$$

Ponendo $\psi = \phi$ in (4.47) otteniamo anche

$$\cos(2\phi) = \cos^2 \phi - \sin^2 \phi = 1 - 2 \sin^2 \phi = 2 \cos^2 \phi - 1 ,$$

dove abbiamo usato (4.45). Analogamente, (4.48) dà

$$\sin(2\phi) = 2 \sin \phi \cos \phi .$$

Sommando $\cos(\phi + \psi)$ e $\cos(\phi - \psi)$ otteniamo invece

$$\cos(\phi + \psi) + \cos(\phi - \psi) = 2 \cos \phi \cos \psi .$$

Ponendo $\alpha = \phi + \psi$ e $\beta = \phi - \psi$, in modo che $\phi = (\alpha + \beta)/2$ e $\psi = (\alpha - \beta)/2$, ricaviamo quindi la prima *formula di prostaferesi* per la somma di coseni:

$$\cos \alpha + \cos \beta = 2 \cos \frac{\alpha + \beta}{2} \cos \frac{\alpha - \beta}{2} .$$

In maniera analoga (esercizio) si ricavano le formule di prostaferesi per la differenza di coseni, e per la somma e differenza di seni:

$$\begin{aligned} \cos \alpha - \cos \beta &= -2 \sin \frac{\alpha + \beta}{2} \sin \frac{\alpha - \beta}{2} , \\ \sin \alpha + \sin \beta &= 2 \sin \frac{\alpha + \beta}{2} \cos \frac{\alpha - \beta}{2} , \\ \sin \alpha - \sin \beta &= 2 \cos \frac{\alpha + \beta}{2} \sin \frac{\alpha - \beta}{2} . \end{aligned}$$

CURIOSITÀ 4.16 Vediamo come si dimostra la formula (4.46) per il coseno della differenza.

Consideriamo i due triangoli rettangoli $P(\psi)AP(\phi)$ e $P(\psi)BP(\phi)$ rappresentati nella Figura 4.28. Nel triangolo $P(\psi)AP(\phi)$, il cateto $AP(\phi)$ è lungo $|\sin \phi - \sin \psi|$, e il cateto $AP(\psi)$ è lungo $|\cos \phi - \cos \psi|$. Se indichiamo con d la lunghezza della corda $P(\psi)P(\phi)$, il Teorema di Pitagora ci dice che

$$d^2 = (\cos \phi - \cos \psi)^2 + (\sin \phi - \sin \psi)^2 .$$

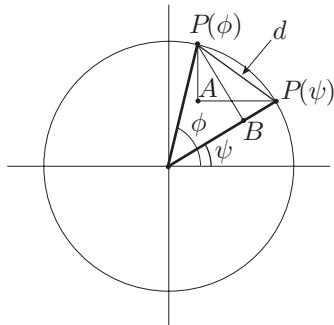
D'altra parte, nel triangolo $P(\psi)BP(\phi)$, il cateto $BP(\phi)$ è lungo $|\sin(\phi - \psi)|$, e il cateto $BP(\psi)$ è lungo $|1 - \cos(\phi - \psi)|$. Una nuova applicazione del Teorema di Pitagora ci dà quindi

$$d^2 = \sin^2(\phi - \psi) + [1 - \cos(\phi - \psi)]^2 .$$

Espandendo i quadrati e confrontando le due espressioni troviamo

$$\begin{aligned} \sin^2 \phi + \cos^2 \phi + \sin^2 \psi + \cos^2 \psi - 2(\cos \phi \cos \psi + \sin \phi \sin \psi) \\ = 1 + \sin^2(\phi - \psi) + \cos^2(\phi - \psi) - 2 \cos(\phi - \psi) , \end{aligned}$$

e (4.46) segue subito da (4.45).

**Figura 4.28** Coseno della differenza.

Osservazione 4.64 Prendiamo un punto $P = (x, y) \neq (0, 0)$ del piano, a distanza r dall’origine. Indichiamo con $\phi \in [0, 2\pi)$ l’angolo fra l’asse positivo delle ascisse e la semiretta uscente dall’origine passante per P . Allora usando i triangoli simili evidenti nella Figura 4.29 otteniamo subito

$$x = r \cos \phi , \quad y = r \sin \phi . \quad (4.49)$$

Quindi conoscendo la distanza r e l’angolo ϕ è possibile ritrovare le coordinate cartesiane del punto P ; la coppia (r, ϕ) determina il punto $P \neq (0, 0)$ altrettanto bene della coppia (x, y) . Per questo motivo, (r, ϕ) vengono dette *coordinate polari* del punto P ; sono definite per tutti i punti del piano tranne il *polo* $(0, 0)$, che è a distanza 0 dall’origine ma al quale non è possibile associare un angolo.

La (4.49) ci dice come recuperare le coordinate cartesiane (x, y) di un punto conoscendone le coordinate polari (r, ϕ) . Viceversa, il Teorema di Pitagora (vedi nuovamente la Figura 4.29) ci dice che

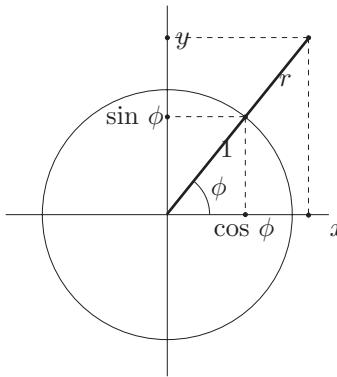
$$r^2 = x^2 + y^2 ,$$

per cui $r = \sqrt{x^2 + y^2}$. Una volta trovato r , l’angolo ϕ è l’unico angolo nell’intervallo $[0, 2\pi)$ tale che

$$\cos \phi = \frac{x}{r} = \frac{x}{\sqrt{x^2 + y^2}} \quad \text{e} \quad \sin \phi = \frac{y}{r} = \frac{y}{\sqrt{x^2 + y^2}} ,$$

per cui siamo in grado di recuperare le coordinate polari a partire dalle coordinate cartesiane. *Attenzione:* non è sufficiente conoscere solo $\cos \phi$ o solo $\sin \phi$ per ricavare ϕ . Infatti, come puoi verificare facilmente, per ogni $\phi \in [0, 2\pi)$ esistono sempre ϕ' , $\phi'' \in [0, 2\pi)$ diversi da ϕ e tali che $\cos \phi' = \cos \phi$ e $\sin \phi'' = \sin \phi$.

ESEMPIO 4.30 Alcuni animali utilizzano sistematicamente le coordinate polari; per esempio, le api. Quando un’ape operaia vuole comunicare alle altre dove si trovano dei fiori rispetto all’alveare, inizia una danza così fatta: si muove inizialmente in linea retta, poi si gira e ritorna al punto di partenza in modo da formare un angolo ϕ con la direzione iniziale, e ripete questo percorso un numero preciso di volte.

**Figura 4.29** Coordinate polari.

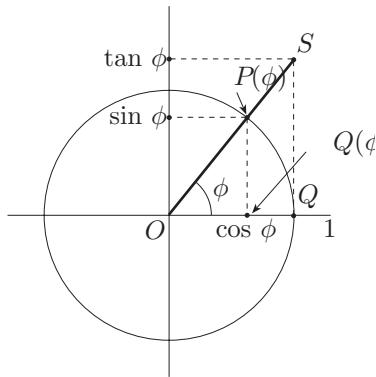
L'angolo ϕ indica la coordinata polare angolare della posizione dei fiori, misurata rispetto alla direzione del sole; e il numero di ripetizioni indica la distanza r dei fiori dall'alveare. In questo modo le altre api capiscono in che direzione e a quale distanza devono volare per trovare i fiori.

Seno e coseno sono le più importanti funzioni trigonometriche (cioè dipendenti da un angolo), ma non le uniche. Un'altra funzione trigonometrica utile è la *tangente* definita da

$$\tan \phi = \frac{\sin \phi}{\cos \phi} ;$$

a volte si scrive $\operatorname{tg} \phi$ invece di $\tan \phi$.

Osservazione 4.65 Il motivo del nome è illustrato nella Figura 4.30: se $\phi \in [0, \pi/2)$, allora il segmento QS tangente in $(1, 0)$ alla solita circonferenza di raggio unitario e centro l'origine, e intersecante la semiretta passante per $P(\phi)$, ha lunghezza esattamente uguale a $\tan \phi$, come si verifica subito usando i triangoli simili.

**Figura 4.30** Tangente.

Dalla definizione e dalle proprietà di seno e coseno si deducono subito le seguenti proprietà della tangente:

- la tangente non è definita dove il coseno si annulla, cioè nei punti $\pi/2 + k\pi$ per $k \in \mathbb{Z}$, che sono delle singolarità per $\tan \phi$;
- $\tan \phi$ è periodica di periodo π , cioè $\tan(\phi + k\pi) = \tan \phi$ per ogni $k \in \mathbb{Z}$;
- $\tan \phi$ è strettamente crescente in ciascun intervallo $(-\pi/2 + k\pi, \pi/2 + k\pi)$, per $k \in \mathbb{Z}$;
- $\tan \phi$ è positiva negli intervalli $(k\pi, \pi/2 + k\pi)$ e negativa in $(-\pi/2 + k\pi, k\pi)$, per $k \in \mathbb{Z}$;
- il grafico di $\tan \phi$ ha un asintoto verticale nelle singolarità, e

$$\forall k \in \mathbb{Z} \quad \lim_{x \rightarrow (\pi/2+k\pi)^{\pm}} \tan \phi = \mp\infty ;$$

- l'immagine di $\tan \phi$ ristretta a ciascun intervallo $(-\pi/2 + k\pi, \pi/2 + k\pi)$ è l'intera retta reale \mathbb{R} .

La Figura 4.31 riporta a sinistra il grafico di $\tan \phi$.

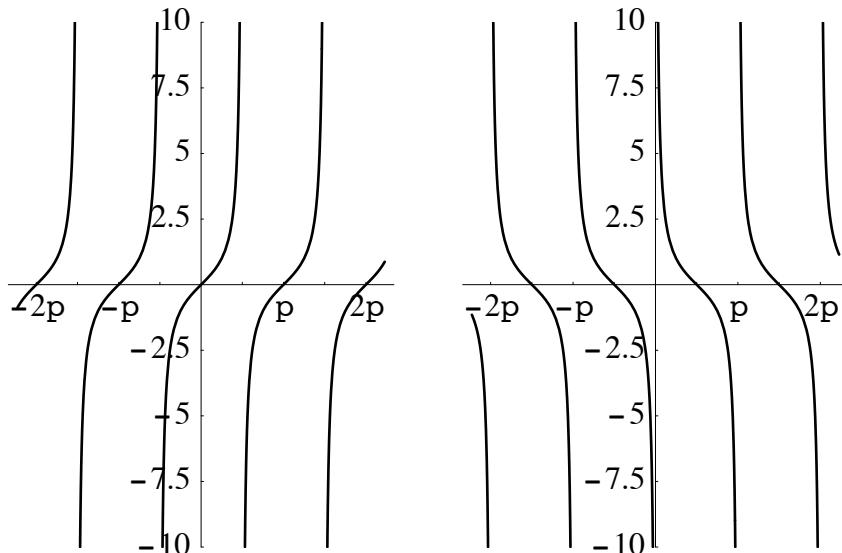


Figura 4.31 Tangente e cotangente.

La cotangente $\cot \phi$ (talvolta indicata anche con $\operatorname{ctg} \phi$ o $\operatorname{cotan} \phi$) è il reciproco della tangente:

$$\cot \phi = \frac{1}{\tan \phi} = \frac{\cos \phi}{\sin \phi} .$$

Le proprietà della tangente implicano le seguenti proprietà della cotangente:

- la cotangente non è definita dove il seno si annulla, cioè nei punti $k\pi$ per $k \in \mathbb{Z}$, che sono delle singolarità per $\cot \phi$;

- $\cot \phi$ è periodica di periodo π , cioè $\cot(\phi + k\pi) = \cot \phi$ per ogni $k \in \mathbb{Z}$;
- $\cot \phi$ è strettamente decrescente in ciascun intervallo $(k\pi, (k+1)\pi)$, per $k \in \mathbb{Z}$;
- $\cot \phi$ è positiva negli intervalli $(k\pi, \pi/2 + k\pi)$ e negativa in $(-\pi/2 + k\pi, k\pi)$, per $k \in \mathbb{Z}$;
- il grafico di $\cot \phi$ ha un asintoto verticale nelle singolarità, e

$$\forall k \in \mathbb{Z} \quad \lim_{x \rightarrow (k\pi)^{\pm}} \cot \phi = \pm\infty ;$$

- l'immagine di $\cot \phi$ ristretta a ciascun intervallo $(k\pi, (k+1)\pi)$ è l'intera retta reale \mathbb{R} .

La Figura 4.31 riporta a destra il grafico di $\cot \phi$.

Osservazione 4.66 Sia $f(x) = mx$ una funzione lineare (con $d = 0$) e $P = (x, y)$ un punto del suo grafico; in particolare, $y = f(x) = mx$. Se (r, ϕ) sono le coordinate polari di P , allora

$$m = \frac{y}{x} = \frac{r \sin \phi}{r \cos \phi} = \tan \phi ,$$

cioè m è la tangente dell'angolo fra l'asse delle ascisse e la retta grafico di f . Per questo motivo, m è chiamato *coefficiente angolare* della retta.

CURIOSITÀ 4.17 Vediamo come si può dimostrare che $\lim_{\phi \rightarrow 0} \frac{\sin \phi}{\phi} = 1$. Consideriamo per semplicità il caso $\phi \rightarrow 0^+$, il caso $\phi \rightarrow 0^-$ essendo analogo. Facendo riferimento alla Figura 4.30, vogliamo confrontare le aree del triangolo $OQ(\phi)P(\phi)$, del settore circolare $OQP(\phi)$, e del triangolo OQS . Con le ovvie notazioni, abbiamo

$$A_{OQ(\phi)P(\phi)} < A_{OQP(\phi)} < A_{OQS} ,$$

in quanto i tre insiemi sono contenuti uno dentro l'altro. Ora, i due cateti di $OQ(\phi)P(\phi)$ misurano rispettivamente $\cos \phi$ e $\sin \phi$; quindi $A_{OQ(\phi)P(\phi)} = \frac{1}{2} \cos \phi \sin \phi$. Analogamente, i due cateti di OQS misurano rispettivamente 1 e $\tan \phi$; quindi $A_{OQS} = \frac{1}{2} \sin \phi / \cos \phi$. Infine, l'area di un settore circolare è proporzionale alla lunghezza dell'arco. Siccome l'area del cerchio di raggio 1, sotteso da un arco di lunghezza 2π , è uguale a π , l'area del settore circolare $OQP(\phi)$, sotteso da un arco di lunghezza ϕ , è $A_{OQS} = \phi\pi/2\pi = \phi/2$. Riassumendo abbiamo ottenuto

$$\frac{1}{2} \sin \phi \cos \phi < \frac{\phi}{2} < \frac{1}{2} \frac{\sin \phi}{\cos \phi} .$$

Dividendo per $\sin \phi$ (che è strettamente positivo per $\phi > 0$ piccolo e non nullo), moltiplicando per 2, e prendendo il reciproco otteniamo

$$\cos \phi < \frac{\sin \phi}{\phi} < \frac{1}{\cos \phi} .$$

A questo punto, facendo tendere ϕ a 0^+ , ricordando che $\cos 0 = 1$ e usando il Teorema dei due carabinieri, otteniamo

$$\lim_{\phi \rightarrow 0} \frac{\sin \phi}{\phi} = 1 .$$

Le funzioni periodiche non possono essere globalmente invertibili, in quanto non sono iniettive. Però, restringendole a opportuni intervalli, si possono trovare delle inverse.

Per esempio, abbiamo visto che il seno è strettamente crescente sull'intervallo $[-\pi/2, \pi/2]$, con immagine uguale a tutto l'intervallo $[-1, 1]$. Siccome le funzioni strettamente crescenti sono invertibili, possiamo definire la funzione inversa *arcoseno* $\arcsin: [-1, 1] \rightarrow [-\pi/2, \pi/2]$; $\arcsin x$ è l'unico angolo nell'intervallo $[-\pi/2, \pi/2]$ il cui seno sia uguale a x . In altre parole, $\arcsin x$ è l'unica soluzione nell'intervallo $[-\pi/2, \pi/2]$ dell'equazione $\sin \phi = x$.

Osservazione 4.67 Attenzione: ti dev'essere ben chiaro che $\arcsin x$ non è l'unica soluzione dell'equazione $\sin \phi = x$; è l'unica soluzione che appartiene all'intervallo $[-\pi/2, \pi/2]$. Per periodicità, tutti gli angoli della forma $\arcsin x + 2k\pi$ con $k \in \mathbb{Z}$ sono soluzioni della stessa equazione, ma ce ne sono anche altre. Per esempio, l'equazione $\sin \phi = 1/2$ ha due soluzioni nell'intervallo $[0, \pi]$: la soluzione $\arcsin x = \pi/4$, e la soluzione $3\pi/4$. Il modo più semplice per capire cosa succede è intersecare il grafico di $\sin x$ con una retta orizzontale, come discusso nel Capitolo 3.

Analogamente, abbiamo visto che il coseno è strettamente decrescente nell'intervallo $[0, \pi]$. Quindi possiamo definire la funzione inversa, chiamata *arcocoseno* $\arccos: [-1, 1] \rightarrow [0, \pi]$; $\arccos x$ è l'unico angolo nell'intervallo $[0, \pi]$ il cui coseno sia uguale a x . In altre parole, $\arccos x$ è l'unica soluzione nell'intervallo $[0, \pi]$ dell'equazione $\cos \phi = x$.

Osservazione 4.68 Anche per l'arcocoseno valgono le stesse avvertenze fatte per l'arcoseno. Inoltre, vale la pena di sottolineare che arcoseno e arcocoseno hanno immagine in intervalli diversi: il seno non è iniettivo nell'intervallo $[0, \pi]$, per cui non possiamo definire un'inversa a valori in $[0, \pi]$. Analogamente, il coseno non è invertibile nell'intervallo $[-\pi/2, \pi/2]$. La Figura 4.32 contiene i grafici di arcoseno e arcocoseno.

Più interessante è l'inversa della tangente. Abbiamo visto che la tangente è strettamente crescente nell'intervallo $(-\pi/2, \pi/2)$, con immagine tutta la retta reale. Quindi esiste la funzione inversa *arcotangente* $\arctan: \mathbb{R} \rightarrow (-\pi/2, \pi/2)$; $\arctan x$ è l'unico angolo nell'intervallo $(-\pi/2, \pi/2)$ la cui tangente sia uguale a x . In altre parole, $\arctan x$ (a volte scritta “atg” o “atan”) è l'unica soluzione nell'intervallo $(-\pi/2, \pi/2)$ dell'equazione $\tan \phi = x$.

L'arcotangente (contrariamente ad arcoseno e arcocoseno) ha il grosso vantaggio di essere definita su tutto l'asse reale. Inoltre dalle proprietà della tangente si deduce subito che $\arctan x$ è strettamente crescente, negativa per $x < 0$, positiva per $x > 0$, e

$$\lim_{x \rightarrow \pm\infty} \arctan x = \pm \frac{\pi}{2};$$

in particolare, l'arcotangente ha un asintoto orizzontale sia a $+\infty$ che a $-\infty$. La Figura 4.33 contiene il grafico dell'arcotangente.

Vediamo ora come trovare una funzione sinusoidale con caratteristiche predefinite, cominciando con un esempio.

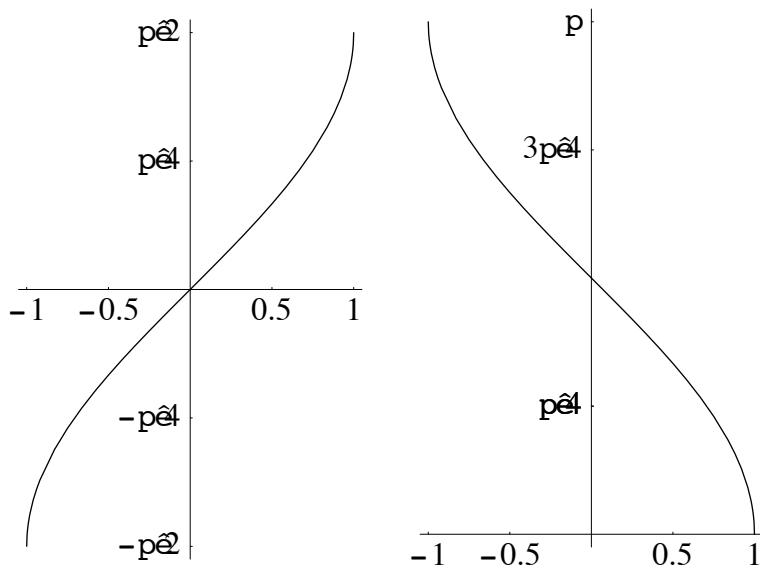


Figura 4.32 Arcoseno e arcocoseno.

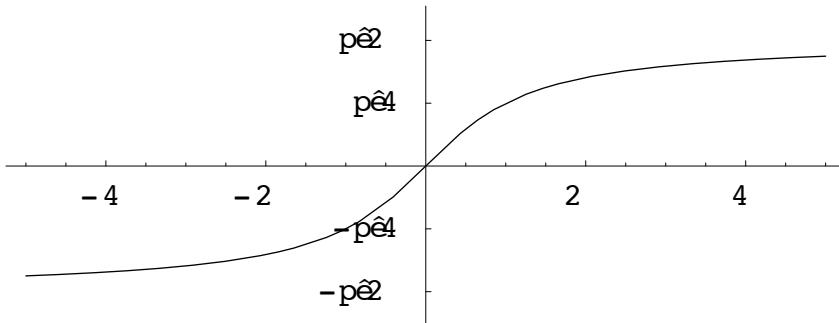


Figura 4.33 Arcotangente.

ESEMPIO 4.31 Molti fenomeni biologici hanno un andamento periodico; per esempio quelli legati alle stagioni. Il tuo assistente ha tenuto sotto osservazione una popolazione di uccelli stanziali nella zona del lago di Massaciuccoli, notando che il numero di individui nella popolazione varia in modo periodico nel corso dell'anno da (circa) 1000 individui a (circa) 1500 individui. Il minimo numero di uccelli è misurato intorno all'inizio di aprile, il massimo circa 6 mesi dopo. Vogliamo trovare una funzione sinusoidale f che rappresenti questo andamento.

Vediamo come modificare la funzione $f_0(x) = \cos x$ in modo da ottenere la funzione voluta. La prima osservazione da fare è che mentre f_0 ha periodo 2π , la funzione f che stiamo cercando deve avere periodo 365 giorni (misuriamo il tempo in giorni, e trascuriamo per semplicità gli anni bisestili). Dunque mentre f_0 è completamente descritta da come si comporta sull'intervallo $[0, 2\pi]$, la

nostra funzione f dev'essere completamente descritta da come si comporta sull'intervallo $[0, 365]$. Ora, se moltiplichiamo le ascisse per $2\pi/365$ trasformiamo l'intervallo $[0, 365]$ nell'intervallo $[0, 2\pi]$; quindi potrebbe essere utile considerare la funzione

$$f_1(x) = f_0\left(\frac{2\pi}{365}x\right) = \cos\left(\frac{2\pi}{365}x\right).$$

La funzione $f_1: \mathbb{R} \rightarrow \mathbb{R}$ ha effettivamente periodo 365: infatti,

$$f_1(x + 365) = \cos\left(\frac{2\pi}{365}(x + 365)\right) = \cos\left(\frac{2\pi}{365}x + 2\pi\right) = f_1(x).$$

Il periodo è sistemato; passiamo alla fase. La funzione f_1 assume il primo massimo nello stesso punto in cui l'assume il coseno, cioè in $x_0 = 0$; invece la nostra funzione deve assumere il primo massimo all'inizio di ottobre, cioè in $x_0 = 274$. Dunque dobbiamo traslare il grafico verso destra di 274 giorni; come abbiamo visto più volte, questo equivale a sottrarre 274 alle ascisse. Quindi giungiamo alla funzione

$$f_2(x) = f_1(x - 274) = \cos\left(\frac{2\pi}{365}(x - 274)\right),$$

che ha periodo e fase giusti. La funzione f_2 ha valore minimo $m = -1$ e valore massimo $M = 1$, che corrispondono a un'ampiezza $(M - m)/2 = 1$ e a un valor medio $(M + m)/2 = 0$. La nostra funzione f deve invece avere valore minimo 1000 e valore massimo 1500, che corrispondono a un'ampiezza $A = (1500 - 1000)/2 = 250$ e a un valor medio $(1500 + 1000)/2 = 1250$. Per modificare l'ampiezza senza cambiare valor medio, fase e periodo, è sufficiente moltiplicare le ordinate; in particolare la funzione

$$f_3(x) = 250f_2(x) = 250 \cos\left(\frac{2\pi}{365}(x - 274)\right)$$

ha ampiezza 250 (e periodo e fase giusti). Infine, per modificare il valor medio senza cambiare ampiezza, fase e periodo basta traslare verticalmente il grafico, per cui arriviamo infine alla funzione

$$f(x) = f_3(x) + 1250 = 250 \cos\left(\frac{2\pi}{365}(x - 274)\right) + 1250.$$

Questa funzione ha periodo 365, e assume massimo in $x_0 = 274$, dove vale proprio $250 + 1250 = 1500$. Inoltre, ha valore minimo (uguale a $-250 + 1250 = 100$) nei punti x_1 tali che

$$\cos\left(\frac{2\pi}{365}(x_1 - 274)\right) = -1,$$

cioè per $(2\pi/365)(x_1 - 274) = (2k + 1)\pi$, con $k \in \mathbb{Z}$, cioè $x_1 = 274 + (365/2) + 365k$ con $k \in \mathbb{Z}$. In particolare, per trovare un valore di x_1 nel primo anno (cioè con $0 \leq x_1 \leq 365$) prendiamo $k = -1$ ottenendo $x_1 = 274 - (365/2) = 91.5$, che è proprio all'inizio di aprile. Dunque la funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ soddisfa tutte le condizioni che avevamo posto. Ne troverai il grafico nella Figura 4.34.

**Figura 4.34** .

Vediamo come adattare il metodo del precedente esempio per ottenere una funzione sinusoidale con periodo P , fase x_0 , ampiezza A e valor medio y_0 . Abbiamo visto che la funzione $\cos(2\pi x/365)$ ha periodo 365; in maniera assolutamente analoga si verifica che la funzione

$$f_1(x) = \cos\left(\frac{2\pi}{P}x\right)$$

ha periodo P . La quantità $\omega = 2\pi/P$ è detta *frequenza angolare* della funzione sinusoidale di periodo P (mentre $f = 1/P$ viene detto *frequenza* della stessa funzione). Spesso si usa esplicitare la frequenza angolare piuttosto che il periodo, per cui si usa scrivere $f_1(x) = \cos(\omega x)$ e ricordarsi che il periodo vale

$$P = \frac{2\pi}{\omega}.$$

Per sistemare la fase, basta traslare le ascisse; quindi la funzione

$$f_2(x) = \cos(\omega(x - x_0))$$

ha fase x_0 e frequenza angolare ω . Infine, per ottenere ampiezza A e valor medio y_0 basta moltiplicare le ordinate per A e poi sommare y_0 ; giungiamo quindi alla funzione sinusoidale generica

$$f(x) = A \cos(\omega(x - x_0)) + y_0. \quad (4.50)$$

Osservazione 4.69 Sommando funzioni sinusoidali di uguali ampiezza e periodo ma differente fase (e valor medio nullo per semplicità) possono avvenire fenomeni curiosi. Usando le formule di prostaferesi otteniamo

$$A \cos(\omega(x - x_0)) + A \cos(\omega(x - x'_0)) = 2A \cos\left(\omega \frac{x'_0 - x_0}{2}\right) \cos\left[\omega\left(x - \frac{x_0 + x'_0}{2}\right)\right].$$

In particolare, se $\omega(x'_0 - x_0)/2 = \pi/2$, cioè $x'_0 - x_0 = \pi/\omega = P/2$ allora la somma si annulla! Questo accade per esempio con la luce: illuminando lo stesso punto con due raggi di luce della stessa intensità (cioè con la stessa ampiezza) e uguale colore (cioè con la stessa frequenza $f = 1/P$), ma con fasi che differiscono di $P/2$, otteniamo il buio! Questo tipo di fenomeni cade sotto il nome di fenomeni di *interferenza*. Un altro esempio di fenomeno di interferenza sono i colori delle penne di un pavone: la struttura fine delle penne cambia in modo diverso da punto a punto la fase della luce che viene riflessa, con il risultato che vediamo colori diversi a seconda del punto d'osservazione che scegliamo.

Non tutti i fenomeni periodici si possono rappresentare con funzioni sinusoidali; per esempio, la somma di due funzioni sinusoidali con uguale periodo ma differente ampiezza (e/o fase) potrebbe non essere una funzione sinusoidale (vedi per esempio la Figura 4.35). Si può però procedere in maniera un poco diversa. Per spiegare come, cominciamo con l'usare (4.49) per espandere il coseno della differenza in (4.50): otteniamo

$$\begin{aligned} A \cos(\omega(x - x_0)) + y_0 &= A \cos(\omega x_0) \cos(\omega x) + A \sin(\omega x_0) \sin(\omega x) + y_0 \\ &= y_0 + a_1 \cos \omega x + b_1 \sin \omega x , \end{aligned}$$

dove abbiamo posto $a_1 = A \cos \omega x_0$ e $b_1 = A \sin \omega x_0$. Più in generale, possiamo considerare i *polinomi trigonometrici*, cioè funzioni della forma

$$\begin{aligned} p(x) &= y_0 + a_1 \cos \omega x + a_2 \cos(2\omega x) + \cdots + a_n \cos(n\omega x) \\ &\quad + b_1 \sin \omega x + b_2 \sin(2\omega x) + \cdots + b_n \sin(n\omega x) , \end{aligned}$$

con $y_0, a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R}$. Sono tutte funzioni di periodo $2\pi/\omega$ (controlla), ma non sono funzioni sinusoidali; vedi per esempio la Figura 4.35.

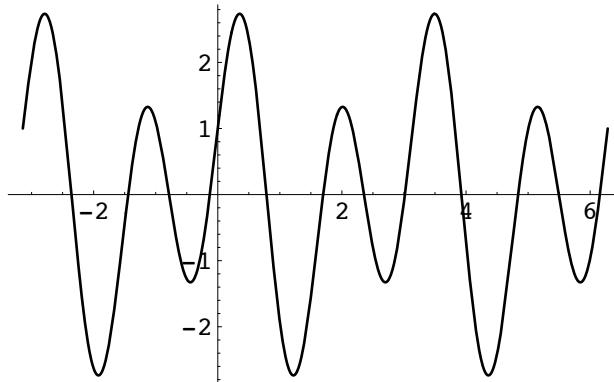


Figura 4.35 $f(x) = \cos(2x) + 2 \sin(4x)$.

L'importanza dei polinomi trigonometrici è dovuta al fatto che *ogni funzione*

periodica¹⁷ di frequenza angolare ω è approssimabile arbitrariamente bene da polinomi trigonometrici di frequenza angolare ω . In altre parole, se $f: \mathbb{R} \rightarrow \mathbb{R}$ è una funzione periodica di frequenza angolare ω , e $\varepsilon > 0$ è un qualsiasi numero (arbitrariamente piccolo), esiste sempre un polinomio trigonometrico p di frequenza angolare ω tale che $|f(x) - p(x)| < \varepsilon$ per ogni $x \in \mathbb{R}$. La tecnica per trovare il polinomio trigonometrico data la funzione si chiama *analisi di Fourier* (o *analisi armonica*), ed è sfortunatamente troppo complessa per essere presentata in questo corso.

ESEMPIO 4.32 Se hai studiato un po' di musica, saprai che quando suoni una corda di una chitarra (per esempio il La) il suono che ottieni non è composto solo dalla frequenza corrispondente al La fondamentale, ma anche dai cosiddetti armonici. Ebbene: il La puro è rappresentato da una sinusoide, mentre gli armonici corrispondono esattamente ai termini successivi della rappresentazione del suono tramite un polinomio trigonometrico. I termini di frequenza angolare 2ω corrispondono al La dell'ottava superiore, i termini di frequenza angolare 3ω forniscono la quinta (cioè il Mi), quelli di frequenza angolare 4ω il La due ottave sopra, e così via.

¹⁷ Che ti capiterà di incontrare; ce ne sono alcune orribili per cui non è vero, ma non intervengono nella pratica sperimentale (almeno per ora).