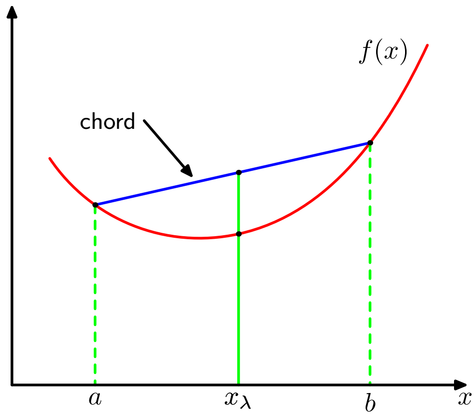


**Figure 2.15** A convex function $f(x)$ is one for which every chord (shown in blue) lies on or above the function (shown in red).

is given by $\lambda f(a) + (1 - \lambda)f(b)$, and the corresponding value of the function is $f(\lambda a + (1 - \lambda)b)$. Convexity then implies

$$f(\lambda a + (1 - \lambda)b) \leqslant \lambda f(a) + (1 - \lambda)f(b). \tag{2.101}$$

*Exercise 2.32*

This is equivalent to the requirement that the second derivative of the function be everywhere positive. Examples of convex functions are $x \ln x$ (for $x > 0$) and $x^2$. A function is called *strictly convex* if the equality is satisfied only for $\lambda = 0$ and $\lambda = 1$. If a function has the opposite property, namely that every chord lies on or below the function, it is called *concave*, with a corresponding definition for *strictly concave*. If a function $f(x)$ is convex, then $-f(x)$ will be concave.

*Exercise 2.33*

Using the technique of proof by induction, we can show from (2.101) that a convex function $f(x)$ satisfies

$$f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \leqslant \sum_{i=1}^{M} \lambda_i f(x_i) \tag{2.102}$$

where $\lambda_i \geqslant 0$ and $\sum_i \lambda_i = 1$, for any set of points $\{x_i\}$. The result (2.102) is known as *Jensen's inequality*. If we interpret the $\lambda_i$ as the probability distribution over a discrete variable $x$ taking the values $\{x_i\}$, then (2.102) can be written

$$f\left(\mathbb{E}[x]\right) \leqslant \mathbb{E}[f(x)] \tag{2.103}$$

where $\mathbb{E}[\cdot]$ denotes the expectation. For continuous variables, Jensen's inequality takes the form

$$f\left(\int \mathbf{x} p(\mathbf{x}) \,\mathrm{d}\mathbf{x}\right) \leqslant \int f(\mathbf{x}) p(\mathbf{x}) \,\mathrm{d}\mathbf{x}. \tag{2.104}$$

We can apply Jensen's inequality in the form (2.104) to the Kullback–Leibler divergence (2.100) to give

$$\mathrm{KL}(p\|q) = -\int p(\mathbf{x}) \ln\left\{\frac{q(\mathbf{x})}{p(\mathbf{x})}\right\} \mathrm{d}\mathbf{x} \geqslant -\ln \int q(\mathbf{x}) \,\mathrm{d}\mathbf{x} = 0 \tag{2.105}$$