

which satisfies $P'(x) = p(x)$, as shown in [Figure 2.6](#).

If we have several continuous variables x_1, \dots, x_D , denoted collectively by the vector \mathbf{x} , then we can define a joint probability density $p(\mathbf{x}) = p(x_1, \dots, x_D)$ such that the probability of \mathbf{x} falling in an infinitesimal volume $\delta\mathbf{x}$ containing the point \mathbf{x} is given by $p(\mathbf{x})\delta\mathbf{x}$. This multivariate probability density must satisfy

$$p(\mathbf{x}) \geq 0 \quad (2.27)$$

$$\int p(\mathbf{x}) d\mathbf{x} = 1 \quad (2.28)$$

in which the integral is taken over the whole of \mathbf{x} space. More generally, we can also consider joint probability distributions over a combination of discrete and continuous variables.

The sum and product rules of probability, as well as Bayes' theorem, also apply to probability densities as well as to combinations of discrete and continuous variables. If \mathbf{x} and \mathbf{y} are two real variables, then the sum and product rules take the form

$$\text{sum rule} \quad p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{y}) d\mathbf{y} \quad (2.29)$$

$$\text{product rule} \quad p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \quad (2.30)$$

Similarly, Bayes' theorem can be written in the form

$$p(\mathbf{y}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{y})p(\mathbf{y})}{p(\mathbf{x})} \quad (2.31)$$

where the denominator is given by

$$p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) d\mathbf{y}. \quad (2.32)$$

A formal justification of the sum and product rules for continuous variables requires a branch of mathematics called *measure theory* (Feller, 1966) and lies outside the scope of this book. Its validity can be seen informally, however, by dividing each real variable into intervals of width Δ and considering the discrete probability distribution over these intervals. Taking the limit $\Delta \rightarrow 0$ then turns sums into integrals and gives the desired result.

2.2.1 Example distributions

There are many forms of probability density that are in widespread use and that are important both in their own right and as building blocks for more complex probabilistic models. The simplest form would be one in which $p(x)$ is a constant, independent of x , but this cannot be normalized because the integral in (2.28) will be divergent. Distributions that cannot be normalized are called *improper*. We can, however, have the uniform distribution that is constant over a finite region, say (c, d) , and zero elsewhere, in which case (2.28) implies

$$p(x) = 1/(d - c), \quad x \in (c, d). \quad (2.33)$$