



Figure 3.1: **The normal distribution:** The normal distribution  $\mathcal{N}(x; \mu, \sigma^2)$  exhibits a classic “bell curve” shape, with the  $x$  coordinate of its central peak given by  $\mu$ , and the width of its peak controlled by  $\sigma$ . In this example, we depict the **standard normal distribution**, with  $\mu = 0$  and  $\sigma = 1$ .

First, many distributions we wish to model are truly close to being normal distributions. The **central limit theorem** shows that the sum of many independent random variables is approximately normally distributed. This means that in practice, many complicated systems can be modeled successfully as normally distributed noise, even if the system can be decomposed into parts with more structured behavior.

Second, out of all possible probability distributions with the same variance, the normal distribution encodes the maximum amount of uncertainty over the real numbers. We can thus think of the normal distribution as being the one that inserts the least amount of prior knowledge into a model. Fully developing and justifying this idea requires more mathematical tools, and is postponed to section 19.4.2.

The normal distribution generalizes to  $\mathbb{R}^n$ , in which case it is known as the **multivariate normal distribution**. It may be parametrized with a positive definite symmetric matrix  $\Sigma$ :

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sqrt{\frac{1}{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp \left( -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right). \quad (3.23)$$