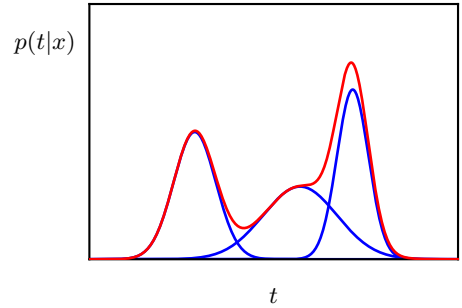


**Figure 3.7** Example of a Gaussian mixture distribution in one dimension showing three Gaussians (each scaled by a coefficient) in blue and their sum in red.



this proves to be the case, as can be seen from Figure 3.6(b). Such superpositions, formed by taking linear combinations of more basic distributions such as Gaussians, can be formulated as probabilistic models known as *mixture distributions*. In this section we will consider Gaussians to illustrate the framework of mixture models. More generally, mixture models can comprise linear combinations of other distributions, for example mixtures of Bernoulli distributions for binary variables. In Figure 3.7 we see that a linear combination of Gaussians can give rise to very complex densities. By using a sufficient number of Gaussians and by adjusting their means and covariances as well as the coefficients in the linear combination, almost any continuous distribution can be approximated to arbitrary accuracy.

We therefore consider a superposition of  $K$  Gaussian densities of the form

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (3.111)$$

which is called a *mixture of Gaussians*. Each Gaussian density  $\mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  is called a *component* of the mixture and has its own mean  $\boldsymbol{\mu}_k$  and covariance  $\boldsymbol{\Sigma}_k$ . Contour and surface plots for a Gaussian mixture in two dimensions having three components are shown in Figure 3.8.

The parameters  $\pi_k$  in (3.111) are called *mixing coefficients*. If we integrate both sides of (3.111) with respect to  $\mathbf{x}$ , and note that both  $p(\mathbf{x})$  and the individual Gaussian components are normalized, we obtain

$$\sum_{k=1}^K \pi_k = 1. \quad (3.112)$$

Also, given that  $\mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \geq 0$ , a sufficient condition for the requirement  $p(\mathbf{x}) \geq 0$  is that  $\pi_k \geq 0$  for all  $k$ . Combining this with the condition (3.112), we obtain

$$0 \leq \pi_k \leq 1. \quad (3.113)$$

We can therefore see that the mixing coefficients satisfy the requirements to be probabilities, and we will show that this probabilistic interpretation of mixture distributions is very powerful.