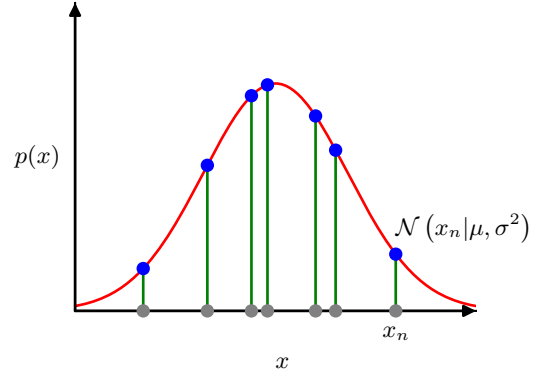


Figure 2.9 Illustration of the likelihood function for the Gaussian distribution shown by the red curve. Here the grey points denote a data set of values $\{x_n\}$, and the likelihood function (2.55) is given by the product of the corresponding values of $p(x)$ denoted by the blue points. Maximizing the likelihood involves adjusting the mean and variance of the Gaussian so as to maximize this product.



set \mathbf{x} is i.i.d., we can therefore write the probability of the data set, given μ and σ^2 , in the form

$$p(\mathbf{x}|\mu, \sigma^2) = \prod_{n=1}^N \mathcal{N}(x_n|\mu, \sigma^2). \quad (2.55)$$

When viewed as a function of μ and σ^2 , this is called the *likelihood function* for the Gaussian and is interpreted diagrammatically in Figure 2.9.

One common approach for determining the parameters in a probability distribution using an observed data set, known as *maximum likelihood*, is to find the parameter values that maximize the likelihood function. This might appear to be a strange criterion because, from our foregoing discussion of probability theory, it would seem more natural to maximize the probability of the parameters given the data, not the probability of the data given the parameters. In fact, these two criteria are related.

Section 2.6.2

To start with, however, we will determine values for the unknown parameters μ and σ^2 in the Gaussian by maximizing the likelihood function (2.55). In practice, it is more convenient to maximize the log of the likelihood function. Because the logarithm is a monotonically increasing function of its argument, maximizing the log of a function is equivalent to maximizing the function itself. Taking the log not only simplifies the subsequent mathematical analysis, but it also helps numerically because the product of a large number of small probabilities can easily underflow the numerical precision of the computer, and this is resolved by computing the sum of the log probabilities instead. From (2.49) and (2.55), the log likelihood function can be written in the form

$$\ln p(\mathbf{x}|\mu, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{n=1}^N (x_n - \mu)^2 - \frac{N}{2} \ln \sigma^2 - \frac{N}{2} \ln(2\pi). \quad (2.56)$$

Maximizing (2.56) with respect to μ , we obtain the maximum likelihood solution given by

Exercise 2.15

$$\mu_{\text{ML}} = \frac{1}{N} \sum_{n=1}^N x_n, \quad (2.57)$$