which is the *sample mean*, i.e., the mean of the observed values $\{x_n\}$. Similarly, maximizing (2.56) with respect to σ^2, we obtain the maximum likelihood solution for the variance in the form

$$\sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{\mathrm{ML}})^2, \tag{2.58}$$

which is the *sample variance* measured with respect to the sample mean μ_{ML}. Note that we are performing a joint maximization of (2.56) with respect to μ and σ^2, but for a Gaussian distribution, the solution for μ decouples from that for σ^2 so that we can first evaluate (2.57) and then subsequently use this result to evaluate (2.58).

2.3.3 Bias of maximum likelihood

The technique of maximum likelihood is widely used in deep learning and forms the foundation for most machine learning algorithms. However, it has some limitations, which we can illustrate using a univariate Gaussian.

We first note that the maximum likelihood solutions μ_{ML} and σ^2_{ML} are functions of the data set values $x_1, \ldots, x_N$. Suppose that each of these values has been generated independently from a Gaussian distribution whose true parameters are μ and σ^2. Now consider the expectations of μ_{ML} and σ^2_{ML} with respect to these data set values. It is straightforward to show that

Exercise 2.16

$$\mathbb{E}[\mu_{\mathrm{ML}}] = \mu \tag{2.59}$$

$$\mathbb{E}[\sigma^2_{\mathrm{ML}}] = \left(\frac{N-1}{N}\right)\sigma^2. \tag{2.60}$$

We see that, when averaged over data sets of a given size, the maximum likelihood solution for the mean will equal the true mean. However, the maximum likelihood estimate of the variance will underestimate the true variance by a factor $(N-1)/N$. This is an example of a phenomenon called *bias* in which the estimator of a random quantity is systematically different from the true value. The intuition behind this result is given by Figure 2.10.

Note that bias arises because the variance is measured relative to the maximum likelihood estimate of the mean, which itself is tuned to the data. Suppose instead we had access to the true mean μ and we used this to determine the variance using the estimator

$$\widehat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)^2. \tag{2.61}$$

Exercise 2.17

Then we find that

$$\mathbb{E}\left[\widehat{\sigma}^2\right] = \sigma^2, \tag{2.62}$$

which is unbiased. Of course, we do not have access to the true mean but only to the observed data values. From the result (2.60) it follows that for a Gaussian distribution, the following estimate for the variance parameter is unbiased:

$$\widetilde{\sigma}^2 = \frac{N}{N-1}\sigma^2_{\mathrm{ML}} = \frac{1}{N-1}\sum_{n=1}^{N}(x_n - \mu_{\mathrm{ML}})^2. \tag{2.63}$$