



Figure 3.2 Histogram plots of the mean of N uniformly distributed numbers for various values of N . We observe that as N increases, the distribution tends towards a Gaussian.

Exercise 3.8

a single real variable, the distribution that maximizes the entropy is the Gaussian. This property applies also to the multivariate Gaussian.

Another situation in which the Gaussian distribution arises is when we consider the sum of multiple random variables. The *central limit theorem* tells us that, subject to certain mild conditions, the sum of a set of random variables, which is of course itself a random variable, has a distribution that becomes increasingly Gaussian as the number of terms in the sum increases (Walker, 1969). We can illustrate this by considering N variables x_1, \dots, x_N each of which has a uniform distribution over the interval $[0, 1]$ and then considering the distribution of the mean $(x_1 + \dots + x_N)/N$. For large N , this distribution tends to a Gaussian, as illustrated in Figure 3.2. In practice, the convergence to a Gaussian as N increases can be very rapid. One consequence of this result is that the binomial distribution (3.9), which is a distribution over m defined by the sum of N observations of the random binary variable x , will tend to a Gaussian as $N \rightarrow \infty$ (see Figure 3.1 for $N = 10$).

The Gaussian distribution has many important analytical properties, and we will consider several of these in detail. As a result, this section will be rather more technically involved than some of the earlier sections and will require familiarity with various matrix identities.

Appendix A

3.2.1 Geometry of the Gaussian

We begin by considering the geometrical form of the Gaussian distribution. The functional dependence of the Gaussian on \mathbf{x} is through the quadratic form

$$\Delta^2 = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}), \quad (3.27)$$

which appears in the exponent. The quantity Δ is called the *Mahalanobis distance* from $\boldsymbol{\mu}$ to \mathbf{x} . It reduces to the Euclidean distance when $\boldsymbol{\Sigma}$ is the identity matrix. The Gaussian distribution is constant on surfaces in \mathbf{x} -space for which this quadratic form is constant.

Exercise 3.11

First, note that the matrix $\boldsymbol{\Sigma}$ can be taken to be symmetric, without loss of generality, because any antisymmetric component would disappear from the exponent. Now consider the eigenvector equation for the covariance matrix

$$\boldsymbol{\Sigma} \mathbf{u}_i = \lambda_i \mathbf{u}_i \quad (3.28)$$