Setting the derivative of (3.20) with respect to μ_k to zero, we obtain

$$\mu_k = -m_k/\lambda. \tag{3.21}$$

We can solve for the Lagrange multiplier λ by substituting (3.21) into the constraint $\sum_k \mu_k = 1$ to give $\lambda = -N$. Thus, we obtain the maximum likelihood solution for μ_k in the form

$$\mu_k^{\mathrm{ML}} = \frac{m_k}{N}, \tag{3.22}$$

which is the fraction of the N observations for which $x_k = 1$.

We can also consider the joint distribution of the quantities $m_1, \ldots, m_K$, conditioned on the parameter vector $\boldsymbol{\mu}$ and on the total number N of observations. From (3.17), this takes the form

$$\mathrm{Mult}(m_1, m_2, \ldots, m_K|\boldsymbol{\mu}, N) = \binom{N}{m_1 m_2 \ldots m_K} \prod_{k=1}^{K} \mu_k^{m_k}, \tag{3.23}$$

which is known as the *multinomial* distribution. The normalization coefficient is the number of ways of partitioning N objects into K groups of size $m_1, \ldots, m_K$ and is given by

$$\binom{N}{m_1 m_2 \ldots m_K} = \frac{N!}{m_1! m_2! \ldots m_K!}. \tag{3.24}$$

Note that two-state quantities can be represented either as binary variables and modelled using the binomial distribution (3.9) or as 1-of-2 variables and modelled using the distribution (3.14) with $K = 2$.

3.2. The Multivariate Gaussian

The Gaussian, also known as the normal distribution, is a widely used model for the distribution of continuous variables. We have already seen that for of a single variable x, the Gaussian distribution can be written in the form

Section 2.3

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \tag{3.25}$$

where μ is the mean and σ^2 is the variance. For a D-dimensional vector $\mathbf{x}$, the multivariate Gaussian distribution takes the form

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \tag{3.26}$$

where $\boldsymbol{\mu}$ is the D-dimensional mean vector, $\boldsymbol{\Sigma}$ is the $D \times D$ covariance matrix, and $\det \boldsymbol{\Sigma}$ denotes the determinant of $\boldsymbol{\Sigma}$.

The Gaussian distribution arises in many different contexts and can be motivated from a variety of different perspectives. For example, we have already seen that for

Section 2.5