$$\boldsymbol{\eta} = \begin{pmatrix} \mu/\sigma^2 \\ -1/2\sigma^2 \end{pmatrix} \tag{3.164}$$

$$\mathbf{u}(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix} \tag{3.165}$$

$$h(\mathbf{x}) = (2\pi)^{-1/2} \tag{3.166}$$

$$g(\boldsymbol{\eta}) = (-2\eta_2)^{1/2} \exp\left(\frac{\eta_1^2}{4\eta_2}\right). \tag{3.167}$$

Finally, we shall sometimes make use of a restricted form of (3.138) in which we choose $\mathbf{u}(\mathbf{x}) = \mathbf{x}$. However, this can be somewhat generalized by noting that if $f(\mathbf{x})$ is a normalized density then

$$\frac{1}{s} f\left(\frac{1}{s}\mathbf{x}\right) \tag{3.168}$$

is also a normalized density, where $s > 0$ is a scale parameter. Combining these, we arrive at a restricted set of exponential family class-conditional densities of the form

$$p(\mathbf{x}|\boldsymbol{\lambda}_k, s) = \frac{1}{s} h\left(\frac{1}{s}\mathbf{x}\right) g(\boldsymbol{\lambda}_k) \exp\left\{\frac{1}{s}\boldsymbol{\lambda}_k^{\mathrm{T}}\mathbf{x}\right\}. \tag{3.169}$$

Note that we are allowing each class to have its own parameter vector $\boldsymbol{\lambda}_k$ but we are assuming that the classes share the same scale parameter s.

3.4.1 Sufficient statistics

Let us now consider the problem of estimating the parameter vector $\boldsymbol{\eta}$ in the general exponential family distribution (3.138) using the technique of maximum likelihood. Taking the gradient of both sides of (3.139) with respect to $\boldsymbol{\eta}$, we have

$$\begin{aligned} &\nabla g(\boldsymbol{\eta}) \int h(\mathbf{x}) \exp\left\{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u}(\mathbf{x})\right\} \mathrm{d}\mathbf{x} \\ &+ \quad g(\boldsymbol{\eta}) \int h(\mathbf{x}) \exp\left\{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u}(\mathbf{x})\right\} \mathbf{u}(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0. \end{aligned} \tag{3.170}$$

Rearranging and making use again of (3.139) then gives

$$-\frac{1}{g(\boldsymbol{\eta})}\nabla g(\boldsymbol{\eta}) = g(\boldsymbol{\eta}) \int h(\mathbf{x}) \exp\left\{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u}(\mathbf{x})\right\} \mathbf{u}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \mathbb{E}[\mathbf{u}(\mathbf{x})]. \tag{3.171}$$

We therefore obtain the result

$$-\nabla \ln g(\boldsymbol{\eta}) = \mathbb{E}[\mathbf{u}(\mathbf{x})]. \tag{3.172}$$

Exercise 3.36

Note that the covariance of $\mathbf{u}(\mathbf{x})$ can be expressed in terms of the second derivatives of $g(\boldsymbol{\eta})$, and similarly for higher-order moments. Thus, provided we can normalize a distribution from the exponential family, we can always find its moments by simple differentiation.