

## Section 3.5

simplicity and flexibility but also suffers from significant limitations, as we will see when we discuss histogram methods in more detail later. Another approach starts, like the von Mises distribution, from a Gaussian distribution over a Euclidean space but now marginalizes onto the unit circle rather than conditioning (Mardia and Jupp, 2000). However, this leads to more complex forms of distribution and will not be discussed further. Finally, any valid distribution over the real axis (such as a Gaussian) can be turned into a periodic distribution by mapping successive intervals of width  $2\pi$  onto the periodic variable  $(0, 2\pi)$ , which corresponds to ‘wrapping’ the real axis around the unit circle. Again, the resulting distribution is more complex to handle than the von Mises distribution.

### 3.4. The Exponential Family

The probability distributions that we have studied so far in this chapter (with the exception of mixture models) are specific examples of a broad class of distributions called the *exponential family* (Duda and Hart, 1973; Bernardo and Smith, 1994). Members of the exponential family have many important properties in common, and it is illuminating to discuss these properties in some generality.

The exponential family of distributions over  $\mathbf{x}$ , given parameters  $\boldsymbol{\eta}$ , is defined to be the set of distributions of the form

$$p(\mathbf{x}|\boldsymbol{\eta}) = h(\mathbf{x})g(\boldsymbol{\eta}) \exp \{ \boldsymbol{\eta}^T \mathbf{u}(\mathbf{x}) \} \quad (3.138)$$

where  $\mathbf{x}$  may be scalar or vector and may be discrete or continuous. Here  $\boldsymbol{\eta}$  are called the *natural parameters* of the distribution, and  $\mathbf{u}(\mathbf{x})$  is some function of  $\mathbf{x}$ . The function  $g(\boldsymbol{\eta})$  can be interpreted as the coefficient that ensures that the distribution is normalized, and therefore, it satisfies

$$g(\boldsymbol{\eta}) \int h(\mathbf{x}) \exp \{ \boldsymbol{\eta}^T \mathbf{u}(\mathbf{x}) \} d\mathbf{x} = 1 \quad (3.139)$$

where the integration is replaced by summation if  $\mathbf{x}$  is a discrete variable.

We begin by taking some examples of the distributions introduced earlier in the chapter and showing that they are indeed members of the exponential family. Consider first the Bernoulli distribution:

$$p(x|\mu) = \text{Bern}(x|\mu) = \mu^x (1 - \mu)^{1-x}. \quad (3.140)$$

Expressing the right-hand side as the exponential of the logarithm, we have

$$\begin{aligned} p(x|\mu) &= \exp \{ x \ln \mu + (1 - x) \ln(1 - \mu) \} \\ &= (1 - \mu) \exp \left\{ \ln \left( \frac{\mu}{1 - \mu} \right) x \right\}. \end{aligned} \quad (3.141)$$

Comparison with (3.138) allows us to identify

$$\eta = \ln \left( \frac{\mu}{1 - \mu} \right) \quad (3.142)$$