L Multivariate normal distributions

L.1 The multivariate normal

Say we have p independent zero mean unit variance normally distributed variates. The joint distribution is just the product of the individual variates:

$$p(x_1, \ldots, x_p) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_1^2}{2}\right) \ldots \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_p^2}{2}\right). \tag{152}$$

This can be expressed concisely in vector terms as:

$$p(\mathbf{x}) = \frac{1}{\sqrt{2\pi}^p} \exp\left(-\frac{\mathbf{x}^T\mathbf{x}}{2}\right). \tag{153}$$

Now scale each variate individually by substituting

$$\mathbf{x}' = \mathbf{A}\mathbf{x}, \tag{154}$$

where $\mathbf{A}$ is a square non-singular matrix. The Jacobian determinant is:

$$\frac{dx}{dx'} = |\mathbf{A}|^{-1}, \tag{155}$$

so

$$p(\mathbf{x}') = \frac{1}{\sqrt{2\pi}^p |\mathbf{A}|} \exp\left(-\frac{(\mathbf{A}^{-1}\mathbf{x}')^T \mathbf{A}^{-1}\mathbf{x}'}{2}\right), \tag{156}$$

$$= \frac{1}{\sqrt{2\pi}^p |\mathbf{A}|} \exp\left(-\frac{1}{2}\mathbf{x}'^T \mathbf{A}^{-1^T} \mathbf{A}^{-1}\mathbf{x}'\right), \tag{157}$$

$$= \frac{1}{\sqrt{2\pi}^p |\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}'^T \mathbf{\Sigma}^{-1}\mathbf{x}'\right), \tag{158}$$

where $\mathbf{\Sigma}$ is now a positive definite matrix. Notice that if $\mathbf{A}$ is diagonal, this is really just a product of p normally distributed variates.

Say we then apply an offset; that is, write

$$\mathbf{x}'' = \mathbf{x}' + \boldsymbol{\mu}, \tag{159}$$

where $\boldsymbol{\mu}$ is another vector. In this case, the Jacobian determinant is the identity matrix, and

$$p(\mathbf{x}'') = \frac{1}{\sqrt{2\pi}^p |\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}'' - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{x}'' - \boldsymbol{\mu})\right). \tag{160}$$

So, the multivariate normal arises as a linear transform of independent normal distributions.

L.2 The matrix normal

In the above, we expressed $\mathbf{x}$ as a vector. We could also express it as a diagonal matrix:

$$\mathbf{X} = \begin{pmatrix} x_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & x_p \end{pmatrix}. \tag{161}$$

In this case,

$$p(\mathbf{X}) = \frac{1}{\sqrt{2\pi}^p} \exp\left(-\frac{\operatorname{tr}\left[\mathbf{X}^T\mathbf{X}\right]}{2}\right). \tag{162}$$

Turning the handle, to a first approximation we would get a distribution over a matrix variate. There's something screwy about the normaliser though - it goes from a p dimensional thing to a $p \times p$ dimensional one.