Gaussian Identities

Michael A. Osborne
mosb@robots.ox.ax.uk

January 27, 2014

The following identities are almost indispensible when dealing with Gaussian distributions, which we denote in the usual way as

$$\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\mathbf{A}) \triangleq \frac{1}{\sqrt{\det 2\pi\mathbf{A}}}\,\exp\left(-\frac{1}{2}\,(\boldsymbol{x}-\boldsymbol{\mu})^{\mathsf{T}}\,\mathbf{A}^{-1}\,(\boldsymbol{x}-\boldsymbol{\mu})\right)\,.$$

Probably the most important properties of a multivariate Gaussian distribution are that its marginals and conditionals are both themselves Gaussian. That is, if we have

$$p(\,\boldsymbol{x},\,\boldsymbol{y}\mid I\,) \triangleq \mathcal{N}\left(\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix};\begin{bmatrix}\boldsymbol{\mu}\\ \boldsymbol{\nu}\end{bmatrix},\begin{bmatrix}\mathbf{A} & \mathbf{C}\\ \mathbf{C}^{\mathsf{T}} & \mathbf{B}\end{bmatrix}\right)\,,$$

then its marginal and conditional distributions are respectively

$$p(\,\boldsymbol{x}\mid I\,) = \mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\mathbf{A}) \tag{1}$$

$$p(\,\boldsymbol{x}\mid\boldsymbol{y},\,I\,) = \mathcal{N}\big(\boldsymbol{x};\boldsymbol{\mu}+\mathbf{C}\,\mathbf{B}^{-1}(\boldsymbol{y}-\boldsymbol{\nu}),\mathbf{A}-\mathbf{C}\,\mathbf{B}^{-1}\,\mathbf{C}^{\mathsf{T}}\big)\,. \tag{2}$$

We now turn to another important property of the Gaussian distribution. If we have

$$\begin{aligned} p(\,\boldsymbol{x}\mid I\,) &\triangleq \mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\mathbf{A})\\ p(\,\boldsymbol{y}\mid\boldsymbol{x},\,I\,) &\triangleq \mathcal{N}(\boldsymbol{y};\mathbf{M}\,\boldsymbol{x}+\boldsymbol{c},\mathbf{L})\,, \end{aligned} \tag{3}$$

then the joint distribution can be written as

$$p(\,\boldsymbol{x},\,\boldsymbol{y}\mid I\,) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix};\begin{bmatrix}\boldsymbol{\mu}\\ \mathbf{M}\,\boldsymbol{\mu}+\boldsymbol{c}\end{bmatrix},\begin{bmatrix}\mathbf{A} & \mathbf{A}\,\mathbf{M}^{\mathsf{T}}\\ \mathbf{M}\,\mathbf{A} & \mathbf{L}+\mathbf{M}\,\mathbf{A}\,\mathbf{M}^{\mathsf{T}}\end{bmatrix}\right)\,, \tag{4}$$

and so, using (1) and (2)

$$p(\,\boldsymbol{y}\mid I\,) = \mathcal{N}\big(\boldsymbol{y};\mathbf{M}\,\boldsymbol{\mu}+\boldsymbol{c},\mathbf{L}+\mathbf{M}\,\mathbf{A}\,\mathbf{M}^{\mathsf{T}}\big) \tag{5}$$

$$p(\,\boldsymbol{x}\mid\boldsymbol{y},\,I\,) = \mathcal{N}(\boldsymbol{x};\boldsymbol{\mu}+\boldsymbol{\Gamma}(\boldsymbol{y}-\mathbf{M}\,\boldsymbol{\mu}-\boldsymbol{c}),\mathbf{A}-\boldsymbol{\Gamma}\,\mathbf{M}\,\mathbf{A}) \tag{6}$$

where

$$\boldsymbol{\Gamma} = \mathbf{A}\,\mathbf{M}^{\mathsf{T}}\left(\mathbf{L}+\mathbf{M}\,\mathbf{A}\,\mathbf{M}^{\mathsf{T}}\right)^{-1}$$