The Bayesian Evidence for Multiple Gaussian Observations

David Pfau

2 November 2009

We present a simple derivation of the evidence $p(y_{1:N})$ of multiple observations from a univariate Gaussian with known variance and unknown mean. Assume N points are drawn from the distribution $y|\mu \sim \mathcal{N}(\mu, \sigma^2)$, while μ itself is drawn according to $\mu \sim \mathcal{N}(0, \tau^2)$. Then the joint probability of $y_{1:N}$ and μ is

$$\begin{aligned} p(y_{1:N}, \mu) = p(\mu) \prod_{i=1}^{N} p(y_i|\mu) &\propto exp\left(-\frac{\mu}{2\tau^2} - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \mu)^2 \right) \qquad (1) \\ &= exp\left(-\frac{1}{2} \left(\frac{1}{\tau^2} + \frac{N}{\sigma^2} \right) \mu^2 + \frac{1}{\sigma^2} \sum_{i=1}^{N} \mu y_i - \frac{1}{2\sigma^2} \sum_{i=1}^{N} y_i^2 \right) \end{aligned}$$

Which is to say, the joint distribution is a Gaussian with mean 0 and precision

$$\Lambda = \begin{pmatrix} \frac{1}{\sigma^2} I_N & \frac{1}{\sigma^2} \mathbf{1}_N \\ \frac{1}{\sigma^2} \mathbf{1}_N^T & \frac{1}{\sigma^2} + \frac{N}{\tau^2} \end{pmatrix} \qquad (2)$$

Where $\mathbf{1}_N$ is the length N vector with all components 1. The marginal variance of a multivariate Gaussian is simply the block of the covariance matrix that corresponds to the desired dimensions, while the mean is simply the mean of the relevant dimensions. So the evidence, which is simply the joint marginalized over μ, has mean 0 and a variance that corresponds to the top N by N block of the joint covariance. We can find this by inverting the precision matrix above (and using a handy block matrix inversion formula I googled):

$$\Sigma_{y_{1:N}} = \sigma^2 I_N + \tau^2 \mathbf{1}_{N \times N} \qquad (3)$$

The determinant and precision for the evidence are given by

$$\begin{aligned} |\Sigma_{y_{1:N}}| &= \sigma^{2N} \left(1 + N \frac{\tau^2}{\sigma^2} \right) \\ \Lambda_{y_{1:N}} &= \frac{1}{\sigma^2} I_N - \frac{\tau^2/\sigma^2}{\sigma^2 + \tau^2} \mathbf{1}_{N \times N} \end{aligned}$$

1 Handy Identities I Cribbed From The Internet

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix} \tag{4}$$

Which means if we write out the joint precision of a and b in block form, we can read off the marginal precisions directly:

$$\begin{aligned} \Lambda_a &= \Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba} \\ \Lambda_b &= \Lambda_{bb} - \Lambda_{ba}\Lambda_{aa}^{-1}\Lambda_{ab} \end{aligned} \tag{5}$$

And the determinant of a full rank plus a rank one matrix is

$$|A + uv^T| = (1 + v^T A^{-1} u)\, |A| \tag{6}$$