Bayes factors for linear models

9.29 Suppose that in addition to the original model (9.1) we have an alternative model $\mathbf{y} = \mathbf{X}_A\boldsymbol{\beta}_A + \boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon}$ distributed as $N(\mathbf{0}, \sigma^2\mathbf{I})$ as before. The prior distribution for $(\boldsymbol{\beta}_A, \sigma^2)$ is $NIG(a, d, \mathbf{m}_A, \mathbf{V}_A)$. The two models make the same assumptions about the error term $\boldsymbol{\epsilon}$, including the same $IG(a, d)$ prior distribution for σ^2. They differ in the matrices $\mathbf{X}$ and $\mathbf{X}_A$ of coefficients, and so try to explain or predict the response variable $\mathbf{y}$ using different regressor variables. Accordingly, they have different parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_A$.

The Bayes factor in favour of the alternative model is the ratio $B = f_A(\mathbf{y})/f(\mathbf{y})$ of the resulting marginal densities for $\mathbf{y}$ under the two models. The denominator is obtained as follows. From (9.2) and (9.11).

$$\begin{aligned} f(\mathbf{y}) &= \int\int f(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}, \sigma^2)\, \mathrm{d}\boldsymbol{\beta}\, \mathrm{d}\sigma^2 \\ &= k \int\int (\sigma^2)^{-(d+n+p+2)/2} \exp\{-Q/(2\sigma^2)\}\, \mathrm{d}\boldsymbol{\beta}\, \mathrm{d}\sigma^2, \end{aligned} \tag{9.43}$$

where

$$k = \frac{(a/2)^{d/2}}{(2\pi)^{(n+p)/2}|\mathbf{V}|^{1/2}\Gamma(d/2)}$$

and Q is given by (9.14). Now the equivalent expression (9.15) allows us to do the integration with respect to $\boldsymbol{\beta}$ in (9.43), to yield

$$\begin{aligned} f(\mathbf{y}) &= k|\mathbf{V}^\star|^{1/2}(2\pi)^{p/2} \int (\sigma^2)^{-(d^\star+2)/2} \exp\{-a^\star/(2\sigma^2)\}\, \mathrm{d}\sigma^2 \\ &= k|\mathbf{V}^\star|^{1/2}(2\pi)^{p/2}(a^\star/2)^{-d^\star/2}\Gamma(d^\star/2) \\ &= \frac{|\mathbf{V}^\star|^{1/2}a^{d/2}\Gamma(d^\star/2)}{|\mathbf{V}|^{1/2}\pi^{n/2}\Gamma(d/2)}(a^\star)^{-d^\star/2}. \end{aligned} \tag{9.44}$$

Notice that $\mathbf{y}$ only appears in (9.44) through $a^\star$. The rest of the expression is the normalizing constant for $f(\mathbf{y})$.

9.30 The analogous expression for $f_A(\mathbf{y})$ adds subscript A to $\mathbf{V}$, $\mathbf{V}^\star$ and $\mathbf{a}^\star$, so that the Bayes factor is

$$B = \frac{|\mathbf{V}|^{1/2}|\mathbf{V}_A^\star|^{1/2}}{|\mathbf{V}_A|^{1/2}|\mathbf{V}^\star|^{1/2}} \cdot \left(\frac{a^\star}{a_A^\star}\right)^{d^\star/2}. \tag{9.45}$$

The four determinants do not depend on the observed data $\mathbf{y}$, and are concerned with the relative strength of prior information and data information about the parameter vectors, as measured by $\mathbf{V}$, $\mathbf{V}_A$, $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}_A'\mathbf{X}_A$. The term involving $\mathbf{y}$ is an increasing function of $a^\star/a_A^\star$, and so favours the alternative model if it leads to a smaller $a_A^\star$ than the original model's $a^\star$. Since $d^\star = d + n$ is the same in both models, the Bayes factor tends to favour the model producing the lower posterior estimate of σ^2. This is intuitively reasonable since σ^2 determines the magnitude of the errors $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ or $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}_A\boldsymbol{\beta}_A$ and so measures the lack of fit of the model to the data. An estimate such as $E(\sigma^2 \mid \mathbf{y}) = a^\star/(d^\star - 2)$ of σ^2 estimates this lack of fit.