Appendix 2: Bayesian Anova

Richard Newton\textsuperscript{1}, Jason Hinds\textsuperscript{2} and Lorenz Wernisch\textsuperscript{1}

\textsuperscript{1}MRC Biostatistics Unit, Robinson Way, Cambridge, CB2 0SR, UK
\textsuperscript{2}Bacterial Microarray Group, Division of Cellular \& Molecular Medicine, St. George’s, University of London, Cranmer Terrace, London, SW17 0RE, UK

Email: Richard Newton\textsuperscript{*} - richard.newton@mrc-bsu.cam.ac.uk; Jason Hinds - j.hinds@sgul.ac.uk; Lorenz Wernisch - lorenz.wernisch@mrc-bsu.cam.ac.uk;

\textsuperscript{*}Corresponding author

We follow the Bayesian treatment of linear models as outlined in Sorensen and Gianola [Sorensen and Gianola(2002)] although the specific forms of posteriors we are interested in deviate slightly from the book.

The objective is to obtain posterior distributions for the coefficients $\theta$ in

\[ y = X\theta + Zu + \epsilon \]

while integrating out the nuisance variables $u$ and $\epsilon$. As priors we assume $\theta \sim N(0, B\sigma^2_\theta)$, $u \sim N(0, A\sigma^2_u)$, and $\epsilon \sim N(0, I\sigma^2_\epsilon)$. It is straightforward to see that, conditioned on $\theta$, the prior predictive distribution of $y - X\theta$ is a Gaussian with mean 0 and covariance matrix $\sigma^2_\epsilon V = ZAZ'\sigma^2_u + I_n\sigma^2_\epsilon$. That is,

\[ p(y \mid \theta, A, \sigma^2_u, \sigma^2_\epsilon) \propto (\sigma^2_\epsilon)^{-n/2} \exp\left(-\frac{1}{2\sigma^2_\epsilon}((y - X\theta)'V^{-1}(y - X\theta))^{-(n+d)/2}\right) \]

\[ V = ZAZ'\frac{\sigma^2_u}{\sigma^2_\epsilon} + I_n \]

(we keep track of all variance terms for later use). The posterior distribution for $\theta$ is (see equation (6.67) in [Sorensen and Gianola(2002)])

\[ p(\theta \mid y, A, \sigma^2_u, B, \sigma^2_\theta, \sigma^2_\epsilon) \propto (\sigma^2_\theta)^{-b/2} (\sigma^2_\epsilon)^{-n/2} \exp\left(-\frac{1}{2\sigma^2_\epsilon}((\theta - \hat{\theta})'W^{-1}(\theta - \hat{\theta}))\right) \]

\[ W = (X'V^{-1}X + B^{-1}\frac{\sigma^2_\theta}{\sigma^2_\theta})^{-1} \]

\[ \hat{\theta} = WX'V^{-1}y \]

where $b$ is the dimension of $\theta$.

We are left with the problem of integrating out variance components $\sigma^2_u, \sigma^2_\epsilon, \sigma^2_\theta$. There is no analytical solution to this integral in its general form. However, making the usual assumption that the error variance $\sigma^2_\epsilon$ is actually closely...
related to the variance factors $\sigma_\theta^2$ and $\sigma_\theta^2$ of the coefficients and setting $\sigma^2 = \sigma_\varepsilon^2 = \sigma_\theta^2$, a conjugate analysis is possible for $\sigma^2$. We assume a prior $p_{ICl}(\sigma^2 \mid \nu_0, \sigma_0^2)$. First note that

$$(y - X\hat{\theta})'V^{-1}(y - X\hat{\theta}) + \theta'B^{-1}\theta = (\theta - \hat{\theta})'W^{-1}(\theta - \hat{\theta}) + S_\theta + S_e$$

with

$$S_\theta = \hat{\theta}'D(D + B^{-1})B^{-1}\hat{\theta}, \quad S_e = (y - X\hat{\theta})^2, \quad D = X'V^{-1}X, \quad \hat{\theta} = D^{-1}X'V^{-1}y$$

where $\hat{\theta}$ is the ML estimate of $\theta$ (after integrating over $u$). The joint distribution of $\theta$ and $\sigma^2 = \sigma_\varepsilon^2 = \sigma_\theta^2$ is

$$p(\theta, \sigma^2 \mid y, A, B) \propto (\sigma^2)^{-(n/2+b/2+\nu_0/2+1)} \exp\left(\frac{(\theta - \hat{\theta})'W^{-1}(\theta - \hat{\theta}) + S_\theta + S_e + \nu_0\sigma_0^2}{2\sigma^2}\right)$$

We obtain

$$p(\theta \mid y, A, B, \nu_0, \sigma_0^2) = \int_0^\infty p(\theta, \sigma^2 \mid y, A, B) d(\sigma^2) = p_t(\theta \mid \hat{\theta}, n + \nu_0, (S_\theta + S_e + \nu_0\sigma_0^2)W) \quad (1)$$

where $n$ is the dimension of $y$ and $b$ is the dimension of $\theta$.

Finally, to derive a likelihood for the optimisation of hyperparameters we start with the predictive likelihood conditioned on the variance components, which is a Gaussian with mean 0 and covariance

$$\sigma_e^2U = XBX'\sigma_\varepsilon^2 + ZAZ'\sigma_u^2 + I\sigma_\theta^2.$$ For further analysis we assume equality of all variance components and use the same prior as above on $\sigma^2$,

$$p(y \mid A, B, \nu_0, \sigma_0) = \int p_N(y \mid A, B, \sigma^2) p_{ICl}(\sigma^2 \mid \nu_0, \sigma_0^2) d(\sigma^2) = p_t(y \mid 0, \nu_0, \sigma_0^2(XBX' + ZAZ' + I_n)) \quad (2)$$

References