IMPLEMENTING CONJUGATE PRIORS WITH DUMMY OBSERVATIONS

1. THE SETUP

Suppose our data satisfy the SNLM

$$\underset{T\times 1}{y} = \underset{T\times k}{X}\beta + \varepsilon \tag{1}$$

with $\{\varepsilon \mid X\} \sim N(0, \sigma^2 I)$.

The **conjugate prior** for this setup with β and σ^2 unknown has pdf

$$\frac{1}{\Gamma(p)(2\pi)^{k/2}}\alpha^p \sigma^{-2(p+1)-k} |\Omega|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(\beta-\bar{\beta})'\Omega^{-1}(\beta-\bar{\beta}) - \frac{\alpha}{\sigma^2}\right). \tag{2}$$

This is known as the **normal-inverse-gamma** distribution, because $\left\{\beta \mid \sigma^2\right\}$ is normal, that is $N(\bar{\beta}, \sigma^2\Omega)$, and σ^{-2} has an unconditional gamma distribution, that is $\Gamma(p, \alpha)$. Ω and α are fixed parameters of the prior.

Note that this prior makes our prior uncertainty about β proportional to the unknown parameter σ^2. This may or may not make sense in a particular application.

With this prior, the posterior distribution implies

$$\begin{gathered}
\beta \mid \left\{\sigma^2, y\right\} \sim N(\beta^*, \Sigma^*) \\
\beta^* = (X'X + \Omega^{-1})^{-1}(X'X\hat{\beta} + \Omega^{-1}\bar{\beta}) \\
\Sigma^* = \sigma^2 (X'X + \Omega^{-1})^{-1},
\end{gathered}$$

where $\hat{\beta} = (X'X)^{-1}X'y$ is the OLS estimate of β. These formulas can be verified by a "completing the square" exercise.

Note that this makes the posterior mean a matrix weighted average of $\hat{\beta}$ and the prior mean $\bar{\beta}$. In fact, the posterior mean conditional on σ^2 does not depend on σ^2, so this conditional posterior mean is also the unconditional posterior mean.

2. DUMMY OBSERVATIONS

Now we re-express this posterior pdf in terms of dummy observations. Let $\tilde{X}$ have the property that $\tilde{X}'\tilde{X} = \Omega^{-1}$ and let $\tilde{y} = \tilde{X}\bar{\beta}$. Then it is straightforward (more completing the square) to verify that the likelihood can be written as

$$(2\pi)^{-k/2}\Gamma(p)^{-1}\alpha^p \sigma^{-(T+k+2(p+1))} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(\hat{u}^{*\prime}\hat{u}^* + (\beta-\beta^*)'X^{*\prime}X^*(\beta-\beta^*) + 2\alpha)}, \tag{3}$$

©2003 by Christopher A. Sims. This document may be reproduced for educational and research purposes, so long as the copies contain this notice and are retained for personal use or distributed free.

where $\beta^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*$ is the OLS estimator of beta based on the expanded data set in which $\tilde{y}$ and $\tilde{X}$ are appended to the real data, and $\hat{u}^*$ is the corresponding extended residual vector, formed with the expanded data and using β^* as the OLS estimator.

The exponential part of this representation of the likelihood is exactly in the form it would have taken if we had observed y^* and X^* as real data, except for the term α/σ^2. To get the likelihood built from dummy observations exactly into the form of the posterior with conjugate prior, we would have to add an additional dummy observation y^{**}, X^{**} with $X^{**} = 0$ and $y^{**} = \sqrt{2\alpha}$, to match this extra term in the exponent.

The $\sigma^{-T-2(p+1)-k}$ factor is what we would have in the likelihood if we had exactly $2(p+1)+k$ dummy observations. Since with any proper prior we will have $p > 0$, it is straightforward to construct $(k+2p+2)\times k$ matrices $\tilde{X}$ that satisfy $\tilde{X}'\tilde{X} = \Omega^{-1}$, so that we match the posterior perfectly with a dummy-observation likelihood. This obviously will not work for p not an integral multiple of $\frac{1}{2}$.

The constant factors in this likelihood are quite different from what we would get from a likelihood generated naively from the dummy observations alone. This is not important for inference on β, σ^2 in this model by itself, but if we need to integrate over β, σ^2 to get the posterior probability of this model when it is competing with other possible models, these constant terms matter. The factors in the posterior that would not be present in a dummy-observation likelihood are $\Gamma(p)^{-1}\alpha^p |\Omega|^{-\frac{1}{2}}$. In comparing models that have the same prior on σ^2, the only part of this that will matter is $|\Omega|^{-\frac{1}{2}}$.

3. Marginal posteriors

If we integrate β out of (3) we arrive at

$$\Gamma(p)^{-1}\alpha^p\sigma^{-(T+2(p+1))}|\Omega|^{-\frac{1}{2}}\left|X^{*\prime}X^*\right|^{-\frac{1}{2}}e^{-\frac{1}{2\sigma^2}(\hat{u}^{*\prime}\hat{u}^*+2\alpha)}\,.$$

This makes the pdf of σ^{-2} proportional to a $\Gamma\left(\frac{1}{2}(T+2p), \alpha+\frac{1}{2}\hat{u}^{*\prime}\hat{u}^*\right)$.

If we integrate σ^2 out of (3) we get

$$(2\pi)^{-k/2}\Gamma(p)^{-1}\Gamma\left(\tfrac{1}{2}(T+k)+p\right)|\Omega|^{-\frac{1}{2}}\left(\tfrac{1}{2}\left(\hat{u}^{*\prime}\hat{u}^*+(\beta-\beta^*)'X^{*\prime}X^*(\beta-\beta^*)\right)+\alpha\right)^{-\frac{1}{2}(T+k)-p}.$$

This is proportional to a multivariate t pdf with $T+2p$ degrees of freedom.

4. Building priors from dummy observations

We have to this point shown how to take a conjugate prior and re-express it in terms of dummy observations. Often it is intuitively appealing to formulate priors directly in terms of dummy observations. This is particularly helpful when prior information seems to be expressible in terms of more or fewer than k dummy observations. For example, we might want to use prior information that the sum of coefficients in a regression is close to one (in a production function, for example). This is easily expressed as a dummy observation with $y^* = \lambda$, $X^* = \lambda\begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}$, where λ is a scale factor determining the precision of our prior beliefs, in relation to σ. It is clear how to introduce this single dummy observation, even though by itself it corresponds to no proper prior. Similarly, in a production function with two independent variables, we might believe the sum of coefficients should be close to one, that the coefficient on K should be close to .3, and the coefficient on L should be close to .7. These three beliefs can easily be combined via three dummy observations, though one would have to give some thought to how to weight them. Of course independent priors making the mean of the K coefficient .3 and the mean of the L coefficient .7 already imply a prior mean of 1 for for their sum, but by adding the dummy observation on their sum, we imply a negative correlation between the two coefficients in our prior. This allows, e.g., for our being much less certain about the individual .7 and .3 values than about their summing to one.

When the dummy observations X^* are full rank, so their number equals or exceeds k, we can, even though we have formulated the prior with dummy observations, put the $\alpha^p \Gamma(p) \left|\Omega\right|^{-\frac{1}{2}}$ term into the likelihood to get correct posteriors on models. When the number of dummy observations is less than k, Ω^{-1} is singular, so there is no way to include it. This is just a special case of the general proposition that when making model comparisons, improper priors create problems and paradoxes.