An Introduction to Modern Bayesian Econometrics

Answers to Selected Exercises.

Most of the exercises in the book ask readers to simulate and calculate. So there are really no answers to give. This is consistent with the spirit in which the book was written which empasized calculation and not formal, and often rather tedious, algebraic derivations. But some questions need mathematical answers and these are offered below.

Page 61, first exercise: The posterior density is found by multiplying the likelihood and the prior and it and its logarithm are

$$p(\tau|y) \propto \tau^{n/2-1}\exp\{-\tau\Sigma y_i^2/2\}; \qquad \log(p(\tau|y)) = \left(\frac{n-2}{2}\right)\log\tau - \frac{\tau\Sigma y_i^2}{2}.$$

Its first two derivatives are

$$\frac{\partial\log(p(\tau|y))}{\partial\tau} = \left(\frac{n-2}{2}\right)\frac{1}{\tau} - \frac{\Sigma y_i^2}{2}; \qquad \frac{\partial^2\log(p(\tau|y))}{\partial\tau^2} = -\left(\frac{n-2}{2}\right)\frac{1}{\tau^2}.$$

If we equate the first of these expressions to zero and solve for τ we find the unique solution $\widehat{\tau} = (n-2)/\Sigma y_i^2$ and since from the second expression we see that the log posterior density is globally concave for $n > 2$ this solution is the posterior mode. The negative hessian at $\widehat{\tau}$ is

$$-H(\widehat{\tau}) = \frac{(\Sigma y_i^2)^2}{2(n-2)}.$$

The information, $I_\tau(\tau)$ is the expectation of the negative hessian, but since this is non-stochastic, given τ, the information is identical to the negative hessian and the observed information is identical to $-H(\widehat{\tau})$ for this model.

From theorem 1.1 a large sample normal approximation to the posterior density of τ is

$$p(\tau|y) \simeq n(\widehat{\tau}, -H(\widehat{\tau}))$$

where $-H(\widehat{\tau})$ is the precision of this approximating normal distribution.

You can easily make a numerical comparsion of exact and approximate posterior distributions by statements – in R – such as

n <- 20; y <- rnorm(n); tauhat <- (n-2)/sum(y^2)................generate a sample of size 20 from $n(0,1)$ and calculate $\widehat{\tau}$.

tv <- seq(0.1,2,length=100)..............................tau values for the density plots *sig <- sqrt(2*tauhat*tauhat/(n-2))*.................standard deviation from the precision

plot(tv,dgamma(tv,n/2,sum(y^2)/2),type="l").............plots the exact posterior density, which is gamma

points(tv,dnorm(tv,tauhat,sig)).....................superimposes the normal approximation.

Try it.

Page 61, second exercise. The likelihood is the product of n poisson mass functions and is $\ell(y;\theta) \propto \theta^{\Sigma y_i} \exp\{-n\theta\}$ which leads to the posterior

$$p(\theta|y) \propto \theta^{n\overline{y}-1} e^{-n\theta} \qquad \text{with logarithm} \quad \log(p(\theta|y)) = (n\overline{y}-1)\log\theta - n\theta$$

and first two derivatives

$$\frac{\partial \log p}{\partial \theta} = \frac{n\overline{y}-1}{\theta} - n; \qquad \frac{\partial^2 \log p}{\partial \theta^2} = -\frac{n\overline{y}-1}{\theta^2}.$$

The log posterior is concave as long as $\Sigma_i y_i > 1$ and in this case the unique posterior mode is $\widehat{\theta} = \overline{y} - 1/n$. At this point the negative hessian is

$$-H(\widehat{\theta}) = \frac{n^2}{(n\overline{y}-1)}.$$

But using $E(\overline{y}|\theta) = \theta$ the information is

$$I_\theta(\theta) = \frac{n\theta - 1}{\theta^2}$$

and the observed information is

$$I_\theta(\widehat{\theta}) = \frac{n^2(n\overline{y}-2)}{(n\overline{y}-1)^2}$$

So two slightly different[1] asymptotic normal approximations to the posterior distribution of θ are

$$p(\theta|y) \simeq n(\widehat{\theta}, -H(\widehat{\theta})) \quad \text{and} \quad n(\widehat{\theta}, I(\widehat{\theta}))$$

A simulation and graphical comparison of the (exact) gamma posterior and the normal approximations could be done in R with, say,

*n <- 10; y <- rpois(n,3); ybar <- mean(y); thetahat <- (n*ybar-1)/n*

tv_ seq(0.1,5,length=100)

plot(tv,dgamma(tv,sum(y)-1,n),type="l",ylim=c(0, 0.8)).......the ylim is to get both curves on the same graph.

points(tv,dnorm(tv,thetahat,sig),pch=1)

[1]Note that the difference between $-H(\widehat{\theta})$ and $I_\theta(\widehat{\theta})$ depends on the ratio of $n\overline{y}-2$ to $n\overline{y}-1$ which is $O(1)$ whereas both precisions are $O(n)$

Note the role of $n\overline{y} = \Sigma y_i$ which is the total number of events. It is this rather than the number of individuals (n) whose magnitude determines whether the asymptotic normal approximation is adequate.

Page 106, chapter 2, exercise 2. This question asks you to work out and plot the predictive distribution of each element of a sequence of iid exponential(λ) variates given all the preceding ones and an initial flat prior on $\log\lambda$. The point of the calculation is to study the convergence of this sequence of distributions as evidence about λ increases. Eventually, when enough data have been seen to make λ known with high accuracy, the predictive distribution of the next observation will be that of an exponential variate with *known* λ.

The calculation is as follows: We require $p(y_n|y_1, y_2, ...y_{n-1})$ and this is

$$\begin{aligned} p(y_n|y_1, y_2, ...y_{n-1}) &= \int p(y_n, \lambda|y_1, y_2, ...y_{n-1})d\lambda \\ &= \int p(y_n|\lambda)p(\lambda|y_1, y_2, ...y_{n-1})d\lambda, \end{aligned}$$

where we have used the fact that $p(y_n|\lambda, y_1, y_2, ...y_{n-1}) = p(y_n|\lambda)$ since if you know λ the earlier data are irrelevant. The first term in the integrand is just the exponential(λ) density function while the second term is the posterior density of λ given the observations through y_{n-1}.This latter follows from Bayes theorem as

$$\begin{aligned} p(\lambda|y_1, y_2, ...y_{n-1}) &\propto p(y_1, y_2, ...y_{n-1}|\lambda)p(\lambda) \\ &= \lambda^{n-1}\exp\{-\lambda s_{n-1}\}/\lambda = \lambda^{n-2}\exp\{-\lambda s_{n-1}\}, \end{aligned}$$

where $s_{n-1} = y_1 + y_2 + ... + y_{n-1}$. Thus,

$$\begin{aligned} p(y_n|y_1, y_2, ...y_{n-1}) &\propto \int \lambda\exp\{-\lambda y_n\}\lambda^{n-2}\exp\{-\lambda s_{n-1}\}d\lambda \\ &= \int \lambda^{n-1}\exp\{-\lambda s_n\}d\lambda = \Gamma(n)s_n^{-n} \end{aligned}$$

where $s_n = y_n + s_{n-1}$. Finally, a straightforward exercise in integration supplies the normalizing constant and the exact predictive density is

$$p(y_n|y_1, y_2, ...y_{n-1}) = (n-1)\frac{s_{n-1}^{n-1}}{(y_n + s_{n-1})^n}; \qquad 0 \le y_n < \infty.$$

To plot this function you would want to define a sequence of values for y_n by, say *yv_seq(0.1,2,len=100)* and then choose n and generate, say, unit exponential variates by, say, *n=100, y <- rexp(n)*. Since you have chosen $\lambda = 1$ the first thing to plot is the distribution towards which the predictive distribution will tend as $n \to \infty$ and this is the unit exponental. You could do this by *plot(yv, dexp(yv), type="l", ylim = c(0,1.5), ylab = "predictive densities", xlab = "y")*. Then you could plot the predictive distribution for, say, y_5 by *n <- 5; s <- sum(y[1:n-1]); points(lv, (n-1)*s^(n-1)*(s+lv)^(-n))* followed by *text(locator(1),"n=5")* to label this curve. Finally you could plot the predictive density of y_{50} to see if it is closer to its limiting form than that for y_5, by using *n <- 50; s <- sum(y[1:n-1]); points(lv,(n-1)*s^(n-1)*(s+lv)^(-n))*.

Chapter 3, exercise 8.

(a) Using the definition $y - X\beta = \varepsilon$ write out $\varepsilon'(P \otimes I_n)\varepsilon$ in partitioned form to see its equivalence to $tr\ SP$.

(b) This is a straightforward application of definition 9, "completing the square".

(c) Using two versions of the likelihood, either (3.56) or (3.57,58) one obtains two versions of the posterior, namely

$$\begin{aligned} p(\beta, P|y) &\propto |P|^{(n-m-1)/2} \exp\{-(1/2)((\beta - \widehat{\beta})' X'(P \otimes I_n) X(\beta - \widehat{\beta})\} \\ &\quad \times \exp\{-(1/2)e'(P \otimes I_n)e\} \\ \text{and} \quad &\propto |P|^{(n-m-1)/2} \exp\{-(1/2) tr\ SP\}. \end{aligned}$$

The first of these shows immediately that, given P, β is $n(\widehat{\beta}, X'(P \otimes I_n)X)$. And the second shows, by comparison with definition 13 above that, given β, P is Wishart distributed. A Gibbs algorithm requires sampling alternately from a multivariate normal, which is an available distribution, and from a Wishart distribution – see definition 13.

(d) When $X_1 = X_2 = ...X_m = Z$ then X is block diagonal with Z in every diagonal block and zeros elsewhere. Thus $X = I_m \otimes Z$.Substituting this expression into the definition of the GLS estimator gives

$$\begin{aligned} \widehat{\beta} &= (X'(P \otimes I_n)X)^{-1} X'(P \otimes I_n) y \\ &\quad \text{But} \\ X'(P \otimes I_n)X &= (I_m \otimes Z'_{k \times n})(P \otimes I_n)(I_m \otimes Z_{n \times k}) = P \otimes Z'Z \\ &\quad \text{and so } \widehat{\beta} \text{ becomes} \\ \widehat{\beta} &= (P^{-1} \otimes (Z'Z)^{-1})(P \otimes Z') \begin{pmatrix} y_1 \\ y_2 \\ . \\ y_m \end{pmatrix} = \begin{pmatrix} (Z'Z)^{-1} Z' y_1 \\ (Z'Z)^{-1} Z' y_2 \\ . \\ (Z'Z)^{-1} Z' y_m \end{pmatrix} \end{aligned}$$

and this amounts to m separate least squares calculations.