Christiano
Econometric Tools for Macroeconomics
Bayesian Methods Exercise

1. Here is a very simple example to convey some of the ideas of Bayesian analysis. Consider a time series, y_t, which is independently and identically distributed $N(\mu, \sigma^2)$, where the value of σ^2 is known to the analyst. The analyst is uncertain about the value of μ and her uncertainty takes the form of a Normal distribution with mean, m, and variance, σ^2/ν. Here, σ^2 is the same parameter that governs the variance of y_t. That it also appears in the prior distribution is done for notational convenience, and obviously is done without loss of generality. Note that the bigger ν is, the more precise is the analyst's prior information about μ. For this reason, ν is referred to as the 'precision' of the analyst's prior. In this example, we can derive a simple easy-to-interpret expression for $p(\mu|y)$, the posterior distribution of μ given the data, $y = y_1, ..., y_T$. No MCMC algorithm or other such method is required because the derivations are analytic.

    (a) Write out the expression for the likelihood of the data, conditional on a value for the one unknown parameter:

    $$p(y|\mu).$$

    Show that one can write

    $$p(y|\mu) = \frac{1}{(2\pi\sigma^2)^{T/2}} \exp\left\{-\frac{1}{2}\frac{T\hat{\sigma}^2}{\sigma^2} - \frac{1}{2}\frac{(\bar{y}-\mu)^2}{\sigma^2/T}\right\},$$

    where $\bar{y}$ and $\hat{\sigma}^2$ denote the sample mean and variance of y_t :

    $$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t,\ \hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2.$$

    Hint:

    $$\sum_{t=1}^{T}(y_t-\mu)^2 = T\hat{\sigma}^2 + T(\bar{y}-\mu)^2.$$

(b) Show that

$$p\left(\mu|y\right) \propto \exp\left\{-\frac{1}{2}\frac{\left(\mu - m^*\right)^2}{\frac{\sigma^2}{T+\nu}}\right\},$$

where $\propto$ means 'is proportional to', where the factor of proportionality does not involve μ. Also,

$$m^* = \bar{y}\left(\frac{T}{T+\nu}\right) + m\left(\frac{\nu}{T+\nu}\right).$$

Thus, the posterior is Normal with mean equal to a weighted average of the sample mean of the data and the mean of the prior. The relative weight on the prior is an increasing function of the precision of the prior distribution.

Hint: note that

$$\frac{\left(\bar{y} - \mu\right)^2}{\sigma^2/T} + \frac{\left(\mu - m\right)^2}{\sigma^2/\nu} = \frac{\left(\mu^2 - 2\mu\bar{y}\right)\frac{\sigma^2}{\nu} + \left(\mu^2 - 2\mu m\right)\frac{\sigma^2}{T}}{\frac{\sigma^2}{T}\frac{\sigma^2}{\nu}} + X,$$

where X does not involve the unknown parameter, μ. With the given definition of m^*, the latter reduces further to

$$\frac{1}{\sigma^2}\left(T + \nu\right)\left(\mu - m^*\right)^2 + \tilde{X},$$

where $\tilde{X}$ does not involve μ. In this way, the joint distribution reduces to:

$$f\left(y, \theta\right) \propto \exp\left\{-\frac{1}{2}\frac{\left(\mu - m^*\right)^2}{\frac{\sigma^2}{T+\nu}}\right\}$$

End of hint.

2. This question explores the MCMC algorithm and the Laplace approximation in a simple example. Technical details about both these objects are discussed in lecture notes.

   To understand the workings of the MCMC algorithm, it is useful to see how it works when you know the function you are trying to approximate. Thus, consider the Weibull probability distribution function (pdf),

$$ba^{-b}\theta^{b-1}e^{-\left(\frac{\theta}{a}\right)^b},\ \theta \geq 0,$$

where a, b are parameters and θ is the random variable. (For an explanation of this pdf, see the MATLAB documentation for $wblpdf(\theta, a, b)$.) Consider $a = 10$, $b = 20$. Graph this pdf over the grid, $[7, 11.5]$, with intervals 0.001 (i.e., graph g on the vertical axis, where $g = wblpdf(x, 10, 20)$, and x on the horizontal axis, where $x = 7 : .001 : 11.5$). Compute the mode of this pdf by finding the element in your grid with the highest value of g. Let f denote the log of the Weibull density function and compute the second derivative of f at the mode point numerically, using the formula,

$$f''(x) = \frac{f(x + 2\varepsilon) - 2f(x) + f(x - 2\varepsilon)}{4\varepsilon^2},$$

for ε small (for example, you could set $\varepsilon = 0.000001$.) Here, x denotes θ^* and f denotes the log of the output of the MATLAB function, *wblpdf*. Set $V = -f''(\theta^*)^{-1}$.[1]

(a) Program up the MCMC algorithm in MATLAB (as indicated by the handout, this is quite simple). Set $M = 1{,}000$ (a very small number!) and try $k = 2, 4, 6$. In each case, generate $\theta^{(1)}, ..., \theta^{(M)}$ as discussed in class. Which value of k implies an acceptance rate closest to the recommended value of around 0.23? Graph the histogram of $\theta^{(1)}, ..., \theta^{(M)}$ and compare that to the actual density function (careful, the density function has the property that the area under the curve is unity and you should normalize the height of the histogram so that it has the same property, otherwise the density function and histogram will be off by an order of magnitude...also, it is best to graph the histogram as a curve rather than bar chart). Note that the MCMC estimate of the distribution is quite volatile (in practice, a device is used to smooth it out). Now,

[1] The strategy for computing the mode of the Weibull and f'' in the text are meant to resemble what is done in practice, when the form of the density function is unknown. In the case of the Weibull, these objects are straightforward to compute analytically. In particular,

$$f'(\theta) = \frac{b-1}{\theta} - b\left(\frac{\theta}{a}\right)^{b-1}\frac{1}{a},\ f''(\theta) = -\frac{b-1}{\theta^2} - (b-1)\,b\left(\frac{\theta}{a}\right)^{b-2}\frac{1}{a^2}.$$

and the mode of f is $\theta^* = ((b-1)/b)^{1/b}\, a$.

set $M = 10,000,000$. You should see that the MCMC estimate of the Weibull and the actual Weibull roughly coincide. Does the value of k make much difference?

Also, graph the Laplace approximation of the Weibull. How well does that do?

(b) We subject the MCMC algorithm to a much tougher test if we posit that the true distribution is bimodal, as in the case of a mixture of two Normals. Suppose the i^{th} Normal has mean and variance, μ_i and σ_i^2, respectively, $i = 1, 2$. Suppose also that the $i = 1$ Normal is selected with probability, π, and the $i = 2$ normal is selected with probability $1 - \pi$. In addition, suppose

$$\mu_1 = -0.06,\ \mu_2 = 0.06,\ \sigma_1 = 0.02,\ \sigma_2 = 0.01,\ \pi = 1/2.$$

The mode of this distribution is the mode of the Normal with $i = 2$. Graph the mixture of Normals distribution and compare it to what you get with the MCMC algorithm. How does the MCMC algorithm work if you make the variance of the jump distribution, V in the lecture, equal to unity? What if you start the algorithm at $\theta^{(1)}$ different from the mode? Graph the Laplace approximation of the mixture of Normals.

3. Following are the equations of the Clarida-Gali-Gertler model.

$$\begin{aligned}
\pi_t &= \beta E_t \pi_{t+1} + \kappa x_t \text{ (Calvo pricing equation)} \\
x_t &= -\left[r_t - E_t \pi_{t+1} - r_t^*\right] + E_t x_{t+1} \text{ (intertemporal equation)} \\
r_t &= \alpha r_{t-1} + (1-\alpha)\left[\phi_\pi \pi_t + \phi_x x_t\right] \text{ (policy rule)} \\
r_t^* &= \rho \Delta a_t + \frac{1}{1+\varphi}(1-\lambda)\tau_t \text{ (natural rate)} \\
y_t^* &= a_t - \frac{1}{1+\varphi}\tau_t \text{ (natural output)} \\
x_t &= y_t - y_t^* \text{ (output gap)} \\
\Delta a_t &= \rho \Delta a_{t-1} + \varepsilon_t^a,\ \tau_t = \lambda \tau_{t-1} + \varepsilon_t^\tau
\end{aligned}$$

Consider the following model parameterization:

$$\begin{aligned}
\beta &= 0.97,\ \phi_x = 0.15,\ \phi_\pi = 1.5,\ \alpha = 0.8,\ \rho = 0.9,\ \lambda = 0.5, \\
\varphi &= 1,\ \theta = 0.75,\ \sigma_a = \sigma_\tau = 0.02.
\end{aligned}$$

Generate $T = 5000$ artificial observations on the 'endogenous' (in the sense of Dynare) variables of the model. These are the variables in the 'var' list. Before doing the simulation, you should add the growth rate of output to the equations of the model and to the var list (call it 'dy'.) That way, Dynare will also simulate output growth. The variables simulated by Dynare are placed in the $n \times T$ matrix, oo_.endo_simul.[2] The n rows of oo_.endo_simul correspond to the $n = 7$ variables in var, *listed in the order in which you have listed them in the var statement* from the first to the last row. To verify the order that Dynare puts the variables in, see how they are ordered in Dynare-created structure, M_.endo_names. Now do Bayesian estimation for the four parameters, $\sigma_a, \sigma_\tau, \lambda, \rho$, using the inverted gamma distribution as the prior on the two standard deviations and the beta distribution as the prior on the two autocorrelations.

(a) Set the mean of the priors over the parameters to the corresponding true values. Set the standard deviation of the inverted gamma to 10 and of the beta to 0.04. (It's hard to interpret these standard deviations directly, but you will see graphs of the priors, which are easier to interpret.) Use 30 observations in the estimation. Adjust the value of k, so that you get a reasonable acceptance rate. I found that $k = 1.5$ works well. Have a look at the posteriors, and notice how, with one exception, they are much tighter than the priors. The exception is lambda, where the posterior and prior are very similar. This is evidence that there is little information in the data about lambda.

(b) Redo (a), but set the mean and standard deviation of the prior on lambda equal to 0.95 and 0.04, respectively. Note how the prior and posterior are again very similar. There is not much information in the data about the value of lambda!

(c) Note how the priors on σ_a and ρ have faint 'shoulders' on the right side. Redo (a), with $M = 4,000$ (M is mh_replic, which controls the number of MCMC replications). Note that the posteriors are now smoother. Actually, $M = 4,000$ is a small number of replications to use in practice.

[2] Here, endo_simul is the matrix, which is a 'field' in the structure, oo_.

(d) Now set the mean of the priors on the standard deviations to 0.1, far from the truth. Set the prior standard deviation on the inverted gamma distributions to 1. Keep the observations at 30, and see how the posteriors compare with the priors. (Reset $M = 1,000$ so that the computations go quickly.) Note that the posteriors move sharply back into the neighborhood of 0.02. Evidently, there is a lot of information in the data about these parameters.

(e) Repeat (a) with 4,000 observations. Compare the priors and posteriors. Note how, with one exception, the posteriors are 'spikes'. The exception, of course, is lambda. Still, the difference between the prior and posterior in this case indicates there is information in the data about lambda.