Bayesian Inference for DSGE Models

Lawrence J. Christiano

Outline

- Bayes' rule.
- Monte Carlo integation: a simple example.
- Markov Chain Monte Carlo (MCMC) algorithm.
- Laplace approximation

Bayesian Inference

- Bayesian inference is about describing the mapping from prior beliefs about θ, summarized in $p(\theta)$, to new posterior beliefs in the light of observing the data, Y^{data}.
- General property of probabilities:

$$p\left(Y^{data},\theta\right)=\left\{\begin{array}{c} p\left(Y^{data}|\theta\right)\times p\left(\theta\right) \\ p\left(\theta|Y^{data}\right)\times p\left(Y^{data}\right) \end{array}\right. ,$$

which implies Bayes' rule:

$$p\left(\theta|Y^{data}\right)=\frac{p\left(Y^{data}|\theta\right)p\left(\theta\right)}{p\left(Y^{data}\right)},$$

mapping from prior to posterior induced by Y^{data}.

Bayesian Inference

- Report features of the posterior distribution, $p\left(\theta|Y^{data}\right)$.
  - The value of θ that maximizes $p\left(\theta|Y^{data}\right)$, 'mode' of posterior distribution.
  - Compare marginal prior, $p\left(\theta_i\right)$, with marginal posterior of individual elements of θ, $g\left(\theta_i|Y^{data}\right)$:

$$g\left(\theta_i|Y^{data}\right) = \int_{\theta_{j\neq i}} p\left(\theta|Y^{data}\right) d\theta_{j\neq i} \text{ (multiple integration!!)}$$

  - Probability intervals about the mode of θ ('Bayesian confidence intervals'), need $g\left(\theta_i|Y^{data}\right)$.

- Marginal likelihood for assessing model 'fit':

$$p\left(Y^{data}\right) = \int_{\theta} p\left(Y^{data}|\theta\right) p\left(\theta\right) d\theta \text{ (multiple integration)}$$

Monte Carlo Integration: Simple Example

- Much of Bayesian inference is about multiple integration.
- Numerical methods for multiple integration:
  - Quadrature integration (example: approximating the integral as the sum of the areas of triangles beneath the integrand).
  - Monte Carlo Integration: uses random number generator.
- Example of Monte Carlo Integration:
  - suppose you want to evaluate

$$\int_a^b f(x)\,dx,\ -\infty \leq a < b \leq \infty.$$

  - select a density function, $g(x)$ for $x \in [a,b]$ and note:

$$\int_a^b f(x)\,dx = \int_a^b \frac{f(x)}{g(x)} g(x)\,dx = E\frac{f(x)}{g(x)},$$

where E is the expectation operator, given $g(x)$.

Monte Carlo Integration: Simple Example

- Previous result: can express an integral as an expectation relative to a (arbitrary, subject to obvious regularity conditions) density function.
- Use the law of large numbers (LLN) to approximate the expectation.
  - step 1: draw x_i independently from density, g, for $i = 1, ..., M$.
  - step 2: evaluate $f(x_i)/g(x_i)$ and compute:

$$\mu_M \equiv \frac{1}{M} \sum_{i=1}^{M} \frac{f(x_i)}{g(x_i)} \rightarrow_{M\rightarrow\infty} E\frac{f(x)}{g(x)}.$$

- Exercise.
  - Consider an integral where you have an analytic solution available, e.g., $\int_0^1 x^2 dx$.
  - Evaluate the accuracy of the Monte Carlo method using various distributions on $[0,1]$ like uniform or Beta.

Monte Carlo Integration: Simple Example

- Standard classical sampling theory applies.
- Independence of $f(x_i)/g(x_i)$ over i implies:

$$var\left(\frac{1}{M}\sum_{i=1}^{M}\frac{f(x_i)}{g(x_i)}\right) = \frac{v_M}{M},$$

$$v_M \equiv var\left(\frac{f(x_i)}{g(x_i)}\right) \simeq \frac{1}{M}\sum_{i=1}^{M}\left[\frac{f(x_i)}{g(x_i)} - \mu_M\right]^2.$$

- Central Limit Theorem
  - Estimate of $\int_a^b f(x)\,dx$ is a realization from a Nomal distribution with mean estimated by μ_M and variance, v_M/M.
  - With 95% probability,

$$\mu_M - 1.96 \times \sqrt{\frac{v_M}{M}} \ \le \ \int_a^b f(x)\,dx \ \le \ \mu_M + 1.96 \times \sqrt{\frac{v_M}{M}}$$

  - Pick g to minimize variance in $f(x_i)/g(x_i)$ and M to minimize (subject to computing cost) v_M/M.

Markov Chain, Monte Carlo (MCMC) Algorithms

- Among the top 10 algorithms "with the greatest influence on the development and practice of science and engineering in the 20th century".
  - Reference: January/February 2000 issue of Computing in Science & Engineering, a joint publication of the American Institute of Physics and the IEEE Computer Society.'
- Developed in 1946 by John von Neumann, Stan Ulam, and Nick Metropolis (see http://www.siam.org/pdf/news/637.pdf)

MCMC Algorithm: Overview

- compute a sequence, $\theta^{(1)}, \theta^{(2)}, ..., \theta^{(M)}$, of values of the $N \times 1$ vector of model parameters in such a way that

$$\lim_{M \to \infty} Frequency \left[\theta^{(i)} \text{ close to } \theta \right] = p \left(\theta | Y^{data} \right).$$

- Use $\theta^{(1)}, \theta^{(2)}, ..., \theta^{(M)}$ to obtain an approximation for
  - $E\theta$, $Var(\theta)$ under posterior distribution, $p\left(\theta|Y^{data}\right)$
  - $g\left(\theta^i|Y^{data}\right) = \int_{\theta_{i \neq j}} p\left(\theta|Y^{data}\right) d\theta d\theta$
  - $p\left(Y^{data}\right) = \int_{\theta} p\left(Y^{data}|\theta\right) p(\theta) d\theta$
  - posterior distribution of any function of θ, $f(\theta)$ (e.g., impulse responses functions, second moments).
- MCMC also useful for computing posterior mode, $\arg\max_{\theta} p\left(\theta|Y^{data}\right)$.

MCMC Algorithm: setting up

- Let $G(\theta)$ denote the log of the posterior distribution (excluding an additive constant):

$$G(\theta) = \log p\left(Y^{data}|\theta\right) + \log p(\theta);$$

- Compute posterior mode:

$$\theta^* = \arg\max_{\theta} G(\theta).$$

- Compute the positive definite matrix, V :

$$V \equiv \left[-\frac{\partial^2 G(\theta)}{\partial\theta\partial\theta'}\right]^{-1}_{\theta=\theta^*}$$

- Later, we will see that V is a rough estimate of the variance-covariance matrix of θ under the posterior distribution.

MCMC Algorithm: Metropolis-Hastings

- $\theta^{(1)} = \theta^*$
- to compute $\theta^{(r)}$, for $r > 1$
  - step 1: select candidate $\theta^{(r)}$, x,

$$\text{draw } \underbrace{x}_{N\times 1} \text{ from } \theta^{(r-1)} + \overbrace{k \times N\left(\underbrace{0}_{N\times 1}, V\right)}^{\text{'jump' distribution'}}, \; k \text{ is a scalar}$$

  - step 2: compute scalar, λ :

$$\lambda = \frac{p\left(Y^{data}|x\right)p\left(x\right)}{p\left(Y^{data}|\theta^{(r-1)}\right)p\left(\theta^{(r-1)}\right)}$$

  - step 3: compute $\theta^{(r)}$:

$$\theta^{(r)} = \begin{cases} \theta^{(r-1)} & \text{if } u > \lambda \\ x & \text{if } u < \lambda \end{cases}, \; u \text{ is a realization from uniform } [0,1]$$

Practical issues

- What is a sensible value for k?
  - set k so that you accept (i.e., $\theta^{(r)} = x$) in step 3 of MCMC algorithm are roughly 23 percent of time
- What value of M should you set?
  - want 'convergence', in the sense that if M is increased further, the econometric results do not change substantially
  - in practice, $M = 10,000$ (a small value) up to $M = 1,000,000$.
  - large M is time-consuming.
    - could use Laplace approximation (after checking its accuracy) in initial phases of research project.
    - more on Laplace below.
- Burn-in: in practice, some initial $\theta^{(i)}$'s are discarded to minimize the impact of initial conditions on the results.
- Multiple chains: may promote efficiency.
  - increase independence among $\theta^{(i)}$'s.
  - can do MCMC utilizing parallel computing (Dynare can do this).

MCMC Algorithm: Why Does it Work?

- Proposition that MCMC works may be surprising.
  - Whether or not it works does *not* depend on the details, i.e., precisely how you choose the jump distribution (of course, you had better use $k > 0$ and V positive definite).
    - Proof: see, e.g., Robert, C. P. (2001), *The Bayesian Choice*, Second Edition, New York: Springer-Verlag.
  - The details may matter by improving the efficiency of the MCMC algorithm, i.e., by influencing what value of M you need.
- Some Intuition
  - the sequence, $\theta^{(1)}, \theta^{(2)}, ..., \theta^{(M)}$, is relatively heavily populated by θ's that have high probability and relatively lightly populated by low probability θ's.
  - Additional intuition can be obtained by positing a simple scalar distribution and using MATLAB to verify that MCMC approximates it well (see, e.g., question 2 in assignment 9).

MCMC Algorithm: using the Results

- To approximate marginal posterior distribution, $g\left(\theta_i | Y^{data}\right)$, of θ_i,
  - compute and display the histogram of $\theta_i^{(1)}, \theta_i^{(2)}, ..., \theta_i^{(M)}$, $i = 1, ..., M$.
- Other objects of interest:
  - mean and variance of posterior distribution θ :

$$E\theta \simeq \bar{\theta} \equiv \frac{1}{M}\sum_{j=1}^{M}\theta^{(j)},\ Var\left(\theta\right) \simeq \frac{1}{M}\sum_{j=1}^{M}\left[\theta^{(j)} - \bar{\theta}\right]\left[\theta^{(j)} - \bar{\theta}\right]'.$$

MCMC Algorithm: using the Results

- More complicated objects of interest:
  - impulse response functions,
  - model second moments,
  - forecasts,
  - Kalman smoothed estimates of real rate, natural rate, etc.
- All these things can be represented as non-linear functions of the model parameters, i.e., $f(\theta)$.
  - can approximate the distribution of $f(\theta)$ using

$$
\begin{aligned}
& f\left(\theta^{(1)}\right), \ldots, f\left(\theta^{(M)}\right) \\
\rightarrow \quad & Ef(\theta) \simeq \bar{f} \equiv \frac{1}{M} \sum_{i=1}^{M} f\left(\theta^{(i)}\right), \\
Var(f(\theta)) \quad \simeq \quad & \frac{1}{M} \sum_{i=1}^{M} \left[f\left(\theta^{(i)}\right) - \bar{f}\right] \left[f\left(\theta^{(i)}\right) - \bar{f}\right]'
\end{aligned}
$$

MCMC: Remaining Issues

- In addition to the first and second moments already discused, would also like to have the marginal likelihood of the data.
- Marginal likelihood is a Bayesian measure of model fit.

MCMC Algorithm: the Marginal Likelihood

- Consider the following sample average:

$$\frac{1}{M}\sum_{j=1}^{M}\frac{h\left(\theta^{(j)}\right)}{p\left(Y^{data}|\theta^{(j)}\right)p\left(\theta^{(j)}\right)},$$

  where $h(\theta)$ is an arbitrary density function over the $N-$ dimensional variable, θ.

By the law of large numbers,

$$\frac{1}{M}\sum_{j=1}^{M}\frac{h\left(\theta^{(j)}\right)}{p\left(Y^{data}|\theta^{(j)}\right)p\left(\theta^{(j)}\right)}\underset{M\rightarrow\infty}{\rightarrow}E\left(\frac{h(\theta)}{p\left(Y^{data}|\theta\right)p(\theta)}\right)$$

MCMC Algorithm: the Marginal Likelihood

$$
\frac{1}{M}\sum_{j=1}^{M}\frac{h\left(\theta^{(j)}\right)}{p\left(Y^{data}|\theta^{(j)}\right)p\left(\theta^{(j)}\right)} \rightarrow_{M\rightarrow\infty} E\left(\frac{h\left(\theta\right)}{p\left(Y^{data}|\theta\right)p\left(\theta\right)}\right)
$$

$$
=\int_{\theta}\left(\frac{h\left(\theta\right)}{p\left(Y^{data}|\theta\right)p\left(\theta\right)}\right)\frac{p\left(Y^{data}|\theta\right)p\left(\theta\right)}{p\left(Y^{data}\right)}d\theta \;\; =\frac{1}{p\left(Y^{data}\right)}.
$$

- When $h\left(\theta\right)=p\left(\theta\right)$, *harmonic mean estimator of the marginal likelihood*.
- Ideally, want an h such that the variance of

$$
\frac{h\left(\theta^{(j)}\right)}{p\left(Y^{data}|\theta^{(j)}\right)p\left(\theta^{(j)}\right)}
$$

  is small (recall the earlier discussion of Monte Carlo integration). More on this below.

Laplace Approximation to Posterior Distribution

- In practice, MCMC algorithm very time intensive.

- Laplace approximation is easy to compute and in many cases it provides a 'quick and dirty' approximation that is quite good.

Let $\theta \in R^N$ denote the $N-$dimensional vector of parameters and, as before,

$$G(\theta) \equiv \log p\left(Y^{data}|\theta\right)p(\theta)$$

$$p\left(Y^{data}|\theta\right) \text{ ~likelihood of data}$$

$$p(\theta) \text{ ~prior on parameters}$$

$$\theta^* \text{ ~maximum of } G(\theta) \text{ (i.e., mode)}$$

Laplace Approximation

Second order Taylor series expansion of
$G(\theta) \equiv \log \left[p\left(Y^{data}|\theta\right) p(\theta) \right]$ about $\theta = \theta^*$:

$$G(\theta) \approx G(\theta^*) + G_\theta(\theta^*)(\theta - \theta^*) - \frac{1}{2}(\theta - \theta^*)' G_{\theta\theta}(\theta^*)(\theta - \theta^*),$$

where

$$G_{\theta\theta}(\theta^*) = -\frac{\partial^2 \log p\left(Y^{data}|\theta\right) p(\theta)}{\partial\theta\partial\theta'}|_{\theta=\theta^*}$$

Interior optimality of θ^* implies:

$$G_\theta(\theta^*) = 0,\ G_{\theta\theta}(\theta^*) \text{ positive definite}$$

Then:

$$\begin{aligned} & p\left(Y^{data}|\theta\right) p(\theta) \\ \simeq \ & p\left(Y^{data}|\theta^*\right) p(\theta^*) \exp\left\{ -\frac{1}{2}(\theta - \theta^*)' G_{\theta\theta}(\theta^*)(\theta - \theta^*) \right\}. \end{aligned}$$

Laplace Approximation to Posterior Distribution

Property of Normal distribution:

$$\int_\theta \frac{1}{(2\pi)^{\frac{N}{2}}} \left|G_{\theta\theta}(\theta^*)\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\theta-\theta^*)' G_{\theta\theta}(\theta^*)(\theta-\theta^*)\right\} d\theta = 1$$

Then,

$$\begin{aligned}
\int p\left(Y^{data}|\theta\right) p(\theta)\, d\theta &\simeq \int p\left(Y^{data}|\theta^*\right) p(\theta^*) \\
&\quad \times \exp\left\{-\frac{1}{2}(\theta-\theta^*)' G_{\theta\theta}(\theta^*)(\theta-\theta^*)\right\} d\theta \\
&= \frac{p\left(Y^{data}|\theta^*\right) p(\theta^*)}{\frac{1}{(2\pi)^{\frac{N}{2}}} \left|G_{\theta\theta}(\theta^*)\right|^{\frac{1}{2}}}.
\end{aligned}$$

Laplace Approximation

- Conclude:

$$p\left(Y^{data}\right) \simeq \frac{p\left(Y^{data}|\theta^*\right)p\left(\theta^*\right)}{\frac{1}{(2\pi)^{\frac{N}{2}}}\left|G_{\theta\theta}\left(\theta^*\right)\right|^{\frac{1}{2}}}.$$

- Laplace approximation to posterior distribution:

$$\frac{p\left(Y^{data}|\theta\right)p\left(\theta\right)}{p\left(Y^{data}\right)} \simeq \frac{1}{(2\pi)^{\frac{N}{2}}}\left|G_{\theta\theta}\left(\theta^*\right)\right|^{\frac{1}{2}} \times \exp\left\{-\frac{1}{2}\left(\theta-\theta^*\right)' G_{\theta\theta}\left(\theta^*\right)\left(\theta-\theta^*\right)\right\}$$

- So, posterior of θ_i (i.e., $g\left(\theta_i|Y^{data}\right)$) is approximately

$$\theta_i \sim N\left(\theta_i^*, \left[G_{\theta\theta}\left(\theta^*\right)^{-1}\right]_{ii}\right).$$

Modified Harmonic Mean Estimator of Marginal Likelihood

- Harmonic mean estimator of the marginal likelihood, $p\left(Y^{data}\right)$:

$$\left[\frac{1}{M}\sum_{j=1}^{M}\frac{h\left(\theta^{(j)}\right)}{p\left(Y^{data}|\theta^{(j)}\right)p\left(\theta^{(j)}\right)}\right]^{-1},$$

  with $h\left(\theta\right)$ set to $p\left(\theta\right)$.
  - In this case, the marginal likelihood is the harmonic mean of the likelihood, evaluated at the values of θ generated by the MCMC algorithm.
  - Problem: the variance of the object being averaged is likely to be high, requiring high M for accuracy.
- When $h\left(\theta\right)$ is instead equated to Laplace approximation of posterior distribution, then $h\left(\theta\right)$ is approximately proportional to $p\left(Y^{data}|\theta^{(j)}\right)p\left(\theta^{(j)}\right)$ so that the variance of the variable being averaged in the last expression is low.

The Marginal Likelihood and Model Comparison

- Suppose we have two models, $Model$ 1 and $Model$ 2.
  - compute $p\left(Y^{data}|Model\ 1\right)$ and $p\left(Y^{data}|Model\ 2\right)$
- Suppose $p\left(Y^{data}|Model\ 1\right) > p\left(Y^{data}|Model\ 2\right)$. Then, posterior odds on Model 1 higher than Model 2.
  - 'Model 1 fits better than Model 2'
- Can use this to compare across two different models, or to evaluate contribution to fit of various model features: habit persistence, adjustment costs, etc.
  - For an application of this and the other methods in these notes, see Smets and Wouters, AER 2007.