Bayesian Econometrics: Meet-up 7

Kevin Arthur Tittel

April 18, 2023

1 Recap Meet-up 6

In this Section we briefly review some main take-aways from Meet-up 6. We continued our discussion on numerical methods used to obtain posterior results (by solving necessary integrals, i.e. $\mathbb{E}_{\theta|y}[h(\theta)] = \int_{-\infty}^{\infty} h(\theta)p(\theta|y)d\theta$) in case it is not possible to obtain these analytically (e.g. posterior means, modes or variances of logit or probit models). In particular, we focused on high-dimensional integrals, where Monte Carlo integration can be used. Monte Carlo integration requires methods that are able to simulate a set of draws $\{\theta^{(m)}, m = 1, ..., M\}$ from the posterior distribution $p(\theta|y)$. Subsequently, the integral of interest can be approximated by Monte Carlo integration (i.e. $\mathbb{E}_{\theta|y}[h(\theta)] \approx \frac{1}{M}\sum_{m=1}^{M} h(\theta^{(m)})$).

In terms of sampling methods for Monte Carlo integration, we have, till now, discussed importance sampling and Gibbs sampling, a type of MCMC sampling. A Gibbs sampler entails an iterative simulation procedure, where the posterior function is obtained in the limit (*True or False Question 4*). The Gibbs sampler divides the parameter of interest θ into d blocks $(\theta_1, ..., \theta_d)$, where for each of these blocks we derive the full conditional posterior distributions $p(\theta_1|\theta_2, ..., \theta_d, y)$, ..., $p(\theta_d|\theta_1, ..., \theta_{d-1}, y)$. Since the Gibbs sampler is an MCMC sampler, the Markov chain components appears through the marginal distributions: for the next iteration $(m + 1)$, we simulate draws $\theta_1^{(m+1)}, ..., \theta_d^{(m+1)}$ from $p(\theta_1|\theta_2^{(m)}, ..., \theta_d^{(m)}, y)$, ..., $p(\theta_d|\theta_1^{(m+1)}, ..., \theta_{d-1}^{(m+1)}, y)$, respectively. After point of convergence $m = m^*$, all draws $\{\theta^m, m \geq m^*\}$ can be used as a sample from the joint posterior $p(\theta_1, ..., \theta_d|y)$.

Important aspects to consider when simulating are i) the simulation error and ii) sampler convergence. First, the error margin of the simulation procedure can be computed for each posterior quantity of interest, given whether the draws are independent (e.g. importance sampling) or correlated (e.g. MCMC sampling). Correlated sample draws tend to cause a larger simulation error variance (*True or False Question 6, Meet-up 5*). Hence, the feature of thinning could be applied, where in a correlated set of draws only each 1 of (thinning value) k draws are kept, with the remaining draws deleted. As a result, to keep an effective sample size EES of the same size M, a higher amount of M^* should be drawn, increasing with a higher k (*True or False Question 5, Meet-up 5*). Trivially, the larger the correlation between subsequent sample draws and/or thinning value, the slower the convergence of the Markov chain (*True or False Question 7, Meet-up 5*). As an example, for a large number of draws M, the Monte Carlo estimate of the mean equals $\mathbb{E}_{\theta|y}[h(\theta)] \approx \frac{1}{M}\sum_{m=1}^{M} h(\theta^{(m)})$, where the sampling error decreases with $\sqrt{M}$ (or, $\sqrt{M/(1+2\sum_{j=1}^{M}\rho_j)}$) for independent (or, correlated) draws (*True or False Question 8, Meet-up 5*).

Second, convergence of a sampler can be assessed using informal tools such as traceplots, or formal tests, such as the Geweke test for single Markov chains, or the Gelman-Rubin test for multiple chains.

2 Questions

For each of the following Gibbs samplers, find the mistakes in the sampling procedure.

Gibbs sampler 1. Linear regression model $y = X\beta+\epsilon, \epsilon \sim N(0, \sigma^2 I_N)$, with independent uninformative priors for β and σ^2, $p(\beta) \propto 1$ and $p(\sigma^2) \propto \sigma^{-2}$:

1. Starting value: set $\beta^{(0)}$ equal to starting value (e.g. $\hat{\beta}$) and set $m = 0$;
2. Simulate $\sigma^{2(m+1)}$ from $p(\sigma^2|\beta^{(m+1)}, y)$, which is an inverted Gamma-2 distribution, so use: $\frac{(y-X\beta^{(m+1)})'(y-X\beta^{(m+1)})}{\sigma^2} \sim \chi^2(N)$;
3. Simulate $\beta^{(m+1)}$ from $p(\beta|\sigma^{2(m+1)}, y)$, which is a multivariate t-distribution: $\beta^{(m+1)} \sim t(\hat{\beta}, \sigma^{2(m+1)}(X'X)^{-1}, N-k)$;
4. Set $m = m + 1$, and go to step 2.

Gibbs sampler 2. Linear regression model $y = X\beta + \epsilon, \epsilon \sim N(0, \sigma^2 I_N)$, with a natural conjugate prior for $\beta|\sigma^2$ and a diffuse prior for σ^2, $p(\beta|\sigma^2) \sim N(b, \sigma^2 B)$ and $p(\sigma^2) \propto \sigma^{-2}$:

1. Starting value: set $\sigma^{(0)}$ equal to starting value (e.g. use $(y - X\beta)'(y - X\beta)/\sigma^2 \sim \chi^2(N + k)$) and set $m = 0$;
2. Simulate $\beta^{(m+1)}$ from $p(\beta|\sigma^{2(m)})$, which is a multivariate normal distribution: $\beta^{(m+1)} \sim N(\hat{\beta}, \sigma^{2(m)}(X'X)^{-1})$;
3. Simulate $\sigma^{2(m+1)}$ from $p(\sigma^2|\beta^{(m+1)})$, which is an inverted Gamma-2 distribution, so use: $\frac{(y-X\beta^{(m+1)})'(y-X\beta^{(m+1)})+(b-\beta^{(m+1)})'B^{-1}(b-\beta^{(m+1)})}{\sigma^2} \sim \chi^2(N + k)$;
4. Set $m = m + 1$, and go to step 2.

3 Discussion Theory

In this Section we continue to discuss theory on MCMC samplers, given the slides of Lecture 4. Having discussed the Gibbs sampler previously, we now move on to the Metropolis-Hastings algorithm.

What is Metropolis-Hastings sampling? Whereas the Gibbs sampler iteratively draws from the known, full conditional distributions of the parameters, $p(\theta_i|\theta_{i-1}, y)$, where $\theta_{-i} = (\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_d)$, Metropolis-Hastings sampling can be applied when these full conditional distributions $p(\theta_i|\theta_{-i}, y)$ (or, the posterior distribution $p(\theta|y)$) turn out to be unknown. Similar to importance sampling, the unknown density function $f(\theta)$ will be approximated by a similar, easier and known function; the so-called candidate-generating density function, $g(\theta|\theta^{(m)})$.

What is the MH-procedure? Since the MH-algorithm is a Markov procedure, the iterative procedure depends on the previous draw $\theta^{(m)}$. Namely, given starting values $\theta^{(0)}$ and $m = 0$, a candidate draw θ^* is simulated from $g(\theta|\theta^{(m)}$. Then, this candidate draw is accepted to be the next official draw $\theta^{(m+1)}$ with probability α, or rejected (and replaced by $\theta^{(m)}$) with probability $1 - \alpha$. The acceptance probability is then determined as follows: the candidate ratio $f(\theta^*)/g(\theta^*|\theta^{(m)})$ is divided by the old ratio $f(\theta^{(m)})/g(\theta^{(m)}|\theta^*)$, and the larger the candidate ratio compared to the old ratio, the larger the acceptance probability. Or, mathematically: $\alpha = \min(\frac{f(\theta^*)g(\theta^{(m)}|\theta^*)}{f(\theta^{(m)})g(\theta^*|\theta^{(m)})}, 1)$.