Bayesian Econometrics: Meet-up 6

Kevin Arthur Tittel

March 29, 2023

1 Recap Meet-up 5

In this Section we briefly review some main take-aways from Meet-up 5. We started to discuss the theory 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).

For low-dimensional integrals, numerical integration can be used due to its fast convergence with small approximation errors. For high-dimensional integrals, Monte Carlo integration can be used, as its computational burden does not exponentially depend on the dimension of the integral, despite its relatively slower convergence with larger approximation errors. 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. $\mathbb{E}_{\theta|y}[h(\theta)] \approx \frac{1}{M}\sum_{m=1}^{M} h(\theta^{(m)})$).

Sampling methods for Monte Carlo integration include i) importance sampling and ii) MCMC methods (e.g. Gibbs sampling, Metropolis-Hastings method). In case the posterior distribution is difficult to sample directly from, but an easier, so-called importance function $g(\theta|y)$ exists, then we use importance sampling and sample $\theta^{(m)}$ from $g(\theta)$. This importance function can be used to sample from the full posterior density function $p(\theta|y)$ with all integrating constants, or from the kernel function $k(\theta|y) = p(\theta)p(y|\theta)$ (*True or False Question 1*). For each sampled draw $\theta^{(m)}$ we compute its importance weight $w(\theta^{(m)})$, which indicates how good of an approximation the draw is for solving the necessary integral, where weights of bad approximations are close to zero (*True or False Question 2*). When computing importance weights and constructing the Monte Carlo integration expression, we distinguish two cases:

- In case the importance function aims to approximate the posterior density function, importance weights are defined as $w(\theta^{(m)}) = p(\theta|y)/g(\theta)$, and the original expression changes to $\mathbb{E}_{\theta|y}[h(\theta)] \approx \frac{1}{M}\sum_{m=1}^{M} w(\theta^{(m)})h(\theta^{(m)})$;
- In case the importance function aims to approximate the kernel function, importance weights are defined as $w(\theta^{(m)}) = p(\theta)p(y|\theta)/g(\theta)$, and the original expression changes to $\mathbb{E}_{\theta|y}[h(\theta)] \approx \frac{1}{M}\sum_{m=1}^{M} \frac{w(\theta^{(m)})h(\theta^{(m)})}{w(\theta^{(m)})}$.

It can be seen that, in case one uses a non-conjugate yet informative prior $p(\beta)$ leading to an unknown posterior kernel $p(\theta)p(y|\theta)$, one may choose $g(\theta)$ to equal the proper likelihood function $p(y|\theta)$ such that importance weights simplify to $w(\theta^{(m)}) = p(\beta)$ (*True or False Question 3*).

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.