Bayesian Econometrics: Meet-up 8

Kevin Arthur Tittel

April 21, 2023

1 Recap Meet-up 7

In this Section we briefly review some main take-aways from Meet-up 7. We continued our discussion on MCMC sampling methods for Monte Carlo integration, in particular Gibbs sampling and Metropolis-Hastings sampling.

Regarding Gibbs sampling, we discussed 2 Gibbs samplers for linear regression models (with uninformative or conjugate priors). Both samplers started with setting starting values for either $\beta^{(0)}$ or $\sigma^{2(0)}$ (or both), and setting the simulation iteration to $m = 0$. Then, draws were simulated for $\sigma^{2(m+1)}$ and $\beta^{(m+1)}$ from inverted Gamma-2 distribution and multivariate normal distributions, respectively. In the details, a diffuse prior induced N degrees of freedom for the χ^2 distribution, opposed to $N + k$ under a conjugate prior. Next, a diffuse prior induced mean $\hat{\beta}$ and covariance $\sigma^2(X'X)^{-1}$ for the normal distribution, opposed to mean $\tilde{\beta}$ and covariance $\sigma^2(X'X + B^{-1})^{-1}$ under a conjugate distribution. Last, if $\beta^{(m+1)}$ (or, $\sigma^{2(m+1)}$) is simulated first in each next iteration $(m + 1)$, the other parameter's previous value $\sigma^{2(m)}$ (or, $\beta^{(m)}$) of iteration (m) is given and used.

Regarding Metropolis-Hastings sampling, we started discussing corresponding theory. Opposed to Gibbs sampling where full conditional distributions of the parameters, $p(\theta_i|\theta_{i-1}, y)$ with $\theta_{-i} = (\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_d)$, are known and used, Metropolis-Hastings assumes unknown full conditional posterior (or posterior) distributions. Similar to importance sampling, the unknown density $f(\theta)$ is approximated by a, so-called, candidate generating density function $g(\theta|\theta^{(m)})$. Then, one iteratively simulates from the candidate generating density function, only accepting a candidate draw (i.e. $\theta^{(m+1)} = \theta^*$) in case

θ^* is "better" compared to the old draw $\theta^{(m)}$. Intuitively, one is more likely to accept θ^* (with probability α) if the candidate ratio $f(\theta^*)/g(\theta^*|\theta^{(m)})$ is larger than the old ratio $f(\theta^{(m)})/g(\theta^{(m)}|\theta^*)$. Or: $\alpha = \min(\frac{f(\theta^*)g(\theta^{(m)}|\theta^*)}{f(\theta^{(m)})g(\theta^*|\theta^{(m)})}, 1)$.

2 Discussion Theory

In this Section we continue to discuss theory on the Metropolis-Hastings sampler, given the slides of Lecture 4.

What are independence and random walk samplers? In principle, the MH-algorithm is an MCMC sampler with a Markov component, that is, the candidate-generating density function depends on the previous draw $\theta^{(m)}$ through $g(\theta|\theta^{(m)}$. However, in case we ignore this Markov component, other choices for the candidate-generating function lead to different versions of the MH-algorithm:

- **Independence MH sampler**: If the candidate density does not depend on $\theta^{(m)}$ (i.e. $g(\theta|\theta^{(m)}) = g(\theta)$, then the candidate density is an independence sampler. Then, we accept the new draw θ^* if its corresponding candidate ratio $f(\theta^*)/g(\theta^*)$ is larger than the old ratio $f(\theta^{(m)})/g(\theta^{(m)})$, or: $\alpha = \min(\frac{f(\theta^*)g(\theta^{(m)})}{f(\theta^{(m)})g(\theta^*)}, 1)$. The closer α is to 100%, the closer the candidate draw is to the target distribution.

- **Random walk MH sampler**: If the candidate density obeys $g(\theta^*|\theta^{(m)}) = g(\theta^{(m)}|\theta^*)$, then the candidate density is a random walk sampler, i.e. $\theta^* = \theta^{(m)} + \eta$ where η follows a symmetric distribution around 0. You can use this if you are unable to find a good independent candidate-generating density function. We accept the new draw with acceptance probability $\alpha = \min \frac{f(\theta^*)}{f(\theta^{(m)})}$. NOTE: The optimal acceptance rate is NOT 100%, as small variance of η might lead to acceptance of every candidate θ^* without reaching the target distribution. Hence, rules of thumb advocate optimal acceptance rates of 50% (1-dimensional sampler) or 25% (multidimensional sampler).

3 Questions

Extra Exercise 4, 7; Exam Q2 2018.