

Computing Session 2: An Introduction to Monte Carlo Integration and Gibbs Sampling

These exercises are taken (or adapted) from the book Bayesian Econometric Methods (BEM) and Matlab code can be found on the BEM website (you can connect through the link for the book on my website or directly to www.econ.iastate.edu/faculty/tobias/)

Exercise 1: Drawing from Standard Distributions.

Simulation-based inference via Monte Carlo integration or the Gibbs sampler requires the researcher to be able to draw from standard distributions. In this exercise we discuss how MATLAB can be used to obtain draws from a variety of standard continuous distributions. Specifically, we obtain draws from the Uniform, Normal, Student-t, Beta, Exponential and Chi-squared distributions, using MATLAB (see the Appendix to BEM or BE for definitions of these distributions). This exercise is designed to be illustrative - MATLAB is capable of generating variates from virtually any distribution that an applied researcher will encounter (and the same applies to other relevant computer languages such as Gauss).

Using MATLAB, obtain sets of 10, 100 and 100,000 draws from the Uniform, standard Normal, Student-t(3) (denoted $t(0, 1, 3)$ in the notation of the Appendix), Beta(3,2), Exponential with mean 5 and $\chi^2(3)$ distributions. For each sample size calculate the mean and standard deviation and compare these quantities to the known means and standard deviations from each distribution.

Solution: This is Exercise 11.2 in BEM.

Exercise 2: Analytical and Monte Carlo Integration in the Normal Linear Regression Model

a) Generate an artificial data set of size $N = 100$ from the Normal linear regression model with an intercept and one other explanatory variable. Set the intercept (β_1) to 0, the slope coefficient (β_2) to 1.0 and $h = 1.0$. Generate the explanatory variable by taking random draws from the $U(0, 1)$ distribution.

b) Calculate the posterior mean and standard deviation for the slope coefficient, β_2 , for this data set using a Normal-Gamma prior with $\underline{\beta} = (0, 1)'$, $\underline{V} = I_2$, $\underline{s}^{-2} = 1$, $\underline{\nu} = 1$.

c) Calculate the Bayes factor comparing the model $M_1 : \beta_2 = 0$ with $M_2 : \beta_2 \neq 0$.

d) Carry out a prior sensitivity analysis by setting $\underline{V} = cI_2$ and repeating parts b) and c) for values of $c = 0.01, 1.0, 100.0, 1 \times 10^6$. How sensitive is the posterior to changes in prior information? How sensitive is the Bayes factor?

e) Repeat part b) using Monte Carlo integration for various values of R . How large does R have to be before you reproduce the results of the previous parts to two decimal places?

f) Calculate the numerical standard errors associated with the posterior mean of the slope coefficient for the models. Does the use seem to give a reliable guide to the accuracy of the approximation provided by Monte Carlo integration?

Solution: This is Exercise 11.3 in BEM.

Exercise 3: Gibbs Sampling from The Bivariate Normal.

The purpose of this question is to learn about the properties of the Gibbs sampler in a simple case.

Assume that you have a model which yields a bivariate Normal posterior,

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$

where $|\rho| < 1$ is the (known) posterior correlation between θ_1 and θ_2 .

(a) Write a program which uses Monte Carlo integration to calculate the posterior means and standard deviations of θ_1 and θ_2 .

(b) Write a program which uses Gibbs sampling to calculate the posterior means and standard deviations of θ_1 and θ_2 .

(c) Set $\rho = 0$ and compare the programs from parts a) and b) for a given number of replications (e.g. $R = 100$) and compare the accuracy of the two algorithms.

(d) Repeat part c) of this question for $\rho = .5, .9, .99$ and $.999$. Discuss how the degree of correlation between θ_1 and θ_2 affects the performance of the Gibbs sampler. Make graphs of the Monte Carlo and Gibbs sampler replications of θ_1 (i.e. make a graph with x-axis being replication number and y-axis being θ_1). What can the graphs you have made tell you about the properties of Monte Carlo and Gibbs sampling algorithms?

(e) Repeat parts c) and d) more replications (e.g. $R = 10,000$) and discuss how Gibbs sampling accuracy improves with number of replications.

Solution: This is Exercise 11.7 in BEM.