Econometrics, Semester II, 2024-25
Homework IV (100 points)

Instructor: M.A. Rahman

Deadline: 7:00 pm, April 24, 2025.

Please read the instructions carefully and follow them while writing answers.

- *Solutions to homework should be typed in LaTeXor written in A4 size loose sheets.*
- *Questions should be answered in order as they appear in the homework. Every new question should begin in a new page. Please number all the pages of your homework solution.*
- *Please leave a margin of one inch from top and one inch from left. Staple the sheets on top-left.*
- *Matlab assignments (if any) and written answers should be together and in order.*
- *Please write your name and names of your group members on the first page of your answer script.*

---

1. **(5+5+5+5 = 20 points)** *Gibbs sampling*: Consider a single observation y = (y_1, y_2) from the distribution,

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

and assume a uniform prior on θ.

   (a) Find the posterior distribution of θ.

   (b) Show that the posterior conditional distributions are

$$\theta_1|\theta_2, y \sim N\big(y_1 + \rho(\theta_2 - y_2), 1 - \rho^2\big) \quad (1)$$
$$\theta_2|\theta_1, y \sim N\big(y_2 + \rho(\theta_1 - y_1), 1 - \rho^2\big). \quad (2)$$

   (c) Assume y = (1, 0.5) and $\rho = 0.8$. Draw 10,000 observations from the conditional posterior distributions using Gibbs sampling. Plot the distribution of θ_1 and θ_2 after a burn-in of 1000.

(d) While doing Gibbs sampling, simultaneously construct two parameters, say η_1 and η_2 which are standardized versions of θ_1 and θ_2, respectively. Plot the distribution of η_1/η_2 after the burn-in. What do you observe? Can you recognize its distribution?

2. **(20 points)** Generate 10,000 draws from a $Beta(3,4)$ distribution with $U(0,1)$ as proposal density with independent chain, where U denotes a Uniform distribution. Report the acceptance rate, mean, variance of the MH draws. Present a trace plot and a histogram of the draws.

   For a $Beta(3,4)$ distribution, what are the theoretical mean and variance? Are the mean and variance from MH draws different from the theoretical counterpart?

3. **(10+5+10+5+5 = 35 points)**. Consider the data in the file "`Mroz Data.xlsx`", first utilized by Mroz (1987) to explore important issues of endogeneity and sample selection, but these will be neglected here to provide a simple example of Tobit model. The data corresponds to 753 married women for the year 1975 from the University of Michigan Panel Study of Income Dynamics. There are many variables in the file, but we will only employ a subset of the variables for our exercise. Our response variable is the number of hours worked (`WHRS`). The covariates include a constant, number of children less than six years old at home (`childl6`), number of children between six and eighteen years old at home (`childg6`), the woman's age (`age`), and the husband's yearly wage (`huswage`).

   Assume the following prior distributions: $\beta \sim N_k(0, 1000 * I_k), \sigma^2 \sim IG(\alpha_0/2, \delta_0/2)$ where $\alpha_0 = 100,000$ and $\delta_0 = 10$. Please run the MCMC chain for 20,000 iterations after a burn-in of 5,000 iterations. Given the information, answer the following questions.

   (a) Fit a Bayesian linear regression model and report the posterior mean and posterior standard deviation of the parameters in a table (the name of the variables should be present in the first column). Are there reasons to believe that a linear regression framework will not be appropriate for this data?

   (b) Write down a Tobit model and the corresponding augmented posterior distribution. Derive the conditional posteriors for (β, σ^2, z).

   (c) Fit a Tobit model and report the posterior mean and standard deviations in a table. Comment on the effect of each variable on the response variable.

   (d) Report the 95% probability interval or credible interval. Explain the difference between confidence interval and credible interval.

   (e) Compute the inefficiency factor for each parameter using the batch-means method (Greenberg, 2012). Comment on the cost of using MCMC for each parameter to get an *iid* draw.

4. **(5+10+5+5 = 25 points)** *Bayesian estimation of probit model*: Consider the file "`hmda.xlsx`", where the data is present in the sheet "`data`" and variable description in the sheet "`desc`". This is a cross section data from the Home Mortgage Disclosure Act (HMDA) pertaining to mortgage applications made in 1990 in the greater Boston metropolitan area.

Our objective is to model the probability of loan denial (i.e., $y = 1$ if `deny = yes`, and 0 otherwise) as a function of the intercept and remaining variables in the file. Note that the following covariates are categorical/binary: `chist, mhist, phist, selfemp, insurance, condomin, afam, single,` and `hschool`.

First, estimate the model using the classical method. Second, fit a binary probit model using Bayesian methods assuming the prior $\beta \sim N(0_k, 100 * I_k)$. Prepare a table where you report the maxmimum likelihood (ML) estimate and standard error of the regression coefficients. In the same table, report the posterior estimates (mean and standard deviation) based on 10,000 MCMC iterations after a burn-in of 2,500 iterations. Present a trace plot of the MCMC draws of the 20 parameters (including the intercept) in one figure.

Do you find any evidence of racial bias in granting loans to black family? What is average covariate effect of loan denial for a black family?

References

Greenberg, E. (2012), *Introduction to Bayesian Econometrics*, 2nd Edition, Cambridge University Press, New York.

Mroz, T. M. (1987), "The Sensitivity of an Empirical Model of Married Women's Hour of Work to Economic and Statistical Assumptions," *Econometrica*, 55, 765–799.