Bayesian Econometrics
Tutorial 03 - Numerical Methods for Bayesian Linear Regression Models

Tutor: Richard Schnorrenberger

richard.schn@stat-econ.uni-kiel.de

Institute for Statistics and Econometrics
Kiel University

Winter Term 2024/25

Review the Concepts and Proofs

- 1. Consider the normal linear regression model. What is the difference between the natural conjugate and the independent normal-gamma prior?
- 2. What is Monte Carlo integration? Give three examples of often-used integrand functions $g(\theta)$.
- 3. Explain the probability integral transform.
- 4. Explain the acceptance-rejection method.
- 5. Explain importance sampling. Why is it particularly helpful for Monte Carlo integration of the normal linear regression model with inequality constraints?
- 6. Explain the Gibbs sampler. Why is it particularly helpful for Monte Carlo integration of the normal linear regression model with independent normal-gamma prior?
- 7. What does Markov chain Monte Carlo (MCMC) mean?
- 8. Why is it important to check convergence of MCMC algorithms?
- 9. Why is it, at least in principle, necessary to use a long-run variance estimator when computing the numerical standard error for a MCMC algorithm?
- 10. What is the Savage-Dickey density ratio?

Exercise 1

Show how you can apply importance sampling to simulate moments of a truncated normal distribution, $x \sim \mathcal{TN}_{[a,b]}(\mu, \sigma^2)$, $a \leq x \leq b$, using only normal random numbers.

Exercise 2

How can you easily simulate draws from the general $t(\mu, \sigma^2, k)$-distribution for integer values of k if your software offers you only standard normal random numbers? (Hint: think of the definition of the t-distribution.)

Recall the CLT under normal random variables: Let $\{x_1, x_2, ..., x_N\}$ be a random sample drawn from a population following a normal distribution, i.e. $X \sim \mathcal{N}(\mu, \sigma^2)$. Then, if $\bar{X}_N$ is the sample mean, the random variable

$$\sqrt{N}\,\frac{\bar{X}_N - \mu}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1)$$

However, if the population variance σ^2 is unknown, we can use its sample counterpart $S_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X}_N)^2$ such that the standardized sample mean now follows the student's t distribution with $k = N - 1$ degrees of freedom:

$$\sqrt{N}\,\frac{\bar{X}_N - \mu}{S_X} \xrightarrow{d} t_k$$

Exercise 3

Suppose you have a posterior distribution of the scalar parameter θ which is logistic with parameters $-\infty < \bar{\alpha} < \infty$ and $0 < \bar{\beta} < \infty$ and cdf

$$F(\theta|\bar{\alpha}, \bar{\beta}) = \left[1 + \exp\left(-\frac{\theta - \bar{\alpha}}{\bar{\beta}}\right)\right]^{-1}.$$

You want to apply Monte Carlo integration to find (i) the mean of θ, (ii) the variance of θ, and (iii) the expected value of $g(\theta) = \exp(\sqrt{|\theta|} - 1)$ but you have access to random numbers from the uniform distribution and the standard normal distribution only.

- (a) Find the posterior pdf.
- (b) Write a pseudo code (i.e., a step-by-step algorithm) that applies the probability integral transform.

Exercise 3

- (c) Consider the acceptance-rejection method. Show that the normal distribution is not a good proposal distribution $h(\theta)$ because the scale factor M is unbounded in this case. (Hint: what happens to $f(\theta)/h(\theta)$ as $\theta \to \infty$?) Next show that the t-distribution should work properly because M is bounded.
- (d) Write a pseudo code that applies the acceptance-rejection method.
- (e) Show in pseudo code how importance sampling can be used. Which importance function do you choose?

Solution to Exercise 3 (e)

Given the discussion in (d) we may choose the $t(\mu, \sigma^2, k)$ pdf as importance function $q(\theta)$ with the same choice of parameters. The pseudo code is then given by:

1. Define posterior parameters $\bar{\alpha}$ and $\bar{\beta}$.
2. Define $\mu = \bar{\alpha}$.
3. Define $\sigma = \bar{\beta}\pi\sqrt{\frac{k-2}{3k}}$.
4. Define $k = 7$.
5. Define number of replications S.
6. Draw S $t(\mu, \sigma^2, k)$-distributed random numbers $y_1, \dots, y_S$ (see Exercise 2).
7. For each y_i compute $p_i = p(y_i)$ and $q_i = q(y_i)$.
8. For each draw i, compute the weight $w_i = p_i / q_i$.
9. To find $\mathrm{E}(\theta|y)$, compute $m_S = \sum_{i=1}^{S} w_i y_i / \sum_{i=1}^{S} w_i$.
10. To find $\mathrm{Var}(\theta|y)$, compute $s_S^2 = \sum_{i=1}^{S} w_i (y_i - m_S)^2 / \sum_{i=1}^{S} w_i$.
11. To find $\mathrm{E}[g(\theta)|y]$, compute $g_i = \exp(\sqrt{|y_i|} - 1)$ for all $i = 1, \dots, \tilde{S}$. Then compute $\hat{g}_S^w = \sum_{i=1}^{S} w_i g_i / \sum_{i=1}^{S} w_i$.

Exercise 4

Consider the normal linear regression model with independent Normal-Gamma prior, $\beta \sim \mathcal{N}(\underline{\beta}, \underline{V})$ and $h \sim Gamma(\underline{s}^{-2}, \underline{\nu})$. Assume the regression model includes an intercept β_1 and one regressor with slope β_2, i.e., $\beta = (\beta_1, \beta_2)'$.

- (a) Explain how prior knowledge can be used to specify the prior parameters $\underline{\beta}$, $\underline{V}$, $\underline{s}^{-2}$, and $\underline{\nu}$.
- (b) State prior and likelihood. Find the posterior pdf.
- (c) Find the conditional posterior of β given h and the data.
- (d) Find the conditional posterior of h given β and the data.

Exercise 4

- (e) (*) Show in pseudo code how the acceptance-rejection method can be used to find the posterior mean and variance matrix of β if only standard normal random numbers are available. Find an efficient proposal distribution that minimizes the acceptance rate.
- (f) (*) Show in pseudo code how importance sampling can be used to find the posterior mean and variance matrix of β if only standard normal random numbers are available.
- (g) Show in pseudo code how Gibbs sampling can be used to find the posterior mean and variance matrix of β if independent normal and gamma random numbers are available.

Solution to Exercise 4 (a)

- $\underline{\beta}$: (i) crude guesses (subjective prior) or (ii) pre-sample results (other studies).
- $\underline{V}$: how certain are you about the values of $\underline{\beta}$? If you are uncertain, then you should set a really high $\underline{V}$, otherwise you can define smaller values.
- $\underline{s}^{-2}$: think about the variation in your dependent variable y. For instance, suppose the error term will be around 10,000. Then $\sigma \approx 5,000$ since in the Normal setting the standard error would $1.96 \times 5,000 = 9,800$. This implies $\underline{s}^{-2} \approx 1/(5,000)^2$.
- $\underline{\nu}$: how much weight do you want to attach to $\underline{s}^{-2}$? If you want to attach little weight to it, then set $\underline{\nu}$ much smaller than the sample size N, where the ratio $\underline{\nu}/N$ kind of determines the relative weight of prior information in comparison to data information.

Solution to Exercise 4 (g)

- Note that we have conditional posteriors available in closed-form solution, however, as econometricians, we are usually interested in the joint posterior density $p(\beta, h|y)$, or maybe in the marginal posteriors, but NOT in $p(\beta|y, h)$ and $p(h|y, \beta)$.
- Since $p(\beta, h|y) \neq p(\beta|y, h)p(h|y, \beta)$, the conditional posteriors do not directly tell us about $p(\beta, h|y)$.
- The Gibbs sampler provides a solution: $p(\beta, h|y) = p(\beta|y, h)p(h|y) = p(h|y, \beta)p(\beta|y)$ implies that taking a random draw from a marginal posterior and then using this draw to generate a random draw from the corresponding conditional posterior is equivalent to directly sampling from the joint posterior $p(\beta, h|y)$.
- To sum up, the Gibbs sampler uses conditional posteriors to produce sequentially/recursively random draws $\beta^{(s)}$ and $h^{(s)}$ for $s = 1, \ldots, S$ which can be averaged to produce estimates of posterior properties just as with Monte Carlo integration.

Solution to Exercise 4 (g)

- **Remark 1**: since the Gibbs sampler needs to be initialized with a random draw $\beta^{(0)}$ or $h^{(0)}$ from $p(\beta|y)$ or $p(h|y)$, where both do not have a closed-form solution, we split the number of replications into S_0 burn-in replications and use the last S_1 replications for Monte Carlo integration. The idea is that the influence of $\beta^{(0)}$ or $h^{(0)}$ in the MCMC dies out after the S_0 burn-in replications.
- **Remark 2**: we could in fact apply importance sampling or the acceptance-rejection method by suggesting a good proposal density for $p(\beta, h|y)$. However, both of these simulation methods become inefficient as the dimension of the integration problem gets large (as the dimension of β gets large). Therefore, let us exploit the fact that the conditional posteriors are available in closed-form solution so that the Gibbs sampler can be applied to efficiently draw from the joint posterior $p(\beta, h|y)$.

Solution to Exercise 4 (g)

The Gibbs sampler pseudo code:

1. Set prior parameters $\underline{\beta}$, $\underline{V}$, $\underline{s}^{-2}$, and $\underline{\nu}$.
2. Define posterior parameters $\bar{\beta}$, $\bar{V}$, $\bar{s}^2$ and $\bar{\nu}$.
3. Define number of replications $S = S_0 + S_1$, where S_0 denotes burn-in replications.
4. Initialize the Gibbs sampler: draw $\beta^{(0)}$ from some marginal distribution $p(\beta|y)$ (e.g., OLS estimates).
5. Generate the first draw of $h^{(1)}|\beta^{(0)}$ from $p(h|y, \beta^{(0)})$.
6. Generate the first draw of $\beta^{(1)}|h^{(1)}$ from $p(\beta|y, h^{(0)})$.
7. Start iteration with loop counter $i = 2 : S$
   - Draw $h^{(i)}|\beta^{(i-1)}$ from $p(h|y, \beta^{(i-1)})$.
   - Draw $\beta^{(i)}|h^{(i)}$ from $p(\beta|y, h^{(i)})$.
8. Discard burn-in draws S_0.
9. Compute posterior mean $E(\beta|y)$ by the sample average $\hat{\beta} = \frac{1}{S_1}\sum_{i=S_0+1}^{S_1} \beta^{(i)}$.
10. Compute posterior variance $Var(\beta|y)$ by the sample variance $\hat{S}^2_{\beta} = \frac{1}{S_1-1}\sum_{i=S_0+1}^{S_1}(\beta^{(i)} - \hat{\beta})^2$.