Bayesian Econometrics
PC Tutorial 03

Tutor: Richard Schnorrenberger

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

Institute for Statistics and Econometrics
Kiel University

Winter Term 2024/25

Exercise 1

Write a Matlab script that uses importance sampling to simulate the mean and variance of a truncated standard normal distribution with truncation bounds $a = -1$ and $b = 1.5$ using only normal random numbers. Compare to the analytically available results shown here.

The *importance sampling* technique can be here applied using the standard normal as the importance sampler:

$$
\begin{aligned}
I &= \int g(\theta) \underbrace{p(\theta|y)}_{\sim \mathcal{TN}_{[a,b]}(0,1)} d\theta = \int \frac{g(\theta)p(\theta|y)}{m(\theta)} \underbrace{m(\theta)}_{\sim \mathcal{N}(0,1)} d\theta = E\left[\frac{g(\theta)p(\theta|y)}{m(\theta)}\right] \\
&\approx \hat{I}_S = \frac{1}{S}\sum_{s=1}^{S} w(\theta^{(s)})g(\theta^{(s)}) \quad \text{such that} \quad \hat{I}_S \xrightarrow{p} E\left[\frac{g(\theta)p(\theta|y)}{m(\theta)}\right]
\end{aligned}
$$

where $w(\theta^{(s)}) = \frac{p(\theta^{(s)}|y)}{m(\theta^{(s)})}$ are the importance sampling weights and they “correct" for sampling from the importance sampler $m(\theta)$ rather than the target pdf $p(\theta|y)$.

Exercise 1

Simulation thus proceeds as follows:

1. Simulate S random numbers $\theta^{(s)}$ from the importance sampler $\mathcal{N}(0, 1)$ with $s = 1, \ldots, S$.
2. For each $\theta^{(s)}$ compute a weight $w(\theta^{(s)})$ such that:

$$w(\theta^{(s)}) = \begin{cases} 1 & \text{if } a \leq \theta^{(s)} \leq b \\ 0 & \text{otherwise} \end{cases}$$

3. Compute the weighted average

$$\bar{\theta} = \frac{\sum_{s=1}^{S} w(\theta^{(s)})\theta^{(s)}}{\sum_{s=1}^{S} w(\theta^{(s)})}$$

4. Compute the weighted variance

$$S_\theta^2 = \frac{\sum_{s=1}^{S} w(\theta^{(s)}) \left(\theta^{(s)} - \bar{\theta}\right)^2}{\sum_{s=1}^{S} w(\theta^{(s)})}$$

Exercise 2

Suppose, as in Exercise 3 of the Pen & Paper, you have a posterior distribution of the scalar parameter θ which is logistic. Assume the parameters are $\bar{\alpha} = 4$ and $\bar{\beta} = 2$. Apply Monte Carlo integration to find (i) the mean of θ, (ii) the variance of θ, and (iii) the expected value of $g(\theta) = \exp(\sqrt{|\theta|} - 1)$. Compare to the analytic results available here.

- (a) Write a Matlab script that applies the probability integral transform.
- (b) Write a Matlab script that applies the acceptance-rejection method using the *t*-distribution as proposal distribution. Plot both the target and proposal pdfs.
- (c) Write a Matlab script that applies importance sampling. Choose the *t*-distribution as importance function. (*) You can play around by reporting the average and variance of the weights for different degrees of freedom *k*.

Exercise 2 (a): pseudo code

Set the cdf $F(\theta)$ equal to $u \sim U(0,1)$ and apply the transformation (inverse cdf):

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

$$\Rightarrow -\log\left(u^{-1} - 1\right) = \frac{\theta - \bar{\alpha}}{\bar{\beta}} \Rightarrow \theta = \bar{\alpha} - \bar{\beta}\log\left(u^{-1} - 1\right)$$

Numerical standard error of the mean estimate $\bar{\theta}$:

$$\frac{\sigma_g}{\sqrt{S}} = \frac{\sqrt{S\,\mathrm{Var}(\hat{g}_S)}}{\sqrt{S}} = \sqrt{\mathrm{Var}(\hat{g}_S)} = \sqrt{\frac{1}{S^2}\sum_{s=1}^{S}\underbrace{\mathrm{Var}(\theta^{(s)})}_{\mathrm{Logistic}(\bar{\alpha},\bar{\beta})}} = \sqrt{\frac{S}{S^2}\frac{\bar{\beta}^2\pi^2}{3}} = \frac{\bar{\beta}\pi}{\sqrt{3S}}$$

Exercise 2 (b): pseudo code

1. Set posterior parameters $\bar{\alpha}$ and $\bar{\beta}$.
2. Define the $t(\mu, \sigma^2, k)$-distribution parameters such that $\mu = \bar{\alpha}$, $\sigma = \bar{\beta}\pi\sqrt{\frac{k-2}{3k}}$ and $k = 7$.
3. Define number of replications S.
4. Draw S $t(\mu, \sigma^2, k)$-distributed random numbers $y_1, \ldots, y_S$.
5. For each y_i compute the target pdf $p_i = p(y_i)$ and the proposal pdf $q_i = q(y_i)$.
6. Compute M as the maximum of all p_i/q_i.
7. Draw S standard uniform random numbers $u_1, \ldots, u_S$.
8. For each draw i, if $u_i \leq p_i/(M\,q_i)$, accept the draw as θ_i. Otherwise discard it.
9. Compute $\tilde{S}$ as the number of accepted draws $\theta_1, \ldots, \theta_{\tilde{S}}$.
10. To find $\mathrm{E}(\theta|y)$, compute $\bar{\theta} = \tilde{S}^{-1} \sum_{i=1}^{\tilde{S}} \theta_i$.
11. To find $\mathrm{Var}(\theta|y)$, compute $S_\theta^2 = (\tilde{S} - 1)^{-1} \sum_{i=1}^{\tilde{S}} (\theta_i - \bar{\theta})^2$.
12. To find $\mathrm{E}[g(\theta)|y]$, compute $g_i = \exp(\sqrt{|\theta_i|} - 1)$ for all $i = 1, \ldots, \tilde{S}$. Then compute $\hat{g}_{\tilde{S}} = \tilde{S}^{-1} \sum_{i=1}^{\tilde{S}} g_i$.

Exercise 2 (c): pseudo code

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

Exercise 3

Consider the textbook example of Canadian house prices with independent normal-gamma prior, $\beta \sim \mathcal{N}(\underline{\beta}, \underline{V})$ and $h \sim Gamma(\underline{s}^{-2}, \underline{\nu})$. Choose the prior parameters according to the textbook.

- (a) Write a Matlab script that includes the following steps: (i) load the data `HPRICE.txt` (ii) compute OLS estimates, (iii) set the priors, (iv) perform Gibbs sampling with $1,000$ burn-in replications and $10,000$ MC posterior replications, and (v) compute posterior means and standard deviations for β.
- (b) Extend your script to report numerical standard errors for β based on Newey-West long-run variances (use the `NeweyWest.m` function).
- (c) Extend your script to report *CD* statistics for convergence based on subsamples $A = 10\%$, $B = 50\%$, and $C = 40\%$.
- (d) (*) Extend your script to report estimated potential scale reductions for β based on $m = 20$ parallel Markov chains.
- (e) (*) Extend your script to report posterior mean and standard deviation of the prediction y^* based on $X^* = (1, 5000, 2, 2, 1)$. Plot the predictive density.