Homework Assignment 1

Bayesian Learning

Professional Master in Economics, Insper

Instructor: Prof. Hedibert F. Lopes
Teaching assistant/grader: Guilherme Piantino
Issued date: May 13th, 2026
Due date: June 1st, 2026 (at most by 7:15pm)

Instructions

- Submit a single PDF (typeset in LaTeX; handwritten work will not be accepted) together with reproducible code in a single `.zip` archive.
- Code may be in R, Python, or Julia. All random-number experiments must be seeded so that results are reproducible. Effective sample sizes (ESS) should be computed using a standard package: `coda` or `mcmcse` in R, `arviz` in Python, or `MCMCChains.jl` in Julia.
- Plots must be labeled and legible. Discussion of results is required throughout; raw numerical output without commentary will not receive full credit.
- Use the dataset $\boldsymbol{x} = (x_1, \ldots, x_n)$ generated as described in Question 2, with `set.seed(20260512)` (or the equivalent in your language) so that the class works from the same simulated sample.
- Questions about the assignment should be directed to the TA, Guilherme Piantino.

Question 1 (20 points)

A useful representation of the Student's t distribution as a scale mixture of normals.

Let $\nu > 0$, $\mu \in \mathbb{R}$, and $\sigma^2 > 0$ be fixed. Suppose that

$$x \mid \lambda \sim \mathcal{N}(\mu, \lambda\sigma^2), \qquad \lambda \sim \mathcal{IG}(\nu/2, \nu/2),$$

where $\mathcal{IG}(a, b)$ denotes the inverse-gamma distribution with shape a and scale b, i.e. with density $p(\lambda) = \frac{b^a}{\Gamma(a)}\lambda^{-a-1}e^{-b/\lambda}$ for $\lambda > 0$.

Show that the marginal distribution of x is Student's t with location μ, scale σ^2, and ν degrees of freedom; that is, $x \sim t_\nu(\mu, \sigma^2)$ with density

$$p(x \mid \mu, \sigma^2, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu\sigma^2}\,\Gamma\left(\frac{\nu}{2}\right)}\left[1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right]^{-(\nu+1)/2}.$$

(In particular, setting $\sigma^2 = 1$ recovers the usual $t_\nu(\mu, 1)$ density. The result is used throughout Question 2.)

Hint. Write the joint density $p(x, \lambda) = p(x \mid \lambda)\, p(\lambda)$, collect powers of λ, and integrate λ out by recognizing the integrand as the (unnormalized) kernel of an inverse-gamma density.

Question 2 (80 points)

Posterior inference for μ under a Student's t likelihood.

Simulate $n = 50$ observations from $t_4(2, 1)$, using `set.seed(20260512)`. Throughout this question:

- Treat $\nu = 4$ and $\sigma^2 = 1$ as **known and fixed.**
- Place the prior $\mu \sim \mathcal{N}(\mu_0, \tau_0^2)$ with **known** hyperparameters $\mu_0 = 0$ and $\tau_0^2 = 10$.
- The only unknown parameter is μ.

The goal in parts (a)–(c) is to obtain Monte Carlo draws from the posterior $p(\mu \mid \boldsymbol{x})$ by three different algorithms, and in part (d) to compare them.

(a) Sampling Importance Resampling (20 points)

Implement SIR to obtain $M = 10{,}000$ approximate draws from the posterior $p(\mu \mid \boldsymbol{x})$, using the Student's t likelihood directly (i.e. without the scale-mixture augmentation of Question 1). For each proposal $q(\mu)$, draw a large pool of candidates $\mu^{(j)} \sim q(\mu)$ for $j = 1, \ldots, N$ (take $N = 100{,}000$), assign importance weights

$$w_j \;\propto\; \frac{p(\boldsymbol{x} \mid \mu^{(j)})\, p(\mu^{(j)})}{q(\mu^{(j)})}, \qquad \sum_{j=1}^{N} w_j = 1,$$

and resample M values with probabilities $\{w_j\}$.

Compare two alternative proposal distributions:

(i) $q_1(\mu) = \text{Uniform}(-10, 10)$;

(ii) $q_2(\mu) = t_3(\bar{x},\, s^2)$, where $\bar{x}$ and s^2 are the sample mean and sample variance of the data.

Proposal (ii) uses information from the data and is expected to be much more efficient; proposal (i) is cruder but trivial to set up. Quantify the trade-off:

- Report the effective sample size of the importance weights, $\text{ESS} = \left(\sum_j w_j^2\right)^{-1}$, for each proposal.
- Plot a histogram of the (normalized) weights for each proposal.
- Plot the resampled posterior for each proposal and overlay them on the same axes.

Comment on which proposal you would use in practice and why.

(b) Random-walk Metropolis (20 points)

Implement a random-walk Metropolis algorithm targeting $p(\mu \mid \boldsymbol{x})$, with Gaussian proposals

$$\mu^{\text{prop}} \mid \mu^{(t)} \sim \mathcal{N}(\mu^{(t)}, c^2),$$

for tuning standard deviations $c \in \{0.1,\ 0.5,\ 1.0,\ 2.0\}$.

For *each* value of c:

- Run the chain for $M = 20{,}000$ iterations after a 5,000-iteration burn-in, starting from $\mu^{(0)} = 0$.
- Report the empirical acceptance rate.
- Plot the trace of μ and the autocorrelation function (ACF) up to lag 50.
- Compute the effective sample size $\text{ESS}(c)$ for the estimator of $\mathbb{E}(\mu \mid \boldsymbol{x})$.

Then produce a single table containing, for each c, the acceptance rate, the ESS, and the cost per effective sample $M/\text{ESS}(c)$. Discuss which tuning standard deviation delivers the best mixing, and how this relates to the well-known "≈ 0.234 optimal acceptance rate" rule of thumb for random walks.

(c) Gibbs sampler via scale-mixture augmentation (25 points)

Using the representation derived in Question 1, write the model as

$$x_i \mid \mu, \lambda_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu,\ \lambda_i\sigma^2), \qquad \lambda_i \overset{\text{iid}}{\sim} \mathcal{IG}(\nu/2,\ \nu/2), \qquad i = 1, \ldots, n,$$

with prior $\mu \sim \mathcal{N}(\mu_0, \tau_0^2)$ as above. Then:

(c.1) Derive the two full conditionals,

$$p(\mu \mid \boldsymbol{\lambda}, \boldsymbol{x}) \qquad \text{and} \qquad p(\lambda_i \mid \mu, \boldsymbol{x}) \ \text{ for } \ i = 1, \ldots, n.$$

Show that both belong to standard families (Gaussian and inverse-gamma, respectively) and report the parameters explicitly.

(c.2) Implement the $(n+1)$-step Gibbs sampler in which each iteration cycles through:

- draw $\lambda_i \mid \mu, \boldsymbol{x}$, for $i = 1, \ldots, n$;
- draw $\mu \mid \boldsymbol{\lambda}, \boldsymbol{x}$.

Run for $M = 20{,}000$ iterations after a 5,000-iteration burn-in.

(c.3) Diagnose the chain: trace plot of μ, ACF of μ, posterior histogram for μ, and ESS for the estimator of $\mathbb{E}(\mu \mid \boldsymbol{x})$.

(d) Synthesis (15 points)

Pull everything together. At a minimum, your write-up should address:

- **SIR (a):** which proposal worked better, and was the gain worth the extra effort? When would you reach for the uniform proposal anyway?
- **Random-walk Metropolis (b):** describe the relationship between the tuning standard deviation c, the acceptance rate, and the effective sample size. Identify the regime in which the chain is "too sticky" (high autocorrelation, high acceptance) versus "too jumpy" (low acceptance), and the sweet spot in between.
- **Gibbs (c):** what was the cost of augmenting the parameter space with $n = 50$ additional latent variables $\lambda_1, \ldots, \lambda_n$? Did the closed-form full conditionals compensate for the extra bookkeeping?
- Provide a single **summary table** comparing the three algorithms (each at its best tuning):

Method	$\widehat{\mathbb{E}}(\mu \mid \boldsymbol{x})$	$\widehat{\mathrm{sd}}(\mu \mid \boldsymbol{x})$	**ESS**	**time / ESS**
SIR with proposal (ii)	______	______	______	______
Random-walk Metropolis (best c)	______	______	______	______
Gibbs	______	______	______	______

- **Final recommendation:** which algorithm would you reach for in a real applied problem with the same structure (Student's t likelihood, conjugate normal prior on the location)? Justify briefly.

End of assignment.