# Homework Assignment 2

## Bayesian Learning

Professional Master in Economics, Insper

---

**Instructor:** Prof. Hedibert F. Lopes
**Teaching assistant/grader:** Guilherme Piantino
**Issued date:** May 25th, 2026
**Due date:** June 8th, 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.
- Both questions are reproductions/extensions of worked examples in Gamerman & Lopes (2006), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd edition (Chapman & Hall/CRC). Question 1 builds on Example 5.1 (coal-mining disasters, Carlin, Gelfand and Smith, 1992); Question 2 builds on Example 6.4 (motorettes time-to-failure, Tanner, 1996).
- Use `set.seed(20260525)` (or the equivalent in your language) at the top of each MCMC run so that the class works from the same simulated streams.
- Questions about the assignment should be directed to the TA, Guilherme Piantino.

---

## Question 1 (50 points)

**Bayesian changepoint analysis of the British coal-mining disasters via a Gibbs sampler (Example 5.1, Gamerman & Lopes, 2006).**

Let $y_1, \ldots, y_n$ be annual counts of disasters in Great Britain from 1851 to 1962, with $n = 112$ (Jarret, 1979; Table 5.1 of Gamerman & Lopes, 2006). The model suspects a single *changepoint* $m \in \{1, \ldots, n\}$ in the Poisson rate:

$$y_i \mid \lambda \overset{\text{ind}}{\sim} \text{Poi}(\lambda), \qquad i = 1, \ldots, m, \qquad y_i \mid \phi \overset{\text{ind}}{\sim} \text{Poi}(\phi), \qquad i = m+1, \ldots, n.$$

The prior is

$$\lambda \sim \mathcal{G}(\alpha, \beta), \qquad \phi \sim \mathcal{G}(\gamma, \delta), \qquad m \sim \text{Uniform}\{1, \ldots, n\},$$

with $\alpha, \beta, \gamma, \delta > 0$ **known**. Write $s_\ell = \sum_{i=1}^{\ell} y_i$ for $\ell = 1, \ldots, n$ (and $s_0 = 0$).

### (a) Full conditionals (10 points)

Show that the joint posterior is proportional to

$$\pi(\lambda, \phi, m) \ \propto \ \lambda^{\alpha + s_m - 1}\, e^{-(\beta+m)\lambda}\, \phi^{\gamma + s_n - s_m - 1}\, e^{-(\delta + n - m)\phi},$$

and **derive** the three full conditionals (cf. equation (5.2) of Gamerman & Lopes, 2006):

$$\begin{aligned}
\pi(\lambda \mid \phi, m, \boldsymbol{y}) &= \mathcal{G}(\alpha + s_m,\ \beta + m), \\
\pi(\phi \mid \lambda, m, \boldsymbol{y}) &= \mathcal{G}(\gamma + s_n - s_m,\ \delta + n - m), \\
\pi(m \mid \lambda, \phi, \boldsymbol{y}) &\propto \lambda^{s_m}\, e^{-m\lambda}\, \phi^{-s_m}\, e^{m\phi}, \qquad m = 1, \ldots, n.
\end{aligned}$$

Comment on the fact that, conditionally on $m$, the rates $\lambda$ and $\phi$ are a posteriori independent.

### (b) Gibbs sampler (15 points)

Implement the three-step Gibbs sampler that, at each iteration, cycles through

$$\lambda \mid \phi, m, \boldsymbol{y} \ \longrightarrow \ \phi \mid \lambda, m, \boldsymbol{y} \ \longrightarrow \ m \mid \lambda, \phi, \boldsymbol{y}.$$

Use the hyperparameters $\alpha = \gamma = 0.5$ and $\beta = \delta = 1$ (a vague Gamma prior centered well below the data scale). Run the chain for $M = 20{,}000$ iterations after a 5,000-iteration burn-in, starting from $(\lambda^{(0)}, \phi^{(0)}, m^{(0)}) = (\bar{y}, \bar{y}, n/2)$. The disasters dataset is widely available; you may use the $n = 112$ counts reproduced in Table 5.1 of Gamerman & Lopes (2006) (also available in R via `boot::coal` or from many public repositories — confirm $n = 112$ and $s_n = \sum y_i = 191$ before proceeding).

For the discrete draw $m \mid \lambda, \phi, \boldsymbol{y}$, evaluate the log of the unnormalized mass on $\{1, \ldots, n\}$, subtract the maximum (for numerical stability), exponentiate, and sample with probabilities proportional to the resulting vector.

### (c) Convergence diagnostics (10 points)

Report, for each of $\lambda$, $\phi$, and $m$:

- the trace plot,
- the autocorrelation function up to lag 50,
- the marginal posterior (histogram for $\lambda, \phi$; bar chart over $\{1, \ldots, n\}$ for $m$),
- the posterior mean, standard deviation, and a 95% credible interval,
- the effective sample size.

Translate the posterior of $m$ back to calendar years (year $= 1850 + m$) and identify the most likely change year. Briefly relate the result to the historical record (Mining Regulation Act of 1887; safety legislation in the early 1900s).

### (d) Analytical marginal vs. Monte Carlo (10 points)

Equation (5.3) of Gamerman & Lopes (2006) gives the *analytical* marginal posterior of $m$:

$$\pi(m) \;\propto\; \frac{\Gamma(\alpha + s_m)\,\Gamma(\gamma + s_n - s_m)}{(m+\beta)^{\alpha+s_m}\,(n-m+\delta)^{\gamma+s_n-s_m}}, \qquad m = 1, \ldots, n.$$

Evaluate $\pi(m)$ on the log scale for $m = 1, \ldots, n$, normalise by summing, and overlay the resulting bars on the histogram of the Gibbs draws from (c). Comment on the agreement. Using

$$\mathbb{E}_\pi(\lambda \mid \boldsymbol{y}) \;=\; \sum_{m=1}^{n} \mathbb{E}(\lambda \mid m, \boldsymbol{y})\,\pi(m), \qquad \mathrm{Var}_\pi(\lambda \mid \boldsymbol{y}) \;=\; \text{(law of total variance)},$$

compare the Rao−Blackwellised estimates of $\mathbb{E}(\lambda \mid \boldsymbol{y})$ and $\mathrm{Var}(\lambda \mid \boldsymbol{y})$ to their raw Monte Carlo counterparts, and likewise for $\phi$. Report the variance reduction.

### (e) Prior sensitivity (5 points)

Re-run the Gibbs sampler for two additional prior pairs:

(i) $\alpha = \gamma = 1$, $\beta = \delta = 1$ (mildly more informative),

(ii) $\alpha = \gamma = 10$, $\beta = \delta = 5$ (clearly more informative).

Summarise how the posteriors of $\lambda$, $\phi$, and $m$ shift, and how much (or how little) the location of the changepoint depends on the prior. One or two paragraphs suffice.

---

## Question 2 (50 points)

**Posterior inference for a censored linear regression via three Metropolis–Hastings algorithms (Example 6.4, Gamerman & Lopes, 2006).**

Times to failure of motorettes were tested at four temperatures (Table 6.2 of Gamerman & Lopes, 2006; data originally from Tanner, 1996, p. 67). Transform the raw failure time $f_i$ (in hours) and the test temperature $t_i$ (in degrees Celsius) by

$$y_i = \log_{10}(f_i), \qquad x_i = \frac{1000}{t_i + 273.2}, \qquad i = 1, \ldots, n, \quad n = 40,$$

and fit the simple linear regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma^2),$$

treating $\sigma^2 = 0.2592$ as **known and fixed** and $p(\boldsymbol{\beta}) \propto 1$ as an improper flat prior on $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$. The data have $m = 17$ uncensored failures and $n - m = 23$ right-censored observations (the bold entries in Table 6.2). Letting $\varepsilon_i(\boldsymbol{\beta}) = y_i - \beta_0 - \beta_1 x_i$ and $\Phi$ denote the standard normal CDF, the (improper) posterior is

$$\pi(\boldsymbol{\beta}) \ \propto \ \exp\left\{ -\frac{1}{2} \sum_{i=1}^{m} \frac{\varepsilon_i^2(\boldsymbol{\beta})}{\sigma^2} \right\} \prod_{i=m+1}^{n} \left[ 1 - \Phi\left( \frac{\varepsilon_i(\boldsymbol{\beta})}{\sigma} \right) \right].$$

The censored factor has no closed form, so analytical posterior inference is unavailable; we use MCMC.

*Data.* Reproduce Table 6.2 of Gamerman & Lopes (2006) verbatim. The uncensored block (the first $m = 17$ observations, in the order of the table) is

$$\begin{aligned}(t, f):\ & (170, 1764),\ (170, 2772),\ (170, 3444),\ (170, 3542),\ (170, 3780),\\ & (170, 4860),\ (170, 5196),\\ & (190, 408),\ (190, 408),\ (190, 1344),\ (190, 1344),\ (190, 1440),\\ & (220, 408),\ (220, 408),\ (220, 504),\ (220, 504),\ (220, 504).\end{aligned}$$

The censored block (the remaining $n - m = 23$) is

$$\begin{aligned}(t, f_c):\ & (150, 8064) \times 10,\\ & (170, 5448) \times 3,\\ & (190, 1680) \times 5,\\ & (220, 528) \times 5,\end{aligned}$$

where $f_c$ is the censoring time. Verify by computing the OLS fit on the uncensored block: you should obtain $\widehat{\boldsymbol{\beta}} \approx (-6.0, 4.3)^\top$ with a strong negative correlation between $\widehat{\beta}_0$ and $\widehat{\beta}_1$.

### (a) Three MH algorithms (25 points)

Implement all three variants from the textbook:

**Alg. 1. Random-walk Metropolis with single-coordinate moves.** At each iteration, given the current state $\boldsymbol{\beta}^{(j)} = (\beta_0^{(j)}, \beta_1^{(j)})$, propose $\beta_0^\star \sim \mathcal{N}(\beta_0^{(j)}, \tau^2)$ and accept/reject (updating $\beta_0$ only), then propose $\beta_1^\star \sim \mathcal{N}(\beta_1^{(j)}, \tau^2)$ and accept/reject. Use $\tau = 0.1$.

**Alg. 2. Random-walk Metropolis with a block move.** Propose $\boldsymbol{\beta}^\star \sim \mathcal{N}_2(\boldsymbol{\beta}^{(j)}, \tau^2 I_2)$ and accept/reject as a block. Use $\tau = 0.1$.

**Alg. 3. Independence Metropolis.** Propose $\boldsymbol{\beta}^\star \sim \mathcal{N}_2(\widehat{\boldsymbol{\beta}}, \sigma^2(X^\top X)^{-1})$, where $\widehat{\boldsymbol{\beta}} = (X^\top X)^{-1}X^\top \boldsymbol{y}$ is the OLS estimator on the uncensored block, $\boldsymbol{y} = (y_1, \ldots, y_m)^\top$, and the design matrix $X$ has rows $(1, x_i)$ for $i = 1, \ldots, m$. The proposal correlation between $\beta_0$ and $\beta_1$ is approximately $-0.999$ — a heavy hint that the single-coordinate move of Alg. 1 will mix poorly.

For each algorithm, run a single chain for $M = 20{,}000$ iterations after a 5,000-iteration burn-in, starting from $\boldsymbol{\beta}^{(0)} = \widehat{\boldsymbol{\beta}}$. Always work on the **log scale** when forming the MH ratio (the censored factor $1 - \Phi(\cdot)$ can underflow in the right tail; use `pnorm(.,lower.tail=FALSE,log.p=TRUE)` in R, or `scipy.stats.norm.logsf` in Python).

## (b) Diagnostics and tuning (15 points)

For each of the three algorithms, report:

- the empirical acceptance rate (per coordinate for Alg. 1, joint for Algs. 2 and 3);
- trace plots of $\beta_0$ and $\beta_1$;
- the ACF of each coordinate up to lag 50;
- a scatter plot of the post-burn-in draws $(\beta_0^{(j)}, \beta_1^{(j)})$;
- the effective sample sizes $\mathrm{ESS}(\beta_0)$ and $\mathrm{ESS}(\beta_1)$, and the cost-per-effective-sample $M/\min\{\mathrm{ESS}\}$.

Then *tune*: re-run Algs. 1 and 2 for $\tau \in \{0.01, 0.05, 0.1, 0.3\}$ and produce, for each algorithm, a table of acceptance rate, ESS, and $M/\mathrm{ESS}$ as a function of $\tau$. Comment on the optimal acceptance rate (roughly 0.44 for one-at-a-time random walks and 0.234 for $d$-dimensional block random walks as $d$ grows).

## (c) Posterior summaries (5 points)

For *your best-mixing* configuration of each algorithm, report the posterior mean, posterior standard deviation, and a 95% credible interval for $\beta_0$, $\beta_1$, and the 10-year-failure log-rate at 130°C, i.e. $y^\star = \beta_0 + \beta_1 x^\star$ with $x^\star = 1000/(130 + 273.2)$. Translate this back to the predicted median failure time $f^\star = 10^{y^\star}$ hours and convert to years (1 year = 8,760 hours).

## (d) Synthesis (5 points)

Produce a single summary table of the three algorithms at their respective best tunings:

| **Algorithm** | **Acc. rate** | ESS($\beta_0$) | ESS($\beta_1$) | **time / ESS** |
|---|---|---|---|---|
| RW-MH, single-move ($\tau^\star$) | ____ | ____ | ____ | ____ |
| RW-MH, block move ($\tau^\star$) | ____ | ____ | ____ | ____ |
| Independence MH (OLS proposal) | ____ | ____ | ____ | ____ |

Discuss:

- Why does the single-coordinate random walk mix so poorly for this target? (Hint: look at the orientation and aspect ratio of the posterior in the scatter plot of (b).)
- Why does the OLS-based independence proposal work so well here, and what would break it? (Consider, for instance, increasing the censoring fraction or shrinking the uncensored block to $m = 5$.)
- Which algorithm would you reach for in a real applied problem with the same structure (censored Gaussian regression, flat prior on the regression coefficients, $\sigma^2$ known)? Justify briefly.

---

*End of assignment.*