5.4 Metropolis-within-Gibbs

Combining Gibbs updates with MH steps

Why Metropolis-within-Gibbs?

- We have two core MCMC tools:
  - **Gibbs sampling**: update a parameter by sampling *exactly* from its full conditional.
  - **Metropolis–Hastings (MH)**: update by proposing then accepting/rejecting when the full conditional is not easy to sample from.
- Many realistic posteriors mix both situations:
  - some full conditionals are standard (Gamma, Normal, etc.) $\Rightarrow$ Gibbs steps,
  - others have no closed form $\Rightarrow$ MH steps.
- **Metropolis-within-Gibbs** (MwG): cycle through parameters, using a Gibbs update when possible, otherwise an MH update.

Why Metropolis-within-Gibbs?

- We have two core MCMC tools:
  - **Gibbs sampling**: update a parameter by sampling *exactly* from its full conditional.
  - **Metropolis–Hastings (MH)**: update by proposing then accepting/rejecting when the full conditional is not easy to sample from.
- Many realistic posteriors mix both situations:
  - some full conditionals are standard (Gamma, Normal, etc.) $\Rightarrow$ Gibbs steps,
  - others have no closed form $\Rightarrow$ MH steps.
- **Metropolis-within-Gibbs** (MwG): cycle through parameters, using a Gibbs update when possible, otherwise an MH update.

Generic MwG pattern (2 parameters)

At iteration i:

$$\theta^{(i)} \sim \pi(\theta \mid \beta^{(i-1)}, y), \qquad \beta^{(i)} \leftarrow \text{MH targeting } \pi(\beta \mid \theta^{(i)}, y)$$

(or the opposite order).

Reminder: Full conditionals

Suppose parameters are $(\theta_1, \dots, \theta_d)$.

- The **full conditional** of θ_k is
$$\pi(\theta_k \mid \theta_{-k}, y),$$
where θ_{-k} denotes all parameters except θ_k.
- In Gibbs sampling, we iterate
$$\theta_k^{(i)} \sim \pi(\theta_k \mid \theta_{-k}^{(\text{most recent})}, y).$$
- In MwG, for parameters where $\pi(\theta_k \mid \theta_{-k}, y)$ is not tractable to sample from, we perform an MH step that leaves that full conditional invariant.

Reminder: Full conditionals

Suppose parameters are $(\theta_1, \ldots, \theta_d)$.

- The **full conditional** of θ_k is

$$\pi(\theta_k \mid \theta_{-k}, y),$$

  where θ_{-k} denotes all parameters except θ_k.
- In Gibbs sampling, we iterate

$$\theta_k^{(i)} \sim \pi(\theta_k \mid \theta_{-k}^{(\text{most recent})}, y).$$

- In MwG, for parameters where $\pi(\theta_k \mid \theta_{-k}, y)$ is not tractable to sample from, we perform an MH step that leaves that full conditional invariant.

Key message

You do *not* need the full conditional in normalised form for MH. You only need it up to a proportionality constant.

Example 5.9: Weibull model + priors

Data model:

$$Y_1, \ldots, Y_N \mid (\beta, \theta) \overset{\text{iid}}{\sim} \text{Weibull}(\beta, \theta), \qquad y_i > 0,\ \beta > 0,\ \theta > 0.$$

Weibull density (as given in the notes):

$$\pi(y \mid \beta, \theta) = \frac{\beta}{\theta} \left(\frac{y}{\theta}\right)^{\beta - 1} \exp\left(-\frac{y^\beta}{\theta}\right).$$

Example 5.9: Weibull model + priors

Data model:

$$Y_1, \ldots, Y_N \mid (\beta, \theta) \overset{\text{iid}}{\sim} \text{Weibull}(\beta, \theta), \qquad y_i > 0,\ \beta > 0,\ \theta > 0.$$

Weibull density (as given in the notes):

$$\pi(y \mid \beta, \theta) = \frac{\beta}{\theta} \left(\frac{y}{\theta}\right)^{\beta-1} \exp\left(-\frac{y^\beta}{\theta}\right).$$

Independent priors:

- $\beta \sim \text{Exp}(\lambda)$, i.e.
$$\pi(\beta) \propto e^{-\lambda\beta}, \qquad \beta > 0.$$
- $\theta \sim \text{Inv-Gamma}(a, b)$ in the form
$$\pi(\theta) \propto \theta^{-(a+1)} \exp\left(-\frac{b}{\theta}\right), \qquad \theta > 0.$$

Likelihood for the Weibull sample

For $y = (y_1, \ldots, y_N)$, the likelihood is

$$\pi(y \mid \beta, \theta) = \prod_{i=1}^{N} \pi(y_i \mid \beta, \theta).$$

Likelihood for the Weibull sample

For $y = (y_1, \ldots, y_N)$, the likelihood is

$$\pi(y \mid \beta, \theta) = \prod_{i=1}^{N} \pi(y_i \mid \beta, \theta).$$

Using the given density:

$$\begin{aligned} \pi(y \mid \beta, \theta) &= \prod_{i=1}^{N} \left[\frac{\beta}{\theta} \left(\frac{y_i}{\theta} \right)^{\beta-1} \exp\left(-\frac{y_i^{\beta}}{\theta} \right) \right] \\ &= \left(\frac{\beta}{\theta} \right)^{N} \left(\prod_{i=1}^{N} \left(\frac{y_i}{\theta} \right)^{\beta-1} \right) \exp\left(-\frac{1}{\theta} \sum_{i=1}^{N} y_i^{\beta} \right). \end{aligned}$$

Likelihood for the Weibull sample

For $y = (y_1, \ldots, y_N)$, the likelihood is

$$\pi(y \mid \beta, \theta) = \prod_{i=1}^{N} \pi(y_i \mid \beta, \theta).$$

Using the given density:

$$\begin{aligned}
\pi(y \mid \beta, \theta) &= \prod_{i=1}^{N} \left[\frac{\beta}{\theta} \left(\frac{y_i}{\theta} \right)^{\beta - 1} \exp\left(-\frac{y_i^{\beta}}{\theta} \right) \right] \\
&= \left(\frac{\beta}{\theta} \right)^{N} \left(\prod_{i=1}^{N} \left(\frac{y_i}{\theta} \right)^{\beta - 1} \right) \exp\left(-\frac{1}{\theta} \sum_{i=1}^{N} y_i^{\beta} \right).
\end{aligned}$$

We often work *up to proportionality* in (β, θ):

$$\pi(y \mid \beta, \theta) \;\propto\; \beta^{N} \, \theta^{-N\beta} \left(\prod_{i=1}^{N} y_i^{\beta - 1} \right) \exp\left(-\frac{1}{\theta} \sum_{i=1}^{N} y_i^{\beta} \right).$$

Posterior (up to proportionality)

By Bayes' rule:

$$\pi(\beta, \theta \mid y) \;\propto\; \pi(y \mid \beta, \theta)\, \pi(\beta)\, \pi(\theta).$$

Posterior (up to proportionality)

By Bayes' rule:

$$\pi(\beta, \theta \mid y) \;\propto\; \pi(y \mid \beta, \theta)\, \pi(\beta)\, \pi(\theta).$$

Substitute the (unnormalised) likelihood and priors:

$$\pi(\beta, \theta \mid y) \propto \left[\beta^{N}\, \theta^{-N\beta} \left(\prod_{i=1}^{N} y_i^{\beta-1}\right) \exp\left(-\frac{1}{\theta}\sum_{i=1}^{N} y_i^{\beta}\right)\right] \cdot e^{-\lambda\beta} \cdot \theta^{-(a+1)} e^{-b/\theta}.$$

Posterior (up to proportionality)

By Bayes' rule:

$$\pi(\beta, \theta \mid y) \;\propto\; \pi(y \mid \beta, \theta)\, \pi(\beta)\, \pi(\theta).$$

Substitute the (unnormalised) likelihood and priors:

$$\pi(\beta, \theta \mid y) \propto \left[\beta^N \, \theta^{-N\beta} \left(\prod_{i=1}^{N} y_i^{\beta-1}\right) \exp\left(-\frac{1}{\theta} \sum_{i=1}^{N} y_i^{\beta}\right)\right] \cdot e^{-\lambda\beta} \cdot \theta^{-(a+1)} e^{-b/\theta}.$$

Collect θ-terms:

$$\pi(\beta, \theta \mid y) \;\propto\; \beta^N \left(\prod_{i=1}^{N} y_i^{\beta-1}\right) e^{-\lambda\beta} \cdot \theta^{-(N\beta+a+1)} \exp\left(-\frac{b + \sum_{i=1}^{N} y_i^{\beta}}{\theta}\right).$$

Full conditional for θ: conjugate

Treat β as fixed. From the posterior expression:

$$\pi(\theta \mid \beta, y) \;\propto\; \theta^{-(N\beta+a+1)} \exp\left(-\frac{b + \sum_{i=1}^{N} y_i^{\beta}}{\theta}\right).$$

Full conditional for θ: conjugate

Treat β as fixed. From the posterior expression:

$$\pi(\theta \mid \beta, y) \;\propto\; \theta^{-(N\beta+a+1)} \exp\left(-\frac{b + \sum_{i=1}^{N} y_i^{\beta}}{\theta}\right).$$

This is the kernel of an inverse-gamma distribution.

Result

$$\theta \mid (\beta, y) \;\sim\; \text{Inv-Gamma}(\alpha_\theta,\; s_\theta)$$

with

$$\alpha_\theta = N\beta + a, \qquad s_\theta = b + \sum_{i=1}^{N} y_i^{\beta},$$

under the parameterisation

$$\pi(\theta) \propto \theta^{-(\alpha_\theta+1)} \exp(-s_\theta/\theta).$$

Full conditional for θ: conjugate

Treat β as fixed. From the posterior expression:

$$\pi(\theta \mid \beta, y) \;\propto\; \theta^{-(N\beta+a+1)} \exp\left(-\frac{b + \sum_{i=1}^{N} y_i^{\beta}}{\theta}\right).$$

This is the kernel of an inverse-gamma distribution.

Result

$$\theta \mid (\beta, y) \;\sim\; \text{Inv-Gamma}(\alpha_\theta,\; s_\theta)$$

with

$$\alpha_\theta = N\beta + a, \qquad s_\theta = b + \sum_{i=1}^{N} y_i^{\beta},$$

under the parameterisation

$$\pi(\theta) \propto \theta^{-(\alpha_\theta+1)} \exp(-s_\theta/\theta).$$

- So θ can be updated by a **Gibbs step**: sample directly from this Inv-Gamma.

Full conditional for β: no easy closed form

Treat θ as fixed. The full conditional kernel is

$$\pi(\beta \mid \theta, y) \;\propto\; \beta^N \left(\prod_{i=1}^{N} y_i^{\beta-1} \right) \theta^{-N\beta} \exp\left(-\frac{1}{\theta} \sum_{i=1}^{N} y_i^{\beta} \right) \exp(-\lambda\beta), \quad \beta > 0.$$

Full conditional for β: no easy closed form

Treat θ as fixed. The full conditional kernel is

$$\pi(\beta \mid \theta, y) \;\propto\; \beta^{N} \Big(\prod_{i=1}^{N} y_i^{\beta-1} \Big)\, \theta^{-N\beta}\, \exp\!\left(-\frac{1}{\theta} \sum_{i=1}^{N} y_i^{\beta} \right) \exp(-\lambda\beta), \quad \beta > 0.$$

- There is **no standard distribution** you can sample from directly here.
- So we use an **MH step** targeting $\pi(\beta \mid \theta, y)$.

Full conditional for β: no easy closed form

Treat θ as fixed. The full conditional kernel is

$$\pi(\beta \mid \theta, y) \;\propto\; \beta^N \Big(\prod_{i=1}^{N} y_i^{\beta-1}\Big)\, \theta^{-N\beta} \exp\!\left(-\frac{1}{\theta}\sum_{i=1}^{N} y_i^{\beta}\right) \exp(-\lambda\beta), \quad \beta > 0.$$

- There is **no standard distribution** you can sample from directly here.
- So we use an **MH step** targeting $\pi(\beta \mid \theta, y)$.

Practical note

Compute acceptance ratios using the **log** full conditional to avoid numerical underflow/overflow.

MH step for β: symmetric random-walk proposal

A simple proposal (as in the notes):

$$\beta' \sim \mathrm{U}\left[\beta^{(i-1)} - \varepsilon,\ \beta^{(i-1)} + \varepsilon\right].$$

MH step for β: symmetric random-walk proposal

A simple proposal (as in the notes):

$$\beta' \sim \mathrm{U}\left[\beta^{(i-1)} - \varepsilon,\ \beta^{(i-1)} + \varepsilon\right].$$

- Support constraint: $\beta > 0$.
  - If $\beta' \leq 0$, reject immediately and set $\beta^{(i)} = \beta^{(i-1)}$.
- Proposal density is symmetric:

$$q(\beta' \mid \beta) = q(\beta \mid \beta').$$

- Therefore the MH acceptance probability simplifies to

$$p_{\text{acc}} = \min\left\{1,\ \frac{\pi(\beta' \mid \theta^{(i-1)}, y)}{\pi(\beta^{(i-1)} \mid \theta^{(i-1)}, y)}\right\}.$$

Why does symmetry help?

In MH,

$$p_{\mathsf{acc}} = \min\left\{1,\ \frac{\pi(\beta' \mid \theta, y)}{\pi(\beta \mid \theta, y)} \cdot \frac{q(\beta \mid \beta')}{q(\beta' \mid \beta)}\right\}.$$

If $q(\beta \mid \beta') = q(\beta' \mid \beta)$, the q-ratio is 1.

Acceptance ratio in log form (recommended)

Define the (unnormalised) log full conditional for β:

$$\ell(\beta; \theta, y) := \log \pi(\beta \mid \theta, y) + \text{constant in } \beta.$$

Acceptance ratio in log form (recommended)

Define the (unnormalised) log full conditional for β:

$$\ell(\beta;\theta,y) := \log \pi(\beta \mid \theta, y) + \text{constant in } \beta.$$

From the kernel:

$$\ell(\beta;\theta,y) = N\log\beta + \sum_{i=1}^{N}(\beta-1)\log y_i - N\beta\log\theta - \frac{1}{\theta}\sum_{i=1}^{N} y_i^{\beta} - \lambda\beta.$$

Acceptance ratio in log form (recommended)

Define the (unnormalised) log full conditional for β:

$$\ell(\beta; \theta, y) := \log \pi(\beta \mid \theta, y) + \text{constant in } \beta.$$

From the kernel:

$$\ell(\beta; \theta, y) = N \log \beta + \sum_{i=1}^{N} (\beta - 1) \log y_i - N\beta \log \theta - \frac{1}{\theta} \sum_{i=1}^{N} y_i^{\beta} - \lambda\beta.$$

Then

$$\log r = \ell(\beta'; \theta, y) - \ell(\beta; \theta, y), \qquad p_{\text{acc}} = \min\{1, \exp(\log r)\}.$$

Acceptance ratio in log form (recommended)

Define the (unnormalised) log full conditional for β:

$$\ell(\beta;\theta,y) := \log \pi(\beta \mid \theta, y) + \text{constant in } \beta.$$

From the kernel:

$$\ell(\beta;\theta,y) = N\log\beta + \sum_{i=1}^{N}(\beta-1)\log y_i - N\beta\log\theta - \frac{1}{\theta}\sum_{i=1}^{N} y_i^{\beta} - \lambda\beta.$$

Then

$$\log r = \ell(\beta';\theta,y) - \ell(\beta;\theta,y), \qquad p_{\text{acc}} = \min\{1, \exp(\log r)\}.$$

- Compute $\sum y_i^{\beta}$ carefully.
- Using logs avoids numerical issues, especially when N is large.

Putting it together: Metropolis-within-Gibbs algorithm

Algorithm (one iteration)

Given current $(\beta^{(i-1)}, \theta^{(i-1)})$:

1. **MH update for β:**
   1. Propose $\beta' \sim \mathrm{U}[\beta^{(i-1)} - \varepsilon,\ \beta^{(i-1)} + \varepsilon]$.
   2. If $\beta' \leq 0$, reject.
   3. Else accept with probability
   $$p_{\text{acc}} = \min\left\{1,\ \frac{\pi(\beta' \mid \theta^{(i-1)}, y)}{\pi(\beta^{(i-1)} \mid \theta^{(i-1)}, y)}\right\}.$$
   4. Set $\beta^{(i)} = \beta'$ if accepted; otherwise $\beta^{(i)} = \beta^{(i-1)}$.
2. **Gibbs update for θ:**
$$\theta^{(i)} \sim \text{Inv-Gamma}\left(N\beta^{(i)} + a,\ b + \sum_{j=1}^{N} y_j^{\beta^{(i)}}\right).$$

Putting it together: Metropolis-within-Gibbs algorithm

Algorithm (one iteration)

Given current $(\beta^{(i-1)}, \theta^{(i-1)})$:

1. **MH update for β:**
   1. Propose $\beta' \sim \mathrm{U}[\beta^{(i-1)} - \varepsilon,\ \beta^{(i-1)} + \varepsilon]$.
   2. If $\beta' \leq 0$, reject.
   3. Else accept with probability

$$p_{\text{acc}} = \min\left\{1,\ \frac{\pi(\beta' \mid \theta^{(i-1)}, y)}{\pi(\beta^{(i-1)} \mid \theta^{(i-1)}, y)}\right\}.$$

   4. Set $\beta^{(i)} = \beta'$ if accepted; otherwise $\beta^{(i)} = \beta^{(i-1)}$.
2. **Gibbs update for θ:**

$$\theta^{(i)} \sim \text{Inv-Gamma}\left(N\beta^{(i)} + a,\ b + \sum_{j=1}^{N} y_j^{\beta^{(i)}}\right).$$

Summary

- Metropolis-within-Gibbs combines:
  - Gibbs steps for parameters with tractable full conditionals,
  - MH steps for parameters without closed-form full conditionals.
- In Example 5.9 (Weibull):
  - $\theta \mid (\beta, y)$ is inverse-gamma $\Rightarrow$ Gibbs update,
  - $\beta \mid (\theta, y)$ has no closed form $\Rightarrow$ MH update.
- Symmetric random-walk proposals make the MH acceptance ratio simple:

$$p_{\text{acc}} = \min\left\{1, \frac{\pi(\beta' \mid \theta, y)}{\pi(\beta \mid \theta, y)}\right\}.$$