Lecture 9B
MTH6102: Bayesian Statistical Methods

Eftychia Solea

Queen Mary University of London

2023

Today's agenda

Today's lecture

- Review
- Metropolis-Hastings in Bayesian inference to generate samples from the posterior pdf.

MH for Bayesian inference

- **Goal:** Generate a sample $\theta_1, \theta_2, \ldots$ from the posterior pdf, $p(\theta \mid y)$.
- $p(\theta \mid y)$ is called the target distribution.
- Last time we saw that posterior densities can be complicated when not using a conjugate prior distribution.
- It is difficult to find the normalising constant with a non-conjugate prior distribution, and hence we cannot simulate directly from $p(\theta \mid y)$.

MH for Bayesian inference

- In these cases, MCMC are helpful.
- Metropolis-Hastings is a special case of a MCMC that can generate a sample $\theta_1, \theta_2, \ldots$ that is **approximately** from $p(\theta \mid y)$.
- The sample $\theta_1, \theta_2, \ldots$ is a Markov chain whose distribution converges to $p(\theta \mid y)$ (under some conditions).

Metropolis-Hastings algorithm

The algorithm constructs $\theta_1, \theta_2, \ldots$ as follows.

Start with arbitrary θ_1. Suppose we have generated $\{\theta_1, \ldots, \theta_i\}$. To generate θ_{i+1} do the following

1. Generate a proposal random variable ψ from distribution $q(\psi \mid \theta_i)$.
2. Compute the acceptance probability

$$r = \min\left\{1, \frac{p(\psi \mid y)q(\theta_i \mid \psi)}{p(\theta_i \mid y)q(\psi \mid \theta_i)}\right\}.$$

3. Set

$$\theta_{i+1} = \begin{cases} \psi & \text{with probability } r, \\ \theta_i & \text{with probability } 1-r. \end{cases}$$

   In practice, generate $U \sim U[0,1]$. If $U < r$, set $\theta_{i+1} = \psi$, otherwise set $\theta_{i+1} = \theta_i$.

Metropolis-Hastings algorithm terminology

- q is the proposal distribution: At each step, we propose a new rv ψ using the conditional distribution $q(\cdot \mid \theta_i)$ that depends on θ_i (not on the past).

- MH accepts ψ with probability

$$r = \min \left\{ 1, \frac{p(\psi \mid y) q(\theta_i \mid \psi)}{p(\theta_i \mid y) q(\psi \mid \theta_i)} \right\}$$

  called the acceptance probability.

- r reflects how likely it is that ψ is from $p(\theta \mid y)$.

Symmetric Metropolis-Hastings algorithm

- The simplest case uses a symmetric proposal distribution, that is $q(\psi \mid \theta_i) = q(\theta_i \mid \psi)$.
- In this case, the acceptance probability simplifies to

$$r = \min\left\{1, \frac{p(\psi \mid y)}{p(\theta_i \mid y)}\right\}.$$

- Does not involve the proposal density at all.
- Some common examples of symmetric q: $\psi \sim N(\theta, b^2)$, $\psi \sim U[\theta - a, \theta + a]$ for some $a > 0$

Metropolis-Hastings algorithm for Bayesian inference

- In Bayesian inference, the posterior density is

$$p(\theta \mid y) = \frac{p(\theta)\, p(y \mid \theta)}{\int p(\theta)\, p(y \mid \theta)\, d\theta} = \frac{p(\theta)\, p(y \mid \theta)}{T}.$$

- It's difficult to find the normalizing constant

$$T = \int p(\theta)\, p(y \mid \theta)\, d\theta.$$

- We don't need to find this: The acceptance probability **does not depend on the normalizing constant**

$$\begin{aligned} r &= \min\left\{1, \frac{p(\psi \mid y) q(\theta_i \mid \psi)}{p(\theta_i \mid y) q(\psi \mid \theta_i)}\right\} \\ &= \min\left\{1, \frac{p(\psi)\, p(y \mid \psi)}{p(\theta_i)\, p(y \mid \theta_i)} \frac{q(\theta_i \mid \psi)}{q(\psi \mid \theta_i)}\right\}. \end{aligned}$$

- so we only need to know $p(\theta \mid y)$ up to a constant.

Metropolis-Hastings algorithm for Bayesian inference

Define $g(\theta) = p(\theta)\, p(y \mid \theta)$, the non-normalized posterior density or the Bayes numerator.

Generate $\theta_1, \theta_2, \ldots$ as follows:

- Start with θ_1, where $g(\theta_1) > 0$.
- For each $i > 1$:
  - Generate $\psi \sim q(\psi \mid \theta_i)$.
  - Let
$$r = \min\left\{1, \frac{g(\psi)}{g(\theta_i)} \frac{q(\theta_i \mid \psi)}{q(\psi \mid \theta_i)}\right\}.$$
  - Set
$$\theta_{i+1} = \begin{cases} \psi & \text{with probability } r \\ \theta_i & \text{with probability } 1 - r \end{cases}$$

Metropolis-Hastings algorithm for Bayesian inference

- Metropolis-Hastings algorithm generates a sequence $\theta^{(1)}, \ldots,$ of dependent or correlated θ values.
  - e.g., θ_{i+1} is correlated with θ_i because ψ has been rejected.

- Also, $\theta^{(1)}, \ldots,$ is Markov chain since each ψ is generated from $q(\psi \mid \theta_i)$ that depends on the last accepted value θ_i.

- In practice we cannot run the Markov chain forever but for some large number of steps N.

Metropolis-Hastings algorithm for Bayesian inference

- But we can still use the sample $\theta^{(i)}$, $i = 1, 2, \ldots, N$ to make inferences about the posterior.
- Under mild conditions, the empirical distribution of $\theta^{(i)}$, $i = 1, 2, \ldots, N$ will approximate well the posterior for large N.
- We can view $\theta^{(i)}$, $i = 1, 2, \ldots, N$ as a sample from the posterior $p(\theta|y)$.
- Hence, we can approximate posterior means, quantiles and other posterior quantities of interest using $\{\theta^{(1)}, \ldots, \theta^{(N)}\}$ for large N.

Example: binomial data/beta prior

- Let $k = 12 \sim \text{binomial}(40, q)$, where q is the probability of success.
- $q \sim \text{beta}(2, 2)$.

1. Apply the Metropolis-Hastings algorithm to simulate from the posterior $p(q|k)$ using normal proposal distribution with with standard deviation $b = 0.05$.
2. Plot the histogram of the chain and compare it with the true posterior
3. Compute the sample posterior mean, sample posterior median and sample equal-tail interval and compare with the true posterior summaries.

Working on the log scale

- The likelihood is typically a product of many terms.

$$p(y \mid \theta) = \prod_{i=1}^{n} p(y_i \mid \theta)$$

- For numerical stability, we usually do the computations using the log of the posterior density.
- So calculate

$$\log\left(p(y \mid \theta)\right) = \sum_{i=1}^{n} \log\left(p(y_i \mid \theta)\right)$$

Symmetric MH using the log scale

- Define $\mathcal{L}(\theta) = \log\left(p(\theta)\, p(y \mid \theta)\right) = \log\left(p(\theta)\right) + \log\left(p(y \mid \theta)\right)$, the log of the posterior density (up to a constant).
- To work on the log scale, the part of the algorithm with the acceptance probability changes.
- Define
$$\delta = \min\left(0, \mathcal{L}(\psi) - \mathcal{L}(\theta_{i-1})\right)$$
- Generate $u \sim \text{Uniform}(0, 1)$
- Set
$$\theta_{i+1} = \begin{cases} \psi & \text{if } \log(u) < \delta \\ \theta_i & \text{otherwise} \end{cases}$$

Example: Normal example with known variance

- $Y_1, \ldots, Y_n$ iid from $N(\theta, \sigma^2)$ where σ^2 is known.
- $\theta \sim N(\mu, \tau^2)$ with τ^2 known.
- Apply the Metropolis-Hastings algorithm on the log-scale to simulate from the posterior $p(\theta|y_1, \ldots, y_n)$ after observing $Y = y = (y_1, \ldots, y_n)$.
- Use $q(\psi \mid \theta) \sim N(\theta, b^2)$ with $b = 2$, and $q(\psi \mid \theta) \sim U(\theta - 4, \theta + 4)$.

Board example: binomial data/beta prior

- Let $k = 12 \sim \text{binomial}(40, q)$, where q is the probability of success.
- $q \sim \text{beta}(2, 2)$.

- Apply the Metropolis-Hastings algorithm on the log-scale to simulate from the posterior $p(q|k)$ using normal proposal distribution with standard deviation $b = 0.06$.