MSA101/MVE187 2022 Lecture 6
MCMC. Random walk. Independent proposal. Convergence. Checking convergence. Burn-in. Smart proposals. Examples.

Petter Mostad

Chalmers University

September 14, 2022

Review: The Metropolis-Hastings algorithm

Given a probability density f that we want to simulate from. Construct a *proposal function* $q(y \mid x)$ which for every x gives a probability density for a proposed new value y. The algorithm starts with a choice of an initial value $x^{(0)}$ for x, and then simulates each $x^{(t)}$ based on $x^{(t-1)}$. Specifically, given $x^{(t)}$,

- Simulate a new value y according to $q(y \mid x^{(t)})$.
- Compute the acceptance probability

$$\rho(x^{(t)}, y) = \min\left(\frac{f(y)q(x^{(t)} \mid y)}{f(x^{(t)})q(y \mid x^{(t)})}, 1\right).$$

- Set

$$x^{(t+1)} = \begin{cases} y & \text{with probability } \rho(x^{(t)}, y) \\ x^{(t)} & \text{with probability } 1 - \rho(x^{(t)}, y) \end{cases}$$

Plan for today

- A simple 1D example.
- Types of proposal functions. How to choose proposal functions.
- More examples.
- Assessing convergence.
- Burn-in and thinning.

First example

Imagine you have a 1D (posterior) density proportional to

$$f(x) = 100x^2 \exp(-x^2 + x) + x^3 \exp(-(x-4)^4)$$

for $x \geq 0$.
Generate an (approximate) sample using Metropolis Hastings.
Explore some effects of changing the proposal function.

- (Concepts: Trace plot. Acceptance rate.)

Symmetric proposal functions ("Metropolis" algorithm)

- The case where the proposal is symmetric:

$$q(x \mid y) = q(y \mid x)$$

  for all pairs (x, y).

- The acceptance probability for MH becomes

$$\rho(x^{(t)}, y) = \min\left(\frac{f(y)q(x^{(t)} \mid y)}{f(x^{(t)})q(y \mid x^{(t)})}, 1\right) = \min\left(\frac{f(y)}{f(x^{(t)})}, 1\right).$$

- Often implemented by adding to the current value $x^{(t)}$ of the Markov chain a random variable with symmetric density $\pi(u) = \pi(-u)$. (Called "random walk"). Often (but not always) a normal random variable with expectation zero.
- Other changes can be included (such as taking the absolute value in the example), as long as we ensure $q(x \mid y) = q(y \mid x)$.

Independent proposal functions

- A simple special case is when $q(y \mid x)$ does not depend on x; i.e. proposals are independently generated from $q(y)$.
- The acceptance probability is

$$\rho(x^{(t)}, y) = \min\left(1, \frac{f(y)q(x^{(t)})}{f(x^{(t)})q(y)}\right).$$

- The sequence of generated values are *not* independent: When the proposed value is not accepted, the new value in the chain is equal to the old.
- Note that the method works well if the proposal distribution is close to the target distribution.
- Note for example that, if the ratio $f(x)/q(x)$ is unbounded, the chain can become stuck in such point where this ratio is too high. Then the convergence can be very bad.

General proposal functions

- Note that the proposal function $q(y \mid x^{(t)})$ can have any form, also asymmetric, as long as the resulting Markov chain becomes ergodic.
- We will see examples of when this is very useful.
- Note also that at each iteration we may (randomly) choose between different proposal functions; the complete procedure becomes a combined proposal function.
- Just remember to make sure the resulting Markov chain is ergodic.

A bivariate example

Assume our (posterior) density is proportional to

$$f(x, y) = \exp(-40(x-0.3)^2-10(y-0.2)^2)+\exp(-15(x-0.6)^2-30(y-0.7)^2)$$

with the limitations $0 < x < 1$ and $0 < y < 1$.
Find an (approximate) sample using Metropolis Hastings.

Using Metropolis Hastings for Bayesian inference

If we go back to the notation y_{pred}, y_{data}, and θ: Predict y_{pred} using the data y_{data} and a model that also involves a parameter vector θ:

- Use that the posterior is proportional to $\pi(y_{\text{data}} \mid \theta)\pi(\theta)$ and use MH to get a posterior sample $\theta_1, \dots, \theta_N$.
- In the MH loop, while we simulate the θ_i, we may also simulate from $\pi(y_{\text{pred}} \mid \theta_i)$, or compute $\pi(y_{\text{pred}} \mid \theta_i)$ for a fixed value of y_{pred}.

Example: Braking distance of cars

We are given data which, for $i = 1, \ldots, 50$ cars (in the 1920s) lists their speed x_i (in mph) and braking distance y_i. From this, we would like to predict the braking distance at speed 21 mph.

- First step of analysis: Decide on Y_{data} and Y_{pred}: Done.
- Next step: Explore the data and the context, and decide on a reasonable model.
- We come up with the following model:

  $$y_i = \theta_1 + \theta_2 \cdot x_i + \theta_3 \cdot x_i^2 + \epsilon_i$$

  where $\epsilon_i \sim \text{Normal}(0, \theta_4^2)$.
- For simplicity, we use a flat (improper) prior on the parameters $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$, with $\theta_4 > 0$.

Example: Braking distance of cars, cont.

- The posterior becomes *proportional to* the likelihood:

$$f(\theta) = \prod_{i=1}^{50} \text{Normal}(y_i; \theta_1 + \theta_2 \cdot x_i + \theta_3 \cdot x_i^2, \theta_4^2)$$

- We simulate using a Random Walk Metropolis Hastings MCMC.
- Note that, together with simulation of the parameters, we also simulate the braking distance at speed 21 at these parameters, as this is what we want to predict. We can then use these values, with Monte Carlo integration, to answer or original question.