Markov Chain Monte Carlo: Metropolis-Hastings

Module DES130: Computational Statistics

Dr. Erick A. Chacón Montalván (echacon@uni.edu.pe)

Escuela de Profesional de Ingeniería Estadística
Facultad de Ingeniería Económica, Estadística y Ciencias Sociales
Universidad Nacional de Ingeniería (UNI)
Lima – Perú

Motivation

Motivation

- Gibbs sampling need knowledge of the conditional distributions which are not always tractable.
- Gibss sampling might not have good mixing in various cases.
- Metropolis-Hastings sampling provides a wider family of algorithms for Markov chain Monte Carlo.
- Most MCMC are based on Metropolis-Hastings algorithm including Gibss sampling.

Metropolis-Hastings

Concepts

- $\pi(\boldsymbol{x})$: A multivariate *target* density function.
- $q(\boldsymbol{x}, \boldsymbol{y})$: The *proposal density* of moving from $\boldsymbol{x}$ to $\boldsymbol{y}$. You can think of this as a conditional proposal density function $\tilde{\pi}(\boldsymbol{y} \mid \boldsymbol{x})$.
- $\alpha(\boldsymbol{x}, \boldsymbol{y})$: The *acceptance probability* of $\boldsymbol{y}$ given $\boldsymbol{x}$. You can think of this as a conditional acceptance probability $\tilde{\alpha}(\boldsymbol{y} \mid \boldsymbol{x})$.
- P_{xy}: The *transition probability* of moving from $\boldsymbol{x}$ to $\boldsymbol{y}$. Notice that to move from $\boldsymbol{x}$ to $\boldsymbol{y}$ we need to (i) propose $\boldsymbol{y}$ given $\boldsymbol{x}$ and (ii) accept $\boldsymbol{y}$ given $\boldsymbol{x}$. Then $P_{xy} = q(\boldsymbol{x}, \boldsymbol{y})\alpha(\boldsymbol{x}, \boldsymbol{y})$ for $\boldsymbol{y}$.

Metropolis-Hastings algorithm

Algorithm

1. Given the current state $\boldsymbol{x}_{n-1} = \boldsymbol{x}$, generate a candidate value $\boldsymbol{y}$ using the transition probability $q(\boldsymbol{x}, \boldsymbol{y})$.
2. Compute the acceptance probability of $\boldsymbol{y}$ given current state $\boldsymbol{x}$ as follows

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} \min\left(\frac{\pi(\boldsymbol{y})q(\boldsymbol{y},\boldsymbol{x})}{\pi(\boldsymbol{x})q(\boldsymbol{x},\boldsymbol{y})}, 1\right) & \text{if } \pi(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{y}) > 0 \\ 1 & \text{if } \pi(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{y}) = 0. \end{cases}$$

3. Set $x_n = \boldsymbol{y}$ with probability $\alpha(\boldsymbol{x}, \boldsymbol{y})$ and set $x_n = \boldsymbol{x}$ otherwise.
4. Repeat until the desired number of samples.

The resulting samples come from $\pi(\boldsymbol{x})$, which is the stationary distribution of the MCMC.

Metropolis-Hastings proof: detailed balance

An easy way to proof that $\pi(\boldsymbol{x})$ is the stationary distribution of the chain is using *detailed balance*.

Detailed balance

$$\begin{aligned}
\pi(\boldsymbol{x})P_{x,y} &= \pi(\boldsymbol{x})q(\boldsymbol{x},\boldsymbol{y})\alpha(\boldsymbol{x},\boldsymbol{y}) \\
&= \pi(\boldsymbol{x})q(\boldsymbol{x},\boldsymbol{y})\min\left(1, \frac{\pi(\boldsymbol{y})q(\boldsymbol{y},\boldsymbol{x})}{\pi(\boldsymbol{x})q(\boldsymbol{x},\boldsymbol{y})}\right) \\
&= \min\left(\pi(\boldsymbol{x})q(\boldsymbol{x},\boldsymbol{y}), \pi(\boldsymbol{y})q(\boldsymbol{y},\boldsymbol{x})\right) \\
&= \pi(\boldsymbol{y})q(\boldsymbol{y},\boldsymbol{x})\min\left(\frac{\pi(\boldsymbol{x})q(\boldsymbol{x},\boldsymbol{y})}{\pi(\boldsymbol{y})q(\boldsymbol{y},\boldsymbol{x})}, 1\right) \\
\pi(\boldsymbol{x})P_{x,y} &= \pi(\boldsymbol{y})P_{y,x}.
\end{aligned}$$

Hence $\pi(\boldsymbol{x})$ is the stationary distribution of the Markov chain.

Metropolis-Hastings proof: marginal

Notice that $P_{xx} = q(\boldsymbol{x}, \boldsymbol{x})\alpha(\boldsymbol{x}, \boldsymbol{x}) + \int q(\boldsymbol{x}, \boldsymbol{z})(1 - \alpha(\boldsymbol{x}, \boldsymbol{z}))d\boldsymbol{z}$. We can assume that $q(\boldsymbol{x}, \boldsymbol{x}) = 0$, then the marginal density is:

Resulting density

$$
\begin{aligned}
\int \pi(\boldsymbol{y})P_{yx}d\boldsymbol{y} &= \int_{y \neq x} \pi(\boldsymbol{y})P_{yx}d\boldsymbol{y} + \pi(\boldsymbol{x})P_{xx} \\
&= \int \pi(\boldsymbol{y})q(\boldsymbol{y}, \boldsymbol{x})\alpha(\boldsymbol{y}, \boldsymbol{x})d\boldsymbol{y} + \int \pi(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{z})(1 - \alpha(\boldsymbol{x}, \boldsymbol{z}))d\boldsymbol{z} \\
&= \int \pi(\boldsymbol{y})q(\boldsymbol{y}, \boldsymbol{x})\alpha(\boldsymbol{y}, \boldsymbol{x})d\boldsymbol{y} + \int \pi(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{z})d\boldsymbol{z} - \int \pi(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{z})\alpha(\boldsymbol{x}, \boldsymbol{z})d\boldsymbol{z} \\
&= \int \min(\pi(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{y}), \pi(\boldsymbol{y})q(\boldsymbol{y}, \boldsymbol{x}))d\boldsymbol{y} + \int \pi(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{z})d\boldsymbol{z} \\
&\quad - \int \min(\pi(\boldsymbol{z})q(\boldsymbol{z}, \boldsymbol{x}), \pi(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{z}))d\boldsymbol{z} \\
&= \int \pi(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{z})d\boldsymbol{z} = \pi(\boldsymbol{x}).
\end{aligned}
$$

Gibbs sampling as a Metropolis-Hastings

Let $x = (x_1, \ldots, x_i, \ldots, x_d)$ and let $y = (x_1, \ldots, x_i^*, \ldots, x_d)$. It is known that

$$\pi(\boldsymbol{x}) = \pi(\boldsymbol{x}_{-i})\pi(x_i \mid \boldsymbol{x}_{-i}).$$

For Gibss sampling that updates each coordinate, we can see that

$$q(\boldsymbol{x}, \boldsymbol{y}) = \pi(x_i^* \mid \boldsymbol{x}_{-i}).$$

Then the acceptance probability is

$$\begin{aligned}
\min\left(\frac{\pi(\boldsymbol{y})q(\boldsymbol{y}, \boldsymbol{x})}{\pi(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{y})}, 1\right) &= \min\left(\frac{\pi(\boldsymbol{y})\pi(x_i \mid \boldsymbol{x}_{-i})}{\pi(\boldsymbol{x})\pi(x_i^* \mid \boldsymbol{x}_{-i})}, 1\right) \\
&= \min\left(\frac{\pi(\boldsymbol{x}_{-i})\pi(x_i^* \mid \boldsymbol{x}_{-i})\pi(x_i \mid \boldsymbol{x}_{-i})}{\pi(\boldsymbol{x}_{-i})\pi(x_i \mid \boldsymbol{x}_{-i})\pi(x_i^* \mid \boldsymbol{x}_{-i})}, 1\right) \\
&= 1.
\end{aligned}$$

Independence sampler

Independence sampler

The proposal density is independent of current state $\boldsymbol{x}$ such as

$$q(\boldsymbol{x}, \boldsymbol{y}) = q(\boldsymbol{y}).$$

This is similar to rejection sampling with proposal density $q(\boldsymbol{y})$ and acceptance probability

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) = \min\left(\frac{\pi(\boldsymbol{y})q(\boldsymbol{x})}{\pi(\boldsymbol{x})q(\boldsymbol{y})}, 1\right) = \min\left(\frac{w(\boldsymbol{y})}{w(\boldsymbol{x})}, 1\right).$$

where $w(\boldsymbol{x}) = \pi(\boldsymbol{x})/q(\boldsymbol{x})$.

Example: mixture of Normal distributions

Let consider the mixture distribution

$$X \sim \begin{cases} N(-1,1), & 0.5, \\ N(3,4), & 0.5. \end{cases}$$

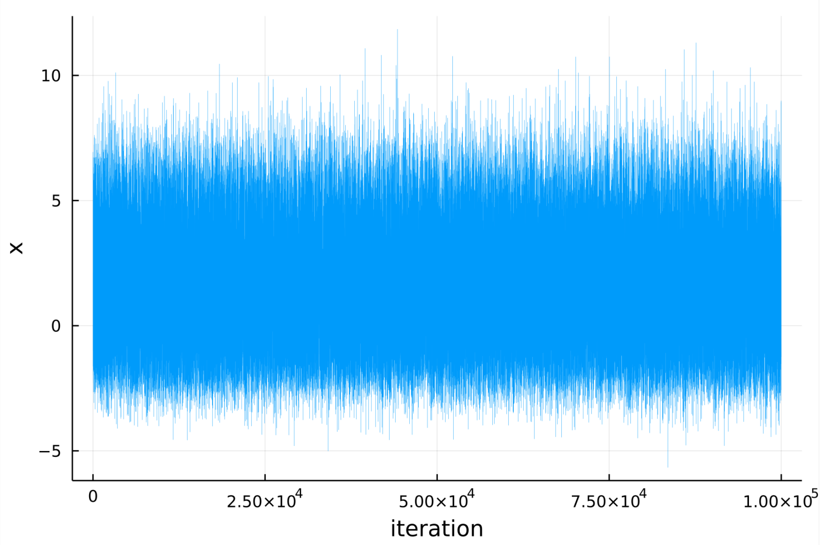
With density

$$\pi(x) = \frac{1}{2}\left\{\frac{1}{\sqrt{2\pi}}\exp(-(x+1)^2/2) + \frac{1}{2\sqrt{2\pi}}\exp(-(x-3)^2/8).\right\}$$

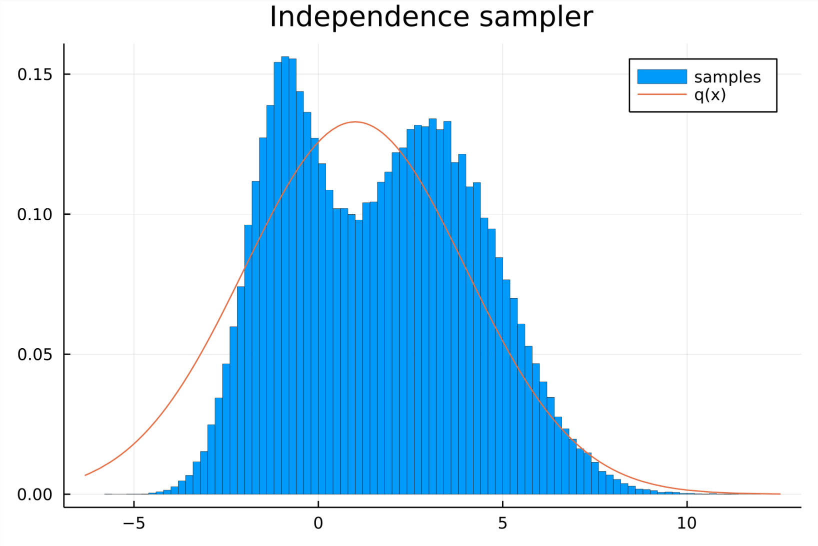
And proposal density $N(1,9)$ such as

$$q(y) = \frac{1}{3\sqrt{2\pi}}\exp(-(y-1)^2/18).$$

Example: mixture of Normal distributions



Example: mixture of Normal distributions



Random walk Metropolis-Hastings

Random walk Metropolis-Hastings

The proposal density is symmetric with respect to current state $\boldsymbol{x}$ such as

$$q(\boldsymbol{x}, \boldsymbol{y}) = q(\boldsymbol{y}, \boldsymbol{x}).$$

Then, the acceptance probability is

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) = \min\left(\frac{\pi(\boldsymbol{y})q(\boldsymbol{y}, \boldsymbol{x})}{\pi(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{y})}, 1\right) = \min\left(\frac{\pi(\boldsymbol{y})}{\pi(\boldsymbol{x})}, 1\right).$$

In this case it is easier to select $q(\boldsymbol{x}, \boldsymbol{y})$ for higher dimensions.

Example: mixture of Normal distributions

Let consider the mixture distribution

$$X \sim \begin{cases} N(-1, 1), & 0.5, \\ N(3, 4), & 0.5. \end{cases}$$

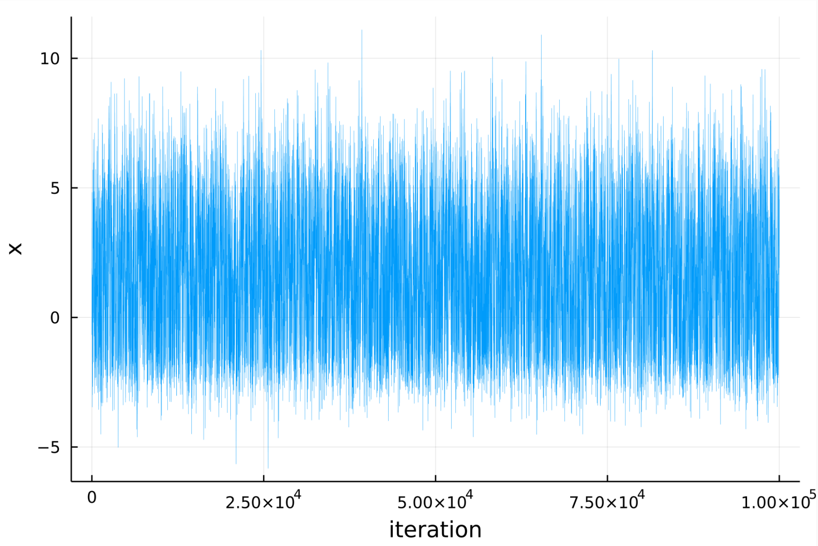
With density

$$\pi(x) = \frac{1}{2}\left\{ \frac{1}{\sqrt{2\pi}} \exp(-(x+1)^2/2) + \frac{1}{2\sqrt{2\pi}} \exp(-(x-3)^2/8) \right\}.$$

And proposal density $N(x, 1)$ such as

$$q(x, y) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(y-x)^2).$$

Example: mixture of Normal distributions



Example: mixture of Normal distributions