STA 4103/5107 Computational Methods in Statistics II

Department of Statistics
Florida State University

Class 8
February 2, 2017

7.3 Metropolis-Hastings Algorithm

Metropolis-Hastings (M-H) Algorithm

- One of the most popular MCMC techniques used in approximate sampling from complicated distributions is the M-H algorithm.

- **Goal: generating samples of a random variable X distributed according to the density $f(x)$.**

- In addition to $f(x)$, we will assume having a density $q(y|x)$ that satisfies the following properties:

  1. It is easy to sample from $q(\cdot|x)$ for all x.
  2. The support of q contains the support of f.
  3. The functional form of $q(y|x)$ is known or $q(y|x)$ is symmetric in y and x. It is not necessary to know the normalizing constant in $q(y|x)$ as long as it does not depend upon x.

M-H Algorithm

- **Algorithm 34 (Metropolis-Hastings Algorithm)** Given $f(x)$ and a choice of $q(y|x)$ that satisfies the above mentioned properties, we choose an initial condition X_0 in the support of $f(x)$. The Markov chain $X_1, X_2, \ldots, X_n$ is constructed iteratively according to the steps:

  1. Generate a candidate $Y \sim q(y|X_t)$.
  2. Update the state to X_{t+1} according to:

$$X_{t+1} = \begin{cases} Y & \text{with probability } \rho(X_t, Y) \\ X_t & \text{with probability } 1 - \rho(X_t, Y) \end{cases}$$

  where $\rho(x, y) = \min\{[f(y)q(x \mid y)]/[f(x)q(y \mid x)], 1\}$.

  $q(y|x)$ is called the proposal density and $\rho(x, y)$ is called the acceptance-rejection function.

Analysis of the Algorithm

- Consider first the case where the ratio
$$[f(y)q(x\,|\,y)]/[f(x)q(y\,|\,x)]$$
values more than one and hence the acceptance-rejection function takes the value one.

- In this case, we set $X_{t+1} = Y$ with probability one.

- In case this ratio goes below one, we set X_{t+1} to Y with probability $\rho(X_t, Y)$.

- Higher the value of $\rho(X_t, Y)$ is, higher are the chances of accepting Y as the new state.

Analysis of the Algorithm

- Note that the normalizing constants in the two densities f and q cancel out and hence are not explicitly needed.

- However, if the normalizing constant for $q(y|x)$ depends upon x, then it does not cancel out and is needed in the expression for ρ.

- In the algorithm, one generates samples from q at every step independently but the elements of the chain are not independent of each other (it is possible to have $X_{t+1} = X_t$ for some t).

- **Theorem 17 For any proposal density that satisfies the three conditions listed earlier, *f* is a stationary probability density of the Markov chain produced by M-H algorithm.**

Special Cases

- 1. In case $q(y|x)$ is symmetric in the two arguments, the expression for the acceptance-rejection function simplifies:

$$\rho(x, y) = \min\{\frac{f(y)}{f(x)}, 1\}.$$

where $f(y)/f(x)$ is often called the likelihood ratio.

- 2. **Independent M-H:** In cases where the proposal density is independent of the current state, i.e. $q(y|x) = q(y)$, then the algorithm is called Independent M-H algorithm. In this case, the acceptance-rejection function becomes:

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

Example

- Implementing the Metropolis-Hastings algorithm to sample a random variable X with the density

$$f(x) = \frac{x^4 e^{-x^3}}{\int_0^\infty x^4 e^{-x^3}\, dx}, \quad x > 0.$$

- Histogram the values attained by the generated Markov chain and compare it to the plot of $f(x)$.

Theorem

- **Theorem 18** For a given density function $f(x)$, and proposal density $q(y|x)$ that satisfies the positivity condition. Let X_t be a Markov chain generated by Metropolis-Hastings algorithm for this setup. Then, for a function $h(x)$ that satisfies

$$\int h(x) f(x) dx < \infty.$$

Then, we have

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} h(X_t) = \int h(x) f(x) dx.$$

Example

- **Example 7** Let X be a gamma random variable with parameters $\alpha > 0$ and $\beta > 0$. Its density function $f(x)$ is given by:

$$f(x \mid \alpha, \beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} \exp(-\frac{x}{\beta}).$$

where, $\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1}\, dt.$

We want to use M-H algorithm to generate samples from $f(x)$.

Note that:

1. If α is a positive integer n, then $\Gamma(n) = (n-1)!$

2. If $\alpha = 1$, the gamma distribution becomes an exponential distribution with mean β.

Example

3. The Chi-square distribution

$$f(x \mid k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$

is a special case of the gamma distribution when $\alpha = k/2$, $\beta = 2$.

4. if α is a positive integer, then X is the sum of α independent random variables, each being exponentially distributed with mean β.

Let n be a positive integer closest to α. Then, a random variable Y with density $f(y|n, \beta)$ can be simulated using n independent exponential random variables.

Example

We will use Y as a proposal in an independent M-H algorithm.

- **Algorithm 35**

1. Generate $Y \sim f(y|n, \beta)$.
2. Update the state to X_{t+1} according to:

$$X_{t+1} = \begin{cases} Y & \text{with probability } \rho(X_t, Y) \\ X_t & \text{with probability } 1 - \rho(X_t, Y) \end{cases}$$

where

$$\rho(x, y) = \min\{\frac{f(y)q(x)}{f(x)q(y)}, 1\} = \min\{(\frac{y}{x})^{\alpha - n}, 1\}.$$