7.2 Introduction to Metropolis-Hastings Algorithm

Suppose that we wish to generate a discrete random variables with p.m.f.

$$
\pi(j) = \frac{h(j)}{c}, \quad j = 1, 2, ..., m,
$$

where $h(j)$, $j = 1, 2, ..., m$, are positive integers, $m$ is very large and the normalizing constant $c = \sum_{j=1}^{m} h(j)$, is hard to compute (or $f(x) = \frac{h(x)}{\int \frac{h(t)}{h(t)} dt}, x \in \mathcal{X}$, for the continuous case).

In order to simulate a sequence of random variables whose distributions converge to $\{\pi(j)\}$, we can find a Markov chain that is easy to produce, with limiting probabilities $\{\pi(j)\}$. We present now a method that accomplishes just that, by constructing a time-reversible Markov chain with desired equilibrium distribution. This method is known in the literature as the Metropolis-Hastings Algorithm (Metropolis et al., Journal of Chemical Physics, 1953, and Hastings, Biometrika, 1970).

Suppose that we can produce easily an irreducible Markov chain with transition matrix $Q = [(q(i, j))]$, for $i, j = 1, 2, ..., m$. We define now a Markov chain $\{X_n : n \geq 0\}$ where if $X_n = i$, we generate a random variable $X$ with p.m.f. $q(i, j)$, $j = 1, 2, ..., m$ (the $i$-th row from the transition matrix). Then if $X = j$, we choose the next state as

$$
X_{n+1} = \begin{cases} 
    j, \text{ with probability } a(i, j) \\
    i, \text{ with probability } 1 - a(i, j)
\end{cases},
$$

where $a(i, j)$, will be determined.

Now the transition probabilities for the Markov chain $\{X_n\}$ become

$$
p_{ij} = q(i, j)a(i, j), \quad \text{if } j \neq i, \quad \text{and}
$$

$$
p_{ii} = q(i, i) + \sum_{k \neq i} q(i, k)(1 - a(i, k)),
$$

and the chain will be time reversible with stationary probabilities $\{\pi(j)\}$ if

$$
\pi(i)p(i, j) = \pi(j)p(j, i), \quad j \neq i,
$$

or

$$
\pi(i)q(i, j)a(i, j) = \pi(j)q(j, i)a(j, i), \quad j \neq i.
$$

The method selects the $a(i, j)$ as

$$
a(i, j) = \min \left\{ \frac{\pi(j)q(j, i)}{\pi(i)q(i, j)}, 1 \right\} = \min \left\{ \frac{h(j)q(j, i)}{h(i)q(i, j)}, 1 \right\},
$$

in order to produce transition probabilities $p_{ij}$ of a time-reversible Markov chain as desired. Notice that calculation of the normalizing constant is no longer an issue.

Hence Metropolis-Hastings can generate realizations from a p.m.f. $\pi(j) = \frac{h(j)}{c}$, $j = 1, 2, ..., m$, as follows:

**Step 1:** Select an irreducible Markov chain with transition probabilities $q(i, j)$, $i, j = 1, 2, ..., m$, and start the target Markov chain with some $X_0 = k$.

**Step 2:** Generate the random variable $X$ with p.m.f. $q(X_n, j)$, $j = 1, 2, ..., m$ ($X_n$ the previous state, and pick the $X_n$ th row of the transition matrix $Q$ as the p.m.f.) and generate $U \sim U(0, 1)$. 

1
Step 3: If
\[ U < \frac{h(X)q(X, X_n)}{h(X_n)q(X, X_n)} \]
then set \( X_{n+1} = X \), otherwise remain at the same state, i.e., \( X_{n+1} = X_n \). Go to the previous step and continue to generate the next state.

7.2.1 Metropolis-Hastings Algorithm for generating continuous random vectors

Let us present the algorithm for the continuous case. The target distribution is
\[ f(x) = \frac{h(x)}{\int h(t)dt}, x \in \mathcal{X}, \]
but only \( h(x) \) is known. The source \( q(x, y) \) must be a density with respect to \( x \). Notice that Hastings algorithm is a special case when symmetry holds for \( q \), i.e., \( q(x, y) = q(y, x) \). Here are the steps:

**Step 1:** Let \( X_n = (x_1, ..., x_p) \) be the current state of the chain and generate the proposed vector \( Y \) from the source density \( q(x, \cdot) \).

**Step 2:** Generate \( U \sim U(0, 1) \).

**Step 3:** If
\[ U < \frac{h(Y)q(Y, X_n)}{h(X_n)q(Y, X_n)} \]
then set \( X_{n+1} = Y \) otherwise \( X_{n+1} = X_n \).