Observations on the Metropolis-Hastings Algorithm

Myron Hlynka & Michelle Cylwa
Department of Mathematics & Statistics
University of Windsor
Windsor, Ontario, Canada, N9B 3P4

Abstract: We present some properties of the Metropolis-Hastings algorithm for constructing a Markov chain with a given limiting probability distribution. In particular, we consider what happens if we apply the Metropolis-Hastings algorithm repeatedly to a "proposal" distribution which has already been updated.

1 Introduction

In MCMC (Markov Chain Monte Carlo) studies, there is extensive use of the Metropolis-Hastings algorithm. See Ross (2007) or Evans and Rosenthal (2004) or Ibe (2009). The algorithm is usually applied to infinite state Markov chains but also works for finite state chains. Most of the results in this paper are derived for the finite state case but also work in the infinite state case.

Suppose the states of a finite state Markov chain are labeled $\{1, 2, \ldots, n\}$. Assume that the limiting vector $\underline{\pi} = (\pi_1, \ldots, \pi_n)$ is known. (We assume throughout that all $\pi_i \neq 0$.) The Metropolis- Hastings algorithm finds a transition matrix with the given limiting vector. The algorithm has two parts. First, we select a "proposal" distribution for moving between states. Second, there is an "acceptance" distribution that can be used with the proposal distribution.

The proposal distribution from a state i consists of values of the ith row of a probability transition matrix $Q = [q_{ij}]$. We refer to Q as the proposal transition matrix. The Metropolis-Hastings algorithm then defines a set of acceptance values α_{ij} and combines the q_{ij} and α_{ij} values to get the final transition probabilities p_{ij}.

Begin with the given $\underline{\pi}$. The user chooses any set of q_{ij} values such that Q is a transition matrix of an irreducible Markov chain.

Next, the acceptance distribution α_{ij} is defined as follows.
If $i \neq j$ and $q_{ij} \neq 0$, define

$$\alpha_{ij} = \min\left\{1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right\}. \tag{1}$$

If $i = j$ or $i \neq j$ and $q_{ij} = 0$, define $\alpha_{ij} = 0$.

Next, we define the probability transition matrix P.
For $i \neq j$, define $p_{ij} = q_{ij}\alpha_{ij}$.
For $i = j$, define p_{ii} so that the sum of each row is 1.
Then $P = [p_{ij}]$ is a transition matrix with limiting vector $\underline{\pi}$.

Since the proposal distribution q_{ij} can be almost anything, there are a huge number of possible Markov chains (and Markov transition matrices) that can result. The most important new results in this paper are Properties 3.1, 3.4, 3.6, 3.7.

2 Initial Example

We begin by applying the Metropolis-Hastings algorithm to the sequence $\{1, 1, 2, 3, 5\}$ to see what happens. The normalized vector is $\underline{\pi} = (\frac{1}{12}, \frac{1}{12}, \frac{2}{12}, \frac{3}{12}, \frac{5}{12})$. The states are labeled $\{1, 2, 3, 4, 5\}$. The "proposal" distribution uses a symmetric cyclic random walk so we expect to get a matrix of birth-death type (tridiagonal form with possible entries in the upper right and lower left corners). Choose $q_{12} = q_{15} = .5$, $q_{23} = q_{21} = .5$, $q_{34} = q_{32} = .5$, $q_{43} = q_{45} = .5$, $q_{54} = q_{51} = .5$. All other $q_{ij} = 0$.

The α values are determined as follows. For $i = 1$, we have $q_{15} = .5$ and $q_{12} = .5$. Thus

$$\alpha_{12} = \min\left\{1, \frac{\pi_2 q_{21}}{\pi_1 q_{12}}\right\} = \min\left\{1, \frac{(1/12)(.5)}{(1/12)(.5)}\right\} = 1.$$

$$\alpha_{15} = \min\left\{1, \frac{\pi_5 q_{51}}{\pi_1 q_{15}}\right\} = \min\left\{1, \frac{(5/12)(.5)}{(1/12)(.5)}\right\} = 1.$$

Thus $p_{12} = q_{12}\alpha_{12} = .5(1) = .5$ and $p_{15} = q_{15}\alpha_{15} = .5(1) = .5$

Next $p_{11} = 1 - p_{12} - p_{15} = 1 - .5 - .5 = 0$. Also $0 = p_{11} = p_{13} = p_{14}$. So row 1 of P is $[0, .5, 0, 0, .5]$.

For $i = 2$, we have $q_{21} = .5$ and $q_{23} = .5$. Thus

$$\alpha_{23} = \min\left\{1, \frac{\pi_3 q_{32}}{\pi_2 q_{23}}\right\} = \min\left\{1, \frac{(2/12)(.5)}{(1/12)(.5)}\right\} = 1.$$

$$\alpha_{21} = \min\left\{1, \frac{\pi_1 q_{21}}{\pi_2 q_{21}}\right\} = \min\left\{1, \frac{(1/12)(.5)}{(1/12)(.5)}\right\} = 1.$$

Thus $p_{23} = q_{23}\alpha_{23} = .5(1) = .5$ and $p_{21} = q_{21}\alpha_{21} = .5(1) = .5$

Next $p_{22} = 1 - p_{21} - p_{23} = 1 - .5 - .5 = 0$. So row 2 of P is $[.5, 0, .5, 0, 0]$.

For $i = 3$, we have $q_{32} = .5$ and $q_{34} = .5$. Thus

$$\alpha_{34} = \min\left\{1, \frac{\pi_4 q_{43}}{\pi_3 q_{34}}\right\} = \min\left\{1, \frac{(3/12)(.5)}{(2/12)(.5)}\right\} = 1.$$

$$\alpha_{32} = \min\left\{1, \frac{\pi_2 q_{32}}{\pi_3 q_{32}}\right\} = \min\left\{1, \frac{(1/12)(.5)}{(2/12)(.5)}\right\} = .5.$$

Thus $p_{34} = q_{34}\alpha_{34} = .5(1) = .5$ and $p_{32} = q_{32}\alpha_{32} = .5(.5) = .25$

Next $p_{33} = 1 - p_{32} - p_{34} = 1 - .25 - .5 = .25$. So row 3 of P is $[0, .25, .25, .5, 0]$.

For $i = 4$, we have $q_{43} = .5$ and $q_{45} = .5$. Thus

$$\alpha_{45} = \min\left\{1, \frac{\pi_5 q_{43}}{\pi_4 q_{45}}\right\} = \min\left\{1, \frac{(5/12)(.5)}{(3/12)(.5)}\right\} = 1.$$

$$\alpha_{43} = \min\left\{1, \frac{\pi_3 q_{43}}{\pi_4 q_{43}}\right\} = \min\left\{1, \frac{(2/12)(.5)}{(3/12)(.5)}\right\} = 2/3.$$

Thus $p_{45} = q_{45}\alpha_{45} = .5(1) = .5$ and $p_{43} = q_{43}\alpha_{43} = .5(2/3) = 1/3$

Next $p_{44} = 1 - p_{43} - p_{45} = 1 - 1/3 - .5 = 1/6$. So row 4 of P is $[0, 0, 1/3, 1/6, .5]$.

For $i = 5$, we have $q_{54} = .5$ and $q_{51} = .5$. Thus

$$\alpha_{54} = \min\left\{1, \frac{\pi_4 q_{45}}{\pi_5 q_{54}}\right\} = \min\left\{1, \frac{(3/12)(.5)}{(5/12)(.5)}\right\} = .6.$$

$$\alpha_{51} = \min\left\{1, \frac{\pi_1 q_{15}}{\pi_5 q_{51}}\right\} = \min\left\{1, \frac{(1/12)(.5)}{(5/12)(.5)}\right\} = 1/5.$$

Thus $p_{51} = q_{51}\alpha_{51} = .5(1/5) = .1$ and $p_{54} = q_{54}\alpha_{54} = (.5)(.6) = .3$

Next $p_{55} = 1 - p_{51} - p_{54} = 1 - .1 - .3 = .6$. So row 5 of P is $[.1, 0, 0, .3, .6]$.

Finally, the complete transition matrix is $P = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{2}{4} & 0 \\ 0 & 0 & \frac{2}{6} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{10} & 0 & 0 & \frac{3}{10} & \frac{6}{10} \end{bmatrix}$.

3 Observations

We are interested in knowing what happens if we apply the Metropolis-Hastings algorithm repeatedly.

Property 3.1. *Suppose the limiting probability vector is $\underline{\pi}$ and the initial proposal probability transition matrix is Q_0, resulting in the probability transition matrix P_0. If we repeat the algorithm the same limiting probability vector $\underline{\pi}$ using $Q_1 = P_0$ as the proposal probability transition matrix, then the resulting probability transition matrix is P_1 where $P_1 = P_0$.*

Proof. We begin with proposal probability transition matrix $Q_0 = [q_{ij}^{(0)}]$ and limiting probability vector $\underline{\pi} = (\pi_1, \ldots, \pi_n)$ and calculate the probability transition matrix P_0

by applying the algorithm. Then for $q_{ij}^{(0)} \neq 0$ and $i \neq j$ we have $p_{ij}^{(0)} = \alpha_{ij} q_{ij}^{(0)}$ where $\alpha_{ij} = \min\left\{1, \dfrac{\pi_j q_{ji}^{(0)}}{\pi_i q_{ij}^{(0)}}\right\}$. Likewise for $q_{ji}^{(0)} \neq 0$ and $i \neq j$ we have $p_{ji}^{(0)} = \alpha_{ji} q_{ji}^{(0)}$ where $\alpha_{ji} = \min\left\{1, \dfrac{\pi_i q_{ij}^{(0)}}{\pi_j q_{ji}^{(0)}}\right\}$.

If we repeat the procedure using $Q_1 = P_0$ as the proposal probability transition matrix and the same limiting probability vector $\underline{\pi}$, then we compute the new probability transition matrix $P_1 = [p_{ij}^{(1)}]$. For $p_{ij}^{(0)} \neq 0$ and $i \neq j$ we have $p_{ij}^{(1)} = \beta_{ij} p_{ij}^{(0)}$ where

$$\begin{aligned}
\beta_{i,j} = \min\left\{1, \frac{\pi_j p_{ji}^{(0)}}{\pi_i p_{ij}^{(0)}}\right\} &= \min\left\{1, \frac{\pi_j}{\pi_i}\frac{\alpha_{ji}}{\alpha_{ij}}\frac{q_{ji}^{(0)}}{q_{ij}^{(0)}}\right\} \\
&= \min\left\{1, \frac{\pi_j}{\pi_i}\frac{q_{ji}^{(0)}}{q_{ij}^{(0)}} \frac{\min\left\{1, \dfrac{\pi_i q_{ij}^{(0)}}{\pi_j q_{ji}^{(0)}}\right\}}{\min\left\{1, \dfrac{\pi_j q_{ji}^{(0)}}{\pi_i q_{ij}^{(0)}}\right\}}\right\}.
\end{aligned}$$

Case 1: $\pi_j q_{ji}^{(0)} = \pi_i q_{ij}^{(0)}$

$\Rightarrow \beta_{ij} = \min\left\{1, 1\dfrac{\min\{1,1\}}{\min\{1,1\}}\right\} = \min\{1,1\} = 1.$

Case 2: $\pi_j q_{ji}^{(0)} > \pi_i q_{ij}^{(0)}$

$$\Rightarrow \beta_{ij} = \min\left\{1, \frac{\pi_j}{\pi_i}\frac{q_{ji}^{(0)}}{q_{ij}^{(0)}} \frac{\dfrac{\pi_i q_{ij}^{(0)}}{\pi_j q_{ji}^{(0)}}}{1}\right\} = \min\{1,1\} = 1.$$

Case 3: $\pi_j q_{ji}^{(0)} < \pi_i q_{ij}^{(0)}$

$$\Rightarrow \beta_{ij} = \min\left\{1, \frac{\pi_j}{\pi_i}\frac{q_{ji}^{(0)}}{q_{ij}^{(0)}} \frac{1}{\dfrac{\pi_j q_{ji}^{(0)}}{\pi_i q_{ij}^{(0)}}}\right\} = \min\{1,1\} = 1.$$

Thus for $i \neq j$ and $p_{ij} \neq 0$ we have $p_{ij}^{(1)} = p_{ij}^{(0)}$. Also, for $i \neq j$, if $p_{ij}^{(0)} = 0$ then $\beta_{ij} = 0$ so $p_{ij}^{(1)} = p_{ij}^{(0)} = 0$. Thus $\forall i, p_{ii}^{(1)} = p_{ii}^{(0)}$ since the rows must sum to 1. Therefore, $P_1 = P_0$.

$\square$

Thus a repetition of the Metropolis-Hastings algorithm does not change the resulting Markov chain. Normally the limiting probability vector of the proposal transition matrix Q will not match the initial limiting probability vector. Even if we choose a proposal matrix Q which happens to have a limiting probability vector that matches the initial limiting probability vector, it is still not necessarily true that $P = Q$.

Information about reversible Markov chains can be found in Ross (2007).

Definition 3.1. *A stationary ergodic Markov chain with transition matrix P and limiting vector $\underline{\pi}$ is reversible iff $\pi_i p_{ij} = \pi_j p_{ji}$ $\forall i, j$.*
We refer to the transition matrix P as reversible if the corresponding Markov chain is reversible.

Property 3.2. *The matrix P obtained by applying the Metropolis-Hastings algorithm is reversible.*

Proof. See Ross (2007). □

Thus only reversible matrices P can be the result of applying the Metropolis-Hastings algorithm. Most transition matrices P are thus excluded as possible output transition matrices.

Property 3.3. *We are given the limiting probability vector $\underline{\pi}$. The proposal probability transition matrix Q may have the same limiting probability vector $\underline{\pi}$, but the Metropolis-Hastings algorithm need not return Q as the calculated probability transition matrix.*

Proof. Use Property 3.2. So if we begin with a transition matrix Q that is non-reversible transition matrix of an irreducible Markov chain, then the Metropolis-Hastings algorithm must generate a new transition matrix P which is different from Q, regardless of whether or not the limiting matrix Q gives a limiting probability vector to match the initial $\underline{\pi}$. □

Example 3.1. *The probability transition matrix* $Q = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$ *has limiting* *probability vector* $\underline{\pi} = \left(\frac{34}{55}, \frac{13}{55}, \frac{5}{55}, \frac{2}{55}, \frac{1}{55}\right)$. *When* Q *and* $\underline{\pi}$ *are used in the Metropolis-Hastings algorithm, the resultant probability transition matrix is* $P = \begin{bmatrix} \frac{38}{51} & \frac{13}{51} & 0 & 0 & 0 \\ \frac{2}{3} & \frac{8}{39} & \frac{5}{39} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{8}{15} & \frac{2}{15} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$ *which also has limiting probability vector* $\underline{\pi}$.

Property 3.4. *Suppose we begin with limiting probability vector* $\underline{\pi}$ *and proposal probability transition matrix* Q. *Then the calculated probability matrix* P *will be equal to the proposal matrix* Q *iff* Q *defines a reversible Markov chain and has limiting probability vector* $\underline{\pi}$.

Proof. First recall that $p_{ij} = \alpha_{ij} q_{ij}$ $\forall i \neq j$. Thus $P = Q$ iff $p_{ii} = q_{ii}$ $\forall i$ and either $\alpha_{ij} = 1$ or $q_{ij} = 0 = p_{ij}$ $\forall i \neq j$.
If Q is reversible with limiting vector $\underline{\pi}$, then $\pi_j q_{ji} = \pi_i q_{ij}$ $\forall i, j$. If $i \neq j$ and $q_{ij} \neq 0$, then $\alpha_{ij} = \min\left\{1, \dfrac{\pi_j q_{ji}}{\pi_i q_{ij}}\right\} = 1$ so $p_{ij} = q_{ij}$. If $i \neq j$ and $q_{ij} = 0$ then $p_{ij} = \alpha_{ij} q_{ij} = 0$ so $p_{ij} = q_{ij}$. Since we have $p_{ij} = q_{ij}$ $\forall i \neq j$, we must also have $p_{ii} = q_{ii}$ $\forall i$ and hence $P = Q$.
Next suppose $P = Q$. From Ross (2007), we know that P is reversible with limiting probability vector to match the original limiting vector. Since $P = Q$, we have Q is reversible as well. $\square$

Suppose we begin with transition matrix Q that we can recognize as reversible, and then compute the limiting probability vector $\underline{\pi}$ (which is easy to compute for reversible Markov chains). If we apply Metropolis-Hastings to Q and $\underline{\pi}$ then we know that $P = Q$.

Example 3.2. *Consider the proposal probability transition matrix* $Q = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$.
We recognize this as a reversible matrix. See Jiang (2009). We compute $\underline{\pi} = (\frac{2}{8}, \frac{3}{8}, \frac{3}{8})$. *Applying the Metropolis-Hastings algorithm to this* Q *and* $\underline{\pi}$ *results in* Q *and satisfies*

$\underline{\pi}Q = \underline{\pi}$.

Property 3.5. *The non-zero nondiagonal entries of P resulting from the application of the Metropolis-Hastings algorithm may only occur in the same positions as the non-zero nondiagonal entries of Q.*

Proof. Non-zero α_{ij} may only occur when $q_{ij} \neq 0$ for $i \neq j$. Thus non-zero p_{ij} may occur in the same positions as the non-zero q_{ij}. Of course for $i \neq j$ if $q_{ij} = 0$ then $p_{ij} = 0$. □

Property 3.6. *If the entry $q_{ij} = 0$ for $i \neq j$ occurs in Q, then the Metropolis-Hastings algorithm results in P with $p_{ij} = 0$ and $p_{ji} = 0$.*

Proof. Since $p_{ij} = \alpha_{ij}q_{ij}$, we have $p_{ij} = 0$. Since P is reversible, we have $\pi_i p_{ij} = \pi_j p_{ji}$ so $p_{ji} = 0$ (since $\pi_i \neq 0$and $\pi_j \neq 0$). □

CONDITION A

Consider an $n \times n$ proposal probability transition matrix Q with the following properties:

1. The row sum is 1.
2. $q_{ij} = q_{ji}, \forall i, j, i \neq j$.
3. Any non-zero entry in the matrix is equal to the constant c.
4. $q_{ii} = 0$ or $q_{ii} = c, \forall i$.

Property 3.7. *Suppose we apply the Metropolis-Hastings algorithm to matrix Q (satisfying Condition A) and limiting probability vector $\underline{\pi} = (\pi_1, \ldots, \pi_n), \pi_i > 0, \forall i$. Then*
(a) the row(s) of P corresponding to the minimum entry of $\underline{\pi}$ are unchanged from the corresponding row(s) of Q.
(b) the column(s) of P corresponding to the maximum entry of $\underline{pi}$ are the same (except possibly for the diagonal entry) as the corresponding row(s) of Q.

Proof. Suppose π_i is the minimum entry of $\underline{\pi}$. Since $\pi_j \geq \pi_i$, $\alpha_{ij} = \min\left\{1, \dfrac{\pi_j q_{ji}}{\pi_i q_{ij}}\right\} = \min\left\{1, \dfrac{\pi_j}{\pi_i}\right\} = 1$. Hence $p_{ij} = \alpha_{ij} q_{ij} = q_{ij}, \forall j \neq i$. The diagonal term ensures the row sum is equal to 1, therefore the i^{th} row of Q and P are equal. □

Although CONDITION A seems restrictive, we note that a symmetric random walk will satisfy such a condition and that is a reasonable initial matrix Q.

From our previous result, we observe that given a limiting probability vector $\underline{\pi}$ and a proposal probability transition matrix Q, it is simple to find P. We illustrate with the following example.

Example 3.3. *Consider, the limiting probability vector* $\underline{\pi} = (\frac{1}{20}, \frac{1}{20}, \frac{2}{20}, \frac{3}{20}, \frac{5}{20}, \frac{8}{20})$, *and the proposal probability transition matrix* $Q = \begin{bmatrix} 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}$. *From Property 3.7, we know that the first two rows of P are the same as the first two rows of Q since they correspond to the minimum entry of $\underline{\pi}$. Also the last column of P matches the last column of Q except perhaps the diagonal entry. From Property 3.6, we know where zero entries will occur and where non-zero entries may occur.*

Using only these two observations, so far we have $P = \begin{bmatrix} 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \star & 0 & * & 0 & \star & 0 \\ 0 & \star & 0 & * & 0 & \frac{1}{2} \\ \star & 0 & \star & 0 & * & 0 \\ 0 & \star & 0 & \star & 0 & * \end{bmatrix}$

For each of the $\star$ terms we have $p_{ij} = \alpha_{ij} q_{ij}$. For the terms below the diagonal $\alpha_{ij} = \frac{\pi_j}{\pi_i}$ since the entries of $\underline{\pi}$ increase as the index increases. For the terms above the diagonal $\alpha_{ij} = 1$ for the same reason.

We can easily fill in the $\star$ terms to find $P = \begin{bmatrix} 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{2}{4} & 0 \\ 0 & \frac{1}{6} & 0 & \frac{2}{6} & 0 & \frac{3}{6} \\ \frac{1}{10} & 0 & \frac{2}{10} & 0 & \frac{7}{10} & 0 \\ 0 & \frac{1}{16} & 0 & \frac{3}{16} & 0 & \frac{12}{16} \end{bmatrix}$.

The diagonal terms are found by using the fact that the rows must sum to 1.

4 Conclusions

Since the Metropolis-Hastings algorithm is widely used so any understanding of the its operation is beneficial. A natural question regarding this algorithm is what happens when it is applied to the output distribution rather than the proposal distribution. This paper indicates that there is no gain.

Further questions arise as to whether computations can be simplified if the initial limiting distribution is strictly decreasing, for example, or perhaps unimodal. Such questions could be the subject of future work.

References

[1] Evans, M. and Rosenthal, J. (2004) *Probability and Statistics: The Science of Uncertainty.* W.H. Freeman & Co.

[2] Ibe, O.C. (2009). *Markov Processes for Stochastic Modeling.* Academic Press.

[3] Jiang, Q. (2009). *Construction of Transition Matrices of Reversible Markov Chains.* M.Sc. Major Paper. Department of Mathematics and Statistics. University of Windsor.

[4] Ross, S. (2007). *Introduction to Probability Models (9th ed.).* Academic Press.