Chapter 2
Markov Chain Monte Carlo Methods over Discrete Sample Space

2.1 Constructing a Connected Markov Chain over a Conditional Sample Space: Markov Basis

In the previous chapter we discussed exact tests for some simple models of contingency tables. As we discussed at the end of Sect. 1.3, the Markov chain Monte Carlo method is general and useful when the cardinality of conditional sample space (fiber) is large. We first consider connectivity of a Markov chain, without fully specifying the transition probabilities.

Consider the independence model of general two-way contingency tables in Sect. 1.3. The fiber is the set of $I \times J$ contingency tables with fixed row sums and column sums:

$$\mathscr{F}_{\boldsymbol{t}} = \{\boldsymbol{x} \geq 0 \mid x_{i+}, i \in [I],\ x_{+j}, j \in [J] \text{ are fixed according to } \boldsymbol{t}\}. \tag{2.1}$$

Let A be the configuration in (1.18). The kernel of A is denoted by $\ker A$. The set of integer vectors in $\ker A$ is called the *integer kernel* of A and is denoted by

$$\ker_{\mathbb{Z}} A = \{\boldsymbol{z} \mid A\boldsymbol{z} = 0, \boldsymbol{z} \in \mathbb{Z}^{\eta}\}, \quad \eta = IJ.$$

An element of $\ker_{\mathbb{Z}} A$ is called a *move* for the configuration A. If $\boldsymbol{x}$ and $\boldsymbol{y}$ belong to the same fiber $\mathscr{F}_{\boldsymbol{t}}$, then $\boldsymbol{y} - \boldsymbol{x}$ is a move, because

$$A(\boldsymbol{y} - \boldsymbol{x}) = A\boldsymbol{y} - A\boldsymbol{x} = \boldsymbol{t} - \boldsymbol{t} = 0. \tag{2.2}$$

Now consider the following integer matrix $\boldsymbol{z} = \boldsymbol{z}(i_1, i_2; j_1, j_2) = \{z_{ij}\}$,

$$z_{ij} = \begin{cases} +1, & (i,j) = (i_1, j_1), (i_2, j_2), \\ -1, & (i,j) = (i_1, j_2), (i_2, j_1), \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

S. Aoki et al., *Markov Bases in Algebraic Statistics*, Springer Series in Statistics 199, DOI 10.1007/978-1-4614-3719-2_2,

The nonzero elements of $\boldsymbol{z}(i_1,i_2;j_1,j_2)$ are depicted as

$$\begin{array}{c|cc|} & j_1 & j_2 \\ \cline{2-3} i_1 & +1 & -1 \\ i_2 & -1 & +1 \\ \cline{2-3} \end{array}\,. \tag{2.4}$$

Adding $\boldsymbol{z}(i_1,i_2;j_1,j_2)$ to a contingency table $\boldsymbol{x}$ does not alter the row sums and the column sums. Hence $\boldsymbol{z}(i_1,i_2;j_1,j_2)$ is a move for A in (1.18); that is, $\boldsymbol{z}(i_1,i_2;j_1,j_2) \in \ker_{\mathbb{Z}} A$. We call a move of the form (2.4) a *basic move* for the independence model of two-way contingency tables. Because of the elements -1 in $\boldsymbol{z}(i_1,i_2;j_1,j_2)$, $\boldsymbol{x}+\boldsymbol{z}$ contains a negative element if $x_{i_2 j_1} = 0$ or $x_{i_1 j_2} = 0$. If both of these elements are positive, then $\boldsymbol{x}+\boldsymbol{z}$ is in $\mathscr{F}_{\boldsymbol{t}}$ if $\boldsymbol{x} \in \mathscr{F}_{\boldsymbol{t}}$. We have "moved" from $\boldsymbol{x}$ to $\boldsymbol{x}+\boldsymbol{z}$ in $\mathscr{F}_{\boldsymbol{t}}$. This is why we call $\boldsymbol{z}(i_1,i_2;j_1,j_2)$ a move. The following is an example of adding a move for the case of $I = J = 3$, $i_1 = j_1 = 1$, $i_2 = j_2 = 2$.

$$\begin{array}{|c|c|c|c} \hline 2 & 1 & 1 & 4 \\ \hline 2 & 0 & 2 & 4 \\ \hline 1 & 2 & 0 & 3 \\ \hline 5 & 3 & 3 & \end{array} + \begin{array}{|c|c|c|} \hline 1 & -1 & 0 \\ \hline -1 & 1 & 0 \\ \hline 0 & 0 & 0 \\ \hline \end{array} = \begin{array}{|c|c|c|c} \hline 3 & 0 & 1 & 4 \\ \hline 1 & 1 & 2 & 4 \\ \hline 1 & 2 & 0 & 3 \\ \hline 5 & 3 & 3 & \end{array}\,.$$

Suppose that we always use the last row I and the last column J in the move and let $i_2 = I$ and $j_2 = J$. Then

$$\{\boldsymbol{z}(i_1,I;j_1,J) \mid 1 \le i_1 \le I-1, 1 \le j_1 \le J-1\}$$

forms a basis of $\ker_{\mathbb{Z}} A$. More precisely the set forms a *lattice basis* of $\ker_{\mathbb{Z}} A$ in the sense that every $\boldsymbol{z} \in \ker_{\mathbb{Z}} A$ is uniquely written as an integer combination of $\boldsymbol{z}(i_1,I;j_1,J)$s. In fact the elements of the last row and the last column of $\boldsymbol{z} = \{z_{ij}\} \in \ker_{\mathbb{Z}} A$ are uniquely determined from the other elements. Hence $\boldsymbol{z} \in \ker_{\mathbb{Z}} A$ can be uniquely written as

$$\boldsymbol{z} = \sum_{i_1=1}^{I-1} \sum_{j_1=1}^{J-1} z_{i_1 j_1} \times \boldsymbol{z}(i_1,I;j_1,J), \tag{2.5}$$

because both sides have the same elements in the first $I-1$ rows and the first $J-1$ columns. This is related to use of the last level as the base level discussed at the end of Chap. 1.

Note that the lattice basis is very simple for the independence model of $I \times J$ tables. However, for the fiber in (2.1) we are requiring nonnegativeness of the frequency vectors. As an example consider the following two elements of the fiber for $I = J = 3$ with $1 = x_{1+} = x_{2+} = x_{+1} = x_{+2}$, $0 = x_{3+} = x_{+3}$.

$$\begin{array}{|c|c|c|c} \hline 1 & 0 & 0 & 1 \\ \hline 0 & 1 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 \\ \hline 1 & 1 & 0 & 2 \end{array}\,, \qquad \begin{array}{|c|c|c|c} \hline 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 \\ \hline 1 & 1 & 0 & 2 \end{array}\,.$$

We see that we cannot add or subtract any of $\boldsymbol{z}(i_1,3;j_1,3)$ to/from these tables without making some cell frequency negative. However, obviously these two tables are connected by the following move:

$$\begin{array}{|c|c|c|}\hline 1 & -1 & 0 \\ \hline -1 & 1 & 0 \\ \hline 0 & 0 & 0 \\ \hline \end{array}\,.$$

This example suggests that we can move around a fiber if we can use all moves of the form (2.3).

Let $\mathscr{B} \subset \ker_{\mathbb{Z}} A$ be a finite set of moves for a configuration A. $\mathscr{B}$ is called a *Markov basis* if for all fibers $\mathscr{F}_{\boldsymbol{t}}$ and for all elements $\boldsymbol{x},\boldsymbol{y} \in \mathscr{F}_{\boldsymbol{t}}$, $\boldsymbol{x} \neq \boldsymbol{y}$, there exist $K > 0$, $\boldsymbol{z}_1,\dots,\boldsymbol{z}_K \in \mathscr{B}$ and $\varepsilon_1,\dots,\varepsilon_K \in \{-1,1\}$, such that

$$\boldsymbol{y} = \boldsymbol{x} + \sum_{k=1}^{K} \varepsilon_k \boldsymbol{z}_k, \qquad \boldsymbol{x} + \sum_{k=1}^{L} \varepsilon_k \boldsymbol{z}_k \in \mathscr{F}_{\boldsymbol{t}}, \qquad L = 1,\dots,K-1. \tag{2.6}$$

The first condition says that by adding or subtracting elements of $\mathscr{B}$, we can move from $\boldsymbol{x}$ to $\boldsymbol{y}$. The second condition says that on the way from $\boldsymbol{x}$ to $\boldsymbol{y}$ we never encounter a negative frequency. Therefore if a Markov basis $\mathscr{B}$ is given, then we can move all over any fiber by adding or subtracting moves from $\mathscr{B}$. Thus connectivity of every fiber is guaranteed by a Markov basis. We define Markov basis again in Chap. 4 for a general configuration A. In this introductory explanation, we give a proof that a Markov basis for the $I \times J$ independence model of two-way contingency tables is given by the set of moves $\boldsymbol{z}(i_1,i_2;j_1,j_2)$. We state this as a theorem.

Theorem 2.1. *Let*

$$\mathscr{B} = \{\boldsymbol{z}(i_1,i_2;j_1,j_2) \mid 1 \le i_1 < i_2 \le I, 1 \le j_1 < j_2 \le J\}.$$

$\mathscr{B}$ forms a Markov basis for the $I \times J$ independence model of two-way contingency tables.

The following proof is a typical "distance reducing argument," that is frequently used in later chapters of this book.

Proof. We argue by contradiction. Suppose that $\mathscr{B}$ is not a Markov basis. Then there exists a fiber $\mathscr{F}_{\boldsymbol{t}}$ and two elements $\boldsymbol{x},\boldsymbol{y} \in \mathscr{F}_{\boldsymbol{t}}$ of the fiber, such that we cannot move from $\boldsymbol{x}$ to $\boldsymbol{y}$ by the moves of $\mathscr{B}$ as in (2.6). Let

$$\mathscr{N}_{\boldsymbol{x}} = \{\boldsymbol{y} \in \mathscr{F}_{\boldsymbol{t}} \mid \text{we cannot move from } \boldsymbol{x} \text{ to } \boldsymbol{y} \text{ by moves of } \mathscr{B}\}.$$

Then $\mathscr{N}_{\boldsymbol{x}}$ is not empty by assumption. For $\boldsymbol{z} = \{z_{ij}\} \in \ker_{\mathbb{Z}} A$, let $|\boldsymbol{z}| = \sum_{i=1}^{I} \sum_{j=1}^{J} |z_{ij}|$ denote its 1-norm. In Sect. 4.3 we define $\deg \boldsymbol{z}$ as $|\boldsymbol{z}|/2$.

Define

$$\boldsymbol{y}^* = \arg\min_{\boldsymbol{y}\in\mathcal{N}_{\boldsymbol{x}}} |\boldsymbol{x}-\boldsymbol{y}|. \tag{2.7}$$

$\boldsymbol{y}^*$ is one of the closest elements of $\mathcal{F}_{\boldsymbol{t}}$ that cannot be reached from $\boldsymbol{x}$ by $\mathcal{B}$:

$$|\boldsymbol{x}-\boldsymbol{y}^*| = \min_{\boldsymbol{y}\in\mathcal{N}_{\boldsymbol{x}}} |\boldsymbol{x}-\boldsymbol{y}|.$$

Now let $\boldsymbol{w} = \boldsymbol{x}-\boldsymbol{y}^*$ and consider the signs of elements of $\boldsymbol{w}$. Because $\boldsymbol{w}$ contains a positive element, let $w_{i_1 j_1} > 0$. Then because $\boldsymbol{w}$ is a move, there exist $j_2 \neq j_1$ with $w_{i_1 j_2} < 0$ and $i_2 \neq i_1$ with $w_{i_2 j_1} < 0$. Hence for $\boldsymbol{y}^* = \{y^*_{ij}\}$ we have $y^*_{i_1 j_2} > 0$, $y^*_{i_2 j_1} > 0$. Then

$$\boldsymbol{y}^* + \boldsymbol{z}(i_1,i_2;j_1,j_2) \in \mathcal{F}_{\boldsymbol{t}}.$$

$\boldsymbol{y}^*$ cannot be reached from $\boldsymbol{x}$ by $\mathcal{B}$, therefore $\boldsymbol{y}^* + \boldsymbol{z}(i_1,i_2;j_1,j_2)$ cannot be reached from $\boldsymbol{x}$ by $\mathcal{B}$ either and $\boldsymbol{y}^* + \boldsymbol{z}(i_1,i_2;j_1,j_2) \in \mathcal{N}_{\boldsymbol{x}}$. Now we check the value of $|\boldsymbol{x} - (\boldsymbol{y}^* + \boldsymbol{z}(i_1,i_2;j_1,j_2))|$.

- If $w_{i_2 j_2} > 0$, then $|\boldsymbol{x} - (\boldsymbol{y}^* + \boldsymbol{z}(i_1,i_2;j_1,j_2))| = |\boldsymbol{x}-\boldsymbol{y}^*| - 4$,
- If $w_{i_2 j_2} \le 0$, then $|\boldsymbol{x} - (\boldsymbol{y}^* + \boldsymbol{z}(i_1,i_2;j_1,j_2))| = |\boldsymbol{x}-\boldsymbol{y}^*| - 2$.

Therefore for both cases, $|\boldsymbol{x} - (\boldsymbol{y}^* + \boldsymbol{z}(i_1,i_2;j_1,j_2))| < |\boldsymbol{x}-\boldsymbol{y}^*|$. However, this contradicts the minimality in (2.7) of $\boldsymbol{y}^*$. □

By this theorem, we can construct a connected Markov chain over any fiber. We choose $i_1, i_2 \in [I]$ and $j_1, j_2 \in [J]$ randomly. We add or subtract $\boldsymbol{z}(i_1,i_2;j_1,j_2)$ to/from the current state $\boldsymbol{x}$ and move to $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{z}(i_1,i_2;j_1,j_2)$ as long as there is no negative frequency in $\boldsymbol{y}$. In the case where $\boldsymbol{y}$ contains a negative element, we choose another set of indices $i_1, i_2 \in [I]$ and $j_1, j_2 \in [J]$ and continue. Then connectivity of every fiber is guaranteed by Theorem 2.1.

Note that in the above explanation we are not precisely specifying the probability distribution of choosing an element $\boldsymbol{z}(i_1,i_2;j_1,j_2)$. Also, when we say "add or subtract," we are not exactly saying which to choose. In fact, we should choose the sign of a move $\boldsymbol{z}(i_1,i_2;j_1,j_2)$ (i.e., whether we add it or subtract it) with probability $1/2$. This is related to the Markov chain symmetry for the Metropolis–Hastings algorithm in the next section. Other than the choice of the sign of a move, the distribution for choosing a move can be arbitrary.

In this section we considered the independence model of two-way contingency tables. We now briefly mention the conditional independence model of three-way contingency tables. As we saw in the previous section, the conditional independence model of three-way contingency tables can be treated as the two-way independence model given each level of the conditioning variable. Therefore a Markov basis for the conditional independence model of three-way contingency tables is given as a union of Markov bases for two-way cases in each slice of the contingency table given the level of the conditioning variable. The two-way independence model and the conditional independence model of three-way contingency tables are actually simple examples. Markov bases for more complicated models of contingency tables are in fact difficult and each model needs separate consideration. One notable exception is the decomposable model studied in Chap. 8.

On the other hand, there exists a general algorithm to compute a Markov basis in the form of the Gröbner basis for any configuration. So is the problem of obtaining a Markov basis already solved by a general algorithm? The answer is yes and no, depending on the viewpoint. The existence of a general algorithm means that the answer is yes from a certain theoretical viewpoint. On the other hand, for practical purposes, the computation of the Gröbner basis for a complicated model is often infeasible in a practical amount of time and in this sense the answer is no. Therefore, both theoretical investigations of Markov bases for specific models and the further general improvements in the algorithms for Gröbner basis computation are very much needed at present.

2.2 Adjusting Transition Probabilities by Metropolis–Hastings Algorithm

In this section we explain how to construct a Markov chain that has a specified distribution as the stationary distribution. A good reference on important facts on Markov chains is Häggström [69].

Consider a Markov chain over a finite sample space $\mathscr{F}$. Suppose that the elements of $\mathscr{F}$ are given as

$$\mathscr{F} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s\}. \tag{2.8}$$

Let $\{Z_t,\ t = 0, 1, 2, \ldots\}$, $Z_t \in \mathscr{F}$, be a Markov chain over $\mathscr{F}$ with the transition probability $Q = (q_{ij})$:

$$q_{ij} = P(Z_{t+1} = \boldsymbol{x}_j \mid Z_t = \boldsymbol{x}_i), \qquad 1 \le i, j \le s.$$

A Markov chain is called *symmetric* if Q is a symmetric matrix ($q_{ij} = q_{ji}$).

Let

$$\boldsymbol{\pi} = (\pi_1, \ldots, \pi_s)$$

denote the initial probability distribution of Z_0 (by standard notation, we consider $\boldsymbol{\pi}$ as a row vector). $\boldsymbol{\pi}$ is called a *stationary distribution* if

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

$\boldsymbol{\pi}$ is the eigenvector from the left of Q with the eigenvalue 1.

It is known that the stationary distribution exists uniquely under the assumption that the Markov chain is irreducible and aperiodic. We only consider Markov chains satisfying these conditions. Under these conditions, starting from an arbitrary state $Z_0 = \boldsymbol{x}_i$, the distribution of Z_t for large t is close to the stationary distribution $\boldsymbol{\pi}$. Therefore if we can construct a Markov chain with the "target" stationary distribution $\boldsymbol{\pi}$, then by running a Markov chain and discarding a large number t of initial steps (called *burn-in steps*), we can consider $Z_{t+1}, Z_{t+2}, \ldots$ as observations from the stationary distribution $\boldsymbol{\pi}$.

For our problem, the target distribution $\boldsymbol{\pi}$ is already given as the hypergeometric distribution over the fiber in (1.31). We want to construct a Markov chain over $\mathscr{F}_{\boldsymbol{t}}$ just for the purpose of sampling from the hypergeometric distribution. For this purpose the Metropolis–Hastings algorithm is very useful. By the algorithm, once we can construct an arbitrary irreducible (i.e., connected) chain over $\mathscr{F}_{\boldsymbol{t}}$, we can easily modify the stationary distribution to the given target distribution $\boldsymbol{\pi}$.

Theorem 2.2 (Metropolis–Hastings algorithm). *Let $\boldsymbol{\pi}$ be a probability distribution on $\mathscr{F}$. Let $R = (r_{ij})$ be the transition probability matrix of an irreducible, aperiodic, and symmetric Markov chain over $\mathscr{F}$. Define $Q = (q_{ij})$ by*

$$
\begin{aligned}
q_{ij} &= r_{ij} \min\left(1, \frac{\pi_j}{\pi_i}\right), \qquad i \neq j,\\
q_{ii} &= 1 - \sum_{j \neq i} q_{ij}.
\end{aligned} \tag{2.9}
$$

Then Q satisfies $\boldsymbol{\pi} = \boldsymbol{\pi} Q$.

This result is a special case of Hastings [82] and the symmetry assumption on R can be removed relatively easily. In this book we only consider symmetric R and the simple statement of the above theorem is sufficient for our purposes.

Proof (Theorem 2.2). It suffices to show that the above Q is "reversible" in the following sense.

$$
\pi_i q_{ij} = \pi_j q_{ji}. \tag{2.10}
$$

In fact, under the reversibility

$$
\pi_i = \pi_i \sum_{j=1}^{s} q_{ij} = \sum_{j=1}^{s} \pi_j q_{ji}
$$

and we have $\boldsymbol{\pi} = \boldsymbol{\pi} Q$. Now (2.10) clearly holds for $i = j$. Also for $i \neq j$

$$
\pi_i q_{ij} = \pi_i r_{ij} \min\left(1, \frac{\pi_j}{\pi_i}\right) = r_{ij} \min(\pi_i, \pi_j);
$$

hence (2.10) holds if $r_{ij} = r_{ji}$. □

Equation (2.10) is often called the *detailed balance* or detailed balance equation.

An important advantage of the Markov chain Monte Carlo method is that it does not need the explicit evaluation of the normalizing constant of the stationary distribution $\boldsymbol{\pi}$. We only need to know $\boldsymbol{\pi}$ up to a multiplicative constant. In fact in (2.9) the stationary distribution only appears in the form of ratios of its elements π_i / π_j and the normalizing constant is canceled.

Another important point in (2.9) is how the transition probability r_{ij} is modified. It is modified by $\min(1, \pi_j / \pi_i)$, which does not depend on how r_{ij} is specified.

In fact (2.9) can be understood as follows. r_{ij} is the proposal transition probability. Suppose that we are at state i and we propose to move to j with the conditional probability r_{ij} by some random mechanism. Then after the proposal, we actually move to j with probability $\min(1, \pi_j/\pi_i)$ (or stay at i with probability $1 - \min(1, \pi_j/\pi_i)$). We can do this even without knowing the value of r_{ij}, as long as it is symmetric. This fact is relevant in the application of the Markov basis, because when a Markov basis element is chosen "randomly," the probability distribution of choosing an element can be arbitrary, as long as there is a positive probability of choosing every element. Irrespective of the distribution, the Metropolis–Hastings algorithm yields a Markov chain whose stationary distribution is $\boldsymbol{\pi}$.

By Theorem 2.2 we only need to construct one Markov chain, which is irreducible, aperiodic, and symmetric. By the Metropolis–Hastings algorithm, we can then modify the transition probability to achieve the desired stationary distribution $\boldsymbol{\pi}$.

In the previous section we obtained a Markov basis for two-way tables. Once a Markov basis is obtained for some model, it is easy to construct an irreducible and symmetric Markov chain over $\mathscr{F}_{A\boldsymbol{x}^o}$, where $\boldsymbol{x}^o$ is the observed frequency vector and $\mathscr{F}_{A\boldsymbol{x}^o}$ is the fiber containing $\boldsymbol{x}^o$. For example, at each step of the Markov chain, randomly choose an element $\boldsymbol{z} \in \mathscr{B}$ of the Markov basis and the sign $\varepsilon \in \{-1, +1\}$. If $\boldsymbol{x} + \varepsilon\boldsymbol{z} \in \mathscr{F}_{\boldsymbol{t}}$ then we move to $\boldsymbol{x} + \varepsilon\boldsymbol{z}$. If $\boldsymbol{x} + \varepsilon\boldsymbol{z} \notin \mathscr{F}_{\boldsymbol{t}}$ we stay at $\boldsymbol{x}$. Then the resulting Markov chain is irreducible and symmetric. It is important to note that this holds irrespective of the distribution of choosing an element from $\mathscr{B}$, as long as each element of $\mathscr{B}$ is chosen with positive probability. On the other hand, the sign of ε should be chosen with probability $1/2$.

We can then apply the Metropolis–Hastings algorithm of Theorem 2.2 to this Markov chain. The resulting algorithm is given as follows.

Algorithm 2.1

Input: Observed frequency vector $\boldsymbol{x}^o$, Markov basis $\mathscr{B}$, number of steps N, configuration A, the null distribution $f(\cdot)$, test statistic $T(\cdot)$, $\boldsymbol{t} = A\boldsymbol{x}^o$.
Output: Estimate of the p-value.
Variables: $\texttt{obs}$, $\texttt{count}$, $\texttt{sig}$, $\boldsymbol{x}$, $\boldsymbol{x}_{next}$.
Step 1: $\texttt{obs} = T(\boldsymbol{x}^o)$, $\boldsymbol{x} = \boldsymbol{x}^o$, $\texttt{count} = 0$, $\texttt{sig} = 0$.
Step 2: Choose $\boldsymbol{z} \in \mathscr{B}$ randomly. Choose $\varepsilon \in \{-1, +1\}$ with probability $\frac{1}{2}$.
Step 3: If $\boldsymbol{x} + \varepsilon\boldsymbol{z} \notin \mathscr{F}_{\boldsymbol{t}}$ then $\boldsymbol{x}_{next} = \boldsymbol{x}$ and go to Step 5. If $\boldsymbol{x} + \varepsilon\boldsymbol{z} \in \mathscr{F}_{\boldsymbol{t}}$ then let u be a uniform random number between 0 and 1.
Step 4: If $u \le \frac{f(\boldsymbol{x}+\varepsilon\boldsymbol{z})}{f(\boldsymbol{x})}$ then let $\boldsymbol{x}_{next} = \boldsymbol{x} + \varepsilon\boldsymbol{z}$ and go to Step 5. If $u > \frac{f(\boldsymbol{x}+\varepsilon\boldsymbol{z})}{f(\boldsymbol{x})}$ then let $\boldsymbol{x}_{next} = \boldsymbol{x}$ and go to Step 5.
Step 5: If $T(\boldsymbol{x}_{next}) \ge \texttt{obs}$ then let $\texttt{sig} = \texttt{sig} + 1$.
Step 6: $\boldsymbol{x} = \boldsymbol{x}_{next}$, $\texttt{count} = \texttt{count} + 1$.
Step 7: If $\texttt{count} < N$ then go to Step 2.
Step 8: The estimate of p-value is $\texttt{sig}/N$.

We should mention one important point concerning the counting of steps. There are two cases where we stay at the same state $\boldsymbol{x}_{next} = \boldsymbol{x}$. One case is that we reject a move $\boldsymbol{z}$ because $\boldsymbol{x} + \varepsilon\boldsymbol{z} \notin \mathscr{F}_{\boldsymbol{t}}$ in Step 3. Another case is that the proposed state is

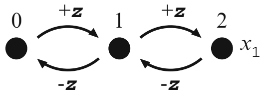

Fig. 2.1 The fiber $\mathscr{F}_2$

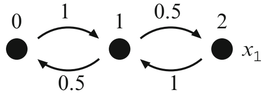

Fig. 2.2 Transition probabilities ignoring rejections in Step 3

rejected because of $u > f(\boldsymbol{x}+\varepsilon\boldsymbol{z})/f(\boldsymbol{x})$ in Step 4. In both cases, we evaluate the value of the test statistic $T(\boldsymbol{x}_{next}) = T(\boldsymbol{x})$ and the counter `count` is increased. For unbiased estimation of the p-value, we need to include both cases in evaluation of T and the counting of the steps.

In Step 3, if $\boldsymbol{x}$ is close to the boundary of $\mathscr{F}_{\boldsymbol{t}}$, then it may be the case that $\boldsymbol{x}+\varepsilon\boldsymbol{z} \notin \mathscr{F}_{\boldsymbol{t}}$ with high probability. In this case we might be tempted to choose $\boldsymbol{z}$ depending on $\boldsymbol{x}$ such that the probability of $\boldsymbol{x}+\varepsilon\boldsymbol{z} \in \mathscr{F}_{\boldsymbol{t}}$ is higher. This is an interesting topic for investigation, although it is not trivial to guarantee the symmetry $r_{ij} = r_{ji}$ if we choose a move depending on the state.

The above point can be illustrated by the following very simple example. Consider a configuration $A = (1,1)$, which is a 1×2 matrix. Let $t = A\boldsymbol{x}$, $\boldsymbol{x} = (x_1, x_2)'$ and consider the fiber with $t = 2$:

$$\mathscr{F}_2 = \{(x_1,x_2) \mid x_1 + x_2 = 2, x_1, x_2 \in \mathbb{N}\}, \qquad \mathbb{N} = \{0,1,2,\ldots\}.$$

Then $\boldsymbol{z} = (1,-1)'$ is a move, which obviously connects $\mathscr{F}_2$. The fiber is depicted as in Fig. 2.1, where the states are labeled by the values of x_1.

Note that $\boldsymbol{z}$ cannot be subtracted from $(0,2)$ and $\boldsymbol{z}$ cannot be added to $(2,0)$, because these operations produce -1. Therefore if we are at $(0,2)$ we can only add $\boldsymbol{z}$. Similarly if we are at $(2,0)$ we can only subtract $\boldsymbol{z}$. Now suppose that we want to sample from the uniform distribution over $\mathscr{F}_2$. Then in the Metropolis–Hastings algorithm, $\min(1, \pi_j/\pi_i) \equiv 1$. Therefore we stay at the same state only because of Step 3 of Algorithm 1. If we ignore the rejections in Step 3 for this example, the transition probabilities of the chain are depicted in Fig. 2.2. The stationary distribution of this chain is given by

$$(\pi(0,2), \pi(1,1), \pi(2,0)) = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right),$$

which is not uniform.

On the other hand if we count the rejections in Step 3, then the Markov chain has self-loops and the transition probabilities of the chain are depicted in Fig. 2.3. For this chain the stationary distribution is the uniform distribution, which was our target.

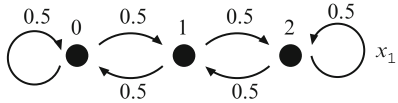

Fig. 2.3 Transition probabilities taking rejections in Step `3` into account

Algorithm 2.1 is a very simple algorithm and various improvements are possible. For example, grouping several steps of Algorithm 2.1 in one step makes the convergence to the stationary distribution faster. This can be achieved as follows.

Algorithm 2.2 Modify Steps `2`, `3`, `4` in Algorithm 2.1 as follows.

Step `2` : Choose $\boldsymbol{z} \in \mathscr{B}$ randomly.
Step `3` : Let $I = \{n \mid \boldsymbol{x} + n\boldsymbol{z} \in \mathscr{F}_{\boldsymbol{t}}\}$.
Step `4` : Choose $\boldsymbol{x}_{\text{next}}$ from $\{\boldsymbol{x} + n\boldsymbol{z} \mid n \in I\}$ according to the probability

$$p_n = \frac{f(\boldsymbol{x} + n\boldsymbol{z})}{\sum_{n \in I} f(\boldsymbol{x} + n\boldsymbol{z})}.$$

Note that both in Algorithms 2.1 and 2.2, the target distribution $f(\cdot)$ appears in the form of the ratio. Hence we do not need to compute the normalizing constant for $f(\cdot)$. Often the computation of the normalizing constant is difficult, therefore this is an important advantage of the Markov chain Monte Carlo method.

Springer

http://www.springer.com/978-1-4614-3718-5

Markov Bases in Algebraic Statistics
Aoki, S.; Hara, H.; Takemura, A.
2012, XII, 300 p., Hardcover
ISBN: 978-1-4614-3718-5