

## 第12章 马尔可夫链蒙特卡罗方法

从本书的前面章节中我们已经看到，模拟是概率统计学中一种强大的技术。如果你无法说服你的朋友在蒙提·霍尔问题 (Monty Hall problem) 中换门是个好主意，你可以在一秒钟内模拟几千次游戏，然后你的朋友将会看到换门成功的几率大约为2/3。如果你不确定如何计算一个随机变量 $X$ 的均值和方差，但你知道如何生成该分布的一组独立同分布的抽样 $X_1, X_2, \dots, X_n$ ，你可以使用模拟抽样的均值和样本方差来近似真实的均值和方差：

$$E(X) \approx \frac{1}{n}(X_1 + \dots + X_n) = \bar{X}_n,$$

$$Var(X) \approx \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2.$$

大数定律告诉我们，如果 $n$ 很大，这些近似将会很好。我们可以通过增加 $n$ 来获得更好的近似，只需要运行计算机更长的时间而不必努力处理可能难以解决的求和或积分。如第10章所讨论的，这种通过生成随机值来近似量的模拟方法，被称为蒙特卡罗 (Monte Carlo) 方法。

使用蒙特卡罗方法进行模拟的关键在于我们必须知道如何高效地生成 $X_1, X_2, \dots, X_n$ 的一组观测样本；例如，考虑模拟一个连续随机变量的分布，其概率密度函数为

$$f(x) \propto x^{3.1}(1-x)^{4.2}, \quad 0 < x < 1.$$

尽管 $f(x)$ 是贝塔 (Beta) 分布 $Beta(4.1, 5.2)$ 的概率密度，但从概率密度函数出发并不能直接得到对应该密度函数的随机变量。贝塔分布的累积分布函数 (CDF) 很难求，更不用说求解其逆函数了。关于贝塔分布的模拟在例12.1.4中有更详细的解释。

实际应用中发现的一些分布远比贝塔分布复杂；我们知道贝塔分布的分布函数的归一化常数表达式依赖于伽马 (Gamma) 函数，而对于其他许多具有某种特征的分布，其分布函数的归一化常数的形式往往未知，在这种情况下，即使是运算速度最快的计算机和最优的算法也无法实现。

本章介绍基于马尔可夫链的一种算法——马尔可夫链蒙特卡罗 (Markov chain Monte Carlo, MCMC) 方法，这是一类高效模拟复杂分布的算法。MCMC方法的发展彻底改变了统计计算和科学计算，大大扩展了可以模拟的分布范围，其中包括高维联合分布。该方法的基本思想是建立一个马尔可夫链，使得感兴趣的分布是马尔可夫链的平稳分布。

在上一章，我们考察了一些特定转移矩阵 $Q$ 的马尔科夫链，并试图得到链的平稳分布 $s$ ；而本章反过来，想从模拟的分布 $s$ 出发，建立一个平稳分布为 $s$ 的马尔科夫链，即在运行该马尔科夫链足够长的时间后，链的平分布近似为 $s$ 。

那么是否能直接设计一个转移矩阵 $Q$ ，使其马尔可夫链的平稳分布成为我们想要的分布？假如可以，这个思路的实施又是否比通过模拟获得该分布的随机样本这一原始思路更容易？然而，MCMC方法展示了高度的一般性，它使得以一种简单的方式建立一个平稳分布已知的马尔可夫链，而不必知道该分布的归一化常数成为可能！

现在MCMC方法广泛应用于生物学、自然科学和物理学，并且已经发展了许多不同的MCMC算法。这里将介绍两个最重要、应用最广泛的MCMC算法：Metropolis-Hastings方法和Gibbs抽样。MCMC方法是一个宏大的、不断增长的统计计算领域，关于它的更多理论、方法和应用方面的知识可参考Brooks、Gelman、Jones和Meng的文章[2]。

## 12.1 Metropolis-Hastings方法

Metropolis-Hastings方法应用十分广泛，它允许从已知状态空间上的任意不可约马尔可夫链开始，经过调整，使其收敛到具有给定平稳分布的新马尔可夫链；这个过程需要在初始链中引入一些可选择性：在初始链上对状态提出转移建议，但是可以接受或不接受该建议。例如，假设初始链处于称为“波士顿”的状态且其潜在转移状态是“旧金山”，则对于新链，它既可以接受建议去“旧金山”，也可以不接受该建议仍留在“波士顿”。通过认真选择接受建议的概率，这种简单的调整确保了新的马尔可夫链收敛到给定的平稳分布。

### 算法12.1.1 Metropolis-Hastings方法

已知 $s = (s_1, \dots, s_M)$ 是状态空间 $\{1, \dots, M\}$ 上给定的平稳分布。假设对任意状态 $i, s_i > 0$  否则，只需将状态空间中任意使 $s_i = 0$ 的状态删除）且马尔可夫链在该状态空间上的转移矩阵 $P = (p_{ij})$ ，直观上看， $P$ 可以代表一条转移过程已知的马尔可夫链，但该链不一定收敛到给定的平稳分布。

我们的目的是通过调整转移矩阵 $P$ ，建立平稳分布为 $s$ 的马尔可夫链 $X_0, X_1, \dots$ ，为此将给出一个Metropolis-Hastings算法。

对任意初始状态 $X_0$  随机的或确定的），设马尔可夫链当前位于状态 $X_n$ ，链的一次转移遵循以下步骤：

1. 若 $X_n = i$ ，依据初始转移矩阵 $P$ 中第 $i$ 行的各转移概率给出转移状态到 $j$ 的建议；
2. 计算接受概率

$$a_{ij} = \min \left( \frac{s_j p_{ji}}{s_i p_{ij}}, 1 \right)$$

3. 抛掷一枚正面朝上概率为 $a_{ij}$ 的硬币；
4. 若硬币正面朝上，则接受建议，从而链在下一时刻转移至状态 $j$ ，即 $X_{n+1} = j$ ；否则，拒绝该建议，链在下一时刻转移至状态 $i$ ，即 $X_{n+1} = i$ 。

也就是说，Metropolis-Hastings方法使马尔可夫链根据初始转移概率 $p_{ij}$ 为下一时刻的状态提出建议，并以 $a_{ij}$ 的概率接受该建议，以 $1 - a_{ij}$ 的概率保持当前状态不变；这个算法的优点就在于不需要知道分布 $s$ 的归一化常数，因为在 $s_j/s_i$ 中常数都会被抵消掉。比如，在一些问题中，我们可能会希望平稳分布在所有状态上是均匀的[即 $s = (1/M, 1/M, \dots, 1/M)$ ]，但是当 $M$ 较大且未知时，则求解 $M$ 将是一个非常困难的计算问题；然而，不管 $M$ 取何值 $s_j/s_i = 1$ ，则通过化简可得 $s \propto (1, 1, \dots, 1)$ ，这样就可以直接计算 $a_{ij}$ 而不用知道 $M$ 的值。

在算法运行过程中，接受概率 $a_{ij}$ 中分母上的元素 $p_{ij}$ 不能为0，否则，初始马尔可夫链将不能建议从状态 $i$ 转移到状态 $j$ ；另外，若 $p_{ii} > 0$ ，则可能出现潜在转移状态 $j$ 就是当前状态 $i$ ，此时，无论建议是否被接受，链都保持在状态 $i$ 。即表面上拒绝留在状态 $i$ 的建议，实际上下一刻仍处于状态 $i$ ，就像一个孩子说的“是的，我会留在我的房间，但不是因为你告诉我才这样做！”）

由Metropolis-Hastings方法构造的马尔可夫链是可逆的，且其平稳分布为 $s$ 。

### Proof:

设 $Q$ 为应用Metropolis-Hastings方法后的马尔可夫链的转移矩阵。只需要验证可逆条件 $s_i q_{ij} = s_j q_{ji}$ 对任意的 $i, j$ 都成立。

当 $i = j$ 时，等式显然成立；

当 $i \neq j$ 时，若 $q_{ij} > 0$ ，则 $p_{ij} > 0$ ，这是因为其逆否命题“若 $p_{ij} = 0$ ，则 $q_{ij} = 0$ ”成立，（即如果状态 $i$ 的潜在转移不能是状态 $j$ ，那么该链就无法从状态 $i$ 转移至状态 $j$ 。）且 $p_{ji} > 0$  否则，接受概率为0）；反过来，若已知 $p_{ij} > 0, p_{ji} > 0$ ，则 $q_{ji} > 0$ 。故 $q_{ij}$ 与 $q_{ji}$ 同时为0，或者同时非零。不妨设它们同时非零，则有

$$q_{ij} = p_{ij} a_{ij},$$

即状态 $i$ 转移至状态 $j$ 等价于建议状态 $i$ 转移至状态 $j$ 并接受这个建议。

当 $s_j p_{ji} \leq s_i p_{ij}$ 时，有

$$a_{ij} = \frac{s_j p_{ji}}{s_i p_{ij}}, a_{ji} = 1,$$

故

$$s_i q_{ij} = s_i p_{ij} a_{ij} = s_i p_{ij} \frac{s_j p_{ji}}{s_i p_{ij}} = s_j p_{ji} = s_j p_{ji} a_{ji} = s_j q_{ji}.$$

同理，当 $s_j p_{ji} > s_i p_{ij}$ 时，在上式中调整 $i$ 和 $j$ 的位置可得 $s_i q_{ij} = s_j q_{ji}$ 。

由此可知，可逆条件成立，且 $s$ 是由转移矩阵 $Q$ 构造的马尔可夫链的平稳分布。

### 12.1.2

Metropolis-Hastings方法是构造平稳分布已知的马尔可夫链的最一般方法。在上述公式中，分布 $s$ 和建议概率的分布 $P$ （以下简称建议分布）都很一般，除了要求在相同的状态空间之外，并没有任何相关性要求；然而，在实际中，建议分布的选择却非常重要，因为它的不同选择对链收敛到其平稳分布的速度有很大影响。

如何选择好的建议分布是一个复杂的问题，这里不详细讨论。直观地，接受概率非常低的建议分布会使马尔可夫链的收敛速度非常缓慢（因为链很少会转移到其他状态）；而高接受概率也可能不是理想的方案，因为这会导致马尔可夫链更倾向于选择变动很小的建议，在状态空间较大时，马尔可夫链将需要很长时间才能遍历整个空间。

这里有几个例子说明如何使用Metropolis-Hastings方法通过分布进行模拟。

### 例12.1.3 Zipf分布的模拟

记 $M \geq 2$ 为整数。随机变量 $X$ 服从参数为 $a > 0$ 的Zipf分布，如果 $X$ 的概率分布函数满足：

$$P(X = k) = \frac{1}{k^a} / \sum_{j=1}^M \left( \frac{1}{j^a} \right),$$

其中， $k = 1, 2, \dots, M$ 。Zipf分布广泛应用于语言学中以研究单词出现频率的规律。

试建立一条平稳分布为Zipf分布的马尔可夫链 $X_0, X_1, \dots$ ，它满足对任意 $n$ ，有 $|X_{n+1} - X_n| \leq 1$ 成立。要求对马尔可夫链每一步的状态转移进行简要描述，比如，对任意 $n$ ，说明状态 $X_n$ 是如何转移至状态 $X_{n+1}$ 的。

#### Solution:

只需要找到一个建议分布，就可以使用Metropolis-Hastings方法。建议分布有很多种选择，一种简单的方法就是在状态空间 $\{1, 2, \dots, M\}$ 上的随机游走。即当 $i \neq 1$ 且 $i \neq M$ 时，状态 $i$ 以概率 $1/2$ 移动到状态 $i - 1$ 或 $i + 1$ ；状态为 $1$ 时，留在状态 $1$ 或移动至状态 $2$ 的概率均为 $1/2$ ；状态为 $M$ 时，状态不变或移动到状态 $M - 1$ 的概率均为 $1/2$ （见图12.1）。

□

设 $P$ 为该链的转移矩阵，且 $P$ 的平稳分布是均匀的。因为 $P$ 是对称矩阵，所以命题11.4.3适用，应用Metropolis-Hastings方法将 $P$ 转换成平稳分布为Zipf分布的马尔可夫链。

设 $X_0$ 为任意起始状态，由以下步骤得到马尔可夫链 $X_0, X_1, \dots$ ，若链当前处于状态 $i$ ，则：

1. 根据转移矩阵 $P$ 给出潜在转移状态 $j$ ；
2. 以概率 $\min\left(\frac{i^a}{j^a}, 1\right)$ 接受该建议。如果接受建议，则状态转移到 $j$ ；否则，留在状态 $i$ 。

这个过程很容易实现，每次移动需要的计算很少；注意，在链转移过程中不需要计算归一化常数。

### 例12.1.4 贝塔分布的模拟

现在回到本章开头提到的贝塔分布的模拟问题。假设我们想生成一个服从 $Beta(a, b)$ 分布的随机变量 $W$ ，即 $W \sim Beta(a, b)$ ，但我们不知道R语言中的 `rbeta` 命令，现在我们只有一组服从 $Unif(0, 1)$ 分布的独立随机变量序列。

- (a) 若 $a$ 和 $b$ 为正整数，如何使用均匀分布对 $W$ 进行模拟？
- (b) 若 $a$ 和 $b$ 为任意正实数，如何使用连续状态空间 $(0,1)$ 上的马尔可夫链近似地模拟 $W$ ? \*\*

### Solution:

(a) 由均匀分布的性质直接模拟贝塔分布比较困难。在前面的章节我们已经证明，如果  $X \sim \text{Gamma}(a, 1)$  且  $Y \sim \text{Gamma}(b, 1)$ ，并且  $X$  与  $Y$  相互独立，则有  $X/(X+Y) \sim \text{Beta}(a, b)$ 。因此，我们可以首先模拟得到伽马分布，然后再生成贝塔分布。

借助指数分布对伽马分布进行模拟。设独立随机变量序列  $X_1, \dots, X_a \sim \text{Exp}(1)$ ，则  $X = X_1 + \dots + X_a \sim \text{Gamma}(a, 1)$ ， $Y$ 也可以得到类似的结论。由于 $\text{Exp}(1)$ 分布的分布函数的逆及均匀分布的性质，有

$$-\ln(1-U) \sim \text{Exp}(1),$$

其中  $U \sim \text{Unif}(0, 1)$ ，由此能很方便地生成多组服从 $\text{Exp}(1)$ 分布的随机变量序列。

(b) 接下来使用Metropolis-Hastings方法对 $W$ 进行模拟。我们之前只介绍了Metropolis-Hastings方法在有限状态空间的应用，其在无限状态空间中的用法是类似的。一个简单的建议链可以是一列服从 $\text{Unif}(0,1)$ 的随机变量序列（独立同分布的随机变量满足马尔可夫性），这样在区间 $(0,1)$ 上的建议状态总是一个新的 $\text{Unif}(0,1)$ 分布，且与之前状态独立。这种生成Metropolis-Hastings链的方法称为**独立抽样法**。

设 $W_0$ 为任意起始状态，并按以下步骤生成马尔可夫链 $\{W_0, W_1, \dots\}$ 。若该链现处于状态 $w, w \in (0, 1)$ ，则：

1. 随机抽取一个服从 $\text{Unif}(0,1)$ 的随机变量来生成建议 $u$ ；
2. 以概率  $\min\left(\frac{u^{a-1}(1-u)^{b-1}}{w^{a-1}(1-w)^{b-1}}, 1\right)$  接受此建议。若接受该建议，则转移至 $u$ ；否则，留在 $w$ 。

上述链的生成过程同样不必求解归一化常数；转移过程中为得到接受概率，贝塔分布 $\text{Beta}(a,b)$ 的概率密度函数起到了 $s$ 的作用，因为 $\text{Beta}(a,b)$ 的概率密度函数是我们想获得的平稳分布，而建议序列是独立于当前状态且服从 $(0,1)$ 区间上均匀分布的随机变量序列，故 $\text{Unif}(0,1)$ 的概率密度函数起到了转移概率 $p_{ij}$ （同理 $p_{ji}$ ）的作用。

运行马尔可夫链可以得到，当 $n$ 足够大时， $W_n, W_{n+1}, W_{n+2}, \dots$ 近似服从 $\text{Beta}(a,b)$ ；注意，序列 $W_n, W_{n+1}, W_{n+2}, \dots$ 不是随机抽取的独立同分布序列，而是相关的随机变量序列。

## 12.1.5 MCMC方法产生相关样本



运用MCMC方法生成马尔可夫链 $\{X_n\}_{n=0,1,2,\dots}$ 时，一个重要的问题是算法运行的时间。这部分是因为通常难以知道在某个具体时刻马尔可夫链的分布与平稳分布的接近程度；另一个原因是 $X_0, X_1, \dots$ 通常是相关的。一些马尔可夫链在转移过程中趋于停留在状态空间的某些状态子集上，而不是遍历整个状态空间；而如果链很容易停留在某个状态，则 $X_n$ 可能与 $X_{n+1}$ 高度正相关。滞后阶为 $h$ 的自相关指的是随着 $n$ 趋于无穷时 $X_n$ 与 $X_{n+h}$ 之间的相关性。高度自相关往往意味着蒙特卡罗模拟具有高方差，故希望随着滞后阶数 $k$ 的增大， $k$ 阶自相关系数能迅速趋于零。

分析马尔可夫链运行时间的长短及诊断其是否已经平稳是非常活跃的研究领域，通常的做法是从不同的起点出发，将链迭代足够多次，然后考察结果的稳定性。

Metropolis-Hastings方法应用非常广泛，通常应用于较大状态空间的情形；也适用于某些看起来似乎与分布模拟毫无关系的问题，如密码破译。

### 例12.1.6 密码破译

最近，马尔可夫链被应用于密码破译，下面介绍一个应用案例。关于这种应用的更多信息，参见Diaconis[7]和Chen和Rosenthal[4]。）替代密码指字母a~z的一个排列 $g$ ，通过使用 $g(\alpha)$ 代替每一个字母 $\alpha$ 给信息加密，比如：

```
1 | abcdefghijklmnopqrstuvwxyz
2 | zyxwvutsrqponmlkjihgfedcba
```

上面第二行列出了 $g(a), g(b), \dots, g(z)$ 的值；由此我们可以得到“statistics”的替代密码是“hggrhgrxh”。我们也可以对大写字母、空格和标点符号进行转化。）上述状态空间为字母a~z的所有排列 $26! \approx 4 \times 10^{26}$ 。这是一个非常大的空间：假设必须要用这些排列中的一组来解码文本，并且每纳秒(ns)处理一个排列，则处理完所有排列仍需要120亿年的时间；因此，一个一个地对每个排列直接进行试验是不可行的，下面将使用随机排列的方式。

(a) 考察这样一个马尔可夫链，从a~z的26个字母中随机选取两个并将第二行相应位置的字母交换。比如，选取的位置为7和20，则原始排列

```
1 | abcdefghijklmnopqrstuvwxyz
2 | zyxwvutsrqponmlkjihgfedcba
```

转化为

```
1 | abcdefghijklmnopqrstuvwxyz
2 | zywxvugsrqponmlkjihfedcba
```

试求排列 $g$ 到排列 $h$ 的一步转移概率，及该马尔可夫链的平稳分布。

(b) 假设现在有一个系统为每个排列 $g$ 赋予一个正的得分 $s(g)$ ；直观上看，在给定 $g$ 是所使用的密码时， $s(g)$ 可以作为衡量破译已观察到的加密文本的可能性。请使用Metropolis-Hastings方法构造一个马尔可夫链，使其平稳分布与得分 $s(g)$ 的分布成比例。

### Solution:

(a) 除非通过交换第二行的两个字母可以使 $g$ 获得 $h$ ，否则自 $g$ 到 $h$ 的一步转移概率为0；假设 $h$ 能够以这种方式由 $g$ 得到，相应的概率为 $\frac{1}{\binom{26}{2}}$ 。由此得到的马尔可夫链是不可约的，因为只要进行足够多次的交换，总能从任意一组排列转移到其他任意排列。其中， $p(g, h) = p(h, g)$ ， $p(g, h)$ 表示状态 $g$ 到状态 $h$ 的转移概率，故转移矩阵是对称矩阵，从而平稳分布均匀地分布在由字母 $a \sim z$ 得到的 $26!$ 个排列上。

(b) 运用(a)中结论，自任意状态 $g$ 出发，由(a)中马尔可夫链得到建议转移状态 $h$ ；接着抛掷一枚正面朝上概率为 $\min\left(\frac{s(h)}{s(g)}, 1\right)$ 的硬币，若出现正面，转移至 $h$ ；反之，则留在 $g$ 。

为了证明该过程具有所期望的平稳分布，可以运用算法12.1.1的一般性证明，也可以直接验证可逆性条件；本题我们选择后者。

需要验证对于所有的 $g$ 和 $h$ ，有 $s(g)q(g, h) = s(h)q(h, g)$ ，其中 $q(g, h)$ 是更新的马尔可夫链自状态 $g$ 到状态 $h$ 的转移概率。

若 $g = h$ 或 $q(g, h) = 0$ ，则该方程明显成立，下面验证 $g \neq h$ 和 $q(g, h) \neq 0$ 的情形； $p(g, h)$ 是(a)中的转移概率，表示自状态 $g$ 转移至状态 $h$ 的潜在转移概率。

首先，考虑 $s(g) \leq s(h)$ 的情形，则由 $q(g, h) = p(g, h)$ 以及

$$q(h, g) = p(h, g) \frac{s(g)}{s(h)} = p(g, h) \frac{s(g)}{s(h)} = q(g, h) \frac{s(g)}{s(h)},$$

故 $s(g)q(g, h) = s(h)q(h, g)$ 。同理，当 $s(h) < s(g)$ 时，将状态 $g$ 和 $h$ 交换顺序，可得到同样的结论。因此，状态 $g$ 的平稳概率与它的得分 $s(g)$ 成正比。

综上所述，在运用Metropolis-Hastings方法时，从一个马尔可夫链开始，该链从长期来看等可能地到达所有密码序列，并构建一个其平稳分布)依据得分将密码进行整理的马尔可夫链，该链从长期来看更频繁地到达最有可能的密码。

下面是另一个具有巨大状态空间的MCMC方法的例子，同样地，这个例子乍看起来似乎与模拟分布没有太多关系。但他却指出MCMC方法不仅可以用于抽样，还可以用于[优化](#)。

### 例12.1.7 背包问题

Bilbo the Burglar在Smaug的巢穴中找到了 $m$ 件宝藏，正在决定偷走哪些宝藏。他不能一次带走所有的宝藏，因为他能够携带的最大重量是 $w$ 。假设第 $j$ 件宝贝价值 $g_j$ 个金币且重量为 $w_j$ ；因此，Bilbo必须选择一个向量 $x = (x_1, \dots, x_m)$ ，其中，如果他带走第 $j$ 件宝贝，则 $x_j$ 取值为1；否则为0，并且 $x_j$ 取值为1的宝物的总重量不能超过 $w$ 。设 $\mathbb{C}$ 为所有这些向量的空间，所以 $\mathbb{C}$ 包含所有二进制向量 $(x_1, \dots, x_m)$ 使得 $\sum_{j=1}^m x_j w_j \leq w$ 。

Bilbo 希望最大化他带走的宝藏的总价值。找到最优解是一个极其困难的问题，称为背包问题，它在计算机科学中有悠久的历史。蛮力解决方案通常是完全不可行的。Bilbo 决定改为使用 MCMC 探索空间 $\mathbb{C}$ ——幸运的是，他带着运行 R 的笔记本电脑。

(a) 考虑以下马尔可夫链。从 $(0, 0, \dots, 0)$ 开始。链的一次移动如下。假设当前状态为 $x = (x_1, \dots, x_m)$ 。选择 $\{1, 2, \dots, m\}$ 中的一个均匀随机数 $J$ ，并通过将 $x_J$ 替换为 $1 - x_J$ （即，切换是否将该宝藏带走）从 $x$ 获得 $y$ 。如果 $y$ 不在 $\mathbb{C}$ 中，停留在 $x$ ；如 $y$ 在 $\mathbb{C}$ 中，移动到 $y$ 。展示 $\mathbb{C}$ 上的均匀分布对于这个链是平稳的。

(b) 展示 (a) 中的链是不可约的，并且它可能是非周期性的（取决于  $w, w_1, \dots, w_m$ ）。

(c) (a) 中的链是获得近似均匀解的有用方法，但 Bilbo 更感兴趣的是找到价值（以金币计）较高的解。在这部分中，目标是构造一个其平稳分布对任何特定高价值解赋予远高于任何特定低价值解的马尔可夫链。

具体来说，假设我们想从分布

$$s(x) \propto e^{\beta V(x)},$$

模拟，其中  $V(x) = \sum_{j=1}^m x_j g_j$  是  $x$  以金币计算的价值，而 $\beta$ 是一个正常数。这个分布的想法是给每个高价值解比每个低价值解更多的指数级概率。创建一个其平稳分布如所期望的马尔可夫链。

解：

(a) 由于对于  $x = y$ ，从  $x$  到  $y$  和从  $y$  到  $x$  的转移概率要么同时为 0 要么同时为  $1/m$ ，所以转移矩阵是对称的。因此，平稳分布在 $\mathbb{C}$ 上是均匀的。

(b) 我们可以通过一次丢弃一个宝藏从任何  $x \in \mathbb{C}$  到达  $(0, 0, \dots, 0)$ 。我们可以通过一次捡起一个宝藏从  $(0, 0, \dots, 0)$  到达任何  $y \in \mathbb{C}$ 。结合这些，我们可以从任何地方到达任何地方，所以链是不可约的。要研究周期性，让我们看一些简单的情况。首先考虑  $w_1 + \dots + w_m < w$  的简单情况，即 Bilbo 可以同时携带所有宝藏。那么所有长度为  $m$  的二进制向量都是允许的。所以  $(0, 0, \dots, 0)$  的周期是 2，因为从那个状态开始，Bilbo 需要捡起然后放下一个宝藏才能回到那个状态。事实上，如果 Bilbo 从  $(0, 0, \dots, 0)$  开始，任何奇数次移动后他都会携带奇数个宝藏。

现在考虑  $w_1 > w$  的情况，即第一个宝藏对 Bilbo 来说太重了。从任何  $x \in \mathbb{C}$ ，链尝试捡起第一个宝贝的概率是  $1/m$ ，如果发生这种情况，链将停留在  $x$ 。所以每个状态的周期是 1。

(c) 我们可以使用 (a) 中的链来提出建议。从  $(0, 0, \dots, 0)$  开始。假设当前状态为  $x = (x_1, \dots, x_m)$ 。然后：



1. 在  $\{1, 2, \dots, m\}$  中选择一个均匀随机  $J$ ，并通过将  $x_J$  替换为  $1 - x_J$  从  $x$  获得  $y$ 。
2. 如果  $y$  不在  $\mathbb{C}$  中，停留在  $x$ 。如果  $y$  在  $\mathbb{C}$  中，抛出一枚硬币，正面朝上的概率为  $\min\left(\frac{e^{\beta V(y)}}{e^{\beta V(x)}}, 1\right)$ 。如果是正面，转移到  $y$ ；如果是反面，停留在  $x$ 。

### 例12.1.8 正态共轭

设  $Y|\theta \sim N(\theta, \sigma^2)$ ，其中  $\sigma^2$  已知，但  $\theta$  未知。使用贝叶斯框架，我们将  $\theta$  视为一个随机变量，其先验分布给定为  $\theta \sim N(\mu, \tau^2)$ ， $\mu$  和  $\tau^2$  为已知常数。因此，我们有以下两级模型：

$$\begin{aligned}\theta &\sim N(\mu, \tau^2) \\ Y|\theta &\sim N(\theta, \sigma^2)\end{aligned}$$

在观察到  $Y$  的值后，使用 Metropolis-Hastings 算法找到  $\theta$  的后验均值和方差。

### Solution:

在观察到  $Y = y$  后，我们可以使用贝叶斯规则更新我们对  $\theta$  的先验不确定性。因为我们对  $\theta$  的后验分布感兴趣，任何与  $\theta$  无关的项都可以被视为归一化常数的一部分。因此，

$$f_{\theta|Y}(\theta|y) \propto f_{Y|\theta}(y|\theta)f_{\theta}(\theta) \propto \exp\left(-\frac{1}{2\sigma^2}(y - \theta)^2\right) \exp\left(-\frac{1}{2\tau^2}(\theta - \mu)^2\right)$$

由于数中  $\theta$  的二次函数，我们认出  $\theta$  的后验 PDF 为一个正态 PDF。后验分布保持在正态族中，这告诉我们正态是正态的共轭先验。通过完成平方（这里我们省略了一个相当繁琐的计算），我们可以获得  $\theta$  后验分布的显式公式：

$$\theta|Y = y \sim N\left(\frac{\sigma^{-2}y + \tau^{-2}\mu}{\sigma^{-2} + \tau^{-2}}, (\sigma^{-2} + \tau^{-2})^{-1}\right)$$

这个公式表明  $\theta$  的后验均值  $E(\theta|Y = y)$  是先验均值  $\mu$  和观察数据  $y$  的加权平均，权重由我们在获得数据之前对  $\theta$  的认识（即  $\tau^2$  的先验方差）和数据的测量精度（即  $\sigma^2$ ）决定。如果在获得数据之前我们已经对  $\theta$  非常确定，则  $\tau^2$  将会很小，使得  $\mu$  的权重很大。另一方面，如果数据非常精确，则  $\sigma^2$  会很小，这时数据  $y$  会被赋予较大的权重。

对于后验方差，如果将精度定义为方差的倒数，那么结果简单地表明  $\theta$  的后验精度是先验精度  $1/\tau^2$  与观测数据精度  $1/\sigma^2$  的和。

这些结论非常有用，但如果我们不知道如何完成平方，或者我们想要检验特定值  $y$ 、 $\sigma^2$ 、 $\mu$ 、 $\tau^2$  的计算结果，也就是说，无法直接获得准确的后验分布时，应该如何计算后验均值和方差呢？这时我们可以通过使用 Metropolis-Hastings 算法来构建一个其平稳分布为  $f_{\theta|Y}(\theta|y)$  的马尔可夫链，以模拟  $\theta$  的后验分布。同样的方法也可以应用于其他更复杂的分布的分析模拟。

生成  $\theta_0, \theta_1, \dots$  的一种 Metropolis-Hastings 方法如下：

1. 如果 $\theta_n = x$ ，则根据一些转移规则提出新的状态 $x'$ 。在连续状态空间中做到这一点的一种方法是生成一个均值为0的正态随机变量 $\varepsilon_n$ 并将其加到当前状态上以得到建议状态：换句话说，我们生成 $\varepsilon_n \sim N(0, d^2)$  ( $d$ 是某个常数)，然后设置 $x' = x + \varepsilon_n$ 。
2. 接受概率为 $a(x, x') = \min\left(1, \frac{f_{\theta|Y}(x'|y)}{f_{\theta|Y}(x|y)}\right)$ ，其中 $f_{\theta|Y}(x'|y)$ 和 $f_{\theta|Y}(x|y)$ 分别是在 $x'$ 和 $x$ 处的后验PDF值。
3. 独立于马尔可夫链，抛掷一枚正面朝上概率为 $a(x, x')$ 的硬币。
4. 如果硬币正面朝上，则接受该建议，状态转移到 $x'$ ，设 $\theta_{n+1} = x'$ ；否则，保持当前状态，设 $\theta_{n+1} = x$ 。

通过运行算法足够次迭代，给定 $Y = 3$ ， $\mu = 0$ ， $\sigma^2 = 1$ ， $\tau^2 = 4$ ，和 $d = 1$ 。图12.2展示了从 $\theta$ 的后验分布中提取的样本绘制的直方图。这个直方图显示 $\theta$ 的后验分布确实看起来像正态分布。我们可以使用样本的均值和方差来估计后验均值和方差。对于我们获得的样本，样本均值为2.4，样本方差为0.8，与理论值相当一致：

$$E(\theta|Y=3) = \frac{\frac{1}{\sigma^2}y + \frac{1}{\tau^2}\mu}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}} = \frac{\frac{1}{1} \cdot 3 + \frac{1}{4} \cdot 0}{\frac{1}{1} + \frac{1}{4}} = 2.4$$

$$Var(\theta|Y=3) = \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1} = \left(\frac{1}{1} + \frac{1}{4}\right)^{-1} = 0.8$$

后验均值更接近观察到的数据而不是先验均值，这是因为 $\tau^2$ 比 $\sigma^2$ 大，对应于先验分布具有相对较高的不确定性。使用本章提供的R代码，你可以看到不同的 $y$ 、 $\mu$ 、 $\sigma^2$ 、 $\tau^2$ 值对后验分布的影响。

为了帮助诊断我们的马尔可夫链是否充分探索了状态空间，我们可以制作迹线图，即将样本 $\theta_n$ 作为 $n$ 的函数绘制。图12.3显示了三个不同的标准差 $d$ 选择对应的迹线图，即 $d = 100$ ， $d = 1$ ，和 $d = 0.01$ 。对于 $d = 100$ 的迹线图，有许多平坦区域，其中链保持在某些状态，这表明 $d$ 太大，因此建议通常被拒绝。另一方面， $d = 0.01$ 太小；我们可以从迹线图看到，链每次移动的步长很小，并且无法远离其起点。对于 $d = 1$ 的迹线图正合适，既没有 $d = 100$ 链的低接受率，也没有 $d = 0.01$ 链的移动限制。

在这个例子中， $\theta$ 的后验分布可以直接分析，因此我们使用MCMC出于示例目的，以显示MCMC获得的结果与理论结果一致。但相同的技术也适用于先验不是共轭的和后验不是命名分布的问题。这展示了MCMC方法的强大灵活性，能够在复杂模型中估计不容易直接计算的后验分布参数。

## 12.2 Gibbs抽样

**Gibbs抽样**也是一种MCMC方法，通过每次从**条件分布**中抽取一个样本来获得**联合分布**的**近似**样本。在每个步骤中，通过从给定其他所有变量的条件分布中抽样来更新一个变量，同时保持其他变量不变。这种方法特别适用于条件分布容易得到的情况。

首先，我们将通过一个包含两个变量的情况来说明Gibbs抽样的工作方式，我们感兴趣的平稳分布是离散随机变量 $X$ 和 $Y$ 的**联合概率质量函数**。根据更新变量的**顺序**，Gibbs抽样有几种不同的形式。我们将介绍两种主要的Gibbs抽样方法：**系统扫描（systematic scan）**和**随机扫描（random scan）**。系统扫描按照**确定**的顺序更新变量，而随机扫描在每个阶段**随机**选择一个变量进行更新。

### 算法12.2.1 系统扫描Gibbs抽样

$X$  和  $Y$  是两个离散随机变量，它们的**联合概率质量函数**为  $p_{X,Y}(x,y) = P(X=x, Y=y)$ 。我们希望构造一个**二维马尔可夫链** $(X_n, Y_n)$ ，其**平稳分布**为  $p_{X,Y}$ 。**系统扫描**Gibbs抽样通过**交替**更新 $X$ 和 $Y$ 分量进行；其过程可描述为，若当前状态处于  $(X_n, Y_n) = (x_n, y_n)$ ，则在保持 $Y$ 不变的同时更新 $X$ ，接着保持 $X$ 不变更新 $Y$ ：

1. 给定  $Y = y_n$ ，从 $X$ 的**条件分布** $P(X|Y = y_n)$ 中抽取  $x_{n+1}$ ，并设置  $X_{n+1} = x_{n+1}$ 。
2. 给定  $X = x_{n+1}$ ，从 $Y$ 的**条件分布** $P(Y|X = x_{n+1})$ 中抽取  $y_{n+1}$ ，并设置  $Y_{n+1} = y_{n+1}$ 。

重复以上步骤，得到的马尔可夫链  $(X_0, Y_0), (X_1, Y_1), (X_2, Y_2), \dots$  的平稳分布即为  $p_{X,Y}$ 。

### 算法12.2.2 随机扫描Gibbs抽样

同上， $X$ 和 $Y$ 是两个具有联合概率质量函数  $p_{X,Y}(x,y) = P(X=x, Y=y)$  的离散随机变量。我们希望构造一个平稳分布为  $p_{X,Y}$  的二维马尔可夫链  $(X_n, Y_n)$ 。**随机扫描**Gibbs抽样在每次移动时**均等概率**选择一个变量进行更新，更新过程根据给定另一变量的条件分布进行：

1. 等概率选择一个要更新的变量。
2. 如果选择更新  $X$  变量，则从  $X$  的条件分布  $P(X|Y = y_n)$  中抽取  $x_{n+1}$ ，并设置  $X_{n+1} = x_{n+1}$ ， $Y_{n+1} = y_n$ ；如果选择更新  $Y$  变量，则从  $Y$  的条件分布  $P(Y|X = x_n)$  中抽取  $y_{n+1}$ ，并设置  $X_{n+1} = x_n$ ， $Y_{n+1} = y_{n+1}$ 。

重复以上步骤，得到的马尔可夫链  $(X_0, Y_0), (X_1, Y_1), (X_2, Y_2), \dots$  的平稳分布即为  $p_{X,Y}$ 。

Gibbs抽样可以自然地推广到**更高维度**。如果想从一个 $d$ 维联合分布中获得一组样本，那么我们构造的马尔可夫链将是一个 **$d$ 维随机向量序列**。在每个阶段，选择随机向量中的一个变量进行更新，并通过从给定其他变量的最新取值的**条件分布**中抽样来更新该变量。同样，既可以按照**系统顺序**循环更新随机向量的各个分量，也可以在每次更新时**等概率**地随机选择一个分量。

Gibbs抽样没有Metropolis-Hastings方法那样灵活，因为我们不能选择一个**建议分布**，但这也使得它在某种意义上更简单。Gibbs和Metropolis-Hastings的方法各有优点，Gibbs抽样强调的是**条件分布**，而Metropolis-Hastings方法则强调的是**接受概率**。然而，我们将在下文中看到，这两种算法又有着紧密的联系。

### Theorem 12.2.3

随机扫描Gibbs抽样是Metropolis-Hastings算法的一个特例，在这个特例中，建议总是被接受的。特别地，这意味着随机扫描Gibbs抽样的平稳分布正是我们感兴趣的分布。

#### Proof:

我们将在两个维度中展示这一点，但是在任何维度上的证明都是类似的。让 $X$ 和 $Y$ 是两个离散随机变量，其联合 $PMF$ 是我们感兴趣的平稳分布。让我们看看Metropolis-Hastings算法告诉我们要做什么，使用以下建议分布：从 $(x, y)$ 开始，通过运行一次随机扫描Gibbs抽样随机更新一个坐标。

为了简化符号，我们写作

$$P(X = x, Y = y) = p(x, y), P(Y = y|X = x) = p(y|x), P(X = x|Y = y) = p(x|y)$$

更正式地，我们应该写作 $p_{Y|X}(y|x)$ 而不是 $p(y|x)$ ，以避免类似于 $p(5|3)$ 这样的问题。但是写作 $p(y|x)$ 更为简洁，并且在此证明中不会引起歧义。

让我们计算从 $(x, y)$ 到 $(x', y')$ 的Metropolis-Hastings接受概率。由于每次建议只更新一个变量，因此状态 $(x, y)$ 与 $(x', y')$ 在至少一个分量上必须相等。假设 $x = x'$ （ $y = y'$ 的情况可以通过对称性处理）。然后接受概率为

$$\frac{p(x, y')p(y|x)^{\frac{1}{2}}}{p(x, y)p(y'|x)^{\frac{1}{2}}} = \frac{p(x)p(y'|x)p(y|x)}{p(x)p(y|x)p(y'|x)} = 1$$

因此，这种Metropolis-Hastings算法总是接受建议！所以它只是毫无变动地运行随机扫描Gibbs抽样。

接下来，我们将研究一些具体的Gibbs抽样例子。

### 例12.2.4 图形着色

设 $G$ 是一个网络（**network**）（也称为图（**Graph**）），其中有 $n$ 个节点（**nodes**），且每对不同的节点间可以存在或不存在连接它们的边（**edges**）。现有一个包含 $k$ 种颜色的集合，比如，若 $k = 7$ ，则颜色集可以是{红色，橙色，黄色，绿色，蓝色，靛蓝，紫色}。一个 $k$ -色网络指的是每种颜色对应一个节点，使得由边连接的两个节点不能是相同的颜色，如图12.4所示的3-色网络）。

□

图形着色是计算机科学中的一个重要命题，它具有广泛的应用，如任务安排和数独游戏等。

假设存在 $k$ -色网络 $G$ 。构造一条马尔可夫链，其状态空间为所有 $k$ -色网络 $G$ 的集合构成的空间。链的转移方式如下：从 $k$ -色网络 $G$ 出发，等概率地从 $n$ 个节点中选取一个节点，接着考察该节点的合法用色，并在所有合法用色中等概率地随机挑选一种颜色给该节点重新上色（这种随机颜色可能与当前颜色相同）。试证明该马尔可夫链可逆并求出它的平稳分布。

### Solution:

设 $C$ 是所有 $k$ -色网络 $G$ 的集合，且对于 $C$ 中任意的 $k$ -色网络 $i$ 和 $j$ ，令 $q_{ij}$ 表示从 $i$ 到 $j$ 的转移概率。下面证明 $q_{ij} = q_{ji}$ ，这意味着平稳分布在集合 $C$ 上是等概率的。

对任意 $k$ -色网络上的节点 $i$ 和节点 $v$ ，令 $L(i, v)$ 表示保持所有其他节点的颜色与它们在 $i$ 中的颜色相同时，节点 $v$ 的合法颜色数目。若 $k$ -色网络 $i$ 与 $j$ 在多于一个节点上的颜色不同，则 $q_{ij} = 0 = q_{ji}$ ；若 $i = j$ ，显然有 $q_{ij} = q_{ji}$ 。如果 $i$ 和 $j$ 恰好只在一个节点 $n$ 处颜色不同，则 $L(i, v) = L(j, v)$ ，故

$$q_{ij} = \frac{1}{n} \frac{1}{L(i, v)} = \frac{1}{n} L(j, v) = q_{ji}$$

因此，转移矩阵是对称矩阵，这表示平稳分布在状态空间上是均匀的。

这是否是Gibbs抽样的一个例子值得深思。可以将图中的每个节点视为一个离散型随机变量，它有 $k$ 个可能的取值。这些节点有联合概率分布，且由边连接的节点不能涂有相同颜色的约束，使节点之间具有复杂的依赖结构。

现在想要对整个 $k$ -色网络空间进行抽样，即从所有节点的联合分布中进行抽样；这非常困难，现在改为抽取某节点在给定所有其他节点时的条件分布。如果联合概率分布在所有合法图形上是均匀的，那么该条件分布在其合法颜色上的分布也是均匀的。因此，我们在算法中的每一步都是从相应的条件分布中抽样，即正在运行一个随机扫描Gibbs采样器。

### 例12.2.5 达尔文雀族

当达尔文访问加拉帕戈斯群岛时，他记录了在每个岛屿上观察到的雀的种类。表12.1是对达尔文统计数据的一个概括，每行和列分别对应一个物种和一个岛屿；表中 $(i, j)$ 处标记1表示在 $j$ 岛上观察到物种 $i$ 。

【

给定这些数据，我们可能想要知道表中得到的0和1是否具有某种关联。例如，行和列之间是否存在依赖关系；一些物种在某些岛屿上一起出现的预期频率是否比偶然情形下更大。这些结构也许能够揭示物种之间合作或竞争的动态关系。一种检验某种结构是否存在的方法是考察大量与表12.1具有相同行和及列和的随机列联表，并将实际观察到的列联表与随机列联表进行比较分析。这是统计学中的一种常用的方法，称为**拟合优度检验**。



但如何生成与表12.1具有相同行和与列和的随机表呢？满足上述约束的随机数表是不可能枚举的；但MCMC方法可以为我们解决这个问题：构造一条马尔可夫链，其状态空间为所有满足上述行和与列和要求的随机数表，且其平稳分布在整个状态空间上是均匀的。

为构造马尔可夫链，需要在不改变行和与列和的条件下制订一种表与表之间的转移规则：从实际观察到的列联表开始，随机选择两行与两列，若它们交汇处的四个点具有以下两种形式之一：

$$\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \text{ 或 } \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}$$

则将以概率1/2切换到相反模式；否则，留在原状态。例如，如果选择1行和3行以及1列与17列，则将以1/2的概率把其交叉处的四个数从 $\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}$ 切换成 $\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}$ 这是一个对称的转移矩阵（对于任意表 $t$ 和 $t'$ ，从 $t$ 到 $t'$ 的转移概率等于从 $t'$ 到 $t$ 的转移概率），且转移过程不改变行和与列和，同时可以表明该马尔可夫链是不可约的。因此，该链的平稳分布在所有满足上述给定的行和与列和的随机表上是均匀的。

为了给出该过程作为Gibbs采样器的解释，考虑除了1行和3行以及1列和17列相交处的四个数值之外，给定表中所有其他数值的条件分布；如果链的平稳分布在所有满足条件的表上是均匀的，那么这四个点的条件分布在满足不改变行和与列和的所有表上（即 $\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}$ 和 $\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}$ ）也一定是均匀的。因此，我们在算法的每一步中都是从四个交汇点的条件分布中进行抽样的。

与Metropolis-Hastings方法一样，Gibbs抽样也适用于连续分布，只需用条件概率密度函数代替条件概率质量函数。

### 例12.2.6

已知每只母鸡下蛋的个数  $N \sim \text{Pois}(\lambda)$ ，每个蛋孵出小鸡的概率为  $p$ ，而  $p$  是未知的；令  $p \sim \text{Beta}(a, b)$ ，且常数  $\lambda$ 、 $a$ 、 $b$  已知。

这里需要注意的是，观察不到  $N$ ，而只能观察到孵化的蛋的个数  $X$ 。分析现已观察到只孵出的  $x$  只小鸡，则如何运用Gibbs抽样求得  $p$  的后验均值  $E(p|X = x)$ 。

### Solution:

由之前章节中的例子可知，当参数  $p$  给定后， $X \sim \text{Pois}(\lambda p)$ ，而  $p$  的后验概率密度函数为

$$f(p|X = x) \propto P(X = x|p)f(p) \propto e^{-\lambda p}(\lambda p)^x p^{a-1}(1-p)^{b-1},$$

式中略去了所有与  $p$  无关的项。

由于这不是一个已命名分布，直接来看我们似乎被它难住了；然而，通过考察条件分布将摆脱这个麻烦。首先，我们希望自己知道什么呢？自然是希望知道无法观测的鸡蛋总数  $N$ 。当已知  $N = n$  和  $p$  的真实值时， $X$  服从二项分布  $\text{Bin}(n, p)$ ；而在给定鸡蛋总数的情况下，贝塔分布是二项分布的共轭分布族。由此，可以直接使用案例8.3.3,写下参数  $p$

的后验分布：

$$p|X = x, N = n \sim \text{Beta}(x + a, n - x + b)。$$

可以看到，当给定鸡蛋总数  $N$  时，问题得到简化；同时这启发我们使用Gibbs抽样来解决这个问题。在以下运算中，交替在给定  $N$  时  $p$  的分布以及给定  $p$  时  $N$  的分布中抽样；整个过程中假定  $X = x$ ，这是由于我们是在已知信息下研究参数的后验分布。

首先，需要给出  $p$  和  $N$  的一个猜测值作为初始值，接着重复以下步骤：

1. 在已知  $N = n$  和  $X = x$  的条件下，从分布  $\text{Beta}(x + a, n - x + b)$  中随机抽取一个样本作为  $p$  的新猜测；
2. 在给定  $p$  及  $X = x$  的条件下，未孵化的鸡蛋数  $Y \sim \text{Pois}(\lambda(1 - p))$ ，则可以从分布  $\text{Pois}(\lambda(1 - p))$  中随机抽取得到  $Y$  的样本值  $y$ ，并将鸡蛋总数  $N$  的新值设为  $N = x + y$ 。

经多次迭代，可以得到  $p$  和  $N$  的大量样本点；由于  $N$  仅是分析过程的辅助变量，故可忽略  $N$  的样本分布图，但这里将  $p$  和  $N$  的样本分布图都绘制了出来：图12.5表示当  $\lambda = 10, a = b = 1$  即  $p$  的先验分布为均匀分布  $\text{Unif}(0, 1)$  时，由  $N$  的样本与  $p$  的后验样本绘制的直方图，其中，观察到  $X = 7$  只孵化的小鸡。由直方图可知，在我们用观察到的数据更新对参数  $p$  的认识后， $p$  最有可能在0.7附近，且不大可能低于0.2；而鸡蛋总数  $N$  至少为7（因为已观察到7个孵化的鸡蛋），最有可能在9附近。

□

对于问题中最初要求的后验均值  $E(p|X = x)$ ，可以用  $p$  的样本均值得到其良好近似；在上述假设下，求得样本均值为0.68。使用本章R语言应用示例中提供的代码，可以尝试不同的  $\lambda, a, b$  以及  $x$  的值，以观察直方图和后验均值的变化情况。

本例的关键在于将未观察到的鸡蛋总数  $N$  添加到模型中来，以便我们能够轻松地求得条件分布，并方便地运用Gibbs抽样。

## 12.3 要点重述

MCMC方法运用马尔可夫链使我们能够从**复杂分布**中进行**抽样**，其基本思路是建立一个平稳分布为  $\pi(x)$  的马尔可夫链，得到  $\pi(x)$  的样本， $\pi(x)$  是感兴趣的概率分布；在运行马尔可夫链很长一段时间之后，马尔可夫链所处的状态序列可作为感兴趣分布的随机样本。

本章讨论的两种MCMC方法分别是**Metropolis-Hastings方法**和**Gibbs抽样**。前者使用任意一个在状态空间上**不可约**的马尔可夫链来提出建议，然后通过接受或拒绝这些**建议**进行转移，最后得到一个平稳分布为给定分布的修正马尔可夫链；Gibbs抽样则是一种用于 **$d$ -维联合分布**的抽样方法，其通过每次更新 **$d$ -维马尔可夫链的一个维度**且保持其他分量不变来实现；Gibbs抽样的迭代过程可以通过**系统扫描**（以固定顺序循环迭代各个分量）或**随机扫描**（每次随机选择迭代的变量）来完成。

