中国科学（A 辑）
第 30 卷 第 12 期 SCIENCE IN CHINA (Series A) 2000 年 12 月

多峰多维密度模拟退火的可逆算法*

龚光鲁
（清华大学数学科学系，北京 100084）

钱敏平 解 军
（北京大学概率统计系，北京 100871）

摘要 在对多峰多维密度作模拟退火时，使用 Markov 链 Monte Carlo 方法常会遇到从逸出一个峰底进入另一个峰底的困难，因为这常常需要指数长的时间，从而在实际上使算法不可能实现．为了克服这个困难，建议用一个可逆设计生成一个 Markov 链，运用它为工具所得到的模拟密度，在相当一般的情形下，可以成功地避免使算法陷入局部峰底的困难．

关键词 Gibbs 采样 多峰性 可逆 Markov 链

Gibbs 采样是一种可以用来模拟相当高维密度的 Markov 链 Monte Carlo 方法（MCMC），它可以用于处理估计、滤波、预测、Bayes 后验分布、统计诊断等等问题，此时的目标密度显现为该 Markov 链的不变密度．然而，Markov 链的发展，从不变密度的一个峰底转移到另一个峰底常常需经历指数长的时间．所以，通常人们使用一系列初值以免被套死在某个峰底．但是，如何组装这些不同初值的 Markov 发展，来得到不变测度 $\pi(X)$ 的一个整体模拟呢？Besag 和 Green[1] 曾用辅助变量来处理多峰性．在本文中，我们建议采用另一种算法，即可逆 Markov 链算法，来构造 Gibbs 采样的整体模拟．在此算法中，Markov 链的“局部遍历性”或亚稳定性[2]，保证了此链在实际上可行的时间内收敛到不变测度的指定的一个峰底区，最后再给出落入各个峰底区的相对比值，以完成整体装配．

令 $\pi(X)=\pi(x_1,\cdots,x_m)$，$X\in\mathscr{X}\subseteq\mathbb{R}^m$ 为一个密度．可以构造一个以 $\pi(X)$ 为平衡分布的不可约 Markov 链．记 $\pi(x_i\mid x_j, j\neq i)$ 为在其他分量给定下，每个分量 x_i 的全条件分布密度，则此 Markov 链的转移密度为

$$\mathrm{P}(X,Y)=\prod_{l=1}^{m}\pi(y_l\mid x_j,j>l,y_j,j<l),\ X,Y\in\mathscr{X},$$

它容许在某个时间生成一个一维随机变量．继续做 m 次，以从一个 m 维随机向量转移到下一个时刻的另一个随机向量．重复这个手续，便生成了一系列 m 维随机向量 $X^{(0)},X^{(1)},\cdots,X^{(t)},\cdots$，它们可以作为具有转移概率 $\mathrm{P}(X,Y)$ 的 Markov 链 X 的轨道现实．

众所周知，在不十分强的条件下，对于一个 π-可积函数 f 有下式：

2000-07-06 收稿

* 国家自然科学基金（批准号：19970120，19971005）和教育部博士点基金及国家“八六三”高技术计划资助项目

$$X^{(t)} \xrightarrow[t \to \infty]{d} X \sim \pi(X);$$
$$\frac{1}{t}\sum_{i=1}^{t} f(X^{(i)}) \xrightarrow[t \to \infty]{} E_\pi\{f(X)\}, \text{ a.e.} \tag{1}$$

对于一个充分大的时刻 t，$X^{(t)}$可以看成为 $\pi(X)$的随机样本，而 $f(X^{(i)})$在 Markov 链的一个轨道上的平均，则是 Markov 链的这个函数的期望的近似．这就是 Gibbs 采样，其收敛性在文献中已有大量的讨论．例如可参见文献[3～10]．可是，如何在实践中判定是否 t 已经达到了充分大呢？我们的想法十分简单，即，首先把 Gibbs 采样的转移函数对称化，以使 Markov 链变为可逆；其次，一旦发现 Markov 链在某个陷阱区达到“局部遍历”就停止；第三，用简单模拟计算 Markov 链落在不同峰区的分布的比值；最后，以上述比值拒绝样本，以得到目标分布．

1　构造可逆 Markov 链

在 MCMC 算法的实际使用中，经过并行地运行链的多个轨道后，如果发现输出并不相容，则必须对于使用 Gibbs 采样的推理十分谨慎．通常这显示了平衡分布 $\pi(X)$具有多峰性质．假定状态空间可以分成“中心”远为分离的两个子集 B_1 和 B_2，并且使得此 Markov 链很难从其中的一个集合转移到另一个集合(转移的概率非常小)，我们建议用以下的程式．

1.1　可逆链在不变分布的一个峰区中的限制

为了叙述方便，我们用离散状态空间情形来阐述构思．一个具有平衡分布 $\pi(X)$的 Markov 链是可逆的，如果它的转移概率$(\mathrm{P}(X,Y))$满足

$$\pi(X)\mathrm{P}(X,Y) = \pi(Y)\mathrm{P}(Y,X),\ X,Y \in \mathscr{X}. \tag{2}$$

熟知，对于一个可逆的 Markov 链，可以把它限制在状态空间的某个子集 B_i 上，使它在 B_i 上仍以 $\pi(X)$为它的不变分布．记

$$\mathrm{P}_i(X,Y) = \begin{cases} \mathrm{P}(X,Y), & Y \neq X,\ X,Y \in B_i, \\ \mathrm{P}(X,X) + \sum\limits_{Y \notin B_i} \mathrm{P}(X,Y), & Y = X,\ X \in B_i \end{cases} \tag{3}$$

和

$$\pi_i(X) = \frac{\pi(X)}{\pi(B_i)},\quad X \in B_i,$$

此处

$$\pi(B_i) = \sum_{X \in B_i} \pi(X).$$

那么 $\pi_i(X)$是$\{\mathrm{P}_i(X,Y),\ X,Y \in B_i\}$的一个概率不变分布．

在两个峰区 B_1 和 B_2 的情形，如果从 B_1 到 B_2 和从 B_2 到 B_1 的概率都很小，则此链常被套在 B_1 或 B_2 中．所以，如从 B_i 中的点出发运行这个链，则在它离开 B_i 前与运行由$(\mathrm{P}_i(X,Y))$ $(i=1,2)$决定的 Markov 链一样．于是此 Markov 链看起来好像在 B_1 与 B_2 中平稳(在实际计算中的“平稳”，意即此链看起来在某区域平稳地振动)，而样本可近似地设想成分别采自 $\pi_i(X)(i=1,2)$．这样我们在只差一个常数比例因子下，得到了来自 $\pi(X)$的样本．而转移为$(\mathrm{P}_i(X,Y))(i=1,2,\cdots)$的 Markov 链趋于平稳的速度要比转移为$(\mathrm{P}(X,Y))$的 Markov 链快得多．

1.2 构造一个可逆的 Markov 链

利用对称推理，可以把任意一个概率转移(P(X,Y))代以用如下的一个可逆的概率转移：

$$\mathrm{P}^*(X,Y)=\frac{1}{2}(\mathrm{P}(X,Y)+\mathrm{P}^-(X,Y)),$$

它们有相同的平衡密度 $\pi(x)$，此处

$$\mathrm{P}^-(X,Y)=\frac{\pi(Y)\mathrm{P}(Y,X)}{\pi(X)}.$$

后一个模型的优点在于($\mathrm{P}^*(X,Y)$)保留了 Gibbs 性质，即在任意时刻只生成一个分量的性质．幸运的是 $\mathrm{P}^-(X,Y)$也是一个 Gibbs 采样型的转移，它在固定的时刻生成一个一维随机变量，而它与(P(X,Y))的不同仅在于随机变量的次序相反．这一事实可由下述引理看出：

引理 1 $\mathrm{P}^-(X,Y)=\prod\limits_{l=1}^{m}\pi(y_l \mid y_j,j>l;\ x_k,k<l).$

证

$$\begin{aligned}
&\frac{\pi(Y)\mathrm{P}(Y,X)}{\pi(X)}\\
&=\frac{\pi(y_1,\cdots,y_m)\pi(x_1 \mid y_2,\cdots,y_m)\pi(x_2 \mid x_1,y_3,\cdots,y_m)\cdots\pi(x_m \mid x_1,\cdots,x_{m-1})}{\pi(x_1,\cdots,x_m)}\\
&=\frac{\pi(x_m \mid x_1,\cdots,x_{m-1})\pi(x_{m-1} \mid x_1,\cdots,x_{m-2},y_m)\cdots\pi(x_1 \mid y_2,\cdots,y_m)\pi(y_1,\cdots,y_m)}{\pi(x_1,\cdots,x_m)}\\
&=\frac{\pi(x_1,\cdots,x_m)\pi(x_1,\cdots,x_{m-1},y_m)\cdots\pi(x_1,y_2,\cdots,y_m)\pi(y_1,\cdots,y_m)}{\pi(x_1,\cdots,x_m)\pi(x_1,\cdots,x_{m-1})\pi(x_1,\cdots,x_{m-2},y_m)\cdots\pi(y_2,\cdots,y_m)}\\
&=\pi(y_m \mid x_1,\cdots,x_{m-1})\pi(y_{m-1} \mid x_1,\cdots,x_{m-2},y_m)\cdots\pi(y_1 \mid y_2,\cdots,y_m).
\end{aligned}$$

引理 1 告诉我们

$$\mathrm{P}^*(X,Y)=\frac{1}{2}\prod_{l=1}^{m}\pi(y_l \mid x_j,j>l,y_k,k<l)+\frac{1}{2}\prod_{l=1}^{m}\pi(y_l \mid y_j,j>l,x_k,k<l). \quad (4)$$

对于模拟来说，这几乎与原来对于 P(X,Y)的 Gibbs 采样差不多，只需抛一个钱币以决定生成一维随机变量的次序．如果钱币的面向上，则按 $y_1,\cdots,y_m$ 的次序从 $X=(x_1,\cdots,x_m)$出发，生成下一个随机向量 $Y=(y_1,\cdots,y_m)$，否则就按相反的次序 $y_m,y_{m-1},\cdots,y_1$ 生成 Y.

1.3 获取两个峰的比

令

$$\frac{\pi_i(X)}{\pi(X)}=c_i,\quad \forall X\in B_i,\quad i=1,2,$$

则有

$$\pi(B_i)=\frac{1}{c_i},\quad \sum_i\pi(B_i)=\sum_i\frac{1}{c_i}=1.$$

利用(2)式及可逆性，当 $\mathrm{P}(X,\tilde{X})>0$ 时，有

$$\frac{\pi(X)}{\pi(\tilde{X})}=\frac{\mathrm{P}(\tilde{X},X)}{\mathrm{P}(X,\tilde{X})},\quad X\in B_1,\tilde{X}\in B_2,$$

由此推出

$$\frac{c_1}{c_2}=\frac{\pi_1(X)}{\pi(X)}\Big/\frac{\pi_2(\tilde{X})}{\pi(\tilde{X})}=\frac{\pi_1(X)\mathrm{P}(X,\tilde{X})}{\pi_2(\tilde{X})\mathrm{P}(\tilde{X},X)}. \tag{5}$$

而当 $\mathrm{P}(X,\tilde{X})=0$ 时，令 $X=Y_1,Y_2,\cdots,Y_k=\tilde{X}$ 是从 X 到 $\tilde{X}$ 的一条路径，则有

$$\pi(X)\prod_{l=1}^{k-1}\mathrm{P}(Y_l,Y_{l+1})=\pi(\tilde{X})\prod_{l=1}^{k-1}\mathrm{P}(Y_{l+1},Y_l),$$

这就导致

$$\frac{c_1}{c_2}=\frac{\pi_1(X)\prod_{l=1}^{k-1}\mathrm{P}(Y_l,Y_{l+1})}{\pi_2(\tilde{X})\prod_{l=1}^{k-1}\mathrm{P}(Y_{l+1},Y_l)}. \tag{6}$$

因为 $\mathrm{P}(X,\tilde{X})$已知，所以比值 c_1/c_2 可由计算 $\pi_i(\cdot)$的值而确定.

在实算中，$\pi_i(X)$的估计 $\hat{\pi}_i(X)$可由下面的引理 2 从一个 $\mathrm{P}_i(X,\cdot)$ 随机变量的简单模拟得到，而此随机变量就是 $\mathrm{P}(X,\cdot)$在 $X\in B_i$ 上的限制.

引理 2　在连续情形，假设转移密度限制在 B_i 上满足 $\mathrm{P}_i(X,Y)>0$，$\forall X,Y\in B_i$，则

$$\pi_i(X)=\left(\mathrm{E}_i\left(\frac{1}{\mathrm{P}_i(\eta,X)}\right)\right)^{-1},$$

此处 E_i 记在概率 $\mathrm{P}_i(X,\cdot)$的期望，若记 η 为遵从这个分布的随机变量. 假设 $\eta^1,\cdots,\eta^n$ 为与此同分布的独立同分布序列，那么 $\pi_i(X)$可用

$$\hat{\pi}_i(X)=\left[\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\mathrm{P}_i(\eta^j,X)}\right]^{-1} \tag{7}$$

来模拟.

证　因为 $\mathrm{P}_i(X,Y)>0$，(2)式即

$$\pi_i(Y)=\frac{\pi_i(X)}{\mathrm{P}_i(Y,X)}\mathrm{P}_i(X,Y),$$

由此导出

$$\pi_i(X)\int_{B_i}\frac{1}{\mathrm{P}_i(Y,X)}\mathrm{P}_i(X,Y)\mathrm{d}Y=1,$$

于是

$$\pi_i(X)=\left[\int_{B_i}\frac{1}{\mathrm{P}_i(Y,X)}\mathrm{P}_i(X,Y)\mathrm{d}Y\right]^{-1}=\left(E_i\left(\frac{1}{\mathrm{P}_i(\eta,X)}\right)\right)^{-1}.$$

应用大数定律，$\pi_i(X)$可用 $1/P_i(Y,X)$在 Y 的样本上取平均来模拟. 即(7)式成立.

选取常出现的点 $X\in B_1$ 和 $\tilde{X}\in B_2$，则此比值可用

$$\left(\frac{c_1}{c_2}\right)^{\wedge}=\frac{\hat{\pi}_1(X)\mathrm{P}(X,\tilde{X})}{\hat{\pi}_2(\tilde{X})\mathrm{P}(\tilde{X},X)} \tag{8}$$

近似给出.

1.4　基本算法

这里列出可逆 Markov 链 Gibbs 采样的算法：

第 1 步　从样本空间选取一系列初值. 它们可以是均匀分布的，也可以是按某些已知条

件分布的.

第 2 步　从这些初值出发，运行多个 Markov 链，用多维观察值的“点团”的数值分析探索(EDA)来作出判定,尽量发现状态空间的各个峰区.

第 3 步　一旦发现峰区，就用聚类方法把状态空间划分成数个局部盆区 $B_1, B_2, \cdots, B_k$. 事实上，出现在第 2 步中的样本轨道的不同区域，就构造了盆区 $B_i(i=1,2,\cdots,k)$.

第 4 步　在每个盆区 B_i，按(4)式构造可逆 Markov 链,再判断在局部盆区的亚稳定性，并且得到在这些局部盆区中的样本.

第 5 步　在每个盆区 B_i 中，选取出现的相对频率较高的点 X_i，并用(7)式计算 $\hat{\pi}_i(X_i)$的值. 再用(8)式来估计两个盆区 B_1 和 $B_i(i=2,\cdots,k)$的比值$\left(\frac{c_1}{c_i}\right)^{\wedge}$. 进一步用

$$\sum_{i=1}^{k} \frac{1}{c_i} = 1,$$

就得到模拟值 $\hat{c}_1, \cdots, \hat{c}_k$.

第 6 步　对不同的盆区，用概率分布

$$\left(\frac{1}{\hat{c}_1}, \frac{1}{\hat{c}_2}, \cdots, \frac{1}{\hat{c}_k}\right)$$

作为接受样本，用

$$\pi(\hat{X}) = \sum_{i=1}^{k} \frac{1}{\hat{c}_i} \pi_i(\hat{X})$$

作为整体分布 $\pi(X)$的估计. $\mathrm{E}_\pi\{f(X)\}$的估计也可以用样本的混合平均给出.

注　对于 Bayes 后验分布 $\pi(X)$，在实际中常因先验分布和似然方法导致其主峰远为分离，从而可以以后验分布的知识，用决定性的爬山方法给出其位置. 一旦知道了位置,我们的算法就较为合适. 这个算法经一个高维混合正态分布检验后,结果很满意，但是由于篇幅所限，暂不把它列出.

2　二维 Ising 模型

通过可逆 Markov 链，我们得到了一个处理多峰问题的程式. 典型的例子出现于某些动力系统，在那里可以观察到亚稳态性质. 而且正如前面所述，当系统长时间地留在某个亚稳态集合中，Markov 链就被套在一个陷阱中，致使我们得不到整体信息. 二维随机 Ising 模型的某些亚稳定性质的研究，为此提供了一个清楚的图例. 本节将用在第 1 节中引入的程式，来模拟二维 Ising 模型，并分析输出.

二维 Ising 模型可建立如下:设 Λ 为 2 维环面格点，即带周期边界的$\{1,\cdots,N\} \times \{1,\cdots,N\}$. 对于每一个格点位置 $x \in \Lambda$ 赋予一个自旋 +1 或 −1，我们抽象地把此自旋记为 $\eta(x)$. 全体$\{\eta(x)\}_{x\in\Lambda}$记成$\eta$，称为一个组态. 定义

$$+\underline{1} = \{\eta(x) = +1,\ \forall\, x \in \Lambda\},$$
$$-\underline{1} = \{\eta(x) = -1,\ \forall\, x \in \Lambda\}.$$

记 S 为所有组态 η 的集合，即 $S = \{-1, +1\}^{\Lambda}$. 对于每一个组态 $\eta \in S$，安排一个 Hamilton 量

$$H(\eta)=-\frac{1}{2}\sum_{\langle x,y\rangle}\eta(x)\eta(y)-\frac{h}{2}\sum_{x\in\Lambda}\eta(x)-\frac{h}{2}N^2+N^2,$$

此处第 1 个求和遍取 Λ 的一对邻域对，并且把 $\langle x,y\rangle$ 与 $\langle y,x\rangle$ 看成相同，易见 $H(-\underline{1})=0$.

离散时间的时间 Ising 模型定义为 $S=\{-1,+1\}^{\Lambda}$ 上的 Markov 链，且具有转移概率

$$\mathrm{P}^{\beta}(\xi,\eta)=\begin{cases}N^{-2}[1+\exp(-\beta(H(\xi)-H(\eta)))]^{-1}, & 若\ \eta=\xi^{x},\ \forall x\in\Lambda,\\ 0, & 若\ \eta\neq\xi^{x},\ \forall x\in\Lambda,\ \eta\neq\xi,\\ N^{-2}\sum\limits_{x\in\Lambda}[1+\exp(\beta(H(\xi)-H(\xi^{x})))]^{-1}, & 若\ \eta=\xi,\end{cases}$$

此处 ξ^{x} 的自旋是在 ξ 中改变且只改变 x 处的自旋 $\xi(x)$，也就是

$$\xi^{x}(y)=\begin{cases}\xi(y), & 若\ y\neq x,\\ -\xi(y), & 若\ y=x,\end{cases}$$

β 是一个参数，在物理中被解释为温度的倒数，而 h 是外场的强度. 我们记随机 Ising 模型为 $\{X_n,n\geqslant 0\}$.

这个 Markov 链 $\{X_n,n\geqslant 0\}$ 是甚高维的. 事实上，每个状态 η 由在 $N\times N$ 个位置上的自旋确定. 而转移概率 $(\mathrm{P}^{\beta}(\xi,\eta))_{(\xi,\eta)}$ 蕴含在每个时刻只在一个位置上有转移，并蕴含该 Markov 链是可逆的. 此 Markov 链也是正常返的，因为状态空间是有限的，且所有状态是连通的. 但是，存在状态空间的盆区，使得 Markov 链长时间陷入某盆区. Ising 模型是典型的多峰问题，可参见文献[2]. 其状态空间 S 有一个多层结构，而对一个固定的水平我们对它的吸引子及其盆区均有清楚的图像. 由此，能够确定它在状态空间中存在多少个峰区. 因此，可限制 Markov 链于每一个盆区生成样本，然后估计分布在两个盆区的比值.

在简单情形，假定 $\beta=2.0$，$N=5$ 和 $h=0$. 给定在状态空间中均匀分布的初始状态，并构造许多并行链. 计算显示，该链从任意组态出发，很快就落入 $+\underline{1}$ 或 $-\underline{1}$，随后就在此二状态附近转移. 我们可以从组态的形状看出 Markov 链的转移特性.

运行步数：30 000

初态

$$\begin{array}{ccccc}+&+&-&-&+\\+&-&-&+&-\\-&-&-&-&+\\+&+&+&+&+\\-&-&+&+&-\end{array}\qquad \text{Harmilton 量：28.00}$$

$$\downarrow$$

$$\begin{array}{ccccc}+&+&+&+&-\\+&+&+&+&+\\+&+&+&+&+\\+&+&+&+&+\\+&+&+&+&+\end{array}\qquad \text{Harmilton 量：4.00　停留步数：81}$$

$$\downarrow$$

$$\begin{array}{ccccc} + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \end{array}$$

Harmilton 量：0　停留步数：29 644

运行步数：30 000

初态

$$\begin{array}{ccccc} + & + & - & - & - \\ - & + & - & - & - \\ - & - & + & - & + \\ - & + & + & + & + \\ + & + & + & - & + \end{array}$$

Harmilton 量：22.00

$$\downarrow$$

$$\begin{array}{ccccc} - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & + \\ - & - & - & - & - \end{array}$$

Harmilton 量：4.00　停留步数：12

$$\downarrow$$

$$\begin{array}{ccccc} - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{array}$$

Harmilton 量：0　停留步数：28 602

本文由模拟得到两个盆区，它们分别集中在 $+\underline{1}$ 与 $-\underline{1}$ 附近．在 $n=30\,000$ 步内此链并未离开它的盆区．因此可以只在这些局部区域内生成样本．在一条实际轨道运行此链更长的时间(步数)，在 $n=1\,759\,052$ 步时，状态 $+\underline{1}$ 和 $-\underline{1}$ 就连通了．但是在 $+\underline{1}$ 的停留次数为 282 381 次，而在 $-\underline{1}$ 的停留次数则是 1 153 683 次，这里显示的两个状态间的比值是错误．结果列于下表：

步数	停留在 $+\underline{1}$	停留在 $-\underline{1}$	比值
30 000	29 644	0	–
30 000	0	28 602	–
1 759 052	282 381	1 153 683	0.245

用第 1 节的算法,假定 B_1 和 B_2 分别为组态 $+\underline{1}$ 和 $-\underline{1}$ 附近的由模拟建立的盆区．我们有(6)式：

$$\frac{c_1}{c_2}=\frac{\pi_1(+\underline{1})\prod_{l=0}^{k-1}\mathrm{P}^{\beta}(\xi_l,\xi_{l+1})}{\pi_2(-\underline{1})\prod_{l=0}^{k-1}\mathrm{P}^{\beta}(\xi_{l+1},\xi_l)},$$

此处 $+\underline{1}=\xi_0,\xi_1,\cdots,\xi_k=-\underline{1}$ 是一条由 $+\underline{1}$ 到 $-\underline{1}$ 的路径．由形式 $(\mathrm{P}^{\beta}(\xi,\eta))_{(\xi,\eta)}$ 引出

$$\frac{c_1}{c_2}=\frac{\pi_1(+\underline{1})}{\pi_2(-\underline{1})}\mathrm{e}^{\beta(H(+\underline{1})-H(-\underline{1}))}.$$

而估计 $\hat{\pi}_1(+\underline{1})$ 和 $\hat{\pi}_2(-\underline{1})$ 都由频率算出，并由此导出

$$\frac{\hat{\pi}_1(+\underline{1})}{\hat{\pi}_2(-\underline{1})}=1.03.$$

进一步知道 $H(-\underline{1})=H(+\underline{1})=0$，因而其估值为

$$\left(\frac{c_1}{c_2}\right)^{\wedge}=\frac{\hat{\pi}_1(+\underline{1})}{\hat{\pi}_2(-\underline{1})}=1.03.$$

这说明从 B_1 到 B_2 的样本接受比值差不多是 1．这个结论与理论计算相符．事实上，Ising 模型的不变分布是 Gibbs 分布

$$\pi(\xi)=\frac{\mathrm{e}^{-\beta H(\xi)}}{Z_\beta},$$

此处 $Z_\beta=\sum_{\eta\in S}\mathrm{e}^{-\beta H(\eta)}$ 为配分函数．易知 $+\underline{1}$ 和 $-\underline{1}$ 都是 Hamilton 量的最低点，即它们占有最大的概率．因为选取 $h=0$，$+\underline{1}$ 和 $-\underline{1}$，它们对应的盆区 B_1 和 B_2 是对称的．B_1 和 B_2 在状态空间的比值恰为 1．

参　考　文　献

1 Besag J, Green P J. Spatial statistics and Bayesian computation. J R Statist Soc B, 1993, 55: 25 ~ 37

2 Chen D, Feng J, Qian M P. The metastable behavior of the two dimensional Ising model. In: Ma Z M, Röckner M, Yan J A, eds. Dirichlet Forms and Stochastic Processes. Berlin: De Guiter, 1995. 73 ~ 86

3 Gelman A, Rubin D B. Inference from iterative simulation using multiple sequences. Statist Soc, 1992, 7(4): 457 ~ 511

4 Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo J M, Berger J O, Dawid A P, et al, eds. Bayesian Statistics 4. Oxford: Oxford University Press, 1992. 169 ~ 193

5 Schervish M J, Carlin B P. On the convergence of successive substitution sampling. J Comp Graph Statist, 1992, 1(2): 111 ~ 127

6 Roberts G O, Polson N G. On the geometric convergence of the Gibbs sampler. J R Statist Soc B, 1994, 56(2): 377 ~ 384

7 Tierney L. Markov chains for exploring posterior distributions. Ann Stat, 1994, 22(4): 1 701 ~ 1 728

8 Diaconis P, Stroock D. Geometric bounds for eigenvalues of Markov chains. Ann Appl Prob, 1991, 1(1): 36 ~ 61

9 Liu J, Wong W H, Kong A. Covariance structure and convergence rate of the Gibbs sampler with various scan. J R Statist Soc B, 1995, 57(1): 157 ~ 169

10 Rosenthal J S. Rates of convergence for Gibbs sampling for variance component models. Ann of Statist 1995, 23(2): 740 ~ 761