第24章：模拟方法

- 为什么需要Monte Carlo方法
- Monte Carlo 方法
  - 产生独立样本
    - 基本方法：概率积分变换（第一部分已讲）
    - 接受—拒绝（舍选）采样
    - 重要性采样
  - 产生相关样本：Markov Chain Monte Carlo
    - Metropolis-Hastings算法
    - Gibbs Sampler

MCMC：Markov Chain Monte Carlo

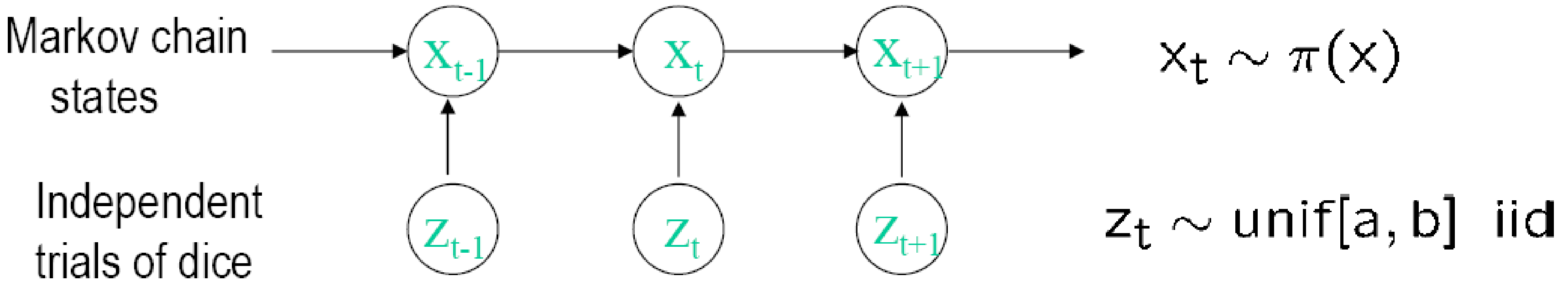
- 重要性采样和接受—拒绝采样都只$q(x)$与$\pi(x)$很相近似时才表现很好

- 在高维空间问题中，标准的采样方法会失败：
  - 接受—拒绝采样：维数增高时，拒绝率→100%
  - 重要性采样：大多数的样本权重→ 0

- 对高维复杂问题，用马尔科夫链（Markov Chain）产生一些列相关样本，实现对分布的采样

MCMC

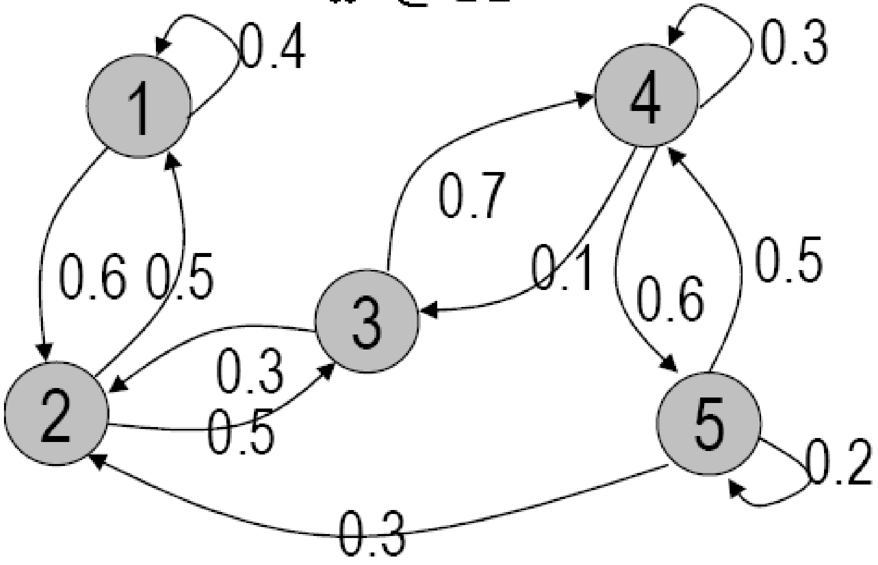
- MCMC：一种利用一定范围内的均匀分布的随机数，对高维空间概率进行采样的通用技术。
- 基本思想：设计一个马尔科夫链，使得其稳定概率为目标分布 $\pi(x)$



马尔科夫链：A Toy Example

- 假设一个孤岛上有5个家庭，他们的总财产为1。令状态表示每年的财产分布，每个家庭与其它一些家庭发生贸易，如家庭1将其收入的60%用于向家庭2购买物品…
- 问：经过一些年后，各家庭的财产情况？

$x \in \Omega$



马尔可夫链

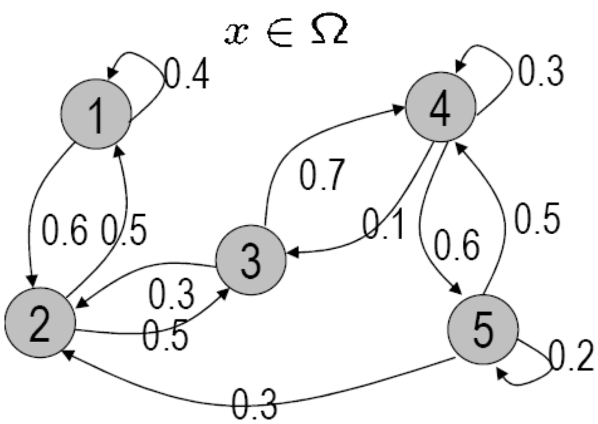
$$(\Omega, \mu_0, P)$$

状态空间　　初始分布　　转移核

$$P = \begin{pmatrix} 0.4 & 0.6 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.0 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.3 & 0.0 & 0.7 & 0.0 \\ 0.0 & 0.0 & 0.1 & 0.3 & 0.6 \\ 0.0 & 0.3 & 0.0 & 0.5 & 0.2 \end{pmatrix}$$

$$\sum_i p_{ij} = 1$$

稳定分布



$$\lim_{n\to\infty} \mu_0 P^n = \pi$$

稳定分布

year			
1	1.0 0.0 0.0 0.0 0.0		0.0 0.0 1.0 0.0 0.0
2	0.4 0.6 0.0 0.0 0.0		0.0 0.3 0.0 0.7 0.0
3	0.46 0.24 0.30 0.0 0.0		0.15 0.0 0.22 0.21 0.42
4	…		…
5			
6	0.23 0.21 0.16 0.21 0.17		0.17 0.16 0.16 0.26 0.25
	0.17 0.20 0.13 0.28 0.21		0.17 0.20 0.13 0.28 0.21

MCMC

- 从任意的初始状态 X_0 开始，通过利用一个转换核，产生一个各态历经的链 X_t，其稳定分布为我们感兴趣的目标分布 $\pi(x)$
  - 保证 X_t 的分布会收敛到来自分布 π 的随机变量
  - 对一个足够大的 T_0 ，X_{T_0} 可以认为是来自分布
  - 从分布 π 产生一系列相关的样本 $X_{T_0}, X_{T_0+1}, \ldots,$ 可以对很多应用足够近似

MCMC

- 为了实现对目标分布的采样，设计的马尔科夫链必须具有如下性质：
  - 不可约性：可以从任意状态到达另外任意状态
  - 循环性：可以无限次访问任意状态
  - 各态历经性：忘记初始状态

- 同时，马尔科夫链的收敛速度应尽可能快

马尔科夫链的设计

- 令$|\Omega| = r$表示系统的状态的数目，则
- 未知数的数目，即P中的元素数目为r^2
- 约束：$2r$个

$$\begin{cases} \pi P = \pi \\ P \times 1 = 1 \end{cases}$$

- 所以设计满足条件马尔可夫链是一个病态问题，自由度很大。

- 细致平衡是系统平衡的充分条件。

细致平衡

- 系统总平衡（稳定分布为π）：

$$\sum_x \pi(x)P(x,y)=\pi(y),\forall y\in\Omega$$

$$\boxed{\pi P=\pi}$$

- 其中 $P(x,y)$表示从状态x到状态y的转移概率

$$P(x,y)=\mathbb{P}\left(X_{t+1}=y \mid X_t=x\right)$$

- 细致平衡（detail balance equation）：

$$\pi(x)P(x,y)=\pi(y)P(y,x)$$

- 如果系统细致平衡，则一定总平衡。

细致平衡

$$\pi(x)P(x,y)=\pi(y)P(y,x)$$

$$\begin{aligned}
\sum_x \pi(x)P(x,y) &= \sum_x \pi(y)P(y,x) \\
&= \pi(y)\underbrace{\sum_x P(y,x)}_{=1} \\
&= \pi(y)
\end{aligned}$$

Metropolis-Hastings方法

- Top 10 algorithms！【top ten】
- published in 1953 [Metropolis]
  - Metropolis et al. (1953), "Equation of State calculation by fast computing machines"
- Extended in 1970 [Hastings]
  - W. K. Hastings (1970), "Monte Carlo Sampling methods using Markov chain and their application"

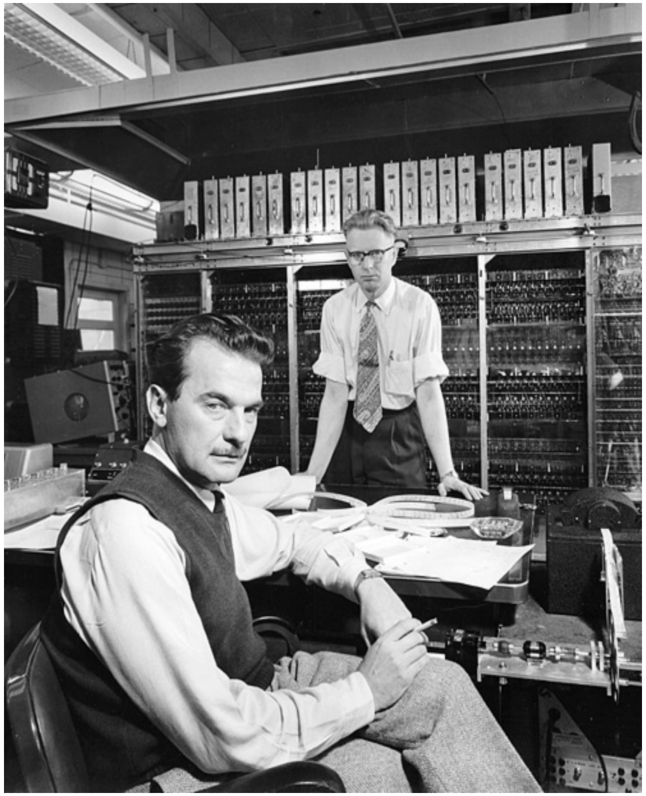
Metropolis-Hastings方法

- 基本思想：设计一个细致平衡的Markov Chain

  $$\pi(x)P\left(x, y\right) = \pi(y)P\left(y, x\right)$$

- Metropolis: 建议分布对称
- Hastings: 将建议分布扩展为一般形式

MANIAC the Computer and the Man



Seated is Nick Metropolis, the background is the MANIAC vacuum tube computer

Metropolis-Hastings方法

$$\boxed{\pi(x)P(x,y)=\pi(y)P(y,x)}$$

- 细致平衡的Markov Chain

$$\begin{cases} \forall y \neq x, & \underbrace{P(x,y)}_{\text{transition kernel}} = \underbrace{q(y|x)}_{\text{proposal}} \cdot \underbrace{\alpha(x,y)}_{\text{accep tan ce rate}} \\ y = x, & P(x,\mathrm{x}) = 1 - \sum_{y \neq x} P(x,y) \end{cases}$$

其中 $\alpha(x,y)=\min\left(1,\dfrac{q(x|y)}{q(y|x)}\dfrac{\pi(y)}{\pi(x)}\right)$

- Metropolis算法中 建议分布对称, $q(x|y)=q(y|x)$ ，所以

$$\alpha(x,y)=\min\left(1,\frac{\pi(y)}{\pi(x)}\right)$$

- 转移矩阵非常稀疏，一个状态只能到少数几个状态，通常不保存矩阵，而是函数形式或者是一个搜索操作

Metropolis-Hastings方法

- Metropolis-Hastings方法满足细致平衡：

$$
\begin{aligned}
\pi(x)P(x,y) &= \pi(x)q(y\mid x)\alpha(x,y) \\
&= \pi(x)q(y\mid x)\min\left(1,\frac{q(x\mid y)}{q(y\mid x)}\frac{\pi(y)}{\pi(x)}\right) \\
&= \min\left(\pi(x)q(y\mid x),\pi(y)q(x\mid y)\right) \qquad \text{（}x,y\text{对称）} \\
&= \pi(y)P(y,x)
\end{aligned}
$$

$$
\boxed{\pi(x)P(x,y) = \pi(y)P(y,x)}
$$

- 所以该马尔可夫链能收敛到目标分布 π

Metropolis-Hastings算法

- 给定在当前状态 $X_t = x$，
- 1、 产生一个建议 $Y_t \sim q(y \mid x)$
- 2、$X_{t+1} = \begin{cases} Y_t & 以概率\,\alpha(x, Y_t)，产生新样本 \\ X_t & 以概率\,1-\alpha(x, Y_t)，等待 \end{cases}$
- 其中 $\alpha(x, y) = \min\left(1, \dfrac{q(x \mid y)}{q(y \mid x)} \dfrac{\pi(y)}{\pi(x)}\right)$
- 如果 $q(x \mid y) / q(y \mid x)$ 比例不对称，则拒绝概率很大，马尔科夫处于等待状态

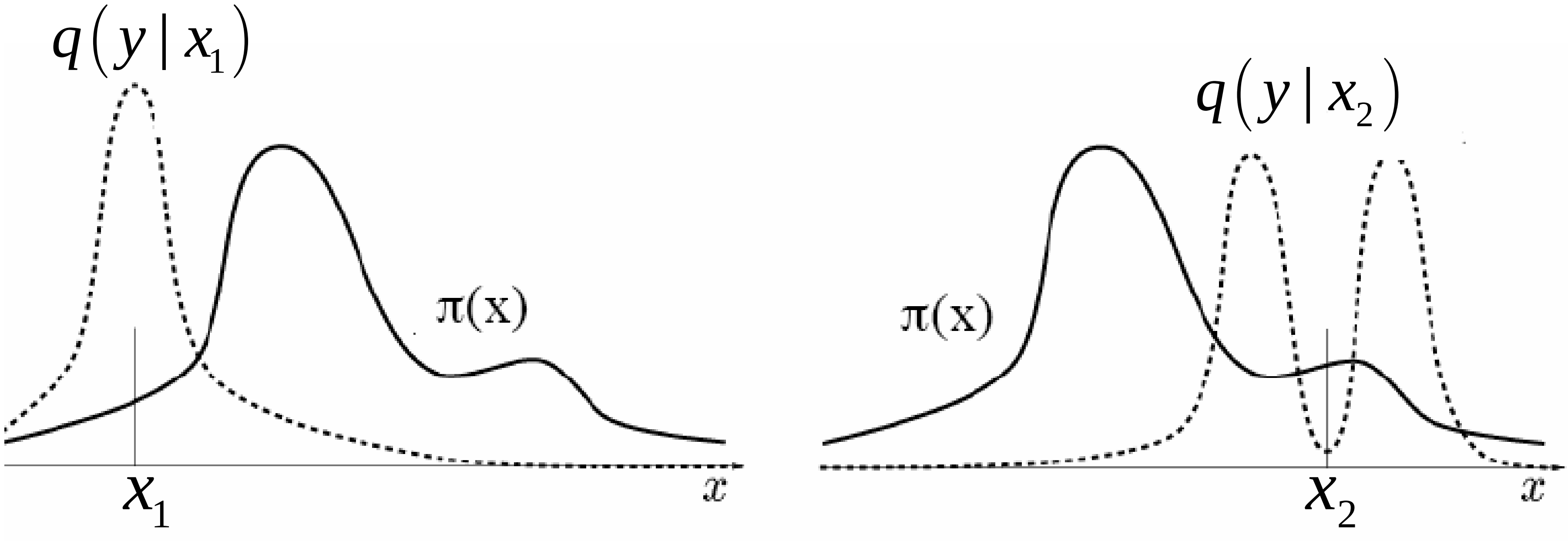
Metropolis-Hastings算法

等价于

- 给定在当前状态 $X_t = x$，
- 1、产生 $Y_t \sim q(y \mid x)$
- 2、$U \sim Uniform[0,1]$
- 3、$X_{t+1} = \begin{cases} Y_t & U \le \alpha(x, Y_t) \\ X_t & \text{其它} \end{cases}$
- 其中 $\alpha(x,y) = \min\left(1, \dfrac{q(x \mid y)}{q(y \mid x)} \dfrac{\pi(y)}{\pi(x)}\right)$

Metropolis-Hastings方法

- 建议分布 $q(y|x_t)$ 取决于当前状态 x_t
  - 可为任意固定的分布
  - 不必与 $\pi(x)$ 相似



例1：能量下降

- 令 $\pi(x)=\frac{1}{Z}\exp[-E(x)],\ \pi(y)=\frac{1}{Z}\exp[-E(y)]$

  $q(y|x)=q(x|y)=uniform[0,1]$

  $\boxed{\alpha(x,y)=\min\left(1,\frac{q(x|y)}{q(y|x)}\frac{\pi(y)}{\pi(x)}\right)}$

- 则

$$\alpha(x,y)=\min\left(1,\frac{1}{\exp[-(E(x)-E(y))]}\right)=\min(1,\exp[\Delta E])$$

$$=\begin{cases}1 & \Delta E>0 \\ \exp[-\Delta E] & \Delta E\le 0\end{cases}$$

，能量下降，接受

- 其中 $\Delta E=E(x)-E(y)$

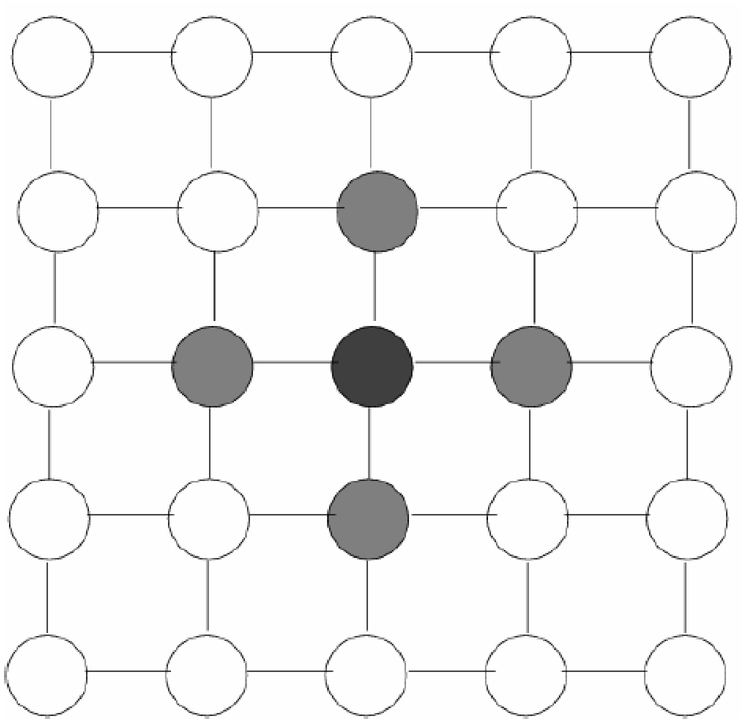
Ising model （1）

- 2D网格上的{-1,+1}随机变量
- 相邻变量相关

- 某个配置x的概率

$$\pi(x) = \frac{1}{Z} \exp\left(-\frac{E(x)}{kT} \right)$$

- 其中
  - Z：归一化常数
  - $E(x)$ ：能量
  - T：温度



Ising model （2）

- 例：Ising模型中 $\pi(x)=\frac{1}{Z}\exp\left(J\sum_{i\sim j}x_i x_j\right)$
  - 其中 $J=-1/kT=0.85$
  - $x_i \sim x_j$ 为相邻像素
  - N=256为每行/列的像素数
- 则 $\sum_{i\sim j}x_i x_j=2N(N-1)-2d_x$
  - 其中 d_x 为边缘的数目（相邻像素分别为-1和+1）
- 所以

$$\pi(x)\propto\exp(-2Jd_x)$$

Ising model （3）

- 用Metropolis方法采样为： $\boxed{\pi(x) = \frac{1}{Z}\exp\left(J\sum_{i\sim j} x_i x_j\right)}$
    - 从当前配置开始：

$$x = \left(x_1, x_2, \ldots, x_{N-1}, x_N, x_{N+1}, \ldots x_{N^2}\right)$$

    - 随机选择一个像素，如第N个像素
    - 得到一个新的建议状态

$$x' = \left(x_1, x_2, \ldots, x_{N-1}, -x_N, x_{N+1}, \ldots x_{N^2}\right)$$

    - 则接受状态x'的概率为

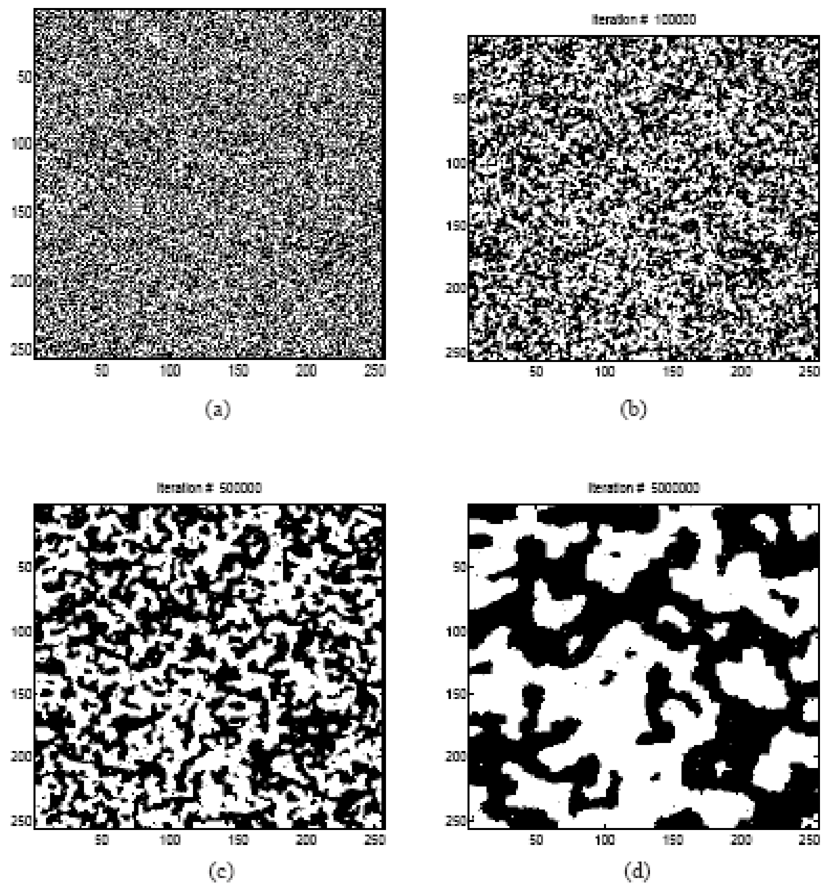
$$\alpha(x,y) = \min\left(1, \frac{\pi(x')}{\pi(x)}\right)$$

$$= \exp\left(-2Jx_j^t\left(x_{j-1}^t + x_{j+1}^t\right)\right), \text{ for } j \neq 1 \text{ or } d$$

Ising model （4）

- 上述例子的一个结果如图所示
  - 参考代码见Ising_ Metropolis.m

随机初始状态



迭代100000次

迭代500000次

迭代5000000次

Random walk Metropolis–Hastings

- 假设目标分布$\pi(x)$定义在d维欧氏空间$\mathbb{R}^d$上。则对当前配置X_t一个自然的扰动是增加一个随机误差，即

$$Y_t = X_t + \varepsilon_t$$

- 其中$\varepsilon_t \sim g$，对不同时间的t为IID，与X_t无关。

- 建议分布的形式为$g(y-x)$，如果g对称，则该马尔科夫链为随机游走。g通常取球形高斯函数。
  - 若没有Metropolis–Hastings中的拒绝规则，则该马尔科夫链会漂移到无穷远，不会再回来$(d \geq 3)$。

- 注意：x的取值范围需为整个实数轴，否则做变换，使得换后的取值范围需为整个实数轴

Random walk Metropolis–Hastings

- 给定在当前状态 $X_t = x$，
- 1、产生 $Y_t \sim g(y-x)$
- 2、$X_{t+1} = \begin{cases} Y_t & \text{以概率 } \alpha(x, Y_t)\text{，产生新样本} \\ X_t & \text{以概率 } 1-\alpha(x, Y_t)\text{，等待} \end{cases}$
- 其中 $\alpha(x, y) = \min\left(1, \dfrac{\pi(y)}{\pi(x)}\right)$

例： Random walk Metropolis–Hastings

- 24.10例：对下述Cauchy分布采样

$$f(x)=\frac{1}{\pi}\frac{1}{1+x^2}$$

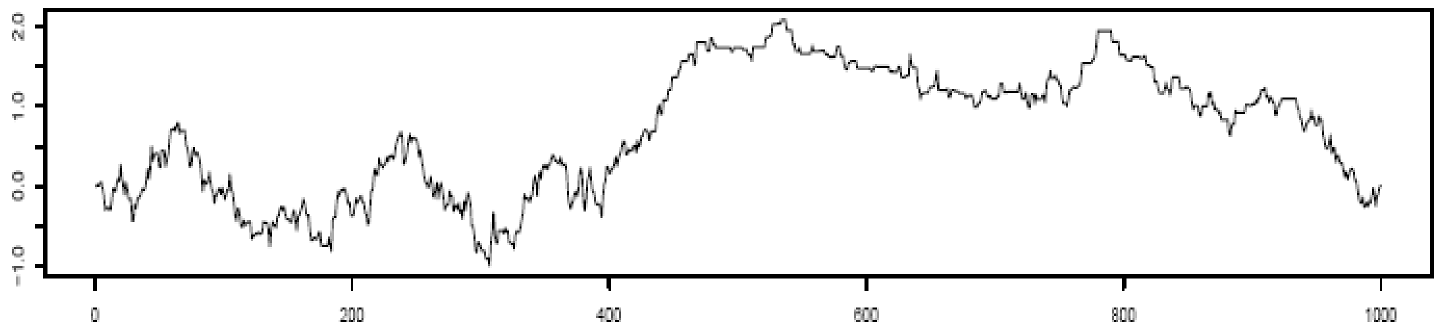
- 采用Random walk Metropolis–Hastings方法
  - 建议分布为 $g(y-x)=N\left(0,b^2\right)$
  - 所以接受率为

$$\alpha(x,y)=\min\left(1,\frac{f(y)}{f(x)}\right)=\min\left(1,\frac{1+x^2}{1+y^2}\right)$$

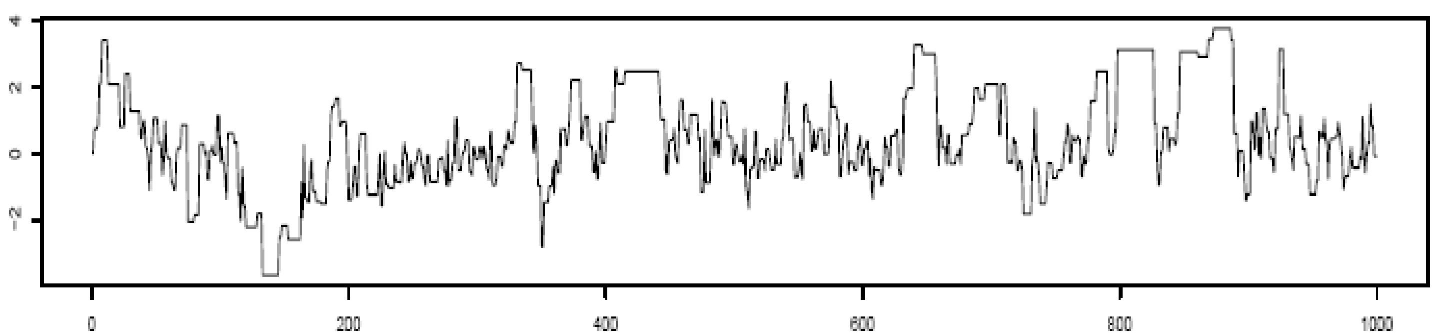
例： Random walk Metropolis–Hastings (续)

- 24.10例(续)： 不同b对应的Metropolis链

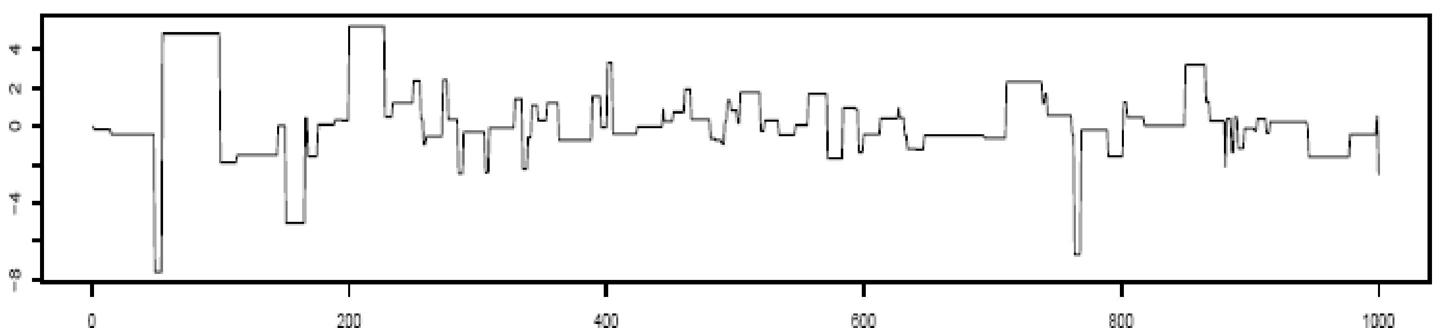
b=0.1



b=1



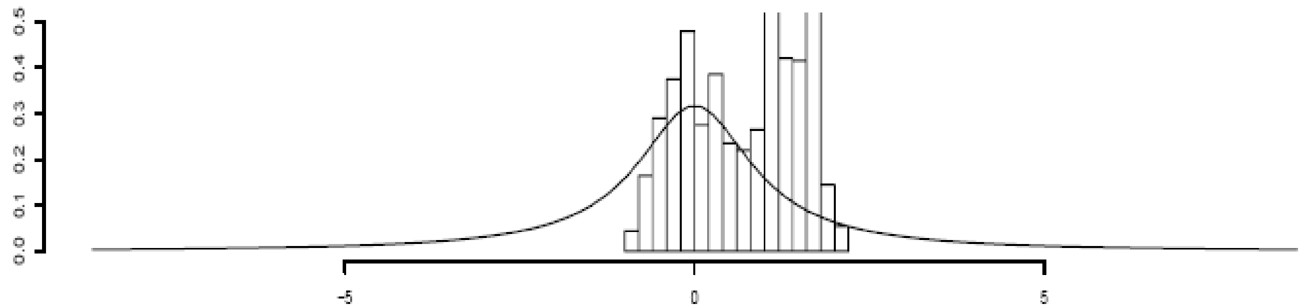
b=10



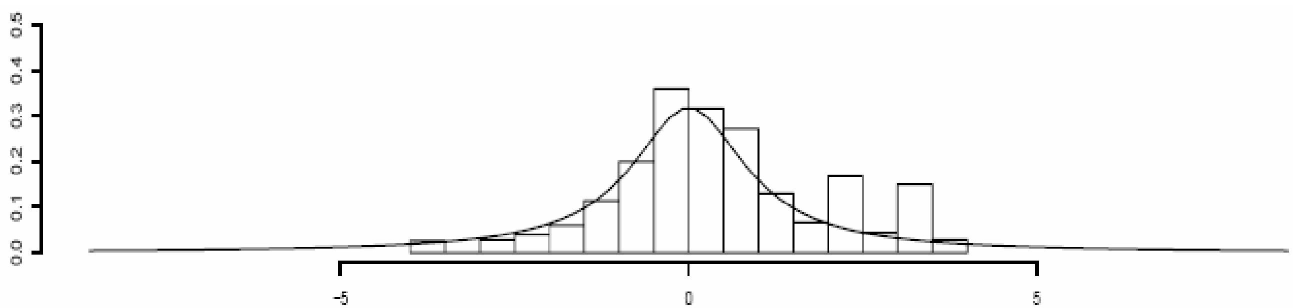
例： Random walk Metropolis–Hastings (续)

- 24.10例(续)： 不同b对应的Metropolis链得到的直方图

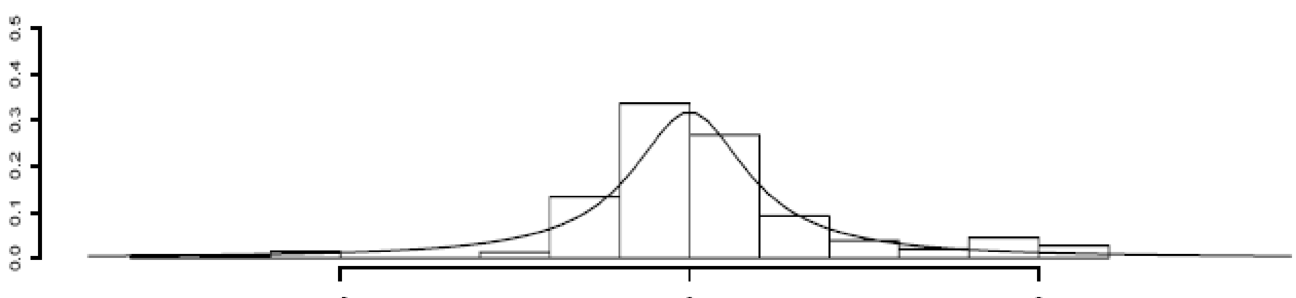
b=0.1



b=1



b=10



例3：Independence Metropolis–Hastings

- 建议分布为一个固定的分布g
- 通常， g应为π的一个近似
- 接受概率为

$$\alpha(x,y)=\min\left(1,\frac{g(x)}{g(y)}\frac{\pi(y)}{\pi(x)}\right)$$

Gibbs sampling

- Metropolis–Hastings在高维问题中广泛使用
  - 比重要性采样和拒绝性采样表现更好
- 在很多应用中，建议通常为局部均匀移动
  - 虽然该方法在概念上和操作上都很简单，但Metropolis算法的通常不是很有效
    - 通常 $q\left(y \mid x\right)$ 相当窄/稀疏
      - 太宽：很多状态被拒绝，因而过程很慢
      - 太窄：随机游走(Random walk)需要很长时间才能采样整个状态空间，因而过程也很缓慢
- Gibbs sampling：对条件分布采样
  - 没有拒绝

Gibbs sampling

- 将高维采样转化为多个一维采样：对条件分布采样
- 在采样过程没有拒绝

- 对 $x \sim \pi(x) = \pi(x_1, \ldots, x_d)$ 的采样通过对

$$x_i \sim \pi(X_i \mid x_1, \ldots, x_{i\text{-}1}, x_{i+1}, \ldots, x_d), i = 1, \ldots, d$$

- 的采样实现

Gibbs sampling

- 给定在当前状态$x^{(t)}=\left(x_1^{(t)},\ldots,x_d^{(t)}\right)$
- 1、 产生$X_1^{(t+1)}\sim\sim\pi\left(X_1\mid x_2^{(t)},\ldots,x_d^{(t)}\right)$
- 2、产生$X_2^{(t+1)}\sim\sim\pi\left(X_2\mid x_1^{(t)},x_3^{(t)},\ldots,x_d^{(t)}\right)$

…

- d、产生$X_d^{(t+1)}\sim\sim\pi\left(X_d\mid x_1^{(t)},x_2^{(t)},\ldots,x_{d-1}^{(t)}\right)$

形式很简单

对MRF，

$$\pi\left(X_i\mid x_1,\ldots,x_{i\text{-}1},x_{i+1},\ldots,x_d\right)=\pi\left(X_i\mid \underline{x_{[-i]}}\right)=\boxed{\pi\left(X_i\mid \underline{x_c}\right)}$$

除i之外的所有节点　　i的邻居节点

Gibbs sampling

- 最后收敛到$\pi\left(x_1,...,x_d\right)$
- $\pi\mathbf{P}=\pi$
- 证明：用归纳法
  - 假设第 t 步满足不变方程，即 $\left(x_i^{(t)},x_{[-i]}^{(t)}\right)\sim\pi\left(x_i,x_{[-i]}\right)$
  - 那么 $x_{[-i]}^{t}$ 为 π 的边缘分布 $\pi\left(x_{[-i]}^{t}\right)$
  - 那么第 $t+1$ 步，选择第 i 个节点调整其状态

$$\pi\left(x_i^{t+1}\mid x_{[-i]}^{t}\right)\pi\left(x_{[-i]}^{t}\right)=\pi\left(x_i^{t+1},x_{[-i]}^{t}\right)$$

  - 仍然满足分布 π

Gibbs sampling

- Gibbs采样为Metropolis–Hastings的一个特例：
  - $x': x_{-i}$固定，x_i变化是的一个可能的取值

$$q(x' \mid x) = \pi(x_i \mid x_{-i})$$

- 根据贝叶斯公式，接受率为

$$\alpha = \frac{\pi(x')}{\pi(x)} \frac{q(x \mid x')}{q(x' \mid x)} = \frac{\pi(x')\pi(x_i \mid x'_{-i})}{\pi(x)\pi(x'_i \mid x_{-i})} \quad \text{建议分布的定义}$$

$$= \frac{\pi(x')\pi(x_i, x'_{-i})\pi(x_{-i})}{\pi(x)\pi(x_i, x_{-i})\pi(x'_{-i})} = \frac{\pi(x_{-i})}{\pi(x'_{-i})} = 1 \quad \text{条件概率的定义}$$

- 所以总是接受

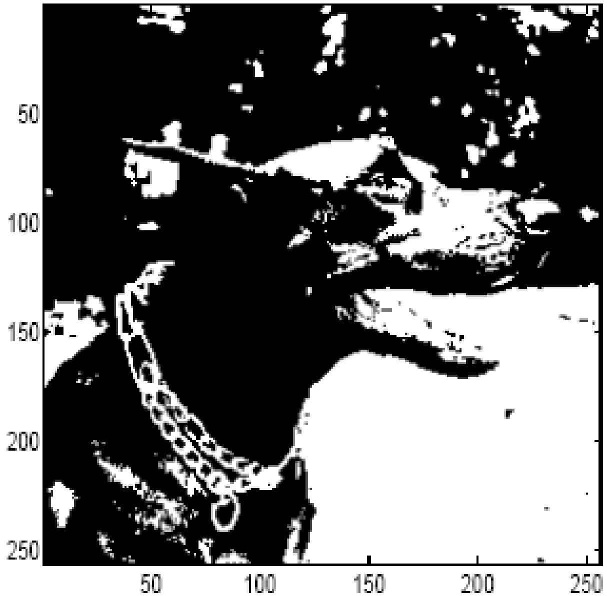
Ising model

- 例：Ising模型中 $\pi(x)=\frac{1}{Z}\exp\left(\beta\sum_{i\sim j}x_i x_j\right)$
- 其中 $\beta=-1/kT=0.85$
  - $x_i \sim x_j$ 为相邻像素
  - N=256为每行/列的像素数
- 则用Gibbs采样方法采样为：
  - 对每个像素n
  - 根据n的邻居m邻居，计算 $w_n=\sum_{m\sim n}x_m$
  - 定义两个概率

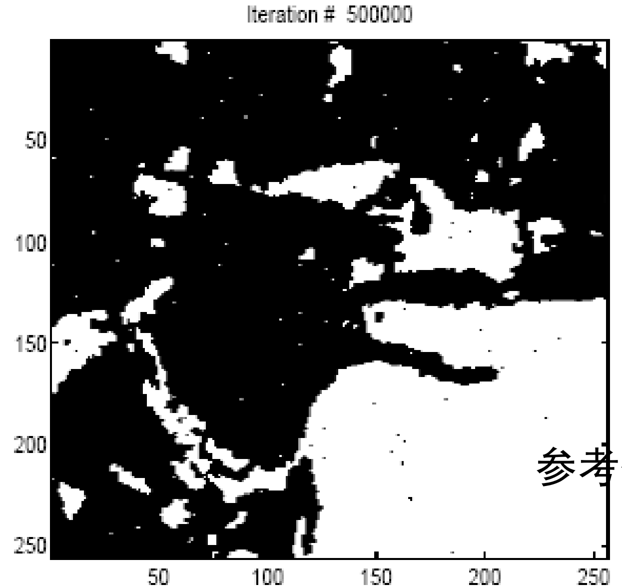
$$p^-=\frac{\exp(-Jw_n)}{\exp(-Jw_n)+\exp(Jw_n)},\quad p^+=1-p^-$$

  - 以概率p^-产生 $x_n^{new}=-1$，以概率p^+产生 $x_n^{new}=1$

Is

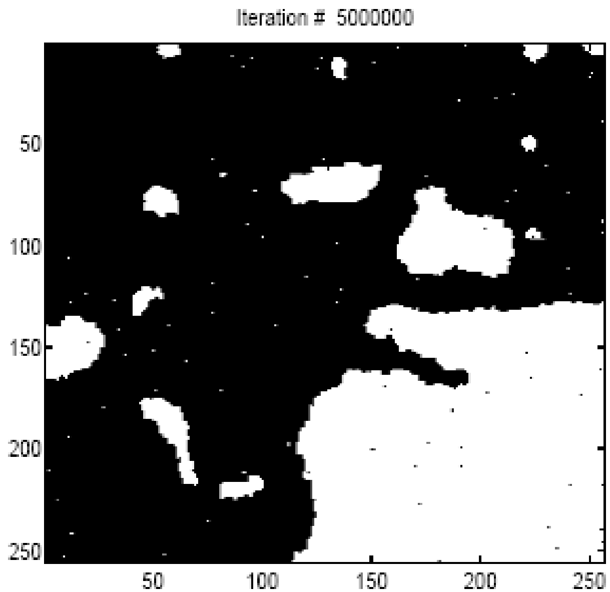


输入初始图像黑白Von Klaus

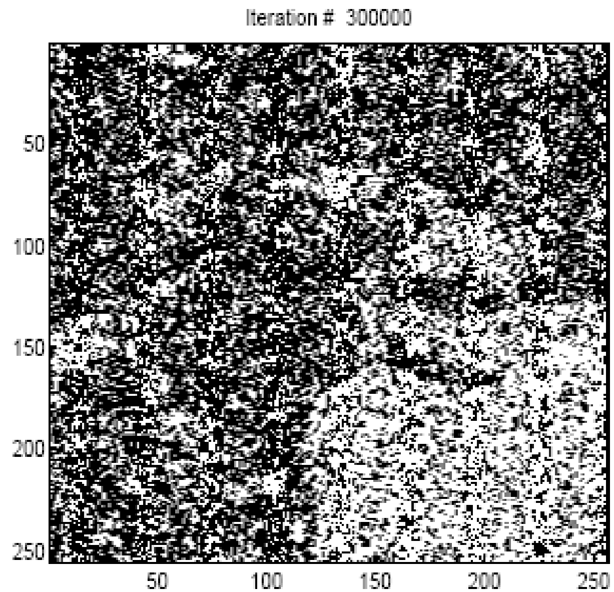


迭代500000次

参考代码见Ising_ Gibss.m



迭代5000000次

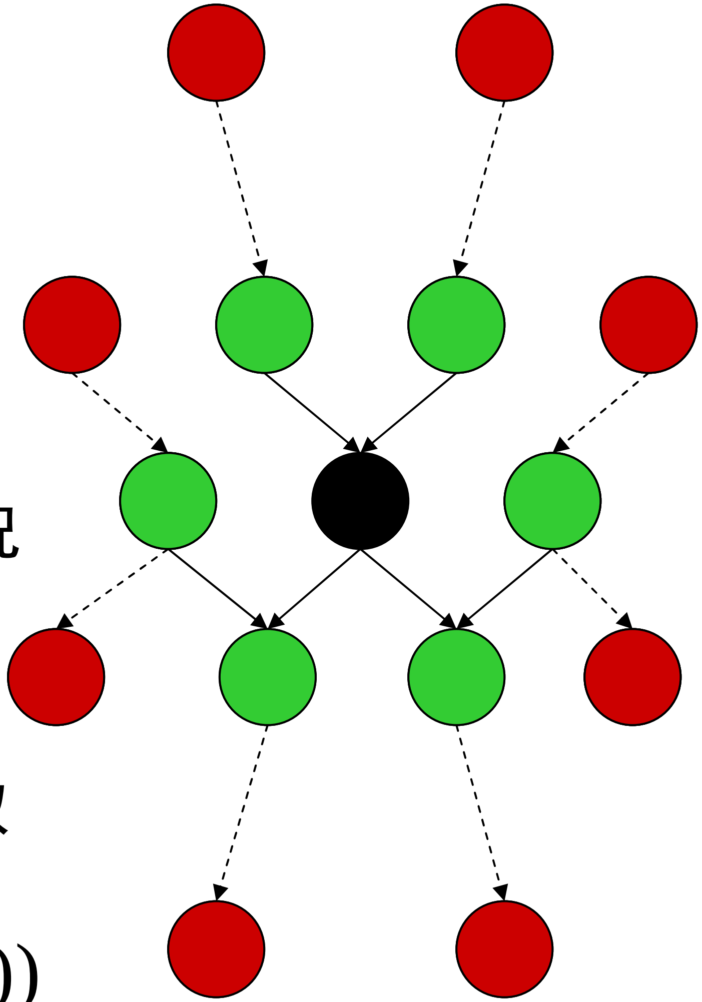


J =0.25，迭代300000次

Markov blanket

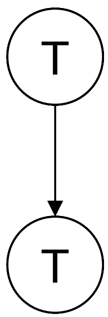
- 在概率图模型（如Bayesian 网络）中，联合分布被写成一些因子的乘积
- 在给定其父亲节点的情况下，变量和其非后继节点独立
- 在给定其*Markov blanket* 的情况下，变量与其他所有节点独立
- 因此，为了对每个节点采样，我们只需对其Markov blanket取条件

$$\pi(x_i \mid x_{[-i]}) = \pi(x_i \mid MB(x_i))$$



例：Gibbs sampling

- 考虑一个简单的2个变量的贝叶斯网络



	a	-a
	0.5	0.5

	b	-b
a	0.8	0.2
-a	0.2	0.8

	b	-b
a	1	1
-a	1	1

- 随机初始化
- 轮流对变量进行采样

Advanced Topics

- 模拟煺火：用于计算全局最优化
- Monte Carlo EM (stochastic E-step)
- 可逆跳转（Reversible jump）：用于模型选择
- Clustering
- Data-driven MCMC
- Adaptive proposal distributions
- …

Advanced Topics

- 怎样度量收敛性
  - First hitting time, mixing time
- 怎样诊断特定马尔科夫收敛性
  - Exact sampling

参考文献

- Jun S. Liu, *Monte Carlo Strategies in Scientific Computing. Springer*, 2005.
  - *www.china-pub.com*上有影印版，中文书名《**科学计算中的蒙特卡罗策略**》

- Christian P. Robert，*Monte Carlo Statistical Methods，Springer-Verlag， 2004.8*
  - ***http://www.ceremade.dauphine.fr/%7Exian/books.html***

- ICCV05 Course：Markov Chain Monte Carlo for Computer Vision
  - http://civs.stat.ucla.edu/MCMC/MCMC_tutorial

Reference links

- Java MCMC applet:
  - http://www.lbreyer.com/mcmc-webdemo.html