第五章 贝叶斯计算（MCMC）

Wang Shujia

Department of Statistics, School of Economics
Shenzhen University

Outline

1. 随机模拟介绍
    - 从蒲丰抛针问题说起
    - 函数积分的 Monte Carlo 法
    - 用 R 直接进行贝叶斯计算的一些技巧
    - 随机变量分布的直接抽样法
2. 马尔可夫链蒙特卡罗迭代法 (MCMC)

贝叶斯方法需要面对的计算问题

1. 获得后验分布 $p(\theta|y)$：$p(\theta|y) \propto \pi(\theta)f(x|\theta)$

贝叶斯方法需要面对的计算问题

1. 获得后验分布 $p(\theta|y)$：$p(\theta|y) \propto \pi(\theta)f(x|\theta)$
   - 抽取后验分布样本点，模拟该分布

贝叶斯方法需要面对的计算问题

1. 获得后验分布 $p(\theta|y)$：$p(\theta|y) \propto \pi(\theta)f(x|\theta)$
   - 抽取后验分布样本点，模拟该分布
2. 计算任意函数 $h(\theta)$ 对后验分布的平均

$$E[h(\theta)|y] = \int h(\theta)p(\theta|y)d\theta$$

贝叶斯方法需要面对的计算问题

1. 获得后验分布 $p(\theta|y)$：$p(\theta|y) \propto \pi(\theta)f(x|\theta)$
   - 抽取后验分布样本点，模拟该分布
2. 计算任意函数 $h(\theta)$ 对后验分布的平均

$$E[h(\theta)|y] = \int h(\theta)p(\theta|y)d\theta$$

   - 期望：$E(\theta|y) = \int \theta p(\theta|y)d\theta$, $(h(\theta) = \theta)$.

贝叶斯方法需要面对的计算问题

1. 获得后验分布 $p(\theta|y)$：$p(\theta|y) \propto \pi(\theta) f(x|\theta)$
   - 抽取后验分布样本点，模拟该分布
2. 计算任意函数 $h(\theta)$ 对后验分布的平均

$$E[h(\theta)|y] = \int h(\theta) p(\theta|y) d\theta$$

   - 期望：$E(\theta|y) = \int \theta p(\theta|y) d\theta$, $(h(\theta) = \theta)$.
   - 预测：$E[\tilde{y}|y] = \int f(\tilde{y}|\theta) p(\theta|y) d\theta$, $(h(\theta) = f(\tilde{y}|\theta))$.

贝叶斯方法需要面对的计算问题

1. 获得后验分布 $p(\theta|y)$：$p(\theta|y) \propto \pi(\theta)f(x|\theta)$
   - 抽取后验分布样本点，模拟该分布
2. 计算任意函数 $h(\theta)$ 对后验分布的平均

$$E[h(\theta)|y] = \int h(\theta)p(\theta|y)d\theta$$

   - 期望：$E(\theta|y) = \int \theta p(\theta|y)d\theta$, $(h(\theta) = \theta)$.
   - 预测：$E[\tilde{y}|y] = \int f(\tilde{y}|\theta)p(\theta|y)d\theta$, $(h(\theta) = f(\tilde{y}|\theta))$.
   - 概率：$P(\theta \in A) = \int_A p(\theta|y)d\theta = \int I_A(\theta)p(\theta|y)d\theta$, $(h(\theta) = I_A(\theta))$.

$$I_A(\theta) = \begin{cases} 1, & \text{if } \theta \in A \\ 0, & \text{if } \theta \notin A \end{cases}$$

如何从后验分布中抽取代表性样本?

- 若后验分布 $p(\theta|y)$ 为常用分布：直接抽取 iid 样本

如何从后验分布中抽取代表性样本?

- 若后验分布 $p(\theta|y)$ 为常用分布：直接抽取 iid 样本
  1. Inverse CDF Method

如何从后验分布中抽取代表性样本?

- 若后验分布 $p(\theta|y)$ 为常用分布：直接抽取 iid 样本
  1. Inverse CDF Method
  2. Rejection Sampling

如何从后验分布中抽取代表性样本?

- 若后验分布 $p(\theta|y)$ 为常用分布：直接抽取 iid 样本
  1. Inverse CDF Method
  2. Rejection Sampling
  3. Importance Sampling

如何从后验分布中抽取代表性样本?

- 若后验分布 $p(\theta|y)$ 为常用分布：直接抽取 iid 样本
  1. Inverse CDF Method
  2. Rejection Sampling
  3. Importance Sampling
     - ★ 缺点：高维有困难

如何从后验分布中抽取代表性样本？

- 若后验分布 $p(\theta|y)$ 为常用分布：直接抽取 iid 样本
  1. Inverse CDF Method
  2. Rejection Sampling
  3. Importance Sampling
     - ★ 缺点：高维有困难
- 若后验分布 $p(\theta|y)$ 为非常见分布：马尔可夫链蒙特卡洛 (MCMC) 法

如何从后验分布中抽取代表性样本?

- 若后验分布 $p(\theta|y)$ 为常用分布：直接抽取 iid 样本
  1. Inverse CDF Method
  2. Rejection Sampling
  3. Importance Sampling
     - 缺点：高维有困难
- 若后验分布 $p(\theta|y)$ 为非常见分布：马尔可夫链蒙特卡洛 (MCMC) 法
  1. The Gibbs Sampler

如何从后验分布中抽取代表性样本?

- 若后验分布 $p(\theta|y)$ 为常用分布：直接抽取 iid 样本
  1. Inverse CDF Method
  2. Rejection Sampling
  3. Importance Sampling
     - ★ 缺点：高维有困难
- 若后验分布 $p(\theta|y)$ 为非常见分布：马尔可夫链蒙特卡洛 (MCMC) 法
  1. The Gibbs Sampler
  2. The Metropolis-Hastings Algorithm

本章目标

目的：用计算机模拟产生后验分布的大量样本，用以计算后验分布及其特征。

1. 计算机模拟简介
   1. 模拟法计算函数的积分
   2. 随机变量分布的直接模拟法（MC）
2. 马尔可夫链蒙特卡罗法 (MCMC)
   1. The Metropolis-Hastings Algorithm（MH 迭代法）
   2. The Gibbs Sampler（吉布斯抽样器）

Outline

Outline

1. 随机模拟介绍
   - 从蒲丰抛针问题说起
   - 函数积分的 Monte Carlo 法
   - 用 R 直接进行贝叶斯计算的一些技巧
   - 随机变量分布的直接抽样法

蒲丰抛针问题

法国科学家蒲丰于 1777 年提出的一种计算圆周率的方法——随机投针法。

1. 取一张白纸，在上面画上许多条间距为 a 的平行线
2. 取一根长度为 L（$L < a$） 的针，随机地向画有平行直线的纸上掷 n 次，观察针与直线相交的次数，记为 m
3. 计算针与直线相交的概率

$$P = \frac{2L}{\pi a} \approx \frac{m}{n}$$

历史上的投针试验

试验者	时间	投掷次数	相交次数	圆周率估计值
Wolf	1850	5000	2532	3.1596
Smith	1855	3204	1219	3.1554
Morgan	1860	600	383	3.137
Fox	1884	1030	489	3.1595
Lazzerini	1901	3408	1808	3.1415929
Reina	1925	2520	859	3.1795

计算机抛针模拟

```
set.seed(1234)
L <- 0.8 # 针的长度，平行线间距 a=1
n <- 1e+06 # 重复 100000 次
u1 <- runif(n) # 取随机数
x <- 1/2 * u1 # x 是针中心到最近的线的距离
u2 <- runif(n)
y <- L/2 * sin(u2 * 2 * pi) # 针的垂直长度的一半
z <- as.numeric(x <= y) # 相交的充要条件是 x<=y
pi.e <- n * L/sum(z) # pi 的估计式
pi.e
[1] 3.140938
```

计算机模拟技术的先驱们

A. A. Markov (1857 – 1936)

John von Neumann, Stanislav Ulam, Nicholas Metropolis

Monte Carlo

- **蒙特卡罗** (MonteCarlo) 方法，或称**计算机随机模拟**方法，是一种基于重复抽取" 随机数" 的计算方法
- 源于美国在第二次世界大战进研制原子弹的 “曼哈顿计划”，涉及多达 10^{23} 个带电原子的估计和计算
- Stanislaw Ulam 在 40 年代末发明，他当时在 Los Alamos 国家实验室研究原子弹
- 冯 · 诺伊曼 (John von Neumann) 用著名的第一代电子计算机 ENIAC 编程计算
- 冯 · 诺伊曼用世界赌城 Monte Carlo 命名

贝叶斯计算发展简史

- prehistory (1763 –1960): Conjugate priors
- 1960s: Numerical quadrature
  - Newton-Cotes methods, Gaussian quadrature, etc.
- 1970s: Expectation-Maximization (“EM”) algorithm
  - iterative mode-finder
- 1980s: Asymptotic methods
  - Laplace’s method, saddlepoint approximations (鞍点逼近)
- 1980s: Noniterative Monte Carlo methods
  - Direct posterior sampling and indirect methods (importance sampling, rejection, etc.)
- 1990s: Markov Chain Monte Carlo (MCMC)，革命性的方法
  - Gibbs sampler, Metropolis-Hastings algorithm
- MCMC methods broadly applicable, but require care in parametrization and convergence diagnosis

Outline

如何用模拟法计算后验分布的积分？

对任意函数 $h(\theta)$ 的后验积分（已知后验分布 $p(\theta|\boldsymbol{y})$)：

1. 网格化 (非随机) 方法：假设参数区间 $[a,b]$ 分为 M 等分，在每个部分选取一点 $\theta^{(i)}$，则

$$\mathrm{E}[h(\theta)|\boldsymbol{y}] = \int_a^b h(\theta)p(\theta|\boldsymbol{y})d\theta \approx \frac{b-a}{M}\sum_{i=1}^{M} h(\theta^{(i)})p(\theta^{(i)}|\boldsymbol{y})$$

如何用模拟法计算后验分布的积分？

对任意函数 $h(\theta)$ 的后验积分（已知后验分布 $p(\theta|\boldsymbol{y})$）：

1. 网格化 (非随机) 方法：假设参数区间 $[a,b]$ 分为 M 等分，在每个部分选取一点 $\theta^{(i)}$，则

$$\mathrm{E}[h(\theta)|\boldsymbol{y}] = \int_a^b h(\theta)p(\theta|\boldsymbol{y})d\theta \approx \frac{b-a}{M}\sum_{i=1}^{M} h(\theta^{(i)})p(\theta^{(i)}|\boldsymbol{y})$$

2. 随机抽样（iid）方法：设 $\theta^{(i)}(i=1,2,\ ,M)$ 为来自后验分布 $p(\theta|\mathbf{y})$ 的 iid 样本，则根据大数定律

$$\bar{h} = \mathrm{E}[h(\theta)|\boldsymbol{y}] = \int h(\theta)p(\theta|\boldsymbol{y})d\theta \approx \frac{1}{M}\sum_{i=1}^{M} h(\theta^{(i)})$$

如何用模拟法计算后验分布的积分？

对任意函数 $h(\theta)$ 的后验积分（已知后验分布 $p(\theta|\boldsymbol{y})$）：

1. 网格化 (非随机) 方法：假设参数区间 $[a,b]$ 分为 M 等分，在每个部分选取一点 $\theta^{(i)}$，则

$$\mathrm{E}[h(\theta)|\boldsymbol{y}] = \int_a^b h(\theta)p(\theta|\boldsymbol{y})d\theta \approx \frac{b-a}{M}\sum_{i=1}^{M} h(\theta^{(i)})p(\theta^{(i)}|\boldsymbol{y})$$

2. 随机抽样（iid）方法：设 $\theta^{(i)}(i=1,2,\ ,M)$ 为来自后验分布 $p(\theta|\mathbf{y})$ 的 iid 样本，则根据大数定律

$$\bar{h} = \mathrm{E}[h(\theta)|\boldsymbol{y}] = \int h(\theta)p(\theta|\boldsymbol{y})d\theta \approx \frac{1}{M}\sum_{i=1}^{M} h(\theta^{(i)})$$

3. 后验分布的推断：

如何用模拟法计算后验分布的积分？

对任意函数 $h(\theta)$ 的后验积分（已知后验分布 $p(\theta|\boldsymbol{y})$）：

1. 网格化 (非随机) 方法：假设参数区间 $[a,b]$ 分为 M 等分，在每个部分选取一点 $\theta^{(i)}$，则

$$\mathrm{E}[h(\theta)|\boldsymbol{y}] = \int_a^b h(\theta)p(\theta|\boldsymbol{y})d\theta \approx \frac{b-a}{M}\sum_{i=1}^{M} h(\theta^{(i)})p(\theta^{(i)}|\boldsymbol{y})$$

2. 随机抽样（iid）方法：设 $\theta^{(i)}(i=1,2,\ ,M)$ 为来自后验分布 $p(\theta|\mathbf{y})$ 的 iid 样本，则根据大数定律

$$\bar{h} = \mathrm{E}[h(\theta)|\boldsymbol{y}] = \int h(\theta)p(\theta|\boldsymbol{y})d\theta \approx \frac{1}{M}\sum_{i=1}^{M} h(\theta^{(i)})$$

3. 后验分布的推断：
   - 期望：$E(\theta|\boldsymbol{y}) = \frac{1}{M}\sum_{i=1}^{M}\theta^{(i)}$

如何用模拟法计算后验分布的积分？

对任意函数 $h(\theta)$ 的后验积分（已知后验分布 $p(\theta|\boldsymbol{y})$）：

1. 网格化 (非随机) 方法：假设参数区间 $[a,b]$ 分为 M 等分，在每个部分选取一点 $\theta^{(i)}$，则

$$\mathrm{E}[h(\theta)|\boldsymbol{y}] = \int_a^b h(\theta)p(\theta|\boldsymbol{y})d\theta \approx \frac{b-a}{M}\sum_{i=1}^{M} h(\theta^{(i)})p(\theta^{(i)}|\boldsymbol{y})$$

2. 随机抽样（iid）方法：设 $\theta^{(i)}(i=1,2,\ ,M)$ 为来自后验分布 $p(\theta|\mathbf{y})$ 的 iid 样本，则根据大数定律

$$\bar{h} = \mathrm{E}[h(\theta)|\boldsymbol{y}] = \int h(\theta)p(\theta|\boldsymbol{y})d\theta \approx \frac{1}{M}\sum_{i=1}^{M} h(\theta^{(i)})$$

3. 后验分布的推断：
   - 期望：$E(\theta|\boldsymbol{y}) = \frac{1}{M}\sum_{i=1}^{M}\theta^{(i)}$
   - 预测：$p(\tilde{y}|\boldsymbol{y}) = \frac{1}{M}\sum_{i=1}^{M} f(\tilde{y}|\theta^{(i)})$

如何用模拟法计算后验分布的积分？

对任意函数 $h(\theta)$ 的后验积分（已知后验分布 $p(\theta|\boldsymbol{y})$）：

1. 网格化 (非随机) 方法：假设参数区间 $[a,b]$ 分为 M 等分，在每个部分选取一点 $\theta^{(i)}$，则

$$\mathrm{E}[h(\theta)|\boldsymbol{y}] = \int_a^b h(\theta)p(\theta|\boldsymbol{y})d\theta \approx \frac{b-a}{M}\sum_{i=1}^{M} h(\theta^{(i)})p(\theta^{(i)}|\boldsymbol{y})$$

2. 随机抽样（iid）方法：设 $\theta^{(i)}(i=1,2,\ ,M)$ 为来自后验分布 $p(\theta|\mathbf{y})$ 的 iid 样本，则根据大数定律

$$\bar{h} = \mathrm{E}[h(\theta)|\boldsymbol{y}] = \int h(\theta)p(\theta|\boldsymbol{y})d\theta \approx \frac{1}{M}\sum_{i=1}^{M} h(\theta^{(i)})$$

3. 后验分布的推断：
   - 期望：$E(\theta|\boldsymbol{y}) = \frac{1}{M}\sum_{i=1}^{M}\theta^{(i)}$
   - 预测：$p(\tilde{y}|\boldsymbol{y}) = \frac{1}{M}\sum_{i=1}^{M} f(\tilde{y}|\theta^{(i)})$
   - 概率：$P(\theta \in A|\boldsymbol{y}) = \frac{1}{M}\{\#\theta^{(i)}\text{of draws} \in A\}$

Monte Carlo 计算偏度和峰度

例 (计算 Beta 分布的偏度和峰度)

假设后验分布 $\theta \sim Beta(5, 10)$，求偏度及峰度。

$$Skew = E\left(\frac{\theta-\mu}{\sigma}\right)^3, \quad Kurt = E\left(\frac{\theta-\mu}{\sigma}\right)^4 - 3$$

```
set.seed(1234)
x<-rbeta(1000,5,10)
u<-(x-mean(x))/sd(x)
skew<-mean(u^3)
kurt<-mean(u^4)-3
c(skew,kurt)

## [1]  0.43107372 -0.02523695
```

- 维数灾难：对高维空间，非随机方法对点的选取有难度
- 关键问题：如何产生后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$ 真正有代表性的随机样本？

Outline

1. 随机模拟介绍
   - 从蒲丰抛针问题说起
   - 函数积分的 Monte Carlo 法
   - 用 R 直接进行贝叶斯计算的一些技巧
   - 随机变量分布的直接抽样法

用 R 直接模拟抽样

一些预处理技巧：

- 后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$ 有显式公式
  - 可以用非正则化密度 $q(\boldsymbol{\theta}|\mathbf{y})$ ($p(\boldsymbol{\theta}|\mathbf{y})$ 剔除常数）
- 把参数变换为实数
  - 参数为正数（如方差），则做对数变换
  - 对成数，做 logit 变换：$\text{logit}(p) = \log[p/(1-p)]$
- 后验分布密度：一般先做 log 变换，使得对样本运算为相加

$$\log p(\boldsymbol{\theta}|\mathbf{y}) = \log \pi(\boldsymbol{\theta}) + \sum_{i=1}^{n} \log f(\mathbf{y}|\boldsymbol{\theta})$$

Outline

1. 随机模拟介绍
   - 从蒲丰抛针问题说起
   - 函数积分的 Monte Carlo 法
   - 用 R 直接进行贝叶斯计算的一些技巧
   - 随机变量分布的直接抽样法

Inverse CDF Method

定理

设随机变量 X 的分布函数 *(CDF)*：$F(x) = P(X \leq x)$，如果 $F(x)$ 连续，严格单调，则

$$Y = F(X) \sim Unif(0,1)$$

应用：设 $u_{1,}u_2,\ldots,u_M$ 是 $Unif(0,1)$ 的 iid 样本，则 X 的 iid 样本为

$$x_i = F^{-1}(u_i)$$

Examples

1. Cauchy(μ, σ)

$$
\begin{aligned}
F(x|\mu,\sigma) &= \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-\mu}{\sigma}\right) \\
F^{-1}(u|\mu,\sigma) &= \mu + \sigma\tan(\pi(u-0.5))
\end{aligned}
$$

2. Exp(β)

$$
\begin{aligned}
F(x|\beta) &= 1-\exp(-x/\beta) \\
F^{-1}(u|\beta) &= -\beta\log(1-u)
\end{aligned}
$$

例：如何模拟产生 Cauchy(1,5) 分布的 n=1000 个样本？

```r
n<-1000
u<-runif(n)
x<-1+5*tan(pi*(u-0.5))
```

优缺点

优点：

- 样本直接来自目标的分布

缺点：

- 分布函数 (CDF) 及其反函数不一定能求出（如 Normal, Beta, Gamma 等）
- 这种模拟方法效果不好（特别对分布尾部）
- 多维变量失效
- 对离散型 rv，还需要其它方法

Rejection Sampling

目的：用筛选法直接抽取后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$ 的样本
找一个容易模拟的分布 $g(\boldsymbol{\theta})$，使得 $p(\boldsymbol{\theta}|\mathbf{y}) \leq cg(\boldsymbol{\theta})$，对所有 $\boldsymbol{\theta}$ 和某个 c 成立。
抽样步骤：

1. 抽取 $\boldsymbol{\theta} \sim g(\boldsymbol{\theta})$，计算接受率

$$r(\boldsymbol{\theta}) = \frac{p(\boldsymbol{\theta}|y)}{cg(\boldsymbol{\theta})} < 1$$

   (r 越大，迭代越有效率)

2. 抽取 $u \sim Unif(0,1)$
   如果 $u \leq r(\boldsymbol{\theta})$，接受 $\boldsymbol{\theta}$ 为 $p(\boldsymbol{\theta}|\mathbf{y})$ 的样本
   如果 $u > r(\boldsymbol{\theta})$，拒绝 $\boldsymbol{\theta}$，并返回 1
   （即以概率 $p = r(\boldsymbol{\theta})$ 接受 $\boldsymbol{\theta}$ 为 $p(\boldsymbol{\theta}|\mathbf{y})$ 的样本）

优缺点

- 优点：
  - 最有用迭代法之一
  - 把 $p(\boldsymbol{\theta}|\mathbf{y})$ 换成非正则化后验密度 $q(\theta|y)$ 也成立，这一点很有用，因为 $p(\boldsymbol{\theta}|\mathbf{y})$ 的常数部分常常未知
  - 多数标准分布都用这种方法：Normal, Gamma, Beta,...
- 缺点：
  - 可能低效率：c 很大时，接受率 ($=1/c$) 很低
  - 源分布 $g(\boldsymbol{\theta})$ 不明确
- 也叫 Acceptance-Rejection Sampling (筛选法)
- Remarks
  - Target density: $p(\boldsymbol{\theta}|\mathbf{y})$
  - Source density: $g(\boldsymbol{\theta})$

例：Normal-Cauchy Model

假设 $Y \sim N(\theta, 1)$，$\theta \sim Cauchy : \pi(\theta) \propto (1+\theta^2)^{-1} = t(1)$，后验分布不是常用分布

现假设参数真值 $\theta = 2$，且有 5 个观察值：y<-rnorm(5,2,1)，试模拟后验分布并求出 θ 的估计值。

R 模拟方法：

1. 定义后验密度
2. 确定源密度 $g(\theta)$： $N(\bar{y}, s^2)$
3. 确定常数 $c = \max[f(\theta|y)/g(\theta)]$
4. 迭代抽样
5. 利用样本作图、计算

Importance sampling

目的：用 Monte Carlo 方法计算 $E(h(\theta)|y)$
记 $q(\theta|y)$ 为没有正则化的后验密度 (核心部分), 无法直接抽取样本
假设有容易抽取样本的常用分布 $g(\theta)$，并模拟出样本：
$\theta^1, \theta^2, \ldots, \theta^M \sim g(\theta)$，则

$$
\begin{aligned}
E(h(\theta)|y) &= \frac{\int h(\theta)q(\theta|y)d\theta}{\int q(\theta|y)d\theta} = \frac{\int [h(\theta)q(\theta|y)/g(\theta)]g(\theta)d\theta}{\int [q(\theta|y)/g(\theta)]g(\theta)d\theta} \\
&= \frac{\frac{1}{M}\sum_{i=1}^{M} h(\theta^i)w(\theta^i)}{\frac{1}{M}\sum_{i=1}^{M} w(\theta^i)}
\end{aligned}
$$

其中

$$
w(\theta^i) = \frac{q(\theta^i|y)}{g(\theta^i)} \propto \frac{\pi(\theta^i)\prod_{i=1}^{n-1} f(y_i|\theta^i)}{g(\theta^i)} f(y_n|\theta^i)
$$

称为重要性比例或权重 (Importance ratios or weights)
当新增一个观察值 y_{n+1}, 对应权重 $w_{n+1}(\theta^i) = w_n(\theta^i)f(y_n|\theta^i)$

例：Normal conjugate

- 模型：$y_1, y_2, \ldots, y_n \sim N(\theta, \sigma^2)$, $\theta \sim N(0, 100)$
- 计算：后验均值 $E(\theta|\mathbf{y})$ (即 $h(\theta) = \theta$)
  - 真值 $\theta = 1$,$\sigma = 1$（已知）
  - 从 $N(1,1)$ 模拟 $n = 100$ 个观察值
- 迭代：
  1. 候选分布抽样：$\theta^1, \ldots, \theta^S \sim g(\theta) = Unif(-5, 5)$
  2. 计算对数权重：
     $\log w_i = \log \phi(\theta^i; 0, 100) + \sum_{j=1}^n \log \phi(y_j; \theta^i, 1) - \log g(\theta^i)$
  3. 计算权重：$w_i = \exp(\log w_i - \max(\log w_i))$
  4. 计算后验均值：$E(\theta|y) = \sum_{i=1}^S \theta^i w_i / (\sum w_i$

Sampling Importance Resampling (SIR)

目的：产生后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$ 的等权重独立样本

- 假设已经产生样本：$\theta^1, \theta^2, \ldots, \theta^M \sim g(\theta)$，和权重 $w(\theta^i) = \frac{q(\theta^i|y)}{g(\theta^i)}$，我们要产生 $p(\boldsymbol{\theta}|\mathbf{y})$ 的 n $(n < M)$ 个样本
- 定义离散型分布：

$$p_j = \frac{w(\theta^j)}{\sum_{i=1}^{M} w(\theta^i)}$$

  - 抽取 θ_1^*: 从 $\{\theta^1, \theta^2, \ldots, \theta^M\}$ 中抽取，每个被抽中的概率为 p_j
  - 抽取 θ_2^*: 从其余 $M-1$ 个抽取第二个（不重复）
  - 以此类推，直到抽取第 n 个，则 $\theta_1^*, \ldots, \theta_n^* \sim p(\theta|y)$

- SIR 是 Weighted Bootstrap 抽样

Outline

Monte Carlo 积分与 MCMC

- Monte Carlo 积分

Monte Carlo 积分与 MCMC

- Monte Carlo 积分
  - Draw **independent** samples from posterior distribution

Monte Carlo 积分与 MCMC

- Monte Carlo 积分
  - Draw **independent** samples from posterior distribution
  - Use sample averages to approximate expectations

Monte Carlo 积分与 MCMC

- Monte Carlo 积分
  - Draw **independent** samples from posterior distribution
  - Use sample averages to approximate expectations
  - BUT it's difficult with high-dimensional posteriors

Monte Carlo 积分与 MCMC

- Monte Carlo 积分
  - Draw **independent** samples from posterior distribution
  - Use sample averages to approximate expectations
  - BUT it's difficult with high-dimensional posteriors
- Markov chain Monte Carlo (MCMC)

Monte Carlo 积分与 MCMC

- Monte Carlo 积分
  - Draw **independent** samples from posterior distribution
  - Use sample averages to approximate expectations
  - BUT it's difficult with high-dimensional posteriors
- Markov chain Monte Carlo (MCMC)
  - Draw samples by running a **markov chain** that is constructed so that its limiting (stationary) distribution is the **joint distribution of interest**

Monte Carlo 积分与 MCMC

- Monte Carlo 积分
  - Draw **independent** samples from posterior distribution
  - Use sample averages to approximate expectations
  - BUT it's difficult with high-dimensional posteriors
- Markov chain Monte Carlo (MCMC)
  - Draw samples by running a **markov chain** that is constructed so that its limiting (stationary) distribution is the **joint distribution of interest**
  - Used when it is not possible (or not computationally efficient) to sample directly

Monte Carlo 积分与 MCMC

- Monte Carlo 积分
  - Draw **independent** samples from posterior distribution
  - Use sample averages to approximate expectations
  - BUT it's difficult with high-dimensional posteriors
- Markov chain Monte Carlo (MCMC)
  - Draw samples by running a **markov chain** that is constructed so that its limiting (stationary) distribution is the **joint distribution of interest**
  - Used when it is not possible (or not computationally efficient) to sample directly
  - Samples are **not independent** ！

马尔可夫链（Markov Chain）

- 马尔科夫链

马尔可夫链（Markov Chain）

- 马尔科夫链
  - 如果 p 维时间序列 $\boldsymbol{\theta}^{(t)}$ $(t = 0, 1, 2, \dots)$ 具有**马氏性**：下一个状态如何，仅依赖于现在的状态，与过去的状态无关，即

$$P(\boldsymbol{\theta}^{(t+1)} \in \Theta | \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(t)}) = P(\boldsymbol{\theta}^{(t+1)} \in \Theta | \boldsymbol{\theta}^{(t)})$$

    则称该时间序列为马尔科夫链。

马尔可夫链（Markov Chain）

- 马尔科夫链
  - 如果 p 维时间序列 $\boldsymbol{\theta}^{(t)}$ $(t = 0, 1, 2, \dots)$ 具有**马氏性**：下一个状态如何，仅依赖于现在的状态，与过去的状态无关，即

$$P(\boldsymbol{\theta}^{(t+1)} \in \Theta | \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(t)}) = P(\boldsymbol{\theta}^{(t+1)} \in \Theta | \boldsymbol{\theta}^{(t)})$$

    则称该时间序列为马尔科夫链。
  - $\boldsymbol{\theta}^{(t)}$ 的所有可能取值组成状态空间 Θ

马尔可夫链（Markov Chain）

- 马尔科夫链
  - 如果 p 维时间序列 $\boldsymbol{\theta}^{(t)}$ $(t = 0, 1, 2, \ldots)$ 具有**马氏性**：下一个状态如何，仅依赖于现在的状态，与过去的状态无关，即

$$P(\boldsymbol{\theta}^{(t+1)} \in \Theta | \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(t)}) = P(\boldsymbol{\theta}^{(t+1)} \in \Theta | \boldsymbol{\theta}^{(t)})$$

    则称该时间序列为马尔科夫链。
  - $\boldsymbol{\theta}^{(t)}$ 的所有可能取值组成状态空间 Θ
- $\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}^{(t-1)} \sim T_t(\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}^{(t-1)})$ 称为**转移分布** (transition distribution)

马尔可夫链（Markov Chain）

- 马尔科夫链
  - 如果 p 维时间序列 $\boldsymbol{\theta}^{(t)}$ $(t=0,1,2,\ldots)$ 具有**马氏性：**下一个状态如何，仅依赖于现在的状态，与过去的状态无关，即
  $$P(\boldsymbol{\theta}^{(t+1)} \in \Theta | \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(t)}) = P(\boldsymbol{\theta}^{(t+1)} \in \Theta | \boldsymbol{\theta}^{(t)})$$
  则称该时间序列为马尔科夫链。
  - $\boldsymbol{\theta}^{(t)}$ 的所有可能取值组成状态空间 Θ
- $\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}^{(t-1)} \sim T_t(\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}^{(t-1)})$ 称为**转移分布** (transition distribution)
  - 如果状态空间是离散的，$p_{jk}^{(t)} = P(\theta^{(t)} = j | \theta^{(t-1)} = k)$ 称为一步转移概率，$M = (p_{jk}^{(t)})$ 称为转移概率矩阵

马尔可夫链（Markov Chain）

- 马尔科夫链
  - 如果 p 维时间序列 $\boldsymbol{\theta}^{(t)}$ $(t=0,1,2,\ldots)$ 具有**马氏性**：下一个状态如何，仅依赖于现在的状态，与过去的状态无关，即
  $$P(\boldsymbol{\theta}^{(t+1)} \in \Theta | \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(t)}) = P(\boldsymbol{\theta}^{(t+1)} \in \Theta | \boldsymbol{\theta}^{(t)})$$
  则称该时间序列为马尔科夫链。
  - $\boldsymbol{\theta}^{(t)}$ 的所有可能取值组成状态空间 Θ
- $\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}^{(t-1)} \sim T_t(\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}^{(t-1)})$ 称为**转移分布** (transition distribution)
  - 如果状态空间是离散的，$p_{jk}^{(t)} = P(\theta^{(t)} = j | \theta^{(t-1)} = k)$ 称为一步转移概率，$M = (p_{jk}^{(t)})$ 称为转移概率矩阵
- 重要结论：稳定分布（极限分布）

马尔可夫链（Markov Chain）

- 马尔科夫链
  - 如果 p 维时间序列 $\boldsymbol{\theta}^{(t)}$ $(t=0,1,2,\dots)$ 具有**马氏性**：下一个状态如何，仅依赖于现在的状态，与过去的状态无关，即

$$P(\boldsymbol{\theta}^{(t+1)} \in \Theta | \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(t)}) = P(\boldsymbol{\theta}^{(t+1)} \in \Theta | \boldsymbol{\theta}^{(t)})$$

    则称该时间序列为马尔科夫链。
  - $\boldsymbol{\theta}^{(t)}$ 的所有可能取值组成状态空间 Θ
- $\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}^{(t-1)} \sim T_t(\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}^{(t-1)})$ 称为**转移分布** (transition distribution)
  - 如果状态空间是离散的，$p_{jk}^{(t)} = P(\theta^{(t)} = j | \theta^{(t-1)} = k)$ 称为一步转移概率，$M = (p_{jk}^{(t)})$ 称为转移概率矩阵
- 重要结论：稳定分布（极限分布）
  - 在一定条件下，马尔科夫链 $\{\boldsymbol{\theta}^{(t)}\}$ 有唯一的极限分布（稳定分布），即当 $t \to \infty$，

$$\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_p^{(t)})' \xrightarrow{\mathrm{d}} \boldsymbol{\theta} \sim p(\boldsymbol{\theta} | \mathbf{y})$$

马尔可夫链蒙特卡洛（MCMC）抽样

如何构造马尔科夫链 $\boldsymbol{\theta}^{(t)}(t \geq 0)$，使得唯一极限分布正好是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$?

- 常用 MCMC 抽样法：

马尔可夫链蒙特卡洛（MCMC）抽样

如何构造马尔科夫链 $\boldsymbol{\theta}^{(t)}(t \geq 0)$，使得唯一极限分布正好是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$?

- 常用 MCMC 抽样法：
  1. The Gibbs Sampler (Geman and Geman, 1984; Gelfand and Smith, 1990; fundamentally changed Bayesian computing)

马尔可夫链蒙特卡洛（MCMC）抽样

如何构造马尔科夫链 $\boldsymbol{\theta}^{(t)}(t \geq 0)$，使得唯一极限分布正好是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$?

- 常用 MCMC 抽样法：
  1. The Gibbs Sampler (Geman and Geman, 1984; Gelfand and Smith, 1990; fundamentally changed Bayesian computing)
  2. The Metropolis-Hastings Algorithm (Hastings (1970))

马尔可夫链蒙特卡洛（MCMC）抽样

如何构造马尔科夫链 $\boldsymbol{\theta}^{(t)}(t \geq 0)$，使得唯一极限分布正好是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$?

- 常用 MCMC 抽样法：
  1. The Gibbs Sampler (Geman and Geman, 1984; Gelfand and Smith, 1990; fundamentally changed Bayesian computing)
  2. The Metropolis-Hastings Algorithm (Hastings (1970))
- 假设马氏链迭代 M 次，放弃前 m 次结果 (叫 Burn-in，因为前面的迭代样本不稳定)，把后面 $(M-m)$ 次结果作为后验分布的样本

Gibbs Sampler

从马尔科夫链中依次抽取 p 维序列 $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots\}$，其极限分布是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$, 其中转移分布 (transition kernel) 是所有条件分布的乘积。

- 参数 $\boldsymbol{\theta}$ 是 p 维向量，而 Gibbs Sampler 依次抽取 1 维样本，再组成 p 维联合样本

Gibbs Sampler

从马尔科夫链中依次抽取 p 维序列 $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots\}$，其极限分布是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$, 其中转移分布 (transition kernel) 是所有条件分布的乘积。

- 参数 $\boldsymbol{\theta}$ 是 p 维向量，而 Gibbs Sampler 依次抽取 1 维样本，再组成 p 维联合样本
- 前提是：要推导出所有**条件后验分布**

Gibbs Sampler

从马尔科夫链中依次抽取 p 维序列 $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots\}$，其极限分布是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$, 其中转移分布 (transition kernel) 是所有条件分布的乘积。

- 参数 $\boldsymbol{\theta}$ 是 p 维向量，而 Gibbs Sampler 依次抽取 1 维样本，再组成 p 维联合样本
- 前提是：要推导出所有**条件后验分布**

Algorithm

Gibbs Sampler

从马尔科夫链中依次抽取 p 维序列 $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots\}$，其极限分布是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$, 其中转移分布 (transition kernel) 是所有条件分布的乘积。

- 参数 $\boldsymbol{\theta}$ 是 p 维向量，而 Gibbs Sampler 依次抽取 1 维样本，再组成 p 维联合样本
- 前提是：要推导出所有**条件后验分布**

Algorithm

1. 初值：$\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)})$

Gibbs Sampler

从马尔科夫链中依次抽取 p 维序列 $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots\}$，其极限分布是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$, 其中转移分布 (transition kernel) 是所有条件分布的乘积。

- 参数 $\boldsymbol{\theta}$ 是 p 维向量，而 Gibbs Sampler 依次抽取 1 维样本，再组成 p 维联合样本
- 前提是：要推导出所有**条件后验分布**

Algorithm

1. 初值：$\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)})$
2. 依次抽取 $\boldsymbol{\theta}^{(j)}$:

$$
\begin{aligned}
\theta_1^{(j)} &\sim p(\theta_1|\theta_2^{(j-1)}, \theta_3^{(j-1)}, \ldots, \theta_p^{(j-1)}, \mathbf{y}) \\
\theta_2^{(j)} &\sim p(\theta_2|\theta_1^{(j)}, \theta_3^{(j-1)}, \ldots, \theta_p^{(j-1)}, \mathbf{y}) \\
&\vdots \\
\theta_p^{(j)} &\sim p(\theta_p|\theta_1^{(j)}, \theta_2^{(j)}, \ldots, \theta_{p-1}^{(j)}, \mathbf{y})
\end{aligned}
$$

Metropolis-Hastings：Motivation

- Gibbs sampling 需要知道所有条件后验分布

Metropolis-Hastings：Motivation

- Gibbs sampling 需要知道所有条件后验分布
  - 所有例子中，条件后验分布都是共轭的

Metropolis-Hastings：Motivation

- Gibbs sampling 需要知道所有条件后验分布
  - 所有例子中，条件后验分布都是共轭的
- 如果不是共轭分布，怎么办？

Metropolis-Hastings：Motivation

- Gibbs sampling 需要知道所有条件后验分布
  - 所有例子中，条件后验分布都是共轭的
- 如果不是共轭分布，怎么办？
  - 可以用 Rejection sampling 抽取条件后验分布的样本

Metropolis-Hastings：Motivation

- Gibbs sampling 需要知道所有条件后验分布
  - 所有例子中，条件后验分布都是共轭的
- 如果不是共轭分布，怎么办？
  - 可以用 Rejection sampling 抽取条件后验分布的样本
  - 如果非共轭参数很少，OK，但... 容易搞混

Metropolis-Hastings：Motivation

- Gibbs sampling 需要知道所有条件后验分布
  - 所有例子中，条件后验分布都是共轭的
- 如果不是共轭分布，怎么办？
  - 可以用 Rejection sampling 抽取条件后验分布的样本
  - 如果非共轭参数很少，OK，但... 容易搞混
- 有几个办法

Metropolis-Hastings：Motivation

- Gibbs sampling 需要知道所有条件后验分布
  - 所有例子中，条件后验分布都是共轭的
- 如果不是共轭分布，怎么办？
  - 可以用 Rejection sampling 抽取条件后验分布的样本
  - 如果非共轭参数很少，OK，但... 容易搞混
- 有几个办法
  1. Auxiliary variables

Metropolis-Hastings：Motivation

- Gibbs sampling 需要知道所有条件后验分布
  - 所有例子中，条件后验分布都是共轭的
- 如果不是共轭分布，怎么办？
  - 可以用 Rejection sampling 抽取条件后验分布的样本
  - 如果非共轭参数很少，OK，但... 容易搞混
- 有几个办法
  1. Auxiliary variables
  2. Slice sampling

Metropolis-Hastings：Motivation

- Gibbs sampling 需要知道所有条件后验分布
  - 所有例子中，条件后验分布都是共轭的
- 如果不是共轭分布，怎么办？
  - 可以用 Rejection sampling 抽取条件后验分布的样本
  - 如果非共轭参数很少，OK，但... 容易搞混
- 有几个办法
  1. Auxiliary variables
  2. Slice sampling
  3. Metropolis-Hastings sampling

Metropolis-Hastings：Motivation

- Gibbs sampling 需要知道所有条件后验分布
  - 所有例子中，条件后验分布都是共轭的
- 如果不是共轭分布，怎么办？
  - 可以用 Rejection sampling 抽取条件后验分布的样本
  - 如果非共轭参数很少，OK，但... 容易搞混
- 有几个办法
  1. Auxiliary variables
  2. Slice sampling
  3. Metropolis-Hastings sampling
- Metropolis-Hastings sampling 应用最广泛

Metropolis-Hastings Algorithm

从马尔科夫链中依次抽取 p 维序列 $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots\}$，其极限分布是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$

Algorithm

Metropolis-Hastings Algorithm

从马尔科夫链中依次抽取 p 维序列 $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \dots\}$，其极限分布是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$

Algorithm

1. 初值：$\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$

Metropolis-Hastings Algorithm

从马尔科夫链中依次抽取 p 维序列 $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots\}$，其极限分布是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$

Algorithm

1. 初值：$\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)})$
2. 第 j 步：从预选分布 (proposal density) 中抽取候选样本:$\boldsymbol{\theta}^* \sim g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})$

Metropolis-Hastings Algorithm

从马尔科夫链中依次抽取 p 维序列 $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots\}$，其极限分布是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$

Algorithm

1. 初值：$\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)})$
2. 第 j 步：从预选分布 (proposal density) 中抽取候选样本:$\boldsymbol{\theta}^* \sim g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})$
3. 计算接受概率

$$\alpha = \min\left\{1, \frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}^{(j-1)}|\mathbf{y})} \frac{g(\boldsymbol{\theta}^{(j-1)}|\boldsymbol{\theta}^*)}{g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})}\right\}$$

Metropolis-Hastings Algorithm

从马尔科夫链中依次抽取 p 维序列 $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots\}$，其极限分布是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$

Algorithm

1. 初值：$\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)})$
2. 第 j 步：从预选分布 (proposal density) 中抽取候选样本:$\boldsymbol{\theta}^* \sim g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})$
3. 计算接受概率

$$\alpha = \min\left\{1, \frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}^{(j-1)}|\mathbf{y})} \frac{g(\boldsymbol{\theta}^{(j-1)}|\boldsymbol{\theta}^*)}{g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})}\right\}$$

4. 接受规则：产生 $U \sim Unif(0,1)$

Metropolis-Hastings Algorithm

从马尔科夫链中依次抽取 p 维序列 $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots\}$，其极限分布是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$

Algorithm

1. 初值：$\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)})$
2. 第 j 步：从预选分布 (proposal density) 中抽取候选样本:$\boldsymbol{\theta}^* \sim g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})$
3. 计算接受概率

$$\alpha = \min\left\{1, \frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}^{(j-1)}|\mathbf{y})} \frac{g(\boldsymbol{\theta}^{(j-1)}|\boldsymbol{\theta}^*)}{g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})}\right\}$$

4. 接受规则：产生 $U \sim Unif(0,1)$
   - 若 $U < \alpha$，则接受 $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^*$

Metropolis-Hastings Algorithm

从马尔科夫链中依次抽取 p 维序列 $\{\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots\}$，其极限分布是后验分布 $p(\boldsymbol{\theta}|\mathbf{y})$

Algorithm

1. 初值：$\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)})$
2. 第 j 步：从预选分布 (proposal density) 中抽取候选样本:$\boldsymbol{\theta}^* \sim g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})$
3. 计算接受概率

$$\alpha = \min\left\{1, \frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}^{(j-1)}|\mathbf{y})} \frac{g(\boldsymbol{\theta}^{(j-1)}|\boldsymbol{\theta}^*)}{g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})}\right\}$$

4. 接受规则：产生 $U \sim Unif(0,1)$
   - 若 $U < \alpha$，则接受 $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^*$
   - 否则，$\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^{(j-1)}$

Proposal density

定理

假设预选分布 $g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})$ 满足一定正则条件，则 *M-H* 抽样得到的序列收敛于稳定分布

Proposal density

定理

假设预选分布 $g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})$ 满足一定正则条件，则 *M-H* 抽样得到的序列收敛于稳定分布

1. 随机漫步 (Random walk chains) 链

Proposal density

定理

假设预选分布 $g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})$ 满足一定正则条件，则 *M-H* 抽样得到的序列收敛于稳定分布

1. 随机漫步 (Random walk chains) 链
   - 预选密度具有如下函数形式：$g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)}) = h(\boldsymbol{\theta}^* - \boldsymbol{\theta}^{(j-1)})$，$h(.)$ 关于原点对称

Proposal density

定理

假设预选分布 $g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})$ 满足一定正则条件，则 *M-H* 抽样得到的序列收敛于稳定分布

1. 随机漫步 (Random walk chains) 链
   - 预选密度具有如下函数形式：$g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)}) = h(\boldsymbol{\theta}^* - \boldsymbol{\theta}^{(j-1)})$，$h(.)$ 关于原点对称
   - $\alpha = \min\left\{1, \frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}^{(j-1)}|\mathbf{y})}\right\}$

Proposal density

定理

假设预选分布 $g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})$ 满足一定正则条件，则 *M-H* 抽样得到的序列收敛于稳定分布

1. 随机漫步 (Random walk chains) 链
   - 预选密度具有如下函数形式：$g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)}) = h(\boldsymbol{\theta}^* - \boldsymbol{\theta}^{(j-1)})$，$h(.)$ 关于原点对称
   - $\alpha = \min\left\{1, \frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}^{(j-1)}|\mathbf{y})}\right\}$
2. 独立链 (Independence chains)

Proposal density

定理

假设预选分布 $g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})$ 满足一定正则条件，则 *M-H* 抽样得到的序列收敛于稳定分布

1. 随机漫步 (Random walk chains) 链
    - 预选密度具有如下函数形式：$g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)}) = h(\boldsymbol{\theta}^* - \boldsymbol{\theta}^{(j-1)})$，$h(.)$ 关于原点对称
    - $\alpha = \min\left\{1, \frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}^{(j-1)}|\mathbf{y})}\right\}$
2. 独立链 (Independence chains)
    - $g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)}) = g(\boldsymbol{\theta}^*)$，与当前序列值无关

Proposal density

定理

假设预选分布 $g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)})$ 满足一定正则条件，则 *M-H* 抽样得到的序列收敛于稳定分布

1. 随机漫步 (Random walk chains) 链
    - 预选密度具有如下函数形式：$g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)}) = h(\boldsymbol{\theta}^* - \boldsymbol{\theta}^{(j-1)})$，$h(.)$ 关于原点对称
    - $\alpha = \min\left\{1, \frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}^{(j-1)}|\mathbf{y})}\right\}$
2. 独立链 (Independence chains)
    - $g(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(j-1)}) = g(\boldsymbol{\theta}^*)$，与当前序列值无关
    - $\alpha = \min\left\{1, \frac{\omega(\boldsymbol{\theta}^*)}{\omega(\boldsymbol{\theta}^{(j-1)})}\right\}$，其中 $\omega(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{y})/g(\boldsymbol{\theta})$

Random walk proposal

- Proposal:

$$\begin{aligned}\alpha^*|\alpha^{(j-1)} &\sim N(\alpha^{(j-1)}, d_1^2)\\ \beta^*|\beta^{(j-1)} &\sim N(\beta^{(j-1)}, d_2^2)\end{aligned}$$

  - Random walk：$x_t = x_{t-1} + e_t,\ e_t \sim N(0, \sigma^2)$

- 接受 $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^*$ 的概率

$$\alpha = \min\left\{1,\ \frac{\pi(\boldsymbol{\theta}^*)f(\mathbf{y}|\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^{(j-1)})f(\mathbf{y}|\boldsymbol{\theta}^{(j-1)})}\right\}$$

例 1：二维正态分布的 Gibbs 抽样

对二元正态分布

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$$

其条件分布为

$$Y|X = x \sim N\big(\mu_1 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), (1 - \rho^2)\sigma_2^2\big)$$

$$X|Y = y \sim N\big(\mu_2 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2), (1 - \rho^2)\sigma_1^2\big)$$

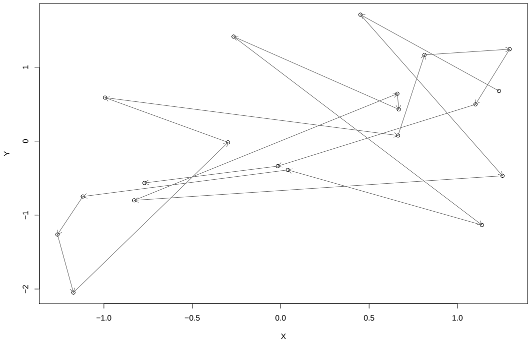
因此可以模拟出联合正态分布的样本：取初始值 $x^{(0)}, y^{(0)}$，然后依次产生 $x^{(k)} \sim \phi(x|y^{(k-1)}$, $y^{(k)} \sim \phi(y|x^{(k)})$。

例 1：二维正态分布的 Gibbs 抽样的 R 代码

```
rbinormal <- function(n, mu1, mu2, sigma1, sigma2, rho) {
  # initialize
  x <- rnorm(1, mu1, sigma1)
  y <- rnorm(1, mu2, sigma2)
  xy <- matrix(nrow = n, ncol = 2, dimnames = list(NULL,
                                                   c("X", "Y")))
  # sample from conditional distributions
  for (i in 1:n) {
    x <- rnorm(1, mu2 + sigma1/sigma2 * rho * (y - mu2),
    sqrt(1 - rho^2) * sigma1)
    y <- rnorm(1, mu1 + sigma2/sigma1 * rho * (x - mu1),
    sqrt(1 - rho^2) * sigma2)
    xy[i, ] <- c(x, y)
  }
  xy
}
```

例 1：Gibbs 抽样的前 20 个样本

```
set.seed(123); n<-20
z<-rbinormal(n,mu1=0,mu2=0,sigma1=1,sigma2=1,rho=0.5)
plot(z)
arrows(z[-n, 1], z[-n, 2], z[-1, 1], z[-1, 2],
                        length = 0.1, col = "gray40")
```



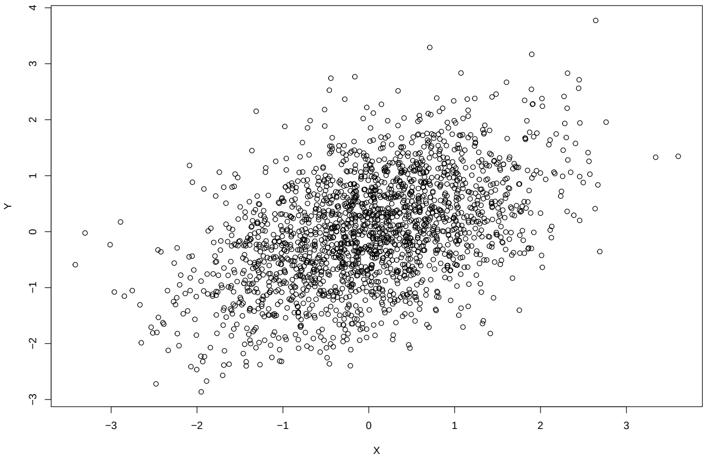
例 1：前 5000 个样本的均值、标准差和相关矩阵

抽取 $N_2(0,1,0,1,0.5)$ 的 5000 个随机样本。

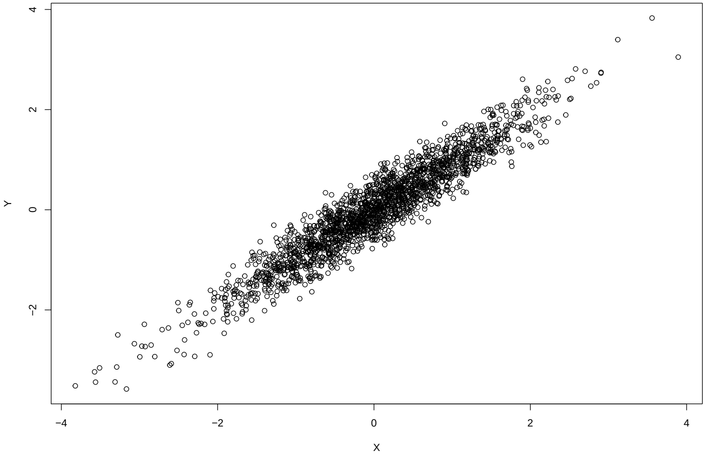
```
> z <- rbinormal(5000, 0, 0, 1, 1, 0.5)
> apply(z, 2, mean) # sample mean
           X            Y
-0.004524327 -0.007134536
> apply(z, 2, sd) # sample sd
       X        Y
1.007888 1.004023
> cor(z) # sample correlation
          X         Y
X 1.0000000 0.5039506
Y 0.5039506 1.0000000
```

例 1：前 2000 个样本散点图（相关系数 =0.5）

```
z <- rbinormal(2000, 0, 0, 1, 1, 0.5)
plot(z)
```



例 1：前 2000 个样本散点图（相关系数 =0.95）



本章小结

Monte Carlo 方法

1. 直接抽样法
   1. Inverse CDF Method
      - 均匀分布 $\rightarrow$ 分布函数的逆函数变换
   2. Rejection Sampling
      - 找一个容易抽取样本的覆盖函数 $g(\theta)$，使得 $p(\theta|y) \le cg(\theta)$，计算接受概率 r，由 $g(\theta)$ 产生的样本，以概率 r 接受为 $p(\theta|y)$ 的样本
   3. Importance Sampling
      - 找一个容易抽取样本的覆盖函数 $g(\theta)$，则 $E[h(\theta|y)]$ 等于对 $g(\theta)$ 的样本进行加权平均
2. 马尔可夫链蒙特卡罗法 (MCMC)
   1. Gibbs Sampler
      - 对所有边缘条件后验分布依次抽样
      - 简单
   2. Metropolis-Hastings
      - 选取合适的预选分布作为马氏链的转移分布
      - Gibbs 是 MH 的特例