

# 贝 叶 斯 统 计

姓名：韦 来 生 编 著

高 等 教 育 出 版 社

第一版，2016

# 贝 叶 斯 统 计

姓名：韦 来 生 编 著

高 等 教 育 出 版 社

第一版，2016

## 第6章 贝叶斯计算方法

### 1 6.1 引言

### 2 6.2 蒙特卡洛抽样方法

- 6.2.1 引言
- 6.2.2 蒙特卡洛抽样方法
- 6.2.2 蒙特卡洛重要性抽样方法

### 3 6.3 马尔科夫链蒙特卡洛方法

- 6.3.1 引言
- 6.3.2 马尔科夫链的定义及性质
- 6.3.3 马氏链的极限定理

### 4 6.4 MCMC 方法简介

- 6.4.1 MCMC 实施中的若干术语
- 6.4.2 利用 MCMC 的输出结果描述目标后验分布
- 6.4.3 MCMC 方法收敛性的诊断

### 5 6.5 Metropolis-Hastings 算法

- 6.5.1 引言
- 6.5.2 M-H 抽样方法
- 6.5.3 M-H 抽样方法的几个变种

### 6 6.6 Gibbs 抽样方法

- 6.6.1 引言
- 6.6.2 Gibbs 抽样方法

### 7 6.7 R 与 WinBUGS 软件简介

## 1 6.1 引言

## 2 6.2 蒙特卡洛抽样方法

- 6.2.1 引言
- 6.2.2 蒙特卡洛抽样方法
- 6.2.2 蒙特卡洛重要性抽样方法

## 3 6.3 马尔科夫链蒙特卡洛方法

- 6.3.1 引言
- 6.3.2 马尔科夫链的定义及性质
- 6.3.3 马氏链的极限定理

## 4 6.4 MCMC 方法简介

- 6.4.1 MCMC 实施中的若干术语
- 6.4.2 利用 MCMC 的输出结果描述目标后验分布
- 6.4.3 MCMC 方法收敛性的诊断

## 5 6.5 Metropolis-Hastings 算法

- 6.5.1 引言
- 6.5.2 M-H 抽样方法
- 6.5.3 M-H 抽样方法的几个变种

## 6 6.6 Gibbs 抽样方法

# 引言

- 在贝叶斯分析中常常需要计算后验分布的期望、方差、分位数或众数等数字特征. 如果先验分布不是共轭先验, 那么后验分布往往不再是标准的分布. 因此需要计算的后验分布数字特征往往没有显式表达, 这就需要一些特殊的计算方法.
- 例 6.1.1** 假设随机变量  $X \sim N(\theta, \sigma^2)$ , 其中  $\sigma^2$  已知. 若选取  $\theta$  的先验分布为柯西分布  $C(\mu, \tau)$ , 则  $\theta$  的后验分布为

$$\pi(\theta|x) \propto \exp\{-(\theta - x)^2/(2\sigma^2)\}(\tau^2 + (\theta - \mu)^2)^{-1}.$$

因此后验期望和后验方差为

$$E^\pi(\theta|x) = \frac{\int_{-\infty}^{\infty} \theta \exp\{-(\theta - x)^2/(2\sigma^2)\}(\tau^2 + (\theta - \mu)^2)^{-1} d\theta}{\int_{-\infty}^{\infty} \exp\{-(\theta - x)^2/(2\sigma^2)\}(\tau^2 + (\theta - \mu)^2)^{-1} d\theta}$$

$$V^\pi(\theta|x) = \frac{\int_{-\infty}^{\infty} \theta^2 \exp\{-(\theta - x)^2/(2\sigma^2)\}(\tau^2 + (\theta - \mu)^2)^{-1} d\theta}{\int_{-\infty}^{\infty} \exp\{-(\theta - x)^2/(2\sigma^2)\}(\tau^2 + (\theta - \mu)^2)^{-1} d\theta} - (E^\pi(\theta|x))^2.$$

# 引言

- 在贝叶斯分析中常常需要计算后验分布的期望、方差、分位数或众数等数字特征. 如果先验分布不是共轭先验, 那么后验分布往往不再是标准的分布. 因此需要计算的后验分布数字特征往往没有显式表达, 这就需要一些特殊的计算方法.
- 例 6.1.1** 假设随机变量  $X \sim N(\theta, \sigma^2)$ , 其中  $\sigma^2$  已知. 若选取  $\theta$  的先验分布为柯西分布  $C(\mu, \tau)$ , 则  $\theta$  的后验分布为

$$\pi(\theta|x) \propto \exp\{-(\theta - x)^2/(2\sigma^2)\}(\tau^2 + (\theta - \mu)^2)^{-1}.$$

因此后验期望和后验方差为

$$E^\pi(\theta|x) = \frac{\int_{-\infty}^{\infty} \theta \exp\{-(\theta - x)^2/(2\sigma^2)\}(\tau^2 + (\theta - \mu)^2)^{-1} d\theta}{\int_{-\infty}^{\infty} \exp\{-(\theta - x)^2/(2\sigma^2)\}(\tau^2 + (\theta - \mu)^2)^{-1} d\theta}$$
$$V^\pi(\theta|x) = \frac{\int_{-\infty}^{\infty} \theta^2 \exp\{-(\theta - x)^2/(2\sigma^2)\}(\tau^2 + (\theta - \mu)^2)^{-1} d\theta}{\int_{-\infty}^{\infty} \exp\{-(\theta - x)^2/(2\sigma^2)\}(\tau^2 + (\theta - \mu)^2)^{-1} d\theta} - (E^\pi(\theta|x))^2.$$

## 计算上更困难的一个例子

- 显然上面这两个积分都没有显式解，但是各种数值积分方法可以被用来高效地逼近这两个积分. 下面的例子给出了一个更困难的问题.
- 例 6.1.2 假设  $X_1, \dots, X_k$  为独立的泊松 (poisson) 随机变量，且  $X_i \sim P(\theta_i), i = 1, \dots, k$ . 如果  $\theta_i$  的先验分布为通过其对数的联合分布给出：

$$\boldsymbol{\nu} = (\log(\theta_1), \dots, \log(\theta_k))' \sim N(\mu \mathbf{1}_k, \tau^2 \{(1 - \rho) \mathbf{I}_k + \rho \mathbf{J}_k\}),$$

其中  $\mathbf{1}_k$  为元素是 1 的  $k$  维列向量， $\mathbf{I}_k$  为  $k$  阶单位阵， $\mathbf{J}_k$  的  $k$  阶元素全为 1 的方阵， $\mu, \tau^2, \rho$  为已知的常数.

- 则由样本分布

$$f(\mathbf{x}|\boldsymbol{\nu}) = \exp\left\{-\sum_{i=1}^k (e^{\nu_i} - \nu_i x_i)\right\} / \prod_{i=1}^k x_i!$$

## 计算上更困难的一个例子

- 显然上面这两个积分都没有显式解，但是各种数值积分方法可以被用来高效地逼近这两个积分。下面的例子给出了一个更困难的问题。
- 例 6.1.2** 假设  $X_1, \dots, X_k$  为独立的泊松 (poisson) 随机变量，且  $X_i \sim P(\theta_i), i = 1, \dots, k$ . 如果  $\theta_i$  的先验分布为通过其对数的联合分布给出：

$$\boldsymbol{\nu} = (\log(\theta_1), \dots, \log(\theta_k))' \sim N(\mu \mathbf{1}_k, \tau^2 \{(1 - \rho) \mathbf{I}_k + \rho \mathbf{J}_k\}),$$

其中  $\mathbf{1}_k$  为元素是 1 的  $k$  维列向量， $\mathbf{I}_k$  为  $k$  阶单位阵， $\mathbf{J}_k$  的  $k$  阶元素全为 1 的方阵， $\mu, \tau^2, \rho$  为已知的常数。

- 则由样本分布

$$f(\mathbf{x}|\boldsymbol{\nu}) = \exp\left\{-\sum_{i=1}^k (e^{\nu_i} - \nu_i x_i)\right\} / \prod_{i=1}^k x_i!$$



## 计算上更困难的一个例子

- 显然上面这两个积分都没有显式解，但是各种数值积分方法可以被用来高效地逼近这两个积分。下面的例子给出了一个更困难的问题。
- 例 6.1.2** 假设  $X_1, \dots, X_k$  为独立的泊松 (poisson) 随机变量，且  $X_i \sim P(\theta_i), i = 1, \dots, k$ . 如果  $\theta_i$  的先验分布为通过其对数的联合分布给出：

$$\boldsymbol{\nu} = (\log(\theta_1), \dots, \log(\theta_k))' \sim N(\mu \mathbf{1}_k, \tau^2 \{(1 - \rho) \mathbf{I}_k + \rho \mathbf{J}_k\}),$$

其中  $\mathbf{1}_k$  为元素是 1 的  $k$  维列向量， $\mathbf{I}_k$  为  $k$  阶单位阵， $\mathbf{J}_k$  的  $k$  阶元素全为 1 的方阵， $\mu, \tau^2, \rho$  为已知的常数。

- 则由样本分布

$$f(\mathbf{x}|\boldsymbol{\nu}) = \exp\left\{-\sum_{i=1}^k (e^{\nu_i} - \nu_i x_i)\right\} / \prod_{i=1}^k x_i!$$

## 计算上更困难的一个例子

- 以及先验分布

$$\pi(\boldsymbol{\nu}) \propto \exp\left\{-\frac{1}{2\tau^2}(\boldsymbol{\nu} - \mu\mathbf{1}_k)'[(1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k]^{-1}(\boldsymbol{\nu} - \mu\mathbf{1}_k)\right\}$$

- 可以得到后验分布

$$\begin{aligned}\pi(\boldsymbol{\nu}|\mathbf{x}) \propto g(\boldsymbol{\nu}|\mathbf{x}) = \exp\left\{-\sum_{i=1}^k(e^{\nu_i} - \nu_i x_i)\right. \\ \left.-\frac{1}{2\tau^2}(\boldsymbol{\nu} - \mu\mathbf{1}_k)'[(1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k]^{-1}(\boldsymbol{\nu} - \mu\mathbf{1}_k)\right\}.\end{aligned}$$

- 因此如果感兴趣的是  $\theta_j$  的后验均值，则需要计算

$$E^\pi(\theta_j|\mathbf{x}) = E^\pi(e^{\nu_j}|\mathbf{x}) = \frac{\int_{\mathbf{R}^k} \exp(\nu_j) g(\boldsymbol{\nu}|\mathbf{x}) d\boldsymbol{\nu}}{\int_{\mathbf{R}^k} g(\boldsymbol{\nu}|\mathbf{x}) d\boldsymbol{\nu}},$$

## 计算上更困难的一个例子

- 以及先验分布

$$\pi(\boldsymbol{\nu}) \propto \exp\left\{-\frac{1}{2\tau^2}(\boldsymbol{\nu} - \mu\mathbf{1}_k)'[(1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k]^{-1}(\boldsymbol{\nu} - \mu\mathbf{1}_k)\right\}$$

- 可以得到后验分布

$$\begin{aligned}\pi(\boldsymbol{\nu}|\mathbf{x}) \propto g(\boldsymbol{\nu}|\mathbf{x}) = \exp\left\{-\sum_{i=1}^k(e^{\nu_i} - \nu_i x_i)\right. \\ \left.-\frac{1}{2\tau^2}(\boldsymbol{\nu} - \mu\mathbf{1}_k)'[(1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k]^{-1}(\boldsymbol{\nu} - \mu\mathbf{1}_k)\right\}.\end{aligned}$$

- 因此如果感兴趣的是  $\theta_j$  的后验均值，则需要计算

$$E^\pi(\theta_j|\mathbf{x}) = E^\pi(e^{\nu_j}|\mathbf{x}) = \frac{\int_{\mathbf{R}^k} \exp(\nu_j) g(\boldsymbol{\nu}|\mathbf{x}) d\boldsymbol{\nu}}{\int_{\mathbf{R}^k} g(\boldsymbol{\nu}|\mathbf{x}) d\boldsymbol{\nu}},$$

## 计算上更困难的一个例子

- 这是两个  $k$  重积分的比值. 当  $k$  越大时就越难处理, 而数值积分方法在这种场合下不再是有效的方法, 这种问题也就是常称的维数灾难问题. 数值逼近中的误差随着维数  $k$  的幂次增加, 最终导致算法失效. 因此数值积分方法在一维和二维积分以外的场合下不是优先使用的方法.
- 由上述两个例子, 可以看到在贝叶斯统计计算问题中, 与后验分布有关的一些积分是很难用数值方法去计算的, 尤其是在高维的情形. 本章后面将介绍的马尔科夫链蒙特卡洛方法提供了一个有效的途径去处理这类问题.

## 计算上更困难的一个例子

- 这是两个  $k$  重积分的比值. 当  $k$  越大时就越难处理, 而数值积分方法在这种场合下不再是有效的方法, 这种问题也就是常称的维数灾难问题. 数值逼近中的误差随着维数  $k$  的幂次增加, 最终导致算法失效. 因此数值积分方法在一维和二维积分以外的场合下不是优先使用的方法.
- 由上述两个例子, 可以看到在贝叶斯统计计算问题中, 与后验分布有关的一些积分是很难用数值方法去计算的, 尤其是在高维的情形. 本章后面将介绍的马尔科夫链蒙特卡洛方法提供了一个有效的途径去处理这类问题.

## 1 6.1 引言

## 2 6.2 蒙特卡洛抽样方法

### • 6.2.1 引言

### • 6.2.2 蒙特卡洛抽样方法

### • 6.2.2 蒙特卡洛重要性抽样方法

## 3 6.3 马尔科夫链蒙特卡洛方法

### • 6.3.1 引言

### • 6.3.2 马尔科夫链的定义及性质

### • 6.3.3 马氏链的极限定理

## 4 6.4 MCMC 方法简介

### • 6.4.1 MCMC 实施中的若干术语

### • 6.4.2 利用 MCMC 的输出结果描述目标后验分布

### • 6.4.3 MCMC 方法收敛性的诊断

## 5 6.5 Metropolis-Hastings 算法

### • 6.5.1 引言

### • 6.5.2 M-H 抽样方法

### • 6.5.3 M-H 抽样方法的几个变种

## 6 6.6 Gibbs 抽样方法

# 引言

- 在贝叶斯统计分析问题中常常要计算后验分布的一些数字特征，如后验期望、后验方差、后验众数和后验分位数等。我们首先将这些数字特征统一在一个积分表达式中，然后研究这个积分怎样计算的问题。
- 设  $p(\mathbf{x}|\theta)$  表示样本的概率函数 (或称为参数  $\theta$  的似然函数)， $\pi(\theta)$  为  $\theta$  的先验分布，则  $\theta$  的后验分布为

$$\pi(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)\pi(\theta)}{\int p(\mathbf{x}|\theta)\pi(\theta)d\theta}$$

# 引言

- 在贝叶斯统计分析问题中常常要计算后验分布的一些数字特征，如后验期望、后验方差、后验众数和后验分位数等。我们首先将这些数字特征统一在一个积分表达式中，然后研究这个积分怎样计算的问题。
- 设  $p(\mathbf{x}|\theta)$  表示样本的概率函数 (或称为参数  $\theta$  的似然函数)， $\pi(\theta)$  为  $\theta$  的先验分布，则  $\theta$  的后验分布为

$$\pi(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)\pi(\theta)}{\int p(\mathbf{x}|\theta)\pi(\theta)d\theta}$$



# 引言

- 我们感兴趣的积分是函数  $h(\theta)$  的后验期望

$$E[h(\theta)|\mathbf{x}] = \int h(\theta)\pi(\theta|\mathbf{x})d\theta = \frac{\int h(\theta)p(\mathbf{x}|\theta)\pi(\theta)d\theta}{\int p(\mathbf{x}|\theta)\pi(\theta)d\theta}. \quad (6.2.1)$$

- 当  $h(\theta) = \theta$  或  $h(\theta) = [\theta - E(\theta|\mathbf{x})]^2$  时, 上式分别表示后验期望和后验方差; 当  $h(\theta)$  是损失函数时, 上式表示后验风险.

# 引言

- 我们感兴趣的积分是函数  $h(\theta)$  的后验期望

$$E[h(\theta)|\mathbf{x}] = \int h(\theta)\pi(\theta|\mathbf{x})d\theta = \frac{\int h(\theta)p(\mathbf{x}|\theta)\pi(\theta)d\theta}{\int p(\mathbf{x}|\theta)\pi(\theta)d\theta}. \quad (6.2.1)$$

- 当  $h(\theta) = \theta$  或  $h(\theta) = [\theta - E(\theta|\mathbf{x})]^2$  时, 上式分别表示后验期望和后验方差; 当  $h(\theta)$  是损失函数时, 上式表示后验风险.

## 蒙特卡洛抽样方法的定义

- 如果式 (6.2.1) 所求的期望没有显式表达, 蒙特卡洛抽样方法也是一个可选的有效计算方法. 为了估计总体均值总体方差和总体分位数等数字特征, 可从总体中抽取足够多的样本, 然后使用样本均值、样本方差和样本分位数来估计相应的总体数字特征. 大数定律保证了所得估计量是相合估计.
- 若可从后验分布  $\pi(\theta|\mathbf{x})$  中产生 *i.i.d.* 观测值  $\theta_1, \theta_2, \dots, \theta_m$ , 则由大数定律可知

$$\bar{h}_m = \frac{1}{m} \sum_{i=1}^m h(\theta_i) \quad (6.2.2)$$

几乎处处收敛到感兴趣的量  $E[h(\theta)|\mathbf{x}]$ . 这一结果保证了在样本量  $m$  足够大时可以使用  $\bar{h}_m$  作为  $E[h(\theta)|\mathbf{x}]$  的估计. 而估计量 (6.2.2) 被称为积分 (6.2.1) 的蒙特卡洛逼近. 这种用估计量 (6.2.2) 去逼近蒙特卡洛积分 (6.2.1) 的方法被称为蒙特卡洛抽样方法.

## 蒙特卡洛抽样方法的定义

- 如果式 (6.2.1) 所求的期望没有显式表达, 蒙特卡洛抽样方法也是一个可选的有效计算方法. 为了估计总体均值总体方差和总体分位数等数字特征, 可从总体中抽取足够多的样本, 然后使用样本均值、样本方差和样本分位数来估计相应的总体数字特征. 大数定律保证了所得估计量是相合估计.
- 若可从后验分布  $\pi(\theta|\mathbf{x})$  中产生 *i.i.d.* 观测值  $\theta_1, \theta_2, \dots, \theta_m$ , 则由大数定律可知

$$\bar{h}_m = \frac{1}{m} \sum_{i=1}^m h(\theta_i) \quad (6.2.2)$$

几乎处处收敛到感兴趣的量  $E[h(\theta)|\mathbf{x}]$ . 这一结果保证了在样本量  $m$  足够大时可以使用  $\bar{h}_m$  作为  $E[h(\theta)|\mathbf{x}]$  的估计. 而估计量 (6.2.2) 被称为积分 (6.2.1) 的蒙特卡洛逼近. 这种用估计量 (6.2.2) 去逼近蒙特卡洛积分 (6.2.1) 的方法被称为蒙特卡洛抽样方法.

## 蒙特卡洛抽样方法的定义

- 如果式 (6.2.1) 所求的期望没有显式表达, 蒙特卡洛抽样方法也是一个可选的有效计算方法. 为了估计总体均值总体方差和总体分位数等数字特征, 可从总体中抽取足够多的样本, 然后使用样本均值、样本方差和样本分位数来估计相应的总体数字特征. 大数定律保证了所得估计量是相合估计.
- 若可从后验分布  $\pi(\theta|\mathbf{x})$  中产生 *i.i.d.* 观测值  $\theta_1, \theta_2, \dots, \theta_m$ , 则由大数定律可知

$$\bar{h}_m = \frac{1}{m} \sum_{i=1}^m h(\theta_i) \quad (6.2.2)$$

几乎处处收敛到感兴趣的量  $E[h(\theta)|\mathbf{x}]$ . 这一结果保证了在样本量  $m$  足够大时可以使用  $\bar{h}_m$  作为  $E[h(\theta)|\mathbf{x}]$  的估计. 而估计量 (6.2.2) 被称为积分 (6.2.1) 的蒙特卡洛逼近. 这种用估计量 (6.2.2) 去逼近蒙特卡洛积分 (6.2.1) 的方法被称为蒙特卡洛抽样方法.

# 蒙特卡洛抽样方法

- 为给出估计的精度或者逼近的误差程度，我们可以使用类似的办法计算  $\hat{h}_m$  标准差. 如果后验方差  $V^\pi[h(\theta)|\mathbf{x}]$  是有限的，则  $\text{Var}(\bar{h}_m) = V^\pi[h(\theta)|\mathbf{x}]/m$ , 样本标准差是

$$s_m = \left\{ \frac{1}{m-1} \sum_{i=1}^m (h(\theta_i) - \bar{h}_m)^2 \right\}^{1/2}.$$

因此可以用  $s_m/\sqrt{m}$  作为  $\bar{h}_m$  标准差的一个估计.

- 用蒙特卡洛逼近方法求积分 (6.2.1) 的另一种方法如下：由于  $\pi(\theta)$  是概率密度函数，从  $\pi(\theta)$  生成 *i.i.d.* 观测值  $\theta_1, \theta_2, \dots, \theta_m$ , 对式 (6.2.1) 最右边一项的分子和分母中的两个积分  $\int h(\theta)p(\mathbf{x}|\theta)\pi(\theta)d\theta$  和  $\int p(\mathbf{x}|\theta)\pi(\theta)d\theta$ , 分别用它们的蒙特卡洛逼近代替，得到  $E[h(\theta)|\mathbf{x}]$  的估计如下：

$$\tilde{h}_m = \sum_{i=1}^m h(\theta_i)p(\mathbf{x}|\theta_i) / \sum_{i=1}^m p(\mathbf{x}|\theta_i).$$

# 蒙特卡洛抽样方法

- 为给出估计的精度或者逼近的误差程度，我们可以使用类似的办法计算  $\hat{h}_m$  标准差. 如果后验方差  $V^\pi[h(\theta)|\mathbf{x}]$  是有限的，则  $\text{Var}(\bar{h}_m) = V^\pi[h(\theta)|\mathbf{x}]/m$ , 样本标准差是

$$s_m = \left\{ \frac{1}{m-1} \sum_{i=1}^m (h(\theta_i) - \bar{h}_m)^2 \right\}^{1/2}.$$

因此可以用  $s_m/\sqrt{m}$  作为  $\bar{h}_m$  标准差的一个估计.

- 用蒙特卡洛逼近方法求积分 (6.2.1) 的另一种方法如下：由于  $\pi(\theta)$  是概率密度函数，从  $\pi(\theta)$  生成 *i.i.d.* 观测值  $\theta_1, \theta_2, \dots, \theta_m$ , 对式 (6.2.1) 最右边一项的分子和分母中的两个积分  $\int h(\theta)p(\mathbf{x}|\theta)\pi(\theta)d\theta$  和  $\int p(\mathbf{x}|\theta)\pi(\theta)d\theta$ , 分别用它们的蒙特卡洛逼近代替，得到  $E[h(\theta)|\mathbf{x}]$  的估计如下：

$$\tilde{h}_m = \sum_{i=1}^m h(\theta_i)p(\mathbf{x}|\theta_i) / \sum_{i=1}^m p(\mathbf{x}|\theta_i).$$

# 蒙特卡洛抽样方法

- 如果需要得到后验期望  $E[h(\theta)|\mathbf{x}]$  的可信区间, 则由

$$\frac{\sqrt{m}(\bar{h}_m - E[h(\theta)|\mathbf{x}])}{s_m} \Bigg| \mathbf{x} \xrightarrow{\mathcal{L}} N(0, 1), \quad \text{当 } n \rightarrow \infty$$

容易得到后验期望  $E[h(\theta)|\mathbf{x}]$  一个渐近水平近似为  $1 - \alpha$  的可信区间

$$\left[ \bar{h}_m - \frac{s_m u_{\alpha/2}}{\sqrt{m}}, \bar{h}_m + \frac{s_m u_{\alpha/2}}{\sqrt{m}} \right].$$

这里  $u_{\alpha/2}$  表示标准正态分布的上侧  $\alpha/2$  分位数.

- 以上的讨论表明, 如果我们需要计算后验均值, 则可以通过从后验分布 (或相关分布) 中产生 *i.i.d* 样本, 然后计算相应的样本均值作为估计. 但这种方法很少能直接使用, 因为大多数情况下后验分布都不是标准分布而难以从中抽样.



# 蒙特卡洛抽样方法

- 如果需要得到后验期望  $E[h(\theta)|\mathbf{x}]$  的可信区间, 则由

$$\frac{\sqrt{m}(\bar{h}_m - E[h(\theta)|\mathbf{x}])}{s_m} \Bigg| \mathbf{x} \xrightarrow{\mathcal{L}} N(0, 1), \quad \text{当 } n \rightarrow \infty$$

容易得到后验期望  $E[h(\theta)|\mathbf{x}]$  一个渐近水平近似为  $1 - \alpha$  的可信区间

$$\left[ \bar{h}_m - \frac{s_m u_{\alpha/2}}{\sqrt{m}}, \bar{h}_m + \frac{s_m u_{\alpha/2}}{\sqrt{m}} \right].$$

这里  $u_{\alpha/2}$  表示标准正态分布的上侧  $\alpha/2$  分位数.

- 以上的讨论表明, 如果我们需要计算后验均值, 则可以通过从后验分布 (或相关分布) 中产生 *i.i.d* 样本, 然后计算相应的样本均值作为估计. 但这种方法很少能直接使用, 因为大多数情况下后验分布都不是标准分布而难以从中抽样.

## 蒙特卡洛重要性抽样方法的引入

- 例 6.2.1 (续例 6.1.1) 由于

$$E^{\pi}(\theta|x) = \frac{\int_0^{\infty} \theta \exp\{-(\theta - x)^2/(2\sigma^2)\}(\tau^2 + (\theta - \mu)^2)^{-1} d\theta}{\int_0^{\infty} \exp\{-(\theta - x)^2/(2\sigma^2)\}(\tau^2 + (\theta - \mu)^2)^{-1} d\theta},$$

因此  $E^{\pi}(\theta|x)$  为在正态分布  $N(x, \sigma^2)$  下

$$h_1(\theta) = \theta(\tau^2 + (\theta - \mu)^2)^{-1} \quad \text{和} \quad h_2(\theta) = (\tau^2 + (\theta - \mu)^2)^{-1}$$

两者的期望之比. 于是由前面的讨论知道, 如果  $\theta_1, \dots, \theta_m$  为从正态分布  $N(x, \sigma^2)$  中产生的 *i.i.d* 样本, 则  $E^{\pi}(\theta|x)$  的估计量为

$$\widehat{E^{\pi}(\theta|x)} = \frac{\sum_{i=1}^m \theta_i (\tau^2 + (\theta_i - \mu)^2)^{-1}}{\sum_{i=1}^m (\tau^2 + (\theta_i - \mu)^2)^{-1}}.$$

- 问题并没有被完美解决. 由于从  $N(x, \sigma^2)$  中抽取的  $\theta$  集中在  $x$  附近, 并没有充分反应出柯西先验对后验分布的贡献, 而应当有显著的样本比例来自于后验分布的尾部.

## 蒙特卡洛重要性抽样方法的引入

- 例 6.2.1 (续例 6.1.1) 由于

$$E^{\pi}(\theta|x) = \frac{\int_0^{\infty} \theta \exp\{-(\theta - x)^2/(2\sigma^2)\}(\tau^2 + (\theta - \mu)^2)^{-1} d\theta}{\int_0^{\infty} \exp\{-(\theta - x)^2/(2\sigma^2)\}(\tau^2 + (\theta - \mu)^2)^{-1} d\theta},$$

因此  $E^{\pi}(\theta|x)$  为在正态分布  $N(x, \sigma^2)$  下

$$h_1(\theta) = \theta(\tau^2 + (\theta - \mu)^2)^{-1} \quad \text{和} \quad h_2(\theta) = (\tau^2 + (\theta - \mu)^2)^{-1}$$

两者的期望之比. 于是由前面的讨论知道, 如果  $\theta_1, \dots, \theta_m$  为从正态分布  $N(x, \sigma^2)$  中产生的 *i.i.d* 样本, 则  $E^{\pi}(\theta|x)$  的估计量为

$$\widehat{E^{\pi}(\theta|x)} = \frac{\sum_{i=1}^m \theta_i (\tau^2 + (\theta_i - \mu)^2)^{-1}}{\sum_{i=1}^m (\tau^2 + (\theta_i - \mu)^2)^{-1}}.$$

- 问题并没有被完美解决. 由于从  $N(x, \sigma^2)$  中抽取的  $\theta$  集中在  $x$  附近, 并没有充分反应出柯西先验对后验分布的贡献, 而应当有显著的样本比例来自于后验分布的尾部.

## 蒙特卡洛重要性抽样方法的引入

- 因此如果把  $E^\pi(\theta|x)$  视为

$$\theta \exp\{-(\theta - x)^2/(2\sigma^2)\} \text{ 和 } \exp\{-(\theta - x)^2/(2\sigma^2)\}$$

在柯西分布下的期望之比, 则一个合适的估计量为

$$\widehat{E^\pi(\theta|x)} = \frac{\sum_{i=1}^m \theta_i \exp\{-(\theta_i - x)^2/(2\sigma^2)\}}{\sum_{i=1}^m \exp\{-(\theta_i - x)^2/(2\sigma^2)\}},$$

其中  $\theta_1, \dots, \theta_m$  为从柯西分布  $C(\mu, \tau)$  中抽取的 *i.i.d* 样本.

- 这样也还是没有令人满意的解决问题. 由于后验分布并没有像柯西分布那样程度的尾部, 因此相对于后验分布的中心而言, 这种做法从尾部抽取了过多的样本. 这样就导致收敛速度变慢且在固定  $m$  时逼近误差增大. 理想地, 为了达到满意的逼近应该直接从后验分布本身抽样. 为此, 上述抽样方法的一个变种—蒙特卡洛重要性抽样方法被提出.

## 蒙特卡洛重要性抽样方法的引入

- 因此如果把  $E^\pi(\theta|x)$  视为

$$\theta \exp\{-(\theta - x)^2/(2\sigma^2)\} \text{ 和 } \exp\{-(\theta - x)^2/(2\sigma^2)\}$$

在柯西分布下的期望之比, 则一个合适的估计量为

$$\widehat{E^\pi(\theta|x)} = \frac{\sum_{i=1}^m \theta_i \exp\{-(\theta_i - x)^2/(2\sigma^2)\}}{\sum_{i=1}^m \exp\{-(\theta_i - x)^2/(2\sigma^2)\}},$$

其中  $\theta_1, \dots, \theta_m$  为从柯西分布  $C(\mu, \tau)$  中抽取的 *i.i.d* 样本.

- 这样也还是没有令人满意的解决问题. 由于后验分布并没有像柯西分布那样程度的尾部, 因此相对于后验分布的中心而言, 这种做法从尾部抽取了过多的样本. 这样就导致收敛速度变慢且在固定  $m$  时逼近误差增大. 理想地, 为了达到满意的逼近应该直接从后验分布本身抽样. 为此, 上述抽样方法的一个变种—蒙特卡洛重要性抽样方法被提出.

# 蒙特卡洛重要性抽样方法

- 设  $g(\theta)$  是概率密度函数，它的支撑集包含  $h(\theta)p(\mathbf{x}|\theta)\pi(\theta)$  的支撑集。若从后验分布直接抽样很困难，而从与后验分布非常接近的分布  $g$  中抽样比较容易，则我们可以将式 (6.2.1) 最右边一项的分子和分母中的积分表示为

$$\begin{aligned}\int h(\theta)p(\mathbf{x}|\theta)\pi(\theta)d\theta &= \int \left\{ \frac{h(\theta)p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\} \cdot g(\theta)d\theta \\ &= E_g \left\{ \frac{h(\theta)p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\} \quad (6.2.4)\end{aligned}$$

- 和

$$\begin{aligned}\int p(\mathbf{x}|\theta)\pi(\theta)d\theta &= \int \left\{ \frac{p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\} \cdot g(\theta)d\theta \\ &= E_g \left\{ \frac{p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\}, \quad (6.2.5)\end{aligned}$$

# 蒙特卡洛重要性抽样方法

- 设  $g(\theta)$  是概率密度函数，它的支撑集包含  $h(\theta)p(\mathbf{x}|\theta)\pi(\theta)$  的支撑集。若从后验分布直接抽样很困难，而从与后验分布非常接近的分布  $g$  中抽样比较容易，则我们可以将式 (6.2.1) 最右边一项的分子和分母中的积分表示为

$$\begin{aligned}\int h(\theta)p(\mathbf{x}|\theta)\pi(\theta)d\theta &= \int \left\{ \frac{h(\theta)p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\} \cdot g(\theta)d\theta \\ &= E_g \left\{ \frac{h(\theta)p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\} \quad (6.2.4)\end{aligned}$$

- 和

$$\begin{aligned}\int p(\mathbf{x}|\theta)\pi(\theta)d\theta &= \int \left\{ \frac{p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\} \cdot g(\theta)d\theta \\ &= E_g \left\{ \frac{p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\}, \quad (6.2.5)\end{aligned}$$

# 蒙特卡洛重要性抽样方法

- 设  $g(\theta)$  是概率密度函数，它的支撑集包含  $h(\theta)p(\mathbf{x}|\theta)\pi(\theta)$  的支撑集。若从后验分布直接抽样很困难，而从与后验分布非常接近的分布  $g$  中抽样比较容易，则我们可以将式 (6.2.1) 最右边一项的分子和分母中的积分表示为

$$\begin{aligned}\int h(\theta)p(\mathbf{x}|\theta)\pi(\theta)d\theta &= \int \left\{ \frac{h(\theta)p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\} \cdot g(\theta)d\theta \\ &= E_g \left\{ \frac{h(\theta)p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\} \quad (6.2.4)\end{aligned}$$

- 和

$$\begin{aligned}\int p(\mathbf{x}|\theta)\pi(\theta)d\theta &= \int \left\{ \frac{p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\} \cdot g(\theta)d\theta \\ &= E_g \left\{ \frac{p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\}, \quad (6.2.5)\end{aligned}$$



# 蒙特卡洛重要性抽样方法

- 由式 (6.2.1), (6.2.4) 和 (6.2.5) 可知,  $E[h(\theta)|\mathbf{x}]$  可表示为

$$E[h(\theta)|\mathbf{x}] = E_g \left\{ \frac{h(\theta)p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\} / E_g \left\{ \frac{p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\}$$

- 设  $\theta_1, \theta_2, \dots, \theta_m$  为从分布  $g(\cdot)$  中生成的 *i.i.d.* 观测值, 则由大数定律可知

$$\frac{1}{m} \sum_{i=1}^m h(\theta_i) \omega(\theta_i) \quad \text{和} \quad \frac{1}{m} \sum_{i=1}^m \omega(\theta_i) \quad (6.2.6)$$

几乎处处收敛到由式 (6.2.4) 和 (6.2.5) 给出的  $E_g \{h(\theta)\omega(\theta)\}$  及  $E_g \{\omega(\theta)\}$ , 此处  $\omega(\theta) = p(\mathbf{x}|\theta)\pi(\theta)/g(\theta)$ .

# 蒙特卡洛重要性抽样方法

- 由式 (6.2.1), (6.2.4) 和 (6.2.5) 可知,  $E[h(\theta)|\mathbf{x}]$  可表示为

$$E[h(\theta)|\mathbf{x}] = E_g \left\{ \frac{h(\theta)p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\} / E_g \left\{ \frac{p(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)} \right\}$$

- 设  $\theta_1, \theta_2, \dots, \theta_m$  为从分布  $g(\cdot)$  中生成的 *i.i.d.* 观测值, 则由大数定律可知

$$\frac{1}{m} \sum_{i=1}^m h(\theta_i) \omega(\theta_i) \quad \text{和} \quad \frac{1}{m} \sum_{i=1}^m \omega(\theta_i) \quad (6.2.6)$$

几乎处处收敛到由式 (6.2.4) 和 (6.2.5) 给出的  $E_g \{h(\theta)\omega(\theta)\}$  及  $E_g \{\omega(\theta)\}$ , 此处  $\omega(\theta) = p(\mathbf{x}|\theta)\pi(\theta)/g(\theta)$ .

# 蒙特卡洛重要性抽样方法

- 因此,  $E[h(\theta)|\mathbf{x}]$  另一个估计为

$$\hat{h}_m = \sum_{i=1}^m h(\theta_i) \omega(\theta_i) / \sum_{i=1}^m \omega(\theta_i)$$

由大数定律可知  $\hat{h}_m$  几乎处处收敛到  $E[h(\theta)|\mathbf{x}]$ , 从而知  $\hat{h}_m$  是  $E[h(\theta)|\mathbf{x}]$  的一个蒙特卡洛逼近.

- 式 (6.2.4) 和式 (6.2.5) 中的函数  $g(\theta)$  称为重要性函数. 这种引入重要性函数  $g(\theta)$ , 用式 (6.2.6) 中的两个量分别去逼近  $E_g\{h(\theta)\omega(\theta)\}$  和  $E_g\{\omega(\theta)\}$  的方法称为蒙特卡洛重要性抽样方法.
- 重要性函数  $g$  的应如何选? 一般说来, 它的选择应当满足下列条件: 首先从分布  $g$  抽样要容易、方便, 使模拟容易实现. 其次, 以积分 (6.2.1) 为例, 选择  $g$  尽可能接近后验分布, 以便提高蒙特卡洛逼近的效果.

# 蒙特卡洛重要性抽样方法

- 因此,  $E[h(\theta)|\mathbf{x}]$  另一个估计为

$$\hat{h}_m = \sum_{i=1}^m h(\theta_i) \omega(\theta_i) / \sum_{i=1}^m \omega(\theta_i)$$

由大数定律可知  $\hat{h}_m$  几乎处处收敛到  $E[h(\theta)|\mathbf{x}]$ , 从而知  $\hat{h}_m$  是  $E[h(\theta)|\mathbf{x}]$  的一个蒙特卡洛逼近.

- 式 (6.2.4) 和式 (6.2.5) 中的函数  $g(\theta)$  称为重要性函数. 这种引入重要性函数  $g(\theta)$ , 用式 (6.2.6) 中的两个量分别去逼近  $E_g\{h(\theta)\omega(\theta)\}$  和  $E_g\{\omega(\theta)\}$  的方法称为蒙特卡洛重要性抽样方法.
- 重要性函数  $g$  的应如何选? 一般说来, 它的选择应当满足下列条件: 首先从分布  $g$  抽样要容易、方便, 使模拟容易实现. 其次, 以积分 (6.2.1) 为例, 选择  $g$  尽可能接近后验分布, 以便提高蒙特卡洛逼近的效果.

# 蒙特卡洛重要性抽样方法

- 因此,  $E[h(\theta)|\mathbf{x}]$  另一个估计为

$$\hat{h}_m = \sum_{i=1}^m h(\theta_i) \omega(\theta_i) / \sum_{i=1}^m \omega(\theta_i)$$

由大数定律可知  $\hat{h}_m$  几乎处处收敛到  $E[h(\theta)|\mathbf{x}]$ , 从而知  $\hat{h}_m$  是  $E[h(\theta)|\mathbf{x}]$  的一个蒙特卡洛逼近.

- 式 (6.2.4) 和式 (6.2.5) 中的函数  $g(\theta)$  称为重要性函数. 这种引入重要性函数  $g(\theta)$ , 用式 (6.2.6) 中的两个量分别去逼近  $E_g\{h(\theta)\omega(\theta)\}$  和  $E_g\{\omega(\theta)\}$  的方法称为蒙特卡洛重要性抽样方法.
- 重要性函数  $g$  的应如何选? 一般说来, 它的选择应当满足下列条件: 首先从分布  $g$  抽样要容易、方便, 使模拟容易实现. 其次, 以积分 (6.2.1) 为例, 选择  $g$  尽可能接近后验分布, 以便提高蒙特卡洛逼近的效果.

## 蒙特卡洛重要性抽样的例子

- **例 6.2.2** 假设  $X_1, \dots, X_n$  为从  $N(\theta, \sigma^2)$  中抽取的 *i.i.d* 样本, 其中  $\theta, \sigma^2$  均未知. 取  $\theta$  和  $\sigma^2$  的先验为独立先验, 其中  $\theta$  服从双指数先验分布, 密度为  $\frac{1}{2}\exp\{-|\theta|/2\}$ ,  $\sigma^2$  有先验密度  $(1 + \sigma^2)^{-2}$ . 这两个先验分布都不是标准先验, 但都是从稳健性考虑而选取的.  $\theta$  的后验均值为感兴趣的量, 用蒙特卡洛重要性抽样方法求积分

$$E^\pi(\theta|\mathbf{x}) = \int_{-\infty}^{\infty} \int_0^{\infty} \theta \pi(\theta, \sigma^2|\mathbf{x}) d\theta d\sigma^2.$$

- **解** 由于  $\pi(\theta, \sigma^2|\mathbf{x})$  不是一个标准分布, 我们先来寻找一个离它较近的分布. 记  $\bar{x} = \sum_{i=1}^n x_i/n$ ,  $s_n^2 = \sum_{i=1}^n (x_i - \bar{x})^2/n$ .

## 蒙特卡洛重要性抽样的例子

- **例 6.2.2** 假设  $X_1, \dots, X_n$  为从  $N(\theta, \sigma^2)$  中抽取的 *i.i.d* 样本, 其中  $\theta, \sigma^2$  均未知. 取  $\theta$  和  $\sigma^2$  的先验为独立先验, 其中  $\theta$  服从双指数先验分布, 密度为  $\frac{1}{2}\exp\{-|\theta|/2\}$ ,  $\sigma^2$  有先验密度  $(1 + \sigma^2)^{-2}$ . 这两个先验分布都不是标准先验, 但都是从稳健性考虑而选取的.  $\theta$  的后验均值为感兴趣的量, 用蒙特卡洛重要性抽样方法求积分

$$E^\pi(\theta|\mathbf{x}) = \int_{-\infty}^{\infty} \int_0^{\infty} \theta \pi(\theta, \sigma^2|\mathbf{x}) d\theta d\sigma^2.$$

- **解** 由于  $\pi(\theta, \sigma^2|\mathbf{x})$  不是一个标准分布, 我们先来寻找一个离它较近的分布. 记  $\bar{x} = \sum_{i=1}^n x_i/n$ ,  $s_n^2 = \sum_{i=1}^n (x_i - \bar{x})^2/n$ .

## 蒙特卡洛重要性抽样的例子

- 因此有

$$\begin{aligned}\pi(\theta, \sigma^2 | \mathbf{x}) &\propto (\sigma^2)^{-n/2} \exp \left\{ -\frac{n}{2\sigma^2} [(\theta - \bar{x})^2 + s_n^2] \right\} \cdot \frac{\exp\{-|\theta|\}}{(1 + \sigma^2)^2} \\ &= \frac{[(\theta - \bar{x})^2 + s_n^2]^{n/2+1}}{(\sigma^2)^{(n/2+2)}} \exp \left\{ -\frac{n}{2\sigma^2} [(\theta - \bar{x})^2 + s_n^2] \right\} \\ &\quad \times [(\theta - \bar{x})^2 + s_n^2]^{-(n/2+1)} \exp\{-|\theta|\} \left( \frac{\sigma^2}{1 + \sigma^2} \right)^2 \\ &\propto g_1(\sigma^2 | \theta, \mathbf{x}) \cdot g_2(\theta | \mathbf{x}) \cdot \exp\{-|\theta|\} \left( \frac{\sigma^2}{1 + \sigma^2} \right)^2,\end{aligned}$$

- 其中  $g_1$  为逆伽玛分布  $\Gamma^{-1}(\alpha, \beta)$ , 其中  $\beta = n[(\theta - \bar{x})^2 + s_n^2]/2$ ,  $\alpha = n/2 + 1$ ;  $g_2$  为广义  $t$  分布  $\mathcal{T}_1(n+1, \bar{x}, s_n^2/(n+1))$ .
- 注意到  $\exp(-|\theta|) \left( \frac{\sigma^2}{1 + \sigma^2} \right)^2$  的尾部对  $g_1 \cdot g_2$  没有太大影响, 故可选择  $g(\theta, \sigma^2 | \mathbf{x}) = g_1(\sigma^2 | \theta, \mathbf{x}) \cdot g_2(\theta | \mathbf{x})$  作为重要性函数.



## 蒙特卡洛重要性抽样的例子

- 因此有

$$\begin{aligned}\pi(\theta, \sigma^2 | \mathbf{x}) &\propto (\sigma^2)^{-n/2} \exp \left\{ -\frac{n}{2\sigma^2} [(\theta - \bar{x})^2 + s_n^2] \right\} \cdot \frac{\exp\{-|\theta|\}}{(1 + \sigma^2)^2} \\&= \frac{[(\theta - \bar{x})^2 + s_n^2]^{n/2+1}}{(\sigma^2)^{(n/2+2)}} \exp \left\{ -\frac{n}{2\sigma^2} [(\theta - \bar{x})^2 + s_n^2] \right\} \\&\quad \times [(\theta - \bar{x})^2 + s_n^2]^{-(n/2+1)} \exp\{-|\theta|\} \left( \frac{\sigma^2}{1 + \sigma^2} \right)^2 \\&\propto g_1(\sigma^2 | \theta, \mathbf{x}) \cdot g_2(\theta | \mathbf{x}) \cdot \exp\{-|\theta|\} \left( \frac{\sigma^2}{1 + \sigma^2} \right)^2,\end{aligned}$$

- 其中  $g_1$  为逆伽玛分布  $\Gamma^{-1}(\alpha, \beta)$ , 其中  $\beta = n[(\theta - \bar{x})^2 + s_n^2]/2$ ,  $\alpha = n/2 + 1$ ;  $g_2$  为广义  $t$  分布  $\mathcal{T}_1(n+1, \bar{x}, s_n^2/(n+1))$ .
- 注意到  $\exp(-|\theta|) \left( \frac{\sigma^2}{1 + \sigma^2} \right)^2$  的尾部对  $g_1 \cdot g_2$  没有太大影响, 故可选择  $g(\theta, \sigma^2 | \mathbf{x}) = g_1(\sigma^2 | \theta, \mathbf{x}) \cdot g_2(\theta | \mathbf{x})$  作为重要性函数.

## 蒙特卡洛重要性抽样的例子

- 因此有

$$\begin{aligned}\pi(\theta, \sigma^2 | \mathbf{x}) &\propto (\sigma^2)^{-n/2} \exp \left\{ -\frac{n}{2\sigma^2} [(\theta - \bar{x})^2 + s_n^2] \right\} \cdot \frac{\exp\{-|\theta|\}}{(1 + \sigma^2)^2} \\ &= \frac{[(\theta - \bar{x})^2 + s_n^2]^{n/2+1}}{(\sigma^2)^{(n/2+2)}} \exp \left\{ -\frac{n}{2\sigma^2} [(\theta - \bar{x})^2 + s_n^2] \right\} \\ &\quad \times [(\theta - \bar{x})^2 + s_n^2]^{-(n/2+1)} \exp\{-|\theta|\} \left( \frac{\sigma^2}{1 + \sigma^2} \right)^2 \\ &\propto g_1(\sigma^2 | \theta, \mathbf{x}) \cdot g_2(\theta | \mathbf{x}) \cdot \exp\{-|\theta|\} \left( \frac{\sigma^2}{1 + \sigma^2} \right)^2,\end{aligned}$$

- 其中  $g_1$  为逆伽玛分布  $\Gamma^{-1}(\alpha, \beta)$ , 其中  $\beta = n[(\theta - \bar{x})^2 + s_n^2]/2$ ,  $\alpha = n/2 + 1$ ;  $g_2$  为广义  $t$  分布  $\mathcal{T}_1(n+1, \bar{x}, s_n^2/(n+1))$ .
- 注意到  $\exp(-|\theta|)(\frac{\sigma^2}{1+\sigma^2})^2$  的尾部对  $g_1 \cdot g_2$  没有太大影响, 故可选择  $g(\theta, \sigma^2 | \mathbf{x}) = g_1(\sigma^2 | \theta, \mathbf{x}) \cdot g_2(\theta | \mathbf{x})$  作为重要性函数.

## 蒙特卡洛重要性抽样的例子

- 因此抽样时需要先从  $g_2$  抽取一个  $\theta$ , 然后在给定此  $\theta$  的条件下, 再从  $g_1$  抽取一个  $\sigma^2$ , 合在一起组成一次抽样  $(\theta, \sigma^2)$ . 在抽取了  $m$  对  $(\theta, \sigma^2)$  后得到  $E^\pi(\theta|\mathbf{x})$  的一个估计

$$\widehat{E^\pi(\theta|\mathbf{x})} = \sum_{i=1}^m \theta_i w(\theta_i, \sigma_i^2 | \mathbf{x}) / \sum_{i=1}^m w(\theta_i, \sigma_i^2 | \mathbf{x}),$$

其中  $w(\theta, \sigma^2 | \mathbf{x}) = f(\mathbf{x} | \theta, \sigma^2) \pi(\theta, \sigma^2) / g(\theta, \sigma^2 | \mathbf{x})$ .

- 关于例 6.2.1 中积分如何达到满意的逼近效果的问题还未解决. 对此问题使用蒙特卡洛重要性抽样方法, 很难找到离目标后验分布比较接近的重要性函数. 因此, 应从后验分布本身抽样. 由于后验分布不是标准分布, 难以从中直接抽样. 我们将在例 6.5.3 中将利用 MCMC 方法, 以目标后验分布  $\pi(\theta|\mathbf{x})$  作为马氏链的稳定分布, 生成随机数, 获得积分较为满意的模拟结果. 在例 6.6.2 中给出了另一种方法, 即利用 Gibbs 抽样方法可获得例 6.2.1 中积分满意的模拟结果.

## 蒙特卡洛重要性抽样的例子

- 因此抽样时需要先从  $g_2$  抽取一个  $\theta$ , 然后在给定此  $\theta$  的条件下, 再从  $g_1$  抽取一个  $\sigma^2$ , 合在一起组成一次抽样  $(\theta, \sigma^2)$ . 在抽取了  $m$  对  $(\theta, \sigma^2)$  后得到  $E^\pi(\theta|\mathbf{x})$  的一个估计

$$\widehat{E^\pi(\theta|\mathbf{x})} = \sum_{i=1}^m \theta_i w(\theta_i, \sigma_i^2 | \mathbf{x}) / \sum_{i=1}^m w(\theta_i, \sigma_i^2 | \mathbf{x}),$$

其中  $w(\theta, \sigma^2 | \mathbf{x}) = f(\mathbf{x} | \theta, \sigma^2) \pi(\theta, \sigma^2) / g(\theta, \sigma^2 | \mathbf{x})$ .

- 关于例 6.2.1 中积分如何达到满意的逼近效果的问题还未解决. 对此问题使用蒙特卡洛重要性抽样方法, 很难找到离目标后验分布比较接近的重要性函数. 因此, 应从后验分布本身抽样. 由于后验分布不是标准分布, 难以从中直接抽样. 我们将在例 6.5.3 中将利用 MCMC 方法, 以目标后验分布  $\pi(\theta|\mathbf{x})$  作为马氏链的稳定分布, 生成随机数, 获得积分较为满意的模拟结果. 在例 6.6.2 中给出了另一种方法, 即利用 Gibbs 抽样方法可获得例 6.2.1 中积分满意的模拟结果.

## 1 6.1 引言

## 2 6.2 蒙特卡洛抽样方法

### • 6.2.1 引言

### • 6.2.2 蒙特卡洛抽样方法

### • 6.2.2 蒙特卡洛重要性抽样方法

## 3 6.3 马尔科夫链蒙特卡洛方法

### • 6.3.1 引言

### • 6.3.2 马尔科夫链的定义及性质

### • 6.3.3 马氏链的极限定理

## 4 6.4 MCMC 方法简介

### • 6.4.1 MCMC 实施中的若干术语

### • 6.4.2 利用 MCMC 的输出结果描述目标后验分布

### • 6.4.3 MCMC 方法收敛性的诊断

## 5 6.5 Metropolis-Hastings 算法

### • 6.5.1 引言

### • 6.5.2 M-H 抽样方法

### • 6.5.3 M-H 抽样方法的几个变种

## 6 6.6 Gibbs 抽样方法

# 引言

- 上一节介绍的蒙特卡洛抽样方法主要使用于后验分布是一维，且是标准分布的情形，此时从中抽样比较容易。
- 它存在如下缺点：(1) 如后验分布不是标准分布使用蒙特卡洛方法中的“重要性函数”，需要选择一个合适的建议密度(即“重要性函数”)，如果“重要性函数”选择不合适，则蒙特卡洛逼近的效果会很差。特别在后验分布是高维(维数大于2)的情形，很难找到合适的“重要性函数”。(2) 标准的蒙特卡洛方法一个严重缺点是在实施中后验分布的完全形式必须已知。对那些后验分布不完全指定或者不直接指定的场合就不能处理了。
- 在上述这些情形下如何从后验分布获取样本？此时马尔科夫链(Markov Chain, 简记 MC) 将被用来解决这个问题。以目标后验分布作为其平稳分布的马尔科夫链生成随机数，利用蒙特卡洛(Monte Carlo, 简记 MC, 即 MCMC 方法中的第二个 MC) 逼近方法，获得相应蒙特卡洛积分的模拟结果。

# 引言

- 上一节介绍的蒙特卡洛抽样方法主要使用于后验分布是一维，且是标准分布的情形，此时从中抽样比较容易。
- 它存在如下缺点：(1) 如后验分布不是标准分布使用蒙特卡洛方法中的“重要性函数”，需要选择一个合适的建议密度 (即“重要性函数”)，如果“重要性函数”选择不合适，则蒙特卡洛逼近的效果会很差。特别在后验分布是高维 (维数大于2) 的情形，很难找到合适的“重要性函数”。  
(2) 标准的蒙特卡洛方法一个严重缺点是在实施中后验分布的完全形式必须已知。对那些后验分布不完全指定或者不直接指定的场合就不能处理了。
- 在上述这些情形下如何从后验分布获取样本？此时马尔科夫链 (Markov Chain, 简记 MC) 将被用来解决这个问题。以目标后验分布作为其平稳分布的马尔科夫链生成随机数，利用蒙特卡洛 (Monte Carlo, 简记 MC, 即 MCMC 方法中的第二个 MC) 逼近方法，获得相应蒙特卡洛积分的模拟结果。

# 引言

- 上一节介绍的蒙特卡洛抽样方法主要使用于后验分布是一维，且是标准分布的情形，此时从中抽样比较容易。
- 它存在如下缺点：(1) 如后验分布不是标准分布使用蒙特卡洛方法中的“重要性函数”，需要选择一个合适的建议密度 (即“重要性函数”)，如果“重要性函数”选择不合适，则蒙特卡洛逼近的效果会很差。特别在后验分布是高维 (维数大于2) 的情形，很难找到合适的“重要性函数”。  
(2) 标准的蒙特卡洛方法一个严重缺点是在实施中后验分布的完全形式必须已知。对那些后验分布不完全指定或者不直接指定的场合就不能处理了。
- 在上述这些情形下如何从后验分布获取样本？此时马尔科夫链 (Markov Chain, 简记 MC) 将被用来解决这个问题。以目标后验分布作为其平稳分布的马尔科夫链生成随机数，利用蒙特卡洛 (Monte Carlo, 简记 MC, 即 MCMC 方法中的第二个 MC) 逼近方法，获得相应蒙特卡洛积分的模拟结果。



# 引言

- 上一节介绍的蒙特卡洛抽样方法主要使用于后验分布是一维，且是标准分布的情形，此时从中抽样比较容易。
- 它存在如下缺点：(1) 如后验分布不是标准分布使用蒙特卡洛方法中的“重要性函数”，需要选择一个合适的建议密度（即“重要性函数”），如果“重要性函数”选择不合适，则蒙特卡洛逼近的效果会很差。特别在后验分布是高维（维数大于2）的情形，很难找到合适的“重要性函数”。  
(2) 标准的蒙特卡洛方法一个严重缺点是在实施中后验分布的完全形式必须已知。对那些后验分布不完全指定或者不直接指定的场合就不能处理了。
- 在上述这些情形下如何从后验分布获取样本？此时马尔科夫链 (Markov Chain, 简记 MC) 将被用来解决这个问题。以目标后验分布作为其平稳分布的马尔科夫链生成随机数，利用蒙特卡洛 (Monte Carlo, 简记 MC, 即 MCMC 方法中的第二个 MC) 逼近方法，获得相应蒙特卡洛积分的模拟结果。

# 马尔科夫链的定义

- **定义 6.3.1** 设  $\{X_n, n \geq 0\}$  是取有限个或可列个值的随机过程, 若  $X_n = i$ , 表示过程在时刻  $n$  状态处于  $i$ ,  $S = \{0, 1, \dots\}$  为状态集. 若对一切  $n$  有

$$\begin{aligned} P(X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i) \\ = P(X_{n+1} = j \mid X_n = i) \end{aligned}$$

则称  $\{X_n, n \geq 0\}$  是离散时间马尔科夫链, 常简称为马氏链.

- 由定义可知, 对过程  $\{X_n, n \geq 0\}$ , 将来状态  $\{X_{n+1}\}$  只与现在状态  $\{X_n\}$  有关, 而与过去状态  $\{X_k, k \leq n-1\}$  无关.
- 条件概率  $P(X_{n+1} = j \mid X_n = i)$  称为马氏链的一步转移概率, 若转移概率与  $n$  无关, 为固定值, 则称马氏链有平稳转移概率, 记为  $p_{ij}$ . 具有平稳转移概率的马氏链它也称为时间齐性马氏链.  $P = (p_{ij}) \forall i, j \in S$  称为马氏链的转移概率矩阵, 满足条件  $p_{ij} \geq 0$ , 且  $\sum_{i=0}^{\infty} p_{ij} = 1$ .

## 马尔科夫链的定义

- **定义 6.3.1** 设  $\{X_n, n \geq 0\}$  是取有限个或可列个值的随机过程, 若  $X_n = i$ , 表示过程在时刻  $n$  状态处于  $i$ ,  $S = \{0, 1, \dots\}$  为状态集. 若对一切  $n$  有

$$\begin{aligned} P(X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i) \\ = P(X_{n+1} = j \mid X_n = i) \end{aligned}$$

则称  $\{X_n, n \geq 0\}$  是离散时间马尔科夫链, 常简称为马氏链.

- 由定义可知, 对过程  $\{X_n, n \geq 0\}$ , 将来状态  $\{X_{n+1}\}$  只与现在状态  $\{X_n\}$  有关, 而与过去状态  $\{X_k, k \leq n-1\}$  无关.
- 条件概率  $P(X_{n+1} = j \mid X_n = i)$  称为马氏链的一步转移概率, 若转移概率与  $n$  无关, 为固定值, 则称马氏链有平稳转移概率, 记为  $p_{ij}$ . 具有平稳转移概率的马氏链它也称为时间齐性马氏链.  $P = (p_{ij}) \forall i, j \in S$  称为马氏链的转移概率矩阵, 满足条件  $p_{ij} \geq 0$ , 且  $\sum_{i=0}^{\infty} p_{ij} = 1$ .

## 马尔科夫链的定义

- **定义 6.3.1** 设  $\{X_n, n \geq 0\}$  是取有限个或可列个值的随机过程, 若  $X_n = i$ , 表示过程在时刻  $n$  状态处于  $i$ ,  $S = \{0, 1, \dots\}$  为状态集. 若对一切  $n$  有

$$\begin{aligned} P(X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i) \\ = P(X_{n+1} = j \mid X_n = i) \end{aligned}$$

则称  $\{X_n, n \geq 0\}$  是离散时间马尔科夫链, 常简称为马氏链.

- 由定义可知, 对过程  $\{X_n, n \geq 0\}$ , 将来状态  $\{X_{n+1}\}$  只与现在状态  $\{X_n\}$  有关, 而与过去状态  $\{X_k, k \leq n-1\}$  无关.
- 条件概率  $P(X_{n+1} = j \mid X_n = i)$  称为马氏链的一步转移概率, 若转移概率与  $n$  无关, 为固定值, 则称马氏链有平稳转移概率, 记为  $p_{ij}$ . 具有平稳转移概率的马氏链它也称为时间齐性马氏链.  $P = (p_{ij}) \forall i, j \in S$  称为马氏链的转移概率矩阵, 满足条件  $p_{ij} \geq 0$ , 且  $\sum_{i=0}^{\infty} p_{ij} = 1$ .

# 1. 马氏链的平稳性

- **定义 6.3.2** 设马尔科夫链有转移概率阵  $P = (p_{ij})$ , 一个概率分布  $\pi = \{\pi_i, i \geq 0\}$  如果满足  $\pi_j = \sum_i \pi_i p_{ij}$ , 则称之为此马尔科夫链的平稳分布.
- 易看出, 如果过程初始状态  $X_0$  有平稳分布  $\pi = \{\pi_i, i \geq 0\}$ , 即  $P(X_0 = j) = \pi_j$ . 则有

$$P(X_1 = j) = \sum_i P(X_1 = j | X_0 = i) P(X_0 = i) = \sum_i \pi_i p_{ij} = \pi_j.$$

- 由归纳法可得

$$P(X_n = j) = \sum_i P(X_n = j | X_{n-1} = i) P(X_{n-1} = i) = \sum_i \pi_i p_{ij}.$$

于是对所有的  $n$ ,  $X_n$  有相同的分布  $\pi$ . 即  $\{X_n, n \geq 0\}$  作为随机过程是平稳的.

# 1. 马氏链的平稳性

- **定义 6.3.2** 设马尔科夫链有转移概率阵  $P = (p_{ij})$ , 一个概率分布  $\pi = \{\pi_i, i \geq 0\}$  如果满足  $\pi_j = \sum_i \pi_i p_{ij}$ , 则称之为此马尔科夫链的平稳分布.
- 易看出, 如果过程初始状态  $X_0$  有平稳分布  $\pi = \{\pi_i, i \geq 0\}$ , 即  $P(X_0 = j) = \pi_j$ . 则有

$$P(X_1 = j) = \sum_i P(X_1 = j | X_0 = i) P(X_0 = i) = \sum_i \pi_i p_{ij} = \pi_j.$$

- 由归纳法可得

$$P(X_n = j) = \sum_i P(X_n = j | X_{n-1} = i) P(X_{n-1} = i) = \sum_i \pi_i p_{ij}.$$

于是对所有的  $n$ ,  $X_n$  有相同的分布  $\pi$ . 即  $\{X_n, n \geq 0\}$  作为随机过程是平稳的.

# 1. 马氏链的平稳性

- **定义 6.3.2** 设马尔科夫链有转移概率阵  $P = (p_{ij})$ , 一个概率分布  $\pi = \{\pi_i, i \geq 0\}$  如果满足  $\pi_j = \sum_i \pi_i p_{ij}$ , 则称之为此马尔科夫链的平稳分布.
- 易看出, 如果过程初始状态  $X_0$  有平稳分布  $\pi = \{\pi_i, i \geq 0\}$ , 即  $P(X_0 = j) = \pi_j$ . 则有

$$P(X_1 = j) = \sum_i P(X_1 = j | X_0 = i) P(X_0 = i) = \sum_i \pi_i p_{ij} = \pi_j.$$

- 由归纳法可得

$$P(X_n = j) = \sum_i P(X_n = j | X_{n-1} = i) P(X_{n-1} = i) = \sum_i \pi_i p_{ij}.$$

于是对所有的  $n$ ,  $X_n$  有相同的分布  $\pi$ . 即  $\{X_n, n \geq 0\}$  作为随机过程是平稳的.

## 2. 不可约性

- **定义 6.3.3** 一个具有可数状态空间  $S$  和转移概率矩阵  $P = (p_{ij})$  的马氏链  $\{X_n\}$  称为是不可约的, 如果对任意两个状态  $i, j \in S$ , 此链从状态  $i$  出发转移到状态  $j$  的概率为正的.

- 即对某个  $n \geq 1$  有

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i) > 0.$$

- 由定义可知, 具有“不可约性”的马氏链意味着从任一状态出发总可到达任一其它状态.



## 2. 不可约性

- **定义 6.3.3** 一个具有可数状态空间  $S$  和转移概率矩阵  $P = (p_{ij})$  的马氏链  $\{X_n\}$  称为是不可约的, 如果对任意两个状态  $i, j \in S$ , 此链从状态  $i$  出发转移到状态  $j$  的概率为正的.
- 即对某个  $n \geq 1$  有

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i) > 0.$$

- 由定义可知, 具有“不可约性”的马氏链意味着从任一状态出发总可到达任一其它状态.

## 2. 不可约性

- **定义 6.3.3** 一个具有可数状态空间  $S$  和转移概率矩阵  $P = (p_{ij})$  的马氏链  $\{X_n\}$  称为是不可约的, 如果对任意两个状态  $i, j \in S$ , 此链从状态  $i$  出发转移到状态  $j$  的概率为正的.
- 即对某个  $n \geq 1$  有

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i) > 0.$$

- 由定义可知, 具有“不可约性”的马氏链意味着从任一状态出发总可到达任一其它状态.

### 3. 非周期性和正常返性

- **定义 6.3.4** 一个马氏链的一个状态  $i$  有周期  $k$ , 如果经过  $k$  的倍数步后一定可以返回到状态  $i$ , 即

$$k(i) = \gcd\{n : P(X_n = i \mid X_0 = i) > 0\},$$

其中  $\gcd$  表示“最大公约数”. 如果返回任一状态的次数的最大公约数是 1, 则称此马氏链是非周期的.

- 非周期的马氏链可以保证其不会陷入循环当中.
- **定义 6.3.5** 对常返状态  $i$ , 令  $T_i = \inf \{n \geq 1 : X_n = i \mid X_0 = i\}$  为首次返回状态  $i$  的时刻, 如果

$$\mu_i = E(T_i) < \infty,$$

则称状态  $i$  是正常返的; 若  $\mu_i = \infty$  时称状态  $i$  是零常返的.

### 3. 非周期性和正常返性

- **定义 6.3.4** 一个马氏链的一个状态  $i$  有周期  $k$ , 如果经过  $k$  的倍数步后一定可以返回到状态  $i$ , 即

$$k(i) = \gcd\{n : P(X_n = i \mid X_0 = i) > 0\},$$

其中  $\gcd$  表示“最大公约数”. 如果返回任一状态的次数的最大公约数是 1, 则称此马氏链是非周期的.

- 非周期的马氏链可以保证其不会陷入循环当中.
- **定义 6.3.5** 对常返状态  $i$ , 令  $T_i = \inf \{n \geq 1 : X_n = i \mid X_0 = i\}$  为首次返回状态  $i$  的时刻, 如果

$$\mu_i = E(T_i) < \infty,$$

则称状态  $i$  是正常返的; 若  $\mu_i = \infty$  时称状态  $i$  是零常返的.

### 3. 非周期性和正常返性

- **定义 6.3.4** 一个马氏链的一个状态  $i$  有周期  $k$ , 如果经过  $k$  的倍数步后一定可以返回到状态  $i$ , 即

$$k(i) = \gcd\{n : P(X_n = i \mid X_0 = i) > 0\},$$

其中  $\gcd$  表示“最大公约数”. 如果返回任一状态的次数的最大公约数是 1, 则称此马氏链是非周期的.

- 非周期的马氏链可以保证其不会陷入循环当中.
- **定义 6.3.5** 对常返状态  $i$ , 令  $T_i = \inf \{n \geq 1 : X_n = i \mid X_0 = i\}$  为首次返回状态  $i$  的时刻, 如果

$$\mu_i = E(T_i) < \infty,$$

则称状态  $i$  是正常返的; 若  $\mu_i = \infty$  时称状态  $i$  是零常返的.

## 4. 遍历性

- **定义 6.3.6** 一个马氏链的状态称为遍历的, 如果它是非周期且正常返的. 如果马氏链的所有状态都是遍历的, 则称此马氏链是遍历的.
- 综上所述, 由马氏链的基本理论可知, 我们需要构造的马氏链必须是不可约、正常返和非周期的. 满足这些正则条件的马氏链存在唯一的平稳分布.

## 4. 遍历性

- **定义 6.3.6** 一个马氏链的状态称为遍历的, 如果它是非周期且正常返的. 如果马氏链的所有状态都是遍历的, 则称此马氏链是遍历的.
- 综上所述, 由马氏链的基本理论可知, 我们需要构造的马氏链必须是不可约、正常返和非周期的. 满足这些正则条件的马氏链存在唯一的平稳分布.

## 马氏链的极限定理

- **定理 6.3.1** 设  $\{X_n, n \geq 0\}$  为一具有可数状态空间  $S$  的马氏链，其转移概率矩阵为  $P$ . 进一步假设它是不可约、非周期，有平稳分布  $\pi = (\pi_i : i \in S)$ , 则有

$$\sum_{j \in S} \left| P(X_n = j) - \pi_j \right| \rightarrow 0, \quad n \rightarrow \infty,$$

对  $X_0$  的任意初始分布  $\pi$ .

- 换言之，对比较大的  $n$ ,  $X_n$  的分布将会接近  $\pi$ . 对一般的状态空间，类似的结果也存在：在合适的条件下，当  $n \rightarrow \infty$  时  $X_n$  的分布将收敛到  $\pi$ .



## 马氏链的极限定理

- **定理 6.3.1** 设  $\{X_n, n \geq 0\}$  为一具有可数状态空间  $S$  的马氏链，其转移概率矩阵为  $P$ . 进一步假设它是不可约、非周期，有平稳分布  $\pi = (\pi_i : i \in S)$ , 则有

$$\sum_{j \in S} \left| P(X_n = j) - \pi_j \right| \rightarrow 0, \quad n \rightarrow \infty,$$

对  $X_0$  的任意初始分布  $\pi$ .

- 换言之，对比较大的  $n$ ,  $X_n$  的分布将会接近  $\pi$ . 对一般的状态空间，类似的结果也存在：在合适的条件下，当  $n \rightarrow \infty$  时  $X_n$  的分布将收敛到  $\pi$ .

## 马氏链的大数定律

- **定理 6.3.2 (马氏链的大数定律)** 假设  $\{X_n, n \geq 0\}$  为一具有可数状态空间  $S$  的马氏链, 其转移概率矩阵为  $P$ . 进一步假设它是不可约的且有平稳分布  $\pi = (\pi_i : i \in S)$ . 则对任何有界函数  $h : S \rightarrow \mathbf{R}$  以及初始值  $X_0$  的任意初始分布有

$$\frac{1}{n} \sum_{i=0}^{n-1} h(X_i) \rightarrow \sum_j h(j) \pi_j, \quad n \rightarrow \infty,$$

依概率成立.

- 当状态空间为不可数, 马氏链  $\{X_n, n \geq 0\}$  为不可约且有平稳分布  $\pi$  时, 也有

$$\frac{1}{n} \sum_{i=0}^{n-1} h(X_i) \rightarrow \int_S h(x) \pi(x) dx, \quad n \rightarrow \infty,$$

## 马氏链的大数定律

- **定理 6.3.2 (马氏链的大数定律)** 假设  $\{X_n, n \geq 0\}$  为一具有可数状态空间  $S$  的马氏链, 其转移概率矩阵为  $P$ . 进一步假设它是不可约的且有平稳分布  $\pi = (\pi_i : i \in S)$ . 则对任何有界函数  $h: S \rightarrow \mathbf{R}$  以及初始值  $X_0$  的任意初始分布有

$$\frac{1}{n} \sum_{i=0}^{n-1} h(X_i) \rightarrow \sum_j h(j) \pi_j, \quad n \rightarrow \infty,$$

依概率成立.

- 当状态空间为不可数, 马氏链  $\{X_n, n \geq 0\}$  为不可约且有平稳分布  $\pi$  时, 也有

$$\frac{1}{n} \sum_{i=0}^{n-1} h(X_i) \rightarrow \int_S h(x) \pi(x) dx, \quad n \rightarrow \infty,$$

## 极限定理的应用

- 这个定理结论是非常有用的. 比如给定集合  $S$  上的概率分布  $\pi$ , 以及  $S$  上的实函数  $h(\theta)$ .
- 设要计算积分  $\mu = \int_S h(\theta)\pi(\theta|\mathbf{x})d\theta$ , 当从后验分布  $\pi(\theta|\mathbf{x})$  中难以直接抽样时, 则可以构造一个马氏链, 使得其状态空间为  $S$  且其平稳分布  $\pi$  就是目标后验分布  $\pi(\cdot|\mathbf{x})$ , 从一初始值  $\theta_0$  出发, 将此链运行一段时间, 如  $0, 1, 2, \dots, n-1$ , 生成随机数(样本)  $\theta_0, \theta_1, \dots, \theta_{n-1}$ ,
- 由马氏链的大数定律可知

$$\bar{\mu}_n = \frac{1}{n} \sum_{j=0}^{n-1} h(\theta_j)$$

为所要求积分  $\mu$  的一相合估计. 这种技术称为马尔科夫链蒙特卡洛 (MCMC) 方法.

## 极限定理的应用

- 这个定理结论是非常有用的. 比如给定集合  $S$  上的概率分布  $\pi$ , 以及  $S$  上的实函数  $h(\theta)$ .
- 设要计算积分  $\mu = \int_S h(\theta)\pi(\theta|\mathbf{x})d\theta$ , 当从后验分布  $\pi(\theta|\mathbf{x})$  中难以直接抽样时, 则可以构造一个马氏链, 使得其状态空间为  $S$  且其平稳分布  $\pi$  就是目标后验分布  $\pi(\cdot|\mathbf{x})$ , 从一初始值  $\theta_0$  出发, 将此链运行一段时间, 如  $0, 1, 2, \dots, n-1$ , 生成随机数(样本)  $\theta_0, \theta_1, \dots, \theta_{n-1}$ ,
- 由马氏链的大数定律可知

$$\bar{\mu}_n = \frac{1}{n} \sum_{j=0}^{n-1} h(\theta_j)$$

为所要求积分  $\mu$  的一相合估计. 这种技术称为马尔科夫链蒙特卡洛 (MCMC) 方法.

## 极限定理的应用

- 这个定理结论是非常有用的. 比如给定集合  $S$  上的概率分布  $\pi$ , 以及  $S$  上的实函数  $h(\theta)$ .
- 设要计算积分  $\mu = \int_S h(\theta)\pi(\theta|\mathbf{x})d\theta$ , 当从后验分布  $\pi(\theta|\mathbf{x})$  中难以直接抽样时, 则可以构造一个马氏链, 使得其状态空间为  $S$  且其平稳分布  $\pi$  就是目标后验分布  $\pi(\cdot|\mathbf{x})$ , 从一初始值  $\theta_0$  出发, 将此链运行一段时间, 如  $0, 1, 2, \dots, n-1$ , 生成随机数(样本)  $\theta_0, \theta_1, \dots, \theta_{n-1}$ ,
- 由马氏链的大数定律可知

$$\bar{\mu}_n = \frac{1}{n} \sum_{j=0}^{n-1} h(\theta_j)$$

为所要求积分  $\mu$  的一相合估计. 这种技术称为马尔科夫链蒙特卡洛 (MCMC) 方法.

- 1 6.1 引言
- 2 6.2 蒙特卡洛抽样方法
  - 6.2.1 引言
  - 6.2.2 蒙特卡洛抽样方法
  - 6.2.2 蒙特卡洛重要性抽样方法
- 3 6.3 马尔科夫链蒙特卡洛方法
  - 6.3.1 引言
  - 6.3.2 马尔科夫链的定义及性质
  - 6.3.3 马氏链的极限定理
- 4 6.4 MCMC 方法简介
  - 6.4.1 MCMC 实施中的若干术语
  - 6.4.2 利用 MCMC 的输出结果描述目标后验分布
  - 6.4.3 MCMC 方法收敛性的诊断
- 5 6.5 Metropolis-Hastings 算法
  - 6.5.1 引言
  - 6.5.2 M-H 抽样方法
  - 6.5.3 M-H 抽样方法的几个变种
- 6 6.6 Gibbs 抽样方法

## 初始值

- **初始值** 初始值被用来初始化一个马氏链. 如果初始值远离后验密度的最高区域, 且算法的迭代次数大小不足以消除初始值的影响, 则它对后验推断可能会造成影响.
- 我们可以通过一些方式降低或者避免初始值的影响, 比如去掉开始一段时间的迭代值、或者从不同的初始值出发获得抽样等等.
- 合理的初始值可以是靠近后验分布的中心位置或者似然函数的最大值点, 但是靠近似然函数的最大值点在一些场合下已经被证明不是一个很好的选择.
- 如果先验分布是有信息的, 则也可以选择先验分布的期望或者众数作为初始值. 一般地, 选择多个从不同初始值开始的链仍然是最推荐的做法.



## 初始值

- **初始值** 初始值被用来初始化一个马氏链. 如果初始值远离后验密度的最高区域, 且算法的迭代次数大小不足以消除初始值的影响, 则它对后验推断可能会造成影响.
- 我们可以通过一些方式降低或者避免初始值的影响, 比如去掉开始一段时间的迭代值、或者从不同的初始值出发获得抽样等等.
- 合理的初始值可以是靠近后验分布的中心位置或者似然函数的最大值点, 但是靠近似然函数的最大值点在一些场合下已经被证明不是一个很好的选择.
- 如果先验分布是有信息的, 则也可以选择先验分布的期望或者众数作为初始值. 一般地, 选择多个从不同初始值开始的链仍然是最推荐的做法.

## 初始值

- **初始值** 初始值被用来初始化一个马氏链. 如果初始值远离后验密度的最高区域, 且算法的迭代次数大小不足以消除初始值的影响, 则它对后验推断可能会造成影响.
- 我们可以通过一些方式降低或者避免初始值的影响, 比如去掉开始一段时间的迭代值、或者从不同的初始值出发获得抽样等等.
- 合理的初始值可以是靠近后验分布的中心位置或者似然函数的最大值点, 但是靠近似然函数的最大值点在一些场合下已经被证明不是一个很好的选择.
- 如果先验分布是有信息的, 则也可以选择先验分布的期望或者众数作为初始值. 一般地, 选择多个从不同初始值开始的链仍然是最推荐的做法.

## 初始值

- **初始值** 初始值被用来初始化一个马氏链. 如果初始值远离后验密度的最高区域, 且算法的迭代次数大小不足以消除初始值的影响, 则它对后验推断可能会造成影响.
- 我们可以通过一些方式降低或者避免初始值的影响, 比如去掉开始一段时间的迭代值、或者从不同的初始值出发获得抽样等等.
- 合理的初始值可以是靠近后验分布的中心位置或者似然函数的最大值点, 但是靠近似然函数的最大值点在一些场合下已经被证明不是一个很好的选择.
- 如果先验分布是有信息的, 则也可以选择先验分布的期望或者众数作为初始值. 一般地, 选择多个从不同初始值开始的链仍然是最推荐的做法.

## 预烧期和抽样步长

- 在 MCMC 机制中用以保证链达到平稳状态所运行的时间称为是 **预烧期**, 其迭代数记为  $B$ .
- 为避免初始值的影响, 预烧期中  $B$  个迭代值将被从样本中去除. 只要链运行的时间足够长, 去除预烧期对后验推断是几乎没有影响的.
- 显然马氏链产生的样本并不是相互独立的, 如果需要独立样本, 则我们可以通过监视产生样本的自相关图, 然后选择 **筛选间隔或抽样步长**  $L > 1$  使得  $L$  步长以后的自相关性很低.
- 我们可以通过每间隔  $L$  个样本抽取一个来获得 (近似) 独立样本.

## 预烧期和抽样步长

- 在 MCMC 机制中用以保证链达到平稳状态所运行的时间称为是 **预烧期**, 其迭代数记为  $B$ .
- 为避免初始值的影响, 预烧期中  $B$  个迭代值将被从样本中去除. 只要链运行的时间足够长, 去除预烧期对后验推断是几乎没有影响的.
- 显然马氏链产生的样本并不是相互独立的, 如果需要独立样本, 则我们可以通过监视产生样本的自相关图, 然后选择 **筛选间隔或抽样步长**  $L > 1$  使得  $L$  步长以后的自相关性很低.
- 我们可以通过每间隔  $L$  个样本抽取一个来获得 (近似) 独立样本.

## 预烧期和抽样步长

- 在 MCMC 机制中用以保证链达到平稳状态所运行的时间称为是 **预烧期**, 其迭代数记为  $B$ .
- 为避免初始值的影响, 预烧期中  $B$  个迭代值将被从样本中去除. 只要链运行的时间足够长, 去除预烧期对后验推断是几乎没有影响的.
- 显然马氏链产生的样本并不是相互独立的, 如果需要独立样本, 则我们可以通过监视产生样本的自相关图, 然后选择 **筛选间隔或抽样步长**  $L > 1$  使得  $L$  步长以后的自相关性很低.
- 我们可以通过每间隔  $L$  个样本抽取一个来获得 (近似) 独立样本.

## 预烧期和抽样步长

- 在 MCMC 机制中用以保证链达到平稳状态所运行的时间称为是 **预烧期**, 其迭代数记为  $B$ .
- 为避免初始值的影响, 预烧期中  $B$  个迭代值将被从样本中去除. 只要链运行的时间足够长, 去除预烧期对后验推断是几乎没有影响的.
- 显然马氏链产生的样本并不是相互独立的, 如果需要独立样本, 则我们可以通过监视产生样本的自相关图, 然后选择 **筛选间隔或抽样步长**  $L > 1$  使得  $L$  步长以后的自相关性很低.
- 我们可以通过每间隔  $L$  个样本抽取一个来获得 (近似) 独立样本.

## 迭代保持数和算法的收敛性

- **迭代保持数  $T$**  设迭代总次数为  $J$ , 迭代保持数为迭代总次数去掉预烧期迭代次数后, 提供给后验贝叶斯分析用的实际样本数, 故有  $T = J - B$ . 如果考虑一个抽样步长  $L$ , 则迭代保持数为去掉预烧期后, 最终的(近似) 独立样本数, 此时  $T = (J - B)/L$ .
- **算法的收敛性** 是指所得到的链是否达到了平稳状态. 如果达到了平稳分布, 则我们得到的样本可以认为是从目标分布中抽取的样本.
- 一般而言, 我们并不清楚必须运行算法多长时间才能认为所得到的链达到了平稳分布. 因此监视链的收敛性是 MCMC 计算方法中的本质问题.



## 迭代保持数和算法的收敛性

- **迭代保持数  $T$**  设迭代总次数为  $J$ , 迭代保持数为迭代总次数去掉预烧期迭代次数后, 提供给后验贝叶斯分析用的实际样本数, 故有  $T = J - B$ . 如果考虑一个抽样步长  $L$ , 则迭代保持数为去掉预烧期后, 最终的(近似) 独立样本数, 此时  $T = (J - B)/L$ .
- **算法的收敛性** 是指所得到的链是否达到了平稳状态. 如果达到了平稳分布, 则我们得到的样本可以认为是从目标分布中抽取的样本.
- 一般而言, 我们并不清楚必须运行算法多长时间才能认为所得到的链达到了平稳分布. 因此监视链的收敛性是 MCMC 计算方法中的本质问题.

## 迭代保持数和算法的收敛性

- **迭代保持数  $T$**  设迭代总次数为  $J$ , 迭代保持数为迭代总次数去掉预烧期迭代次数后, 提供给后验贝叶斯分析用的实际样本数, 故有  $T = J - B$ . 如果考虑一个抽样步长  $L$ , 则迭代保持数为去掉预烧期后, 最终的(近似) 独立样本数, 此时  $T = (J - B)/L$ .
- **算法的收敛性** 是指所得到的链是否达到了平稳状态. 如果达到了平稳分布, 则我们得到的样本可以认为是从目标分布中抽取的样本.
- 一般而言, 我们并不清楚必须运行算法多长时间才能认为所得到的链达到了平稳分布. 因此监视链的收敛性是 MCMC 计算方法中的本质问题.

## 蒙特卡洛 (MC) 误差

- 在 MCMC 输出结果分析中，一个必须报告和监控的量就是**蒙特卡洛误差**。蒙特卡洛误差度量了每一个估计因为随机模拟而导致的波动性。
- 在计算感兴趣的参数时，其精度应该随着样本量递增，因而蒙特卡洛误差必然很低，它和样本量大小成反比并且用户自己可以控制。因此增加迭代次数，感兴趣的量的估计精度也会增加。
- 常用的估计蒙特卡洛误差的方法有两种：组平均 (batch mean) 方法和窗口估计量 (window estimator) 方法。第一种方法简单容易操作，但是第二种方法更精确。

## 蒙特卡洛 (MC) 误差

- 在 MCMC 输出结果分析中，一个必须报告和监控的量就是**蒙特卡洛误差**。蒙特卡洛误差度量了每一个估计因为随机模拟而导致的波动性。
- 在计算感兴趣的参数时，其精度应该随着样本量递增，因而蒙特卡洛误差必然很低，它和样本量大小成反比并且用户自己可以控制。因此增加迭代次数，感兴趣的量的估计精度也会增加。
- 常用的估计蒙特卡洛误差的方法有两种：组平均 (batch mean) 方法和窗口估计量 (window estimator) 方法。第一种方法简单容易操作，但是第二种方法更精确。

## 蒙特卡洛 (MC) 误差

- 在 MCMC 输出结果分析中，一个必须报告和监控的量就是**蒙特卡洛误差**。蒙特卡洛误差度量了每一个估计因为随机模拟而导致的波动性。
- 在计算感兴趣的参数时，其精度应该随着样本量递增，因而蒙特卡洛误差必然很低，它和样本量大小成反比并且用户自己可以控制。因此增加迭代次数，感兴趣的量的估计精度也会增加。
- 常用的估计蒙特卡洛误差的方法有两种：组平均 (batch mean) 方法和窗口估计量 (window estimator) 方法。第一种方法简单容易操作，但是第二种方法更精确。

## 利用 MCMC 的输出结果描述目标后验分布

- MCMC 输出结果给我们提供了一类随机样本

$$\theta_1, \theta_2, \dots, \theta_t, \dots, \theta_T.$$

这组样本可以看成从目标后验分布  $\pi(\theta|\mathbf{x})$  中生成的样本。

- 设  $g(\theta)$  是感兴趣的参数  $\theta$  的函数，从这组样本我们可以进行下列工作：
- (1) 获得  $g(\theta)$  的后验 Summary: 后验均值、后验标准差、后验中位数、后验分位数等。例如，利用传统的估计方法和上述样本可分别获得  $g(\theta)$  的后验期望估计

$$\hat{E}(g(\theta)|\mathbf{x}) = \overline{g(\theta)} = \frac{1}{T} \sum_{t=1}^T g(\theta_t),$$

## 利用 MCMC 的输出结果描述目标后验分布

- MCMC 输出结果给我们提供了一类随机样本

$$\theta_1, \theta_2, \dots, \theta_t, \dots, \theta_T.$$

这组样本可以看成从目标后验分布  $\pi(\theta|\mathbf{x})$  中生成的样本。

- 设  $g(\theta)$  是感兴趣的参数  $\theta$  的函数，从这组样本我们可以进行下列工作：
- (1) 获得  $g(\theta)$  的后验 Summary: 后验均值、后验标准差、后验中位数、后验分位数等。例如，利用传统的估计方法和上述样本可分别获得  $g(\theta)$  的后验期望估计

$$\hat{E}(g(\theta)|\mathbf{x}) = \overline{g(\theta)} = \frac{1}{T} \sum_{t=1}^T g(\theta_t),$$

## 利用 MCMC 的输出结果描述目标后验分布

- MCMC 输出结果给我们提供了一类随机样本

$$\theta_1, \theta_2, \dots, \theta_t, \dots, \theta_T.$$

这组样本可以看成从目标后验分布  $\pi(\theta|\mathbf{x})$  中生成的样本。

- 设  $g(\theta)$  是感兴趣的参数  $\theta$  的函数，从这组样本我们可以进行下列工作：
- (1) 获得  $g(\theta)$  的后验 Summary: 后验均值、后验标准差、后验中位数、后验分位数等。例如，利用传统的估计方法和上述样本可分别获得  $g(\theta)$  的后验期望估计

$$\hat{E}(g(\theta)|\mathbf{x}) = \overline{g(\theta)} = \frac{1}{T} \sum_{t=1}^T g(\theta_t),$$



## 利用 MCMC 的输出结果描述目标后验分布

- 以及后验标准差的估计

$$\widehat{SD}(g(\theta)|\mathbf{x}) = \left\{ \frac{1}{T-1} \sum_{t=1}^T \left[ g(\theta_t) - \widehat{E}(g(\theta)|\mathbf{x}) \right]^2 \right\}^{1/2}.$$

最后，还可以获得后验众数的估计，即在后验密度图中使后验密度达到最大的点.

- (2) 利用后验分位数，例如 2.5% 和 97.5% 分位数，构造可信水平 95% 的可信区间.
- (3) 计算和监测参数间的相关性.
- (4) 产生边缘后验分布图 (直方图、密度函数图等).

## 利用 MCMC 的输出结果描述目标后验分布

- 以及后验标准差的估计

$$\widehat{SD}(g(\theta)|\mathbf{x}) = \left\{ \frac{1}{T-1} \sum_{t=1}^T \left[ g(\theta_t) - \widehat{E}(g(\theta)|\mathbf{x}) \right]^2 \right\}^{1/2}.$$

最后，还可以获得后验众数的估计，即在后验密度图中使后验密度达到最大的点.

- (2) 利用后验分位数，例如 2.5% 和 97.5% 分位数，构造可信水平 95% 的可信区间.
- (3) 计算和监测参数间的相关性.
- (4) 产生边缘后验分布图 (直方图、密度函数图等).

## 利用 MCMC 的输出结果描述目标后验分布

- 以及后验标准差的估计

$$\widehat{SD}(g(\theta)|\mathbf{x}) = \left\{ \frac{1}{T-1} \sum_{t=1}^T \left[ g(\theta_t) - \widehat{E}(g(\theta)|\mathbf{x}) \right]^2 \right\}^{1/2}.$$

最后，还可以获得后验众数的估计，即在后验密度图中使后验密度达到最大的点.

- (2) 利用后验分位数，例如 2.5% 和 97.5% 分位数，构造可信水平 95% 的可信区间.
- (3) 计算和监测参数间的相关性.
- (4) 产生边缘后验分布图 (直方图、密度函数图等).

# MCMC 方法收敛性的诊断

- 无论使用哪一种抽样方法，都要确定所得到的马氏链的收敛性，即需要确定马氏链达到收敛状态时迭代的次数（达到收敛状态前的那一段链称为“预烧期”样本）。
- 通常没有一个全能的方法确定马氏链的收敛性，监视链的收敛性有许多方法。但是每种方法都是针对收敛性问题的不同方面提出的。
- 因此，在绝大多数情况下，为了保证链的收敛性必须应用几种不同的方法去诊断。下面将介绍几种常用的诊断方法。

## MCMC 方法收敛性的诊断

- 无论使用哪一种抽样方法，都要确定所得到的马氏链的收敛性，即需要确定马氏链达到收敛状态时迭代的次数（达到收敛状态前的那一段链称为“预烧期”样本）。
- 通常没有一个全能的方法确定马氏链的收敛性，监视链的收敛性有许多方法。但是每种方法都是针对收敛性问题的不同方面提出的。
- 因此，在绝大多数情况下，为了保证链的收敛性必须应用几种不同的方法去诊断。下面将介绍几种常用的诊断方法。

## MCMC 方法收敛性的诊断

- 无论使用哪一种抽样方法，都要确定所得到的马氏链的收敛性，即需要确定马氏链达到收敛状态时迭代的次数（达到收敛状态前的那一段链称为“预烧期”样本）。
- 通常没有一个全能的方法确定马氏链的收敛性，监视链的收敛性有许多方法。但是每种方法都是针对收敛性问题的不同方面提出的。
- 因此，在绝大多数情况下，为了保证链的收敛性必须应用几种不同的方法去诊断。下面将介绍几种常用的诊断方法。

# 1. 蒙特卡洛 (MC) 误差

- 诊断马氏链收敛性的最简单的方法就是监视 MC 误差.
- 因为较小的 MC 误差表明在计算感兴趣的量时精度较高.
- 因此 MC 误差越小表明马氏链的收敛性越好.

# 1. 蒙特卡洛 (MC) 误差

- 诊断马氏链收敛性的最简单的方法就是监视 MC 误差.
- 因为较小的 MC 误差表明在计算感兴趣的量时精度较高.
- 因此 MC 误差越小表明马氏链的收敛性越好.



# 1. 蒙特卡洛 (MC) 误差

- 诊断马氏链收敛性的最简单的方法就是监视 MC 误差.
- 因为较小的 MC 误差表明在计算感兴趣的量时精度较高.
- 因此 MC 误差越小表明马氏链的收敛性越好.

## 2. 样本路径图

- 另外一种监控方式是使用样本路径图 (trace plot): 马氏链迭代次数对生成的值作图. 如果所有的值都在一个区域里且没有明显的周期性和趋势性, 那么我们可以假设收敛性已经达到.
- 为避免链陷入目标分布的某个局部区域, 通常作几个平行的链, 它们的初始值非常分散. 在经过一段时间后, 如果他们的样本路径图都稳定下来, 而且混合在一起无法区别, 这时可以判定收敛性已经达到.
- 这个想法可以通过将多个链的样本路径图画在同一个图上来检查. 图 6.4.1 给出了一个明显没有达到收敛的例子. 图 6.4.2 则看起来更令人相信链达到了平稳分布, 波动比较稳定, 没有明显的周期性和趋势性.

## 2. 样本路径图

- 另外一种监控方式是使用样本路径图 (trace plot): 马氏链迭代次数对生成的值作图. 如果所有的值都在一个区域里且没有明显的周期性和趋势性, 那么我们可以假设收敛性已经达到.
- 为避免链陷入目标分布的某个局部区域, 通常作几个平行的链, 它们的初始值非常分散. 在经过一段时间后, 如果他们的样本路径图都稳定下来, 而且混合在一起无法区别, 这时可以判定收敛性已经达到.
- 这个想法可以通过将多个链的样本路径图画在同一个图上来检查. 图 6.4.1 给出了一个明显没有达到收敛的例子. 图 6.4.2 则看起来更令人相信链达到了平稳分布, 波动比较稳定, 没有明显的周期性和趋势性.

## 2. 样本路径图

- 另外一种监控方式是使用样本路径图 (trace plot): 马氏链迭代次数对生成的值作图. 如果所有的值都在一个区域里且没有明显的周期性和趋势性, 那么我们可以假设收敛性已经达到.
- 为避免链陷入目标分布的某个局部区域, 通常作几个平行的链, 它们的初始值非常分散. 在经过一段时间后, 如果他们的样本路径图都稳定下来, 而且混合在一起无法区别, 这时可以判定收敛性已经达到.
- 这个想法可以通过将多个链的样本路径图画在同一个图上来检查. 图 6.4.1 给出了一个明显没有达到收敛的例子. 图 6.4.2 则看起来更令人相信链达到了平稳分布, 波动比较稳定, 没有明显的周期性和趋势性.

图 6.4.1

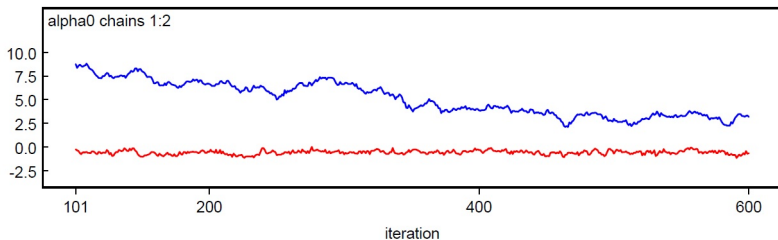


图 6.4.1 没有达到收敛的链的路径图

图 6.4.2

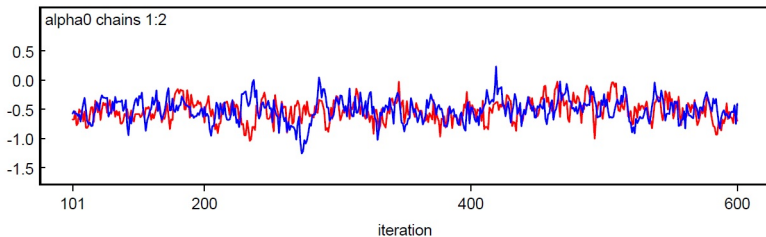


图 6.4.2 达到收敛的链的路径图

### 3. 遍历均值图

- 还有一种很有用的图方法是将马氏链的累积均值对迭代次数作图就得到此链的遍历均值图 (ergodic mean plot).
- 这里累积均值是指此量直至当前迭代的平均值.
- 如果累积均值在经过一些迭代后基本稳定, 则表明算法已经达到收敛 (见图 6.4.3).

### 3. 遍历均值图

- 还有一种很有用的图方法是将马氏链的累积均值对迭代次数作图就得到此链的遍历均值图 (ergodic mean plot).
- 这里累积均值是指此量直至当前迭代的平均值.
- 如果累积均值在经过一些迭代后基本稳定, 则表明算法已经达到收敛 (见图 6.4.3).



### 3. 遍历均值图

- 还有一种很有用的图方法是将马氏链的累积均值对迭代次数作图就得到此链的遍历均值图 (ergodic mean plot).
- 这里累积均值是指此量直至当前迭代的平均值.
- 如果累积均值在经过一些迭代后基本稳定, 则表明算法已经达到收敛 (见图 6.4.3).

图 6.4.3

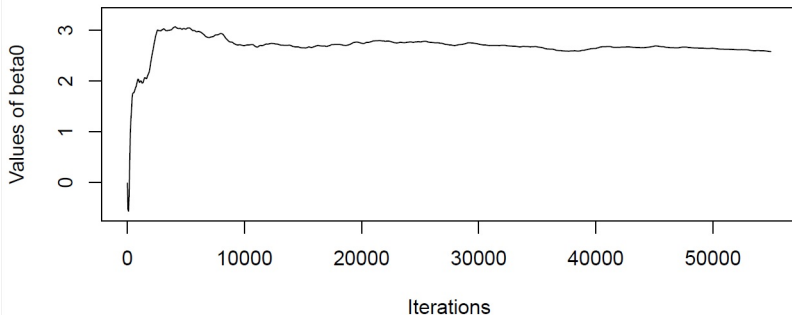


图 6.4.3 达到收敛的链的路径图

## 4. 自相关函数 (ACF) 图和 Gelman-Rubin 方法

- 诊断马氏链收敛性通过监视自相关函数图 (Autocorrelations function plot) 也是很有用的.
- 链的迭代次数对 ACF 作图, 因为较低或者较高的自相关性分别表明了马氏链的快或慢的收敛性.
- Gelman-Rubin 方法是一种常用的收敛性检验方法.

## 4. 自相关函数 (ACF) 图和 Gelman-Rubin 方法

- 诊断马氏链收敛性通过监视自相关函数图 (Autocorrelations function plot) 也是很有用的.
- 链的迭代次数对 ACF 作图, 因为较低或者较高的自相关性分别表明了马氏链的快或慢的收敛性.
- Gelman-Rubin 方法是一种常用的收敛性检验方法.

## 4. 自相关函数 (ACF) 图和 Gelman-Rubin 方法

- 诊断马氏链收敛性通过监视自相关函数图 (Autocorrelations function plot) 也是很有用的.
- 链的迭代次数对 ACF 作图, 因为较低或者较高的自相关性分别表明了马氏链的快或慢的收敛性.
- Gelman-Rubin 方法是一种常用的收敛性检验方法.

## 5. Gelman-Rubin 方法

- 许多统计检验工具也被开发出来用于收敛诊断 (Cowles and Carlin, 1996; Brooks and Roberts, 1998). CODA (Best et al., 1996) 和 BOA (Smith, 2005) 软件程序也被开发用于实施这些工具.
- Gelman-Rubin 方法是一种常用的收敛性检验方法. 详细介绍参看 Gelman & Rubin (1992) 或 Rizzo (2008). 参看本韦和张 (2013) 203-206 页的介绍.

## 5. Gelman-Rubin 方法

- 许多统计检验工具也被开发出来用于收敛诊断 (Cowles and Carlin, 1996; Brooks and Roberts, 1998). CODA (Best et al., 1996) 和 BOA (Smith, 2005) 软件程序也被开发用于实施这些工具.
- Gelman-Rubin 方法是一种常用的收敛性检验方法. 详细介绍参看 Gelman & Rubin (1992) 或 Rizzo (2008). 参看本韦和张 (2013) 203-206 页的介绍.

- 1 6.1 引言
- 2 6.2 蒙特卡洛抽样方法
  - 6.2.1 引言
  - 6.2.2 蒙特卡洛抽样方法
  - 6.2.2 蒙特卡洛重要性抽样方法
- 3 6.3 马尔科夫链蒙特卡洛方法
  - 6.3.1 引言
  - 6.3.2 马尔科夫链的定义及性质
  - 6.3.3 马氏链的极限定理
- 4 6.4 MCMC 方法简介
  - 6.4.1 MCMC 实施中的若干术语
  - 6.4.2 利用 MCMC 的输出结果描述目标后验分布
  - 6.4.3 MCMC 方法收敛性的诊断
- 5 6.5 Metropolis-Hastings 算法
  - 6.5.1 引言
  - 6.5.2 M-H 抽样方法
  - 6.5.3 M-H 抽样方法的几个变种
- 6 6.6 Gibbs 抽样方法



# 引言

- 从一般的后验分布中进行抽样的流行做法是采用马尔科夫链蒙特卡洛 (MCMC) 方法. MCMC 抽样策略是建立一个不可约的、非周期的马氏链, 并且其平稳分布即为感兴趣的目标后验分布.
- 核心问题是确定从当前值转移到下一个值的规则. 其主要任务是生成满足上述正则条件的一个马氏链  $\{x_i, i = 1, \dots, J\}$ .
- Metropolis-Hastings 算法是一类最常用的 MCMC 抽样方法, 它首先由 Metropolis (1953) 提出, 后来由 Hastings (1970) 加以推广的.
- 本节我们介绍 Metropolis-Hastings 抽样方法及其几个特殊情形, 它们在贝叶斯推断中有着广泛的应用.

# 引言

- 从一般的后验分布中进行抽样的流行做法是采用马尔科夫链蒙特卡洛 (MCMC) 方法. MCMC 抽样策略是建立一个不可约的、非周期的马氏链, 并且其平稳分布即为感兴趣的目标后验分布.
- 核心问题是确定从当前值转移到下一个值的规则. 其主要任务是生成满足上述正则条件的一个马氏链  $\{x_i, i = 1, \dots, J\}$ .
- Metropolis-Hastings 算法是一类最常用的 MCMC 抽样方法, 它首先由 Metropolis (1953) 提出, 后来由 Hastings (1970) 加以推广的.
- 本节我们介绍 Metropolis-Hastings 抽样方法及其几个特殊情形, 它们在贝叶斯推断中有着广泛的应用.

# 引言

- 从一般的后验分布中进行抽样的流行做法是采用马尔科夫链蒙特卡洛 (MCMC) 方法. MCMC 抽样策略是建立一个不可约的、非周期的马氏链, 并且其平稳分布即为感兴趣的目标后验分布.
- 核心问题是确定从当前值转移到下一个值的规则. 其主要任务是生成满足上述正则条件的一个马氏链  $\{x_i, i = 1, \dots, J\}$ .
- Metropolis-Hastings 算法是一类最常用的 MCMC 抽样方法, 它首先由 Metropolis (1953) 提出, 后来由 Hastings (1970) 加以推广的.
- 本节我们介绍 Metropolis-Hastings 抽样方法及其几个特殊情形, 它们在贝叶斯推断中有着广泛的应用.

# 引言

- 从一般的后验分布中进行抽样的流行做法是采用马尔科夫链蒙特卡洛 (MCMC) 方法. MCMC 抽样策略是建立一个不可约的、非周期的马氏链, 并且其平稳分布即为感兴趣的目标后验分布.
- 核心问题是确定从当前值转移到下一个值的规则. 其主要任务是生成满足上述正则条件的一个马氏链  $\{x_i, i = 1, \dots, J\}$ .
- Metropolis-Hastings 算法是一类最常用的 MCMC 抽样方法, 它首先由 Metropolis (1953) 提出, 后来由 Hastings (1970) 加以推广的.
- 本节我们介绍 Metropolis-Hastings 抽样方法及其几个特殊情形, 它们在贝叶斯推断中有着广泛的应用.

# Metropolis-Hastings (M-H) 抽样方法

- 为简化记号, 在本节以下用  $f(\cdot)$  表示目标分布, 以  $g(\cdot)$  表示提议分布.
- 设我们希望从目标分布  $f(\cdot)$  中抽样, M-H 算法从初始值  $x_0$  出发, 指定一个从当前值  $x_t$  转移到下一个值  $x_{t+1}$  的规则, 从而产生马氏链  $\{x_i, i = 0, 1, \dots\}$ .
- 具体来说, 在给定当前值  $x_t$ , 从一个提议分布  $g(\cdot|x_t)$  产生一个候选点  $x'$ , 若候选点  $x'$  被接受, 则链从状态  $x'$  转移到链的下一时刻  $t+1$ , 令  $x_{t+1} = x'$ ; 否则链停留在状态  $x_t$ , 令  $x_{t+1} = x_t$ .

# Metropolis-Hastings (M-H) 抽样方法

- 为简化记号, 在本节以下用  $f(\cdot)$  表示目标分布, 以  $g(\cdot)$  表示提议分布.
- 设我们希望从目标分布  $f(\cdot)$  中抽样, M-H 算法从初始值  $x_0$  出发, 指定一个从当前值  $x_t$  转移到下一个值  $x_{t+1}$  的规则, 从而产生马氏链  $\{x_i, i = 0, 1, \dots\}$ .
- 具体来说, 在给定当前值  $x_t$ , 从一个提议分布  $g(\cdot|x_t)$  产生一个候选点  $x'$ , 若候选点  $x'$  被接受, 则链从状态  $x'$  转移到链的下一时刻  $t+1$ , 令  $x_{t+1} = x'$ ; 否则链停留在状态  $x_t$ , 令  $x_{t+1} = x_t$ .

## Metropolis-Hastings (M-H) 抽样方法

- 为简化记号, 在本节以下用  $f(\cdot)$  表示目标分布, 以  $g(\cdot)$  表示提议分布.
- 设我们希望从目标分布  $f(\cdot)$  中抽样, M-H 算法从初始值  $x_0$  出发, 指定一个从当前值  $x_t$  转移到下一个值  $x_{t+1}$  的规则, 从而产生马氏链  $\{x_i, i = 0, 1, \dots\}$ .
- 具体来说, 在给定当前值  $x_t$ , 从一个提议分布  $g(\cdot|x_t)$  产生一个候选点  $x'$ , 若候选点  $x'$  被接受, 则链从状态  $x'$  转移到链的下一时刻  $t+1$ , 令  $x_{t+1} = x'$ ; 否则链停留在状态  $x_t$ , 令  $x_{t+1} = x_t$ .

# M-H 算法

- 候选点  $x'$  是否被接受为链的下一个值, 根据接受概率的表达式  $\alpha(x_t, x') = \min\{1, A\}$  来确定, 其中

$$A = \frac{f(x')g(x_t|x')}{f(x_t)g(x'|x_t)}.$$

- 需要注意的一点是上式分子和分母中的密度函数可分别用“密度函数的核”代替, 故密度函数中的正则化常数因子可省略掉, 以便简化计算.
- M-H算法生成满足正则条件的马氏链迭代过程如下:



# M-H 算法

- 候选点  $x'$  是否被接受为链的下一个值, 根据接受概率的表达式  $\alpha(x_t, x') = \min\{1, A\}$  来确定, 其中

$$A = \frac{f(x')g(x_t|x')}{f(x_t)g(x'|x_t)}.$$

- 需要注意的一点是上式分子和分母中的密度函数可分别用“密度函数的核”代替, 故密度函数中的正则化常数因子可省略掉, 以便简化计算.
- M-H算法生成满足正则条件的马氏链迭代过程如下:

# M-H 算法

- 候选点  $x'$  是否被接受为链的下一个值, 根据接受概率的表达式  $\alpha(x_t, x') = \min\{1, A\}$  来确定, 其中

$$A = \frac{f(x')g(x_t|x')}{f(x_t)g(x'|x_t)}.$$

- 需要注意的一点是上式分子和分母中的密度函数可分别用“密度函数的核”代替, 故密度函数中的正则化常数因子可省略掉, 以便简化计算.
- M-H算法生成满足正则条件的马氏链迭代过程如下:

# M-H算法生成马氏链迭代过程

- (1) 选择一个提议分布  $g(\cdot|x_t)$ ;
- (2) 从提议分布中生成初始值  $x_0$ ;
- (3) 对  $t = 1, 2, \dots$  重复下列步骤
  - (a) 从提议分布  $g(\cdot|x_t)$  中产生一个候选值  $x'$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 计算接受概率, 若  $U \leq A$  则接受  $x'$  且令  $x_{t+1} = x'$ , 否则令  $x_{t+1} = x_t$ .
  - (d) 增加  $t$ , 返回到 (a)

# M-H算法生成马氏链迭代过程

- (1) 选择一个提议分布  $g(\cdot|x_t)$ ;
- (2) 从提议分布中生成初始值  $x_0$ ;
- (3) 对  $t = 1, 2, \dots$  重复下列步骤
  - (a) 从提议分布  $g(\cdot|x_t)$  中产生一个候选值  $x'$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 计算接受概率, 若  $U \leq A$  则接受  $x'$  且令  $x_{t+1} = x'$ , 否则令  $x_{t+1} = x_t$ .
  - (d) 增加  $t$ , 返回到 (a)

# M-H算法生成马氏链迭代过程

- (1) 选择一个提议分布  $g(\cdot|x_t)$ ;
- (2) 从提议分布中生成初始值  $x_0$ ;
- (3) 对  $t = 1, 2, \dots$  重复下列步骤
  - (a) 从提议分布  $g(\cdot|x_t)$  中产生一个候选值  $x'$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 计算接受概率, 若  $U \leq A$  则接受  $x'$  且令  $x_{t+1} = x'$ , 否则令  $x_{t+1} = x_t$ .
  - (d) 增加  $t$ , 返回到 (a)

# M-H算法生成马氏链迭代过程

- (1) 选择一个提议分布  $g(\cdot|x_t)$ ;
- (2) 从提议分布中生成初始值  $x_0$ ;
- (3) 对  $t = 1, 2, \dots$  重复下列步骤
  - (a) 从提议分布  $g(\cdot|x_t)$  中产生一个候选值  $x'$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 计算接受概率, 若  $U \leq A$  则接受  $x'$  且令  $x_{t+1} = x'$ , 否则令  $x_{t+1} = x_t$ .
  - (d) 增加  $t$ , 返回到 (a)

# M-H算法生成马氏链迭代过程

- (1) 选择一个提议分布  $g(\cdot|x_t)$ ;
- (2) 从提议分布中生成初始值  $x_0$ ;
- (3) 对  $t = 1, 2, \dots$  重复下列步骤
  - (a) 从提议分布  $g(\cdot|x_t)$  中产生一个候选值  $x'$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 计算接受概率, 若  $U \leq A$  则接受  $x'$  且令  $x_{t+1} = x'$ , 否则令  $x_{t+1} = x_t$ .
  - (d) 增加  $t$ , 返回到 (a)

# M-H算法生成马氏链迭代过程

- (1) 选择一个提议分布  $g(\cdot|x_t)$ ;
- (2) 从提议分布中生成初始值  $x_0$ ;
- (3) 对  $t = 1, 2, \dots$  重复下列步骤
  - (a) 从提议分布  $g(\cdot|x_t)$  中产生一个候选值  $x'$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 计算接受概率, 若  $U \leq A$  则接受  $x'$  且令  $x_{t+1} = x'$ , 否则令  $x_{t+1} = x_t$ .
  - (d) 增加  $t$ , 返回到 (a)



# M-H算法生成马氏链迭代过程

- (1) 选择一个提议分布  $g(\cdot|x_t)$ ;
- (2) 从提议分布中生成初始值  $x_0$ ;
- (3) 对  $t = 1, 2, \dots$  重复下列步骤
  - (a) 从提议分布  $g(\cdot|x_t)$  中产生一个候选值  $x'$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 计算接受概率, 若  $U \leq A$  则接受  $x'$  且令  $x_{t+1} = x'$ , 否则令  $x_{t+1} = x_t$ .
  - (d) 增加  $t$ , 返回到 (a)

## M-H方法的一个注释

- Robert 和 Casella (1999) 指出, 接受概率并非越大越好, 因为这可能导致较慢的收敛性. Gelman 等 (1996) 建议当参数维数是 1 时, 接受概率应略小于 0.5 是最优的, 当维数大于 5 时, 接受概率应降至 0.25 左右.
- 为说明 M-H 抽样方法产生的马氏链具有平稳分布  $f$ , 则可以通过说明此链的转移核 (或转移概率) 和  $f$  一起满足细致平衡方程. 此结论的证明见 Robert 和 Casella (2004) 定理 7.2, 或参看韦来生和张伟平 (2013).

## M-H方法的一个注释

- Robert 和 Casella (1999) 指出, 接受概率并非越大越好, 因为这可能导致较慢的收敛性. Gelman 等 (1996) 建议当参数维数是 1 时, 接受概率应略小于 0.5 是最优的, 当维数大于 5 时, 接受概率应降至 0.25 左右.
- 为说明 M-H 抽样方法产生的马氏链具有平稳分布  $f$ , 则可以通过说明此链的转移核 (或转移概率) 和  $f$  一起满足细致平衡方程. 此结论的证明见 Robert 和 Casella (2004) 定理 7.2, 或参看韦来生和张伟平 (2013).

## 提议分布应满足的条件

- 提议分布  $g$  的选择除了使得产生的马氏链满足不可约, 正常返, 非周期且具有平稳分布  $f$  等正则化条件外, 还应满足:
- (1) 提议分布的支撑集包含目标分布的支撑集.
- (2) 容易从中抽样, 常取为已知的分布, 如正态或  $t$  分布等.
- (3) 提议分布应使接受概率容易计算.
- (4) 提议分布的尾部要比目标分布的尾部厚.
- (5) 新的候选点被拒绝的频率不高.

## 提议分布应满足的条件

- 提议分布  $g$  的选择除了使得产生的马氏链满足不可约, 正常返, 非周期且具有平稳分布  $f$  等正则化条件外, 还应满足:
- (1) 提议分布的支撑集包含目标分布的支撑集.
- (2) 容易从中抽样, 常取为已知的分布, 如正态或  $t$  分布等.
- (3) 提议分布应使接受概率容易计算.
- (4) 提议分布的尾部要比目标分布的尾部厚.
- (5) 新的候选点被拒绝的频率不高.

## 提议分布应满足的条件

- 提议分布  $g$  的选择除了使得产生的马氏链满足不可约, 正常返, 非周期且具有平稳分布  $f$  等正则化条件外, 还应满足:
- (1) 提议分布的支撑集包含目标分布的支撑集.
- (2) 容易从中抽样, 常取为已知的分布, 如正态或  $t$  分布等.
- (3) 提议分布应使接受概率容易计算.
- (4) 提议分布的尾部要比目标分布的尾部厚.
- (5) 新的候选点被拒绝的频率不高.

## 提议分布应满足的条件

- 提议分布  $g$  的选择除了使得产生的马氏链满足不可约, 正常返, 非周期且具有平稳分布  $f$  等正则化条件外, 还应满足:
- (1) 提议分布的支撑集包含目标分布的支撑集.
- (2) 容易从中抽样, 常取为已知的分布, 如正态或  $t$  分布等.
- (3) 提议分布应使接受概率容易计算.
- (4) 提议分布的尾部要比目标分布的尾部厚.
- (5) 新的候选点被拒绝的频率不高.

## 提议分布应满足的条件

- 提议分布  $g$  的选择除了使得产生的马氏链满足不可约, 正常返, 非周期且具有平稳分布  $f$  等正则化条件外, 还应满足:
- (1) 提议分布的支撑集包含目标分布的支撑集.
- (2) 容易从中抽样, 常取为已知的分布, 如正态或  $t$  分布等.
- (3) 提议分布应使接受概率容易计算.
- (4) 提议分布的尾部要比目标分布的尾部厚.
- (5) 新的候选点被拒绝的频率不高.



## 提议分布应满足的条件

- 提议分布  $g$  的选择除了使得产生的马氏链满足不可约, 正常返, 非周期且具有平稳分布  $f$  等正则化条件外, 还应满足:
- (1) 提议分布的支撑集包含目标分布的支撑集.
- (2) 容易从中抽样, 常取为已知的分布, 如正态或  $t$  分布等.
- (3) 提议分布应使接受概率容易计算.
- (4) 提议分布的尾部要比目标分布的尾部厚.
- (5) 新的候选点被拒绝的频率不高.

## 提议分布应满足的条件

- 提议分布  $g$  的选择除了使得产生的马氏链满足不可约, 正常返, 非周期且具有平稳分布  $f$  等正则化条件外, 还应满足:
- (1) 提议分布的支撑集包含目标分布的支撑集.
- (2) 容易从中抽样, 常取为已知的分布, 如正态或  $t$  分布等.
- (3) 提议分布应使接受概率容易计算.
- (4) 提议分布的尾部要比目标分布的尾部厚.
- (5) 新的候选点被拒绝的频率不高.

## 当目标分布为后验分布时的 M-H 算法

- 前面所述的 M-H 算法可直接在贝叶斯框架下实现, 只要将  $\mathbf{x}$  用感兴趣的参数  $\theta$  代替, 将目标分布  $f(\mathbf{x})$  用后验分布  $\pi(\theta|\mathbf{x})$  代替. 方法概括如下:
  - (1) 选择一个提议分布  $g(\cdot|\theta_t)$ ;
  - (2) 从提议分布中生成初始值  $\theta_0$ ;
  - (3) 对  $t = 1, 2, \dots$  重复下列步骤
    - (a) 从提议分布  $g(\cdot|\theta_t)$  中产生一个候选值  $\theta'$ .
    - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
    - (c) 计算接受概率, 若
$$U \leq \frac{\pi(\theta'|\mathbf{x})g(\theta_t|\theta')}{\pi(\theta_t|\mathbf{x})g(\theta'|\theta_t)},$$
则接受  $\theta'$  且令  $\theta_{t+1} = \theta'$ , 否则令  $\theta_{t+1} = \theta_t$ .
    - (d) 增加  $t$ , 返回到 (a).

## 当目标分布为后验分布时的 M-H 算法

- 前面所述的 M-H 算法可直接在贝叶斯框架下实现，只要将  $\mathbf{x}$  用感兴趣的参数  $\theta$  代替，将目标分布  $f(\mathbf{x})$  用后验分布  $\pi(\theta|\mathbf{x})$  代替. 方法概括如下:
- (1) 选择一个提议分布  $g(\cdot|\theta_t)$ ;
- (2) 从提议分布中生成初始值  $\theta_0$ ;
- (3) 对  $t = 1, 2, \dots$  重复下列步骤
  - (a) 从提议分布  $g(\cdot|\theta_t)$  中产生一个候选值  $\theta'$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 计算接受概率, 若
$$U \leq \frac{\pi(\theta'|\mathbf{x})g(\theta_t|\theta')}{\pi(\theta_t|\mathbf{x})g(\theta'|\theta_t)},$$
则接受  $\theta'$  且令  $\theta_{t+1} = \theta'$ , 否则令  $\theta_{t+1} = \theta_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 当目标分布为后验分布时的 M-H 算法

- 前面所述的 M-H 算法可直接在贝叶斯框架下实现，只要将  $\mathbf{x}$  用感兴趣的参数  $\theta$  代替，将目标分布  $f(\mathbf{x})$  用后验分布  $\pi(\theta|\mathbf{x})$  代替. 方法概括如下：
- (1) 选择一个提议分布  $g(\cdot|\theta_t)$ ;
- (2) 从提议分布中生成初始值  $\theta_0$ ;
- (3) 对  $t = 1, 2, \dots$  重复下列步骤
  - (a) 从提议分布  $g(\cdot|\theta_t)$  中产生一个候选值  $\theta'$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 计算接受概率, 若
$$U \leq \frac{\pi(\theta'|\mathbf{x})g(\theta_t|\theta')}{\pi(\theta_t|\mathbf{x})g(\theta'|\theta_t)},$$
则接受  $\theta'$  且令  $\theta_{t+1} = \theta'$ , 否则令  $\theta_{t+1} = \theta_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 当目标分布为后验分布时的 M-H 算法

- 前面所述的 M-H 算法可直接在贝叶斯框架下实现，只要将  $\mathbf{x}$  用感兴趣的参数  $\theta$  代替，将目标分布  $f(\mathbf{x})$  用后验分布  $\pi(\theta|\mathbf{x})$  代替. 方法概括如下：
- (1) 选择一个提议分布  $g(\cdot|\theta_t)$ ;
- (2) 从提议分布中生成初始值  $\theta_0$ ;
- (3) 对  $t = 1, 2, \dots$  重复下列步骤
  - (a) 从提议分布  $g(\cdot|\theta_t)$  中产生一个候选值  $\theta'$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 计算接受概率, 若
$$U \leq \frac{\pi(\theta'|\mathbf{x})g(\theta_t|\theta')}{\pi(\theta_t|\mathbf{x})g(\theta'|\theta_t)},$$
则接受  $\theta'$  且令  $\theta_{t+1} = \theta'$ , 否则令  $\theta_{t+1} = \theta_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 当目标分布为后验分布时的 M-H 算法

- 前面所述的 M-H 算法可直接在贝叶斯框架下实现，只要将  $\mathbf{x}$  用感兴趣的参数  $\theta$  代替，将目标分布  $f(\mathbf{x})$  用后验分布  $\pi(\theta|\mathbf{x})$  代替. 方法概括如下：
- (1) 选择一个提议分布  $g(\cdot|\theta_t)$ ;
- (2) 从提议分布中生成初始值  $\theta_0$ ;
- (3) 对  $t = 1, 2, \dots$  重复下列步骤
  - 从提议分布  $g(\cdot|\theta_t)$  中产生一个候选值  $\theta'$ .
  - 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - 计算接受概率, 若
$$U \leq \frac{\pi(\theta'|\mathbf{x})g(\theta_t|\theta')}{\pi(\theta_t|\mathbf{x})g(\theta'|\theta_t)},$$
则接受  $\theta'$  且令  $\theta_{t+1} = \theta'$ , 否则令  $\theta_{t+1} = \theta_t$ .
- (d) 增加  $t$ , 返回到 (a).

## 当目标分布为后验分布时的 M-H 算法

- 前面所述的 M-H 算法可直接在贝叶斯框架下实现，只要将  $\mathbf{x}$  用感兴趣的参数  $\theta$  代替，将目标分布  $f(\mathbf{x})$  用后验分布  $\pi(\theta|\mathbf{x})$  代替. 方法概括如下:
- (1) 选择一个提议分布  $g(\cdot|\theta_t)$ ;
- (2) 从提议分布中生成初始值  $\theta_0$ ;
- (3) 对  $t = 1, 2, \dots$  重复下列步骤
  - 从提议分布  $g(\cdot|\theta_t)$  中产生一个候选值  $\theta'$ .
  - 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - 计算接受概率, 若
$$U \leq \frac{\pi(\theta'|\mathbf{x})g(\theta_t|\theta')}{\pi(\theta_t|\mathbf{x})g(\theta'|\theta_t)},$$
则接受  $\theta'$  且令  $\theta_{t+1} = \theta'$ , 否则令  $\theta_{t+1} = \theta_t$ .
- (d) 增加  $t$ , 返回到 (a).



## 当目标分布为后验分布时的 M-H 算法

- 前面所述的 M-H 算法可直接在贝叶斯框架下实现，只要将  $\mathbf{x}$  用感兴趣的参数  $\theta$  代替，将目标分布  $f(\mathbf{x})$  用后验分布  $\pi(\theta|\mathbf{x})$  代替. 方法概括如下:
- (1) 选择一个提议分布  $g(\cdot|\theta_t)$ ;
- (2) 从提议分布中生成初始值  $\theta_0$ ;
- (3) 对  $t = 1, 2, \dots$  重复下列步骤
  - 从提议分布  $g(\cdot|\theta_t)$  中产生一个候选值  $\theta'$ .
  - 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - 计算接受概率, 若

$$U \leq \frac{\pi(\theta'|\mathbf{x})g(\theta_t|\theta')}{\pi(\theta_t|\mathbf{x})g(\theta'|\theta_t)},$$

则接受  $\theta'$  且令  $\theta_{t+1} = \theta'$ , 否则令  $\theta_{t+1} = \theta_t$ .

- 增加  $t$ , 返回到 (a).

## 当目标分布为后验分布时的 M-H 算法

- 前面所述的 M-H 算法可直接在贝叶斯框架下实现，只要将  $\mathbf{x}$  用感兴趣的参数  $\theta$  代替，将目标分布  $f(\mathbf{x})$  用后验分布  $\pi(\theta|\mathbf{x})$  代替. 方法概括如下:
- (1) 选择一个提议分布  $g(\cdot|\theta_t)$ ;
- (2) 从提议分布中生成初始值  $\theta_0$ ;
- (3) 对  $t = 1, 2, \dots$  重复下列步骤
  - 从提议分布  $g(\cdot|\theta_t)$  中产生一个候选值  $\theta'$ .
  - 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - 计算接受概率, 若
$$U \leq \frac{\pi(\theta'|\mathbf{x})g(\theta_t|\theta')}{\pi(\theta_t|\mathbf{x})g(\theta'|\theta_t)},$$
则接受  $\theta'$  且令  $\theta_{t+1} = \theta'$ , 否则令  $\theta_{t+1} = \theta_t$ .
- (d) 增加  $t$ , 返回到 (a).

## M-H 算法的例子

- 下例说明如何利用 M-H 算法生成 Rayleigh (瑞利) 分布的样本.
- 例 6.5.1 使用 Metropolis-Hastings 抽样方法从瑞利分布中抽样. 瑞利分布的密度为

$$f(x) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, \quad x \geq 0, \sigma > 0.$$

瑞利分布适用于使用寿命受到快速老化的建模.

- 解 取自由度 (df) 为  $X_t$  的  $\chi^2$  分布为提议分布, 则使用 M-H 抽样方法的步骤如下:
- (1) 令  $g(\cdot|X)$  为  $\chi^2(df = X)$ .

## M-H 算法的例子

- 下例说明如何利用 M-H 算法生成 Rayleigh (瑞利) 分布的样本.
- 例 6.5.1** 使用 Metropolis-Hastings 抽样方法从瑞利分布中抽样. 瑞利分布的密度为

$$f(x) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, \quad x \geq 0, \sigma > 0.$$

瑞利分布适用于使用寿命受到快速老化的建模.

- 解 取自由度 (df) 为  $X_t$  的  $\chi^2$  分布为提议分布, 则使用 M-H 抽样方法的步骤如下:
- (1) 令  $g(\cdot|X)$  为  $\chi^2(df = X)$ .

## M-H 算法的例子

- 下例说明如何利用 M-H 算法生成 Rayleigh (瑞利) 分布的样本.
- 例 6.5.1** 使用 Metropolis-Hastings 抽样方法从瑞利分布中抽样. 瑞利分布的密度为

$$f(x) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, \quad x \geq 0, \sigma > 0.$$

瑞利分布适用于使用寿命受到快速老化的建模.

- 解** 取自由度 (df) 为  $X_t$  的  $\chi^2$  分布为提议分布, 则使用 M-H 抽样方法的步骤如下:
- (1) 令  $g(\cdot|X)$  为  $\chi^2(df = X)$ .

## M-H 算法的例子

- 下例说明如何利用 M-H 算法生成 Rayleigh (瑞利) 分布的样本.
- 例 6.5.1** 使用 Metropolis-Hastings 抽样方法从瑞利分布中抽样. 瑞利分布的密度为

$$f(x) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, \quad x \geq 0, \sigma > 0.$$

瑞利分布适用于使用寿命受到快速老化的建模.

- 解** 取自由度 (df) 为  $X_t$  的  $\chi^2$  分布为提议分布, 则使用 M-H 抽样方法的步骤如下:
- (1) 令  $g(\cdot|X)$  为  $\chi^2(df = X)$ .

## 续例 6.5.1

- (2) 从  $\chi^2(1)$  中产生  $X_0$ , 并存在  $x[1]$  中.
- (3) 对  $i = 2, \dots, N$ , 重复
  - (a) 从  $\chi^2(df = X_t) = \chi^2(df = x[i - 1])$  中产生  $Y$ .
  - (b) 产生  $U \sim U(0, 1)$ .
  - (c) 由  $X_t = x[i - 1]$ , 计算

$$r(X_t, Y) = \frac{f(Y)g(X_t|Y)}{f(X_t)g(Y|X_t)},$$

其中  $f$  为瑞利密度.  $g(Y|X_t)$  为  $\chi^2(df = X_t)$  的密度在  $Y$  处的值,  $g(X_t|Y)$  为  $\chi^2(df = Y)$  的密度在  $X_t$  处的值. 若  $U \leq r(X_t, Y)$ , 则接受  $Y$ , 令  $X_{t+1} = Y$ ; 否则令  $X_{t+1} = X_t$ . 将  $X_{t+1}$  存在  $x[i]$  里.

- (d) 增加  $t$ , 返回到 (a).

## 续例 6.5.1

- (2) 从  $\chi^2(1)$  中产生  $X_0$ , 并存在  $x[1]$  中.
- (3) 对  $i = 2, \dots, N$ , 重复
  - (a) 从  $\chi^2(df = X_t) = \chi^2(df = x[i - 1])$  中产生  $Y$ .
  - (b) 产生  $U \sim U(0, 1)$ .
  - (c) 由  $X_t = x[i - 1]$ , 计算

$$r(X_t, Y) = \frac{f(Y)g(X_t|Y)}{f(X_t)g(Y|X_t)},$$

其中  $f$  为瑞利密度.  $g(Y|X_t)$  为  $\chi^2(df = X_t)$  的密度在  $Y$  处的值,  $g(X_t|Y)$  为  $\chi^2(df = Y)$  的密度在  $X_t$  处的值. 若  $U \leq r(X_t, Y)$ , 则接受  $Y$ , 令  $X_{t+1} = Y$ ; 否则令  $X_{t+1} = X_t$ . 将  $X_{t+1}$  存在  $x[i]$  里.

- (d) 增加  $t$ , 返回到 (a).



## 续例 6.5.1

- (2) 从  $\chi^2(1)$  中产生  $X_0$ , 并存在  $x[1]$  中.
- (3) 对  $i = 2, \dots, N$ , 重复
  - (a) 从  $\chi^2(df = X_t) = \chi^2(df = x[i - 1])$  中产生  $Y$ .
  - (b) 产生  $U \sim U(0, 1)$ .
  - (c) 由  $X_t = x[i - 1]$ , 计算

$$r(X_t, Y) = \frac{f(Y)g(X_t|Y)}{f(X_t)g(Y|X_t)},$$

其中  $f$  为瑞利密度.  $g(Y|X_t)$  为  $\chi^2(df = X_t)$  的密度在  $Y$  处的值,  $g(X_t|Y)$  为  $\chi^2(df = Y)$  的密度在  $X_t$  处的值. 若  $U \leq r(X_t, Y)$ , 则接受  $Y$ , 令  $X_{t+1} = Y$ ; 否则令  $X_{t+1} = X_t$ . 将  $X_{t+1}$  存在  $x[i]$  里.

- (d) 增加  $t$ , 返回到 (a).

## 续例 6.5.1

- (2) 从  $\chi^2(1)$  中产生  $X_0$ , 并存在  $x[1]$  中.
- (3) 对  $i = 2, \dots, N$ , 重复
  - (a) 从  $\chi^2(df = X_t) = \chi^2(df = x[i - 1])$  中产生  $Y$ .
  - (b) 产生  $U \sim U(0, 1)$ .
  - (c) 由  $X_t = x[i - 1]$ , 计算

$$r(X_t, Y) = \frac{f(Y)g(X_t|Y)}{f(X_t)g(Y|X_t)},$$

其中  $f$  为瑞利密度.  $g(Y|X_t)$  为  $\chi^2(df = X_t)$  的密度在  $Y$  处的值,  $g(X_t|Y)$  为  $\chi^2(df = Y)$  的密度在  $X_t$  处的值. 若  $U \leq r(X_t, Y)$ , 则接受  $Y$ , 令  $X_{t+1} = Y$ ; 否则令  $X_{t+1} = X_t$ . 将  $X_{t+1}$  存在  $x[i]$  里.

- (d) 增加  $t$ , 返回到 (a).

## 续例 6.5.1

- (2) 从  $\chi^2(1)$  中产生  $X_0$ , 并存在  $x[1]$  中.
- (3) 对  $i = 2, \dots, N$ , 重复
  - (a) 从  $\chi^2(df = X_t) = \chi^2(df = x[i - 1])$  中产生  $Y$ .
  - (b) 产生  $U \sim U(0, 1)$ .
  - (c) 由  $X_t = x[i - 1]$ , 计算

$$r(X_t, Y) = \frac{f(Y)g(X_t|Y)}{f(X_t)g(Y|X_t)},$$

其中  $f$  为瑞利密度.  $g(Y|X_t)$  为  $\chi^2(df = X_t)$  的密度在  $Y$  处的值,  $g(X_t|Y)$  为  $\chi^2(df = Y)$  的密度在  $X_t$  处的值. 若  $U \leq r(X_t, Y)$ , 则接受  $Y$ , 令  $X_{t+1} = Y$ ; 否则令  $X_{t+1} = X_t$ . 将  $X_{t+1}$  存在  $x[i]$  里.

- (d) 增加  $t$ , 返回到 (a).

## 续例 6.5.1

- (2) 从  $\chi^2(1)$  中产生  $X_0$ , 并存在  $x[1]$  中.
- (3) 对  $i = 2, \dots, N$ , 重复
  - (a) 从  $\chi^2(df = X_t) = \chi^2(df = x[i - 1])$  中产生  $Y$ .
  - (b) 产生  $U \sim U(0, 1)$ .
  - (c) 由  $X_t = x[i - 1]$ , 计算

$$r(X_t, Y) = \frac{f(Y)g(X_t|Y)}{f(X_t)g(Y|X_t)},$$

其中  $f$  为瑞利密度.  $g(Y|X_t)$  为  $\chi^2(df = X_t)$  的密度在  $Y$  处的值,  $g(X_t|Y)$  为  $\chi^2(df = Y)$  的密度在  $X_t$  处的值. 若  $U \leq r(X_t, Y)$ , 则接受  $Y$ , 令  $X_{t+1} = Y$ ; 否则令  $X_{t+1} = X_t$ . 将  $X_{t+1}$  存在  $x[i]$  里.

- (d) 增加  $t$ , 返回到 (a).

## 续例 6.5.1

- 在密度  $f$  中的常数可以在计算  $r$  中抵消, 因此

$$r(x_t, y) = \frac{f(y)g(x_t|y)}{f(x_t)g(y|x_t)} = \frac{ye^{-y^2/2\sigma^2}}{x_te^{-x_t^2/2\sigma^2}} \times \frac{\Gamma(x_t/2)2^{x_t/2}x_t^{y/2-1}e^{-x_t/2}}{\Gamma(y/2)2^{y/2}y^{x_t/2-1}e^{-y/2}}.$$

在此例中, 我们还是通过计算整个密度在某点的值来计算  $r$ . 下面的代码计算瑞利密度在某点的值:

- 下面我们产生  $\sigma = 4$  的瑞利分布随机数. 使用的提议分布为自由度是  $x_t = x[i-1]$  的  $\chi^2(df = x_t)$  分布:
- 在计算  $r(X_t, Y)$  中, 分子和分母分别用变量 `num` 和 `den` 表示. 记数变量 `k` 记录了候选点被拒绝的次数  $k = 4070$ , 大约 40% 的候选点被拒绝了. 因此这个方法产生链的效率不高. 我们使用样本对时间作图 (称为 `trace plot`), 来观测其样本路径图 (图 6.5.1). 注意在候选点被拒绝的时间点上链没有移动, 因此图中有很多短的水平平移.

## 续例 6.5.1

- 在密度  $f$  中的常数可以在计算  $r$  中抵消, 因此

$$r(x_t, y) = \frac{f(y)g(x_t|y)}{f(x_t)g(y|x_t)} = \frac{ye^{-y^2/2\sigma^2}}{x_te^{-x_t^2/2\sigma^2}} \times \frac{\Gamma(x_t/2)2^{x_t/2}x_t^{y/2-1}e^{-x_t/2}}{\Gamma(y/2)2^{y/2}y^{x_t/2-1}e^{-y/2}}.$$

在此例中, 我们还是通过计算整个密度在某点的值来计算  $r$ . 下面的代码计算瑞利密度在某点的值:

- 下面我们产生  $\sigma = 4$  的瑞利分布随机数. 使用的提议分布为自由度是  $x_t = x[i-1]$  的  $\chi^2(df = x_t)$  分布:
- 在计算  $r(X_t, Y)$  中, 分子和分母分别用变量 `num` 和 `den` 表示. 记数变量 `k` 记录了候选点被拒绝的次数  $k = 4070$ , 大约 40% 的候选点被拒绝了. 因此这个方法产生链的效率不高. 我们使用样本对时间作图 (称为 `trace plot`), 来观测其样本路径图 (图 6.5.1). 注意在候选点被拒绝的时间点上链没有移动, 因此图中有很多短的水平平移.

## 续例 6.5.1

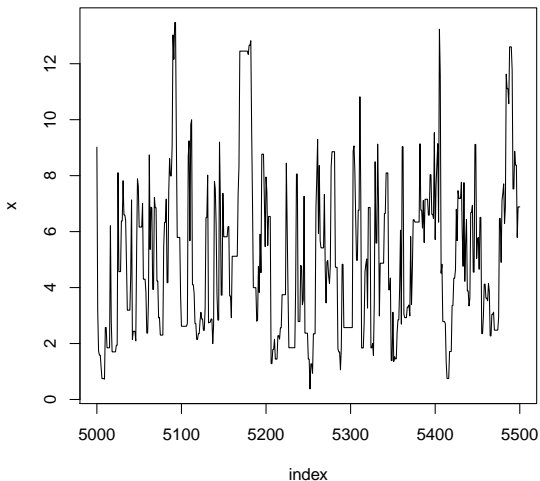
- 在密度  $f$  中的常数可以在计算  $r$  中抵消, 因此

$$r(x_t, y) = \frac{f(y)g(x_t|y)}{f(x_t)g(y|x_t)} = \frac{ye^{-y^2/2\sigma^2}}{x_te^{-x_t^2/2\sigma^2}} \times \frac{\Gamma(x_t/2)2^{x_t/2}x_t^{y/2-1}e^{-x_t/2}}{\Gamma(y/2)2^{y/2}y^{x_t/2-1}e^{-y/2}}.$$

在此例中, 我们还是通过计算整个密度在某点的值来计算  $r$ . 下面的代码计算瑞利密度在某点的值:

- 下面我们产生  $\sigma = 4$  的瑞利分布随机数. 使用的提议分布为自由度是  $x_t = x[i-1]$  的  $\chi^2(df = x_t)$  分布:
- 在计算  $r(X_t, Y)$  中, 分子和分母分别用变量 `num` 和 `den` 表示. 记数变量 `k` 记录了候选点被拒绝的次数  $k = 4070$ , 大约 40% 的候选点被拒绝了. 因此这个方法产生链的效率不高. 我们使用样本对时间作图 (称为 `trace plot`), 来观测其样本路径图 (图 6.5.1). 注意在候选点被拒绝的时间点上链没有移动, 因此图中有很多短的水平平移.

图 6.5.1

图 6.5.1 瑞利分布由  $M-H$  算法产生的马氏链部分路径图



## 续例 6.5.1

- 本例中我们的目的是说明 M-H 抽样方法的应用，对瑞利分布，有更高效率的产生随机数方法. 比如瑞利分布的分位数可以表示为

$$x_q = F^{-1}(q) = \sigma[-2\ln(1 - q)]^{1/2}, 0 < q < 1.$$

因此可以使用逆变换方法生成随机数.

- 利用 R 代码可获得比较  $\sigma = 4$  的瑞利分布的理论分位数和生成链的分位数拟合程度的 QQ 图，和以  $\sigma = 4$  的瑞利分布为目标分布的链生成样本的直方图.
- 图 6.5.2 的左边是样本的直方图和密度函数曲线图，右边的 QQ 图是判断由链产生的样本分位数和目标分布的理论分位数拟合好坏的一种方法. 由于 QQ 图上的点基本上集中在一条直线附近，这显示样本分位数和理论分位数是高度近似一致的. 生成图的 R 代码见教材.

## 续例 6.5.1

- 本例中我们的目的是说明 M-H 抽样方法的应用，对瑞利分布，有更高效率的产生随机数方法. 比如瑞利分布的分位数可以表示为

$$x_q = F^{-1}(q) = \sigma[-2\ln(1 - q)]^{1/2}, 0 < q < 1.$$

因此可以使用逆变换方法生成随机数.

- 利用 R 代码可获得比较  $\sigma = 4$  的瑞利分布的理论分位数和生成链的分位数拟合程度的 QQ 图，和以  $\sigma = 4$  的瑞利分布为目标分布的链生成样本的直方图.
- 图 6.5.2 的左边是样本的直方图和密度函数曲线图，右边的 QQ 图是判断由链产生的样本分位数和目标分布的理论分位数拟合好坏的一种方法. 由于 QQ 图上的点基本上集中在一条直线附近，这显示样本分位数和理论分位数是高度近似一致的. 生成图的 R 代码见教材.

## 续例 6.5.1

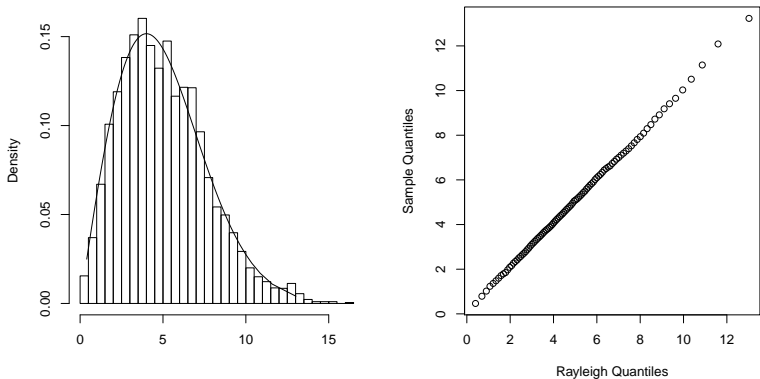
- 本例中我们的目的是说明 M-H 抽样方法的应用，对瑞利分布，有更高效率的产生随机数方法. 比如瑞利分布的分位数可以表示为

$$x_q = F^{-1}(q) = \sigma[-2\ln(1 - q)]^{1/2}, 0 < q < 1.$$

因此可以使用逆变换方法生成随机数.

- 利用 R 代码可获得比较  $\sigma = 4$  的瑞利分布的理论分位数和生成链的分位数拟合程度的 QQ 图，和以  $\sigma = 4$  的瑞利分布为目标分布的链生成样本的直方图.
- 图 6.5.2 的左边是样本的直方图和密度函数曲线图，右边的 QQ 图是判断由链产生的样本分位数和目标分布的理论分位数拟合好坏的一种方法. 由于 QQ 图上的点基本上集中在一条直线附近，这显示样本分位数和理论分位数是高度近似一致的. 生成图的 R 代码见教材.

图 6.5.2

图 6.5.2  $M-H$  链样本的直方图和  $M-H$  链分位数的  $QQ$  图

# 1. Metropolis 抽样方法

- 根据提议分布  $g$  的不同选择, M-H 抽样方法衍生出了几个不同的变种: Metropolis 抽样方法, 随机游动 Metropolis 抽样方法, 独立抽样方法和逐分量 M-H 抽样方法等. 下面将分别介绍这些方法.
- **Metropolis 抽样方法** 在 Metropolis 抽样方法中, 提议分布是对称的. 即  $g(\cdot|X_n)$  满足

$$g(X|Y) = g(Y|X),$$

因此接受概率为

$$\alpha(X_t, Y) = \min \left\{ 1, \frac{f(Y)}{f(X_t)} \right\}.$$

# 1. Metropolis 抽样方法

- 根据提议分布  $g$  的不同选择, M-H 抽样方法衍生出了几个不同的变种: Metropolis 抽样方法, 随机游动 Metropolis 抽样方法, 独立抽样方法和逐分量 M-H 抽样方法等. 下面将分别介绍这些方法.
- **Metropolis 抽样方法** 在 Metropolis 抽样方法中, 提议分布是对称的. 即  $g(\cdot|X_n)$  满足

$$g(X|Y) = g(Y|X),$$

因此接受概率为

$$\alpha(X_t, Y) = \min \left\{ 1, \frac{f(Y)}{f(X_t)} \right\}.$$

## 2. 随机游动 Metropolis方法

- **随机游动 Metropolis方法** 随机游动 Metropolis 抽样方法是 Metropolis 方法的一个特例. 假设候选点  $Y$  从一个对称的提议分布  $g(Y|X_t) = g(|X_t - Y|)$  中产生的. 则在每一次迭代中, 从  $g(\cdot)$  中产生一个随机增量  $Z$ , 然后  $Y = X_t + Z$ . 比如随机增量  $Z$  可以从均值为 0 的正态分布中产生, 此时候选点  $Y|X_t \sim N(X_t, \sigma^2)$ ,  $\sigma^2 > 0$ .
- 随机游动 Metropolis 算法下得到的链, 其收敛性常常对刻度参数的选择比较敏感. 当增量的方差太大时, 大部分的候选点会被拒绝, 此时算法的效率很低. 如果增量的方差太小, 则候选点就几乎都被接受, 因此此时随机游动 Metropolis 算法下得到的链就几乎是随机游动了, 效率也较低. Robert et al (1996) 提出一种选择刻度参数的方法是监视接受率, 拒绝率应该在区间  $[0.15, 0.5]$  之内才可以保证得到的链有较好的性质.

## 2. 随机游动 Metropolis方法

- **随机游动 Metropolis方法** 随机游动 Metropolis 抽样方法是 Metropolis 方法的一个特例. 假设候选点  $Y$  从一个对称的提议分布  $g(Y|X_t) = g(|X_t - Y|)$  中产生的. 则在每一次迭代中, 从  $g(\cdot)$  中产生一个随机增量  $Z$ , 然后  $Y = X_t + Z$ . 比如随机增量  $Z$  可以从均值为 0 的正态分布中产生, 此时候选点  $Y|X_t \sim N(X_t, \sigma^2)$ ,  $\sigma^2 > 0$ .
- 随机游动 Metropolis 算法下得到的链, 其收敛性常常对刻度参数的选择比较敏感. 当增量的方差太大时, 大部分的候选点会被拒绝, 此时算法的效率很低. 如果增量的方差太小, 则候选点就几乎都被接受, 因此此时随机游动 Metropolis 算法下得到的链就几乎是随机游动了, 效率也较低. Robert et al (1996) 提出一种选择刻度参数的方法是监视接受率, 拒绝率应该在区间  $[0.15, 0.5]$  之内才可以保证得到的链有较好的性质.



## 随机游动-Metropolis 算法的例子

- **例 6.5.2** (随机游动 Metropolis) 使用提议分布  $N(X_t, \sigma^2)$  和随机游动 Metropolis 算法产生自由度为  $\nu = 4$  的  $t$  分布随机数, 并对标准差  $\sigma$  的几个不同值重复此过程, 从中并选出  $\sigma$  的一个最佳值, 使链的收敛性最好. 丢弃链的前 500 个迭代 (预烧期) 值, 比较 4 个生成链观测值的十分位数和自由度为  $\nu$  的  $t$  分布理论上的十分位数的拟合状况, 并找出与理论分布十分位数拟合最好的那个链.
- $t_\nu$  的密度正比于  $(1 + x^2/\nu)^{-(\nu+1)/2}$ , 因此

$$\alpha(x_t, y) = \min \left\{ 1, \frac{f(y)}{f(x_t)} \right\} = \min \left\{ 1, \frac{(1 + y^2/\nu)^{-(\nu+1)/2}}{(1 + x_t^2/\nu)^{-(\nu+1)/2}} \right\}.$$

下面我们仍然使用  $dt$  来计算  $t$  密度在给定点处的值.

## 随机游动-Metropolis 算法的例子

- **例 6.5.2** (随机游动 Metropolis) 使用提议分布  $N(X_t, \sigma^2)$  和随机游动 Metropolis 算法产生自由度为  $\nu = 4$  的  $t$  分布随机数, 并对标准差  $\sigma$  的几个不同值重复此过程, 从中并选出  $\sigma$  的一个最佳值, 使链的收敛性最好. 丢弃链的前 500 个迭代 (预烧期) 值, 比较 4 个生成链观测值的十分位数和自由度为  $\nu$  的  $t$  分布理论上的十分位数的拟合状况, 并找出与理论分布十分位数拟合最好的那个链.
- $t_\nu$  的密度正比于  $(1 + x^2/\nu)^{-(\nu+1)/2}$ , 因此

$$\alpha(x_t, y) = \min \left\{ 1, \frac{f(y)}{f(x_t)} \right\} = \min \left\{ 1, \frac{(1 + y^2/\nu)^{-(\nu+1)/2}}{(1 + x_t^2/\nu)^{-(\nu+1)/2}} \right\}.$$

下面我们仍然使用  $dt$  来计算  $t$  密度在给定点处的值.

## 随机游动-Metropolis 算法步骤

- (1) 令提议分布  $g(\cdot|\theta_t)$  为正态分布  $N(X_t, \sigma^2)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, N$  重复下列步骤:
  - (a) 从提议分布  $N(X_t, \sigma^2)$  中产生一个候选值  $y$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{f(y)}{f(x_t)} = \frac{(1 + y^2/\nu)^{-(\nu+1)/2}}{(1 + x_t^2/\nu)^{-(\nu+1)/2}}$$
若  $U \leq A$  则接受  $y$  且令  $X_{t+1} = y$ , 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 随机游动-Metropolis 算法步骤

- (1) 令提议分布  $g(\cdot|\theta_t)$  为正态分布  $N(X_t, \sigma^2)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, N$  重复下列步骤:
  - (a) 从提议分布  $N(X_t, \sigma^2)$  中产生一个候选值  $y$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{f(y)}{f(x_t)} = \frac{(1 + y^2/\nu)^{-(\nu+1)/2}}{(1 + x_t^2/\nu)^{-(\nu+1)/2}}$$
若  $U \leq A$  则接受  $y$  且令  $X_{t+1} = y$ , 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 随机游动-Metropolis 算法步骤

- (1) 令提议分布  $g(\cdot|\theta_t)$  为正态分布  $N(X_t, \sigma^2)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, N$  重复下列步骤:
  - (a) 从提议分布  $N(X_t, \sigma^2)$  中产生一个候选值  $y$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{f(y)}{f(x_t)} = \frac{(1 + y^2/\nu)^{-(\nu+1)/2}}{(1 + x_t^2/\nu)^{-(\nu+1)/2}}$$
若  $U \leq A$  则接受  $y$  且令  $X_{t+1} = y$ , 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 随机游动-Metropolis 算法步骤

- (1) 令提议分布  $g(\cdot|\theta_t)$  为正态分布  $N(X_t, \sigma^2)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, N$  重复下列步骤:
  - (a) 从提议分布  $N(X_t, \sigma^2)$  中产生一个候选值  $y$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{f(y)}{f(x_t)} = \frac{(1 + y^2/\nu)^{-(\nu+1)/2}}{(1 + x_t^2/\nu)^{-(\nu+1)/2}}$$
若  $U \leq A$  则接受  $y$  且令  $X_{t+1} = y$ , 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 随机游动-Metropolis 算法步骤

- (1) 令提议分布  $g(\cdot|\theta_t)$  为正态分布  $N(X_t, \sigma^2)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, N$  重复下列步骤:
  - (a) 从提议分布  $N(X_t, \sigma^2)$  中产生一个候选值  $y$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{f(y)}{f(x_t)} = \frac{(1 + y^2/\nu)^{-(\nu+1)/2}}{(1 + x_t^2/\nu)^{-(\nu+1)/2}}$$
若  $U \leq A$  则接受  $y$  且令  $X_{t+1} = y$ , 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 随机游动-Metropolis 算法步骤

- (1) 令提议分布  $g(\cdot|\theta_t)$  为正态分布  $N(X_t, \sigma^2)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, N$  重复下列步骤:
  - (a) 从提议分布  $N(X_t, \sigma^2)$  中产生一个候选值  $y$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{f(y)}{f(x_t)} = \frac{(1 + y^2/\nu)^{-(\nu+1)/2}}{(1 + x_t^2/\nu)^{-(\nu+1)/2}}$$
若  $U \leq A$  则接受  $y$  且令  $X_{t+1} = y$ , 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).



## 随机游动-Metropolis 算法步骤

- (1) 令提议分布  $g(\cdot|\theta_t)$  为正态分布  $N(X_t, \sigma^2)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, N$  重复下列步骤:
  - (a) 从提议分布  $N(X_t, \sigma^2)$  中产生一个候选值  $y$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{f(y)}{f(x_t)} = \frac{(1 + y^2/\nu)^{-(\nu+1)/2}}{(1 + x_t^2/\nu)^{-(\nu+1)/2}}$$
若  $U \leq A$  则接受  $y$  且令  $X_{t+1} = y$ , 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 续例 6.5.2

- 对提议分布的方差分别取为  $\sigma = 0.05, 0.5, 2, 16$  四种情况, 输出 4 个不同链的拒绝率, 并生成链的轨迹图.
- 上述四种方差的选择, 4 个拒绝率为 0.0035, 0.1000, 0.4390, 0.9010. 只有第三个链的拒绝率在区间  $[0.15, 0.5]$  之间. 我们可以在不同的提议分布方差下, 检查所得链的收敛性.
- 由图 6.5.3 可以看出:  $\sigma = 0.05$  时, 增量太小, 几乎每个候选点都被接受了, 链在 2000 次迭代后还没有收敛.  $\sigma^2 = 0.5$  时, 链的收敛较慢.  $\sigma = 2$  时, 链很快收敛. 而当  $\sigma = 16$  时, 接受的概率太小, 链虽然收敛了, 但是效率很低.

## 续例 6.5.2

- 对提议分布的方差分别取为  $\sigma = 0.05, 0.5, 2, 16$  四种情况, 输出 4 个不同链的拒绝率, 并生成链的轨迹图.
- 上述四种方差的选择, 4 个拒绝率为 0.0035, 0.1000, 0.4390, 0.9010. 只有第三个链的拒绝率在区间  $[0.15, 0.5]$  之间. 我们可以在不同的提议分布方差下, 检查所得链的收敛性.
- 由图 6.5.3 可以看出:  $\sigma = 0.05$  时, 增量太小, 几乎每个候选点都被接受了, 链在 2000 次迭代后还没有收敛.  $\sigma^2 = 0.5$  时, 链的收敛较慢.  $\sigma = 2$  时, 链很快收敛. 而当  $\sigma = 16$  时, 接受的概率太小, 链虽然收敛了, 但是效率很低.

## 续例 6.5.2

- 对提议分布的方差分别取为  $\sigma = 0.05, 0.5, 2, 16$  四种情况, 输出 4 个不同链的拒绝率, 并生成链的轨迹图.
- 上述四种方差的选择, 4 个拒绝率为 0.0035, 0.1000, 0.4390, 0.9010. 只有第三个链的拒绝率在区间  $[0.15, 0.5]$  之间. 我们可以在不同的提议分布方差下, 检查所得链的收敛性.
- 由图 6.5.3 可以看出:  $\sigma = 0.05$  时, 增量太小, 几乎每个候选点都被接受了, 链在 2000 次迭代后还没有收敛.  $\sigma^2 = 0.5$  时, 链的收敛较慢.  $\sigma = 2$  时, 链很快收敛. 而当  $\sigma = 16$  时, 接受的概率太小, 链虽然收敛了, 但是效率很低.

## 6.5 Metropolis-Hastings 算法

## 6.5.3 M-H 抽样方法的几个变种

## 续例 6.5.2

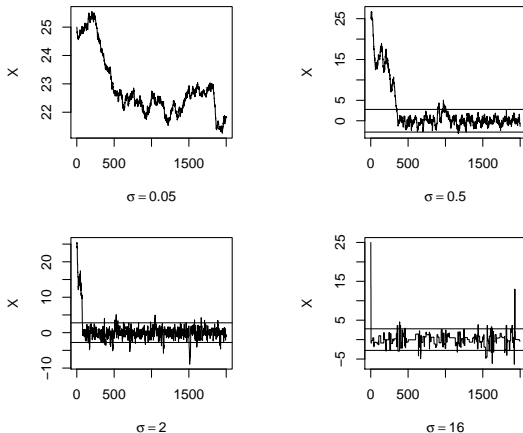


图 6.5.3 由不同方差的提议分布生成随机游动 M-H 链

## 续例 6.5.2

- 一般说来在 MCMC 问题中, 我们不能获得目标分布的理论分位数用于比较的目的.
- 但在本例中理论分位数可以通过从具有一定自由度的  $t$  分布中直接获得, 故可用于比较.
- 忽略预烧期的 500 个值, 通过应用函数获得模拟分位数的值. 目标分布的理论分位数和 4 个链  $rw1$ ,  $rw2$ ,  $rw3$  和  $rw4$  的样本分位数见表 6.5.1. 由表可见  $rw3$  的样本分位数与理论分位数比较接近.

## 续例 6.5.2

- 一般说来在 MCMC 问题中，我们不能获得目标分布的理论分位数用于比较的目的.
- 但在本例中理论分位数可以通过从具有一定自由度的  $t$  分布中直接获得, 故可用于比较.
- 忽略预烧期的 500 个值，通过应用函数获得模拟分位数的值. 目标分布的理论分位数和 4 个链  $rw1$ ,  $rw2$ ,  $rw3$  和  $rw4$  的样本分位数见表 6.5.1. 由表可见  $rw3$  的样本分位数与理论分位数比较接近.

## 续例 6.5.2

- 一般说来在 MCMC 问题中，我们不能获得目标分布的理论分位数用于比较的目的.
- 但在本例中理论分位数可以通过从具有一定自由度的  $t$  分布中直接获得, 故可用于比较.
- 忽略预烧期的 500 个值，通过应用函数获得模拟分位数的值. 目标分布的理论分位数和 4 个链  $rw1$ ,  $rw2$ ,  $rw3$  和  $rw4$  的样本分位数见表 6.5.1. 由表可见  $rw3$  的样本分位数与理论分位数比较接近.



## 续例 6.5.2

表 6.5.1 目标分布的理论分位数和 4 个链的样本分位

	$Q$	$rw1$	$rw2$	$rw3$	$rw4$
5%	-2.132	23.589	-1.474	-2.053	-2.929
10%	-1.533	23.707	-1.108	-1.525	-2.045
20%	-0.941	23.939	-0.402	-1.077	-0.927
30%	-0.569	24.105	0.043	-0.625	-0.500
40%	-0.271	24.211	0.513	-0.333	-0.187
50%	0.000	24.459	0.936	-0.045	-0.025
60%	0.271	24.704	1.766	0.261	0.300
70%	0.569	24.877	4.270	0.533	0.622
80%	0.941	24.992	13.652	0.936	1.051
90%	1.533	25.129	16.868	1.482	1.892
95%	2.132	25.210	19.250	1.860	2.495

## 例 6.5.3 (续例 6.1.1)

- **例 6.5.3 (续例 6.1.1)** 设随机变量  $X \sim N(\theta, \sigma^2)$ , 其中  $\sigma^2$  已知, 取  $\theta$  的先验分布为柯西分布  $C(\mu, \tau)$ , 则  $\theta$  的后验分布

$$\pi(\theta|x) \propto \exp \left\{ -(\theta - x)^2 / (2\sigma^2) \right\} (\tau^2 + (\theta - \mu)^2)^{-1}. \quad (6.5.2)$$

设提议分布为  $\theta \sim N(X_t, \eta^2)$ , 使用随机游动 Metropolis 算法产生以目标分布为  $\pi(\theta|x)$  的随机数, 当  $\eta$  取几个不同值时讨论链的收敛性, 去掉适当的预烧期, 求链的样本均值和样本方差.

- **解** 由于我们不能直接从此后验分布中产生随机数, 故使用随机游动的 Metropolis 算法产生以后验分布 (6.5.2) 作为其平稳分布的马氏链, 生成随机数. 算法如下:
- (1) 令提议分布  $g(\cdot)$  为  $N(X_t, \eta^2)$ ,  $X_t$  为正态分布的均值.

## 例 6.5.3 (续例 6.1.1)

- **例 6.5.3 (续例 6.1.1)** 设随机变量  $X \sim N(\theta, \sigma^2)$ , 其中  $\sigma^2$  已知, 取  $\theta$  的先验分布为柯西分布  $C(\mu, \tau)$ , 则  $\theta$  的后验分布

$$\pi(\theta|x) \propto \exp \left\{ -(\theta - x)^2 / (2\sigma^2) \right\} (\tau^2 + (\theta - \mu)^2)^{-1}. \quad (6.5.2)$$

设提议分布为  $\theta \sim N(X_t, \eta^2)$ , 使用随机游动 Metropolis 算法产生以目标分布为  $\pi(\theta|x)$  的随机数, 当  $\eta$  取几个不同值时讨论链的收敛性, 去掉适当的预烧期, 求链的样本均值和样本方差.

- **解** 由于我们不能直接从此后验分布中产生随机数, 故使用随机游动的 Metropolis 算法产生以后验分布 (6.5.2) 作为其平稳分布的马氏链, 生成随机数. 算法如下:
- (1) 令提议分布  $g(\cdot)$  为  $N(X_t, \eta^2)$ ,  $X_t$  为正态分布的均值.

## 例 6.5.3 (续例 6.1.1)

- **例 6.5.3 (续例 6.1.1)** 设随机变量  $X \sim N(\theta, \sigma^2)$ , 其中  $\sigma^2$  已知, 取  $\theta$  的先验分布为柯西分布  $C(\mu, \tau)$ , 则  $\theta$  的后验分布

$$\pi(\theta|x) \propto \exp \left\{ -(\theta - x)^2 / (2\sigma^2) \right\} (\tau^2 + (\theta - \mu)^2)^{-1}. \quad (6.5.2)$$

设提议分布为  $\theta \sim N(X_t, \eta^2)$ , 使用随机游动 Metropolis 算法产生以目标分布为  $\pi(\theta|x)$  的随机数, 当  $\eta$  取几个不同值时讨论链的收敛性, 去掉适当的预烧期, 求链的样本均值和样本方差.

- **解** 由于我们不能直接从此后验分布中产生随机数, 故使用随机游动的 Metropolis 算法产生以后验分布 (6.5.2) 作为其平稳分布的马氏链, 生成随机数. 算法如下:
- (1) 令提议分布  $g(\cdot)$  为  $N(X_t, \eta^2)$ ,  $X_t$  为正态分布的均值.

## 续例 6.5.3

- (2) 从提议分布中产生初始值  $X_0$ , 并存在  $x[1]$  中.
- (3) 对  $i = 2, \dots, N$ , 重复
  - (a) 从  $N(X_t, \eta^2)$  中产生候选值  $Y$ .
  - (b) 从  $U(0, 1)$  产生随机数  $U$ .
  - (c) 令  $X_t = x[i - 1]$ , 计算接受概率

$$\alpha(X_t, Y) = \min\{1, f(Y)/f(X_t)\},$$

其中目标分布  $f(\cdot)$  就是  $\theta$  的后验分布 (6.5.2) (与  $\theta$  无关的常数因子省略), 而

$$\frac{f(Y)}{f(X_t)} = \frac{\exp\{- (Y - x)^2 / (2\sigma^2)\} (\tau^2 + (Y - \mu)^2)^{-1}}{\exp\{- (X_t - x)^2 / (2\sigma^2)\} (\tau^2 + (X_t - \mu)^2)^{-1}}.$$

若  $U \leq \alpha(X_t, Y)$ , 则接受  $Y$ , 令  $X_{t+1} = Y$ ; 否则令  $X_{t+1} = X_t$ .

- (d) 增加  $t$ , 返回到 (a).

## 续例 6.5.3

- (2) 从提议分布中产生初始值  $X_0$ , 并存在  $x[1]$  中.
- (3) 对  $i = 2, \dots, N$ , 重复
  - (a) 从  $N(X_t, \eta^2)$  中产生候选值  $Y$ .
  - (b) 从  $U(0, 1)$  产生随机数  $U$ .
  - (c) 令  $X_t = x[i - 1]$ , 计算接受概率

$$\alpha(X_t, Y) = \min\{1, f(Y)/f(X_t)\},$$

其中目标分布  $f(\cdot)$  就是  $\theta$  的后验分布 (6.5.2) (与  $\theta$  无关的常数因子省略), 而

$$\frac{f(Y)}{f(X_t)} = \frac{\exp\{- (Y - x)^2 / (2\sigma^2)\} (\tau^2 + (Y - \mu)^2)^{-1}}{\exp\{- (X_t - x)^2 / (2\sigma^2)\} (\tau^2 + (X_t - \mu)^2)^{-1}}.$$

若  $U \leq \alpha(X_t, Y)$ , 则接受  $Y$ , 令  $X_{t+1} = Y$ ; 否则令  $X_{t+1} = X_t$ .

- (d) 增加  $t$ , 返回到 (a).

## 续例 6.5.3

- (2) 从提议分布中产生初始值  $X_0$ , 并存在  $x[1]$  中.
- (3) 对  $i = 2, \dots, N$ , 重复
  - (a) 从  $N(X_t, \eta^2)$  中产生候选值  $Y$ .
  - (b) 从  $U(0, 1)$  产生随机数  $U$ .
  - (c) 令  $X_t = x[i - 1]$ , 计算接受概率

$$\alpha(X_t, Y) = \min\{1, f(Y)/f(X_t)\},$$

其中目标分布  $f(\cdot)$  就是  $\theta$  的后验分布 (6.5.2) (与  $\theta$  无关的常数因子省略), 而

$$\frac{f(Y)}{f(X_t)} = \frac{\exp\{- (Y - x)^2 / (2\sigma^2)\} (\tau^2 + (Y - \mu)^2)^{-1}}{\exp\{- (X_t - x)^2 / (2\sigma^2)\} (\tau^2 + (X_t - \mu)^2)^{-1}}.$$

若  $U \leq \alpha(X_t, Y)$ , 则接受  $Y$ , 令  $X_{t+1} = Y$ ; 否则令  $X_{t+1} = X_t$ .

- (d) 增加  $t$ , 返回到 (a).

## 续例 6.5.3

- (2) 从提议分布中产生初始值  $X_0$ , 并存在  $x[1]$  中.
- (3) 对  $i = 2, \dots, N$ , 重复
  - (a) 从  $N(X_t, \eta^2)$  中产生候选值  $Y$ .
  - (b) 从  $U(0, 1)$  产生随机数  $U$ .
  - (c) 令  $X_t = x[i - 1]$ , 计算接受概率

$$\alpha(X_t, Y) = \min\{1, f(Y)/f(X_t)\},$$

其中目标分布  $f(\cdot)$  就是  $\theta$  的后验分布 (6.5.2) (与  $\theta$  无关的常数因子省略), 而

$$\frac{f(Y)}{f(X_t)} = \frac{\exp\{- (Y - x)^2 / (2\sigma^2)\} (\tau^2 + (Y - \mu)^2)^{-1}}{\exp\{- (X_t - x)^2 / (2\sigma^2)\} (\tau^2 + (X_t - \mu)^2)^{-1}}.$$

若  $U \leq \alpha(X_t, Y)$ , 则接受  $Y$ , 令  $X_{t+1} = Y$ ; 否则令  $X_{t+1} = X_t$ .

- (d) 增加  $t$ , 返回到 (a).



## 续例 6.5.3

- (2) 从提议分布中产生初始值  $X_0$ , 并存在  $x[1]$  中.
- (3) 对  $i = 2, \dots, N$ , 重复
  - (a) 从  $N(X_t, \eta^2)$  中产生候选值  $Y$ .
  - (b) 从  $U(0, 1)$  产生随机数  $U$ .
  - (c) 令  $X_t = x[i - 1]$ , 计算接受概率

$$\alpha(X_t, Y) = \min\{1, f(Y)/f(X_t)\},$$

其中目标分布  $f(\cdot)$  就是  $\theta$  的后验分布 (6.5.2) (与  $\theta$  无关的常数因子省略), 而

$$\frac{f(Y)}{f(X_t)} = \frac{\exp\{- (Y - x)^2 / (2\sigma^2)\} (\tau^2 + (Y - \mu)^2)^{-1}}{\exp\{- (X_t - x)^2 / (2\sigma^2)\} (\tau^2 + (X_t - \mu)^2)^{-1}}.$$

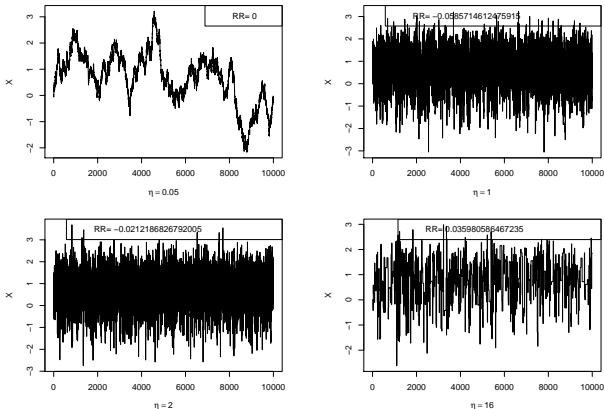
若  $U \leq \alpha(X_t, Y)$ , 则接受  $Y$ , 令  $X_{t+1} = Y$ ; 否则令  $X_{t+1} = X_t$ .

- (d) 增加  $t$ , 返回到 (a).

## 6.5 Metropolis-Hastings 算法

## 6.5.3 M-H 抽样方法的几个变种

## 续例 6.5.3

图 6.5.4 提议分布取不同方差随机游动 *Metropolis* 算法生成的链

## 续例 6.5.3

- 对提议分布四种不同标准差  $\eta$  的选择, 只有第二个链的拒绝率 0.3099 落在区间  $[0.15, 0.5]$  之间. 第三个链的拒绝率比 0.5 略大一点, 故第二个链收敛性最好, 第三个链比第二个链略差, 第四个链较差, 第一个链最差. 由图 6.5.4 可见马氏链的轨迹图显示的收敛性也与上述判断一致. 故提议分布的方差  $\eta = 1$  生成链的效果最好.
- 对提议分布的方差  $\eta = 1$  的马氏链 (即前面所述的第二个链) 去掉开始的 500 个观测值 (预烧期) 后, 由剩下的观测值得链的样本均值和样本标准差.
- 以公式 (6.5.2) 作为目标分布的马氏链获得的样本均值和样本标准差分别为  $m = 0.6955276$  和  $s = 0.9249807$ , 这就是例 6.1.1 中后验分布的后验均值和标准差的蒙特卡洛逼近.

## 续例 6.5.3

- 对提议分布四种不同标准差  $\eta$  的选择, 只有第二个链的拒绝率 0.3099 落在区间  $[0.15, 0.5]$  之间. 第三个链的拒绝率比 0.5 略大一点, 故第二个链收敛性最好, 第三个链比第二个链略差, 第四个链较差, 第一个链最差. 由图 6.5.4 可见马氏链的轨迹图显示的收敛性也与上述判断一致. 故提议分布的方差  $\eta = 1$  生成链的效果最好.
- 对提议分布的方差  $\eta = 1$  的马氏链 (即前面所述的第二个链) 去掉开始的 500 个观测值 (预烧期) 后, 由剩下的观测值得链的样本均值和样本标准差.
- 以公式 (6.5.2) 作为目标分布的马氏链获得的样本均值和样本标准差分别为  $m = 0.6955276$  和  $s = 0.9249807$ , 这就是例 6.1.1 中后验分布的后验均值和标准差的蒙特卡洛逼近.

## 续例 6.5.3

- 对提议分布四种不同标准差  $\eta$  的选择, 只有第二个链的拒绝率 0.3099 落在区间  $[0.15, 0.5]$  之间. 第三个链的拒绝率比 0.5 略大一点, 故第二个链收敛性最好, 第三个链比第二个链略差, 第四个链较差, 第一个链最差. 由图 6.5.4 可见马氏链的轨迹图显示的收敛性也与上述判断一致. 故提议分布的方差  $\eta = 1$  生成链的效果最好.
- 对提议分布的方差  $\eta = 1$  的马氏链 (即前面所述的第二个链) 去掉开始的 500 个观测值 (预烧期) 后, 由剩下的观测值得链的样本均值和样本标准差.
- 以公式 (6.5.2) 作为目标分布的马氏链获得的样本均值和样本标准差分别为  $m = 0.6955276$  和  $s = 0.9249807$ , 这就是例 6.1.1 中后验分布的后验均值和标准差的蒙特卡洛逼近.

## 一个简单的投资模型

**例 6.5.4** (贝叶斯推断：一个简单的投资模型) 一般，不同的投资所得的回报是不独立的. 为了减少风险，因此使用投资组合以保证有价证券的回报是负相关的. 这里不讨论回报的相关性，而是对每天组合里的每个证券的收益进行排序. 假设有 5 种股票被跟踪记录了 250 交易日每天的表现，在每一个交易日，收益最大的股票被标记出来. 用  $X_i$  表示股票  $i$  在 250 个交易日中胜出天数，则记录到得的频数  $(x_1, \dots, x_5)$  为 r.v.  $(X_1, \dots, X_5)$  的观测值. 基于历史数据，假设这 5 种股票在任何给定的一个交易日能胜出的先验机会比率为

$$1 : (1 - \beta) : (1 - 2\beta) : 2\beta : \beta,$$

这里  $\beta \in (0, 0.5)$  是一个未知的参数. 在有了当前这 250 个交易日的的数据后，使用贝叶斯方法对此比例进行更新.

## 续例 6.5.4

- 解 由前所述,  $\mathbf{X} = (X_1, \dots, X_5)$  在给定  $\beta$  的条件下服从多项分布, 概率向量为

$$\mathbf{p} = \mathbf{p}(\beta) = \left( \frac{1}{3}, \frac{1-\beta}{3}, \frac{1-2\beta}{3}, \frac{2\beta}{3}, \frac{\beta}{3} \right).$$

- 令  $\beta$  的先验分布  $\pi(\beta)$  为  $(0, 0.5)$  上的均匀分布,  $\mathbf{p}$  的似然函数 (即样本  $\mathbf{x}$  的分布) 为  $l(\mathbf{p}|\mathbf{x})$ , 因此后验分布为
$$\pi(\beta|x_1, \dots, x_5) \propto l(\mathbf{p}(\beta)|\mathbf{x}) \cdot \pi(\beta) \propto (1-\beta)^{x_2} (1-2\beta)^{x_3} (2\beta)^{x_4} \beta^{x_5}.$$
- 我们不能直接从此后验分布中产生随机数. 一种估计  $\beta$  的方法是用随机游动 Metropolis 算法产生一个马氏链, 使其平稳分布为此后验分布, 提议分布为对称的均匀分布, 然后从此链中产生目标分布的随机数来估计  $\beta$ .

## 续例 6.5.4

- 解 由前所述,  $\mathbf{X} = (X_1, \dots, X_5)$  在给定  $\beta$  的条件下服从多项分布, 概率向量为

$$\mathbf{p} = \mathbf{p}(\beta) = \left( \frac{1}{3}, \frac{1-\beta}{3}, \frac{1-2\beta}{3}, \frac{2\beta}{3}, \frac{\beta}{3} \right).$$

- 令  $\beta$  的先验分布  $\pi(\beta)$  为  $(0, 0.5)$  上的均匀分布,  $\mathbf{p}$  的似然函数 (即样本  $\mathbf{x}$  的分布) 为  $l(\mathbf{p}|\mathbf{x})$ , 因此后验分布为

$$\pi(\beta|x_1, \dots, x_5) \propto l(\mathbf{p}(\beta)|\mathbf{x}) \cdot \pi(\beta) \propto (1-\beta)^{x_2} (1-2\beta)^{x_3} (2\beta)^{x_4} \beta^{x_5}.$$

- 我们不能直接从此后验分布中产生随机数. 一种估计  $\beta$  的方法是用随机游动 Metropolis 算法产生一个马氏链, 使其平稳分布为此后验分布, 提议分布为对称的均匀分布, 然后从此链中产生目标分布的随机数来估计  $\beta$ .



## 续例 6.5.4

- 解 由前所述,  $\mathbf{X} = (X_1, \dots, X_5)$  在给定  $\beta$  的条件下服从多项分布, 概率向量为

$$\mathbf{p} = \mathbf{p}(\beta) = \left( \frac{1}{3}, \frac{1-\beta}{3}, \frac{1-2\beta}{3}, \frac{2\beta}{3}, \frac{\beta}{3} \right).$$

- 令  $\beta$  的先验分布  $\pi(\beta)$  为  $(0, 0.5)$  上的均匀分布,  $\mathbf{p}$  的似然函数 (即样本  $\mathbf{x}$  的分布) 为  $l(\mathbf{p}|\mathbf{x})$ , 因此后验分布为

$$\pi(\beta|x_1, \dots, x_5) \propto l(\mathbf{p}(\beta)|\mathbf{x}) \cdot \pi(\beta) \propto (1-\beta)^{x_2} (1-2\beta)^{x_3} (2\beta)^{x_4} \beta^{x_5}.$$

- 我们不能直接从此后验分布中产生随机数. 一种估计  $\beta$  的方法是用随机游动 Metropolis 算法产生一个马氏链, 使其平稳分布为此后验分布, 提议分布为对称的均匀分布, 然后从此链中产生目标分布的随机数来估计  $\beta$ .

## 续例 6.5.4

- 此时，接受的概率为

$$\alpha(X_t, Y) = \min \left\{ 1, \frac{f(Y)}{f(X_t)} \right\}.$$

其中  $f(\cdot) = \pi(\cdot | x_1, \dots, x_5)$  为目标分布，且

$$\frac{f(Y)}{f(X_t)} = \frac{(1-Y)^{x_2}(1-2Y)^{x_3}(2Y)^{x_4}Y^{x_5}}{(1-X_t)^{x_2}(1-2X_t)^{x_3}(2X_t)^{x_4}X_t^{x_5}}. \quad (1)$$

- 下面使用随机游动 Metropolis 算法生产随机数. 这里需要两个均匀分布的随机变量，其中一个对称均匀分布用于产生提议分布，而另一个均匀分布用于决定接受还是拒绝候选点.

## 续例 6.5.4

- 此时，接受的概率为

$$\alpha(X_t, Y) = \min \left\{ 1, \frac{f(Y)}{f(X_t)} \right\}.$$

其中  $f(\cdot) = \pi(\cdot | x_1, \dots, x_5)$  为目标分布，且

$$\frac{f(Y)}{f(X_t)} = \frac{(1-Y)^{x_2}(1-2Y)^{x_3}(2Y)^{x_4}Y^{x_5}}{(1-X_t)^{x_2}(1-2X_t)^{x_3}(2X_t)^{x_4}X_t^{x_5}}. \quad (1)$$

- 下面使用随机游动 Metropolis 算法生产随机数. 这里需要两个均匀分布的随机变量，其中一个对称均匀分布用于产生提议分布，而另一个均匀分布用于决定接受还是拒绝候选点.

## 随机游动-Metropolis 算法步骤

- (1) 令提议分布为对称均匀分布  $U(-w, w)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, m$  重复下列步骤:
  - (a) 从提议分布  $U(-w, w)$  中产生  $v$ , 令  $y = X_t + v$  为候选值.
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中

$$A = \frac{f(y)}{f(x_t)}$$

如 (1) 所示. 若  $U \leq A$  则接受  $y$  且令  $X_{t+1} = y$ , 否则令  $X_{t+1} = X_t$ .

- (d) 增加  $t$ , 返回到 (a).

## 随机游动-Metropolis 算法步骤

- (1) 令提议分布为对称均匀分布  $U(-w, w)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, m$  重复下列步骤:
  - (a) 从提议分布  $U(-w, w)$  中产生  $v$ , 令  $y = X_t + v$  为候选值.
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中

$$A = \frac{f(y)}{f(x_t)}$$

如 (1) 所示. 若  $U \leq A$  则接受  $y$  且令  $X_{t+1} = y$ , 否则令  $X_{t+1} = X_t$ .

- (d) 增加  $t$ , 返回到 (a).

## 随机游动-Metropolis 算法步骤

- (1) 令提议分布为对称均匀分布  $U(-w, w)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, m$  重复下列步骤:
  - (a) 从提议分布  $U(-w, w)$  中产生  $v$ , 令  $y = X_t + v$  为候选值.
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{f(y)}{f(x_t)}$$
如 (1) 所示. 若  $U \leq A$  则接受  $y$  且令  $X_{t+1} = y$ , 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 随机游动-Metropolis 算法步骤

- (1) 令提议分布为对称均匀分布  $U(-w, w)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, m$  重复下列步骤:
  - (a) 从提议分布  $U(-w, w)$  中产生  $v$ , 令  $y = X_t + v$  为候选值.
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{f(y)}{f(x_t)}$$
如 (1) 所示. 若  $U \leq A$  则接受  $y$  且令  $X_{t+1} = y$ , 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 随机游动-Metropolis 算法步骤

- (1) 令提议分布为对称均匀分布  $U(-w, w)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, m$  重复下列步骤:
  - (a) 从提议分布  $U(-w, w)$  中产生  $v$ , 令  $y = X_t + v$  为候选值.
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中

$$A = \frac{f(y)}{f(x_t)}$$

如 (1) 所示. 若  $U \leq A$  则接受  $y$  且令  $X_{t+1} = y$ , 否则令  $X_{t+1} = X_t$ .

- (d) 增加  $t$ , 返回到 (a).



## 随机游动-Metropolis 算法步骤

- (1) 令提议分布为对称均匀分布  $U(-w, w)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, m$  重复下列步骤:
  - (a) 从提议分布  $U(-w, w)$  中产生  $v$ , 令  $y = X_t + v$  为候选值.
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中

$$A = \frac{f(y)}{f(x_t)}$$

如 (1) 所示. 若  $U \leq A$  则接受  $y$  且令  $X_{t+1} = y$ , 否则令  $X_{t+1} = X_t$ .

(d) 增加  $t$ , 返回到 (a).

## 随机游动-Metropolis 算法步骤

- (1) 令提议分布为对称均匀分布  $U(-w, w)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, m$  重复下列步骤:
  - (a) 从提议分布  $U(-w, w)$  中产生  $v$ , 令  $y = X_t + v$  为候选值.
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中

$$A = \frac{f(y)}{f(x_t)}$$

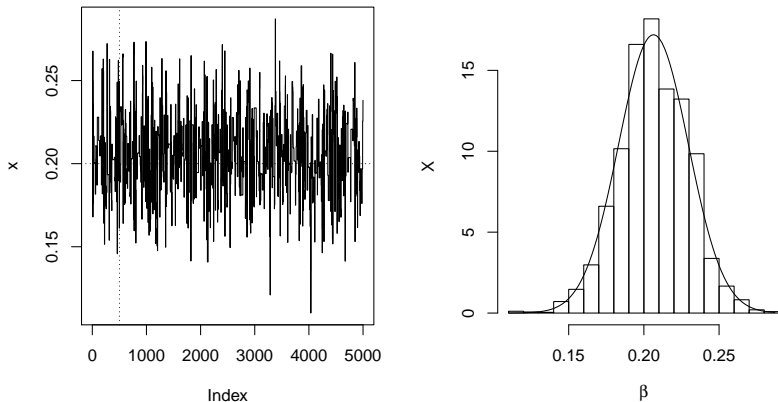
如 (1) 所示. 若  $U \leq A$  则接受  $y$  且令  $X_{t+1} = y$ , 否则令  $X_{t+1} = X_t$ .

- (d) 增加  $t$ , 返回到 (a).

## 6.5 Metropolis-Hastings 算法

## 6.5.3 M-H 抽样方法的几个变种

图 6.5.5

图 6.5.5 随机游动 Metropolis 算法生成  $\beta$  的链和直方图

## 续例 6.5.4

- 图 6.5.5 的左图表明：链的路径图显示链已经近似地收敛到目标分布. 去掉链的预烧期，生成的链可以用于估计  $\beta$ . 由图 6.5.5 的右边的直方图似乎可以看出  $\beta$  的均值接近 0.2 .
- 由链的样本产生的 5 种股票的胜出的天数、胜出频率、以及 MCMC 方法估计多项分布先验概率比的列表如下：

胜出天数： 84, 64, 48, 39, 15

胜出频率： 0.336, 0.256, 0.192, 0.156, 0.060

先验机会比： 0.333, 0.267, 0.200, 0.133, 0.067

生成链的均值： 0.2164363

生成链的标准差： 0.02319191

## 续例 6.5.4

- 图 6.5.5 的左图表明：链的路径图显示链已经近似地收敛到目标分布. 去掉链的预烧期，生成的链可以用于估计  $\beta$ . 由图 6.5.5 的右边的直方图似乎可以看出  $\beta$  的均值接近 0.2 .
- 由链的样本产生的 5 种股票的胜出的天数、胜出频率、以及 MCMC 方法估计多项分布先验概率比的列表如下：

胜出天数： 84, 64, 48, 39, 15

胜出频率： 0.336, 0.256, 0.192, 0.156, 0.060

先验机会比： 0.333, 0.267, 0.200, 0.133, 0.067

生成链的均值： 0.2164363

生成链的标准差： 0.02319191

## 续例 6.5.4

- 图 6.5.5 的左图表明：链的路径图显示链已经近似地收敛到目标分布. 去掉链的预烧期，生成的链可以用于估计  $\beta$ . 由图 6.5.5 的右边的直方图似乎可以看出  $\beta$  的均值接近 0.2 .
- 由链的样本产生的 5 种股票的胜出的天数、胜出频率、以及 MCMC 方法估计多项分布先验概率比的列表如下：

胜出天数： 84, 64, 48, 39, 15

胜出频率： 0.336, 0.256, 0.192, 0.156, 0.060

先验机会比： 0.333, 0.267, 0.200, 0.133, 0.067

生成链的均值： 0.2164363

生成链的标准差： 0.02319191

## 续例 6.5.4

- 图 6.5.5 的左图表明：链的路径图显示链已经近似地收敛到目标分布. 去掉链的预烧期，生成的链可以用于估计  $\beta$ . 由图 6.5.5 的右边的直方图似乎可以看出  $\beta$  的均值接近 0.2 .
- 由链的样本产生的 5 种股票的胜出的天数、胜出频率、以及 MCMC 方法估计多项分布先验概率比的列表如下：

胜出天数： 84, 64, 48, 39, 15

胜出频率： 0.336, 0.256, 0.192, 0.156, 0.060

先验机会比： 0.333, 0.267, 0.200, 0.133, 0.067

生成链的均值： 0.2164363

生成链的标准差： 0.02319191

## 续例 6.5.4

- 图 6.5.5 的左图表明：链的路径图显示链已经近似地收敛到目标分布. 去掉链的预烧期，生成的链可以用于估计  $\beta$ . 由图 6.5.5 的右边的直方图似乎可以看出  $\beta$  的均值接近 0.2 .
- 由链的样本产生的 5 种股票的胜出的天数、胜出频率、以及 MCMC 方法估计多项分布先验概率比的列表如下：

胜出天数： 84, 64, 48, 39, 15

胜出频率： 0.336, 0.256, 0.192, 0.156, 0.060

先验机会比： 0.333, 0.267, 0.200, 0.133, 0.067

生成链的均值： 0.2164363

生成链的标准差： 0.02319191



## 续例 6.5.4

- 图 6.5.5 的左图表明：链的路径图显示链已经近似地收敛到目标分布. 去掉链的预烧期，生成的链可以用于估计  $\beta$ . 由图 6.5.5 的右边的直方图似乎可以看出  $\beta$  的均值接近 0.2 .
- 由链的样本产生的 5 种股票的胜出的天数、胜出频率、以及 MCMC 方法估计多项分布先验概率比的列表如下：

胜出天数： 84, 64, 48, 39, 15

胜出频率： 0.336, 0.256, 0.192, 0.156, 0.060

先验机会比： 0.333, 0.267, 0.200, 0.133, 0.067

生成链的均值： 0.2164363

生成链的标准差： 0.02319191

## 续例 6.5.4

- 图 6.5.5 的左图表明：链的路径图显示链已经近似地收敛到目标分布. 去掉链的预烧期，生成的链可以用于估计  $\beta$ . 由图 6.5.5 的右边的直方图似乎可以看出  $\beta$  的均值接近 0.2 .
- 由链的样本产生的 5 种股票的胜出的天数、胜出频率、以及 MCMC 方法估计多项分布先验概率比的列表如下：

胜出天数： 84, 64, 48, 39, 15

胜出频率： 0.336, 0.256, 0.192, 0.156, 0.060

先验机会比： 0.333, 0.267, 0.200, 0.133, 0.067

生成链的均值： 0.2164363

生成链的标准差： 0.02319191

## 续例 6.5.4

- 图 6.5.5 的左图表明：链的路径图显示链已经近似地收敛到目标分布. 去掉链的预烧期，生成的链可以用于估计  $\beta$ . 由图 6.5.5 的右边的直方图似乎可以看出  $\beta$  的均值接近 0.2 .
- 由链的样本产生的 5 种股票的胜出的天数、胜出频率、以及 MCMC 方法估计多项分布先验概率比的列表如下：

胜出天数： 84, 64, 48, 39, 15

胜出频率： 0.336, 0.256, 0.192, 0.156, 0.060

先验机会比： 0.333, 0.267, 0.200, 0.133, 0.067

生成链的均值： 0.2164363

生成链的标准差： 0.02319191

### 3. 独立抽样方法

- Metropolis-Hastings 抽样方法的另一个特殊情形是独立抽样.
- 独立抽样中的提议分布不依赖于链的前一步状态值. 因此  $g(Y|X_t) = g(Y)$ , 接受概率为

$$\alpha(X_t, Y) = \min \left\{ 1, \frac{f(Y)g(X_t)}{f(X_t)g(Y)} \right\}.$$

- 独立抽样方法容易实施, 而且在提议分布和目标分布很接近时也趋于表现很好, 但是当提议分布和目标分布差别很大时, 其表现就较差.
- Robert (1996) 讨论了独立抽样的收敛性, 并且说道: “独立抽样算法作为单独的算法很少是有用的”. 但是不管怎么样, 我们仍然用下例来说明这种方法的应用. 因为独立抽样方法在混合的 MCMC 方法中是比较有用的.

### 3. 独立抽样方法

- Metropolis-Hastings 抽样方法的另一个特殊情形是独立抽样.
- 独立抽样中的提议分布不依赖于链的前一步状态值. 因此  $g(Y|X_t) = g(Y)$ , 接受概率为

$$\alpha(X_t, Y) = \min \left\{ 1, \frac{f(Y)g(X_t)}{f(X_t)g(Y)} \right\}.$$

- 独立抽样方法容易实施, 而且在提议分布和目标分布很接近时也趋于表现很好, 但是当提议分布和目标分布差别很大时, 其表现就较差.
- Robert (1996) 讨论了独立抽样的收敛性, 并且说道: “独立抽样算法作为单独的算法很少是有用的”. 但是不管怎么样, 我们仍然用下例来说明这种方法的应用. 因为独立抽样方法在混合的 MCMC 方法中是比较有用的.

### 3. 独立抽样方法

- Metropolis-Hastings 抽样方法的另一个特殊情形是独立抽样.
- 独立抽样中的提议分布不依赖于链的前一步状态值. 因此  $g(Y|X_t) = g(Y)$ , 接受概率为

$$\alpha(X_t, Y) = \min \left\{ 1, \frac{f(Y)g(X_t)}{f(X_t)g(Y)} \right\}.$$

- 独立抽样方法容易实施, 而且在提议分布和目标分布很接近时也趋于表现很好, 但是当提议分布和目标分布差别很大时, 其表现就较差.
- Robert (1996) 讨论了独立抽样的收敛性, 并且说道: “独立抽样算法作为单独的算法很少是有用的”. 但是不管怎么样, 我们仍然用下例来说明这种方法的应用. 因为独立抽样方法在混合的 MCMC 方法中是比较有用的.

### 3. 独立抽样方法

- Metropolis-Hastings 抽样方法的另一个特殊情形是独立抽样.
- 独立抽样中的提议分布不依赖于链的前一步状态值. 因此  $g(Y|X_t) = g(Y)$ , 接受概率为

$$\alpha(X_t, Y) = \min \left\{ 1, \frac{f(Y)g(X_t)}{f(X_t)g(Y)} \right\}.$$

- 独立抽样方法容易实施, 而且在提议分布和目标分布很接近时也趋于表现很好, 但是当提议分布和目标分布差别很大时, 其表现就较差.
- Robert (1996) 讨论了独立抽样的收敛性, 并且说道: “独立抽样算法作为单独的算法很少是有用的”. 但是不管怎么样, 我们仍然用下例来说明这种方法的应用. 因为独立抽样方法在混合的 MCMC 方法中是比较有用的.

## 独立抽样方法的例子

- **例 6.5.5 (独立抽样)** 假设从一个正态混合分布

$$p N(\mu_1, \sigma_1^2) + (1 - p) N(\mu_2, \sigma_2^2)$$

中观测到一个样本  $\mathbf{z} = (z_1, \dots, z_n)$ , 求  $p$  的估计.

- **解** 显然, 混合正态的密度为

$$f(\mathbf{z}|\mathbf{p}) = pf_1(\mathbf{z}|\mathbf{p}) + (1 - p)f_2(\mathbf{z}|\mathbf{p}),$$

其中  $f_1, f_2$  分别为两个正态的密度.

- 设  $p$  的先验分布  $\pi(p)$  为  $(0, 1)$  上的均匀分布  $U(0, 1)$ , 则  $p$  的后验分布

$$\pi(p|\mathbf{z}) \propto f(\mathbf{z}|\mathbf{p})\pi(p) = \prod_{j=1}^n [pf_1(z_j|\mathbf{p}) + (1 - p)f_2(z_j|\mathbf{p})]$$

- 提议分布的支撑应和  $p$  的取值范围  $(0, 1)$  相同. 在没有先验信息的情况下, 这里使用贝塔分布  $Be(1, 1)$  作为提议分布.



## 独立抽样方法的例子

- **例 6.5.5 (独立抽样)** 假设从一个正态混合分布

$$p N(\mu_1, \sigma_1^2) + (1 - p) N(\mu_2, \sigma_2^2)$$

中观测到一个样本  $\mathbf{z} = (z_1, \dots, z_n)$ , 求  $p$  的估计.

- **解** 显然, 混合正态的密度为

$$f(\mathbf{z}|p) = pf_1(\mathbf{z}|p) + (1 - p)f_2(\mathbf{z}|p),$$

其中  $f_1, f_2$  分别为两个正态的密度.

- 设  $p$  的先验分布  $\pi(p)$  为  $(0, 1)$  上的均匀分布  $U(0, 1)$ , 则  $p$  的后验分布

$$\pi(p|\mathbf{z}) \propto f(\mathbf{z}|p)\pi(p) = \prod_{j=1}^n [pf_1(z_j|p) + (1 - p)f_2(z_j|p)]$$

- 提议分布的支撑应和  $p$  的取值范围  $(0, 1)$  相同. 在没有先验信息的情况下, 这里使用贝塔分布  $Be(1, 1)$  作为提议分布.

## 独立抽样方法的例子

- **例 6.5.5 (独立抽样)** 假设从一个正态混合分布

$$p N(\mu_1, \sigma_1^2) + (1 - p) N(\mu_2, \sigma_2^2)$$

中观测到一个样本  $\mathbf{z} = (z_1, \dots, z_n)$ , 求  $p$  的估计.

- **解** 显然, 混合正态的密度为

$$f(\mathbf{z}|p) = pf_1(\mathbf{z}|p) + (1 - p)f_2(\mathbf{z}|p),$$

其中  $f_1, f_2$  分别为两个正态的密度.

- 设  $p$  的先验分布  $\pi(p)$  为  $(0, 1)$  上的均匀分布  $U(0, 1)$ , 则  $p$  的后验分布

$$\pi(p|\mathbf{z}) \propto f(\mathbf{z}|p)\pi(p) = \prod_{j=1}^n [pf_1(z_j|p) + (1 - p)f_2(z_j|p)]$$

- 提议分布的支撑应和  $p$  的取值范围  $(0, 1)$  相同. 在没有先验信息的情况下, 这里使用贝塔分布  $Be(1, 1)$  作为提议分布.

## 续例 6.5.5

- 候选点  $Y$  被接受的概率为

$$\alpha(X_t, Y) = \min \left\{ 1, \frac{\pi(Y|\mathbf{z})g(X_t)}{\pi(X_t|\mathbf{z})g(Y)} \right\}.$$

其中  $g$  提议分布, 其密度函数  $g(y) \propto y^{a-1}(1-y)^{b-1}$ ,  $\pi(\cdot|\mathbf{z})$  为目标分布. 此处

$$\frac{\pi(y|\mathbf{z})g(x_t)}{\pi(x_t|\mathbf{z})g(y)} = \frac{\prod_{j=1}^n [yf_1(z_j|y) + (1-y)f_2(z_j|y)] \cdot x_t^{a-1}(1-x_t)^{b-1}}{\prod_{j=1}^n [x_tf_1(z_j|x_t) + (1-x_t)f_2(z_j|x_t)] \cdot y^{a-1}(1-y)^{b-1}}.$$

- 下面我们进行模拟, 提议分布取为  $Be(a, b)$ , 当取  $a = b = 1$  时就是均匀分布  $U(0, 1)$ . 观测数据从下述正态混合中产生

$$0.2N(0, 1) + 0.8N(5, 1)$$

## 续例 6.5.5

- 候选点  $Y$  被接受的概率为

$$\alpha(X_t, Y) = \min \left\{ 1, \frac{\pi(Y|\mathbf{z})g(X_t)}{\pi(X_t|\mathbf{z})g(Y)} \right\}.$$

其中  $g$  提议分布, 其密度函数  $g(y) \propto y^{a-1}(1-y)^{b-1}$ ,  $\pi(\cdot|\mathbf{z})$  为目标分布. 此处

$$\frac{\pi(y|\mathbf{z})g(x_t)}{\pi(x_t|\mathbf{z})g(y)} = \frac{\prod_{j=1}^n [yf_1(z_j|y) + (1-y)f_2(z_j|y)] \cdot x_t^{a-1}(1-x_t)^{b-1}}{\prod_{j=1}^n [x_tf_1(z_j|x_t) + (1-x_t)f_2(z_j|x_t)] \cdot y^{a-1}(1-y)^{b-1}}.$$

- 下面我们进行模拟, 提议分布取为  $Be(a, b)$ , 当取  $a = b = 1$  时就是均匀分布  $U(0, 1)$ . 观测数据从下述正态混合中产生

$$0.2N(0, 1) + 0.8N(5, 1)$$

## 续例 6.5.5

- 候选点  $Y$  被接受的概率为

$$\alpha(X_t, Y) = \min \left\{ 1, \frac{\pi(Y|\mathbf{z})g(X_t)}{\pi(X_t|\mathbf{z})g(Y)} \right\}.$$

其中  $g$  提议分布, 其密度函数  $g(y) \propto y^{a-1}(1-y)^{b-1}$ ,  $\pi(\cdot|\mathbf{z})$  为目标分布. 此处

$$\frac{\pi(y|\mathbf{z})g(x_t)}{\pi(x_t|\mathbf{z})g(y)} = \frac{\prod_{j=1}^n [yf_1(z_j|y) + (1-y)f_2(z_j|y)] \cdot x_t^{a-1}(1-x_t)^{b-1}}{\prod_{j=1}^n [x_tf_1(z_j|x_t) + (1-x_t)f_2(z_j|x_t)] \cdot y^{a-1}(1-y)^{b-1}}.$$

- 下面我们进行模拟, 提议分布取为  $Be(a, b)$ , 当取  $a = b = 1$  时就是均匀分布  $U(0, 1)$ . 观测数据从下述正态混合中产生

$$0.2N(0, 1) + 0.8N(5, 1)$$

## 独立抽样的算法步骤

- (1) 令提议分布为贝塔分布  $Be(a, b)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, N$  重复下列步骤:
  - (a) 从提议分布  $Be(a, b)$  中产生一个候选值  $y$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[t-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{\pi(y|\mathbf{z})g(x_t)}{\pi(x_t|\mathbf{z})g(y)}$$
如前所示. 若  $U \leq A$  接受  $y$ , 令  $X_{t+1} = y$ ; 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 独立抽样的算法步骤

- (1) 令提议分布为贝塔分布  $Be(a, b)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, N$  重复下列步骤:
  - (a) 从提议分布  $Be(a, b)$  中产生一个候选值  $y$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[t-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{\pi(y|\mathbf{z})g(x_t)}{\pi(x_t|\mathbf{z})g(y)}$$
如前所示. 若  $U \leq A$  接受  $y$ , 令  $X_{t+1} = y$ ; 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 独立抽样的算法步骤

- (1) 令提议分布为贝塔分布  $Be(a, b)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, N$  重复下列步骤:
  - (a) 从提议分布  $Be(a, b)$  中产生一个候选值  $y$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[t-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{\pi(y|\mathbf{z})g(x_t)}{\pi(x_t|\mathbf{z})g(y)}$$
如前所示. 若  $U \leq A$  接受  $y$ , 令  $X_{t+1} = y$ ; 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).



## 独立抽样的算法步骤

- (1) 令提议分布为贝塔分布  $Be(a, b)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, N$  重复下列步骤:
  - (a) 从提议分布  $Be(a, b)$  中产生一个候选值  $y$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[t-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{\pi(y|z)g(x_t)}{\pi(x_t|z)g(y)}$$
如前所示. 若  $U \leq A$  接受  $y$ , 令  $X_{t+1} = y$ ; 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 独立抽样的算法步骤

- (1) 令提议分布为贝塔分布  $Be(a, b)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, N$  重复下列步骤:
  - (a) 从提议分布  $Be(a, b)$  中产生一个候选值  $y$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[t-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{\pi(y|\mathbf{z})g(x_t)}{\pi(x_t|\mathbf{z})g(y)}$$
如前所示. 若  $U \leq A$  接受  $y$ , 令  $X_{t+1} = y$ ; 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 独立抽样的算法步骤

- (1) 令提议分布为贝塔分布  $Be(a, b)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, N$  重复下列步骤:
  - (a) 从提议分布  $Be(a, b)$  中产生一个候选值  $y$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i - 1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{\pi(y|\mathbf{z})g(x_t)}{\pi(x_t|\mathbf{z})g(y)}$$
如前所示. 若  $U \leq A$  接受  $y$ , 令  $X_{t+1} = y$ ; 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 独立抽样的算法步骤

- (1) 令提议分布为贝塔分布  $Be(a, b)$ .
- (2) 从提议分布中生成初始值  $X_0$ , 并存入  $X[1]$  中.
- (3) 对  $t = 2, 3, \dots, N$  重复下列步骤:
  - (a) 从提议分布  $Be(a, b)$  中产生一个候选值  $y$ .
  - (b) 从均匀分布  $U(0, 1)$  中生成随机数  $U$ .
  - (c) 由  $X_t = X[i-1]$ , 计算接受概率  $\alpha(x_t, y) = \min\{1, A\}$ , 其中
$$A = \frac{\pi(y|\mathbf{z})g(x_t)}{\pi(x_t|\mathbf{z})g(y)}$$
如前所示. 若  $U \leq A$  接受  $y$ , 令  $X_{t+1} = y$ ; 否则令  $X_{t+1} = X_t$ .
  - (d) 增加  $t$ , 返回到 (a).

## 图 6.5.6

链的路径图和丢掉 100 个预烧期样本后的直方图如下:

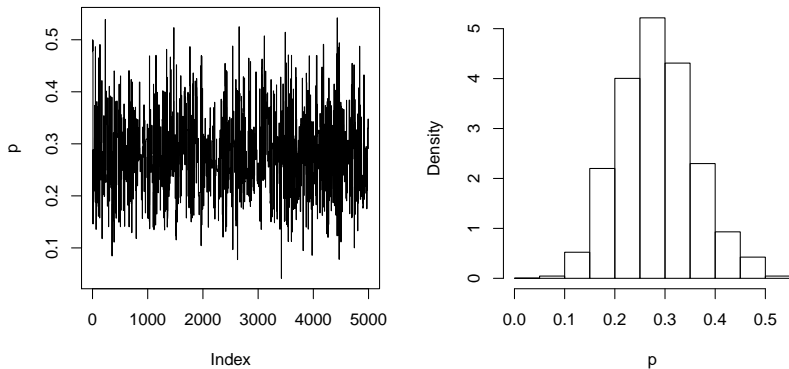


图 6.5.6 提议分布为  $Be(1,1)$  时独立抽样生成的链和直方图

## 续例 6.5.5

- 丢掉 100 个预烧期样本后, 由图 6.5.6 左图可见链的时间状态图显示链混合的很好, 很快收敛到平稳分布. 图 6.5.6 右图是生成链的直方图, 由图可见保留样本的样本均值是 0.2553, 它就是  $p$  的估计值.

## 4. 逐分量的 M-H 抽样方法

- 当状态空间为多维时, 不整体更新  $\mathbf{X}_n$ , 而是对其分量进行逐个更新, 即称为逐分量的 M-H 抽样方法. 这样做更方便和更有效率.
- 记

$$\mathbf{X}_n = (X_{n,1}, \dots, X_{n,k}),$$

$$\mathbf{X}_{n,-i} = (X_{n,1}, \dots, X_{n,i-1}, X_{n,i+1}, \dots, X_{n,k}).$$

分别表示在第  $n$  步链的状态, 及在第  $n$  步除第  $i$  个分量外其他分量的状态.  $f(\mathbf{x}) = f(x_1, \dots, x_k)$  为目标分布,

$$f(x_i | \mathbf{x}_{-i}) = f(\mathbf{x}) / \int f(x_1, \dots, x_k) dx_i$$

表示  $X_i$  对其他分量的条件密度.

## 4. 逐分量的 M-H 抽样方法

- 当状态空间为多维时, 不整体更新  $\mathbf{X}_n$ , 而是对其分量进行逐个更新, 即称为逐分量的 M-H 抽样方法. 这样做更方便和更有效率.
- 记

$$\mathbf{X}_n = (X_{n,1}, \dots, X_{n,k}),$$

$$\mathbf{X}_{n,-i} = (X_{n,1}, \dots, X_{n,i-1}, X_{n,i+1}, \dots, X_{n,k}).$$

分别表示在第  $n$  步链的状态, 及在第  $n$  步除第  $i$  个分量外其他分量的状态.  $f(\mathbf{x}) = f(x_1, \dots, x_k)$  为目标分布,

$$f(x_i | \mathbf{x}_{-i}) = f(\mathbf{x}) / \int f(x_1, \dots, x_k) dx_i$$

表示  $X_i$  对其他分量的条件密度.



## 逐分量的 M-H 抽样方法

- 逐分量的 M-H 抽样方法由  $k$  步构成：令  $X_{n,i}$  表示在第  $n$  次迭代后  $\mathbf{X}_n$  第  $i$  个分量的状态，则在第  $n+1$  步迭代的第  $i$  步中，使用 M-H 算法更新  $X_{n,i}$ 。做法如下：
- 对  $i = 1, \dots, k$ ，从第  $i$  个提议分布  $q_i(\cdot | X_{n,i}, \mathbf{X}_{n,-i}^*)$  中产生  $Y_i$ ，这里

$$\mathbf{X}_{n,-i}^* = (X_{n+1,1}, \dots, X_{n+1,i-1}, X_{n,i+1}, \dots, X_{n,k}).$$

- 然后以概率

$$\alpha(\mathbf{X}_{n,-i}^*, X_{n,i}, Y_i) = \min \left\{ 1, \frac{f(Y_i | \mathbf{X}_{n,-i}^*) q_i(X_{n,i} | Y_i, \mathbf{X}_{n,-i}^*)}{f(X_{n,i} | \mathbf{X}_{n,-i}^*) q_i(Y_i | X_{n,i}, \mathbf{X}_{n,-i}^*)} \right\}$$

若  $Y_i$  被接受，则令  $X_{n+1,i} = Y_i$ ；否则令  $X_{n+1,i} = X_{n,i}$ 。

## 逐分量的 M-H 抽样方法

- 逐分量的 M-H 抽样方法由  $k$  步构成：令  $X_{n,i}$  表示在第  $n$  次迭代后  $\mathbf{X}_n$  第  $i$  个分量的状态，则在第  $n+1$  步迭代的第  $i$  步中，使用 M-H 算法更新  $X_{n,i}$ 。做法如下：
- 对  $i = 1, \dots, k$ ，从第  $i$  个提议分布  $q_i(\cdot | X_{n,i}, \mathbf{X}_{n,-i}^*)$  中产生  $Y_i$ ，这里

$$\mathbf{X}_{n,-i}^* = (X_{n+1,1}, \dots, X_{n+1,i-1}, X_{n,i+1}, \dots, X_{n,k}).$$

- 然后以概率

$$\alpha(\mathbf{X}_{n,-i}^*, X_{n,i}, Y_i) = \min \left\{ 1, \frac{f(Y_i | \mathbf{X}_{n,-i}^*) q_i(X_{n,i} | Y_i, \mathbf{X}_{n,-i}^*)}{f(X_{n,i} | \mathbf{X}_{n,-i}^*) q_i(Y_i | X_{n,i}, \mathbf{X}_{n,-i}^*)} \right\}$$

若  $Y_i$  被接受，则令  $X_{n+1,i} = Y_i$ ；否则令  $X_{n+1,i} = X_{n,i}$ 。

## 逐分量的 M-H 抽样方法

- 逐分量的 M-H 抽样方法由  $k$  步构成：令  $X_{n,i}$  表示在第  $n$  次迭代后  $\mathbf{X}_n$  第  $i$  个分量的状态，则在第  $n+1$  步迭代的第  $i$  步中，使用 M-H 算法更新  $X_{n,i}$ 。做法如下：
- 对  $i = 1, \dots, k$ ，从第  $i$  个提议分布  $q_i(\cdot | X_{n,i}, \mathbf{X}_{n,-i}^*)$  中产生  $Y_i$ ，这里

$$\mathbf{X}_{n,-i}^* = (X_{n+1,1}, \dots, X_{n+1,i-1}, X_{n,i+1}, \dots, X_{n,k}).$$

- 然后以概率

$$\alpha(\mathbf{X}_{n,-i}^*, X_{n,i}, Y_i) = \min \left\{ 1, \frac{f(Y_i | \mathbf{X}_{n,-i}^*) q_i(X_{n,i} | Y_i, \mathbf{X}_{n,-i}^*)}{f(X_{n,i} | \mathbf{X}_{n,-i}^*) q_i(Y_i | X_{n,i}, \mathbf{X}_{n,-i}^*)} \right\}$$

若  $Y_i$  被接受，则令  $X_{n+1,i} = Y_i$ ；否则令  $X_{n+1,i} = X_{n,i}$ 。

## 逐分量的 M-H 抽样方法的例子

- **例 6.5.6** 考虑 54 位老年人的智力测试成绩 (Wechsler Adult Intelligence Scale, WAIS, 0-20分). 研究的兴趣在于发现老年痴呆症与 WAIS 成绩的关系.
- 解 我们采用如下简单的 logistic 回归模型:

$$Y_i \sim B(1, \pi_i), \quad \ln \frac{\pi_i}{1 - \pi_i} = \beta_0 + x_i \beta_1, \quad i = 1, \dots, 54.$$

其中  $Y_i = 1$  表示第  $i$  个人老年痴呆,  $x_i$  表示第  $i$  个人的 WAIS 成绩.

- 似然函数为

$$\begin{aligned} f(\mathbf{y}|\beta_0, \beta_1) &= \prod_{i=1}^n \left( \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{1-y_i} \\ &= \exp \left\{ n\bar{y}\beta_0 + \beta_1 \sum_{i=1}^n x_i y_i - \sum_{i=1}^n \ln \left( 1 + e^{\beta_0 + x_i \beta_1} \right) \right\}. \end{aligned}$$

## 逐分量的 M-H 抽样方法的例子

- **例 6.5.6** 考虑 54 位老年人的智力测试成绩 (Wechsler Adult Intelligence Scale, WAIS, 0-20分). 研究的兴趣在于发现老年痴呆症与 WAIS 成绩的关系.
- **解** 我们采用如下简单的 logistic 回归模型:

$$Y_i \sim B(1, \pi_i), \quad \ln \frac{\pi_i}{1 - \pi_i} = \beta_0 + x_i \beta_1, \quad i = 1, \dots, 54.$$

其中  $Y_i = 1$  表示第  $i$  个人老年痴呆,  $x_i$  表示第  $i$  个人的 WAIS 成绩.

- 似然函数为

$$\begin{aligned} f(\mathbf{y}|\beta_0, \beta_1) &= \prod_{i=1}^n \left( \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{1-y_i} \\ &= \exp \left\{ n\bar{y}\beta_0 + \beta_1 \sum_{i=1}^n x_i y_i - \sum_{i=1}^n \ln \left( 1 + e^{\beta_0 + x_i \beta_1} \right) \right\}. \end{aligned}$$

## 逐分量的 M-H 抽样方法的例子

- **例 6.5.6** 考虑 54 位老年人的智力测试成绩 (Wechsler Adult Intelligence Scale, WAIS, 0-20分). 研究的兴趣在于发现老年痴呆症与 WAIS 成绩的关系.
- **解** 我们采用如下简单的 logistic 回归模型:

$$Y_i \sim B(1, \pi_i), \quad \ln \frac{\pi_i}{1 - \pi_i} = \beta_0 + x_i \beta_1, \quad i = 1, \dots, 54.$$

其中  $Y_i = 1$  表示第  $i$  个人老年痴呆,  $x_i$  表示第  $i$  个人的 WAIS 成绩.

- 似然函数为

$$\begin{aligned} f(\mathbf{y}|\beta_0, \beta_1) &= \prod_{i=1}^n \left( \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{1-y_i} \\ &= \exp \left\{ n\bar{y}\beta_0 + \beta_1 \sum_{i=1}^n x_i y_i - \sum_{i=1}^n \ln \left( 1 + e^{\beta_0 + x_i \beta_1} \right) \right\}. \end{aligned}$$

## 续例 6.5.6

- 考虑  $\beta_0, \beta_1$  的先验分布  $\pi$  为独立的正态分布, 即  $\pi(\beta_0, \beta_1) = \pi_1(\beta_0) \pi_2(\beta_1)$ , 而

$$\beta_j \sim N(\mu_{\beta_j}, \sigma_j^2), \quad j = 0, 1.$$

其中当  $\mu_{\beta_j} = 0, \sigma_j^2$  很大时, 上述先验分布接近无信息先验.

- 因此后验分布为

$$\begin{aligned} \pi(\beta_0, \beta_1 | \mathbf{y}) &\propto f(\mathbf{y} | \beta_0, \beta_1) \pi(\beta_0, \beta_1) \\ &\propto \exp \left\{ \sum_{i=1}^n \left[ (\beta_0 + \beta_1 x_i) y_i - \ln (1 + e^{\beta_0 + x_i \beta_1}) \right] \right. \\ &\quad \left. - \frac{(\beta_0 - \mu_{\beta_0})^2}{2\sigma_0^2} - \frac{(\beta_1 - \mu_{\beta_1})^2}{2\sigma_1^2} \right\}. \end{aligned}$$

## 续例 6.5.6

- 考虑  $\beta_0, \beta_1$  的先验分布  $\pi$  为独立的正态分布, 即  $\pi(\beta_0, \beta_1) = \pi_1(\beta_0) \pi_2(\beta_1)$ , 而

$$\beta_j \sim N(\mu_{\beta_j}, \sigma_j^2), \quad j = 0, 1.$$

其中当  $\mu_{\beta_j} = 0, \sigma_j^2$  很大时, 上述先验分布接近无信息先验.

- 因此后验分布为

$$\begin{aligned} \pi(\beta_0, \beta_1 | \mathbf{y}) &\propto f(\mathbf{y} | \beta_0, \beta_1) \pi(\beta_0, \beta_1) \\ &\propto \exp \left\{ \sum_{i=1}^n \left[ (\beta_0 + \beta_1 x_i) y_i - \ln (1 + e^{\beta_0 + x_i \beta_1}) \right] \right. \\ &\quad \left. - \frac{(\beta_0 - \mu_{\beta_0})^2}{2\sigma_0^2} - \frac{(\beta_1 - \mu_{\beta_1})^2}{2\sigma_1^2} \right\}. \end{aligned}$$



## 续例 6.5.6

- 下面我们使用逐分量的 M-H 算法从此分布中产生随机数. 提议分布取为

$$\boldsymbol{\beta}' \sim N_2(\boldsymbol{\beta}, \text{diag}\{\bar{s}_{\beta_0}^2, \bar{s}_{\beta_1}^2\}) \Leftrightarrow \beta'_0 \sim N(\beta_0, \bar{s}_{\beta_0}^2), \beta'_1 \sim N(\beta_1, \bar{s}_{\beta_1}^2).$$

此处  $\boldsymbol{\beta}' = (\beta'_0, \beta'_1)^\tau$ , 而  $\boldsymbol{\beta} = (\beta_0, \beta_1)^\tau$ .

- 在 M-H 算法中, 按分量进行逐个更新, 其优势在于应用方便, 不需要考虑调节参数. 算法如下:

对  $t = 1, \dots, T$ :

- (1) 令  $\boldsymbol{\beta} = (\beta_0^{(t-1)}, \beta_1^{(t-1)})^\tau$ .
- (2) 从提议分布  $N(\beta_0, \bar{s}_{\beta_0}^2)$  产生候选点  $\beta'_0$ .

## 续例 6.5.6

- 下面我们使用逐分量的 M-H 算法从此分布中产生随机数. 提议分布取为

$$\boldsymbol{\beta}' \sim N_2(\boldsymbol{\beta}, \text{diag}\{\bar{s}_{\beta_0}^2, \bar{s}_{\beta_1}^2\}) \Leftrightarrow \beta'_0 \sim N(\beta_0, \bar{s}_{\beta_0}^2), \beta'_1 \sim N(\beta_1, \bar{s}_{\beta_1}^2).$$

此处  $\boldsymbol{\beta}' = (\beta'_0, \beta'_1)^\tau$ , 而  $\boldsymbol{\beta} = (\beta_0, \beta_1)^\tau$ .

- 在 M-H 算法中, 按分量进行逐个更新, 其优势在于应用方便, 不需要考虑调节参数. 算法如下:

对  $t = 1, \dots, T$ :

- (1) 令  $\boldsymbol{\beta} = (\beta_0^{(t-1)}, \beta_1^{(t-1)})^\tau$ .
- (2) 从提议分布  $N(\beta_0, \bar{s}_{\beta_0}^2)$  产生候选点  $\beta'_0$ .

## 续例 6.5.6

- 下面我们使用逐分量的 M-H 算法从此分布中产生随机数. 提议分布取为

$$\boldsymbol{\beta}' \sim N_2(\boldsymbol{\beta}, \text{diag}\{\bar{s}_{\beta_0}^2, \bar{s}_{\beta_1}^2\}) \Leftrightarrow \beta'_0 \sim N(\beta_0, \bar{s}_{\beta_0}^2), \beta'_1 \sim N(\beta_1, \bar{s}_{\beta_1}^2).$$

此处  $\boldsymbol{\beta}' = (\beta'_0, \beta'_1)^\tau$ , 而  $\boldsymbol{\beta} = (\beta_0, \beta_1)^\tau$ .

- 在 M-H 算法中, 按分量进行逐个更新, 其优势在于应用方便, 不需要考虑调节参数. 算法如下:

对  $t = 1, \dots, T$ :

- (1) 令  $\boldsymbol{\beta} = (\beta_0^{(t-1)}, \beta_1^{(t-1)})^\tau$ .
- (2) 从提议分布  $N(\beta_0, \bar{s}_{\beta_0}^2)$  产生候选点  $\beta'_0$ .

## 续例 6.5.6

- 下面我们使用逐分量的 M-H 算法从此分布中产生随机数. 提议分布取为

$$\boldsymbol{\beta}' \sim N_2(\boldsymbol{\beta}, \text{diag}\{\bar{s}_{\beta_0}^2, \bar{s}_{\beta_1}^2\}) \Leftrightarrow \beta'_0 \sim N(\beta_0, \bar{s}_{\beta_0}^2), \beta'_1 \sim N(\beta_1, \bar{s}_{\beta_1}^2).$$

此处  $\boldsymbol{\beta}' = (\beta'_0, \beta'_1)^\tau$ , 而  $\boldsymbol{\beta} = (\beta_0, \beta_1)^\tau$ .

- 在 M-H 算法中, 按分量进行逐个更新, 其优势在于应用方便, 不需要考虑调节参数. 算法如下:

对  $t = 1, \dots, T$ :

- (1) 令  $\boldsymbol{\beta} = (\beta_0^{(t-1)}, \beta_1^{(t-1)})^\tau$ .
- (2) 从提议分布  $N(\beta_0, \bar{s}_{\beta_0}^2)$  产生候选点  $\beta'_0$ .

## 续例 6.5.6

- (3) 令  $\beta' = (\beta'_0, \beta_1^{(t-1)})^\tau$ , 接受概率  $\alpha_0(\beta, \beta') = \min\{1, A\}$ , 其中

$$A = \frac{\pi(\beta'_0, \beta_1 | \mathbf{y})}{\pi(\beta_0, \beta_1 | \mathbf{y})} = \frac{f(\mathbf{y} | \beta'_0, \beta_1) \pi(\beta'_0, \beta_1)}{f(\mathbf{y} | \beta_0, \beta_1) \pi(\beta_0, \beta_1)}.$$

- (4) 以概率  $\alpha_0(\beta, \beta')$  接受  $\beta = \beta'$ , 即将  $\beta$  的第一个分量  $\beta_0$  更新为  $\beta'_0$ ; 否则  $\beta$  保持不变.
- (5) 从提议分布  $N(\beta_1, \bar{s}_{\beta_1}^2)$  产生候选点  $\beta'_1$ .

- (6) 令  $\beta' = (\beta_0, \beta'_1)^\tau$ , 接受概率  $\alpha_1(\beta, \beta') = \min\{1, B\}$ , 其中

$$B = \frac{\pi(\beta_0, \beta'_1 | \mathbf{y})}{\pi(\beta_0, \beta_1 | \mathbf{y})} = \frac{f(\mathbf{y} | \beta_0, \beta'_1) \pi(\beta_0, \beta'_1)}{f(\mathbf{y} | \beta_0, \beta_1) \pi(\beta_0, \beta_1)}.$$

- (7) 以概率  $\alpha_1(\beta, \beta')$  接受  $\beta = \beta'$ , 即将  $\beta$  的第二个分量  $\beta_1$  更新为  $\beta'_1$ ; 否则  $\beta$  保持不变.
- (8) 令  $\beta^{(t)} = \beta$ .

## 续例 6.5.6

- (3) 令  $\beta' = (\beta'_0, \beta_1^{(t-1)})^\tau$ , 接受概率  $\alpha_0(\beta, \beta') = \min\{1, A\}$ , 其中

$$A = \frac{\pi(\beta'_0, \beta_1 | \mathbf{y})}{\pi(\beta_0, \beta_1 | \mathbf{y})} = \frac{f(\mathbf{y} | \beta'_0, \beta_1) \pi(\beta'_0, \beta_1)}{f(\mathbf{y} | \beta_0, \beta_1) \pi(\beta_0, \beta_1)}.$$

- (4) 以概率  $\alpha_0(\beta, \beta')$  接受  $\beta = \beta'$ , 即将  $\beta$  的第一个分量  $\beta_0$  更新为  $\beta'_0$ ; 否则  $\beta$  保持不变.
- (5) 从提议分布  $N(\beta_1, \bar{s}_{\beta_1}^2)$  产生候选点  $\beta'_1$ .

- (6) 令  $\beta' = (\beta_0, \beta'_1)^\tau$ , 接受概率  $\alpha_1(\beta, \beta') = \min\{1, B\}$ , 其中

$$B = \frac{\pi(\beta_0, \beta'_1 | \mathbf{y})}{\pi(\beta_0, \beta_1 | \mathbf{y})} = \frac{f(\mathbf{y} | \beta_0, \beta'_1) \pi(\beta_0, \beta'_1)}{f(\mathbf{y} | \beta_0, \beta_1) \pi(\beta_0, \beta_1)}.$$

- (7) 以概率  $\alpha_1(\beta, \beta')$  接受  $\beta = \beta'$ , 即将  $\beta$  的第二个分量  $\beta_1$  更新为  $\beta'_1$ ; 否则  $\beta$  保持不变.
- (8) 令  $\beta^{(t)} = \beta$ .

## 续例 6.5.6

- (3) 令  $\beta' = (\beta'_0, \beta_1^{(t-1)})^\tau$ , 接受概率  $\alpha_0(\beta, \beta') = \min\{1, A\}$ , 其中

$$A = \frac{\pi(\beta'_0, \beta_1 | \mathbf{y})}{\pi(\beta_0, \beta_1 | \mathbf{y})} = \frac{f(\mathbf{y} | \beta'_0, \beta_1) \pi(\beta'_0, \beta_1)}{f(\mathbf{y} | \beta_0, \beta_1) \pi(\beta_0, \beta_1)}.$$

- (4) 以概率  $\alpha_0(\beta, \beta')$  接受  $\beta = \beta'$ , 即将  $\beta$  的第一个分量  $\beta_0$  更新为  $\beta'_0$ ; 否则  $\beta$  保持不变.
- (5) 从提议分布  $N(\beta_1, \bar{\sigma}_{\beta_1}^2)$  产生候选点  $\beta'_1$ .

- (6) 令  $\beta' = (\beta_0, \beta'_1)^\tau$ , 接受概率  $\alpha_1(\beta, \beta') = \min\{1, B\}$ , 其中

$$B = \frac{\pi(\beta_0, \beta'_1 | \mathbf{y})}{\pi(\beta_0, \beta_1 | \mathbf{y})} = \frac{f(\mathbf{y} | \beta_0, \beta'_1) \pi(\beta_0, \beta'_1)}{f(\mathbf{y} | \beta_0, \beta_1) \pi(\beta_0, \beta_1)}.$$

- (7) 以概率  $\alpha_1(\beta, \beta')$  接受  $\beta = \beta'$ , 即将  $\beta$  的第二个分量  $\beta_1$  更新为  $\beta'_1$ ; 否则  $\beta$  保持不变.
- (8) 令  $\beta^{(t)} = \beta$ .

## 续例 6.5.6

- (3) 令  $\beta' = (\beta'_0, \beta_1^{(t-1)})^\tau$ , 接受概率  $\alpha_0(\beta, \beta') = \min\{1, A\}$ , 其中

$$A = \frac{\pi(\beta'_0, \beta_1 | \mathbf{y})}{\pi(\beta_0, \beta_1 | \mathbf{y})} = \frac{f(\mathbf{y} | \beta'_0, \beta_1) \pi(\beta'_0, \beta_1)}{f(\mathbf{y} | \beta_0, \beta_1) \pi(\beta_0, \beta_1)}.$$

- (4) 以概率  $\alpha_0(\beta, \beta')$  接受  $\beta = \beta'$ , 即将  $\beta$  的第一个分量  $\beta_0$  更新为  $\beta'_0$ ; 否则  $\beta$  保持不变.
- (5) 从提议分布  $N(\beta_1, \bar{s}_{\beta_1}^2)$  产生候选点  $\beta'_1$ .

- (6) 令  $\beta' = (\beta_0, \beta'_1)^\tau$ , 接受概率  $\alpha_1(\beta, \beta') = \min\{1, B\}$ , 其中

$$B = \frac{\pi(\beta_0, \beta'_1 | \mathbf{y})}{\pi(\beta_0, \beta_1 | \mathbf{y})} = \frac{f(\mathbf{y} | \beta_0, \beta'_1) \pi(\beta_0, \beta'_1)}{f(\mathbf{y} | \beta_0, \beta_1) \pi(\beta_0, \beta_1)}.$$

- (7) 以概率  $\alpha_1(\beta, \beta')$  接受  $\beta = \beta'$ , 即将  $\beta$  的第二个分量  $\beta_1$  更新为  $\beta'_1$ ; 否则  $\beta$  保持不变.
- (8) 令  $\beta^{(t)} = \beta$ .



## 续例 6.5.6

- (3) 令  $\beta' = (\beta'_0, \beta_1^{(t-1)})^\tau$ , 接受概率  $\alpha_0(\beta, \beta') = \min\{1, A\}$ , 其中

$$A = \frac{\pi(\beta'_0, \beta_1 | \mathbf{y})}{\pi(\beta_0, \beta_1 | \mathbf{y})} = \frac{f(\mathbf{y} | \beta'_0, \beta_1) \pi(\beta'_0, \beta_1)}{f(\mathbf{y} | \beta_0, \beta_1) \pi(\beta_0, \beta_1)}.$$

- (4) 以概率  $\alpha_0(\beta, \beta')$  接受  $\beta = \beta'$ , 即将  $\beta$  的第一个分量  $\beta_0$  更新为  $\beta'_0$ ; 否则  $\beta$  保持不变.
- (5) 从提议分布  $N(\beta_1, \bar{s}_{\beta_1}^2)$  产生候选点  $\beta'_1$ .

- (6) 令  $\beta' = (\beta_0, \beta'_1)^\tau$ , 接受概率  $\alpha_1(\beta, \beta') = \min\{1, B\}$ , 其中

$$B = \frac{\pi(\beta_0, \beta'_1 | \mathbf{y})}{\pi(\beta_0, \beta_1 | \mathbf{y})} = \frac{f(\mathbf{y} | \beta_0, \beta'_1) \pi(\beta_0, \beta'_1)}{f(\mathbf{y} | \beta_0, \beta_1) \pi(\beta_0, \beta_1)}.$$

- (7) 以概率  $\alpha_1(\beta, \beta')$  接受  $\beta = \beta'$ , 即将  $\beta$  的第二个分量  $\beta_1$  更新为  $\beta'_1$ ; 否则  $\beta$  保持不变.
- (8) 令  $\beta^{(t)} = \beta$ .

## 续例 6.5.6

- (3) 令  $\beta' = (\beta'_0, \beta_1^{(t-1)})^\tau$ , 接受概率  $\alpha_0(\beta, \beta') = \min\{1, A\}$ , 其中

$$A = \frac{\pi(\beta'_0, \beta_1 | \mathbf{y})}{\pi(\beta_0, \beta_1 | \mathbf{y})} = \frac{f(\mathbf{y} | \beta'_0, \beta_1) \pi(\beta'_0, \beta_1)}{f(\mathbf{y} | \beta_0, \beta_1) \pi(\beta_0, \beta_1)}.$$

- (4) 以概率  $\alpha_0(\beta, \beta')$  接受  $\beta = \beta'$ , 即将  $\beta$  的第一个分量  $\beta_0$  更新为  $\beta'_0$ ; 否则  $\beta$  保持不变.
- (5) 从提议分布  $N(\beta_1, \bar{s}_{\beta_1}^2)$  产生候选点  $\beta'_1$ .

- (6) 令  $\beta' = (\beta_0, \beta'_1)^\tau$ , 接受概率  $\alpha_1(\beta, \beta') = \min\{1, B\}$ , 其中

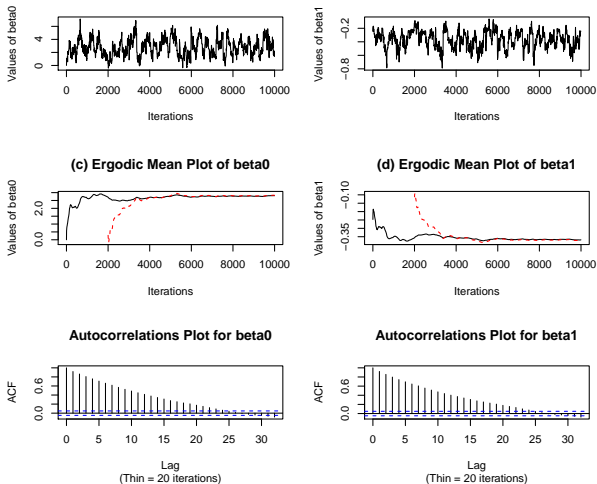
$$B = \frac{\pi(\beta_0, \beta'_1 | \mathbf{y})}{\pi(\beta_0, \beta_1 | \mathbf{y})} = \frac{f(\mathbf{y} | \beta_0, \beta'_1) \pi(\beta_0, \beta'_1)}{f(\mathbf{y} | \beta_0, \beta_1) \pi(\beta_0, \beta_1)}.$$

- (7) 以概率  $\alpha_1(\beta, \beta')$  接受  $\beta = \beta'$ , 即将  $\beta$  的第二个分量  $\beta_1$  更新为  $\beta'_1$ ; 否则  $\beta$  保持不变.
- (8) 令  $\beta^{(t)} = \beta$ .

## 6.5 Metropolis-Hastings 算法

## 6.5.3 M-H 抽样方法的几个变种

## 续例 6.5.6

图 6.5.8 逐分量 M-H 算法 logistic 回归参数  $\beta_0$  和  $\beta_1$  的诊断图

## 续例 6.5.6

- 逐分量随机游动 M-H 算法 logistic 回归参数的诊断图, 以及去掉预烧期后由链获得的  $\beta$  的两个分量  $\beta_0$  和  $\beta_1$  的样本均值和样本标准差如下:
- 图 6.5.8 最上面两个小图, 即  $\beta_0$  和  $\beta_1$  的时间状态图显示链的混合较好. 中间两个小图, 即  $\beta_0$  和  $\beta_1$  的遍历均值图, 去掉前 2000 次迭代 (即预烧期) 后趋于平稳, 也表明链的收敛性较好. 最下面两个小图, 即链的自相关函数图中显示马氏链的抽样步长  $L = 20$  次迭代, 也就是每间隔 20 个样本抽取一个, 获得的样本的自相关性就很低了, 这几种诊断方法都显示链的收敛性较好. 最后,  $\beta_0$  和  $\beta_1$  的链的样本均值分别是 2.8793214,  $-0.3737996$ . 样本标准差分别是 1.3479687, 0.1287645.

## 续例 6.5.6

- 逐分量随机游动 M-H 算法 logistic 回归参数的诊断图, 以及去掉预烧期后由链获得的  $\beta$  的两个分量  $\beta_0$  和  $\beta_1$  的样本均值和样本标准差如下:
- 图 6.5.8 最上面两个小图, 即  $\beta_0$  和  $\beta_1$  的时间状态图显示链的混合较好. 中间两个小图, 即  $\beta_0$  和  $\beta_1$  的遍历均值图, 去掉前 2000 次迭代 (即预烧期) 后趋于平稳, 也表明链的收敛性较好. 最下面两个小图, 即链的自相关函数图中显示马氏链的抽样步长  $L = 20$  次迭代, 也就是每间隔 20 个样本抽取一个, 获得的样本的自相关性就很低了, 这几种诊断方法都显示链的收敛性较好. 最后,  $\beta_0$  和  $\beta_1$  的链的样本均值分别是 2.8793214,  $-0.3737996$ . 样本标准差分别是 1.3479687, 0.1287645.

- 1 6.1 引言
- 2 6.2 蒙特卡洛抽样方法
  - 6.2.1 引言
  - 6.2.2 蒙特卡洛抽样方法
  - 6.2.2 蒙特卡洛重要性抽样方法
- 3 6.3 马尔科夫链蒙特卡洛方法
  - 6.3.1 引言
  - 6.3.2 马尔科夫链的定义及性质
  - 6.3.3 马氏链的极限定理
- 4 6.4 MCMC 方法简介
  - 6.4.1 MCMC 实施中的若干术语
  - 6.4.2 利用 MCMC 的输出结果描述目标后验分布
  - 6.4.3 MCMC 方法收敛性的诊断
- 5 6.5 Metropolis-Hastings 算法
  - 6.5.1 引言
  - 6.5.2 M-H 抽样方法
  - 6.5.3 M-H 抽样方法的几个变种
- 6 6.6 Gibbs 抽样方法

# 引言

- Gibbs 抽样最早是由 Geman 和 Geman (1984) 提出并被用于 Gibbs 格子点分布，因而由此得名的。它是 M-H 抽样的另一种特殊情形。
- Gibbs 抽样方法特别适合于目标分布是多元的场合，其最令人感兴趣的方面是为了产生不可约的、非周期、并以高维空间的目标分布作为其平稳分布的马氏链，只需要从一些一元分布中进行抽样就可以了。即将从多元目标分布中抽样转化为从一元目标分布抽样，这是 Gibbs 抽样的重要性所在。

# 引言

- Gibbs 抽样最早是由 Geman 和 Geman (1984) 提出并被用于 Gibbs 格子点分布，因而由此得名的。它是 M-H 抽样的另一种特殊情形。
- Gibbs 抽样方法特别适合于目标分布是多元的场合，其最令人感兴趣的方面是为了产生不可约的、非周期、并以高维空间的目标分布作为其平稳分布的马氏链，只需要从一些一元分布中进行抽样就可以了。即将从多元目标分布中抽样转化为从一元目标分布抽样，这是 Gibbs 抽样的重要性所在。



# Gibbs 抽样方法

- 令  $\mathbf{X} = (X_1, \dots, X_k)$  为  $\mathbf{R}^k$  中的随机变量, 其联合分布  $f(\mathbf{x}) = f(x_1, \dots, x_k)$  为目标抽样分布. 定义  $k-1$  维的随机变量

$$\mathbf{X}_{-j} = (X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_k),$$

- 记  $X_j | \mathbf{X}_{-j}$  的满条件分布密度为  $f(x_j | \mathbf{x}_{-j})$ ,  $j = 1, \dots, k$ . 则 Gibbs 抽样方法是从这  $k$  个条件分布中产生候选点, 以解决直接从  $f$  中进行抽样的困难. 算法如下:
- (1) 在  $t = 0$  时, 初始化  $\mathbf{X}(0)$ ;

# Gibbs 抽样方法

- 令  $\mathbf{X} = (X_1, \dots, X_k)$  为  $\mathbf{R}^k$  中的随机变量, 其联合分布  $f(\mathbf{x}) = f(x_1, \dots, x_k)$  为目标抽样分布. 定义  $k-1$  维的随机变量

$$\mathbf{X}_{-j} = (X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_k),$$

- 记  $X_j | \mathbf{X}_{-j}$  的满条件分布密度为  $f(x_j | \mathbf{x}_{-j})$ ,  $j = 1, \dots, k$ . 则 Gibbs 抽样方法是从这  $k$  个条件分布中产生候选点, 以解决直接从  $f$  中进行抽样的困难. 算法如下:
- (1) 在  $t = 0$  时, 初始化  $\mathbf{X}(0)$ ;

# Gibbs 抽样方法

- 令  $\mathbf{X} = (X_1, \dots, X_k)$  为  $\mathbf{R}^k$  中的随机变量, 其联合分布  $f(\mathbf{x}) = f(x_1, \dots, x_k)$  为目标抽样分布. 定义  $k-1$  维的随机变量

$$\mathbf{X}_{-j} = (X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_k),$$

- 记  $X_j | \mathbf{X}_{-j}$  的满条件分布密度为  $f(x_j | \mathbf{x}_{-j})$ ,  $j = 1, \dots, k$ . 则 Gibbs 抽样方法是从这  $k$  个条件分布中产生候选点, 以解决直接从  $f$  中进行抽样的困难. 算法如下:
- (1) 在  $t = 0$  时, 初始化  $\mathbf{X}(0)$ ;

# Gibbs 抽样方法

(2) 对  $t = 1, 2, \dots, J$  重复下列步骤

(a) 令  $(x_1, x_2, \dots, x_k) = \mathbf{X}(t-1)$ .

(b) 对每个分量  $j = 1, \dots, k$ ,

(i) 从  $f(X_j | \mathbf{x}_{-j})$  中产生候选点  $X_j^*(t)$ .

(ii) 更新  $x_j = X_j^*(t)$ .

(c) 令  $\mathbf{X}(t) = (X_1^*(t), \dots, X_k^*(t))$  (每个候选点都被接受)

(d) 增加  $t$  返回到 (a).

# Gibbs 抽样方法

- (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
- (a) 令  $(x_1, x_2, \dots, x_k) = \mathbf{X}(t-1)$ .
  - (b) 对每个分量  $j = 1, \dots, k$ ,
    - (i) 从  $f(X_j | \mathbf{x}_{-j})$  中产生候选点  $X_j^*(t)$ .
    - (ii) 更新  $x_j = X_j^*(t)$ .
  - (c) 令  $\mathbf{X}(t) = (X_1^*(t), \dots, X_k^*(t))$  (每个候选点都被接受)
  - (d) 增加  $t$  返回到 (a).

# Gibbs 抽样方法

- (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
- (a) 令  $(x_1, x_2, \dots, x_k) = \mathbf{X}(t-1)$ .
  - (b) 对每个分量  $j = 1, \dots, k$ ,
    - (i) 从  $f(X_j | \mathbf{x}_{-j})$  中产生候选点  $X_j^*(t)$ .
    - (ii) 更新  $x_j = X_j^*(t)$ .
  - (c) 令  $\mathbf{X}(t) = (X_1^*(t), \dots, X_k^*(t))$  (每个候选点都被接受)
  - (d) 增加  $t$  返回到 (a).

# Gibbs 抽样方法

- (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
- (a) 令  $(x_1, x_2, \dots, x_k) = \mathbf{X}(t-1)$ .
  - (b) 对每个分量  $j = 1, \dots, k$ ,
    - (i) 从  $f(X_j | \mathbf{x}_{-j})$  中产生候选点  $X_j^*(t)$ .
    - (ii) 更新  $x_j = X_j^*(t)$ .
  - (c) 令  $\mathbf{X}(t) = (X_1^*(t), \dots, X_k^*(t))$  (每个候选点都被接受)
  - (d) 增加  $t$  返回到 (a).

# Gibbs 抽样方法

- (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
- (a) 令  $(x_1, x_2, \dots, x_k) = \mathbf{X}(t-1)$ .
  - (b) 对每个分量  $j = 1, \dots, k$ ,
    - (i) 从  $f(X_j | \mathbf{x}_{-j})$  中产生候选点  $X_j^*(t)$ .
    - (ii) 更新  $x_j = X_j^*(t)$ .
  - (c) 令  $\mathbf{X}(t) = (X_1^*(t), \dots, X_k^*(t))$  (每个候选点都被接受)
  - (d) 增加  $t$  返回到 (a).



# Gibbs 抽样方法

- (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
- (a) 令  $(x_1, x_2, \dots, x_k) = \mathbf{X}(t-1)$ .
  - (b) 对每个分量  $j = 1, \dots, k$ ,
    - (i) 从  $f(X_j | \mathbf{x}_{-j})$  中产生候选点  $X_j^*(t)$ .
    - (ii) 更新  $x_j = X_j^*(t)$ .
  - (c) 令  $\mathbf{X}(t) = (X_1^*(t), \dots, X_k^*(t))$  (每个候选点都被接受)
  - (d) 增加  $t$  返回到 (a).

# Gibbs 抽样方法

- (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
- (a) 令  $(x_1, x_2, \dots, x_k) = \mathbf{X}(t-1)$ .
  - (b) 对每个分量  $j = 1, \dots, k$ ,
    - (i) 从  $f(X_j | \mathbf{x}_{-j})$  中产生候选点  $X_j^*(t)$ .
    - (ii) 更新  $x_j = X_j^*(t)$ .
  - (c) 令  $\mathbf{X}(t) = (X_1^*(t), \dots, X_k^*(t))$  (每个候选点都被接受)
  - (d) 增加  $t$  返回到 (a).

# Gibbs 抽样方法

- 注意在上述算法 (b) 步抽样中, 各个分量依次被更新:

$$x_1(t) \sim f(x_1 | x_2(t-1), x_3(t-1), \dots, x_k(t-1));$$

$$x_2(t) \sim f(x_2 | x_1(t), x_3(t-1), \dots, x_k(t-1))$$

$$\vdots$$

$$x_k(t) \sim f(x_k | x_1(t), x_2(t), \dots, x_{k-1}(t)).$$

- 从一元分布

$$f(x_j | x_1(t), x_2(t), \dots, x_{j-1}(t), x_{j+1}(t-1), \dots, x_k(t-1))$$

中抽样是比较容易的, 因为  $f(x_j | \mathbf{x}_{-j}) \propto f(\mathbf{x})$ , 其中除了变量  $x_j$  外, 其他变量都是常数.

# Gibbs 抽样方法

- 注意在上述算法 (b) 步抽样中, 各个分量依次被更新:

$$x_1(t) \sim f(x_1 | x_2(t-1), x_3(t-1), \dots, x_k(t-1));$$

$$x_2(t) \sim f(x_2 | x_1(t), x_3(t-1), \dots, x_k(t-1))$$

$$\vdots$$

$$x_k(t) \sim f(x_k | x_1(t), x_2(t), \dots, x_{k-1}(t)).$$

- 从一元分布

$$f(x_j | x_1(t), x_2(t), \dots, x_{j-1}(t), x_{j+1}(t-1), \dots, x_k(t-1))$$

中抽样是比较容易的, 因为  $f(x_j | \mathbf{x}_{-j}) \propto f(\mathbf{x})$ , 其中除了变量  $x_j$  外, 其他变量都是常数.

# Gibbs 抽样方法

- Gibbs 抽样方法的合理性并不需要通过验证其是 M-H 抽样方法的一种特例来证明. Gibbs 抽样方法的一个特别之处就是满条件分布可以唯一确定联合分布. 这就是著名的 **Hammersley-Clifford 定理**, 详见韦和张 (2013). 定理 5.5.2.
- 前面所述的 Gibbs 抽样方法可直接在贝叶斯框架下实现. 令  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ , 而

$$\boldsymbol{\theta}_{-j} = (\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_k).$$

- 将  $\mathbf{x}$  用感兴趣的参数  $\boldsymbol{\theta}$  代替, 将目标分布  $f(\mathbf{x})$  用后验分布  $\pi(\boldsymbol{\theta}|\mathbf{x})$  代替, 将满条件分布  $f(x_j | \mathbf{x}_{-j})$ ,  $j = 1, \dots, k$  用  $\pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{x})$ ,  $j = 1, \dots, k$  代替.

# Gibbs 抽样方法

- Gibbs 抽样方法的合理性并不需要通过验证其是 M-H 抽样方法的一种特例来证明. Gibbs 抽样方法的一个特别之处就是满条件分布可以唯一确定联合分布. 这就是著名的 **Hammersley-Clifford 定理**, 详见韦和张 (2013). 定理 5.5.2.
- 前面所述的 Gibbs 抽样方法可直接在贝叶斯框架下实现. 令  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ , 而

$$\boldsymbol{\theta}_{-j} = (\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_k).$$

- 将  $\mathbf{x}$  用感兴趣的参数  $\boldsymbol{\theta}$  代替, 将目标分布  $f(\mathbf{x})$  用后验分布  $\pi(\boldsymbol{\theta}|\mathbf{x})$  代替, 将满条件分布  $f(x_j | \mathbf{x}_{-j})$ ,  $j = 1, \dots, k$  用  $\pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{x})$ ,  $j = 1, \dots, k$  代替.

# Gibbs 抽样方法

- Gibbs 抽样方法的合理性并不需要通过验证其是 M-H 抽样方法的一种特例来证明. Gibbs 抽样方法的一个特别之处就是满条件分布可以唯一确定联合分布. 这就是著名的 **Hammersley-Clifford 定理**, 详见韦和张 (2013). 定理 5.5.2.
- 前面所述的 Gibbs 抽样方法可直接在贝叶斯框架下实现. 令  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ , 而

$$\boldsymbol{\theta}_{-j} = (\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_k).$$

- 将  $\mathbf{x}$  用感兴趣的参数  $\boldsymbol{\theta}$  代替, 将目标分布  $f(\mathbf{x})$  用后验分布  $\pi(\boldsymbol{\theta}|\mathbf{x})$  代替, 将满条件分布  $f(x_j | \mathbf{x}_{-j})$ ,  $j = 1, \dots, k$  用  $\pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{x})$ ,  $j = 1, \dots, k$  代替.

# Gibbs 抽样方法

- 因此，在贝叶斯分析中，Gibbs 抽样方法产生马氏链  $\{\boldsymbol{\theta}^{(i)}, i = 1, 2, \dots\}$  的方法概括如下：
  - (1) 选择参数的一个初始值  $\boldsymbol{\theta}^{(0)}$ ;
  - (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
    - (a) 令  $(\theta_1, \theta_2, \dots, \theta_k) = \boldsymbol{\theta}^{(t-1)}$
    - (b) 对每个分量  $j = 1, \dots, k$ 
      - (i) 从  $\pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{x})$  中产生候选点  $\theta_j^{(t)}$ .
      - (ii) 更新  $\theta_j = \theta_j^{(t)}$ .
    - (c) 令  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \dots, \theta_k^{(t)})$ .
    - (d) 增加  $t$  返回到 (a).



# Gibbs 抽样方法

- 因此，在贝叶斯分析中，Gibbs 抽样方法产生马氏链  $\{\boldsymbol{\theta}^{(i)}, i = 1, 2, \dots\}$  的方法概括如下：
- (1) 选择参数的一个初始值  $\boldsymbol{\theta}^{(0)}$ ;
- (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
  - (a) 令  $(\theta_1, \theta_2, \dots, \theta_k) = \boldsymbol{\theta}^{(t-1)}$
  - (b) 对每个分量  $j = 1, \dots, k$ 
    - (i) 从  $\pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{x})$  中产生候选点  $\theta_j^{(t)}$ .
    - (ii) 更新  $\theta_j = \theta_j^{(t)}$ .
  - (c) 令  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \dots, \theta_k^{(t)})$ .
  - (d) 增加  $t$  返回到 (a).

# Gibbs 抽样方法

- 因此，在贝叶斯分析中，Gibbs 抽样方法产生马氏链  $\{\boldsymbol{\theta}^{(i)}, i = 1, 2, \dots\}$  的方法概括如下：
- (1) 选择参数的一个初始值  $\boldsymbol{\theta}^{(0)}$ ;
- (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
  - (a) 令  $(\theta_1, \theta_2, \dots, \theta_k) = \boldsymbol{\theta}^{(t-1)}$
  - (b) 对每个分量  $j = 1, \dots, k$ 
    - (i) 从  $\pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{x})$  中产生候选点  $\theta_j^{(t)}$ .
    - (ii) 更新  $\theta_j = \theta_j^{(t)}$ .
  - (c) 令  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \dots, \theta_k^{(t)})$ .
  - (d) 增加  $t$  返回到 (a).

# Gibbs 抽样方法

- 因此，在贝叶斯分析中，Gibbs 抽样方法产生马氏链  $\{\boldsymbol{\theta}^{(i)}, i = 1, 2, \dots\}$  的方法概括如下：
- (1) 选择参数的一个初始值  $\boldsymbol{\theta}^{(0)}$ ;
- (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
  - (a) 令  $(\theta_1, \theta_2, \dots, \theta_k) = \boldsymbol{\theta}^{(t-1)}$
  - (b) 对每个分量  $j = 1, \dots, k$ 
    - (i) 从  $\pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{x})$  中产生候选点  $\theta_j^{(t)}$ .
    - (ii) 更新  $\theta_j = \theta_j^{(t)}$ .
  - (c) 令  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \dots, \theta_k^{(t)})$ .
  - (d) 增加  $t$  返回到 (a).

# Gibbs 抽样方法

- 因此，在贝叶斯分析中，Gibbs 抽样方法产生马氏链  $\{\boldsymbol{\theta}^{(i)}, i = 1, 2, \dots\}$  的方法概括如下：
- (1) 选择参数的一个初始值  $\boldsymbol{\theta}^{(0)}$ ;
- (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
  - (a) 令  $(\theta_1, \theta_2, \dots, \theta_k) = \boldsymbol{\theta}^{(t-1)}$
  - (b) 对每个分量  $j = 1, \dots, k$ 
    - (i) 从  $\pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{x})$  中产生候选点  $\theta_j^{(t)}$ .
    - (ii) 更新  $\theta_j = \theta_j^{(t)}$ .
  - (c) 令  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \dots, \theta_k^{(t)})$ .
  - (d) 增加  $t$  返回到 (a).

# Gibbs 抽样方法

- 因此，在贝叶斯分析中，Gibbs 抽样方法产生马氏链  $\{\boldsymbol{\theta}^{(i)}, i = 1, 2, \dots\}$  的方法概括如下：
- (1) 选择参数的一个初始值  $\boldsymbol{\theta}^{(0)}$ ;
- (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
  - (a) 令  $(\theta_1, \theta_2, \dots, \theta_k) = \boldsymbol{\theta}^{(t-1)}$
  - (b) 对每个分量  $j = 1, \dots, k$ 
    - (i) 从  $\pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{x})$  中产生候选点  $\theta_j^{(t)}$ .
    - (ii) 更新  $\theta_j = \theta_j^{(t)}$ .
  - (c) 令  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \dots, \theta_k^{(t)})$ .
  - (d) 增加  $t$  返回到 (a).

# Gibbs 抽样方法

- 因此，在贝叶斯分析中，Gibbs 抽样方法产生马氏链  $\{\boldsymbol{\theta}^{(i)}, i = 1, 2, \dots\}$  的方法概括如下：
- (1) 选择参数的一个初始值  $\boldsymbol{\theta}^{(0)}$ ;
- (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
  - (a) 令  $(\theta_1, \theta_2, \dots, \theta_k) = \boldsymbol{\theta}^{(t-1)}$
  - (b) 对每个分量  $j = 1, \dots, k$ 
    - (i) 从  $\pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{x})$  中产生候选点  $\theta_j^{(t)}$ .
    - (ii) 更新  $\theta_j = \theta_j^{(t)}$ .
  - (c) 令  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \dots, \theta_k^{(t)})$ .
  - (d) 增加  $t$  返回到 (a).

# Gibbs 抽样方法

- 因此，在贝叶斯分析中，Gibbs 抽样方法产生马氏链  $\{\boldsymbol{\theta}^{(i)}, i = 1, 2, \dots\}$  的方法概括如下：
- (1) 选择参数的一个初始值  $\boldsymbol{\theta}^{(0)}$ ;
- (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
  - (a) 令  $(\theta_1, \theta_2, \dots, \theta_k) = \boldsymbol{\theta}^{(t-1)}$
  - (b) 对每个分量  $j = 1, \dots, k$ 
    - (i) 从  $\pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{x})$  中产生候选点  $\theta_j^{(t)}$ .
    - (ii) 更新  $\theta_j = \theta_j^{(t)}$ .
  - (c) 令  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \dots, \theta_k^{(t)})$ .
  - (d) 增加  $t$  返回到 (a).

# Gibbs 抽样方法

- 因此，在贝叶斯分析中，Gibbs 抽样方法产生马氏链  $\{\boldsymbol{\theta}^{(i)}, i = 1, 2, \dots\}$  的方法概括如下：
- (1) 选择参数的一个初始值  $\boldsymbol{\theta}^{(0)}$ ;
- (2) 对  $t = 1, 2, \dots, J$  重复下列步骤
  - (a) 令  $(\theta_1, \theta_2, \dots, \theta_k) = \boldsymbol{\theta}^{(t-1)}$
  - (b) 对每个分量  $j = 1, \dots, k$ 
    - (i) 从  $\pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{x})$  中产生候选点  $\theta_j^{(t)}$ .
    - (ii) 更新  $\theta_j = \theta_j^{(t)}$ .
  - (c) 令  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \dots, \theta_k^{(t)})$ .
  - (d) 增加  $t$  返回到 (a).



# Gibbs 抽样方法

- 注意在上述算法 (b) 步抽样中, 各个分量依次被更新:
  - 1) 抽取  $\theta_1^{(t)} \sim \pi(\theta_1 | \theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, \mathbf{x})$ ;
  - 2) 抽取  $\theta_2^{(t)} \sim \pi(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, \mathbf{x})$ ;
  - $\vdots$
  - k) 抽取  $\theta_k^{(t)} \sim \pi(\theta_k | \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{k-1}^{(t)}, \mathbf{x})$ .
- 下例说明如何利用 Gibbs 抽样生成二元正态分布的样本.

# Gibbs 抽样方法

- 注意在上述算法 (b) 步抽样中, 各个分量依次被更新:

1) 抽取  $\theta_1^{(t)} \sim \pi(\theta_1 \mid \theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, \mathbf{x});$

2) 抽取  $\theta_2^{(t)} \sim \pi(\theta_2 \mid \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, \mathbf{x});$

$\vdots$

k) 抽取  $\theta_k^{(t)} \sim \pi(\theta_k \mid \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{k-1}^{(t)}, \mathbf{x}).$

- 下例说明如何利用 Gibbs 抽样生成二元正态分布的样本.

# Gibbs 抽样方法

- 注意在上述算法 (b) 步抽样中, 各个分量依次被更新:
  - 1) 抽取  $\theta_1^{(t)} \sim \pi(\theta_1 | \theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, \mathbf{x})$ ;
  - 2) 抽取  $\theta_2^{(t)} \sim \pi(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, \mathbf{x})$ ;
  - $\vdots$
  - k) 抽取  $\theta_k^{(t)} \sim \pi(\theta_k | \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{k-1}^{(t)}, \mathbf{x})$ .
- 下例说明如何利用 Gibbs 抽样生成二元正态分布的样本.

# Gibbs 抽样方法

- 注意在上述算法 (b) 步抽样中, 各个分量依次被更新:
  - 1) 抽取  $\theta_1^{(t)} \sim \pi(\theta_1 | \theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, \mathbf{x})$ ;
  - 2) 抽取  $\theta_2^{(t)} \sim \pi(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, \mathbf{x})$ ;
  - $\vdots$
  - k) 抽取  $\theta_k^{(t)} \sim \pi(\theta_k | \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{k-1}^{(t)}, \mathbf{x})$ .
- 下例说明如何利用 Gibbs 抽样生成二元正态分布的样本.

## Gibbs 抽样生成二元正态分布随机数的例子

- 例 6.6.1 使用 Gibbs 抽样产生二元正态分布

$$N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$$

的随机数.

- 解 在二元正态场合,  $X_1|X_2$  以及  $X_2|X_1$  仍然服从正态分布, 且易知

$$E[X_1|X_2 = x_2] = \mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2),$$

$$\text{Var}[X_1|X_2 = x_2] = (1 - \rho^2)\sigma_1^2.$$

- 类似可得  $X_2|X_1$  的分布. 因此

$$f(x_1|x_2) \sim N\left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), (1 - \rho^2)\sigma_1^2\right),$$

$$f(x_2|x_1) \sim N\left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), (1 - \rho^2)\sigma_2^2\right).$$

## Gibbs 抽样生成二元正态分布随机数的例子

- 例 6.6.1 使用 Gibbs 抽样产生二元正态分布

$$N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$$

的随机数.

- 解 在二元正态场合,  $X_1|X_2$  以及  $X_2|X_1$  仍然服从正态分布, 且易知

$$E[X_1|X_2 = x_2] = \mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2),$$

$$\text{Var}[X_1|X_2 = x_2] = (1 - \rho^2)\sigma_1^2.$$

- 类似可得  $X_2|X_1$  的分布. 因此

$$f(x_1|x_2) \sim N\left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), (1 - \rho^2)\sigma_1^2\right),$$

$$f(x_2|x_1) \sim N\left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), (1 - \rho^2)\sigma_2^2\right).$$

## Gibbs 抽样生成二元正态分布随机数的例子

- **例 6.6.1** 使用 Gibbs 抽样产生二元正态分布

$$N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$$

的随机数.

- **解** 在二元正态场合,  $X_1|X_2$  以及  $X_2|X_1$  仍然服从正态分布, 且易知

$$E[X_1|X_2 = x_2] = \mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2),$$

$$\text{Var}[X_1|X_2 = x_2] = (1 - \rho^2)\sigma_1^2.$$

- 类似可得  $X_2|X_1$  的分布. 因此

$$f(x_1|x_2) \sim N\left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), (1 - \rho^2)\sigma_1^2\right),$$

$$f(x_2|x_1) \sim N\left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), (1 - \rho^2)\sigma_2^2\right).$$

## 续例 6.6.1

从而使用 Gibbs 算法如下

- (1) 令  $(x_1, x_2) = \mathbf{x}(t-1)$ ;
- (2) 从  $f(x_1|x_2)$  中产生候选点  $X_1^*(t)$ .
- (3) 更新  $x_1 = X_1^*(t)$ .
- (4) 从  $f(x_2|x_1)$  中产生  $X_2^*(t)$ .
- (5) 更新  $x_2 = X_2^*(t)$ .
- (6) 令  $X(t) = (X_1^*(t), X_2^*(t))$ .

生成链的散点图如下:



## 续例 6.6.1

从而使用 Gibbs 算法如下

- (1) 令  $(x_1, x_2) = \mathbf{x}(t-1)$ ;
- (2) 从  $f(x_1|x_2)$  中产生候选点  $X_1^*(t)$ .
- (3) 更新  $x_1 = X_1^*(t)$ .
- (4) 从  $f(x_2|x_1)$  中产生  $X_2^*(t)$ .
- (5) 更新  $x_2 = X_2^*(t)$ .
- (6) 令  $X(t) = (X_1^*(t), X_2^*(t))$ .

生成链的散点图如下:

## 续例 6.6.1

从而使用 Gibbs 算法如下

- (1) 令  $(x_1, x_2) = \mathbf{x}(t-1)$ ;
- (2) 从  $f(x_1|x_2)$  中产生候选点  $X_1^*(t)$ .
- (3) 更新  $x_1 = X_1^*(t)$ .
- (4) 从  $f(x_2|x_1)$  中产生  $X_2^*(t)$ .
- (5) 更新  $x_2 = X_2^*(t)$ .
- (6) 令  $X(t) = (X_1^*(t), X_2^*(t))$ .

生成链的散点图如下:

## 续例 6.6.1

从而使用 Gibbs 算法如下

- (1) 令  $(x_1, x_2) = \mathbf{x}(t-1)$ ;
- (2) 从  $f(x_1|x_2)$  中产生候选点  $X_1^*(t)$ .
- (3) 更新  $x_1 = X_1^*(t)$ .
- (4) 从  $f(x_2|x_1)$  中产生  $X_2^*(t)$ .
- (5) 更新  $x_2 = X_2^*(t)$ .
- (6) 令  $X(t) = (X_1^*(t), X_2^*(t))$ .

生成链的散点图如下:

## 续例 6.6.1

从而使用 Gibbs 算法如下

- (1) 令  $(x_1, x_2) = \mathbf{x}(t-1)$ ;
- (2) 从  $f(x_1|x_2)$  中产生候选点  $X_1^*(t)$ .
- (3) 更新  $x_1 = X_1^*(t)$ .
- (4) 从  $f(x_2|x_1)$  中产生  $X_2^*(t)$ .
- (5) 更新  $x_2 = X_2^*(t)$ .
- (6) 令  $X(t) = (X_1^*(t), X_2^*(t))$ .

生成链的散点图如下:

## 续例 6.6.1

从而使用 Gibbs 算法如下

- (1) 令  $(x_1, x_2) = \mathbf{x}(t-1)$ ;
- (2) 从  $f(x_1|x_2)$  中产生候选点  $X_1^*(t)$ .
- (3) 更新  $x_1 = X_1^*(t)$ .
- (4) 从  $f(x_2|x_1)$  中产生  $X_2^*(t)$ .
- (5) 更新  $x_2 = X_2^*(t)$ .
- (6) 令  $X(t) = (X_1^*(t), X_2^*(t))$ .

生成链的散点图如下:

## 续例 6.6.1

从而使用 Gibbs 算法如下

- (1) 令  $(x_1, x_2) = \mathbf{x}(t-1)$ ;
- (2) 从  $f(x_1|x_2)$  中产生候选点  $X_1^*(t)$ .
- (3) 更新  $x_1 = X_1^*(t)$ .
- (4) 从  $f(x_2|x_1)$  中产生  $X_2^*(t)$ .
- (5) 更新  $x_2 = X_2^*(t)$ .
- (6) 令  $X(t) = (X_1^*(t), X_2^*(t))$ .

生成链的散点图如下：

图6.6.1

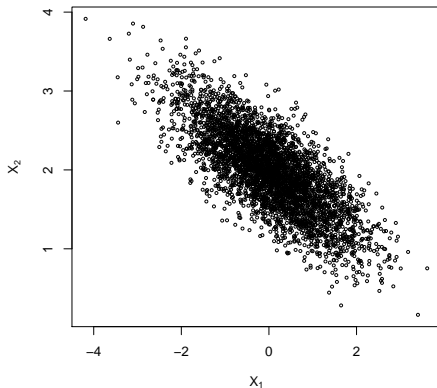


图 6.6.1 Gibbs 抽样生成的二元正态链的散点图

## 续例 6.6.1

- 产生的链开始的 1000 个观测值被丢弃掉，剩下的观测值存在  $\mathbf{x}$  中，由此观察值获得样本均值、样本协方差矩阵和样本相关系数矩阵如下所示。各参数的样本估计值离真值很近，散点图 6.5.7 也显示出二元正态所具有的椭圆对称性和负相关性特征。
- 模拟样本均值为  $-0.0735532, 2.0297997$
- 样本协方差矩阵
$$\begin{pmatrix} 0.9609837 & -0.3506705 \\ -0.3506705 & 0.2327129 \end{pmatrix}$$
- 样本相关系数矩阵
$$\begin{pmatrix} 1.0000000 & -0.7415338 \\ -0.7415338 & 1.0000000 \end{pmatrix}$$



## 续例 6.6.1

- 产生的链开始的 1000 个观测值被丢弃掉，剩下的观测值存在  $\mathbf{x}$  中，由此观察值获得样本均值、样本协方差矩阵和样本相关系数矩阵如下所示。各参数的样本估计值离真值很近，散点图 6.5.7 也显示出二元正态所具有的椭圆对称性和负相关性特征。
- 模拟样本均值为  $-0.0735532, 2.0297997$
- 样本协方差矩阵
$$\begin{pmatrix} 0.9609837 & -0.3506705 \\ -0.3506705 & 0.2327129 \end{pmatrix}$$
- 样本相关系数矩阵
$$\begin{pmatrix} 1.0000000 & -0.7415338 \\ -0.7415338 & 1.0000000 \end{pmatrix}$$

## 续例 6.6.1

- 产生的链开始的 1000 个观测值被丢弃掉，剩下的观测值存在  $\mathbf{x}$  中，由此观察值获得样本均值、样本协方差矩阵和样本相关系数矩阵如下所示。各参数的样本估计值离真值很近，散点图 6.5.7 也显示出二元正态所具有的椭圆对称性和负相关性特征。
- 模拟样本均值为  $-0.0735532, 2.0297997$
- 样本协方差矩阵
$$\begin{array}{cc} 0.9609837 & -0.3506705 \\ -0.3506705 & 0.2327129 \end{array}$$
- 样本相关系数矩阵
$$\begin{array}{cc} 1.0000000 & -0.7415338 \\ -0.7415338 & 1.0000000 \end{array}$$

## 续例 6.6.1

- 产生的链开始的 1000 个观测值被丢弃掉，剩下的观测值存在  $\mathbf{x}$  中，由此观察值获得样本均值、样本协方差矩阵和样本相关系数矩阵如下所示。各参数的样本估计值离真值很近，散点图 6.5.7 也显示出二元正态所具有的椭圆对称性和负相关性特征。
- 模拟样本均值为  $-0.0735532, 2.0297997$
- 样本协方差矩阵
$$\begin{array}{cc} 0.9609837 & -0.3506705 \\ -0.3506705 & 0.2327129 \end{array}$$
- 样本相关系数矩阵
$$\begin{array}{cc} 1.0000000 & -0.7415338 \\ -0.7415338 & 1.0000000 \end{array}$$

## 例 6.6.2

- **例 6.6.2 (续例 6.1.1)** 设  $X|\theta \sim N(\theta, \sigma^2)$ , 其中  $\sigma^2$  已知. 设  $\theta$  的先验为如下的分层先验:  $\theta|\lambda \sim N(\mu, \tau^2/\lambda)$ ,  $\lambda \sim \Gamma(\frac{1}{2}, \frac{1}{2})$ , 其中  $\mu$  和  $\tau$  已知. 用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 并求  $\theta$  的后验期望  $\mu^\pi(x)$  和后验方差  $V^\pi(x)$  的模拟结果. 并与例 6.1.1 进行比较.
- **解** (1) 首先, 此处的分层先验与例 6.1.1 中  $\theta$  的先验分布, 即柯西分布  $C(\mu, \lambda)$  本质上是相同的. 说明如下: 易见  $\theta$  的分层先验密度分别为

$$\pi_1(\theta|\lambda) = \left(\frac{\lambda}{2\pi\tau^2}\right)^{1/2} \exp\left\{-\frac{\lambda(\theta-\mu)^2}{2\tau^2}\right\}, \quad \theta \in \mathbf{R}, \quad (6.6.1)$$

$$\pi_2(\lambda) = (2\pi)^{-1/2} \lambda^{\frac{1}{2}-1} \exp\{-\lambda/2\}, \quad 0 < \lambda < \infty. \quad (6.6.2)$$

## 例 6.6.2

- **例 6.6.2 (续例 6.1.1)** 设  $X|\theta \sim N(\theta, \sigma^2)$ , 其中  $\sigma^2$  已知. 设  $\theta$  的先验为如下的分层先验:  $\theta|\lambda \sim N(\mu, \tau^2/\lambda)$ ,  $\lambda \sim \Gamma(\frac{1}{2}, \frac{1}{2})$ , 其中  $\mu$  和  $\tau$  已知. 用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 并求  $\theta$  的后验期望  $\mu^\pi(x)$  和后验方差  $V^\pi(x)$  的模拟结果. 并与例 6.1.1 进行比较.
- **解** (1) 首先, 此处的分层先验与例 6.1.1 中  $\theta$  的先验分布, 即柯西分布  $C(\mu, \lambda)$  本质上是相同的. 说明如下: 易见  $\theta$  的分层先验密度分别为

$$\pi_1(\theta|\lambda) = \left( \frac{\lambda}{2\pi\tau^2} \right)^{1/2} \exp \left\{ -\frac{\lambda(\theta - \mu)^2}{2\tau^2} \right\}, \quad \theta \in \mathbf{R}, \quad (6.6.1)$$

$$\pi_2(\lambda) = (2\pi)^{-1/2} \lambda^{\frac{1}{2}-1} \exp\{-\lambda/2\}, \quad 0 < \lambda < \infty. \quad (6.6.2)$$

## 续例 6.6.2

- 因此,  $\theta$  和  $\lambda$  的联合先验为  $\pi(\theta, \lambda) = \pi_1(\theta|\lambda)\pi_2(\lambda)$ , 故由 2.6 节可知  $\theta$  的规范先验为

$$\begin{aligned}\pi(\theta) &= \int_0^\infty \pi(\theta, \lambda) d\lambda = \int_0^\infty c \cdot \exp \left\{ -\lambda \left[ \frac{\tau^2 + (\theta - \mu)^2}{2\tau^2} \right] \right\} d\lambda \\ &= \int_0^\infty c \cdot e^{-\lambda B} d\lambda = \frac{c}{B} = \frac{\tau}{\pi[\tau^2 + (\theta - \mu)^2]}, \quad \theta \in \mathbf{R}. \quad (6.6.3)\end{aligned}$$

- 此处  $B = [\tau^2 + (\theta - \mu)^2]/(2\tau^2)$ ,  $c = 1/(2\pi\tau)$ . 式 (6.6.3) 就是柯西分布  $C(\mu, \tau)$  的密度. 因此分层联合先验分布与出现在例 6.1.1 中  $\theta$  的先验分布-柯西分布本质是一样的.

## 续例 6.6.2

- (2) 其次, 为了使用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 需要导出如下两个满条件分布:

$$\theta | \lambda, x \sim N\left(\frac{\tau^2}{\tau^2 + \lambda\sigma^2}x + \frac{\lambda\sigma^2}{\tau^2 + \lambda\sigma^2}\mu, \frac{\tau^2\sigma^2}{\tau^2 + \lambda\sigma^2}\right), \quad (6.6.4)$$

$$\lambda | \theta, x \sim \text{Exp}\left(\frac{\tau^2 + (\theta - \mu)^2}{2\tau^2}\right). \quad (6.6.5)$$

- (3) 用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 算法如下:
  - (a) 令  $(x_1, x_2) = \mathbf{X}(t-1)$ .
  - (b) 从  $\pi(\theta | \lambda, x)$  中产生候选点  $X_1^*(t)$ .
  - (c) 更新  $x_1 = X_1^*(t)$ .
  - (d) 从  $\pi_2(\lambda | \theta, x_1)$  中产生  $X_2^*(t)$ .
  - (e) 更新  $x_2 = X_2^*(t)$ .
  - (f) 令  $\mathbf{X}(t) = (X_1^*(t), X_2^*(t))$ .

## 续例 6.6.2

- (2) 其次, 为了使用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 需要导出如下两个满条件分布:

$$\theta | \lambda, x \sim N\left(\frac{\tau^2}{\tau^2 + \lambda\sigma^2}x + \frac{\lambda\sigma^2}{\tau^2 + \lambda\sigma^2}\mu, \frac{\tau^2\sigma^2}{\tau^2 + \lambda\sigma^2}\right), \quad (6.6.4)$$

$$\lambda | \theta, x \sim \text{Exp}\left(\frac{\tau^2 + (\theta - \mu)^2}{2\tau^2}\right). \quad (6.6.5)$$

- (3) 用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 算法如下:
  - (a) 令  $(x_1, x_2) = \mathbf{X}(t-1)$ .
  - (b) 从  $\pi(\theta | \lambda, x)$  中产生候选点  $X_1^*(t)$ .
  - (c) 更新  $x_1 = X_1^*(t)$ .
  - (d) 从  $\pi_2(\lambda | \theta, x_1)$  中产生  $X_2^*(t)$ .
  - (e) 更新  $x_2 = X_2^*(t)$ .
  - (f) 令  $\mathbf{X}(t) = (X_1^*(t), X_2^*(t))$ .



## 续例 6.6.2

- (2) 其次, 为了使用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 需要导出如下两个满条件分布:

$$\theta | \lambda, x \sim N\left(\frac{\tau^2}{\tau^2 + \lambda\sigma^2}x + \frac{\lambda\sigma^2}{\tau^2 + \lambda\sigma^2}\mu, \frac{\tau^2\sigma^2}{\tau^2 + \lambda\sigma^2}\right), \quad (6.6.4)$$

$$\lambda | \theta, x \sim \text{Exp}\left(\frac{\tau^2 + (\theta - \mu)^2}{2\tau^2}\right). \quad (6.6.5)$$

- (3) 用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 算法如下:
  - (a) 令  $(x_1, x_2) = \mathbf{X}(t-1)$ .
  - (b) 从  $\pi(\theta | \lambda, x)$  中产生候选点  $X_1^*(t)$ .
  - (c) 更新  $x_1 = X_1^*(t)$ .
  - (d) 从  $\pi_2(\lambda | \theta, x_1)$  中产生  $X_2^*(t)$ .
  - (e) 更新  $x_2 = X_2^*(t)$ .
  - (f) 令  $\mathbf{X}(t) = (X_1^*(t), X_2^*(t))$ .

## 续例 6.6.2

- (2) 其次, 为了使用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 需要导出如下两个满条件分布:

$$\theta | \lambda, \mathbf{x} \sim N\left(\frac{\tau^2}{\tau^2 + \lambda\sigma^2}\bar{x} + \frac{\lambda\sigma^2}{\tau^2 + \lambda\sigma^2}\mu, \frac{\tau^2\sigma^2}{\tau^2 + \lambda\sigma^2}\right), \quad (6.6.4)$$

$$\lambda | \theta, \mathbf{x} \sim \text{Exp}\left(\frac{\tau^2 + (\theta - \mu)^2}{2\tau^2}\right). \quad (6.6.5)$$

- (3) 用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 算法如下:
  - (a) 令  $(x_1, x_2) = \mathbf{X}(t-1)$ .
  - (b) 从  $\pi(\theta | \lambda, \mathbf{x})$  中产生候选点  $X_1^*(t)$ .
  - (c) 更新  $x_1 = X_1^*(t)$ .
  - (d) 从  $\pi_2(\lambda | \theta, \mathbf{x}_1)$  中产生  $X_2^*(t)$ .
  - (e) 更新  $x_2 = X_2^*(t)$ .
  - (f) 令  $\mathbf{X}(t) = (X_1^*(t), X_2^*(t))$ .

## 续例 6.6.2

- (2) 其次, 为了使用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 需要导出如下两个满条件分布:

$$\theta | \lambda, x \sim N\left(\frac{\tau^2}{\tau^2 + \lambda\sigma^2}x + \frac{\lambda\sigma^2}{\tau^2 + \lambda\sigma^2}\mu, \frac{\tau^2\sigma^2}{\tau^2 + \lambda\sigma^2}\right), \quad (6.6.4)$$

$$\lambda | \theta, x \sim \text{Exp}\left(\frac{\tau^2 + (\theta - \mu)^2}{2\tau^2}\right). \quad (6.6.5)$$

- (3) 用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 算法如下:
  - (a) 令  $(x_1, x_2) = \mathbf{X}(t-1)$ .
  - (b) 从  $\pi(\theta | \lambda, x)$  中产生候选点  $X_1^*(t)$ .
  - (c) 更新  $x_1 = X_1^*(t)$ .
  - (d) 从  $\pi_2(\lambda | \theta, x_1)$  中产生  $X_2^*(t)$ .
  - (e) 更新  $x_2 = X_2^*(t)$ .
  - (f) 令  $\mathbf{X}(t) = (X_1^*(t), X_2^*(t))$ .

## 续例 6.6.2

- (2) 其次, 为了使用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 需要导出如下两个满条件分布:

$$\theta | \lambda, x \sim N\left(\frac{\tau^2}{\tau^2 + \lambda\sigma^2}x + \frac{\lambda\sigma^2}{\tau^2 + \lambda\sigma^2}\mu, \frac{\tau^2\sigma^2}{\tau^2 + \lambda\sigma^2}\right), \quad (6.6.4)$$

$$\lambda | \theta, x \sim \text{Exp}\left(\frac{\tau^2 + (\theta - \mu)^2}{2\tau^2}\right). \quad (6.6.5)$$

- (3) 用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 算法如下:
  - (a) 令  $(x_1, x_2) = \mathbf{X}(t-1)$ .
  - (b) 从  $\pi(\theta | \lambda, x)$  中产生候选点  $X_1^*(t)$ .
  - (c) 更新  $x_1 = X_1^*(t)$ .
  - (d) 从  $\pi_2(\lambda | \theta, x_1)$  中产生  $X_2^*(t)$ .
  - (e) 更新  $x_2 = X_2^*(t)$ .
  - (f) 令  $\mathbf{X}(t) = (X_1^*(t), X_2^*(t))$ .

## 续例 6.6.2

- (2) 其次, 为了使用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 需要导出如下两个满条件分布:

$$\theta | \lambda, x \sim N\left(\frac{\tau^2}{\tau^2 + \lambda\sigma^2}x + \frac{\lambda\sigma^2}{\tau^2 + \lambda\sigma^2}\mu, \frac{\tau^2\sigma^2}{\tau^2 + \lambda\sigma^2}\right), \quad (6.6.4)$$

$$\lambda | \theta, x \sim \text{Exp}\left(\frac{\tau^2 + (\theta - \mu)^2}{2\tau^2}\right). \quad (6.6.5)$$

- (3) 用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 算法如下:
  - (a) 令  $(x_1, x_2) = \mathbf{X}(t-1)$ .
  - (b) 从  $\pi(\theta | \lambda, x)$  中产生候选点  $X_1^*(t)$ .
  - (c) 更新  $x_1 = X_1^*(t)$ .
  - (d) 从  $\pi_2(\lambda | \theta, x_1)$  中产生  $X_2^*(t)$ .
  - (e) 更新  $x_2 = X_2^*(t)$ .
  - (f) 令  $\mathbf{X}(t) = (X_1^*(t), X_2^*(t))$ .

## 续例 6.6.2

- (2) 其次, 为了使用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 需要导出如下两个满条件分布:

$$\theta | \lambda, x \sim N\left(\frac{\tau^2}{\tau^2 + \lambda\sigma^2}x + \frac{\lambda\sigma^2}{\tau^2 + \lambda\sigma^2}\mu, \frac{\tau^2\sigma^2}{\tau^2 + \lambda\sigma^2}\right), \quad (6.6.4)$$

$$\lambda | \theta, x \sim \text{Exp}\left(\frac{\tau^2 + (\theta - \mu)^2}{2\tau^2}\right). \quad (6.6.5)$$

- (3) 用 Gibbs 抽样方法生成  $(\theta, \lambda)$  的随机数, 算法如下:
  - (a) 令  $(x_1, x_2) = \mathbf{X}(t-1)$ .
  - (b) 从  $\pi(\theta | \lambda, x)$  中产生候选点  $X_1^*(t)$ .
  - (c) 更新  $x_1 = X_1^*(t)$ .
  - (d) 从  $\pi_2(\lambda | \theta, x_1)$  中产生  $X_2^*(t)$ .
  - (e) 更新  $x_2 = X_2^*(t)$ .
  - (f) 令  $\mathbf{X}(t) = (X_1^*(t), X_2^*(t))$ .

## 续例 6.6.2

- 样本路径图 6.6.2 也显示 Gibbs 抽样方法的效果很好.
- 显然由 Gibbs 抽样方法获得  $\theta$  的后验均值  $m = 0.6993006$  和后验标准差  $s = 0.9148281$  与例 6.5.3 中的结果相近. 本例中 Gibbs 抽样方法相比于例 6.5.3 中的 Metropolis 随机游动方法效率要高, 因为此处不需要考虑生成马氏链的效率问题.

## 续例 6.6.2

- 样本路径图 6.6.2 也显示 Gibbs 抽样方法的效果很好.
- 显然由 Gibbs 抽样方法获得  $\theta$  的后验均值  $m = 0.6993006$  和后验标准差  $s = 0.9148281$  与例 6.5.3 中的结果相近. 本例中 Gibbs 抽样方法相比于例 6.5.3 中的 Metropolis 随机游动方法效率要高, 因为此处不需要考虑生成马氏链的效率问题.



## 6.6 Gibbs 抽样方法

## 6.6.2 Gibbs 抽样方法

图 6.6.2

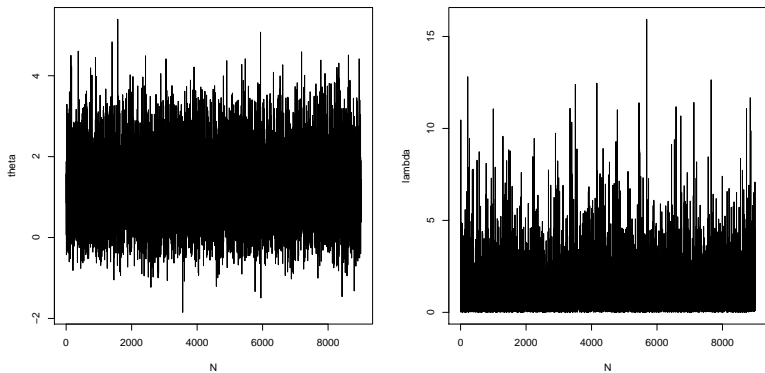


图 6.6.2 样本路径图

- 1 6.1 引言
- 2 6.2 蒙特卡洛抽样方法
  - 6.2.1 引言
  - 6.2.2 蒙特卡洛抽样方法
  - 6.2.2 蒙特卡洛重要性抽样方法
- 3 6.3 马尔科夫链蒙特卡洛方法
  - 6.3.1 引言
  - 6.3.2 马尔科夫链的定义及性质
  - 6.3.3 马氏链的极限定理
- 4 6.4 MCMC 方法简介
  - 6.4.1 MCMC 实施中的若干术语
  - 6.4.2 利用 MCMC 的输出结果描述目标后验分布
  - 6.4.3 MCMC 方法收敛性的诊断
- 5 6.5 Metropolis-Hastings 算法
  - 6.5.1 引言
  - 6.5.2 M-H 抽样方法
  - 6.5.3 M-H 抽样方法的几个变种
- 6 6.6 Gibbs 抽样方法

## WinBUGS 软件简介

- R ( S 和 Splus 的开源版本) 是一套完整的数据处理、计算和制图软件系统. 功能包括: 数据存储和处理系统; 数组运算工具 (其向量、矩阵运算功能尤其强大); 完整连贯的统计分析工具; 优秀的统计制图功能; 简便而强大的编程语言.
- 对贝叶斯计算来说, 可以直接使用 Gibbs 抽样和 Metropolis 算法进行编程, 前面都是如此做的. 对于高强度的计算问题, 可以使用 Fortran 或 C 语言然后从 R 中通过接口调用.
- BUGS (Bayesian inference Using Gibbs Sampling) 是一个免费软件, 其允许用户指定复杂的多层模型并使用 MCMC 算法来估计模型中的未知参数. WinBUGS 是 BUGS 在 Windows 平台下的版本, 而 openBUGS 是 BUGS 的开源版本. 还有一些其它的版本, 可以访问 <http://www.mrc-bsu.cam.ac.uk/bugs/> 查看详细介绍.

## WinBUGS 软件简介

- R ( S 和 Splus 的开源版本) 是一套完整的数据处理、计算和制图软件系统. 功能包括: 数据存储和处理系统; 数组运算工具 (其向量、矩阵运算功能尤其强大); 完整连贯的统计分析工具; 优秀的统计制图功能; 简便而强大的编程语言.
- 对贝叶斯计算来说, 可以直接使用 Gibbs 抽样和 Metropolis 算法进行编程, 前面都是如此做的. 对于高强度的计算问题, 可以使用 Fortran 或 C 语言然后从 R 中通过接口调用.
- BUGS (Bayesian inference Using Gibbs Sampling) 是一个免费软件, 其允许用户指定复杂的多层模型并使用 MCMC 算法来估计模型中的未知参数. WinBUGS 是 BUGS 在 Windows 平台下的版本, 而 openBUGS 是 BUGS 的开源版本. 还有一些其它的版本, 可以访问 <http://www.mrc-bsu.cam.ac.uk/bugs/> 查看详细介绍.

## WinBUGS 软件简介

- R ( S 和 Splus 的开源版本) 是一套完整的数据处理、计算和制图软件系统. 功能包括: 数据存储和处理系统; 数组运算工具 (其向量、矩阵运算功能尤其强大); 完整连贯的统计分析工具; 优秀的统计制图功能; 简便而强大的编程语言.
- 对贝叶斯计算来说, 可以直接使用 Gibbs 抽样和 Metropolis 算法进行编程, 前面都是如此做的. 对于高强度的计算问题, 可以使用 Fortran 或 C 语言然后从 R 中通过接口调用.
- BUGS (Bayesian inference Using Gibbs Sampling) 是一个免费软件, 其允许用户指定复杂的多层模型并使用 MCMC 算法来估计模型中的未知参数. WinBUGS 是 BUGS 在 Windows 平台下的版本, 而 openBUGS 是 BUGS 的开源版本. 还有一些其它的版本, 可以访问 <http://www.mrc-bsu.cam.ac.uk/bugs/> 查看详细介绍.

## WinBUGS 软件简介

- WinBUGS, 它与使用其他软件得到的分析结果是一致的, 即便是它们使用不同的 MCMC 算法. 使用 WinBUGS 软件进行贝叶斯分析的详细介绍也可以参考 Ntzoufras (2009).
- WinBUGS 软件最新版本是 WinBUGS 1.4.3. 下载地址是如前所述的网站, 即 <http://www.mrc-bsu.cam.ac.uk/bugs/>
- 双击 WinBUGS 版本的安装文件即可完成 WinBUGS 的安装. WinBUGS 软件图标位于安装目录下, 为操作方便, 建议用快捷方式将其拖至桌面. 双击该图标将其打开, 阅读自动弹出的许可证协议窗口后将其关闭. 单击左上角主菜单的 File 项, 在下拉菜单中选择 New, 打开一个新的文件窗口, 在此区域编写程序.

## WinBUGS 软件简介

- WinBUGS, 它与使用其他软件得到的分析结果是一致的, 即便是它们使用不同的 MCMC 算法. 使用 WinBUGS 软件进行贝叶斯分析的详细介绍也可以参考 Ntzoufras (2009).
- WinBUGS 软件最新版本是 WinBUGS 1.4.3. 下载地址是如前所述的网站, 即 <http://www.mrc-bsu.cam.ac.uk/bugs/>
- 双击 WinBUGS 版本的安装文件即可完成 WinBUGS 的安装. WinBUGS 软件图标位于安装目录下, 为操作方便, 建议用快捷方式将其拖至桌面. 双击该图标将其打开, 阅读自动弹出的许可证协议窗口后将其关闭. 单击左上角主菜单的 File 项, 在下拉菜单中选择 New, 打开一个新的文件窗口, 在此区域编写程序.

## WinBUGS 软件简介

- WinBUGS, 它与使用其他软件得到的分析结果是一致的, 即便是它们使用不同的 MCMC 算法. 使用 WinBUGS 软件进行贝叶斯分析的详细介绍也可以参考 Ntzoufras (2009).
- WinBUGS 软件最新版本是 WinBUGS 1.4.3. 下载地址是如前所述的网站, 即 <http://www.mrc-bsu.cam.ac.uk/bugs/>
- 双击 WinBUGS 版本的安装文件即可完成 WinBUGS 的安装. WinBUGS 软件图标位于安装目录下, 为操作方便, 建议用快捷方式将其拖至桌面. 双击该图标将其打开, 阅读自动弹出的许可证协议窗口后将其关闭. 单击左上角主菜单的 File 项, 在下拉菜单中选择 New, 打开一个新的文件窗口, 在此区域编写程序.



## 使用R 调用WinBUGS

- 另外一种方式. 使用 R 里的包 R2WinBUGS 来调用 WinBUGS. 安装好 WinBUGS 和此包后, 就可以在 R 里使用 WinBUGS 进行贝叶斯分析了. 在 R 里使用 WinBUGS 进行贝叶斯分析的主要函数为 `bugs`, 在运行此函数后, WinBUGS 会自动打开并运行相应的过程, 然后自动关闭.
- 在 R 里还可以调用包 `coda` 来作更多的分析. 如绘制自相关图, 诊断收敛性等. 请参考 `coda` 的说明文档.

## 使用R 调用WinBUGS

- 另外一种方式. 使用 R 里的包 R2WinBUGS 来调用 WinBUGS. 安装好 WinBUGS 和此包后, 就可以在 R 里使用 WinBUGS 进行贝叶斯分析了. 在 R 里使用 WinBUGS 进行贝叶斯分析的主要函数为 `bugs`, 在运行此函数后, WinBUGS 会自动打开并运行相应的过程, 然后自动关闭.
- 在 R 里还可以调用包 `coda` 来作更多的分析. 如绘制自相关图, 诊断收敛性等. 请参考 `coda` 的说明文档.