

基于 MCMC 算法的 贝叶斯统计方法

刘金山 夏 强\编著

 科学出版社

基于 MCMC 算法的 贝叶斯统计方法

刘金山 夏 强\编著



科学出版社
北京

内 容 简 介

随着计算机、互联网等信息技术的发展,马尔可夫链蒙特卡罗(MCMC)模拟技术使贝叶斯统计方法得以应用于许多领域的复杂问题.

本书在介绍常用MCMC算法的基础上,着重介绍计算贝叶斯后验估计的MCMC方法和新发展的贝叶斯随机搜索模型选择方法,特别是MCMC方法在贝叶斯数据分析中的应用.为了便于读者掌握MCMC方法,书中提供了大量的数据分析案例及相应的算法程序、图表和模拟分析结果.

本书适合概率统计、计量经济、生物信息、环境统计和医学统计等学科的高校教师、研究人员、高年级本科生、硕士和博士研究生参阅,也适合数据分析和统计软件应用领域的有关科技人员使用.

图书在版编目(CIP)数据

基于MCMC算法的贝叶斯统计方法/刘金山,夏强编著. —北京:科学出版社, 2016.3

ISBN 978-7-03-047488-9

I. ①基… II. ①刘…②夏… III. ①贝叶斯方法-统计方法
IV. ①O212.8

中国版本图书馆CIP数据核字(2016)第043875号

责任编辑:郭勇斌 彭婧煜/责任校对:刘亚琦

责任印制:张 伟/封面设计:黄华斌

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

北京教图印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2016年3月第 一 版 开本:720×1000 1/16

2016年3月第一次印刷 印张:21

字数:420 000

定价:108.00 元

(如有印装质量问题,我社负责调换)

序

统计学作为分析和处理数据的一门经验科学，其发展总是与计算能力的发展紧密相关。数十年来以计算机为代表的现代信息技术的发展和广泛应用，极大地促进了处理大数据和复杂模型的能力的发展。

受益于这种先进技术最多的统计学领域是应用贝叶斯推断。虽然贝叶斯方法受到许多使用者的欢迎，但是它的发展受到计算方法的制约。近年来，这个问题发生了历史性的变化。一种相对简便且效果极佳的马尔可夫链蒙特卡罗 (MCMC) 模拟技术使贝叶斯方法在许多复杂的实际问题 and 大数据处理中得以应用。更可喜的是，这种技术在一般场合下都可以应用，无需采用深入的统计理论对问题进行解释和理解。

基于 MCMC 技术的现代统计方法的进一步发展的潜力是巨大的，有关这种技术的文献从 20 世纪 90 年代开始大量出现。未来，随着计算方法的进一步发展和大数据时代的到来，该技术方法在经济、金融、环境、生物、医学等许多领域的数据分析中必定有更加广阔的前景，无论在理论上和应用上都必将有更进一步的发展。

杨振海

2015 年 12 月于北京

目 录

序	
引言	1
第 1 章 贝叶斯统计中的 MCMC 方法	4
1.1 Gibbs 算法	4
1.1.1 基本 Gibbs 抽样	4
1.1.2 Gibbs 抽样的收敛性	10
1.1.3 潜变模型的 Gibbs 抽样	12
1.1.4 分层模型的 Gibbs 抽样	18
1.2 M-H 算法	21
1.2.1 基本 M-H 算法	21
1.2.2 M-H 算法的收敛性	23
1.2.3 独立链和随机游走链 M-H 算法	24
1.3 混合算法	32
1.4 逃逸算法	34
1.5 可逆跳 MCMC 算法	37
1.6 MCMC 算法的加速收敛	38
1.6.1 分组移动和多网格 MC 抽样	39
1.6.2 MCMC 算法的协方差改进	41
1.6.3 评分数据例子	42
1.6.4 其他加速收敛方法	45
1.7 收敛性诊断	49
1.7.1 收敛到平稳分布	49
1.7.2 收敛到后验均值	50
第 2 章 贝叶斯后验估计	54
2.1 计算后验均值	54
2.1.1 基本 MC 方法	54
2.1.2 标准误差的估计	57
2.1.3 MC 估计的改进	59
2.1.4 控制模拟误差	63
2.2 计算后验众数	65

2.2.1	矩阵微分	65
2.2.2	Lindley-Smith 最优化	66
2.2.3	随机近似方法	68
2.3	估计边缘后验密度	71
2.3.1	边缘后验密度	71
2.3.2	核估计方法	73
2.3.3	重要性加权密度估计	74
2.3.4	基于 K-L 散度的有效性	80
第 3 章	贝叶斯模型选择	91
3.1	贝叶斯模型比较	91
3.1.1	边缘似然	91
3.1.2	预测密度	94
3.2	随机搜索模型选择方法	95
3.2.1	线性回归模型的变量选择	96
3.2.2	自回归模型的滞后项选择	100
3.3	贝叶斯模型平均	106
3.4	模型选择的可逆跳 MCMC 方法	109
3.4.1	可逆跳 MCMC 方法	110
3.4.2	变阶数时间序列模型选择方法	112
第 4 章	多元贝叶斯回归模型	118
4.1	多元线性回归模型	118
4.1.1	共轭先验下的贝叶斯推断	119
4.1.2	广义先验下的贝叶斯推断	122
4.1.3	实例分析	125
4.2	向量自回归模型	129
4.2.1	VAR 模型定义	129
4.2.2	似然函数和先验分布	131
4.2.3	VAR 模型约束的随机搜索	133
4.2.4	模拟实验	136
4.3	SUR 模型的 Gibbs 抽样	137
4.3.1	基本 SUR 模型	138
4.3.2	数值模拟	139
4.4	SUR 模型的分层贝叶斯分析	140
4.4.1	分层模型的 Gibbs 抽样	141
4.4.2	具有相关序列误差的 SUR 模型	142

4.4.3	具有时变参数的 SUR 模型	145
4.4.4	实例分析	147
4.5	SUR 模型的变系数 MCMC 方法	149
4.5.1	SUR 模型的变换形式	149
4.5.2	模型参数变换	152
4.5.3	模拟实验	153
4.5.4	实例分析	157
4.6	非参数 SUR 模型的 MCMC 方法	162
4.6.1	基函数展开及分层分析	163
4.6.2	条件后验分布	164
4.6.3	MCMC 抽样方法	166
4.6.4	模拟实验	168
4.6.5	实例分析	171
4.7	多元空间回归模型	172
4.7.1	公共分量模型	173
4.7.2	空间回归模型	175
4.7.3	先验分布	178
4.7.4	后验推断	179
4.7.5	实例分析	182
第 5 章	一些非线性时间序列模型	185
5.1	门限自回归模型	185
5.1.1	模型参数估计	185
5.1.2	模型选择的 RJMCMC 方法	186
5.1.3	抽样方案	187
5.1.4	模拟举例	188
5.2	门限自回归模型的变点分析	190
5.2.1	模型的定义	190
5.2.2	模型参数的贝叶斯推断	191
5.2.3	模型后验概率的估计	194
5.2.4	模拟举例	195
5.3	多变点门限自回归模型	198
5.3.1	模型的参数估计	198
5.3.2	门限估计的 MCMC 方法	203
5.3.3	模拟举例	205
5.4	门限自回归滑动平均模型	208

5.4.1	模型参数的贝叶斯推断	208
5.4.2	模型选择的 RJMCMC 方法	210
5.4.3	模拟举例	210
5.5	带有外生变量的门限自回归滑动平均模型	212
5.5.1	模型参数的贝叶斯推断	213
5.5.2	模拟举例	216
第 6 章	其他贝叶斯模型	218
6.1	空气污染模型	218
6.1.1	贝叶斯模型分析	218
6.1.2	实际数据分析	222
6.1.3	结论	229
6.2	货币汇率模型	230
6.2.1	双门限 GARCH 模型	230
6.2.2	参数的贝叶斯推断	231
6.2.3	实证分析	233
6.3	大脑图像分割模型	236
6.3.1	功能性磁共振成像	236
6.3.2	分层混合模型的贝叶斯分析	237
6.3.3	分层混合模型的 RJMCMC 方法	240
6.3.4	大脑 FMRI 图像分割	242
6.4	河流水文数据模型	245
6.4.1	变点水文时间序列概述	245
6.4.2	贝叶斯变点时序分析	246
6.4.3	贝叶斯变点回归分析	254
6.4.4	结论	258
6.5	随机波动模型	259
6.5.1	随机波动模型简介	259
6.5.2	随机波动模型的贝叶斯推断	260
6.5.3	实证分析	265
参考文献		270
附录 A	常用统计分布	284
A.1	一元统计分布	284
A.1.1	二项分布	284
A.1.2	贝塔分布	284
A.1.3	正态分布	285

A.1.4	伽马分布	285
A.1.5	逆伽马分布	286
A.1.6	学生 t 分布	286
A.1.7	F 分布	287
A.2	多元统计分布	287
A.2.1	多元正态分布	287
A.2.2	多元 t 分布	288
A.3	随机矩阵分布	288
A.3.1	矩阵正态分布	288
A.3.2	Wishart 分布	290
A.3.3	逆 Wishart 分布	290
A.3.4	矩阵 T 分布	291
附录 B	统计软件简介	292
B.1	R 软件介绍	292
B.1.1	数据操作	292
B.1.2	统计分析的常见 R 函数及概率分布函数	294
B.1.3	可视化	295
B.2	WinBUGS 软件介绍	296
B.2.1	软件简介	296
B.2.2	编程要点	297
B.2.3	软件操作	297
B.2.4	举例	298
B.3	R 与 Winbugs 接口	299
B.3.1	R2WinBUGS 软件包介绍	299
B.3.2	coda 软件包介绍	300
B.3.3	举例	300
附录 C	4.6.4 节模拟实验的 R 程序	302

引 言

贝叶斯方法在统计学中的地位可用已故中国科学院院士陈希孺的一段话来形容:“托马斯·贝叶斯这个生性孤僻,哲学气味重于数学气味的学术怪杰,以其一篇遗作的思想重大地影响了两个世纪以后的统计学术界,顶住了统计学的半边天。”

贝叶斯原理基于由观测数据 D 的一个特定的概率模型导出的似然函数 $L(D|\theta)$, 这里假定 θ 是随机的,它具有先验分布 $\pi(\theta)$. 关于 θ 的推断基于由贝叶斯定理得到的后验分布 $\pi(\theta|D)$, 其函数形式为

$$\pi(\theta|D) = \frac{L(D|\theta)\pi(\theta)}{\int_{\Omega} L(D|\theta)\pi(\theta)d\theta},$$

其中 Ω 表示 θ 的参数空间. 由上式可知, $\pi(\theta|D)$ 正比例于似然和先验的乘积, 即

$$\pi(\theta|D) \propto L(D|\theta)\pi(\theta),$$

称后者为后验分布的核. 因此, 后验分布 $\pi(\theta|D)$ 综合了 θ 的样本信息 $L(D|\theta)$ 和先验信息 $\pi(\theta)$. 数值 $m(D) = \int_{\Omega} L(D|\theta)\pi(\theta)d\theta$ 是 $\pi(\theta|D)$ 的规范化常数, 也常称其为数据的边缘分布或先验预测分布.

尽管贝叶斯推断模式简单, 并且概率形式优美, 然而在贝叶斯分析中, 一般只知道后验分布的核, 而难以获得精确的后验分布, 这是因为在多数统计模型或应用中, $m(D)$ 不存在解析形式表达式, 因此 $\pi(\theta|D)$ 也不存在解析表达式. 此时, 如何从后验分布 $\pi(\theta|D)$ 推断 θ 是贝叶斯统计中面临的重要问题, 而对于贝叶斯算法的研究成为现代贝叶斯统计中的热点问题.

在现代贝叶斯算法中有两个重要挑战, 即如何从后验分布中抽取样本和如何利用现代马尔可夫链蒙特卡罗 (Markov Chain Monte Carlo, MCMC) 方法计算所感兴趣的后验量. 当贝叶斯模型太复杂, 或维数过高时, 传统的解析近似或数值积分方法往往会失效, 此时可以自然地采用近似蒙特卡罗 (Monte Carlo, MC) 模拟方法. 随着 MCMC 技术的发展, 出现了许多行之有效的 MC 方法. 近年来, MCMC 技术的发展为贝叶斯方法的应用开辟了广阔的前景, 使得贝叶斯推断理论和方法的研究得到复兴. MCMC 方法已经成为一种特别流行的处理复杂统计问题的工具, 尤其在常常需要复杂的高维积分运算的贝叶斯分析领域更是如此. 以 MCMC 为代表的现代贝叶斯统计方法已广泛应用于几乎所有学科, 并取得显著成果.

已有一些文献介绍基本 MCMC 方法及其现代发展情况. 本书在介绍几个常用 MCMC 算法的基础上, 重点介绍利用来自后验分布的样本计算后验估计的 MCMC 方法, 特别是近期发展的贝叶斯变量选择方法. 本书特别强调 MCMC 方法在贝叶斯数据分析中的应用. 本书内容共分 6 章.

第 1 章介绍几个常用 MCMC 抽样算法, 其中包括 Gibbs 算法、M-H (Metropolis-Hastings) 算法、混合 (Hybrid) 算法、逃逸 (Hit-and-Run) 算法和可逆跳 MCMC 算法. 本章还介绍几个关于加速 MCMC 抽样算法的有用工具, 包括分组移动和多网格 MC 抽样, 以及 MCMC 算法的协方差改进方法, 并讨论了马尔可夫链的收敛性诊断问题.

第 2 章讨论如何利用 MCMC 抽样算法估计所感兴趣参数的后验均值、后验众数和边缘后验密度. 本章还讨论关于 MC 估计的模拟效率的改进问题. 为评价 MC 估计的模拟效率, 给出几个用于计算模拟标准误差的有效方法, 讨论如何在计算 MC 估计时控制模拟误差等问题. 对于边缘后验密度的估计, 提供几种有效的方法, 利用散度评价边缘后验密度估计的有效性.

第 3 章讨论贝叶斯模型选择问题, 包括模型比较和变量选择问题. 贝叶斯变量选择方法是近期发展的基于 MCMC 算法的现代统计方法. 与传统的模型选择方法 (如 AIC 和 BIC 方法) 相比, 贝叶斯变量选择方法可以更好地利用样本信息和先验信息, 且具有更大的灵活性. 本章将系统讨论贝叶斯模型选择方法, 包括贝叶斯模型比较、随机搜索变量选择方法、贝叶斯模型平均以及模型选择的可逆跳 MCMC 算法.

第 4 章至第 6 章讨论基于 MCMC 算法的贝叶斯方法在若干领域的应用, 特别是在贝叶斯回归分析和时间序列分析中的应用问题. 在多元回归分析中, 重点讨论在计量经济等领域中具有广泛应用背景的多元相依回归模型, 或称为似乎不相关回归模型 (Seemingly Unrelated Regression Model, SUR 模型) 的贝叶斯统计推断及其应用. 在时间序列分析中, 重点讨论门限及变点非线性时间序列模型的贝叶斯推断及其应用. 这些方面的内容多数是作者近年来的研究成果. 具体内容如下:

第 4 章系统介绍基于 MCMC 算法的多元贝叶斯回归模型, 其中包括多元线性回归模型、向量自回归模型、线性 SUR 模型、分层 SUR 模型、SUR 模型的参数变换方法、非参数 SUR 模型和多元空间回归模型的贝叶斯推断.

第 5 章介绍基于 MCMC 算法的非线性时间序列模型, 其中包括单变点门限自回归模型、门限自回归模型的变点分析、多变点门限自回归模型、门限自回归滑动平均模型、带有外生变量的门限自回归滑动平均模型.

第 6 章主要介绍基于 MCMC 算法的现代贝叶斯方法在若干实际模型中的应用, 分别介绍空气污染模型、人民币汇率市场分析模型、人类大脑图像分割模型、河流水文数据模型、金融和经济等领域中的随机波动模型.

为了便于读者掌握 MCMC 方法, 书中提供了大量的数据分析案例以及部分算法程序、图表和模拟分析结果. 利用这些案例和程序, 读者可以快速掌握基本 MCMC 方法以及在科研和许多实际问题分析中应用的现代贝叶斯方法. 本书的算法程序主要是 R 软件程序及部分 WinBugs 算法程序. 这两个软件都是贝叶斯计算中的常用统计软件, 且可在互联网上自由下载, 免费使用.

本书撰写过程中参考了大量的国内外文献资料, 其主要内容源自国家自然科学基金项目 (编号:11171117) 和国家社会科学基金项目 (编号:12CTJ019) 的研究结果, 其中有一些结果已发表在中外学术期刊上. 作者感谢历年来参加作者主持的关于基于 MCMC 算法的贝叶斯方法的讨论班的研究生陈镇坤、张韬、卢素、曾婉红、龙健颜、黄来华、万稚慧、唐熙等同学, 他们的工作对本书的完成有很大的帮助, 特别是张韬同学对本书的初稿做了很多有益的工作, 陈镇坤同学对本书中的许多案例设计了算法程序, 并进行抽样运算, 由于篇幅所限, 这些算法程序没有放在书中.

作 者

2015 年 12 月

第 1 章 贝叶斯统计中的 MCMC 方法

在近代贝叶斯统计推断中, 利用蒙特卡罗 (MC) 抽样方法计算高维后验积分的研究得到迅猛发展. 该模拟方法包括重要性 MC 抽样, Gibbs 抽样以及混合 MC 抽样等算法, 其中最为人们熟知的是马尔可夫链蒙特卡罗 (MCMC) 算法. 本章介绍几个常用 MCMC 算法, 包括 Gibbs 算法、M-H 算法、混合算法、逃逸算法和可逆跳 MCMC 算法. 还介绍几个关于加速 MCMC 抽样算法的有用工具, 包括分组移动和多网格 MCMC 抽样, 以及 MCMC 估计的协方差改进方法. 最后, 讨论用 MCMC 算法得到的马尔可夫 (Markov) 链的收敛性诊断问题.

1.1 Gibbs 算法

1.1.1 基本 Gibbs 抽样

Gibbs 抽样方法也许是贝叶斯计算中最为人们熟悉的一种 MCMC 算法. Gibbs 抽样的思想最早由 Grenander (1983) 提出, 而正式术语是由 Geman 和 Geman (1984) 引进的. 把 Gibbs 抽样用于贝叶斯推断问题的具有里程碑意义的文献是 Gelfand 和 Smith (1990). Casella 和 George (1992) 是关于 Gibbs 抽样的一篇优秀文献.

Gibbs 抽样的基本思想是对高维参数进行后验推断时, 通过参数向量 θ 的分量的条件分布族来构造 Markov 链 $\{\theta^{(i)}\}$, 使它的不变分布为目标分布. 设 $\theta = (\theta_1, \dots, \theta_p)'$ 是 p 维参数向量, $\pi(\theta|D)$ 是观察到数据集 D 后 θ 的后验分布. 则基本 Gibbs 抽样方法如下:

第 0 步. 任意选取一个初始点 $\theta^{(0)} = (\theta_{1,0}, \theta_{2,0}, \dots, \theta_{p,0})'$, 并置 $i = 0$;

第 1 步. 按下列方法生成 $\theta^{(i+1)} = (\theta_{1,i+1}, \theta_{2,i+1}, \dots, \theta_{p,i+1})'$:

生成 $\theta_{1,i+1} \sim \pi(\theta_1 | \theta_{2,i}, \dots, \theta_{p,i}, D)$,

生成 $\theta_{2,i+1} \sim \pi(\theta_2 | \theta_{1,i+1}, \theta_{3,i}, \dots, \theta_{p,i}, D)$,

.....

生成 $\theta_{p,i+1} \sim \pi(\theta_p | \theta_{1,i+1}, \theta_{2,i+1}, \dots, \theta_{p-1,i+1}, D)$;

第 2 步. 置 $i = i + 1$, 并返回到第 1 步.

在这个算法过程中, θ 的每一个分量按照自然顺序生成, 每一个循环需要生成 p 个随机变量. 下面通过几个例子说明 Gibbs 抽样方法.

例 1.1.1 假设后验分布 $\pi(\theta|D)$ 是一个二元正态分布 $N(\mu, \Sigma)$, 其中

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

其中 μ_j, σ_j 和 ρ 已知, $j = 1, 2$. 则 Gibbs 抽样要求按下列分布通过逐步迭代方式抽取样本

$$\begin{aligned} \theta_1 &\sim N\left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(\theta_2 - \mu_2), \sigma_1^2(1 - \rho^2)\right), \\ \theta_2 &\sim N\left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(\theta_1 - \mu_1), \sigma_2^2(1 - \rho^2)\right). \end{aligned}$$

令 $\{\theta^{(i)} = (\theta_{1,i}, \theta_{2,i})', i \geq 0\}$ 表示由 Gibbs 抽样方法生成的关于上述二元正态分布的马尔可夫链. 若从平稳分布出发, 即令 $\theta^{(0)} \sim N(\mu, \Sigma)$, 则 $\{\theta_{1,i}, i \geq 0\}$ 和 $\{\theta_{2,i}, i \geq 0\}$ 中的任何一个都是一个 AR(1) 过程.

为了说明这一点, 令 $\{z_{1,i}, z_{2,i}, i \geq 0\}$ 为两个服从 $N(0, 1)$ 分布的独立同分布 (iid) 随机变量序列, 则 Gibbs 抽样结构蕴涵着

$$\begin{aligned} \theta_{1,0} &= \mu_1 + \sigma_1 z_{1,0}, \\ \theta_{2,0} &= \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(\theta_{1,0} - \mu_1) + \sigma_2 \sqrt{1 - \rho^2} z_{2,0}, \end{aligned}$$

及

$$\begin{aligned} \theta_{1,i+1} &= \mu_1 + \rho \frac{\sigma_1}{\sigma_2}(\theta_{2,i} - \mu_2) + \sigma_1 \sqrt{1 - \rho^2} z_{1,i+1}, \\ \theta_{2,i+1} &= \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(\theta_{1,i+1} - \mu_1) + \sigma_2 \sqrt{1 - \rho^2} z_{2,i+1}, \end{aligned} \quad (1.1.1)$$

其中 $i \geq 0$. 现在, 考虑第一个分量 $\theta_{1,i+1}$. 由 (1.1.1) 式, 对于 $i \geq 0$,

$$\begin{aligned} \theta_{1,i+1} &= \mu_1 + \rho \frac{\sigma_1}{\sigma_2} \left[\rho \frac{\sigma_2}{\sigma_1}(\theta_{1,i} - \mu_1) + \sigma_2 \sqrt{1 - \rho^2} z_{2,i} \right] \\ &\quad + \sigma_1 \sqrt{1 - \rho^2} z_{1,i+1} \\ &= \mu_1 + \rho^2(\theta_{1,i} - \mu_1) + \rho\sigma_1 \sqrt{1 - \rho^2} z_{2,i} \\ &\quad + \sigma_1 \sqrt{1 - \rho^2} z_{1,i+1}. \end{aligned} \quad (1.1.2)$$

设 $\psi = \rho^2, \sigma_1^{*2} = \sigma_1^2(1 - \rho^4)$, 并令 $\{z_i^*, i \geq 0\}$ 为服从 $N(0, 1)$ 分布的 iid 随机变量序列. 由于 $z_{1,i}$ 与 $z_{2,i+1}$ 独立同服从分布 $N(0, 1)$, 因此, 可以把 (1.1.2) 式表示为

$$\begin{aligned} \theta_{1,0} &= \mu_1 + \sigma_1 z_0^*, \\ \theta_{1,i+1} &= \mu_1 + \psi(\theta_{1,i} - \mu_1) + \sigma_1^* z_{i+1}^*, \quad i \geq 0. \end{aligned} \quad (1.1.3)$$

因此, $\{\theta_{1,i}, i \geq 0\}$ 是一个 AR(1) 过程, 其延迟为 1, 自相关系数为 $\psi = \rho^2$. 类似地, $\{\theta_{2,i}, i \geq 0\}$ 也是一个 AR(1) 过程, 其延迟为 1, 自相关系数为 $\psi = \rho^2$. 所不同的是, 要用 $\sigma_2^* = \sigma_2 \sqrt{1 - \rho^4}$ 代替 (1.1.3) 式中的 σ_1^* .

Roberts 和 Sahu (1997) 对于一个一般多元正态目标分布 $\pi(\theta|D)$, 得到了一个类似结果, 即通过 Gibbs 抽样得到的马尔可夫链是一个多元 AR(1) 过程.

例 1.1.2 考虑多元线性回归模型

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon, \quad (1.1.4)$$

其中 $\varepsilon \sim N(0, \sigma^2)$, 假定系数参数 β_j 满足单调约束

$$0 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_p.$$

系数参数空间为

$$\Omega = \{\beta = (\beta_1, \beta_2, \cdots, \beta_p)' : 0 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_p, \beta \in \mathbf{R}^p\}. \quad (1.1.5)$$

考虑采用贝叶斯方法估计参数 (β, σ^2) . 取 (β, σ^2) 的联合先验分布为

$$\pi(\beta, \sigma^2) \propto \frac{1}{\sigma^2} \pi(\beta_p | \mu_0, \sigma_0^2), \quad (1.1.6)$$

其中 $\sigma^2 > 0$, $\beta \in \Omega$, $\pi(\beta_p | \mu_0, \sigma_0^2)$ 是均值为 μ_0 方差为 σ_0^2 的正态分布密度. 由 (1.1.6) 式, (β, σ^2) 的基于数据 D 的后验分布为

$$\begin{aligned} \pi(\beta, \sigma^2 | D) &= \frac{\exp\{-(\beta_p - \mu_0)^2 / 2\sigma_0^2\}}{c(D)(\sigma^2)^{(n+2)/2}} \\ &\times \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij}\beta_j\right)^2\right\}, \end{aligned} \quad (1.1.7)$$

其中 $\sigma^2 > 0$, $\beta \in \Omega$, $(y_i, x_{i1}, \cdots, x_{ip})$ 是 $(y, x_1, \cdots, x_p)$ 的第 i 次观察值, $c(D)$ 是规范化常数. 由于约束的限制, 不可能得到参数的后验解析结果, 如 β_j 的后验均值和后验标准差. 然而, 对后验分布 (1.1.7) 式采用 Gibbs 抽样是直接的. 特别是, 可按以下程序进行 Gibbs 抽样运算:

$$\beta_j | \beta_1, \cdots, \beta_{j-1}, \beta_{j+1}, \cdots, \beta_p, \sigma^2, D \sim N(\theta_j, \delta_j^2), \quad j = 1, 2, \cdots, p-1, \quad (1.1.8)$$

约束条件为 $\beta_{j-1} \leq \beta_j \leq \beta_{j+1} (\beta_0 = 0)$;

$$\beta_p | \beta_1, \cdots, \beta_{p-1}, \sigma^2, D \sim N(\psi\theta_p + (1-\psi)\mu_0, (1-\psi)\sigma_0^2), \quad (1.1.9)$$

约束条件为 $\beta_p \geq \beta_{p-1}$;

$$\sigma^2 | \beta, D \sim IG \left\{ \frac{n}{2}, \frac{1}{2} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij} \beta_j \right)^2 \right\}. \quad (1.1.10)$$

在 (1.1.8) 式和 (1.1.9) 式中, $\psi = \sigma_0^2 / (\sigma_0^2 + \delta_p^2)$,

$$\begin{aligned} \theta_j &= \left(\sum_{i=1}^n x_{ij}^2 \right)^{-1} \left[\sum_{i=1}^n \left(y_i - \sum_{l \neq j} x_{il} \beta_l \right) x_{ij} \right], \\ \delta_j^2 &= \left(\sum_{i=1}^n x_{ij}^2 \right)^{-1} \sigma^2, \quad j = 1, 2, \dots, p. \end{aligned} \quad (1.1.11)$$

在 (1.1.10) 式中, $IG(\xi, \eta)$ 表示参数为 (ξ, η) 的逆伽马 (Gamma) 分布, 其密度函数为

$$\pi(\sigma^2 | \xi, \eta) \propto (\sigma^2)^{-(\xi+1)} e^{-\eta/\sigma^2}.$$

例 1.1.3 考虑下列分层正态模型

$$\begin{cases} x_{ij} \sim N(\theta_i, \sigma^2), & j = 1, \dots, n_i, i = 1, \dots, k, \\ \theta_i \sim N(\mu, \tau^2), & i = 1, \dots, k, \\ \mu \sim N(\mu_0, \sigma_\mu^2), \\ \sigma^2 \sim IG(a_1, b_1), \tau^2 \sim IG(a_2, b_2), \sigma_\mu^2 \sim IG(a_3, b_3), \end{cases} \quad (1.1.12)$$

其中 a_i, b_i 是给定的超参数.

设观察数据为 $x = \{(x_{i1}, x_{i2}, \dots, x_{in_i}), i = 1, \dots, k\}$, 并记 $\theta = (\theta_1, \dots, \theta_k)$. 参数 $\theta_i, \mu, \sigma^2, \tau^2, \sigma_\mu^2$ 的条件后验密度分别为

$$\begin{aligned} \pi(\theta_i | \mu, \sigma^2, \tau^2, \sigma_\mu^2, x) &\propto f(x | \theta, \mu, \sigma^2, \tau^2, \sigma_\mu^2) \pi(\theta_i | \mu, \tau^2) \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_j (x_{ij} - \theta_i)^2 \right\} \exp \left\{ -\frac{1}{2\tau^2} (\theta_i - \mu)^2 \right\} \\ &\propto \exp \left\{ -\frac{n_i \tau^2 + \sigma^2}{2\sigma^2 \tau^2} \left[\theta_i - \left(\frac{\sigma^2}{n_i \tau^2 + \sigma^2} \mu + \frac{n_i \tau^2}{n_i \tau^2 + \sigma^2} \bar{x}_i \right) \right]^2 \right\}, \\ &\quad i = 1, 2, \dots, k, \end{aligned}$$

$$\begin{aligned} \pi(\mu | \theta, \sigma^2, \tau^2, \sigma_\mu^2, x) &\propto \pi(\theta | \mu, \tau^2) \pi(\mu | \sigma_\mu^2) \\ &\propto \exp \left\{ -\frac{1}{2\tau^2} \sum_i (\theta_i - \mu)^2 \right\} \exp \left\{ -\frac{1}{2\sigma_\mu^2} (\mu - \mu_0)^2 \right\} \\ &\propto \exp \left\{ -\frac{k\sigma_\mu^2 + \tau^2}{2\sigma_\mu^2 \tau^2} \left[\mu - \left(\frac{\tau^2}{k\sigma_\mu^2 + \tau^2} \mu_0 + \frac{k\sigma_\mu^2}{k\sigma_\mu^2 + \tau^2} \bar{\theta} \right) \right]^2 \right\}, \end{aligned}$$

$$\begin{aligned}
 \pi(\sigma^2|\theta, \mu, \tau^2, \sigma_\mu^2, x) &\propto f(x|\theta, \mu, \sigma^2, \tau^2, \sigma_\mu^2)\pi(\sigma^2) \\
 &\propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{ij} (x_{ij} - \theta_i)^2\right\} \frac{1}{(\sigma^2)^{a_1+1}} \exp\left\{-\frac{b_1}{\sigma^2}\right\} \\
 &\propto \frac{1}{(\sigma^2)^{\frac{n}{2}+a_1+1}} \exp\left\{-\frac{\frac{1}{2} \sum_{ij} (x_{ij} - \theta_i)^2 + b_1}{\sigma^2}\right\},
 \end{aligned}$$

$$\begin{aligned}
 \pi(\tau^2|\theta, \mu, \sigma^2, \sigma_\mu^2, x) &\propto \pi(\theta|\mu, \tau^2)\pi(\tau^2) \\
 &\propto \frac{1}{(\tau^2)^{\frac{k}{2}}} \exp\left\{-\frac{1}{2\tau^2} \sum_i (\theta_i - \mu)^2\right\} \frac{1}{(\tau^2)^{a_2+1}} \exp\left\{-\frac{b_2}{\tau^2}\right\} \\
 &\propto \frac{1}{(\tau^2)^{\frac{k}{2}+a_2+1}} \exp\left\{-\frac{\frac{1}{2} \sum_i (\theta_i - \mu)^2 + b_2}{\tau^2}\right\},
 \end{aligned}$$

$$\begin{aligned}
 \pi(\sigma_\mu^2|\theta, \mu, \sigma^2, \tau^2, x) &\propto \pi(\sigma_\mu^2|\mu)\pi(\mu) \\
 &\propto \frac{1}{(\sigma_\mu^2)^{a_3+1}} \exp\left\{-\frac{b_3}{\sigma_\mu^2}\right\} \frac{1}{\sigma_\mu} \exp\left\{-\frac{1}{2\sigma_\mu^2}(\mu - \mu_0)^2\right\} \\
 &\propto \frac{1}{(\sigma_\mu^2)^{\frac{1}{2}+a_3+1}} \exp\left\{-\frac{\frac{1}{2}(\mu - \mu_0)^2 + b_3}{\sigma_\mu^2}\right\},
 \end{aligned}$$

其中 $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$, $\bar{\theta} = \frac{1}{k} \sum_{i=1}^k \theta_i$, $n = \sum_{i=1}^k n_i$. 即有以下条件后验分布

$$\theta_i|\mu, \sigma^2, \tau^2, \sigma_\mu^2, x \sim N\left(\frac{\sigma^2}{n_i\tau^2 + \sigma^2}\mu + \frac{n_i\tau^2}{n_i\tau^2 + \sigma^2}\bar{x}_i, \frac{\sigma^2\tau^2}{n_i\tau^2 + \sigma^2}\right), \quad i = 1, \dots, k,$$

$$\mu|\theta, \sigma^2, \tau^2, \sigma_\mu^2, x \sim N\left(\frac{\tau^2}{k\sigma_\mu^2 + \tau^2}\mu_0 + \frac{k\sigma_\mu^2}{k\sigma_\mu^2 + \tau^2}\bar{\theta}, \frac{\tau^2\sigma_\mu^2}{k\sigma_\mu^2 + \tau^2}\right),$$

$$\sigma^2|\theta, \mu, \tau^2, \sigma_\mu^2, x \sim IG\left(\frac{n}{2} + a_1, \frac{1}{2} \sum_{ij} (x_{ij} - \theta_i)^2 + b_1\right),$$

$$\tau^2|\theta, \mu, \sigma^2, \sigma_\mu^2, x \sim IG\left(\frac{k}{2} + a_2, \frac{1}{2} \sum_i (\theta_i - \mu)^2 + b_2\right),$$

$$\sigma_\mu^2|\theta, \mu, \sigma^2, \tau^2, x \sim IG\left(\frac{1}{2} + a_3, \frac{1}{2}(\mu - \mu_0)^2 + b_3\right).$$