

湖北大学

硕士学位论文

MCMC方法及应用

姓名：张忠诚

申请学位级别：硕士

专业：基础数学

指导教师：张绍义

20080501

摘 要

在统计学中,会经常遇到积分计算问题,特别是高维积分的计算,用传统的数值方法往往很难解决高维积分计算问题,随着计算机的迅速发展,我们可通过随机模拟的方法解决高维积分计算问题。随机模拟方法适用的范围非常广泛,它既能求解确定性的问题,也能求解随机性的问题以及科学研究中理论性的问题。如计算高维积分、求解代数方程组和计算逆矩阵等。

在随机模拟中,关键是随机样本的抽取。本文主要介绍了均匀分布抽样、已知分布抽样、经验分布抽样以及随机向量抽样的一般方法,通过这些抽样方法,我们可以对一些低维联合分布进行抽样,但直接从一个任意的高维联合分布中产生样本常常是比较困难的,这样使得基于样本的方法具有了局限性。MCMC(Markov Chain Monte Carlo)方法是一种简单易行、应用广泛的计算随机模拟方法,该方法的核心思想是构造一个概率转移矩阵,建立一个以分布 $\pi(x)$ 为平稳分布的 Markov 链来得到 $\pi(x)$ 的样本,通过随机抽样得到的这些样本就可进行各种统计推断。本文主要探讨了 MCMC 随机模拟的思想和一般方法,对应用较为广泛的 MCMC 抽样中的 Gibbs 抽样方法、Metropolis 抽样方法进行了深入研究,给出了实现 Gibbs 抽样、Metropolis 抽样的具体步骤和算法,给出了基于 Bayes 参数估计的 Gibbs 抽样、Metropolis 抽样具体步骤和算法。通过实例进一步说明了如何通过 MCMC 抽样解决实际问题。

关键词: 随机模拟; MCMC 抽样; Bayes 参数估计; Gibbs 抽样; Metropolis 抽样

Abstract

In statistics, we often deal with the problems of integral calculation, particularly the high-dimensional integral calculation. Usually, it is often very difficult to solve the high-dimensional integral calculation by using the traditional numerical methods. With the rapid development of computer, we can use the method of stochastic simulation to solve the problems of high-dimensional integral calculation. The method of stochastic simulation is applied in many fields extensively. It can be used to solve the certainty problem, stochastic randomness problem, and other theoretical problems in scientific research, such as calculating high-dimensional integral, solving algebraic equations and calculating inverse matrix and so on.

In stochastic simulation, the key problem is how to extract the random samples. This paper mainly introduced the methods of uniform distribution sampling, known distribution sampling, empirical distribution sampling and random vector sampling. Based on these sampling methods, we can sample for some low-dimensional distribution. However, it is very difficult to sample from an arbitrary joint distribution of high-dimensional. So the sample methods are limited. MCMC (Markov Chain Monte Carlo) is a simple, widely used method of stochastic simulation. The core idea of this method is to construct a probability shift matrix and to establish markov chain with stationary distribution $\pi(x)$ to obtain the sample of $\pi(x)$. So we can do various kinds of statistical inference through random sampling of these samples.

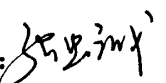
This paper mainly discusses the idea of MCMC stochastic simulation and the general approach. Gibbs sampling method and Metropolis sampling method in the wider application of MCMC sampling are deeply researched, then the concrete steps and algorithms are given to realize Gibbs sampling and Metropolis sampling. Based on the Bayes parameters estimation, the concrete steps and algorithms of Gibbs sampling and Metropolis sampling are also given. Last, examples are further given to illustrate how to solve the practical application problems through MCMC sampling.

Keywords: stochastic simulation; MCMC sampling ; Bayes parameters estimation; Gibbs sampling; Metropolis sampling

湖北大学学位论文原创性声明和使用授权说明

原创性声明

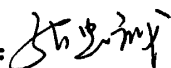
本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

论文作者签名：
日期：2008年5月31日

学位论文使用授权说明

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：

按照学校要求提交学位论文的印刷本和电子版本；学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，并提供目录检索与阅览服务；学校可以允许采用影印、缩印、数字化或其它复制手段保存学位论文；在不以赢利为目的的前提下，学校可以公开学位论文的部分或全部内容。（保密论文在解密后遵守此规定）

作者签名：

日期：2008.5.31

指导教师签名：

日期：

MCMC方法及应用

1 引言

随着现代计算机技术的出现和飞速发展,用计算机模拟概率过程,实现多次模拟试验并统计计算结果,进而可获得所求问题的近似结果.计算机的大存储量、高运算速度使得在短时间内,获得精度极高且内容丰富的模拟结果.人们便把这种计算机随机模拟方法称为蒙特卡罗方法.

随机模拟方法^[1-5]属于试验数学的一个分支.它是一种具有独特风格的数值计算方法,此方法是以概率统计理论为主要基础理论,以随机抽样作为主要手段的广义的数值计算方法.它们用随机数进行统计试验,把得到的统计特征值、均值、概率等,作为所求问题的数值解.蒙特卡罗方法不仅在处理具有概率性质的问题^[6-10]方面获得广泛的应用,对于具有确定性问题的计算也因其程序简单等优点获得了广泛的应用^[11-20].

蒙特卡罗(Monte Carlo)方法以随机模拟和统计试验为手段,是一种从随机变量的概率分布中,通过随机选择数字的方法产生一种符合该随机变量概率分布特性的随机数值序列,作为输入变量序列进行特定的模拟试验、求解的方法^[21-30].在应用Monte Carlo方法时,要求产生的随机数序列应符合该随机变量特定的概率分布.而产生各种特定的、不均匀的概率分布的随机数序列,可行的方法是先产生一种均匀分布的随机数序列,然后再设法转换成特定要求的概率分布的随机数序列,以此作为数字模拟试验的输入变量序列进行模拟求解.基本步骤如下:

(1)建立概率模型,即对所研究的问题构造一个符合其特点的概率模型(随机事件,随机变量等).包括对确定性问题,须把具体问题变为概率问题,建立概率模型;

(2)产生随机数序列,作为系统的抽样输入进行大量的数字模拟试验,得到大量的模拟试验值;

(3)对模拟试验结果进行统计处理(计算频率、均值等特征值),给出所求问题的解和解的精度的估计.

Monte Carlo法可以用来求解某些确定性问题,特别是对工程技术中的数学模型问题(如计算高维积分、求解代数方程组和计算逆矩阵等)在一般的解析法或数值法求解遇到困难时尤其有效.本文主要探讨随机模拟的思想以及一般方法,着重研究基于MCMC随机模拟的Gibbs抽样方法和Metropolis抽样方法,以及如何实现Gibbs抽样和Metropolis抽样.

2 随机数的产生

2.1 均匀分布随机数的产生

均匀随机数是指理论上没有规律可循、在指定的区间内每个数的出现几率相等、无法根据之前的数来预测下一个数的数列.它是产生其它概率分布的随机数的基础和关键.至今应用最为广泛的均匀随机数生成器是基于Lehmer于1951年首先提出的线性同余伪随机数生成器(LCG, Linear Congruential pseudorandom Generator), 其迭代公式为

$$X_{i+1} = a * X_i + c(\text{mod } m) \quad (2-1)$$

其中, $a(0 < a < m)$ 、 $c(0 \leq c < m)$ 和 m 三个参数分别叫做乘数、增量和模; $i = 0$ 时 X_0 称作种子.

若对任意的正整数 i 均有 $X_{i+T} = X_i$, 则最小正整数 T 称为LCG的周期.在一个周期内 T 个模 m 的非负整数的取值是两两不同的, 因此LCG的最大可能的周期必为 m .将任一周期内的 T 点数据对 m 归一化, 即令 $r_j = X_j/m$, 遂可得分布于 $(0, 1)$ 间的均匀随机数序列 $\{r_j\}$.

2.2 任意概率分布的随机数生成

有了均匀概率分布的伪随机数, 就可以通过各种变换及映射关系来得到任意概率分布的伪随机数, 主要的方法有反函数法、变换法和舍选法等.下面分别介绍这三种方法.

2.2.1 反函数法

通过反函数法产生任意分布伪随机数的方法是最常用的方法之一, 其原理是: 已知 $(0,1)$ 区间上均匀分布的伪随机数 r , 将所需的概率分布的伪随机数函数 $F(x)$ 进行反变换, 得到 $F(x)$ 的反函数 F^{-1} , 令 $X = F^{-1}(r)$, 则 X 就是服从概率分布函数为 $F(x)$ 的伪随机数.因此, 只要知道所需概率分布函数的反函数, 就可以从 $(0, 1)$ 均匀分布的伪随机数产生服从所需分布的随机数.

定理1 设 $F(x)$ 是连续且严格单调上升的分布函数, 它的反函数存在, 记为 $F^{-1}(x)$, 即 $F[F^{-1}(x)] = x$.

(1)若随机变量 ξ 的分布函数为 $F(x)$, 则 $F(\xi) \sim U(0, 1)$;

(2)若随机变量 $R \sim U(0, 1)$, 则 $F^{-1}(R)$ 的分布函数为 $F(x)$.

证明 设随机变量 $F(\xi)$ 的分布函数为 $F_1(u)$.

当 $u \in [0, 1]$ 时,

$$F_1(u) = P\{F(\xi) \leq u\} = P\{\xi \leq F^{-1}(u)\} = F[F^{-1}(u)] = u$$

当 $u < 0$ 时, $F_1(u) = 0$; 当 $u > 1$ 时, $F_1(u) = 1$. 所以

$$F(\xi) \sim U(0, 1)$$

设随机变量 $F^{-1}(R)$ 的分布函数为 $F_2(x)$, 则

$$F_2(x) = P\{F^{-1}(R) \leq x\} = P\{R \leq F(x)\} = F_R[F(x)] = F(x)$$

所以 $F^{-1}(R)$ 的分布函数为 $F(x)$.

例1 已知 $r \sim U(0, 1)$, 求服从指数分布的随机变量 X .

解 指数分布函数为:

$$F(x) = 1 - e^{-\lambda x}, x > 0$$

所以, 易求得 $F(x)$ 的反函数为: $F^{-1}(y) = \lambda \ln(1 - y)$, $r \sim U(0, 1)$, 令 $X = F^{-1}(r) = \frac{1}{\lambda} \ln(1 - r)$, 则 X 服从概率分布函数为 $F(x)$ 的指数分布. 即 X 就是所求的随机变量.

2.2.2 变换法

变换法通过一个变换将一个分布的随机数变换成为不同分布产生的随机数, 例如常用的线性变换能够把一个有限区间 $[a, b]$ 上的分布变换到任意实数区间 $[u, v]$ 上. 对于每一个 X 值, 都能够根据下式给出 Y 值

$$Y = \frac{X(u - v)}{b - a} + u \quad (2-2)$$

定理2 设随机变量 X 具有密度函数 $f(x)$, $Y = g(X)$ 是随机变量 X 的函数, 又设 $x = g^{-1}(y) = h(y)$ 存在且有一阶连续导数, 则 $Y = g(X)$ 的密度函数为

$$p(y) = f[h(y)]|h'(y)| \quad (2-3)$$

定理3 设随机向量 (X, Y) 具有二维联合密度 $f(x, y)$, 令

$$\begin{cases} u = g_1(x, y) \\ v = g_2(x, y) \end{cases}$$

设 g_1, g_2 的反变换存在惟一, 记为

$$\begin{cases} x = h_1(u, v) \\ y = h_2(u, v) \end{cases}$$

并设 h_1, h_2 的一阶偏导数存在;函数变换的Jacobi行列式

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \neq 0$$

则随机变量 U, V 的二维联合密度为

$$p(u, v) = f[h_1(u, v), h_2(u, v)]|J| \quad (2-4)$$

如果 R_1, R_2 是两个独立的均匀随机变量, 则 R_1, R_2 的二维函数 $\xi = g_1(R_1, R_2)$ 的随机数可通过二维变换公式由两个均匀随机数产生.

例2 用二维变换抽样法产生标准正态随机数.

解 设 r_1, r_2 为相互独立的均匀分布随机数, 令

$$u = \sqrt{-2\ln(r_1)} \sin(2\pi r_2), \quad v = \sqrt{-2\ln(r_1)} \cos(2\pi r_2) \quad (2-5)$$

则 u, v 相互独立且为 $N(0, 1)$ 随机数.

事实上, 由(2-5)式, 可解出

$$\begin{cases} r_1 = \exp\{-\frac{1}{2}(u^2 + v^2)\} \\ r_2 = \frac{1}{2\pi} \arctg \frac{v}{u} \end{cases}$$

变换的Jacobi行列式为

$$\begin{aligned} J &= \begin{vmatrix} \frac{\partial r_1}{\partial u} & \frac{\partial r_1}{\partial v} \\ \frac{\partial r_2}{\partial u} & \frac{\partial r_2}{\partial v} \end{vmatrix} = \begin{vmatrix} -ur_1 & -vr_1 \\ -\frac{1}{2\pi} \frac{v}{u^2 + v^2} & \frac{1}{2\pi} \frac{u}{u^2 + v^2} \end{vmatrix} \\ &= -\frac{r_1}{2\pi} \frac{u^2}{u^2 + v^2} - \frac{r_1}{2\pi} \frac{v^2}{u^2 + v^2} \\ &= -\frac{1}{2\pi} \exp\{-\frac{1}{2}(u^2 + v^2)\} \end{aligned}$$

由(2-4)式可知随机变量 U, V 的二维联合密度为

$$p(u, v) = f[h_1(u, v), h_2(u, v)]|J| = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$$

显然 U, V 均服从 $N(0, 1)$ 分布, 且相互独立.

上均匀分布的随机变量, 所得的 Y_1, Y_2 是相互独立的服从期望值 $\mu = 0$, 方差 $\sigma^2 = 1$ 的正态分布的随机变量.

2.2.3 舍选法

用反函数法需要知道所求概率分布函数的反函数,当反函数不存在或难以求出时,反函数法便难以使用.这时可以考虑使用舍选法.

舍选法是Von Neuman为克服反函数法和变换法的困难最早提出来的.它的基本思想是:按照给定的分布密度函数 $f(x)$,对均匀分布的随机数序列 $\{r_i\}$ 进行舍选.舍选的原则是在 $f(x)$ 大的地方,保留较多的随机数 r_i ;在 $f(x)$ 小的地方,保留较少的随机数 r_i ,使得到的子样本中 r_i 的分布满足分布密度函数的要求.

生成分布概率密度函数为 $f(x)$ 的伪随机数 X 的步骤如下:

设随机变量 X 的概率密度函数为 $f(x)$,又存在实数 $a < b$,使得 $P(a < X < b) = 1$.

(1)选取常数 λ ,使得 $\lambda f(x) \leq 1, x \in (a, b)$;

(2)产生在 $(0, 1)$ 上均匀分布的随机数 r_1 和 r_2 ,令 $y = a + (b - a)r_1$;

(3)比较 r_2 与 $\lambda f(y)$,若 $r_2 \leq f(y)$,则令 $x = y$;否则剔除 r_1 和 r_2 ,重返步骤(2).

如此重复循环,产生的随机数序列 x_1, x_2, \dots, x_n 的分布由概率密度 $f(x)$ 确定.

若不存在有限区间 (a, b) ,使 $\int_a^b f(x)dx = 1$,可选取有限区间 (a_1, b_1) ,使得 $\int_{a_1}^{b_1} f(x)dx \geq 1 - \varepsilon$,其中, ε 是很小的正数,如取 $a_1 = \mu - 3\sigma, b_1 = \mu + 3\sigma$,有 $\int_{a_1}^{b_1} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx > 1 - 0.003$,对于区间 (a_1, b_1) 应用舍选法,仅会出现较小的系统误差.

2.3 经验分布抽样法

当对实际过程进行模拟计算时,调查得来的数据 x_1, \dots, x_n 来自哪类总体分布 $F(x)$ 是未知的.但由观测数据可求出经验分布函数 $F_n(x)$,直接由经验分布函数或观测数据出发,产生总体 $F(x)$ 随机数.

设已知观测数据 x_1, \dots, x_n 来自某总体,其分布函数为 $F(x)$.将 x_1, \dots, x_n 排序: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, n 个点将 $[x_{(1)}, x_{(n)}]$ 分为 $n-1$ 个小区间,假定数据落入每个小区间的概率均为 $\frac{1}{n-1}$,且每个小区间是均匀分布的.具体抽样如下:

(1)产生 $r \sim U(0, 1)$,记 $p = (n-1)r$,令 $I = [p] + 1$;

(2)令 $x = x_{(I)} + (p - [p])(x_{(I+1)} - x_{(I)})$,则 x 为近似地服从分布函数 $F(x)$ 的随机数.

2.4 随机向量抽样法

设随机向量 $X = (X_1, \dots, X_n)$ 具有联合密度函数 $f(x_1, \dots, x_n)$,若 X 的各分量相互独立.对分量 X_1, \dots, X_n 分别独立的进行抽样.但在实际问题中, X 的各个分量经常是相关的,

我们可采用条件分布抽样方法. 设 (X_1, \dots, X_n) 的联合概率密度为:

$$f(x_1, \dots, x_n) = f_{X_1}(x_1)f(x_2|x_1) \cdots f(x_n|x_1, \dots, x_{n-1}) \quad (2-6)$$

其中 $f_{X_1}(x_1)$ 为 X_1 的边际密度, $f(x_k|x_1, \dots, x_{k-1})$ 为在已知 $X_1 = x_1, \dots, X_{k-1} = x_{k-1}$ 条件下 X_k 条件密度.于是可以先取一个 $f_{X_1}(x)$ 随机数 x_1 ;然后在 x_1 固定的条件下,生成一个 $f(\cdot|x_1)$ 随机数 x_2 ;在 x_1, x_2 固定的条件下,生成一个 $f(\cdot|x_1, x_2)$ 随机数 x_3 ;如此下去,最后,在 x_1, \dots, x_{n-1} 固定的条件下,生成一个 $f(\cdot|x_1, \dots, x_{n-1})$ 随机数 x_n .这样得到的 (x_1, \dots, x_n) 就是随机向量 $X = (X_1, \dots, X_n)$ 的一个随机数.当然在生成各个条件密度的随机数时,仍然可以使用Von Neuman取样原则.

若 $X = (X_1, \dots, X_n)$ 为离散型随机向量,只要将概率函数 $p(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$ 代替密度函数即可.

然而在 n 很大时,更为常用的是Gibbs取样法,它是一种基本的动态Monte Carlo方法,即Markov链Monte Carlo方法.

3 MCMC抽样

3.1 MCMC 方法的基本思想

在Bayes计算中, 我们进行积分运算通常都是需要用到分析或数值逼近的方法, 其中包括基于样本(sample-based) 的各种Monte Carlo 抽样, 如重要抽样、分层抽样、关联抽样等, 这种方法实际上就是从后验分布抽样以估计感兴趣的参数. 但是直接从一个任意的高维联合分布中产生样本常常是比较困难的, 这样使得基于样本的方法具有了局限性. Markov Chain Monte Carlo(MCMC) 方法是最近发展起来的一种简单而行之有效的Bayes 计算方法. 该方法的核心思想就是通过建立一个以 $\pi(x)$ 为平稳分布的Markov链, 对 $\pi(x)$ 进行抽样, 然后基于这些样本做各种统计推断. 比如, 若我们通过抽样得到了 $\pi(x)$ 的样本 $X^{(1)}, \dots, X^{(n)}$, 则

$$E_{\pi}f = \int_D f(x)\pi(x)dx \quad (3-1)$$

其中 D 为状态空间, 于是便可得估计

$$\hat{f}_n = \frac{1}{n} \sum_{i=1}^n f(X^{(i)}) \quad (3-2)$$

这便是Monte Carlo积分. 当 $X^{(1)}, \dots, X^{(n)}$ 独立时, 根据大数定律有:

$$\hat{f}_n \xrightarrow{a.s.} E_{\pi}f, n \rightarrow \infty \quad (3-3)$$

$X^{(1)}, \dots, X^{(n)}$ 是平稳分布为 $\pi(x)$ 的Markov 过程的样本时, (3-3)式也成立.

我们知道, Markov 链是一个随机变量序列 $\{X^{(0)}, X^{(1)} \dots\}$, 在任一时刻 $t(t \geq 0)$, 序列中下一时刻 $t+1$ 处的 $X^{(t+1)}$ 由条件分布 $F(x|X^{(t)})$ 产生, 它只依赖于时刻 t 的状态 $X^{(t)}$, 而与 t 以前的状态 $\{X^{(0)}, X^{(1)} \dots, X^{(t-1)}\}$ 无关. 若该Markov 链满足不可约遍历的条件, 无论初始值 $X^{(0)}$ 取什么, $X^{(t)}$ 的分布都收敛到同一个分布, 即前面所说的平稳分布.

一般地, 令 $X_{t \geq 0}^{(t)}$ 为状态空间 D 上的Markov 链, 其一步转移概率函数为

$$p(x', x) = P(x' \rightarrow x) = P(X^{(t+1)} = x | X^{(t)} = x') \text{ (离散)} \quad (3-4)$$

或

$$P(x' \rightarrow B) = \int_B p(x', x)dx' \text{ (连续)} \quad (3-5)$$

$p(., .)$ 就是该Markov 链的转移核. 通常假定 $p(., .)$ 与 t 无关. t 步转移概率函数为

$$p(t : x', x) = P(X^{(t+s)} = x | X^{(s)} = x') \quad (3-6)$$

记 $X^{(0)}$ 的分布为 $\mu(x) = P(X^{(0)} = x)$, 则经过 t 步后 $X^{(t)}$ 的边际分布记为

$$\mu'(x) = P(X^{(t)} = x) \quad (3-7)$$

如果 $\pi(x)$ 满足

$$\int p(x', x)\pi(x')dx' = \pi(x), \forall x \in D \quad (3-8)$$

则 $\pi(x)$ 就是转移核 $p(.,.)$ 的平稳分布.

作为起始状态, $X^{(0)}$ 最好具有分布 $\pi(x)$, 那么, 由平稳分布的定义, 这就保证了任一 $X^{(t)}$ 的边际分布也是 $\pi(x)$. 然而, 当我们难以直接从 $\pi(x)$ 抽样而需要应用MCMC方法时, 我们并不需要起始状态的边际分布就是城对. 从不同的 $X^{(0)}$ 出发, Markov链经过一段时间的迭代后, 可以认为各个时刻的边际分布都是平稳分布 $\pi(x)$, 即该Markov链收敛了. 而在收敛出现以前的一段时间, 比如 m 次迭代中, 各状态的边际分布还不能认为是 $\pi(x)$. 因而, 我们在应用(3-2)式估计 $E_{\pi}f$ 时, 应把前面的 m 个迭代值去掉, 而用后面的 $n-m$ 个迭代结果来估计. 另外从模拟的角度来看, 我们构造的转移核应该使已知的概率分布 $\pi(x)$ 为平稳分布, 因此, 在应用MCMC方法时, 转移核的构造具有至关重要的作用. 不同的MCMC方法, 其转移核的构造方法是不同的.

MCMC方法的实施步骤概括为如下三步:

- (1) 在状态空间 D 上建立一个以 $\pi(x)$ 为平稳分布, 转移核为 $p(.,.)$ 的Markov链;
- (2) 由 D 中某一点 $X^{(0)}$ 出发, 用(1)步中的Markov链产生点序列 $X^{(1)}, \dots, X^{(n)}$;
- (3) 对某个 m 和足够大的 n , 用下式估计任一函数 $f(x)$

$$\hat{E}_{\pi}f = \frac{1}{n-m} \sum_{i=m+1}^n f(X^{(i)}) \quad (3-9)$$

3.2 Metropolis sampler抽样方法

Monte Carlo方法解决实际问题的最关键一步, 是确定一个随机变量 x 及与其有关的统计量 $g(x)$, 使得 $g(x)$ 的数学期望 $E(g(x)) = \int g(x)f(x)dx$ 正好等于所要求的值, 其 $f(x)$ 为随机变量 x 的密度函数. 密度函数 $f(x)$ 有两个重要条件必须满足, 一个是非负条件, 另一个是归一条件. 这两个约束条件给Monte Carlo方法带来一定问题, 影响了Monte Carlo方法应用领域的进一步扩大.

在统计物理中, 计算物理系综的平均观察量, 是Monte Carlo方法最重要的应用领域之一. 平均观察量的一般形式如下:

$$A = \frac{\int A(x)\pi(x)dx}{\int \pi(x)dx} \quad (3-10)$$

其中 x 表示相空间中的点; $\pi(x)$ 表示物理系综的分布; $A(x)$ 表示 x 的某种观察量.由上式可看出,计算 A 实际上是计算两个积分,一个积分是 $\int A(x)\pi(x)dx$,另一个积分是 $\int \pi(x)dx$.

作为物理系综分布, $\pi(x)$ 满足非负条件,但不满足归一条件,因此,为用一般Monte Carlo方法计算 $\int A(x)\pi(x)dx$,需要对 $\pi(x)$ 进行归一,即用 $\pi(x)/\int \pi(x)dx$ 作为密度函数.然而,由于 $\int \pi(x)dx$ 也是未知的, x 又是高维空间中的点, $\int \pi(x)dx$ 的计算同 $\int A(x)\pi(x)dx$ 的计算相比,其困难程度是同等的,因此,对 $\pi(x)$ 进行归一的办法实际上是行不通的.为解决非归一分布抽样问题,由Metropolis等人于1953年给出了一种抽样方法,这一方法影响了这一问题研究的相当长时间,甚至一直影响到现在.

对于非归一分布 $\pi(x)$,Metropolis抽样方法的一般原理是,确定一个Markov过程,使得它的全概率的极限与分布 $\pi(x)$ 成正比.即确定一个转移函数 $p(x',x)$,满足条件:

$$(1)\pi(x) = \int \pi(x')p(x',x)dx'$$

$$(2)\text{对于任意的分布}\pi_0(x),\text{定义: }\pi_{m+1}(x) = \int \pi_m(x')p(x',x)dx'$$

则有: $\lim_{m \rightarrow \infty} \pi_m(x) = \lambda \pi(x)$,其中 λ 为某一常数.

为抽取分布 $\pi(x)$ 的渐近子样 $\{x_m\}_{m=1}^N = \{x_1, \dots, x_N\}$,对于满足转移核:

$$p_{ij} = \begin{cases} p_{ij}^* \min(1, \frac{\pi_j}{\pi_i}), & j \neq i \\ 1 - \sum_{j \neq i} p_{ij}, & j = i \end{cases} \quad (3-11)$$

其中 $P^* = (p_{ij}^*)$ 是一个对称转移矩阵(预选矩阵).Metropolis抽样方法如下:

(1)由任意分布 $\pi_0(x)$ 中抽样产生 x'_0 ,令 $x_0 = x'_0, m = 0$;

(2)对于确定的 x_m ,由分布 $p^*(x_m, x)$ 中抽样产生 x'_{m+1} ,并用下式确定 x_{m+1} :

$$x_{m+1} = \begin{cases} x'_{m+1}, & \pi(x_m) \leq \pi(x'_{m+1}) \\ x'_{m+1}, & \pi(x_m) > \pi(x'_{m+1}), \xi \leq \frac{\pi(x'_{m+1})}{\pi(x_m)} \\ x_m, & \pi(x_m) > \pi(x'_{m+1}), \xi > \frac{\pi(x'_{m+1})}{\pi(x_m)} \end{cases} \quad (3-12)$$

其中 $\xi \sim U[0, 1]$.

(3)令 $m = m + 1$,当 $m < N$ 时重复(2);当 $m = N$ 时终止.

3.3 Gibbs抽样方法

3.3.1 基本思想

Gibbs方式更新的思想,在用于高维总体或复杂总体的取样时,主要是通过 π 的条件分布族,构造一个不可约正常常返的马氏链 X_n ,使它以 π 为平稳分布.Gibbs抽样方法,将

高维总体化为一系列一维分布进行取样.

设给定一个 m 维联合分布 $\pi(x_1, \dots, x_m)$, 构造如下转移核:

$$p_{xy} = p(x, y) = \prod_{k=1}^m \pi(y_k | y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_m) \quad (3-13)$$

其中 $x = (x_1, \dots, x_m), y = (y_1, \dots, y_m), x_i, y_i \in D, D$ 是高维空间中的一个区域, 而 $\pi(y_k | y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_m)$ 是在除第 k 个分量外, 将第1至第 $k-1$ 个分量固定为 y_1, \dots, y_{k-1} , 并将第 $k+1$ 至 m 个分量固定为 x_{k+1}, \dots, x_m 的条件下, 第 k 个分量在 y_k 处的条件分布. 下证 $p(x, y)$ 确实是一个概率转移矩阵, 即 $\sum_y p(x, y) = 1$.

$$\sum_{y_1} \pi(y_1 | x_2, \dots, x_m) = \sum_{y_1} \frac{\pi(y_1, x_2, \dots, x_m)}{\pi(x_2, \dots, x_m)} = \frac{\pi(x_2, \dots, x_m)}{\pi(x_2, \dots, x_m)} = 1$$

$$\sum_y p(x, y) = \sum_{y_m} \cdots \sum_{y_1} \pi(y_1 | x_2, \dots, x_m) \pi(y_2 | y_1, x_3, \dots, x_m) \cdots \pi(y_m | y_1, \dots, y_{m-1}) = 1$$

下面验证 $\pi(x_1, \dots, x_m)$ 是以 $p(x, y)$ 为转移阵的马氏链的平稳分布.

$$\begin{aligned} \sum_x \pi(x) p(x, y) &= \sum_x \pi(x) \prod_{k=1}^m \pi(y_k | y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_m) \\ &= \sum_{x_2, \dots, x_m} \left[\sum_{x_1} \pi(x_1, \dots, x_m) \right] \frac{\pi(y_1, x_2, \dots, x_m)}{\sum_{x_1} \pi(x_1, \dots, x_m)} \\ &\quad \prod_{k=2}^m \pi(y_k | y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_m) \\ &= \sum_{x_2, \dots, x_m} \pi(y_1, x_2, \dots, x_m) \prod_{k=2}^m \pi(y_k | y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_m) \\ &= \sum_{x_3, \dots, x_m} \left[\sum_{x_2} \pi(y_1, x_2, \dots, x_m) \right] \frac{\pi(y_1, y_2, x_3, \dots, x_m)}{\sum_{x_2} \pi(y_1, x_2, \dots, x_m)} \\ &\quad \prod_{k=3}^m \pi(y_k | y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_m) \\ &= \sum_{x_3, \dots, x_m} \pi(y_1, y_2, x_3, \dots, x_m) \prod_{k=3}^m \pi(y_k | y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_m) \\ &= \cdots = \pi(y_1, \dots, y_m) = \pi(y) \end{aligned}$$

3.3.2 构造Markov链的Gibbs抽样方法

从已得到的Markov链在时刻 n 的样 X_n , 去求转移到时刻 $n+1$ 的样本 X_{n+1} 时, 转移 $p(x, y)$ 是一维的条件分布乘积, 而这些一维条件分布的样本是较为容易生成的, 可按如下步骤逐个得到 X_{n+1} 样本的各个分量:

(1)先得到服从分布 $\{\pi(y_1 | x_2, \dots, x_m) : y_1 \in D\}$ 的随机变量 (记为 $X_{n+1,1}$) 的一个样本 y_1 (在用Von Neumann取舍原则取样时, 对于 $\pi(y_1 | x_2, \dots, x_m) = \frac{\pi(y_1, x_2, \dots, x_m)}{\pi(+\infty, x_2, \dots, x_m)}$, 只

需假定对固定的 (x_2, \dots, x_m) , 存在 $p_0(y_1)$ 及常数 C , 满足 $\pi(y_1, x_2, \dots, x_m) \leq Cp_0(y_1)$, 并对 $U[0, 1]$ 均匀随机数 U , 看 $\frac{\pi(U, x_2, \dots, x_m)}{Cp_0(U)}$ 是否不小于 U , 来决定取舍样本 U ;

(2)用同样的方法, 再得到服从分布 $\{\pi(y_2|y_1, x_3, \dots, x_m) : y_2 \in D\}$ 的随机变量(记为 $X_{n+1,2}$)的一个样本 y_2 ;

(3)依此下去, 得到服从分布 $\{\pi(y_k|y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_m) : y_k \in D\}$ 的随机变量(记为 $X_{n+1,k}$)的一个样本 y_k ;

(4)最后得到服从分布 $\{\pi(y_m|y_1, \dots, y_m) : y_m \in D\}$ 的随机变量(记为 $X_{n+1,m}$)的一个样本 y_m , 则 $y = (y_1, \dots, y_m)$ 就是 X_{n+1} 的一个样本.

我们可任取 $X_0 = y(0)$, 按上面方法得到 X_1, \dots, X_n 的样本 $y(1), \dots, y(n)$. 当 n 充分大时, 由Markov链 X_n 的分布近似于 $\pi(\cdot)$, $y(n)$ 可近似地认为是分布 $\pi(\cdot)$ 的一个样本.

3.4 基于Bayes参数估计的Gibbs抽样方法

依据Bayes思想, 当参数的估计 $\hat{\theta}$ 为后验均值 $\hat{\theta}_E = E(\theta|X)$ 时, 可以使得后验均方差达到最小. 所以, 实际中取后验均值作为 θ 的Bayes估计值. 这里参数 θ 既可以是一元的, 也可以是多元的. 设 X 为观测样本, $\theta = (\theta_1, \theta_2, \dots, \theta_m)$ 为参数向量, 记参数向量 $\theta_{(j)} = (\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_m)$, 密度函数 $\pi(\theta_j|\theta_{(j)}, X)$ 称为参数 $\theta_j(j = 1, \dots, p)$ 的全条件后验分布密度(Full conditional Posterior distributions density). 在推导参数的全条件后验分布时有结论: 对于 N 个观测点 $X_i(i = 1, \dots, N)$, 记 $X = (X_1, \dots, X_N)$, 设 θ 分布中的参数, 给定参数 θ 的先验密度 $\pi(\theta)$, $\pi(X|\theta)$ 为极大似然函数, 后验密度:

$$\pi(\theta|X) = cp(X|\theta)\pi(\theta) \quad (3-14)$$

同时, $\pi(\theta_j|\theta_{(j)}, X)$ 仅与后验密度 $\pi(\theta|X)$ 中与 θ_j 有关的项成比例. 由此可进行Gibbs抽样:

(1)通过联合后验分布, 求出所有参数的全条件后验分布 $\pi(\theta_j|\theta_{(j)}, X)(j = 1, \dots, m)$;

(2)给出任意的一组 θ 的初始值 $\theta_0 = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_m^{(0)})$, 从 $\pi(\theta_1|\theta_2^{(0)}, \dots, \theta_m^{(0)}, X)$ 中抽一个样本 $\theta_1^{(1)}$;

(3)从 $\pi(\theta_2|\theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_m^{(0)}, X)$ 中抽一个样本 $\theta_2^{(1)}$;

(4)如此下去, 从 $\pi(\theta_m|\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{m-1}^{(1)}, X)$ 中抽一个样本 $\theta_m^{(1)}$.

这样的—个完整的步骤称之为一个迭代(iteration). 并用 $\theta^{(1)}$ 表示为生成的向量 $\theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_m^{(1)})$. 经过 n 次迭代后得到 $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n)}$, 形成一个马尔可夫链, 这里 n 通常很大, 一般大于5000次.

在很弱的正则条件下当 $t \rightarrow \infty$ 时, 有 $\theta^{(t)} \xrightarrow{d} \theta \sim \pi(\theta|X)$, $t = 1, 2, \dots, n$, 且 $1/n \sum_{i=1}^n g(\theta^{(i)}) \xrightarrow{a.s.} E_{\theta|X}(g(\theta))$, 当然也有 $\theta_j^{(t)} \xrightarrow{d} \pi(\theta_j|X)$, $1/n \sum_{i=1}^n g(\theta_j^{(i)}) \xrightarrow{a.s.} E_{\theta_j|X}(g(\theta_j))$, 其中 $g(\cdot)$ 为某可测函数. 因此上述 Markov 链达到均衡状态后的样本 $\theta^{(i)}$ 就可作为 $\pi(\theta|X)$ 的样本.

3.5 基于 Bayes 参数估计的 Metropolis 抽样方法

基于数据集 X 参数向量 θ 的贝叶斯推断后验可以借助后验密度 $\pi(\theta|X)$ 得到, 通过贝叶斯原理, 有

$$\pi(\theta|X) = c p(X|\theta) \pi(\theta) \quad (3-15)$$

式中: c 是标准化常数; $p(X|\theta)$ 是以 θ 为条件的似然函数; $\pi(x)$ 是 θ 的先验密度. 贝叶斯方法要求统计推断必须是基于参数的后验分布, 然而直接处理后验是很困难的. 但是, 如果可从后验分布抽取参数向量的样本, 关于参数向量的统计推断就可以使用一般的蒙特卡罗方法实现. MCMC 方法的目的是提供一种从参数后验分布中抽取样本的一种机制. 由于从后验中直接取样是很困难的, 从而可以用 MCMC 方法建立马尔可夫链, 使它的稳定分布和后验分布相同, 当马尔可夫链收敛时, 模拟值可以看作是从后验分布中抽取的样本.

Metropolis 算法从建议密度 $q(\cdot|\theta)$ 生成备选 θ' , 该建议密度必须满足一定的性质. 备选 θ' 以概率 $T(\theta, \theta')$ 被接受, 而接受概率为

$$T(\theta, \theta') = \min \left\{ 1, \frac{\pi(\theta'|x) q(\theta|\theta')}{\pi(\theta|x) q(\theta'|\theta)} \right\} \quad (3-16)$$

算法如下:

- (1) 给定现在状态 $\theta^{(i)}$, 从建议密度 $q(\cdot|\theta^{(i)})$ 生成备选值 θ' ;
- (2) 按照 (3-16) 式计算接受概率 $T(\theta^{(i)}, \theta')$;
- (3) 以概率 $T(\theta^{(i)}, \theta')$ 接受备选值, 即 $\theta^{(i+1)} = \theta'$; 反之, 拒绝备选值, 即 $\theta^{(i+1)} = \theta^{(i)}$;
- (4) 重复前面的步骤, 获得 $\{\theta^{(0)}, \theta^{(1)}, \dots\}$, 剔除前面的 d 个值, 则 $\{\theta^{(d+1)}, \theta^{(d+2)}, \dots\}$ 都具有相同的后验密度 $\pi(\theta|X)$.

Metropolis 算法只需知道 $\pi(\theta)$ 的核即可, 这往往是很方便的. 选取 $q(\theta, \theta')$ 是比较灵活的, 常用 $q(\theta, \theta')$ 为多元正态、多元 t -分布等.

注意 Metropolis 算法在低维时用起来很方便, 但在 θ 维数较高时选择适合的 $q(\theta, \theta')$ 有时是很困难的, 它与 Gibbs 抽样方法结合起来可以优势互补.

4 应用举例

(1) 多重积分的计算

设 k 重积分

$$I_k = \int \cdots \int_D f(x_1, \cdots, x_k) dx_1 \cdots dx_k$$

其中 D 是 k 维积分区域. 设 $g(x_1, \cdots, x_k)$ 是 D 上一个概率密度函数. 令函数:

$$Z(x_1, \cdots, x_k) = \begin{cases} \frac{f(x_1, \cdots, x_k)}{g(x_1, \cdots, x_k)}, & g(x_1, \cdots, x_k) \neq 0 \\ 0, & g(x_1, \cdots, x_k) = 0 \end{cases}$$

k 重积分可改写为:

$$\begin{aligned} I_k &= \int \cdots \int_D Z(x_1, \cdots, x_k) g(x_1, \cdots, x_k) dx_1 \cdots dx_k \\ &= E(Z(X_1, \cdots, X_k)) \end{aligned}$$

即 I_k 是随机变量 $\eta = Z(X_1, \cdots, X_k)$ 的数学期望. 从联合密度为 $g(x_1, \cdots, x_k)$ 的分布中随机抽取 N 个点 $(x_{i1}, \cdots, x_{ik})(i = 1, \cdots, N)$, 并计算 N 个函数值 $\eta_i = Z(x_{i1}, \cdots, x_{ik})(i = 1, \cdots, N)$, 那么其平均为

$$\bar{\eta} = \frac{1}{N} \sum_{i=1}^N Z(x_{i1}, \cdots, x_{ik})$$

上式就是 I_k 的估计公式.

特别, 若选取 $g(x_1, \cdots, x_k)$ 为 D 上的均匀分布:

$$g(x_1, \cdots, x_k) = \begin{cases} \frac{1}{V_D}, & (x_1, \cdots, x_k) \in D \\ 0, & \text{其他} \end{cases}$$

其中 V_D 是区域 D 的体积, 这时有:

$$Z(x_1, \cdots, x_k) = \frac{f(x_1, \cdots, x_k)}{g(x_1, \cdots, x_k)} = V_D f(x_1, \cdots, x_k)$$

首先产生区域 D 上的 k 维均匀随机数 $(r_{i1}, \cdots, r_{ik})(i = 1, \cdots, N)$, 计算函数值 $f(r_{i1}, \cdots, r_{ik})(i = 1, \cdots, N)$, 及 $\frac{1}{N} \sum_{i=1}^N f(r_{i1}, \cdots, r_{ik})$, 则有:

$$I_k = \frac{V_D}{N} \sum_{i=1}^N f(r_{i1}, \cdots, r_{ik})$$

在实际问题中, 积分区域 D 可以是很一般的 k 维区域, 产生 D 上均匀随机及计算体积 V_D 都是一件不易的事情. 处理的办法是取一充分大的一维区间 $[a, b]$, 使得 $D \subset$

$[a, b] \times [a, b] \times \cdots \times [a, b]$, 即长为 $b - a$ 的 k 维正方体区域把 D 包含在其中, 只需产生 k 个在 $[a, b]$ 区间上相互独立的均匀随机数 ξ_1, \dots, ξ_k , 记 $\xi = (\xi_1, \dots, \xi_k)$, 则当在 $\xi \in D$ 的条件下, ξ 在 D 内服从均匀分布.

对 D 内任一子区域 G , 有:

$$\begin{aligned} P(\xi \in G | \xi \in D) &= P(\xi \in G) / P(\xi \in D) \\ &= \frac{V_G}{(b-a)^k} / \frac{V_D}{(b-a)^k} = \frac{V_G}{V_D} \end{aligned}$$

其中 V_G 表示子区域 G 的体积. 于是有:

$$\begin{aligned} E[f(\xi) | \xi \in D] &= \frac{1}{V_D} \int_D \cdots \int_D f(x_1, \dots, x_k) dx_1 \cdots dx_k = \frac{I_k}{V_D} \\ I_k &= V_D E[f(\xi) | \xi \in D] \end{aligned}$$

计算一般区域 D 上多重积分 I_k 的算法如下:

- (1) 赋初值: ξ 落入 D 的次数 $m = 0$, 试验次数 $n = 0$, 并规定试验总次数 N ;
- (2) 产生 k 个相互独立服从 $[a, b]$ 区间上的均匀随机 $\xi = (\xi_1, \dots, \xi_k)$, 置 $n = n + 1$;
- (3) 判断 $n \leq N$ 是否成立: 若成立转(4); 否则停止抽样, 转(5);
- (4) 检验 k 维空间的点 ξ 是否落入区域 D , 若 $\xi \in D$, 置 $m = m + 1$, 并令 $\eta_m = \xi$, 计算 $f(\eta_m)$, 转(2); 否则舍去 ξ , 转(2), 重新产生 k 维均匀随机数;
- (5) 计算 $V_D = \frac{m}{N}(b-a)^k$, $E[f(\xi) | \xi \in D] = \frac{1}{m} \sum_{i=1}^m f(\eta_i)$, 则:

$$I_k = \frac{1}{N}(b-a)^k \sum_{i=1}^m f(\eta_i)$$

用蒙特卡罗方法计算积分值时, 误差的阶数为 $O(\frac{1}{\sqrt{N}})$, 它与多重积分的重数 k 无关, 而用其他数值方法计算多重积分时, 其误差与重数 k 是有关的, 可见当 $k > 3$ 时, 使用蒙特卡罗方法计算多重积分将显示出其独特的优越性.

(2) Gibbs 随机抽样

设 x, y 的联合分布为:

$$p(x, y) = \frac{n!}{(n-x)!x!} y^{x+\alpha-1} (1-y)^{n-x+\beta-1}$$

其中 $x = 0, 1, \dots, n, 0 \leq y \leq 1$.

注意到这儿 x 是离散的, 但 y 是连续的, 它们的联合分布很复杂, 但条件密度很简单. 当把 y 视为固定参数时, x 的条件分布是二项分布, 即 $x|y \sim B(n, y)$, 同理, $y|x \sim \text{Beta}(x+\alpha, n-x+\beta)$. 下面用 Gibbs 方法得到该分布的随机数, 假定 $n = 10, \alpha = 1, \beta = 2$, 初始值 $y_1 = \frac{1}{2}$, 迭代步骤如下:

(1) x_1 通过 $B(n, y_1) = B(10, \frac{1}{2})$ 中抽取, 得 $x_1 = 5$;

(2) y_2 通过 $Beta(x_1 + \alpha, n - x_1 + \beta) = Beta(5 + 1, 10 - 5 + 2)$ 中抽取, 得 $y_2 = 0.593$;

(3) x_2 通过 $B(n, y_2) = B(10, 0.593)$ 中抽取, 得 $x_2 = 3$;

(4) y_3 通过 $Beta(x_2 + \alpha, n - x_2 + \beta) = Beta(4, 9)$ 中抽取, 得 $y_3 = 0.4453$;

(5) x_3 通过 $B(n, y_3) = B(10, 0.4453)$ 中抽取, 得 $x_3 = 7$;

这样得到三个二元随机数组 $(5, 0.5)$, $(3, 0.593)$, $(7, 0.4453)$, 这个过程可以持续下去, 便可得到一条足够长的链. 显然, 这条链的最初的一些值是依赖于初值 y_1 的选择, 但随着链的长度增加, 这种依赖性逐渐消失, 即可认为后面的随机数是符合 $p(x, y)$ 的随机数.

5 结束语

在用蒙特卡罗方法解算问题时, 一般需要这样几个过程: 构造或描述概率过程, 对于本身就具有随机性质的问题, 主要是正确地描述和模拟这个概率过程. 对于本来不是随机性质的确定性问题, 比如计算定积分、解线性方程组、偏微分方程边值问题等; 要用蒙特卡罗方法求解, 就必须事先构造一个人为的概率过程, 它的某些参量正好是所要求问题的解. 即将不具有随机性质的问题, 转化为随机性质的问题. 这构成了蒙特卡罗方法研究与应用上的重要问题之一. 然后建立各种估计量, 使其期望值是所要求解问题的解. 最后根据所构造的概率模型编制计算程序并进行计算, 获得计算结果. 与其他的数值计算方法相比, 蒙特卡罗方法有这样几个优点: (1) 收敛速度与问题维数无关, 换句话说, 要达到同一精度, 用蒙特卡罗方法选取的点数与维数无关; 计算时间仅与维数成比例. 但一般数值方法, 比如在计算多重积分时, 达到同样的误差, 点数与维数的幂次成比例, 即计算量要随维数的幂次方而增加. 这一特性, 决定了对多维问题的适用性. (2) 受问题的条件限制的影响小. (3) 程序结构简单, 在电子计算机上实现蒙特卡罗计算时, 程序结构清晰简单, 便于编制和调试. (4) 具有其他数值计算方法不能替代的作用. 蒙特卡罗方法的弱点是收敛速度慢, 误差大的概率性质. 因此, 已有人将数值方法与蒙特卡罗方法联合起来使用, 克服这种局限性, 取得了一定的效果. 随着计算机技术的飞速发展, 蒙特卡罗方法在不同的科学领域将会发挥更大的作用.

参考文献

- [1] 茆诗松,王静龙,濮晓龙. 高等数理统计[M].北京: 高等教育出版社,1998.
- [2] 钱敏平,龚光鲁.应用随机过程[M].北京: 北京大学出版社,1998.
- [3] 高惠璇.统计计算[M].北京:北京大学出版社,1995.
- [4] 王沫然. Matlab与科学计算[M]. 北京:电子工业出版社,2003:242-243.
- [5] 张栋栋,张德然.概率论思维及其智力品质的培养[J]. 大学数学,2005,21(5):103-108.
- [6] 林正炎,苏中根,张立新.当前概率学科中的研究机遇[J].数学进展,2004,133(2):1-6.
- [7] 黎锁平.运用蒙特卡罗方法求解随机性问题[J].甘肃工业大学学报(自然科学版),2001,27(2):95-97.
- [8] 刘忠,茆诗松.分组数据的Bayes分析—Gibbs抽样方法[J].应用概率统计, 1997,13(2):211-216.
- [9] Fang K T, Wang Y. Number Theoretic Methods in Statistics[M]. Chapman & Hall, New York, 1994. Halton J H. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals[J]. Number. Math. 1960, 2: 84-90.
- [10] 张孝令, 刘福升等. 贝叶斯动态模型及其预测[M]. 山东科学技术出版社, 1992.
- [11] Engle R F, Patton A J. What good is a volatility model[J]. Quantitative Finance, 2001, 1: 237-245.
- [12] Meyer R, Yu J. Bugs for a Bayesian analysis of stochastic volatility models[J]. Econometrics Journal, 2000, 3(2): 198-215.
- [13] Bai X, Russell J R, Tiao G C. Kurtosis of GARCH and stochastic volatility models with non-normal innovations[J]. Journal of Econometrics, 2003, 114(2): 349-360.
- [14] Liesenfeld R, Jung R C. Stochastic volatility models: Conditional normality versus heavy-tailed distributions[J]. Journal of Applied Econometrics, 2000, 15(2): 137-160.
- [15] Engle R. New frontiers for ARCH models[J]. Journal of Applied Econometrics, 2002, 17(5): 425-446.
- [16] Lu Huitian, Kolarik William J, Lu Susan S. Real-time performance reliability prediction. IEEE Transactions on Reliability, 2001, 50(4): 353-357.
- [17] Gebraeel Nagi Z, Lawley Mark A, Li Rong, et al. Residual life distributions from component degradation signals: a Bayesian approach. IIE Transactions, 2005, 37(6): 543-557.

- [18] Meeker William Q, Escobar Luis A, Lu C Joseph. Accelerated degradation tests: modeling and analysis. *Technometrics*, 1998, 40(2): 89-99.
- [19] 茆诗松, 王静龙, 濮晓龙. 高等数理统计. 北京: 高等教育出版社, 1998.
- [20] Lu C Joseph, Meeker William Q. Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, 1993, 35(2): 161-174.
- [21] Lindley D V, Smith A F M. Bayes estimation for the linear model [J]. *Journal of the Royal Statistical Society Series B*(1369-7412), 1972, 34: 1-41.
- [22] 韩明. 多层先验分布的构造及其应用[J]. *运筹与管理*, 1997, 6(3): 31-40.
- [23] 乔世君, 张世英. 用Gibbs 抽样算法计算定数截尾时Weibull 分布的贝叶斯估计[J]. *数理统计与管理*, 2000, 19(2): 35-40.
- [24] 汤银才, 费鹤良. 基于Gibbs 抽样的Weibull 分布序进应力加速寿命试验的Bayes 分析[J]. *数理统计与应用概率*, 1998, 13(1): 81-88.
- [25] Zongwu Cai, Jianqing Fan and Qiwei Yao, Functional-coefficient regression models for nonlinear times series, *JASA*, 95(451)(2000), 94 1-956.
- [26] Chin-Tsang Chiang, Rice, J.A. and Wu, C.O., Smoothing spline estimation for varying coefficient models with repeatedly measure dependent variables, *JASA*, 96(454)(2001), 605-619.
- [27] Colin O. Wu, Chin-Tsang Chiang and Donald R. Hoover, Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data, *JASA*, 93(444)(1998), 1388-1403.
- [28] Fan, J. and Zhang, J. T., Two-step estimation of functional linear models with applications to longitudinal data, *J. R. Statist. Soc. B*, 62 (2000), 303-322.
- [29] Green, P. J., Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, 82(1995), 711-732.
- [30] Kotz S, 吴喜之. 现代贝叶斯统计学[M]. 北京: 中国统计出版社, 2000.

致 谢

在整个论文写作过程中,我的导师张绍义教授给予了悉心指导.并且张老师在百忙之中仔细审阅和校对了全稿,提出了许多宝贵的意见.张老师严谨的治学态度和一丝不苟的精神将使我终身受益.在此谨向张老师表示诚挚的感谢!