

湖北大学

硕士学位论文

马尔可夫链蒙特卡罗算法

姓名：田菲

申请学位级别：硕士

专业：概率论与数理统计

指导教师：张绍义

20070601

摘 要

本文在对 MCMC 算法的起源, 应用以及其相关的基本问题 (随机样本生成法, 静态 Monte Carlo 算法) 做了介绍后, 一方面讨论了 MCMC 算法的构造方法, 另一方面讨论了 Markov 链的定性收敛和定量收敛, 并用耦合构造的方法证明了其中的一些结论, 得到了若干有意义的的结果.

本文第一部分介绍了 MCMC 算法的起源及应用范围, 讨论了若干种随机样本生成法和静态 Monte Carlo 算法, 第二部分介绍了 MCMC 算法的构造方法, 专门讨论了 Metropolis-Hasting 采样法, 第三部分讨论了 Markov 链收敛性几种方法, 包括定量收敛和定性收敛. 第四部分着重讨论了 Markov 链定量收敛的一个推广的结论及证明, 得出了若干结果, 具有较好的理论及现实指导意义.

关键词: MCMC, 蒙特卡罗, 马尔可夫链, Metropolis-Hasting 采样法, 定量收敛, 耦合构造.

Abstract

This paper surveys various results about Markov Chains Monte Carlo algorithms. It began with an introduction to Markov Chain Monte Carlo (MCMC) algorithms, which provided the motivation and context for the following theory. Then, the sufficient conditions for geometric and uniform ergodicity were presented, along with quantitative bounds on the rate of convergence to stationarity. We obtained some important results.

The first part of this paper introduced the motivation and context of MCMC, discussed some sample algorithms and classical Monte Carlo algorithms. The second part introduced the constructing of MCMC algorithms. The third part discussed some methods about Markov Chain convergence times. The last part discussed the quantitative convergence rates, proved the convergence using coupling constructions, and obtained some important results.

Key Words: MCMC, Monte Carlo, Markov Chain, Metropolis-Hasting Algorithm, Quantitative Convergence, Coupling Constructions.

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：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。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。本声明的法律后果由本人承担。

论文作者签名：田菲

签名日期：2007年5月28日

学位论文使用授权说明

本人完全了解湖北大学关于收集、保存、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；学校有权保留学位论文的印刷本和电子版，并提供目录检索与阅览服务；学校可以采用影印、缩印、数字化或其它复制手段保存论文；在不以赢利为目的前提下，学校可以公布论文的部分或全部内容。（保密论文在解密后遵守此规定）

论文作者签名：田菲

签名日期：2007年5月28日

导师签名：张维文

签名日期：2007年5月28日

第一章 序言

1899 年, 瑞利等人最早提出基于统计概念的计算方法 - 蒙特卡诺方法的思想, 二十世纪二十年代柯朗 (德)、冯·诺伊曼 (美) 等人发展了这个方法, 后在电子计算机上获得广泛应用. 1906 年俄国数学家马尔可夫, 首次提出了“马尔可夫链”的数学模型.

本世纪 90 年代以来, 对“复杂性”的研究较为瞩目, 很多应用问题都存在着分析对象比较复杂与正确识别模型结构的困难, 这时用 Markov 链的样本, 来对不变分布, Gibbs 分布, Gibbs 场, 高维分布或样本空间非常大的离散分布等做采样, 并用以做随机模拟的方法, 即 MCMC 方法的问世对建立可实际应用的统计模型开辟了广阔的前景. 根据 MCMC 理论, 通过使用专用统计软件进行 MCMC 模拟, 可解决许多复杂性问题, 它相对于决定性算法, 显示出了其巨大的优越性. 此外, 得益于 MCMC 理论的运用, 使得贝叶斯 (Bayes) 统计得到了再度复兴, 以往被认为不可能实施计算的统计方法变得是很轻而易举了.

在许多很复杂的统计问题中, 有时很难对各种统计方法进行理论分析, 为了评估它们的优劣, 常见的实用方法是做随机模拟: 即设法按问题的要求和条件去构造出一系列的模拟样本, 用它们的样本频率代替相应的概率作统计分析和推断, 观察由这些模拟样品所作出的推断的正确率. 因为在概率论初期发展时, 随机模拟的原型常常来自博采, 于是人们就以博采之都 Monte Carlo 作为随机模拟方法的别称. 久而久之, Monte Carlo 方法作为名称倒比随机模拟方法更为广泛的常用了.

用 Markov 链的样本, 来对不变分布, Gibbs 分布, Gibbs 场, 高维分布或样本空间非常大的离散分布等作采样, 并用以作随机模拟的方法, 统称为 Markov Chain Monte Carlo (MCMC) 方法. 这是动态的 Monte Carlo 方法. 由于这种方法的问世, 使随机模拟在很多领域的计算中, 相对于决定性算法, 显示出它的巨大优越性. 而有时随机模拟与决定性算法的结合使用, 会显示出更多的长处.

MCMC 至少可以用在以下几个层面:

(1) 用于生成较复杂的随机数:

- ①. 实现对高维分布 (或高维格点分布) π 的取样, 得到 π 随机数.
- ②. 是实现重要度采样的一种方法. 对 $|f(x)|$ 的重要度采样, 就是取得随机数

$$\pi = \pi(x) = \frac{|f(x)|}{\int |f(y)|dy}.$$

对于 $f(x) \geq 0$, 作以 $\pi = \pi(x) = \frac{|f(x)|}{\int |f(y)|dy}$ 为极限分布的 Markov 链 X_n , 利用遍历定理可以由这个 Markov 链的一条轨道, 得到分布密度 $\pi(x)$ 的估计.

(2) 实现高维积分 (或项数极多的求和) 的数值计算 (典例是 Gibbs 分布的各种泛函的平均值的计算).

(3) 用模拟方法估计最可几轨道. 例如, 如果模拟了 100 条轨道, 那么就能以大概率推断, 最可几轨道就在这些轨道的附近. 当统计的分布未知时, 可以用模拟方法从频率估计置信限.

(4) 用被估计参数的 Bayes 分布的取样, 来估计参数.

(5) 求复杂样本空间上函数的极值 (模拟退火).

鉴于以上情况, 本文第一部分介绍了 MCMC 算法的起源及应用范围, 讨论了若干种随机样本生成法和静态 Monte Carlo 算法, 第二部分介绍了 MCMC 算法的构造方法, 专门讨论了 Metropolis-Hasting 采样法, 第三部分讨论了 Markov 链收敛性几种方法, 包括定量收敛和定性收敛. 第四部分着重讨论了 Markov 链定量收敛的一个推广的结论及证明.

第二章 随机样本生成法

随机变量（或随机向量）的样本简称为随机数。由于在统计中常用的是独立样本列，不妨假设随机数之间都是独立的，生成随机数的方法，也称为随机数的取样法。

2.1 随机数的生成方法

2.1.1 均匀随机变量的计算机模拟

在 $[0, 1]$ 上均匀分布的随机变量的独立样本称为均匀随机数 ($U(0, 1)$ 随机数)。

在计算机上产生的称之为“伪随机数”的数列，是一种具有非常长周期的，且能通过数理统计中的独立性与均匀性假设检验的数列。实践证明，伪随机数是均匀随机数的一种可行的近似。这种伪随机数虽然不是独立同分布的 $U(0, 1)$ 随机变量的样本，而是在 $[0, 1]$ 中取值的周期数列，但是由于它可以像均匀随机数一样地通过数理统计中的独立性与均匀性假设检验，而且它的周期非常长，以至在计算机实际运算过程中不会出现重复，所以在实际计算中它能很好的替代均匀随机数。最普遍用以产生伪随机数的方法是同余法。典型的例子如下：

$$y_{n+1} = 5^{13} y_n (\text{mod } 2^{36}), y_0 = 1, x_n = 2^{-36} y_n \quad (\text{周期约为 } 2 \times 10^{10}.)$$

2.1.2 分布函数 $F(x)$ 的随机数

(反函数法) 分布函数为 $F(x)$ 的独立随机变量列的样本，称为 $F(x)$ 随机数。若 $F(x)$ 严格单调增加， ξ 是均匀随机数，则 $F^{-1}(\xi)$ 是 $F(x)$ 随机数，其中 F^{-1} 为 F 的反函数。

2.1.3 正态随机数

$N(0, 1)$ 随机数称为标准正态随机数。生成标准正态随机数有一个比反函数的方法更为简单的实践方法，就是利用中心极限定理。设 η_1, \dots, η_{12} 为均匀随机

数,且它们是独立的,由中心极限定理,可以认为 $\xi = \eta_1 + \dots + \eta_{12} - 6 \approx N(0, 1)$, 即用 $\xi = \eta_1 + \dots + \eta_{12} - 6$ 近似的作为标准正态随机数. 在实际计算中 $\eta_i (1 \leq i \leq 12)$ 还应该用伪随机数代替.

命题 2.1(生成标准正态随机数的 Box-Muller 方法), 取两个独立的均匀随机数 η_1, η_2 , 令

$$\xi_1 = \sqrt{-2 \ln \eta_1} \cos(2\pi\eta_2).$$

$$\xi_2 = \sqrt{-2 \ln \eta_1} \sin(2\pi\eta_2).$$

则 ξ_1, ξ_2 为相互独立的标准正态随机数.

2. 1. 4 Von Neuman 取舍原则

假定我们要生成密度为 $p(x)$ 的随机数. 为此取一个参考分布密度 $p_0(x)$, 使它满足:

(1) $p_0(x)$ 随机数容易生成, 例如 $p_0(x)$ 为正态密度, 均匀密度, 指数密度, 及它们的混合密度等;

(2) $p_0(x)$ 与 $p(x)$ 的取值范围差不多, 且存在 C , 使 $p(x) \leq Cp_0(x)$.

则有以下命题:

命题 2.2 设随机变量 η 具有密度 $p_0(x)$, 而随机变量 $U \sim U[0, 1]$ 且与 η 独立, 则

$$P(\eta \leq x \mid \frac{p(\eta)}{Cp_0(\eta)} \geq U) = \int_{-\infty}^x p(v)dv.$$

证明 对 η 的取值用推广了的全概率公式 ($P(A) = \int P(A|\eta = y)p_0(y)dy$), 得到

$$\begin{aligned} \text{左边} &= \frac{P(\eta \leq x \mid \frac{p(\eta)}{Cp_0(\eta)} \geq U)}{P(\frac{p(\eta)}{Cp_0(\eta)} \geq U)} \\ &= \frac{\int_{-\infty}^x P(U \leq \frac{p(y)}{Cp_0(y)})p_0(y)dy}{\int_{-\infty}^{\infty} P(U \leq \frac{p(y)}{Cp_0(y)})p_0(y)dy} \\ &= \frac{\int_{-\infty}^x \frac{1}{C}p(y)dy}{\int_{-\infty}^{\infty} \frac{1}{C}p(y)dy} = \text{右边}. \end{aligned}$$

取舍原则的具体作法是:

(1) 独立的生成 n 个独立的 $p_0(x)$ 随机数 η_1, \dots, η_n 与 n 个与之独立的 $U[0, 1]$ 随机数 U_1, \dots, U_n .

(2) 对于 $i = 1, 2, \dots$, 如果有 $\frac{p(\eta_i)}{C p_0(\eta_i)} \geq U_i$, 就保留 η_i , 否则舍弃 η_i .

由命题 2.2, 所有这样保留下来的 η_i 就成为一系列独立的 $p(x)$ 随机数 (当然个数比 n 小很多). 这种取舍方法称为 Von Neuman 取舍原则.

取舍原则可以改良为以下叙述:

如果 $p(x) = \gamma h(x)$, 只要存在 C , 使 $h(x) \leq C p_0(x)$, 那么我们可以在取舍原则中用 $h(x)$ 代替 $p(x)$, 得到 $p(x)$ 随机数. 具体为: 独立的生成 n 个独立的 $p_0(x)$ 随机数 η_1, \dots, η_n 与 n 个与之独立的独立 $U[0, 1]$ 随机数 U_1, \dots, U_n , 如果

$$\frac{p(\eta_i)}{C p_0(\eta_i)} \geq U_i.$$

就保留 η_i , 否则舍弃 η_i , 那么所有保留的是相互独立的 $p(x)$ 随机数.

2.1.5 多维随机数

对于已知的分布密度, 可以利用条件密度, 把生成多维随机数归结为生成一系列一维随机数: 设随机向量 (X_1, \dots, X_d) 的密度为 $f(x_1, x_2, \dots, x_d)$, 则有表达式

$$f(x_1, x_2, \dots, x_d) = f_{X_1}(x_1) f(x_2|x_1) \dots f(x_d|x_1, \dots, x_{d-1}).$$

其中 $f_{X_1}(x_1)$ 为 X_1 的边缘密度, $f(x_k|x_1, \dots, x_{k-1})$ 为在已知 $X_1 = x_1, \dots, X_{k-1} = x_{k-1}$ 条件下 X_k 的条件密度. 于是可以先取一个 f_{X_1} 随机数 x_1 ; 然后, 在 x_1 固定的情形下, 生成一个 $f(\cdot, x_1)$ 随机数 x_2 ; 再在 x_1, x_2 固定的情形下, 生成一个 $f(\cdot, x_1, x_2)$ 随机数 x_3 ; 最后, 在 x_1, x_2, \dots, x_{d-1} 固定的情形下, 生成一个 $f(\cdot, x_1, \dots, x_{d-1})$ 随机数 x_d . 这样得到的 (x_1, \dots, x_d) 就是向量 (X_1, \dots, X_d) 的一个随机数.

2.2 静态的 Monte Carlo 方法

通过构造独立同分布随机数, 计算积分的 Monte Carlo 方法, 称为静态 Monte Carlo 方法. 其思想可以在本节中, 通过估计最简单的积分 $\int_a^b f(x) dx$ 得到阐明. 对于高维积分, 其思路与一维积分是一样的.

2.2.1 用频率估计概率来计算积分的 Monte Carlo 方法

假定 $0 \leq f(x) \leq M$, 那么由积分的面积含义有

$$\int_a^b f(x)dx = |S| \quad (\text{其中 } |S| \text{ 为 } S = (x, y) : a \leq x \leq b, 0 \leq y \leq f(x) \text{ 的面积}).$$

考虑平面区域 $\Omega = [a, b] \times [0, M]$ 上的均匀随机变量 ξ , 则

$$p = P(\xi \in S) = \frac{1}{(b-a)M} \int_a^b f(x)dx.$$

对于 N 个独立的 Ω -均匀随机数 $\xi_i (i \leq N)$, 记

$$N_s = \xi_1, \dots, \xi_N \text{ 中落在 } S \text{ 中的频数}.$$

于是, 利用大数定理便知

$$\hat{I} = (b-a)M \frac{N_s}{N}.$$

是积分 $I = \int_a^b f(x)dx$ 的相合估计, 即对于任意的 $\varepsilon > 0$, 当 $N \rightarrow \infty$ 时, 有

$$P(|(b-a)M \frac{N_s}{N} - \int_a^b f(x)dx| > \varepsilon) \rightarrow 0.$$

又由于 N_s 服从参数为 (p, N) 的二项分布, 所以有

$$E\hat{I} = \frac{(b-a)M}{N} NP = \int_a^b f(x)dx.$$

即 \hat{I} 是积分 $\int_a^b f(x)dx$ 的无偏估计. 此估计的方差为

$$\begin{aligned} \text{var}\hat{I} &= (b-a)^2 M^2 \frac{p(1-p)}{N} \\ &= (b-a)^2 M^2 \frac{\frac{I}{(b-a)M} (1 - \frac{I}{(b-a)M})}{N} \\ &= \frac{1}{N} [(b-a)M - I] \\ &= o(\frac{1}{N}) \end{aligned}$$

又由于方差代表平均平方误差, 故有积分的估计 \hat{I} 的误差为 $o(\frac{1}{N})$.

2.2.2 用样本函数的平均值估计的期望来计算积分的 Monte Carlo 方法

——期望法

期望法的核心思想是把积分看成某个随机变量的期望. 最常见的是看成 $[a, b]$ 上均匀随机变量的期望. 设 $\eta \sim U[a, b]([a, b]$ 上的均匀分布).

$$I = \int_a^b f(x)dx = (b-a)Ef(\eta)$$

于是对 N 个独立的 $[a, b]$ 上的均匀随机数, 可以用矩估计

$$\hat{I} = (b-a) \frac{f(\eta_1) + \dots + f(\eta_N)}{N}$$

作为 $I = \int_a^b f(x)dx$ 的估计, 由于

$$E\hat{I} = \frac{(b-a)NEf(\eta)}{N} = (b-a)Ef(\eta) = I$$

故 \hat{I} 是无偏估计. 而且

$$\begin{aligned} \text{var}(\hat{I}) &= (b-a)^2 \frac{\text{var}[f(\eta)]}{N} \\ &= (b-a)^2 \frac{1}{N} Ef(\eta)^2 - [Ef(\eta)]^2 \\ &= (b-a)^2 \frac{1}{N} \int_a^b f(x)^2 \frac{dx}{b-a} - \frac{1}{N} I^2 \\ &\leq \frac{1}{N} [(b-a)MI - I^2] \\ &= \text{var}(\hat{I}) \end{aligned}$$

可见频率法比期望法更有效.

2. 2. 3 减少方差的技术 -- 重要度采样法

用 Monte Carlo 方法计算积分 $\int_a^b f(x)dx$ 时, 未必一定要使用均匀随机数. 事实上, 从 $[a, b]$ 上取值的任意一种随机数出发, 都可以得到 $\int_a^b f(x)dx$ 的估计量. 而且在 $f(x) \geq 0$, 显见值 $f(x)$ 大的 x 对于积分 $\int_a^b f(x)dx$ 有更大的贡献, 由此得到启发, 所用的随机数的分布密度的形状越像 $f(x)$, 则越合理. 这个思想就是重要度采样法.

1. g- 采样法

假定分布密度 $g(x)$ 在 $f(x)$ 非零处恒正, 则积分

$$I = \int_a^b f(x)dx = \int_a^b \frac{f(x)}{g(x)} g(x)dx.$$

对于密度为 $g(x)$ 的 “g- 随机数” ξ , 有

$$I = E\left[\frac{f(\xi)}{g(\xi)}\right].$$

于是对于 N 个独立的 g- 随机数 ξ_1, \dots, ξ_N , 关于积分 $I = \int_a^b f(x)dx$ 可取估计

$$\hat{I}^{(g)} = \frac{1}{N} \left[E\left[\frac{f(\xi_1)}{g(\xi_1)}\right] + \dots + E\left[\frac{f(\xi_N)}{g(\xi_N)}\right] \right]$$

显见它也是无偏的相合估计. 利用 Schwartz 不等式得到

$$\begin{aligned} & \int_a^b \left[\frac{f(x)}{g(x)} \right]^2 g(x) dx \\ &= \int_a^b g(x) dx \int_a^b \left[\frac{f(x)}{g(x)} \right]^2 g(x) dx \\ &\geq \left[\int_a^b \frac{f(x)}{\sqrt{g(x)}} \sqrt{g(x)} dx \right]^2 \\ &= \left[\int_a^b f(x) dx \right]^2. \end{aligned}$$

而且上式当且仅当在 $\sqrt{g(x)} = c \frac{f(x)}{\sqrt{g(x)}}$ 时 (即 $g(x) = cf(x)$ 时) 达到极小值 $[\int_a^b f(x) dx]^2$. 这说明 $\text{var}(\hat{f}^{(g)}) = \frac{1}{N} [\int_a^b \left[\frac{f(x)}{g(x)} \right]^2 g(x) dx - I^2]$ 的最小值在 $g_0(x) = cf(x)$ 时取到. 又因为 $g_0(x)$ 为密度. 故 $c = \frac{1}{\int_a^b f(x) dx}$ 是无估计误差的精确值.

综上所述可知, 要使方差达到最小, 就应该用 $g_0(x) = cf(x)$ 作为参考密度. 由此我们可以得到下面的认识, 即只要密度 $g(x)$ 的形状与被积分的函数近似, 用 $\hat{f}^{(g)}$ 作为 $\int_a^b f(x) dx$ 的估计, 就会降低方差. 这就是下面的概念.

定义 2.1 分布密度为 $g(x) = \frac{|f(x)|}{\int_a^b |f(x)| dx}$ 的 g - 采样, 称为关于 $f(x)$ 的重要度采样.

重要度采样不能直接通过取舍原则实现. 近似的实现重要度采样可以采用 Markov 链 Monte Carlo 方法.

在实践中人们也往往按照重要度采样的思路, 灵活的寻找常用的已知类型的密度 g , 使它在峰值附近与 $|f(x)|$ 较接近, 以便达到降低估计的方差的目的.

2. 修正的重要度采样法

对于 g - 采样. 假定存在非负函数 $h(x)$, 满足

$$\alpha = \int_a^b h(x) g(x) dx > 0.$$

而且 α 已知, 那么我们可以采用 $h(x)$ 作为修正乘积因子. 显见对于 g - 随机数 ξ 有

$$I = \int_a^b f(x) dx = \frac{\alpha E\left[\frac{f(\xi)}{g(\xi)}\right]}{Eh(\xi)}.$$

如果先放弃对于估计的无偏性要求, 而只要求估计的相合性, 则对于 N 个独立的 g - 随机数 ξ_1, \dots, ξ_N , 我们可以通过比值, 构造 $I = \int_a^b f(x) dx$ 的如下的估

计量

$$\bar{I} = \alpha \frac{\frac{f(\xi_1)}{g(\xi_1)} + \dots + \frac{f(\xi_N)}{g(\xi_N)}}{h(\xi_1) + \dots + h(\xi_N)}.$$

显见它是 $I = \int_a^b f(x)dx$ 的相合估计. 在某些假定下, 它是渐进无偏的, 即

$$\lim_{N \rightarrow \infty} E(\bar{I}) = I.$$

而且 \bar{I} 保留了重要度采样的特性, 即当 $h(x) = c \frac{f(x)}{g(x)}$ 时, \bar{I} 就是 $I = \int_a^b f(x)dx$. 于是, 只要当 $h(x)$ 与 $\frac{f(x)}{g(x)}$ 近似, 就会降低方差. 注意对于给定的分布密度 $g(x)$, $h(x)$ 应选取的尽量与 $\frac{f(x)}{g(x)}$ 近似. 这个修正乘积因子 $h(x)$ 是用来再一次降低由于密度 $g(x)$ 与被积函数 $f(x)$ 的倍数不够像所带来的失误而设置的. 而当 $h(x) \equiv 1$ 时, 就退化为 g - 采样, 这相当于对 g - 采样不再做修正.

第三章 MCMC 算法的构造

MCMC(Markov Chain Monte Carlo 算法) 是一种简单有效的计算方法, 在统计物理, Bayes 统计计算, 显著性检验, 极大似然估计等领域有着广泛的应用. MCMC 的基本思路是: 通过建立一个平稳分布为 $\pi(x)$ 的 Markov 链来得到 $\pi(x)$ 的样本, 基于这样就可以做各种统计推断.

3.1 问题的提出

对一给定的状态空间 χ , 给定的密度函数 π_μ , 满足 $0 < \int_\chi \pi_\mu < \infty$ (一般情况下 χ 是 R^d 上的一个开子集, π_μ 是 Lebesgue 测度), 由密度函数 π_μ , 可给出 χ 上的一概率测度 $\pi(\cdot)$, 即

$$\pi(A) = \frac{\int_A \pi_\mu(x) dx}{\int_\chi \pi_\mu(x) dx} \quad (3.1).$$

又定义函数 $f: \chi \rightarrow R$ 关于 $\pi(\cdot)$ 的期望为:

$$\pi(f) = E_\pi[f(x)] = \frac{\int_\chi f(x) \pi_\mu(x) dx}{\int_\chi \pi_\mu(x) dx}. \quad (3.2)$$

但若 χ 是高维的, 且 π_μ 是一复合函数, 则 (3.2) 式的直接积分法是不可行的.

对高维积分问题, 经典的 Monte Carlo 算法是去模拟 *i.i.d* 的随机变量 $Z_1, Z_2, \dots, Z_N \sim \pi$, 用

$$\hat{\pi} = \left(\frac{1}{N}\right) \sum_{i=1}^N f(Z_i). \quad (3.3)$$

来估计 $\pi(f)$. 由于

$$E[\hat{\pi}(f)] = \frac{1}{N} N E[f(x)] = E_\pi[f(X)] = \pi(f).$$

故 $\hat{\pi}(f)$ 是无偏估计. 又

$$\text{var} \hat{\pi}(f) = \frac{1}{N^2} [N \text{var} f(x)] = \frac{1}{N} [\pi(f^2) - [\pi(f)]^2] \sim O\left(\frac{1}{N}\right).$$

若 $\pi(f^2) < \infty$, 则由经典的中心极限定理, $\hat{\pi}(f) - \pi(f)$ 的误差为平凡的极限分布. 但问题是, 若 π_μ 是复合的, 将非常难去估计 $\pi(\cdot)$ 的 *i.i.d* 的随机变量.

MCMC 解决这一问题的方法是: 构造 χ 上一个以 $\pi(\cdot)$ 为平稳分布的 Markov 链. 为此, 首先定义一个 Markov 链以 $P(x, dy)((x, y) \in \chi)$ 为转移概率的平稳分布.

定义 3.1 平稳分布 一个链的若满足性质: 对任意的 $n \geq 0$, 任意的 k , 有 $\{\phi_0, \dots, \phi_k\}$ 与 $\{\phi_n, \dots, \phi_{n+k}\}$ 的分布相同, 则称该链的分布为平稳分布.

注: χ 上一 σ -有限测度 π 若满足

$$\pi P(dy) = \int_{x \in \chi} \pi(dx) P(x, dy) = \pi(dy). \quad (3.4)$$

则称 π 为不变测度. 一个以 $\pi(\cdot)$ 为平稳分布的 Markov 链一定含有不变测度 $\pi(\cdot)$.

这样, 如果我们对 Markov 链运行足够长的时间 (以任意点为起点), 对足够大的 n , X_n 的分布将会近似的趋向平稳: $L(X_n) \approx \pi(\cdot)$, 此时可令 $Z_1 = X_n$, 再重新运行 Markov 链, 可依次得到 Z_2, Z_3, \dots 即可由 (3.3) 式得到 $\pi(f)$ 的估计 $\hat{\pi}(f)$.

表面看起来似乎很难找到这样的 Markov 链, 然后直接得出 $\pi(f)$ 的估计. 但在事实上, 在后面的介绍中我们可以看到 Markov 链的构造通常是出乎意料的容易的.

3.2 Metropolis-Hasting 采样法

构造 MCMC 的方法有许多, 这里我主要介绍 Metropolis-Hasting 采样法.

基本思路:

任意选择一个不可约的转移概率 $q(x, y)$ 以及一个转移概率 $\alpha(x, y)$ ($0 < \alpha(x, y) < 1$), 对任一组合 (x, x') , 定义:

$$\begin{aligned} p(x, x') &= q(x, x')\alpha(x, x') & x \neq x'. \\ p(x, x') &= 1 - \int_{x \neq x'} q(x, x')\alpha(x, x')dx' & x = x'. \end{aligned}$$

易见 $p(x, x')$ 构成一个概率转移核.

此方法的实施比较直观, 如果链在时刻 t 处于状态 x , 即 $X_t = x$, 则首先由 $q(\cdot|x)$ 产生一个潜在的转移 $x \rightarrow x'$, 然后以概率 $\alpha(x, x')$ 接受 x' 作为链下一时刻的状态值, 而以概率 $1 - \alpha(x, x')$ 拒绝转移到 x' , 从而链在下一时刻仍然处于状态 x .

我们的目标是使 $\pi(x)$ 成为马氏链的平稳分布, 下面就介绍在给定 $q(x,y)$ 后, 如何选择 $\alpha(x,y)$.

一个常用的选择:

$$\alpha(x, x') = \min\{1, \frac{\pi(x')q(x', x)}{\pi(x)q(x, x')}\}.$$

此时有

$$p(x, x') = \begin{cases} q(x, x') & \pi(x')q(x', x) \geq \pi(x)q(x, x'); \\ q(x, x') \frac{\pi(x')}{\pi(x)} & \pi(x')q(x', x) < \pi(x)q(x, x'). \end{cases}$$

定理 3.1 由上述过程产生的 Markov 链是可逆的, 即

$$\pi(x')p(x', x) = \pi(x)p(x, x'). \quad (3.5)$$

且 $\pi(x)$ 是 Markov 链的平稳分布.

证明 若 $x = x'$, 则上式显然成立. 下面设 $x \neq x'$, 则

$$\begin{aligned} & \pi(x)p(x, x') \\ &= \pi(x)q(x, x') \min\{1, \frac{\pi(x')q(x', x)}{\pi(x)q(x, x')}\} \\ &= \min\{\pi(x)q(x, x'), \pi(x')q(x', x)\} \\ &= \pi(x')q(x', x) \min\{\frac{\pi(x)q(x, x')}{\pi(x')q(x', x)}, 1\} \\ &= \pi(x')p(x', x). \end{aligned}$$

所以 (3.5) 成立.

因为 (3.5) 式成立, 所以有:

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

(最后一个等号成立是因为 $p(x', x)$ 是一个概率核.)

所以, $\pi(x)$ 是 Markov 链的平稳分布, 证毕.

Metropolis-Hasting 采样法的具体步骤:

1. 任意选取 Markov 链的一个初始状态 $X_0 = x$.
2. 由转移核 $q(\cdot, x)$ 产生一个尝试移动 x' .

3. 生成 $U(0,1)$ 随机数 u , 如果 $u < \alpha(x, x')$, 则令 $X_1 = x'$, 否则保持当前状态不变, 即 $X_1 = X_0 = x$.

4. 重复上述步骤, 依次生成 X_2, X_3, \dots, X_n .

附: 两种常用的 $q(x', x)$ 选择:

(1) Metropolis 选择

Metropolis 曾经考虑对称分布, 即

$$q(x', x) = q(x, x') \quad \forall x, x'.$$

此时 $\alpha(x, x') = \min\{1, \frac{\pi(x')}{\pi(x)}\}$.

对称的分布是很常用的, 比如当 x 给定时, $q(x, x')$ 可以取成正态分布, 它以 x 为均值, 方差为常数.

(2) 独立抽样

如果 $q(x, x')$ 与当前状态 x 无关, 即 $q(x, x') = q(x')$, 则由此分布所导出的 Metropolis-Hasting 算法称为独立抽样. $\alpha(x, x') = \min\{1, \frac{\omega(x')}{\omega(x)}\}$, 其中 $\omega(x) = \frac{\pi(x)}{q(x)}$. 一般, 独立抽样的效果可能很好, 也可能很不好. 通常, 要使独立抽样有好的效果, $q(x)$ 应该接近 $\pi(x)$.

第四章 Markov 链的收敛性

我们已经知道了如何去构造 MCMC 算法,但它们是否收敛到分布 $\pi(\cdot)$? 以多快的速度收敛? 这就是本节所要讨论的 Markov 链的收敛问题.

令 $P^n(x, A)$ 为 Markov 链的 n 步收敛概率,

$$P^n(x, A) = P[X_n \in A | X_0 = x]$$

则 MCMC 的收敛问题变为: 当 n 多大时, $P^n(x, A) \rightarrow \pi(A)$?

4.1 全变差范数

我们用全变差范数来度量测度间的距离, 定义如下

定义 4.1. 两个概率测度 $\nu_1(\cdot), \nu_2(\cdot)$ 间的全变差范数为:

$$\|\nu_1(\cdot) - \nu_2(\cdot)\| = \sup_A |\nu_1(A) - \nu_2(A)|.$$

则 Markov 链的收敛问题变为: 是否

$$\lim_{n \rightarrow \infty} \|P^n(x, A) - \pi(A)\| = 0?$$

对给定的 $\varepsilon > 0$, 当 n 至少为多大时有

$$\|P^n(x, A) - \pi(A)\| < \varepsilon.$$

这些问题将在以后的章节中陆续得到讨论.

首先来介绍一些全变差范数的性质.

命题 4.1. (a) $\|\nu_1(\cdot) - \nu_2(\cdot)\| = \sup_{f: X \rightarrow [0,1]} |\int f d\nu_1 - \int f d\nu_2|$.

(b) 对任意 $a < b$, 有 $\|\nu_1(\cdot) - \nu_2(\cdot)\| = \frac{1}{b-a} \sup_{f: X \rightarrow [a,b]} |\int f d\nu_1 - \int f d\nu_2|$. 且特别的有, $\|\nu_1(\cdot) - \nu_2(\cdot)\| = \frac{1}{2} \sup_{f: X \rightarrow [-1,1]} |\int f d\nu_1 - \int f d\nu_2|$.

(c) 如果对 Markov 链的核 P , $\pi(\cdot)$ 是平稳分布, 则 $\|P^n(x, \cdot) - \pi(\cdot)\|$ 关于 n 是 i.e. 不增的, 即对 $n \in N$, 有 $\|P^n(x, \cdot) - \pi(\cdot)\| \leq \|P^{n-1}(x, \cdot) - \pi(\cdot)\|$.

(d) 更一般的, 令 $(\nu_i P)(A) = \int \nu_i(dx) P(x, A)$, 我们有 $\|(\nu_1 P)(\cdot) - (\nu_2 P)(\cdot)\| \leq \|\nu_1(\cdot) - \nu_2(\cdot)\|$.

(e) 令 $t(n) = 2 \sup_{x \in \chi} \|P^n(x, \cdot) - \pi(\cdot)\|$, 其中 $\pi(\cdot)$ 是平稳分布, 则 t 具有次可乘性, 即对 $m, n \in N$, 有 $t(m+n) \leq t(m)t(n)$ a.e.

(f) 如果对某个 σ 有限测度 $\rho(\cdot)$, $\mu(\cdot)$ 和 $\nu(\cdot)$ 分别有密度 g 和 h , 令 $M = \max(g, h)$, $m = \min(g, h)$, 则

$$\|\mu(\cdot) - \nu(\cdot)\| = \frac{1}{2} \int_{\chi} (M - m) d\rho = 1 - \int_{\chi} m d\rho.$$

(g) 对给定的概率测度 $\mu(\cdot)$ 和 $\nu(\cdot)$, 有联合定义的随机变量 X 和 Y 使得 $X \sim \mu(\cdot)$, $Y \sim \nu(\cdot)$, 且有 $P[X = Y] = 1 - \|\mu(\cdot) - \nu(\cdot)\|$.

4.2 渐进收敛

即使一个 Markov 有平稳分布 $\pi(\cdot)$, 它也有可能不收敛到该平稳分布.

反例: 令 $\chi = 1, 2, 3$, $\pi[1] = \pi[2] = \pi[3] = \frac{1}{3}$. 令 $P(1, 1) = P(1, 2) = P(2, 1) = P(2, 2) = \frac{1}{2}$, $P(3, 3) = 1$, 可验证知 $\pi(\cdot)$ 为平稳分布. 但是, 如果 $X_0 = 1$, $X_n \in 1, 2$, 则对所有的 n 有 $P(X_n = 3) = 0$, 故 $P(X_n = 3) \not\rightarrow \pi[3]$, X_n 的分布不收敛于 $\pi(\cdot)$. (事实上, 这里的平稳分布不是唯一的, X_n 的分布收敛于另一个不同的平稳分布 $\pi[1] = \pi[2] = \frac{1}{2}$).

上述例子是“可约的”, 在这种情况下 Markov 链不管经过多少步转移也不可能从状态 1 转移到状态 3. 下面介绍“不可约”的定义:

定义 4.2. 一个链称为 ϕ -不可约链, 如果 χ 上存在一个非零的 σ -有限测度 ϕ , 使得对所有的 $A \subseteq \chi$ 有 $\phi(A) > 0$, 并且对所有 $x \in \chi$, 存在正整数 $n=n(x, A)$ 使得 $P^n(x, A) > 0$.

即使是 ϕ -不可约链也不一定收敛到平稳分布, 比如说在周期情形下, 如下面的一个简单例子.

设 $\chi = 1, 2, 3$, $\pi[1] = \pi[2] = \pi[3] = \frac{1}{3}$, 令 $P(1, 2) = P(2, 3) = P(3, 1) = 1$. 则 $\pi(\cdot)$ 为平稳分布, 且该链为 ϕ -不可约链 [如 $\phi(\cdot) = \delta_1(\cdot)$]. 但是, 如果令 $X_0 = 1$, 且只要 n 是 3 的倍数就有 $X_n = 1$, 则 $P(X_n = 1)$ 的取值在 0 和 1 之间震荡, 故 $P(X_n = 1) \not\rightarrow \pi[3]$, 即 ϕ -不可约链没有收敛到平稳分布.

为了避免上述问题, 现引入非周期性的概念.

定义 4.3. 一个有平稳分布 $\pi(\cdot)$ 的 Markov 链是非周期的, 如果这里不存在 $d \geq 2$ 以及不相交的子集 $\chi_1, \chi_2, \dots, \chi_d \subseteq \chi$, 使得对所有的 $x \in \chi_i$ ($1 \leq i \leq d-1$)

有 $P(x, \chi_{i-1}) = 1$, 并且 $P(x, \chi_1) = 1$ 对所有 $x \in \chi_d$, 使得 $\pi(\chi_1) > 0$ (因此对所有 i 有 $\pi(\chi_i) > 0$). (否则, 链称为周期的, 有周期 d , 周期分解为 $\chi_1, \chi_2, \dots, \chi_d$).

综上所述定义, 可得到如下定理:

定理 4.1. 在一个有可数生成 σ -域的状态空间上, 若 Markov 链是 ϕ -不可约的, 非周期的, 且有平稳分布 $\pi(\cdot)$, 则对 π -a.e., $x \in \chi$, 有

$$\lim_{n \rightarrow \infty} \|P^n(x, \cdot) - \pi(\cdot)\| = 0.$$

特别的, 对所有测度 $A \subseteq \chi$, 有 $\lim_{n \rightarrow \infty} P^n(x, A) = \pi(A)$.

引理 4.1. 在定理 4.1 的条件下, 如果 $h: \chi \rightarrow R$ 且 $\pi(|h|) < \infty$, 则强大数定理也是成立的, 如下:

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} \right) \sum_{i=1}^n h(X_i) = \pi(h). \quad w.p. \quad (4.1)$$

由于对 MCMC 算法, 通常建立的要求是 $\pi(\cdot)$ 是平稳分布, 且一般会直接明确链是 ϕ -不可约的 (如一定区域上的 Lebesgue 测度), 非周期性也一般成立 (如对 Metropolis 算法或 Gibbs 算法), 故定理 1 被广泛的应用于 MCMC 算法中.

引理 4.2. (在周期的情形下) 如果一个 Markov 链是 ϕ -不可约的, 有周期 $d \geq 2$, 且有平稳分布 $\pi(\cdot)$, 则对 π -a.e., $x \in \chi$, 有

$$\lim_{n \rightarrow \infty} \|\sum_{i=n}^{n+d-1} P^i(x, \cdot) - \pi(\cdot)\| = 0. \quad (4.2)$$

并且强大数定理 (4.1) 也是同样成立的.

4.3 一致遍历

定理 4.1 给出了定性收敛到平稳分布的条件, 但是并没有指出以多大的速率定量收敛, 和定量收敛有关的一个性质就是一致遍历.

定义 4.4. 一个有平稳分布 $\pi(\cdot)$ 的 Markov 链是一致遍历的, 如果存在 $\rho < 1, M < \infty$, 使得

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq M\rho^n, \quad n = 1, 2, 3, \dots$$

关于一致收敛有下列命题成立:

命题 4.2. 一个有平稳分布 $\pi(\cdot)$ 的 Markov 链是一致遍历的, 当且仅当存在 $n \in N$, 使得下式成立:

$$\sup_{x \in \chi} \|P^n(x, \cdot) - \pi(\cdot)\| < \frac{1}{2}.$$

证明: “ \Rightarrow ” 若链是一致遍历的, 则

$$\lim_{n \rightarrow \infty} \sup_{x \in \chi} \|P^n(x, \cdot) - \pi(\cdot)\| \leq \lim_{n \rightarrow \infty} M\rho^n = 0.$$

故对足够大的 n , 有 $\sup_{x \in \chi} \|P^n(x, \cdot) - \pi(\cdot)\| < \frac{1}{2}$.

“ \Leftarrow ” 若存在 $n \in N$, 使得 $\sup_{x \in \chi} \|P^n(x, \cdot) - \pi(\cdot)\| < \frac{1}{2}$, 则由命题 4.1(e), 有 $d(n) = 2 \sup_{x \in \chi} \|P^n(x, \cdot) - \pi(\cdot)\| \equiv \beta < 1$. 故对所有的 $j \in N$, 有 $d(jn) \leq (d(n))^j = \beta^j$. 因此, 由命题 4.1(c) 有

$$\begin{aligned} & \|P^m(x, \cdot) - \pi(\cdot)\| \\ & \leq \|P^{\lfloor \frac{m}{n} \rfloor n}(x, \cdot) - \pi(\cdot)\| \\ & \leq \frac{1}{2} d(\lfloor \frac{m}{n} \rfloor n) \\ & \leq \beta \lfloor \frac{m}{n} \rfloor \\ & \leq \beta^{-1} (\beta^{\frac{1}{n}})^m. \end{aligned}$$

令 $M = \beta^{-1}, \rho = \beta^{\frac{1}{n}}$, 则该链是一致遍历的.

定义 4.5. 子集 $C \subseteq \chi$ 是小集, 如果存在正整数 $n_0, \varepsilon > 0$, χ 上一概率测度 $\nu(\cdot)$ 使得下面最小化条件成立:

$$P^{n_0}(x, \cdot) \geq \varepsilon \nu(\cdot) \quad x \in C \quad (4.3)$$

特别的, 对所有测度 $A \subseteq \chi$, 所有 $x \in C$, i.e. 有 $P^{n_0}(x, A) \geq \varepsilon \nu(A)$ i.e..

关于一致遍历最重要的结论为:

定理 4.2. 设 Markov 链有平稳概率分布 $\pi(\cdot)$, 若在特别的空间 $C = \chi$ 上 (即整个状态空间为小集时), 存在正整数 $n_0, \varepsilon > 0$, 概率测度 $\nu(\cdot)$ 使得最小化条件 (4.3) 成立. 则该 Markov 链是一致遍历的, 并且对所有的 $x \in \chi$ 有 $\|P^n(x, \cdot) - \pi(\cdot)\| \leq (1 - \varepsilon)^{\lfloor \frac{n}{n_0} \rfloor}$, 这里 $\lfloor r \rfloor$ 是不超过 r 的最大整数.

定理 4.2 给出了 $\|P^n(x, \cdot) - \pi(\cdot)\|$ 定量收敛到一有界值的条件, 注意它必须是 $\leq (1 - \varepsilon)^{\lfloor \frac{n}{n_0} \rfloor}$. 因此, 一旦 n_0 和 ε 已知, 就可以找到 n_* , 使得 $\|P^{n_*}(x, \cdot) - \pi(\cdot)\| \leq 0.01$. 这些在 MCMC 中有着非常重要广泛的应用.

对离散状态空间的情形, 令

$$\varepsilon_{n_0} = \sum_{y \in \chi} \inf_{x \in C} p^{n_0}(x, y) > 0.$$

则同样有 $\|P^n(x, \cdot) - \pi(\cdot)\| \leq (1 - \varepsilon)^{\lfloor \frac{n}{n_0} \rfloor}$.

4.4 几何遍历

比一致遍历条件稍弱的是几何遍历.

定义 4.6. 一个有平稳分布 $\pi(\cdot)$ 的 Markov 链是几何遍历的, 如果当 $M(x) < \infty$ 时, 对 π -a.e., $x \in \chi$, 存在 $\rho < 1$ 满足

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq M(x)\rho^n, \quad n = 1, 2, 3, \dots$$

定义 4.7. 称一个 Markov 链满足转移条件, 如果存在常数 $0 < \lambda < 1, b < \infty$, 以及函数 $V: \chi \rightarrow [1, \infty]$, 使得下式成立

$$PV \leq \lambda V + bI_C \quad \text{i.e.} \quad (4.4)$$

即对所有 $x \in \chi$, 有 $\int_{\chi} P(x, dy)V(y) \leq \lambda V(x) + bI_C(x)$

关于几何遍历最主要的定理为:

定理 4.3. 对一个 ϕ -不可约的, 非周期的 Markov 链, 有平稳分布 $\pi(\cdot)$, 设存在 $C \subset \chi, \varepsilon > 0$, 以及概率测度 $\nu(\cdot)$ 使得最小化条件 (4.3) 成立. 又进一步设存在常量 $0 < \lambda < 1, b < \infty$, 函数 $V: \chi \rightarrow [1, \infty]$ (至少有一个 $V(x) < \infty$) 满足转移条件 (4.4), 则该 Markov 链的几何遍历的.

MCMC 的几何遍历性是一个非常重要的性质. 它与中心极限定理结合在 MCMC 中有着非常广泛的运用. 定理 4.1, 4.2, 4.3 的证明在文献 [14] 中已经给出.

第五章 Markov 链的定量收敛

5.1 定量收敛率

定义 5.1. (二元转移条件) 存在函数 $h: \chi \times \chi \rightarrow [1, \infty)$, $\alpha > 1$ 使得

$$\bar{P}h(x, y) \leq h(x, y)/\alpha, \quad (x, y) \notin C \times C \quad (5.1)$$

成立, 其中

$$\bar{P}h(x, y) = \int_{\chi} \int_{\chi} h(z, w) P(x, dz) P(y, dw).$$

定义 5.2. 对 $(x, y) \in C \times C$, 定义

$$B = \max[1, \alpha(1 - \varepsilon) \sup_{C \times C} \bar{R}h]. \quad (5.2)$$

其中

$$\bar{R}h(x, y) = \int_{\chi} \int_{\chi} (1 - \varepsilon)^2 h(z, w) (P(x, dz) - \varepsilon v(dz)) (P(y, dw) - \varepsilon v(dw)).$$

定理 5.1. 已知状态空间 χ 上的一 Markov 链, 有转移核 P , 设有 $C \subseteq \chi$, $h: \chi \times \chi \rightarrow [1, \infty)$ 概率分布 $v(\cdot)$, $\alpha > 1, n_0 \in N, \varepsilon > 0$, 使得 (4.3), (5.1) 成立, 定义 B 如 (5.2). 则对任意联合初始分布 $L(X_0), L(X'_0)$. 任意整数 $1 \leq j \leq k$, 如果 $\{X_n\}, \{X'_n\}$ 是 $L(X_0), L(X'_0)$ 上的两个 Markov 链, 则

$$\|L(X_k) - L(X'_k)\| \leq (1 - \varepsilon)^j + \alpha^{-k} B^{j-1} E[h(X_0, X'_0)].$$

定理 5.1 的证明将在后面用耦合构造的方法给出. 这个定理的各种变型和推广已经广泛的应用于 MCMC 算法中, 虽然很难直接将定理 5.1 直接应用于实际的 MCMC 算法, 但是它确实建立了严格的, 精确的, 可行的公理去确保 Markov 链的定量收敛, 因此具有很好的理论指导意义.

5.2 用耦合的方法证明定理

5.2.1. 耦合不等式

关于耦合最基本的思想即为下式, 设在空间 χ 上有两个联合定义的随机变量 X, Y , 如果令 $L(X), L(Y)$ 为它们分别的概率分布, 则有

$$\begin{aligned}
 \|L(X_k) - L(X'_k)\| &= \sup_A |P(X_k \in A) - P(X'_k \in A)| \\
 &= \sup_A |P(X_k \in A, X_k = X'_k) + P(X_k \in A, X_k \neq X'_k) \\
 &\quad - P(X'_k \in A, X_k = X'_k) - P(X'_k \in A, X_k \neq X'_k)| \\
 &= \sup_A |P(X_k \in A, X_k \neq X'_k) - P(X'_k \in A, X_k \neq X'_k)| \\
 &\leq P(X_k \neq X'_k).
 \end{aligned} \tag{5.3}$$

5.2.2. 耦合的构造

设 C 是小集, 从 $X_0 = x, X'_0 \sim \pi(\cdot), n = 0$ 开始, 一直重复下列步骤:

对给定的 X_n, X'_n

1. 若 $X_n = X'_n$, 则令 $X_{n+1} = X'_{n+1} \sim P(X_n, \cdot)$, 用 $n+1$ 代替 n .

2. 若 $X_n \neq X'_n$ 且 $(X_n, X'_n) \in C \times C$ 则

(a) w.p. ε 选择 $X_{n+n_0} = X'_{n+n_0} \sim \nu(\cdot)$

(b) w.p. $1 - \varepsilon$ 条件独立的选择

$$X_{n+n_0} \sim \frac{1}{1-\varepsilon} [P^{n_0}(X_n, \cdot) - \varepsilon \nu(\cdot)].$$

$$X'_{n+n_0} \sim \frac{1}{1-\varepsilon} [P^{n_0}(X'_n, \cdot) - \varepsilon \nu(\cdot)].$$

若 $X_n \neq X'_n$ 且 $(X_n, X'_n) \notin C \times C$ 则独立的选择

$$X_{n+1} \sim P(X_n, \cdot), X'_{n+1} \sim P(X'_n, \cdot),$$

用 $n+1$ 代替 n .

由此构造得到的 $\{X_n\}, \{X'_n\} \quad n = 1, 2, 3, \dots$ 即是 $L(X_0), L(X'_0)$ 上的两个 Markov 链.

5.2.3. 定理 5.1 的证明

(i) 令 $N_k = \{m : 0 \leq m \leq k, (X_m, X'_m) \in C \times C\}$, 令 (X_n, X'_n) 到达 $C \times C$ 的时刻为 τ_1, τ_2, \dots 对任意整数 $j (0 \leq j \leq k)$ 有

$$P[X_k \neq X'_k] = P[X_k \neq X'_k, N_{k-1} \geq j] + P[X_k \neq X'_k, N_{k-1} < j]. \tag{5.4}$$

由于事件 $\{X_k \neq X'_k, N_{k-1} \geq j\} \supseteq \{X_t \neq X'_t, (X_t, X'_t) \in C \times C, t = 1, \dots, j\}$.

而由耦合的构造知

$$\begin{aligned} & P\{X_t = X'_t, (X_t, X'_t) \in C \times C\} = \varepsilon \\ \Rightarrow & P[X_k \neq X'_k, N_{k-1} \geq j] \leq (1 - \varepsilon)^j. \end{aligned} \quad (5.5)$$

为证明 (5.5) 式第 2 部分的有界性, 令

$$M_k = \alpha^k B^{-N_{k-1}} h(X_k, X'_k) I(X_k \neq X'_k), \quad k = 0, 1, 2, \dots \text{ 其中 } N_{-1} = 0$$

(ii) 引理: $E[M_{k+1} \mid X_0, \dots, X_k, X'_0, \dots, X'_k] \leq M_k$ i.e. 即 $\{M_k\}$ 是上鞅.

证明: 若 $(X_k, X'_k) \notin C \times C$, 则由 N_k 的定义 $N_k = N_{k-1}$, 故

$$\begin{aligned} & E[M_{k+1} \mid X_0, \dots, X_k, X'_0, \dots, X'_k] \\ &= \alpha^{k+1} B^{-N_{k-1}} E[h(X_{k+1}, X'_{k+1}) I(X_{k+1} \neq X'_{k+1}) \mid X_k, X'_k] \\ &\leq \alpha^{k+1} B^{-N_{k-1}} E[h(X_{k+1}, X'_{k+1}) \mid X_k, X'_k] I(X_k \neq X'_k) \\ &= M_k \alpha E[h(X_{k+1}, X'_{k+1}) \mid X_k, X'_k] / h(X_k, X'_k) \\ &= M_k \alpha \bar{P} h(X_k, X'_k) / h(X_k, X'_k) \leq M_k. \end{aligned}$$

同样, 若 $(X_k, X'_k) \in C \times C$, 则 $N_k = N_{k-1} + 1$, 不妨设 $X_k \neq X'_k$ (若 $X_k = X'_k$ 则 $M_{k+1} = M_k = 0$) 则有

$$\begin{aligned} & E[M_{k+1} \mid X_0, \dots, X_k, X'_0, \dots, X'_k] \\ &= \alpha^{k+1} B^{-N_{k-1}-1} E[h(X_{k+1}, X'_{k+1}) I(X_{k+1} \neq X'_{k+1}) \mid X_k, X'_k] \\ &= \alpha^{k+1} B^{-N_{k-1}-1} (1 - \varepsilon) (\bar{R}h)(X_k, X'_k) \\ &= M_k \alpha B_{-1} (1 - \varepsilon) (\bar{R}h)(X_k, X'_k) / h(X_k, X'_k) \leq M_k. \end{aligned}$$

综上有 M_k 是上鞅.

(iii) 由定义知 $B \geq 1$,

$$\begin{aligned} & P[X_k \neq X'_k, N_{k-1} < j] = P[X_k \neq X'_k, N_{k-1} \leq j - 1] \\ &\leq P[X_k \neq X'_k, B^{-N_{k-1}} \geq B^{-(j-1)}] \\ &= P[I(X_k \neq X'_k) B^{-N_{k-1}} \geq B^{-(j-1)}] \\ &\leq B^{j-1} E[I(X_k \neq X'_k)] B^{-N_{k-1}} \quad (\text{由 Markov 不等式}) \\ &\leq B^{j-1} E[I(X_k \neq X'_k)] B^{-N_{k-1}} h(X_k, X'_k) \quad (h \geq 1) \\ &= \alpha^{-k} B^{j-1} E[M_k] \leq \alpha^{-k} B^{j-1} E[M_0] \quad (\text{由 } M_k \text{ 是上鞅}) \\ &= \alpha^{-k} B^{j-1} E[h(X_0, X'_0)]. \quad (\text{由 } M_0 \text{ 的定义}) \end{aligned} \quad (5.6)$$

联合 (5.3)(5.4)(5.5)(5.6) 可得

$$\|L(X_k) - L(X'_k)\| \leq (1 - \varepsilon)^j + \alpha^{-k} B^{j-1} E[h(X_0, X'_0)].$$

定理证毕.

参考文献

- [1] Andrieu C., Fort G.. Explicit control of subgeometric ergodicity [J]. *Rapport de Recherche*, 2005, 05: 17
- [2] Baxendale P H.. Renewal theory and computable convergence rates for geometrically ergodic Markov chains[J]. *Annals of Applied Probability*, 2005, 15: 700–738
- [3] Douc R., Moulines E.. Quantitative bounds on convergence of time-inhomogeneous Markov chains[J]. *Annals of Applied Probability*, 2004, 14: 1643–1665
- [4] Douc R, Fort G., Moulines E., Soulier P.. Practical drift conditions for subgeometric rates of convergence[J]. *Ann Appl Probab*, 2004, 14: 1353–1377
- [5] Fort G., Moulines E. Polynomial ergodicity of Markov transition kernels[J]. *Stochastic Processes and their Applications*, 2003, 103: 57–99
- [6] Fort G., Moulinier E.. V -subgeometric ergodicity for a Hastings–Metropolis algorithm[J]. *Statist Probab Lett*, 2000, 49: 401–410
- [7] Fort G., Roberts G O.. Subgeometric ergodicity of strong Markov processes[J]. *Annals of Applied Probability*, 2005, 15: 1565–1589
- [8] Jarner S., Roberts G.O.. Polynomial convergence rates of Markov chains[J]. *Annals of Applied Probability*, 2001, 12: 224–247
- [9] Klok S A., Veretennikov A Y.. Sub-exponential mixing rates for a class of Markov chains[J]. *Math Commun*, 2004, 9: 9–26
- [10] Lund R B., Meyn S P., Tweedie R.. Computable exponential convergence rates for stochastically ordered Markov processes[J]. *Annals of Applied Probability*, 1996, 6: 218–237
- [11] Meyn S P., Tweedie R L.. Computable bounds for convergence rates of Markov chains[J]. *Annals of Applied Probability*, 1994, 4 : 981–1011
- [12] Meyn S P., Tweedie R L.. Markov chains and Stochastic Stability [M]. *London: Springer–Verlag London Ltd*, 1993

- [13] Nummelin E., Tuominen P.. The rates of convergence in Orey's theorem for Harris recurrent Markov Chains with applications to renewal theory[J]. *Stoch Proc Appl*, 1983, 15: 295–311
- [14] Roberts G O, Rosenthal J S.. General State Space and MCMC algorithms[J]. *Probab surv*, 2004, 1: 20–71
- [15] Rosenthal J S.. Geometric convergence rates for Time-sampled Markov chains[J]. *J Theor Prob*, 2003, 16: 671–688
- [16] Rosenthal J S.. Minorization conditions and convergence rates for Markov chains Monte Carlo[J]. *J Amer Sta Assoc*, 1995, 90: 558–688
- [17] Tuomionen P., Tweedie R.. Subgeometric rates of convergence of f-ergodic Markov chains[J]. *Advances in Applied Probability* , 1994, 26: 775–798
- [18] 龚光鲁, 钱敏平. 应用随机过程教程及在算法和智能计算中的随机模型 [M]. 北京: 清华大学出版社, 2004
- [19] 钱敏平, 龚光鲁. 随机过程论 [M]. 北京: 北京大学出版社, 1997
- [20] 王寿仁. 概率论基础和随机过程 [M]. 北京: 科学出版社, 2000
- [21] 严士健. 无穷粒子马尔可夫过程引论 [M]. 北京: 北京师范大学出版社, 1996

致 谢

这篇论文从选题到最后校稿成文都凝结了导师张绍义教授的大量心血，作者在此向导师致以诚挚的谢意！三年来张老师不仅在学业上对我精心指导，诲人不倦；而且他严谨的治学态度，缜密而富有创新的思考方式，温和、大度、远见博识、平易近人等人格魅力都将使我受益终身！衷心的祝愿张老师身体健康，合家幸福！

三年的学习过程中，得到万成高教授，刘莉副教授等老师的教学和指导，这些都是我人生中的宝贵财富，在此谨向各位老师表示衷心的感谢！

感谢张水利，张韧，危黎黎等同学的无私帮助和真诚合作！

感谢我的家人三年中给予我无私的关怀和帮助，无私的爱！

最后，谨向评审这篇论文的各位学者、老师致以诚挚的谢意！

马尔可夫链蒙特卡罗算法

作者：[田菲](#)
学位授予单位：[湖北大学](#)

本文链接：http://d.g.wanfangdata.com.cn/Thesis_Y1223101.aspx