第 6 讲：贝叶斯蒙特卡洛计算

张伟平

目录

1.1 贝叶斯推断例子

假设 $X_1, \ldots, X_k$ 为独立的普阿松 (poisson) 随机变量，且 $X_i \sim Poisson(\theta_i), i = 1, \ldots, k$. 如果 θ_i 的先验分布为通过其对数的联合分布给出:

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

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，μ, τ^2, ρ 为已知的常数.

则由样本分布

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

以及先验分布

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

可以得到后验分布

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

因此如果感兴趣的是 θ_j 的后验均值，则需要计算

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

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.

1.2 EM 方法

- E-M (Expectation-Maximization) 算法是由 Dempster et al. (1977) 提出的一种最优化方法，常用于不完全数据求极大似然估计.

- 设 $Y|\theta$ 有密度 $f(y|\theta)$, 令 θ 的先验分布为 $\pi(\theta)$，由此得到的后验分布记为 $\pi(\theta|y)$. 当计算 $\pi(\theta|y)$ 数字特征非常困难时，有一些“数据扩张”(data augmentation) 的方法或许可以用来解决此类困难.

- 其想法是将观测到的数据 y 与缺失数据或者隐变量数据 z 扩张为“完全”数据 $\mathbf{x} = (y, z)$, 使得扩张后的后验分布 $\pi(\theta|\mathbf{x}) = \pi(\theta|y, z)$ 在计算上是容易处理的.

- 由于目的是最大化后验分布，E-M 算法只能计算后验众数.

EM 算法

- 记 $p(z|y,\hat{\theta})$ 为 Z 在给定 $y,\hat{\theta}$ 时候的预测分布，$\hat{\theta}^{(i)}$ 为在第 i 次迭代时候 θ 的估计值，则 E-M 算法的基本步骤如下：

(1) 计算 $Q(\theta|\theta^{(i)}) = E[\log \pi(\theta|y,z)|y,\hat{\theta}^{(i)}]$;

(2) 最大化 $Q(\theta|\theta^{(i)})$ 记其最大值为 $\hat{\theta}^{(i+1)}$;

(3) 重复 (1) 和 (2) 直至达到收敛要求.

- 这种先求期望，然后求最大值的步骤即为 E-M 算法名称的来源.

Stochastic EM 算法

- 是 MCEM 的一种，其思想通过从 $p(z|y,\hat{\theta})$ 中随机抽样方式来生成缺失变量 z 的值，然后带入到完全后验分布中。

(1) 计算 $z^{(i)} = E[Z|y, \hat{\theta}^{(i)}]$;

(2) 扩充观测数据 y 为 $(y, z^{(i)})$, 最大化 $\pi(\theta|y, z^{(i)})$, 记其最大值为 $\hat{\theta}^{(i+1)}$;

(3) 使用 $\hat{\theta}^{(i+1)}$ 和 (1), 获得 $z^{(i+1)}$, 然后再代入 (2), 如此重复下去直至达到收敛要求.

课本 P194 例 6.3.1

↑Example

↓Example

1.3 蒙特卡洛抽样方法

- 假设感兴趣的量为如下形式的有限期望

$$E_f h(X) = \int_{\mathscr{X}} h(x) f(x) dx.$$

直接抽样 若能从 f 中产生 $i.i.d$ 观测 $X_1, X_2, \cdots, X_m$，则

$$\bar{h}_m = \frac{1}{m} \sum_{i=1}^{m} h(X_i) \to E_f h(X)$$

这种使用随机样本来逼近的方法被称为蒙特卡洛抽样方法.

重要性抽样 假设 f 难以直接抽样，而我们可以从另一各个分布 g 中容易抽样，则

$$
\begin{aligned}
E_f h(X) &= \int_{\mathscr{X}} h(x) f(x) dx = \int_{\mathscr{X}} h(x) \frac{f(x)}{g(x)} g(x) dx \\
&= \int_{\mathscr{X}} \{h(x) w(x)\} g(x) dx = E_g[h(x) w(x)],
\end{aligned}
$$

其中 $w(x) = f(x)/g(x)$. 从 g 中产生 $i.i.d$ 样本 $X_1, \cdots, X_m$, 则一个合适的估计量为

$$
\widehat{hw}_m = \frac{1}{m} \sum_{i=1}^{m} h(X_i) w(X_i).
$$

抽样分布 g 称为**重要性函数**.

- 这种方法的一个问题是重要性函数 g 与 f 不相似，因此 f 的中心可能在 g 的尾部，使得积分值只有少部分数据决定
- 第二个问题是重要性抽样方法并没有给出 f 的随机样本，这可以通过拒绝抽样方法来解决（但是合适的提议分布比较难找）

1.4 MCMC 算法基础

- 蒙特卡洛算法一般来说不容易实施。例如，重要性函数一般难以找到合适的；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.
- 因此其他一些替代算法来产生近似的蒙特卡洛样本，最流行的就是蒙特卡洛马尔可夫链算法 (MCMC)，其通过从一个平稳分布是目标分布的马尔可夫链中抽样

1.4.1 马氏链基本定义

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, $S = \{0, 1, \cdots\}$ 为状态集. 若对一切 n 有

Definition

$$P\left(X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, \cdots, X_{n-1} = i_{n-1}, X_n = i\right)$$
$$= P\left(X_{n+1} = j \mid X_n = i\right) = p_{ij}$$

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 $\{X_n,\ n \geq 0\}$ 的将来状态 $\{X_{n+1}\}$ 只与现在状态 $\{X_n\}$ 有关, 而与过去状态 $\{X_k, k \leq n-1\}$ 无关.
- $P = \left(p_{ij}\right)_{i,j \in S}$ 称为马氏链的转移概率矩阵, 满足条件 $p_{ij} \geq$

0, 且 $\sum_{j \in S} p_{ij} = 1$.

- 条件概率 $P(X_{n+1} = j | X_n = i)$ 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**. 对时间齐次的马氏链，记 π_n 表示 n 时刻马氏链处于各状态的分布向量，则有

$$\pi_n = P^n \pi_0$$

在 MCMC 中，我们仅关注时间齐次的马氏链。

(**平稳分布**) 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 $\pi_j = \sum_i \pi_i p_{ij},$, 即 $\pi = P\pi$, 则称之为此马尔科夫链的平稳分布 .

Definition

- 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 $\pi = P\pi$, 即对任意 j 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 $\pi_n = P^n\pi = P\pi = \pi$, 即

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

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

对连续的状态空间 $\mathcal{S}$, 转移核 $P(x, A)$ 定义为从任意 $x \in \mathcal{S}$ 转移到可测集 $A \subset \mathcal{S}$ 的概率. 转移核密度 $p(x, y)$ 定义为满足下式的非负函数

$$P(x, A) = \int_{y \in A} p(x, y) dy, \forall x \in \mathcal{S},$$

定理 1. 设 π 为一个取值于可数状态空间 $\mathcal{S}$ 上的一个概率分布，如果一个马氏链的转移核满足对 $n = 0, 1, \ldots$ 有

$$\pi_j P(X_{n+1} = i | X_n = j) = \pi_i P(X_{n+1} = j | X_n = i), \forall i,\ j \in S$$

则 π 为该马氏链的平稳分布。称上式为细致平衡方程 *(detailed balance equation).*

事实上，如果 π 满足细致平衡方程，则两边对 j 求和有

$$\sum_{j \in S} \pi_j P(X_{n+1} = i | X_n = j) = \sum_{j \in S} \pi_i P(X_{n+1} = j | X_n = i) = \pi_i$$

即 $P\pi = \pi$。

对连续的状态空间 $\mathcal{S}$, 一个概率分布 π 称为是具有转移核密度 $p(x, y)$ 的平稳分布, 如果对任意的 $y \in \mathcal{S}$ 有

$$\pi(y) = \int p(x, y)\pi(x)dx$$

或者等价地

$$\pi(A) = \int P(x, A)\pi(x)dx$$

对任意可测集 $A \subset \mathcal{S}$.

- 可以证明, 如果此马氏链还是不可约、遍历的 (非周期、正常返), 则 π 是其唯一的平稳分布。

(**不可约性**) 一个具有可数状态空间 $\mathcal{S}$ 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

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

Definition

- 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.

(**周期与非周期**) 称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

$$k(i) = gcd\{n: \ P\big(X_n = i \ \big| \ X_0 = i\big) > 0\big), \qquad \text{Definition}$$

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.

(**正常返**] 对常返状态 i, 令 $T_i = \inf\{n \geq 1 : X_n = i | X_0 = i\}$ 为首次返回状态 i 的时刻, 如果

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

Definition

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

- 正常返性保证了马氏链的转移概率的极限不是 0.

(**遍历性**) 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

Definition

- 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、遍历**的. 满足这些正则条件的马氏链存在唯一的平稳分布.

1.4.2 马氏链的极限定理

定理 2. 设 $\{X_n,\ n \geq 0\}$ 为一具有可数状态空间 $\mathcal{S}$ 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

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

对 X_0 的任意初始分布 π.

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

定理 3. *(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 $\mathcal{S}$ 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 $h : S \to \mathbb{R}$ 以及初始值 X_0 的任意初始分布有

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

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 π 时，也有

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

- 这个定理结论是非常有用的. 比如给定集合 $\mathcal{S}$ 上的概率分布 π,

以及 $\mathcal{S}$ 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 $\mathcal{S}$ 且其平稳分布 π 就是目标后验分布 $\pi(\cdot|\mathbf{x})$,** 从一初始值 θ_0 出发，将此链运行一段时间，如 $0, 1, 2, \ldots, n-1$, 生成随机数 (样本) $\theta_0, \theta_1, \cdots, \theta_{n-1}$,

- 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 μ 的一相合估计. 这种技术称为马尔科夫链蒙特卡洛 (MCMC) 方法.

1.4.3 MCMC 实施中的若干术语

当从后验分布 $\pi(\theta|\mathbf{x})$ 中难以直接抽样时，我们需要

- 构造一个不可约、非周期的马氏链使其平稳分布就是目标后验分布 $\pi(\cdot|\mathbf{x})$,
- 从该马氏链中产生足够长时间的状态，使其达到平稳状态
- 利用平稳状态下的样本，计算后验分布的 Summary (均值、中位数、标准差、分位数，相关系数等), 获得相应的蒙特卡洛积分的模拟结果.

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

- 在 MCMC 机制中用以保证链达到平稳状态所运行的时间称为**预烧期**, 其迭代数记为 B.
- 为避免初始值的影响，预烧期中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.

蒙特卡洛误差 (MC error)

- 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** $\bar{\mu}_T - \mu$. 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. 由于 μ 未知, 因此需要其他方法来度量 MC 误差.

- 由于 $\sqrt{T}(\bar{\mu}_T - \mu) \to N(0, \sigma_h^2)$(在正则条件下), 其中 $\sigma_h^2 := \mathrm{var}_\pi \{h(\theta_1)\} + 2\sum_{i=2}^{\infty} \mathrm{cov}_\pi \{h(\theta_1), h(\theta_i)\}$; 脚表 π 表示期望在后验分布 π 下计算. (Jones, G. L. (2004); Roberts, G. O. and Rosenthal, J. S. (2004))

- 基于上述 CLT, 因此我们使用 σ_h 来度量估计量 $\bar{\mu}_T$ 中的 MC 误差, 其估计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，感兴趣量的估计精度也会增加.

- 由于 σ_h 未知, 估计此量的方法有很多, 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.

(非重叠) 组平均方法

首先将生成的 T 个样本分成 K 个组, 每个组 $\nu = T/K$ 个, 常取为 30 或 50. ν 和 K 都要比较大, 以使得方差的估计量是相合的以及减少自相关性. 在计算 MC 误差时,

- 计算组内均值

$$\overline{\mu}_b = \frac{1}{\nu} \sum_{t=(b-1)\nu+1}^{b\nu} h(\theta_{(t)}), b = 1, \ldots, K$$

- 则 σ_h^2 的组平均估计为

$$\hat{\sigma}_h^2 = \frac{\nu}{K-1} \sum_{b=1}^{K} (\bar{\mu}_b - \bar{\mu}_T)^2$$

- MC 误差估计

$$MCE(\bar{\mu}_T) = \frac{\hat{\sigma}_h}{\sqrt{T}} = \sqrt{\frac{1}{K(K-1)} \sum_{b=1}^{K} [\overline{g(X)}_b - \overline{g(X)}]^2}$$

窗口估计量方法

窗口估计方法基于 Roberts（1996）对自相关样本的样本方差表示

$$MCE(\bar{\mu}_T) = \frac{\hat{\sigma}}{\sqrt{T}}\sqrt{1 + 2\sum_{k=1}^{\infty}\hat{\rho}_k(h)}$$

其中 $\hat{\sigma}$ 是 $h(\theta_1)$ 的后验标准差的估计. 其中 $\hat{\rho}_k(h)$ 是估计 $h(\theta_(k))$ 与 $h(\theta_{(t+k)})$ 之间的 k 阶自相关系数. 显然，k 越大，自相关将接近 0. 因此，取一个窗口 w，使得其后的自相关系数都很小。即

$$MCE(\bar{\mu}_T) = \frac{\hat{\sigma}}{\sqrt{T}}\sqrt{1 + 2\sum_{k=1}^{w}\hat{\rho}_k(h)}$$

其中 $\hat{\sigma} = \frac{1}{T-1}\sum_{t=1}^{T}(h(\theta_t) - \bar{\mu}_T)^2$.

1.4.4 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.

监视蒙特卡洛误差

- 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.

样本路径图

- 另外一种监控方式是使用样本路径图 (trace plot). 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. 下图给出了一个明显没有达到收敛的例子. 而另一个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.

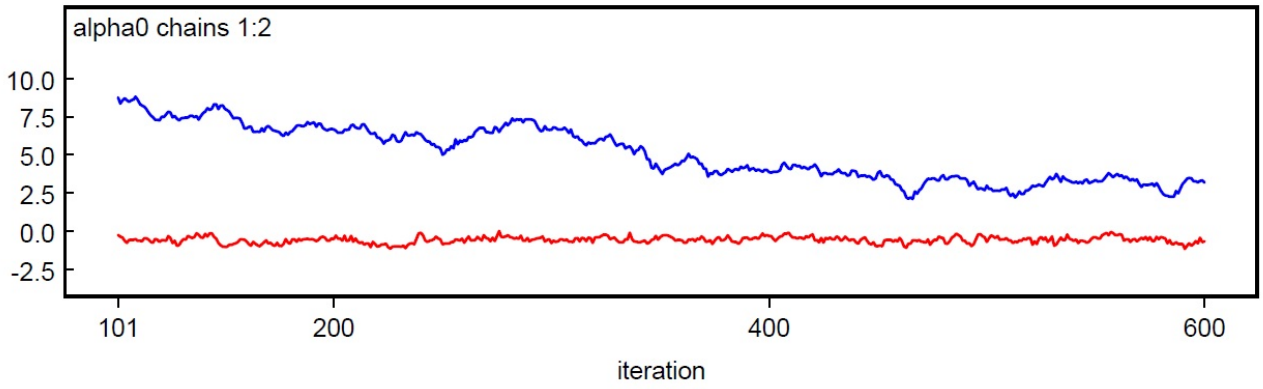

图 5.1 没有达到收敛的链的路径图

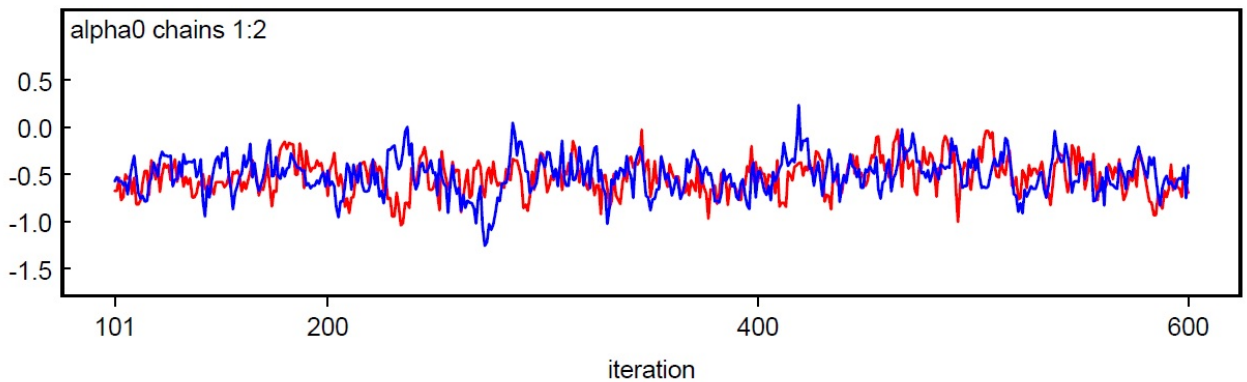

图 5.2 达到收敛的链的路径图

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).
- 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.

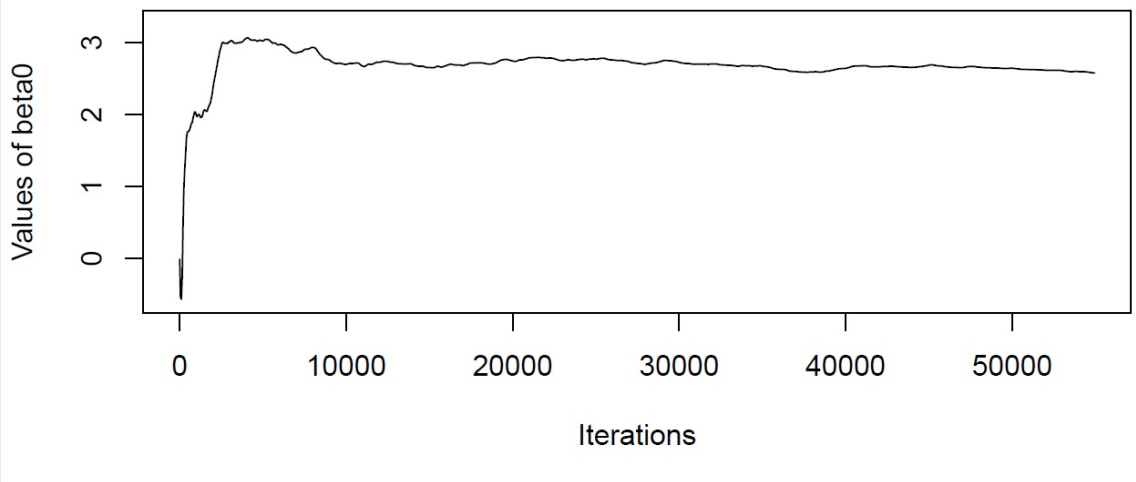

图 5.3 达到收敛的链的路径图

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 方法监测潜在尺度缩减因子 $\sqrt{\hat{R}}$ 是否收敛到 1 (见课本或者课程 R 实施例子)
- 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.

1.5 MCMC 算法实施

- MCMC 实施的核心问题是确定满足要求 (不可约、非周期、正常返且平稳分布为目标抽样分布) 的一个马氏链
- Metropolis-Hastings 算法是一类最常用的用以构造满足要求马氏链的 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 抽样方法及其几个变种, 它们在贝叶斯推断中有着广泛的应用.

1.5.1 Hastings (M-H) 抽样方法

为简化记号, 我们用 $f(\cdot)$ 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 $\{x_i,\ i=0,1,\cdots\}$.

- 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', 然后计算接受概率来决定是否将 x' 作为序列的下一个值:

  (1) 从提议分布 $g(\cdot|x_t)$ 产生一个候选点 x'

  (2) 计算接受概率

$$\alpha(x_t,x')=\min\{\frac{f(x')g(x_t|x')}{f(x_t)g(x'|x_t)},1\}.$$

(3) 以概率 $\alpha(x_t, x')$ 接受 $x_{t+1} = x'$；否则链停留在状态 $x_{t+1} = x_t$.

- 需要注意的一点是接受概率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 抽样方法的理论依据：

记 $q_{ij} = g(x_{t+1} = j | x_t = i)$ 容易看出，M-H 抽样方法产生的样本序列 $x_0, x_1, \ldots$ 为一马氏链，其转移核为

$$p_{ij} = q_{ij}\alpha(i,j) + \delta_i(j)(1 - r(i))$$

其中 $r(i) = \sum_j q_{ij}\alpha(i,j)$，$\alpha(i,j) = \min\{1, f_j q_{ji}/f_i q_{ij}\}$，$\delta_i(j)$ 为 dirac-delta 函数。

另一方面，对 $i \neq j$ 有

$$f_i q_{ij}\alpha(i,j) = f_i q_{ij} \min\{1, \frac{f_j q_{ji}}{f_i q_{ij}}\} = f_j q_{ji} \min\{1, \frac{f_i q_{ij}}{f_j q_{ji}}\} = f_j q_{ji}$$

以及 $i = j$ 时

$$f_i \delta_i(j)(1 - r(i)) = f_j \delta_j(i)(1 - r(j))$$

从而满足细致平衡方程

$$p_{ij} f_i = p_{ji} f_j, \forall i, j$$

从而 f 为该链的平稳分布。可以进一步验证此链是不可约、遍历的，因此 f 是其唯一的平稳分布。

注： 如果是连续的状态空间，则细致平衡方程中的求和改为积分，p_{ij} 解释为从一个状态 $i = x$ 转移到另一状态 $j = y$ 的转移核，即 $P(X_{t+1} \in A | X_t = x) = \int_A p(y|x)dy$

注： **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：

- 提议分布的支撑集包含目标分布的支撑集.
- 容易从中抽样, 常取为已知的分布，如正态或 t 分布等.
- 提议分布应使接受概率容易计算.
- 提议分布的尾部要比目标分布的尾部厚.
- 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.

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.

- **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

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

- **随机游走 Metropolis** 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 $Y|X_t \sim N(X_t, \sigma^2)$, $\sigma^2 > 0$.

- **独立抽样方法**独立抽样中的提议分布不依赖于链的前一步状态值. 因此 $g(Y|X_n) = g(Y)$, 接受概率为

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

- **逐分量的 M-H 抽样方法**当状态空间为 k 维 $(k > 1)$ 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.

  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 $i = 1, \cdots, k$, 从第 i 个提议分布 $q_i(y|X_{n,i}, \mathbf{X}^*_{n,-i})$ 中

产生 Y_i, 这里

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

– 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 $X_{n+1,i} = X_{n,i}$.

1.5.2 Gibbs 抽样方法

Gibbs **抽样方法** 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（全条件分布）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 算法利用每个变量的全条件分布进行抽样，基本步骤如下

记 $X_1, \ldots, X_p$ 的联合密度为 $f(x_1, \ldots, x_p)$,

$$X_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_p)$$

并记 $X_i|X_{-i}$ 的 **（全）条件密度**为

$$f(x_i|\mathbf{x}_{-i}) = f(x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p)$$

(Systematic scan) Gibbs 算法从一个初始值 $\mathbf{x}^{(0)} = (x_1^{(0)}, \ldots, x_p^{(0)})$ 出发, 迭代产生

$$
\begin{aligned}
x_1^{(t)} &\sim f(x_1 | x_2^{(t-1)}, x_3^{(t-1)}, \ldots, x_p^{(t-1)}) \\
x_2^{(t)} &\sim f(x_2 | x_1^{(t)}, x_3^{(t-1)}, \ldots, x_p^{(t-1)}) \\
&\vdots \\
x_i^{(t)} &\sim f(x_i | x_1^{(t)}, \ldots, x_{i-1}^{(t)}, x_{i+1}^{(t-1)}, \ldots, x_p^{(t-1)}) \\
&\vdots \\
x_p^{(t)} &\sim f(x_p | x_1^{(t)}, \ldots, x_{p-1}^{(t)})
\end{aligned}
$$

(Random scan) Gibbs 算法从一个初始值 $\mathbf{x}^{(0)} = (x_1^{(0)}, \ldots, x_p^{(0)})$ 出发, 对 $t = 1, 2, \ldots, T$ 迭代产生

1. 从 $\{1, 2, \ldots, p\}$ 中随机抽取 j

2. 抽取 $X_j^{(t)} \sim f(x_j | x_1^{(t-1)}, \ldots, x_{j-1}^{(t-1)}, x_{j+1}^{(t-1)}, \ldots, x_p^{(t-1)})$, 并对 $k \neq j$, 令 $X_k^{(t)} := X_k^{(t-1)}$

问题

1. Gibbs 抽样器只使用全条件分布, 那么全条件分布是否可以确定联合分布?

2. 生成的样本序列为马氏链, 目标联合分布 f 是否为该链的唯一平稳分布? 使用全部样本还是后面的样本进行推断?

设随机变量 $X = (X_1, \ldots, X_p)$ 的联合密度为 $f(x_1, \ldots, x_p)$，边际密度为 $f_i(x_i)$，称 X 满足正性条件，如果 $f_i(x_i) > 0 (i = 1, \ldots, p)$ 的意味着 $f(x_1, \ldots, x_p) > 0$.

Definition

定理 4. *(Hammersley-Clifford)* 设 $(X_1, \ldots, X_p)$ 满足正性条件，且有联合密度 $f(x_1, \ldots, x_p)$. 则对所有 $(\xi_1, \ldots, \xi_p) \in supp(f)$ 有

$$f(x_1, \ldots, x_p) \propto \prod_{j=1}^{p} \frac{f(x_j | x_1, \ldots, x_{j-1}, \xi_{j+1}, \ldots, \xi_p)}{f(\xi_j | x_1, \ldots, x_{j-1}, \xi_{j+1}, \ldots, \xi_p)}$$

注意此定理并没有保证对任意组全条件分布存在联合分布.

设 $X_1|X_2 \sim Exp(\lambda X_2)$, $X_2|X_1 \sim Exp(\lambda X_1)$, 考察联合分布.

↑Example
↓Example

由 Hammersley-Clifford 定理, 我们有

$$\begin{aligned} f(x_1, x_2) &\propto \frac{f_{X_1|X_2}(x_1|\xi_2) \cdot f_{X_2|X_1}(x_2|x_1)}{f_{X_1|X_2}(\xi_1|\xi_2) \cdot f_{X_2|X_1}(\xi_2|x_1)} \\ &\propto \exp(-\lambda x_1 x_2) \end{aligned}$$

而 $\iint \exp(-\lambda x_1 x_2)\, dx_1 dx_2 = +\infty$, 即在给定全条件密度下不存在联合密度.

显然,Gibbs 抽样下的转移核为

$$
\begin{aligned}
K\left(\mathbf{x}^{(t-1)}, \mathbf{x}^{(t)}\right) = & f\left(x_1^{(t)} | x_2^{(t-1)}, \ldots, x_p^{(t-1)}\right) \\
& \cdot f\left(x_2^{(t)} | x_1^{(t)}, x_3^{(t-1)}, \ldots, x_p^{(t-1)}\right) \\
& \cdots \\
& \cdot f\left(x_p^{(t)} | x_1^{(t)}, \ldots, x_{p-1}^{(t)}\right)
\end{aligned}
$$

则可以证明, 如果联合密度 $f(x_1, \ldots, x_p)$ 满足正性条件, 则 Gibbs 抽样器产生一个不可约, 常返的马氏链, f 为该马氏链的平稳分布.

数据扩张

- Gibbs 抽样方法只有在可以从全条件密度中容易进行抽样时候才是可行的
- 一种有助于使得全条件密度容易抽样的方法是**反边际化** (demarginalisation): 引入辅助随机变量 $Z_1, \ldots, Z_r$, 使得 f 为 $(X_1, \ldots, X_p, Z_1, \ldots, Z_r)$ 的边际密度, 即

$$f(x_1, \ldots, x_p) = \int f_{X,Z}(x_1, \ldots, x_p, z_1, \ldots, z_r)\, d(z_1, \ldots, z_r)$$

  辅助变量的选取使得 $f_{X,Z}$ 下的全条件密度容易抽样.
- 在许多情形下, 辅助变量 $Z_1, \ldots, Z_r$ 是" 自然" 的

设样本 $X_1, \ldots, X_n$ 来自混合密度 $f(x) = \sum_{k=1}^{K} \pi_k \phi_{(\mu_k, 1/\tau)}(x)$, 其中 K, τ 已知, ϕ 为正态密度函数. $(\pi_1, \ldots, \pi_K) \sim Dirichlet(\alpha_1, \ldots, \alpha_K)$, $\mu_1, \ldots, \mu_K$ $i.i.d \sim N(\mu_0, 1/\tau_0)$. 求参数 $\pi_1, \ldots, \pi_K, \mu_1, \ldots, \mu_K$.

↑Example

↓Example

容易写出后验密度

$$f(\mu_1, \ldots, \mu_K, \pi_1, \ldots, \pi_K | x_1, \ldots, x_n) \propto \left(\prod_{k=1}^{K} \pi_k^{\alpha_k - 1}\right) \cdot \left(\prod_{k=1}^{K} \exp\left(-\tau_0 (\mu_k - \mu_0)^2 / 2\right)\right) \cdot \left(\sum_{k=1}^{K} \pi_k \exp\left(-\tau (x_i - \mu_k)^2 / 2\right)\right)$$

该联合密度看起来形式复杂, 其全条件密度不易抽样.

使用数据扩张方法: 引入辅助变量 $Z_1, \ldots, Z_n$ 表示每个观测来自的总体标签, 即

$$P(Z_i = k) = \pi_k \qquad X_i | Z_i = k \sim N(\mu_k, 1/\tau)$$

此时联合密度函数为

$$
\begin{aligned}
& f\left(x_1, \ldots, x_n, z_1, \ldots, z_n, \mu_1, \ldots, \mu_K, \pi_1, \ldots, \pi_K\right) \\
\propto & \left(\prod_{k=1}^{K} \pi_k^{\alpha_k - 1}\right) \cdot \left(\prod_{k=1}^{K} \exp\left(-\tau_0 \left(\mu_k - \mu_0\right)^2 / 2\right)\right) \\
& \cdot \left(\prod_{i=1}^{n} \pi_{z_i} \exp\left(-\tau \left(x_i - \mu_{z_i}\right)^2 / 2\right)\right)
\end{aligned}
$$

因此

$$
\begin{aligned}
& P(Z_i = k | x_1, \ldots, x_n, \mu_1, \ldots, \mu_K, \pi_1, \ldots, \pi_K) \\
& \quad = \frac{\pi_k \phi_{(\mu_k, 1/\tau)}(x_i)}{\sum_{t=1}^{K} \pi_k \phi_{(\mu_i, 1/\tau)}(x_i)}
\end{aligned}
$$

$$\mu_k|x_1,\ldots,x_n,Z_1,\ldots,Z_n,\pi_1,\ldots,\pi_K$$
$$\sim N\left(\frac{\tau\left(\sum_{i:Z_i=k}x_i\right)+\tau_o\mu_0}{|\{i:Z_i=k\}|\,\tau+\tau_0},\frac{1}{|\{i:Z_i=k\}|\,\tau+\tau_0}\right)$$

$$\pi_1,\ldots,\pi_K|x_1,\ldots,x_n,Z_1,\ldots,Z_n,\mu_1,\ldots,\mu_K$$
$$\sim \text{ Dirichlet } (\alpha_1+|\{i:Z_i=1\}|,\ldots,\alpha_K+|\{i:Z_i=K\}|)$$

因此, Gibbs 抽样算法如下:
选取初始值 $\mu_1^{(0)},\ldots,\mu_K^{(0)},\pi_1^{(0)},\ldots,\pi_K^{(0)}$, 迭代 $t=1,2,\ldots,T$:
1. 对 $i=1,\ldots,n$: 从离散分布

$$P\left(Z_i^{(t)}=k\right)=\frac{\pi_k^{(t-1)}\phi_{\left(\mu_k^{(t-1)},1/\tau\right)}(x_i)}{\sum_{l=1}^{K}\pi_l^{(t-1)}\phi_{\left(\mu_i^{(t-1)},1/\tau\right)}(x_i)},k=1,\ldots,K$$

中抽取 $Z_i^{(t)}$.

2. 对 $k = 1, \ldots, K$, 抽取

$$\mu_k^{(t)} \sim N\left(\frac{\tau\left(\sum_{i:Z_i^{(t)}=k} x_i\right) + \tau_o\mu_0}{\left|\left\{i : Z_i^{(t)} = k\right\}\right|\tau + \tau_0}, \frac{1}{\left|\left\{i : Z_i^{(t)} = k\right\}\right|\tau + \tau_0}\right)$$

3. 抽取

$$(\pi_1^{(t)}, \ldots, \pi_K^{(t)})$$
$$\sim \text{ Dirichlet } \left(\alpha_1 + \left|\left\{i : Z_i^{(t)} = 1\right\}\right|, \ldots, \alpha_K + \left|\left\{i : Z_i^{(t)} = K\right\}\right|\right)$$

1.5.3 Hamiltonian Monte Carlo

许多 MCMC 算法中的随机游走做法使得马氏链收敛到平稳分布非常慢. Hamiltonian/Hybrid Monte Carlo (HMC) 是一种 MCMC 方法, 它采用物理系统动力学而不是概率分布来产生马氏链的未来状态提议值. 这使马氏链可以更有效地探索目标分布, 从而加快收敛速度.

Hamiltonian systems

在一个哈密顿系统中, 我们考虑一个物体在时刻 t 处于位置 $\mathbf{x}$ 以及具有动量 (或速度, 如果假设单位质量) $\mathbf{v}$. 此时系统总能量为 $H(\mathbf{x}, \mathbf{v}) = U(\mathbf{x}) + K(\mathbf{v})$, 其中 U 为势能, K 是动能. 此类系统满足以下哈密顿方程组 (描述动能和势能的转换):

$$\frac{dx_i}{dt} = \frac{\partial H}{\partial v_i} = \frac{\partial K(\mathbf{v})}{\partial v_i}$$
$$\frac{dv_i}{dt} = -\frac{\partial H}{\partial x_i} = -\frac{\partial U(\mathbf{x})}{\partial x_i}$$

因为 U 仅依赖于 x, K 仅依赖于 v.

因此当我们有了 $\frac{\partial U(\mathbf{x})}{\partial x_i}$ 和 $\frac{\partial K(\mathbf{v})}{\partial v_i}$ 的表达式, 以及一组初始条件 (比如在初始时刻 t_0 的位置 $\mathbf{x}_0$ 和初始动量 $\mathbf{v}_0$), 则有可能通过模拟系统在时长 T 内的演变来预测物体在 $t = t_0 + T$ 时刻的位置和动量.

模拟哈密顿动力学: Leap Frog 方法

哈密顿方程描述了物体随连续时间的运动. 为了在计算机上数值模拟哈密顿动力学, 必须通过离散化时间来近似哈密顿方程. 我们将时长 T 分成一系列较小的长度为 δ 的间隔. 离散化时间的方法有多种, 包括 Euler 方法和 Leap Frog 方法. 下面我们介绍 Leap Frog 方法:

1. 首先使用 Taylor 展开, 计算

$$v_i(t + \delta/2) = v_i(t) - (\delta/2)\frac{\partial U(\mathbf{x})}{\partial x_i(t)}$$

2. 更新位置

$$x_i(t+\delta) = x_i(t) + \delta \frac{\partial K(\mathbf{v})}{\partial v_i(t+\delta/2)}$$

3. 更新动量

$$v_i(t+\delta) = v_i(t+\delta/2) - (\delta/2) \frac{\partial U(\mathbf{x})}{\partial x_i(t+\delta)}$$

Hamiltonian / Hibrid Monte Carlo 的主要思想是构造哈密顿函数 $H(\mathbf{x}, \mathbf{v})$, 使得所得哈密顿动力学能够有效地探索目标分布 $p(\mathbf{x})$. 如何选择这样的哈密顿函数? 使用统计力学的规范分布 (canonical distribution) 概念将 $H(\mathbf{x}, \mathbf{v})$ 与 $p(\mathbf{x})$ 关联起来非常简单:

$$\begin{aligned} p(\mathbf{x}, \mathbf{v}) &\propto e^{-H(\mathbf{x}, \mathbf{v})} \\ &= e^{-U(\mathbf{x})} e^{-K(\mathbf{v})} \\ &= p(\mathbf{x}) p(\mathbf{v}) \end{aligned}$$

从而可以使用哈密顿动力学来对 $\mathbf{x}$ 和 $\mathbf{v}$ 的联合分布进行抽样, 抽样后只需要 $\mathbf{x}$ 而忽略掉 $\mathbf{v}$ 即可.(类似于数据扩张里的辅助变量) 由于 $\mathbf{x}$ 和 $\mathbf{v}$ 相互独立, 我们可以使用任何分布来抽样得到 $\mathbf{v}$. 正态分布是一种常用的分布:

$$p(\mathbf{v}) \propto e^{-\mathbf{v}'\mathbf{v}/2}$$

这等价于取 $K(\mathbf{v}) = \mathbf{v}'\mathbf{v}/2$.

因此, 给定目标分布 $p(\mathbf{x})$, 我们可以定义

$$U(\mathbf{x}) = -\log p(\mathbf{x})$$

如果偏导数 $\frac{\partial U(\mathbf{x})}{\partial x_i}$ 可以得到, 则可以使用上述 leap frog 算法进行模拟哈密顿动力学.

在哈密顿蒙特卡洛 (HMC) 中, 我们使用哈密顿动力学作为产生马氏链的提议分布. 选定初始值 $\mathbf{x}_0, \mathbf{v}_0$, 我们使用 leap frog 算法模拟一个短时间的哈密顿动力过程, 使用模拟的最后位置和动量作为提议

的状态变量 $\mathbf{x}^*, \mathbf{v}^*$, 计算接收概率

$$\begin{aligned}\alpha &= \min\{1, \frac{p(\mathbf{x}^*, \mathbf{v}^*)}{p(\mathbf{x}_0, \mathbf{v}_0)}\} \\ &= \min\{1, exp[-U(\mathbf{x}^*) + U(\mathbf{x}_0) - K(\mathbf{v}^*) + K(\mathbf{v}_0)]\}\end{aligned}$$

并随机接受. 如果被拒绝, 则保持状态不变. 为了尽可能跑遍目标分布所有区域, 我们可以通过随机化动量来变化总能量. 因此, HMC 方法总结如下:

HMC

1. 令 $t=0$, 产生初始位置 $\mathbf{x}^{(0)} \sim \pi^{(0)}$
2. 重复 $t=1,\ldots,T$
   - 令 $t=t+1$, 产生一个新的动量 $\mathbf{v}_0 \sim p(\mathbf{v})$
   - 令 $\mathbf{x}_0 = \mathbf{x}^{(t-1)}$
   - 使用 Leap Frog 算法, 从 $(\mathbf{x}_0, \mathbf{v}_0)$ 出发, 时长 L 和步长 δ, 得到 $\mathbf{x}^*, \mathbf{v}^*$
   - 计算 Metroplis 接收概率

$$\alpha = \min\{1, exp(-U(\mathbf{x}^*) + U(\mathbf{x}_0) - K(\mathbf{v}^*) + K(\mathbf{v}_0))\}$$

   - 以概率 α 接受 $\mathbf{x}^{(t)} = \mathbf{x}^*$, 否则 $\mathbf{x}^{(t)} = \mathbf{x}^{(t-1)}$.

HMC 方法实施中需要指定多余参数 L,δ 的值. HMC 使用势能 $U(\mathbf{x})$ 及其梯度, 因此需要梯度存在且计算时间可以接受. 相比于 MH 方法, HMC 方法

- 连续生成的两个状态之间的距离一般比较大, 因此抽取代表性样本所需迭代次数较少
- 单次迭代的成本较高 (需要模拟哈密顿动力学过程), 但是仍然显著地更有效
- 大部分情况下会接受新状态
- 当分布有分开的局部最小值时候无法进行全部抽样 (陷在一个局部最小值处), 可以考虑不同初始值.

其他的一些新方法参加阅读材料.

HMC 实施软件

Stan:

- No U-Turn Sampler (NUTS2): Adaptive Hamiltonian Monte Carlo
- Implemented in Stan (rstan: mc-stan.org)
- Stan figures out gradient for you via autodiff

R package hmclearn:

- takes user-defined log posterior and gradient functions as inputs
- includes parameters to enable parallel processing as well as multiple chains
- a variety of Bayesian graphical functions are provided