采样——MCMC

Mia Feng

2018 年 4 月 21 日

1 概述

MCMC：粗暴的采样模拟方式，用于模拟直接计算困难的分布。用于采样，数值积分等等。蒙特卡洛采样方法的一个重要好处是：估计值的精度与 x 的维度无关，而是与采样次数有关。实际问题中采样二十次左右就能达到较好的精度。

求解目标：用多次采样得到的频率分布近似原概率分布。即本来对复杂的 $f(x)$ 做积分，但是因为 $f(x)$ 比较复杂所以显式积分困难。迂回方法是构造统计量 $\frac{f(x)}{p(x)}$，通过对 $x \sim p(x)$ 进行采样（后文叙述中的目标分布 $\pi(x)$ 指的就是它），求取统计量 $\frac{f(x)}{p(x)}$ 的期望得到数值积分值。

$$\theta = \int_a^b f(x)\ dx = \int_a^b \frac{f(x)}{p(x)} p(x)\ dx \approx \frac{1}{n} \sum_{i=0}^{n-1} \frac{f(x_i)}{p(x_i)} \tag{1}$$

求解思路：微积分思想（recall：学习微积分的时候，用无数个划分的小矩形的面积来近似面积，但是当时的小矩形是来自均匀分布的）。实际上，来自均匀分布的可能性很小，此时需要考虑对复杂分布如何模拟采样，一旦完成对复杂分布的描述就可以完成数值积分。

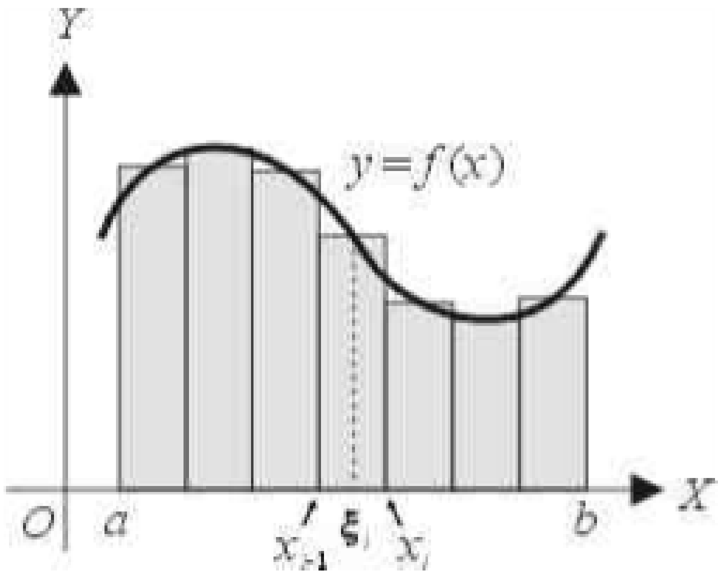

图 1: 矩形近似

求解方法：Markov Chain，蒙特卡洛积分，Metroplis-Hasting，Gibbs。
思维导图：详见 MCMC.xmind

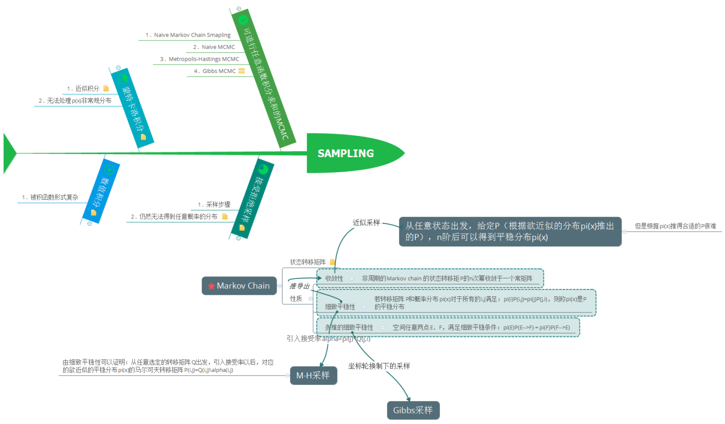

图 2: 思维导图

1.1 基础概念

提议密度（Proposal Density） 记提议密度为 $q\left(x\right)$，其需要满足：

- 对 $q\left(x\right)$ 采样比较容易。
- $q\left(x\right)$ 的形状接近 $\pi\left(x\right)$，具有

$$\pi\left(x\right)\leq Mq\left(x\right),\forall x \tag{2}$$

其中 k 为常数。

1.2 马尔科夫链的性质

马尔可夫矩阵的收敛性 可以参考 MIT 的线性代数课程中对马尔可夫矩阵的讲述。马尔可夫矩阵中各元素大于 0 且小于 1，而且矩阵是对称矩阵。马尔可夫矩阵的特征值中有一个为 1，其余都是比 1 小的正数，所以马尔可夫矩阵的 n 次幂收敛至一个常数，这也是为什么马尔科夫链一定会收敛，最终可以模拟一个平稳分布的原因。而且，每个马尔科夫链对应的稳定分布是唯一的。（唯一性证明请自行百度，尚未查阅这方面的资料）

马尔科夫链的细致平稳条件 如果非周期马尔科夫链的状态转移矩阵 P 和概率分布 $\pi\left(x\right)$ 对于所有的 i,j 满足

$$\pi\left(i\right)P\left(i,j\right)=\pi\left(j\right)P\left(j,i\right) \tag{3}$$

则称概率分布 $\pi\left(x\right)$ 是状态转移矩阵 P 的平稳分布。所以，利用对应的状态转移矩阵 P，就可以模拟平稳复杂分布 π。但是对于任意的平稳分布 π，P 的构造比较困难。MCMC 采用迂回的方式解决了这个问题。

多维数据的马尔可夫链的细致平稳条件 平面上任意两点 E,F，满足细致平稳条件

$$\pi\left(E\right)P\left(E\rightarrow F\right)=\pi\left(F\right)P\left(F\rightarrow E\right) \tag{4}$$

取上一状态的条件概率分布即可作为马尔科夫链的状态转移概率。固定 $n-1$ 维，只改变一维。这样按坐标轴轮换完成整个特征维的状态转移后，算完成一次。

1.3 采样方法

接受——拒绝采样 可以参考 Metropolis 接受准则，核心思想是当能量增加时以一定概率接收，而不是一味的拒绝。

当公式 (1) 中的 $p(x)$ 不是常见分布时，无法根据 $p(x)$ 对 x 直接进行采样。此时接受——拒绝采样可以完成对 x 的模拟采样。

选定提议密度 $q(x)$，如高斯分布、均匀分布等。先根据提议分布 $q(x)$ 采样得到一个样本 x_0，其对应的实际的概率是 u_0。再从均匀分布 $(0, kq(x_0))$ 中采样得到一值 u，若 $u < u_0$，接受这次采样值 x_0，反之拒绝。步骤为：

- 产生样本 $x_0 \sim q(x)$ 以及 $u \sim Uniform(0, kq(x_0))$。
- 若 $u \leq p(x_0)$，则接受 x_0，反之拒绝。

注：为与 MCMC 中的接受率叙述保持一致，这里对接受拒绝准则重述如下：

步骤为：

- 产生样本 $x_0 \sim q(x)$ 以及 $u \sim Uniform(0, 1)$。
- 若 $u \leq \frac{p(x_0)}{kq(x_0)}$，则接受 x_0，反之拒绝。

其中，接受率 $\alpha = \frac{p(x_0)}{kq(x_0)}$，是图 3 中红色曲线的面积与蓝色曲线的面积的比值。其理论值为

$$\alpha^* = \frac{\int_a^b p(x)\ dx}{k\int_a^b q(x)\ dx} = \frac{1}{k} \tag{5}$$

当然，$k, q(x)$ 的选取要满足提议密度的性质：$kq(x) \geq p(x)$。但是满足这个条件比较困难，MCMC 中 Metropolis-Hastings 方法将解决这个问题。

而且，高维空间的 k 将非常大，计算所得的接受率极小，相应的拒绝率会非常高，这导致了采样效率不高 [**?**]。

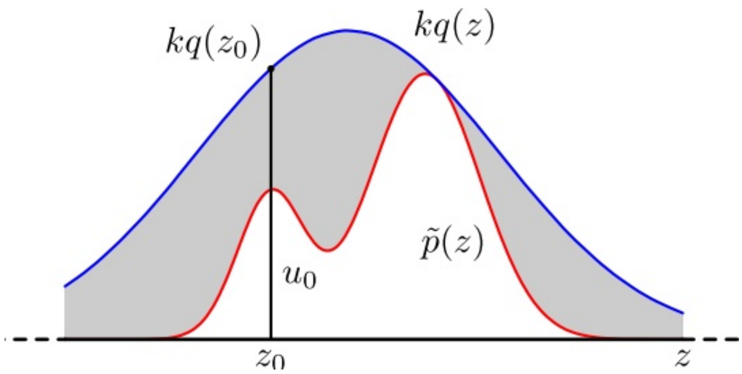

图 3: 接受拒绝采样

重要性采样 为了提升简单接受——拒绝采样方法的采样效率。无需严格满足 $p(x) \le kq(x)$。

$$\theta = \int_a^b f(x)\ dx = \int_a^b \frac{f(x)}{p(x)} \frac{p(x)}{q(x)} q(x)\ dx \tag{6}$$

重要性 w 记为：

$$w = \frac{p(x)}{q(x)} \tag{7}$$

其余的定义不变。

则原数值积分值 θ 的近似计算步骤为：

- 产生样本 $x_1, \cdots, x_n \sim q(x)$。
- 计算对积分值 θ 的无偏估计为

$$\hat{I} = \frac{1}{n} \sum_{i=1}^{n} \frac{f(x_i)}{p(x_i)} w(x_i) \tag{8}$$

其中 $w(x_i) = \frac{p(x_i)}{q(x_i)}$ 实际中常用有偏估计来近似积分值 θ，公式为：

$$\hat{I} = \frac{\sum\limits_{i=1}^{n} \frac{f(x_i)}{p(x_i)} w(x_i)}{\sum\limits_{i=1}^{n} w(x_i)} \tag{9}$$

与接受——拒绝采样中类似，这里的提议密度 $q(x)$ 也是任意的。只是当 $q(x)$ 与 $p(x)$ 更接近，并且 $q(x)$ 尾部较厚的时候方差更小（更近均匀

分布)，提议密度的采样效率更高。(这里不再需要满足 $p(x) \leq kq(x)$)。当然，只有当 $q(x)$ 与 $p(x)$ 更接近时，$w(x)$ 的取值范围会更小一些，否则其取值范围还是很大，还是没法完成对任意分布的近似。

马尔科夫链采样 选定平稳分布 $\pi(x)$ 的状态转移矩阵 P 之后，根据状态转移矩阵 P，不断完成上一状态到下一状态的转移得到此时的条件概率，利用此时的条件概率得到采样值，重复这个转移过程直至条件概率收敛。收敛之后的采样集可以作为采样结果输出用于蒙特卡洛模拟求和。设计不同的马尔科夫状态转移矩阵 P（下文中记为 Q)，就得到了下述几种不同的采样方法。

MCMC 采样——Metropolis 采样 接受——拒绝原理是相同的，与 MH 采样方法不同在 Metropolis 的转移矩阵 Q 是对称矩阵，即由上一状态转移到当前状态和由当前状态转移到上一状态的概率是相等的。

通过比较采样自均匀分布 $(0,1)$ 的值 u 与接受率 $\alpha(i,j)$，决定是否接受转移。

因为采用马尔科夫链，这里的接受率定义为：

$$\alpha = \min\left(1, \frac{\pi(j)}{\pi(i)}\right) \tag{10}$$

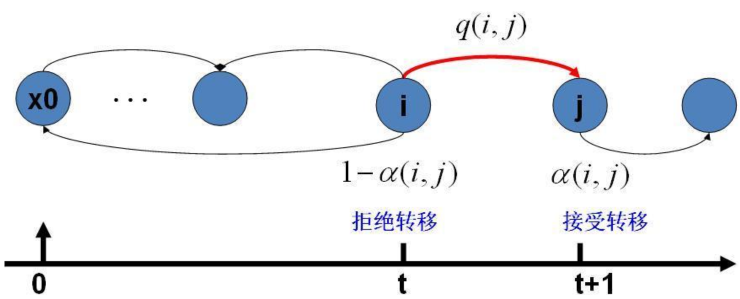

图 4: 马氏链转移和接受概率

MCMC 采样——Metropolis-Hastings 采样 接受——拒绝原理是相同的，但是这里的提议分布是根据马尔科夫链状态转移后得到的。状态转移矩阵 Q 是非对称矩阵。Metropolis 采样可以看做 Metropolis-

Hastings 采样的一个特例。接受率定义为：

$$\alpha = \min\left(1, \frac{\pi(j)\,Q(j,i)}{\pi(i)\,Q(i,j)}\right) \tag{11}$$

MCMC 采样——Gibbs 采样 与马尔可夫随机场模型比较接近为什么待补充。无需计算接受率，所以采样效率提升。无需寻找合适的提议密度。适用于高维特征下的采样。具体步骤如下：

1）输入平稳分布$\pi(x_1, x_2, \ldots, x_n)$或者对应的所有特征的条件概率分布，设定状态转移次数阈值n_1，需要的样本个数n_2

2）随机初始化初始状态值$(x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}$

3）for $t = 0$ to $n_1 + n_2 - 1$:

a) 从条件概率分布$P(x_1|x_2^{(t)}, x_3^{(t)}, \ldots, x_n^{(t)})$中采样得到样本$x_1^{t+1}$

b) 从条件概率分布$P(x_2|x_1^{(t+1)}, x_3^{(t)}, x_4^{(t)}, \ldots, x_n^{(t)})$中采样得到样本$x_2^{t+1}$

c)...

d) 从条件概率分布$P(x_j|x_1^{(t+1)}, x_2^{(t+1)}, \ldots, x_{j-1}^{(t+1)}, x_{j+1}^{(t)} \ldots, x_n^{(t)})$中采样得到样本$x_j^{t+1}$

e)...

f) 从条件概率分布$P(x_n|x_1^{(t+1)}, x_2^{(t+1)}, \ldots, x_{n-1}^{(t+1)})$中采样得到样本$x_n^{t+1}$

样本集$\{(x_1^{(n_1)}, x_2^{(n_1)}, \ldots, x_n^{(n_1)}), \ldots, (x_1^{(n_1+n_2-1)}, x_2^{(n_1+n_2-1)}, \ldots, x_n^{(n_1+n_2-1)})\}$即为我们需要的平稳分布对应的样本集。

图 5: Gibbs 采样步骤

2 算法实现

2.1 验证马尔科夫链的稳定性

代码：SteadyMarkovTranforMatrix

2.2 Metropolis 采样

原文中说的是代码实现的是 MH 采样 [?]，但根据 Metropolis 算法和 MH 算法的区别（转移矩阵 Q 是否对称，我认为其代码实际实现的是 Metropolis 算法）代码：Metropolis.py

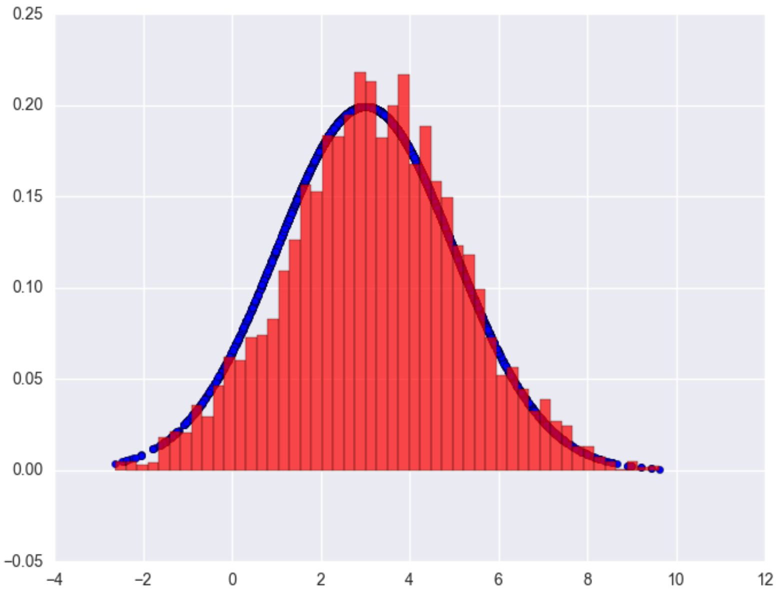

图 6: Metropolis 采样分布直方图

2.3 二维 Gibbs 采样

代码：Gibbs.py

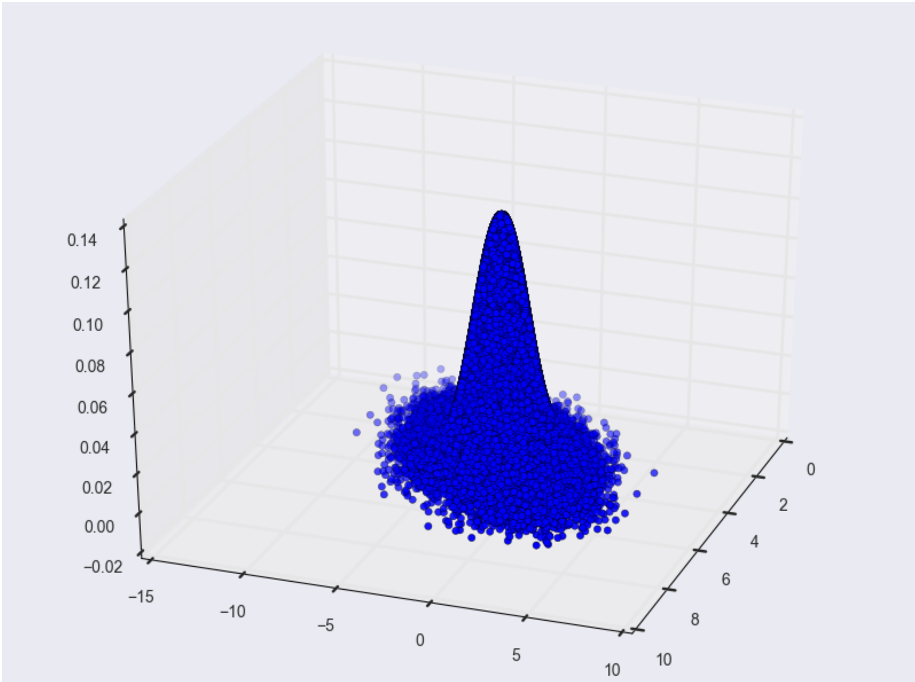

图 7: 二维分布散点图

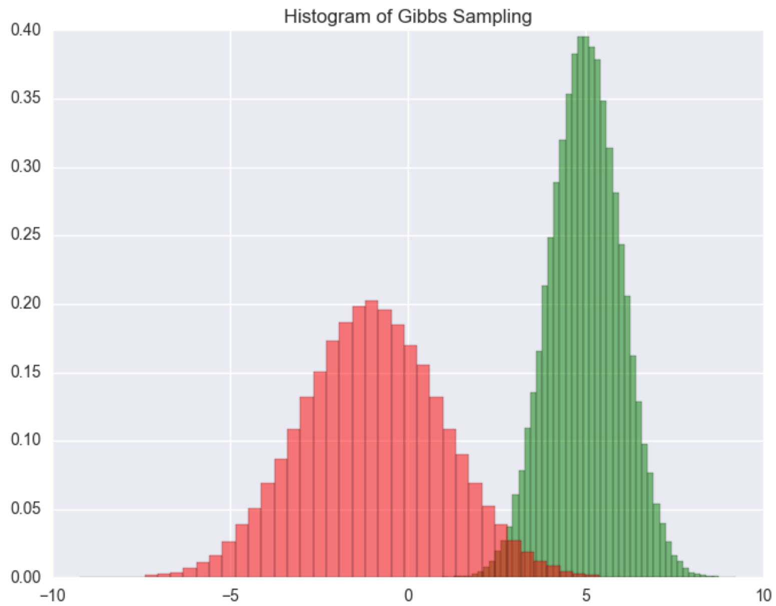

图 8: Gibbs 采样分布直方图

参考文献

[1] 刘建平 Pinard. Mcmc 采样. https://zhuanlan.zhihu.com/p/30003899. Blog, 2017.

[2] 徐晔烨. 抽样方法. https://blog.csdn.net/ustbxy/article/details/45458725. Blog, 2015.