

在[MCMC采样和M-H采样](#)中，我们讲到M-H采样已经可以很好的解决蒙特卡罗方法需要的任意概率分布的样本集问题。但是M-H采样有两个缺点：一是需要计算接受率，在高维情况下计算量非常大，同时由于接受率的原因导致算法收敛时间变长。二是有些高维数据，特征的条件概率分布方便求解，但特征的联合分布很难求解。因此需要改进M-H算法，来解决上面提到的两个问题，下面我们详细介绍Gibbs采样方法。

1.细致平衡条件

[MCMC采样和M-H采样](#)中我们讲到细致平衡条件，即如果非周期马尔可夫链状态转移矩阵P和概率分布 $\pi(x)$ 对于所有的i,j满足下列方程，则称概率分布 $\pi(x)$ 是状态转移矩阵P的平稳分布。

$$\pi(i)P(i, j) = \pi(j)P(j, i)$$

在M-H采样中我们通过引入接受率使细致平稳条件满足，现在我们换一种方法。从二维数据的分布开始，假设 $\pi(x_1, x_2)$ 是一个二维联合的数据分布，观察第一个特征维度相同的两个点 $A(x_1^{(1)}, x_2^{(1)})$ 和 $B(x_1^{(1)}, x_2^{(2)})$ ，容易发现下面两式成立

$$\begin{aligned}\pi(x_1^{(1)}, x_2^{(1)})\pi(x_2^{(2)}|x_1^{(1)}) &= \pi(x_1^{(1)})\pi(x_2^{(1)}|x_1^{(1)})\pi(x_2^{(2)}|x_1^{(1)}) \\ \pi(x_1^{(1)}, x_2^{(2)})\pi(x_2^{(1)}|x_1^{(1)}) &= \pi(x_1^{(1)})\pi(x_2^{(1)}|x_1^{(1)})\pi(x_2^{(2)}|x_1^{(1)})\end{aligned}$$

可以发现，上面两式的右边相等，因此我们有

$$\pi(x_1^{(1)}, x_2^{(1)})\pi(x_2^{(2)}|x_1^{(1)}) = \pi(x_1^{(1)}, x_2^{(2)})\pi(x_2^{(1)}|x_1^{(1)})$$

也就是

$$\pi(A)\pi(x_2^{(2)}|x_1^{(1)}) = \pi(B)\pi(x_2^{(1)}|x_1^{(1)})$$

通过上式，可以发现在 $x_1^{(1)}$ 这条直线上，如果用条件概率分布 $\pi(x_2|x_1^{(1)})$ 作为马尔可夫链的状态转移概率，则任意两个点之间的转移满足细致平稳条件。同样，在 $x_2 = x_2^{(1)}$ 这条直线上，如果用条件概率分布 $\pi(x_1|x_2^{(1)})$ 作为马尔可夫链的状态转移概率，则任意两个点之间的状态转移也满足细致平稳条件。基于上面的发现，我们可以构造分布 $\pi(x_1, x_2)$ 的马尔可夫链对应的状态转移矩阵P

$$\begin{aligned}if \ x_1^{(A)} = x_1^{(B)} = x_1^{(1)} \quad P(A \rightarrow B) &= \pi(x_2^{(B)}|x_1^{(1)}) \\ if \ x_2^{(A)} = x_2^{(C)} = x_2^{(1)} \quad P(A \rightarrow C) &= \pi(x_1^{(C)}|x_2^{(1)}) \\ else \quad P(A \rightarrow D) &= 0\end{aligned}$$

有了上面状态转移矩阵，我们很容易验证平面上任意两点E,F可以满足细致平稳条件

$$\pi(E)P(E \rightarrow F) = \pi(F)P(F \rightarrow E)$$

2.二维Gibbs采样

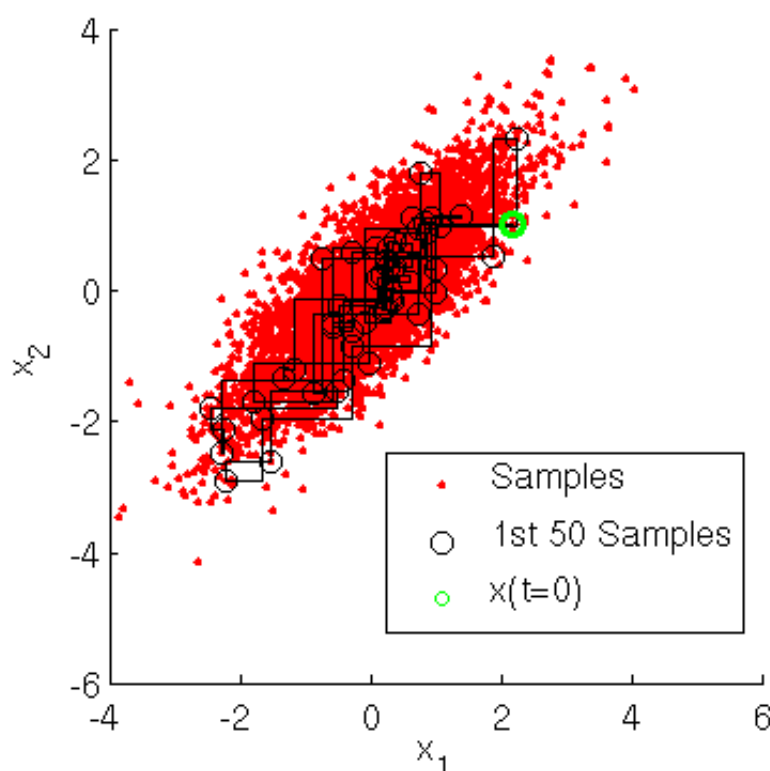
根据上面提到的状态转移矩阵，我们就可以得到二维Gibbs采样，这个采样需要两维度之间的条件概率，具体过程如下

- 输入平稳分布 $\pi(x_1, x_2)$ ，设定状态转移次数阈值 n_1 ，需要的样本个数 n_2 。
- 随机初始化初始状态值 $x_1^{(0)}$ 和 $x_2^{(0)}$ 。
- for t=0 to n_1+n_2-1
 - 从条件概率分布 $P(x_2|x_1^{(t)})$ 中采样得到样本 $x_2^{(t+1)}$ 。
 - 从条件概率分布 $P(x_1|x_2^{(t+1)})$ 中采样得到样本 $x_1^{(t+1)}$ 。

样本集 $\{(x_1^{(n_1)}, x_2^{(n_1)}), (x_1^{(n_1+1)}, x_2^{(n_1+1)}), \dots, (x_1^{(n_1+n_2-1)}, x_2^{(n_1+n_2-1)})\}$ 即为我们需要的平稳分布所对应的样本集。整个采样过程中，通过轮换坐标轴的方式进行采样，采样的过程为

$$(x_1^{(1)}, x_2^{(1)}) \rightarrow (x_1^{(1)}, x_2^{(2)}) \rightarrow (x_1^{(2)}, x_2^{(2)}) \rightarrow \dots \rightarrow (x_1^{(n_1+n_2-1)}, x_2^{(n_1+n_2-1)})$$

用下图可以直观的看出，采样是在两个坐标轴上不断变换的。当然，坐标轴轮换不是必须的，也可以每次随意选择一个坐标轴进行采样。



3.多维Gibbs采样

通过上面对二维Gibbs采样的介绍，我们同样可以推广到多维的情况。比如n维的概率分布 $\pi(x_1, x_2, \dots, x_n)$ ，可以通过在n个坐标轴轮换采样，来得到新的样本。对于轮换的任意一个坐标轴 x_i 的转移，马尔可夫链的状态转移概率为 $P(x_i|x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ 。即固定n-1个坐标轴，在某一个坐标轴上移动。具体算法过程如下

- 输入平稳分布 $\pi(x_1, x_2, \dots, x_n)$ 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 。
- 随机初始化初始状态值 $x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}$ 。
- for t=0 to n_1+n_2-1
 - 从条件概率分布 $P(x_1|x_2^{(t)}, x_3^{(t)}, \dots, x_n^{(t)})$ 中采样得到样本 $x_1^{(t+1)}$ 。
 - 从条件概率分布 $P(x_2|x_1^{(t+1)}, x_3^{(t)}, \dots, x_n^{(t)})$ 中采样得到样本 $x_2^{(t+1)}$ 。

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

样本集 $\{(x_1^{(n1)}, x_2^{(n1)}, \dots, x_n^{(n1)}), \dots, (x_1^{(n1+n2-1)}, x_2^{(n1+n2-1)}, \dots, x_n^{(n1+n2-1)})\}$ 即为我们需要的平稳分布所对应的样本集。同样轮换坐标轴不是必须的，可以随机选择某一个坐标轴进行转移，不过常用的Gibbs采样的实现都是基于坐标轴轮换的。

4. Gibbs采样总结

由于Gibbs采样在高维特征时的优势，目前通常意义上的MCMC采样都是用Gibbs采样。Gibbs采样要求数据至少有两个维度，一维概率分布的采样无法用Gibbs采样实现，这时可以用M-H方法采样。通过Gibbs采样来获取概率分布的样本集，通过蒙特卡罗方法来用样本集求和，两者一起奠定了MCMC算法在高维数据模拟求和时的作用。

5. 推广

更多内容请关注公众号谓之小一，若有疑问可在公众号后台提问，随时回答，欢迎关注，内容转载请注明出处。

「谓之小一」希望提供给读者别处看不到的内容，关于互联网、数据挖掘、机器学习、书籍、生活.....

- 知乎：@谓之小一
- 公众号：@谓之小一
- GitHub：@weizhixiaoyi
- 技术博客：<https://weizhixiaoyi.com>





精之小一

长按关注微信公众号

由锤子便签发送 via Smartisan Notes