在MCMC之马尔可夫链之中我们介绍到，给定一个概率分布π，很难直接找到对应的马尔可夫链状态转移矩阵P。只要解决这个问题，我们便可以找到一种通用的概率分布采样方法，进而用于蒙特卡罗模拟。下面我们来介绍如何找到马尔可夫链所对应的状态转移矩阵P。

1.马尔可夫链细致平稳条件

解决平稳分布π所对应的马尔可夫链状态转移矩阵P之前，我们先看一下马尔可夫链的细致平稳条件。其定义为：如果非周期马尔可夫链的状态转移矩阵P和概率分布π(x)对于所有的i,j满足下列方程，则概率分布π(x)是状态转移矩阵P的平稳分布。

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

证明如下，由细致平稳条件有

$$\sum_{i=1}^{\infty}\pi(i)P(i,j)=\sum_{i=1}^{\infty}\pi(j)P(j,i)=\pi(j)\sum_{i=1}^{\infty}P(j,i)=\pi(j)$$

将上式用矩阵表示为

$$\pi P=\pi$$

上式满足马尔可夫链的收敛性质，也就是说，只要我们找到可以使概率分布π(x)满足细致平稳分布的矩阵P即可。不过仅仅从细致平稳条件还是很难找到合适的矩阵P，比如我们的目标平稳分布使π(x)，随机找一个马尔可夫链状态转移矩阵Q，他是很难满足细致平稳条件的，即

$$\pi(i)Q(i,j)\neq\pi(j)Q(j,i)$$

那么有什么办法可以使这个等式相等呢?

2.MCMC采样

由于一般情况下，目标平稳分布π(x)和某一马尔可夫链状态转移矩阵Q不满足细致平稳条件，即

$$\pi(i)Q(i,j)\neq\pi(j)Q(j,i)$$

我们对上式进行一些变换，使细致平稳条件成立。方法是引入一个α(i,j)，使得上式等式能够成立，即

$$\pi(i)Q(i,j)\alpha(i,j)=\pi(j)Q(j,i)\alpha(j,i)$$

问题是什么样的α可以使上式成立？其实很简单，只要满足

$$\alpha(i,j)=\pi(j)Q(j,i);\ \alpha(j,i)=\pi(i)Q(i,j)$$

这样，我们便找到使分布π(x)对应的马尔可夫链状态转移矩阵P，满足

$$P(i,j)=Q(i,j)\alpha(i,j)$$

从上面可以得到，目标矩阵P可以通过任意一个马尔可夫链状态转移矩阵Q乘以α(i,j)得到。α(i,j)我们一般称之为接受率，取值在[0,1]之间，可以理解为一个概率值。也就是说，目标矩阵P可以通过任意一个马尔可夫链状态转移矩阵Q以一定的接受率得到。

其实很像我们在MCMC之蒙特卡罗方法中提到的接受-拒绝采样，那里是以常用分布通过一定的接受-拒绝概率得到一个非常见分布。这里是通过常见的马尔可夫链状态转移矩阵Q通过一定的接受-拒绝概率得到目标转移矩阵P，两者解决问题的思路是相同的。下面，我们来总结下MCMC的采样过程

- 输入任意选定的马尔可夫链状态转移矩阵Q，平稳分布π(x)，设定状态转移次数阈值n1，需要的样本个数n2。
- 从任意简单概率分布采样得到初始值x_0。
- for t=0 to n1+n2-1
  - 从条件概率分布$Q(x|x_t)$中采样得到样本x_*。
  - 从均匀分布采样u ~ uniform[0, 1]。
  - 如果$u < \alpha(x_t, x_*) = \pi(x_*)Q(X_*, x_t)$，则接受转移$x_t -> x_*$，即$x_{t+1} = x_*$。
  - 否则不接受转移，即$t = max(t-1, 0)$。

样本集$(x_{n1}, x_{n1+1}, ..., x_{n1+n2-1})$即为我们需要的平稳分布样本集。

上述过程便是MCMC采样理论，但很难在实际应用，为什么呢? 因为$\alpha(x_t, x_*)$可能非常小，比如0.1，导致大部分采样值都被拒绝转移，采样效率很低。可能我们采样可上百万次，马尔科夫链还没有收敛。实际应用中，我们可以通过M-H采样方法进行采样。

3.M-H采样

M-H采样解决了MCMC采样接受率过低的问题，我们首先回到MCMC采样的细致平稳条件

$$\pi(i)Q(i,j)\alpha(i,j) = \pi(j)Q(j,i)\alpha(j,i)$$

采样效率过低的原因是α(i,j)太小，比如0.1，α(j,i)为0.2，即

$$\pi(i)Q(i,j) * 0.1 = \pi(j)Q(j,i) * 0.2$$

如果两边同时扩大5倍，细致平稳条件仍然是满足的，即

$$\pi(i)Q(i,j) * 0.5 = \pi(j)Q(j,i) * 1$$

这样我们可以对接受率做如下改进，即

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

通过上述的转换，我们便可在实际应用中使用M-H算法进行采样，M-H采样算法过程如下所示

- 输入任意选定的马尔可夫链状态转移矩阵Q，平稳分布π(x)，设定状态转移次数阈值n1，需要的样本个数n2。
- 从任意简单概率分布采样得到初始值x_0。
- for t=0 to n1+n2-1
  - 从条件概率分布$Q(x|x_t)$中采样得到样本x_*。
  - 从均匀分布采样u ~ uniform[0, 1]。
  - 如果$u < \alpha(x_t, x_*) = min\{\frac{\pi(j)Q(j,i)}{\pi(i)Q(i,j)}, 1\}$，则接受转移$x_t -> x_*$，即$x_{t+1} = x_*$。
  - 否则不接受转移，即$t = max(t-1, 0)$。

样本集$(x_{n1}, x_{n1+1}, ..., x_{n1+n2-1})$即为我们需要的平稳分布样本集。

很多时候，我们选择的马尔科夫链状态转移矩阵Q如果是对称的，即满足Q(i,j)=Q(j,i)，这时我们的接受率可以进一步简化为

$$\alpha(i,j) = min(\frac{\pi(j)}{\pi(i)}, 1)$$

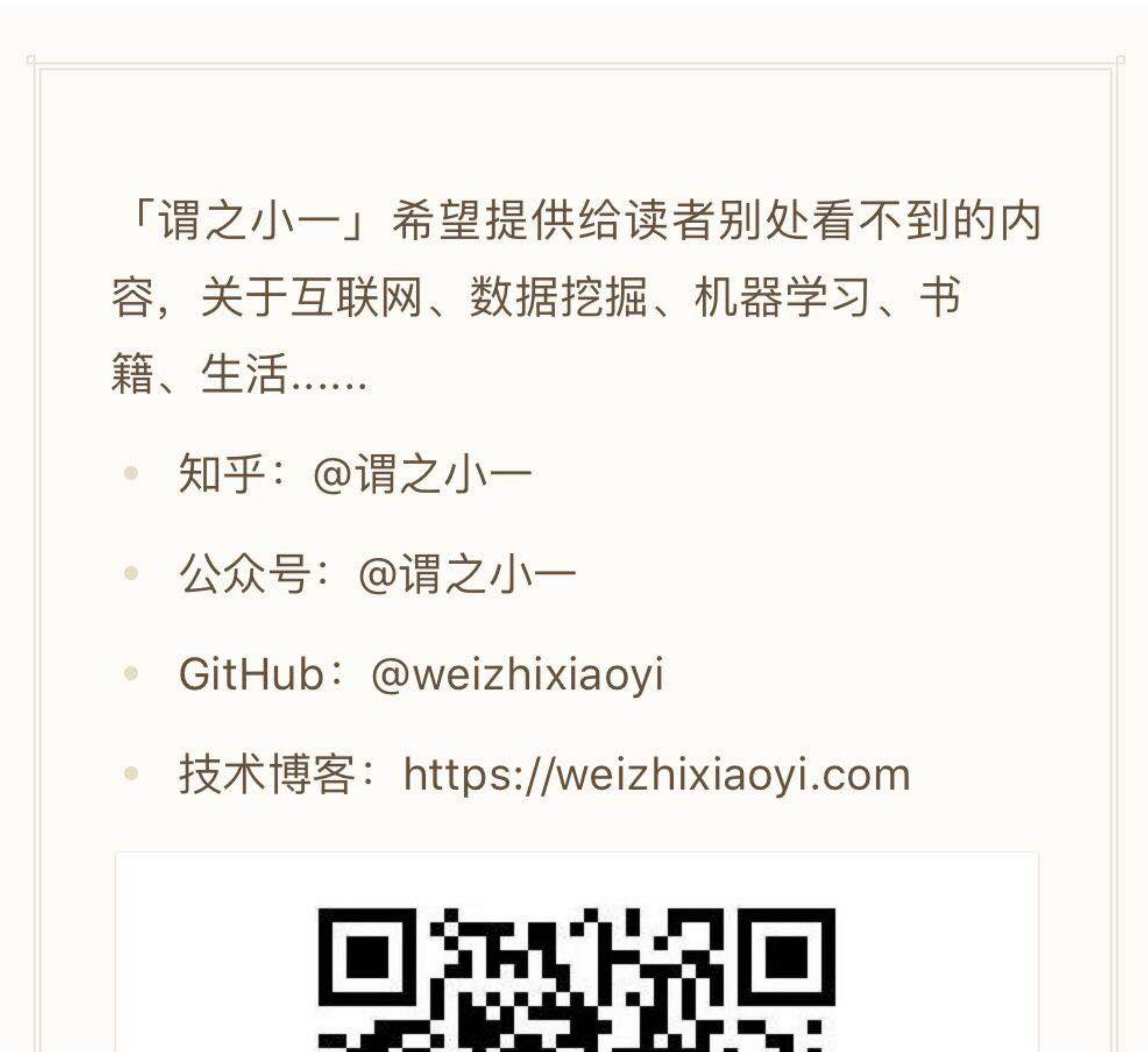
4.M-H采样总结

M-H采样解决了使用蒙特卡罗方法需要的任意概率分布样本集的问题，因此在实际生产环境中得到广泛应用。但在大数据情况下，M-H面临如下问题

- **数据特征非常多**：因为M-H采样由于接受率$\frac{\pi(j)Q(j,i)}{\pi(i)Q(i,j)}$的存在，在高维计算时需要很长的计算时间，算法效率很低。同时α(i,j)一般小于1，有时候辛苦计算出来的结果却被拒绝，能不能做到不拒绝转移呢?
- **特征维度比较大**：很多时候我们很难求出目标的各特征维度联合分布，但是可以方便求出各个特征之间的条件概率分布。这时能不能只用各维度之间的条件概率分布去方便的采样呢?

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。



长按关注微信公众号