[lda数学八卦mcmc-和-gibbs-sampling](http://www.flickering.cn/%E6%95%B0%E5%AD%A6%E4%B9%8B%E7%BE%8E/2014/06/lda%E6%95%B0%E5%AD%A6%E5%85%AB%E5%8D%A6mcmc-%E5%92%8C-gibbs-sampling/)

1.1 随机模拟

对那些用确定算法不可行或不可能解决的问题，蒙特卡罗方法常常为人们带来希望。

一般而言均匀分布 *Uniform*(0,1)的样本是相对容易生成的。 通过线性同余发生器可以生成伪随机数。

而我们常见的概率分布，无论是连续的还是离散的分布，都可以基于*Uniform*(0,1)的样本生成。例如正态分布可以通过著名的 Box-Muller 变换得到。

当*p*(*x*)的形式很复杂，或者 *p*(**x**) 是个高维的分布的时候，样本的生成就可能很困难了。

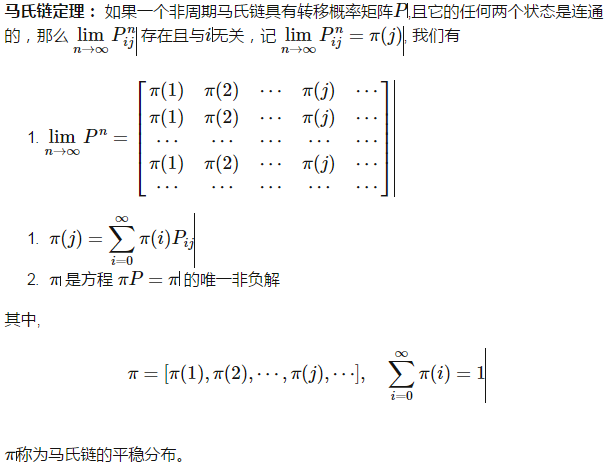
1.2 马氏链

 MCMC(Markov Chain Monte Carlo) **马氏链及其平稳分布**

马氏链的数学定义很简单

*P*(*Xt*+1=*x*|*Xt*,*Xt*−1,⋯)=*P*(*Xt*+1=*x*|*Xt*)

也就是状态转移的概率只依赖于前一个状态。



两个状态*i*,*j*是连通并非指*i* 可以直接一步转移到*j*(*Pij*>0),而是指 *i* 可以通过有限的*n*步转移到达*j*(*Pnij*>0)。马氏链的任何两个状态是连通的含义是指存在一个*n*, 使得矩阵*Pn* 中的任何一个元素的数值都大于零。

由马氏链收敛的定理, 概率分布*πi*(*x*)将收敛到平稳分布 *π*(*x*)。假设到第*n*步的时候马氏链收敛。如果我们从一个具体的初始状态 *x*0 开始,沿着马氏链按照概率转移矩阵做跳转，那么我们得到一个转移序列 *x*0,*x*1,*x*2,⋯*xn*,*xn*+1⋯, 由于马氏链的收敛行为， *xn*,*xn*+1,⋯ 都将是平稳分布 *π*(*x*) 的样本。

1.3 **Markov Chain Monte Carlo**

如果我们能构造一个转移矩阵为*P*的马氏链，使得该马氏链的平稳分布恰好是*p*(*x*), 那么我们从任何一个初始状态*x*0出发沿着马氏链转移, 得到一个转移序列 *x*0,*x*1,*x*2,⋯*xn*,*xn*+1⋯,， 如果马氏链在第*n*步已经收敛了，于是我们就得到了 *π*(*x*) 的样本*xn*,*xn*+1⋯。

**定理：[细致平稳条件]**如果非周期马氏链的转移矩阵*P*和分布*π*(*x*) 满足

*π*(*i*)*Pij*=*π*(*j*)*Pji*for all*i*,*j*(1)

则 *π*(*x*) 是马氏链的平稳分布，上式被称为细致平稳条件(detailed balance condition)。

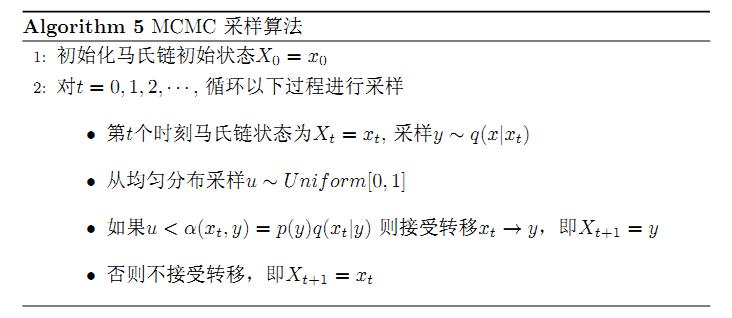
譬如，我们引入一个 *α*(*i*,*j*), 我们希望

*p*(*i*)*q*(*i*,*j*)*α*(*i*,*j*)=*p*(*j*)*q*(*j*,*i*)*α*(*j*,*i*) (∗)(2)

取什么样的 *α*(*i*,*j*) 以上等式能成立呢？最简单的，按照对称性，我们可以取

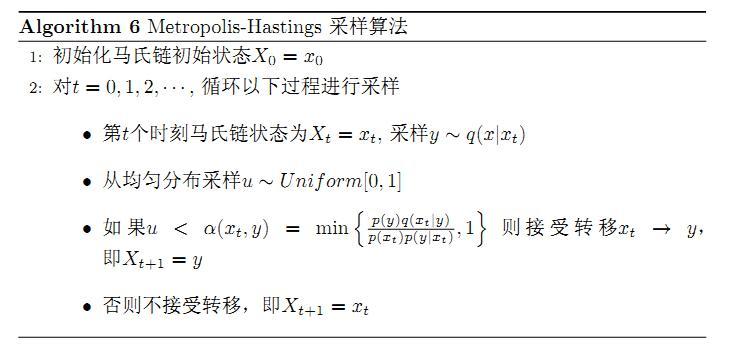
*α*(*i*,*j*)=*p*(*j*)*q*(*j*,*i*)，*α*(*j*,*i*)=*p*(*i*)*q*(*i*,*j*)

在改造 *Q* 的过程中引入的 *α*(*i*,*j*)称为接受率，物理意义可以理解为在原来的马氏链上，从状态 *i* 以*q*(*i*,*j*) 的概率转跳转到状态*j* 的时候，我们以*α*(*i*,*j*)的概率接受这个转移，于是得到新的马氏链*Q*′的转移概率为*q*(*i*,*j*)*α*(*i*,*j*)。



以上的 MCMC 采样算法已经能很漂亮的工作了，不过它有一个小的问题：马氏链*Q*在转移的过程中的接受率 *α*(*i*,*j*) 可能偏小，这样采样过程中马氏链容易原地踏步，拒绝大量的跳转，这使得马氏链遍历所有的状态空间要花费太长的时间，收敛到平稳分布*p*(*x*)的速度太慢。

看，我们提高了接受率，而细致平稳条件并没有打破！这启发我们可以把细致平稳条件(\*\*) 式中的*α*(*i*,*j*),*α*(*j*,*i*) 同比例放大，使得两数中最大的一个放大到1，这样我们就提高了采样中的跳转接受率。



1.4 Gibbs Sampling

对于高维的情形，由于接受率 *α*的存在(通常 *α*<1), 以上 Metropolis-Hastings 算法的效率不够高。

我们先看看二维的情形，假设有一个概率分布 *p*(*x*,*y*), 考察*x*坐标相同的两个点*A*(*x*1,*y*1),*B*(*x*1,*y*2)，我们发现

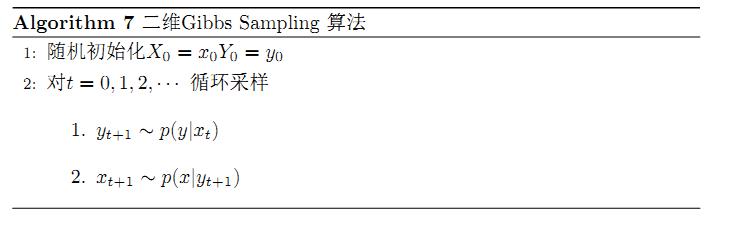
*p*(*x*1,*y*1)*p*(*y*2|*x*1)=*p*(*x*1)*p*(*y*1|*x*1)*p*(*y*2|*x*1)

*p*(*x*1,*y*2)*p*(*y*1|*x*1)=*p*(*x*1)*p*(*y*2|*x*1)*p*(*y*1|*x*1)

所以得到

*p*(*x*1,*y*1)*p*(*y*2|*x*1)=*p*(*x*1,*y*2)*p*(*y*1|*x*1) (∗∗∗)(4)

基于以上等式，我们发现，在 *x*=*x*1 这条平行于 *y*轴的直线上，如果使用条件分布 *p*(*y*|*x*1)做为任何两个点之间的转移概率，那么任何两个点之间的转移满足细致平稳条件。



所以*n*维空间中对于概率分布 *p*(*x*1,*x*2,⋯,*xn*) 可以如下定义转移矩阵

1. 如果当前状态为(*x*1,*x*2,⋯,*xn*)，马氏链转移的过程中，只能沿着坐标轴做转移。沿着 *xi* 这根坐标轴做转移的时候，转移概率由条件概率 *p*(*xi*|*x*1,⋯,*xi*−1,*xi*+1,⋯,*xn*) 定义；
2. 其它无法沿着单根坐标轴进行的跳转，转移概率都设置为 0。

