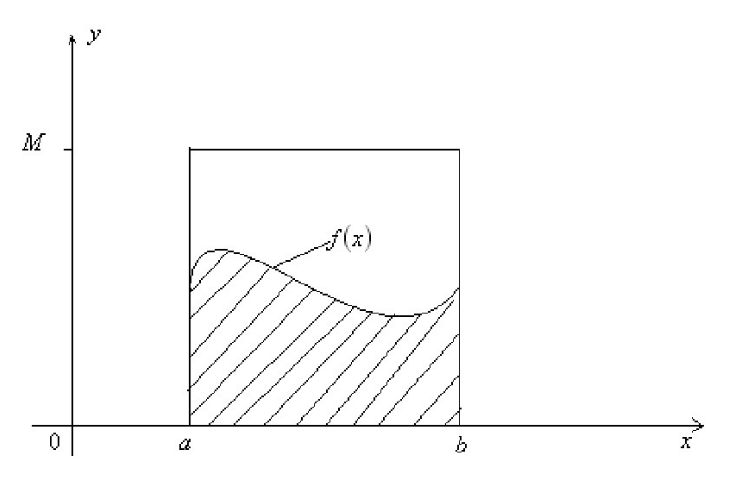
**MCMC（Markov Chain Monte Carlo）**

回顾：

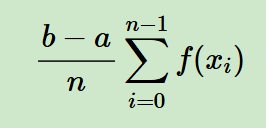
1、Monte Carlo



（1）概率分布采样

如果我们很难求解出f(x)的原函数，那么这个积分比较难求解。

我们可以采样[a,b]区间的n个值：x0,x1,...xn−1,用它们的均值来代表[a,b]区间上所有的f(x)的值。这样我们上面的定积分的近似求解为:



（分布均匀）

（2）分布不均问题

绝大部分情况，x在[a,b]之间不是均匀分布的

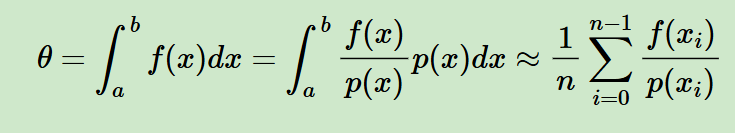
Example：







我们通过x在[a,b]的概率分布函数p(x)，可以求出Θ：



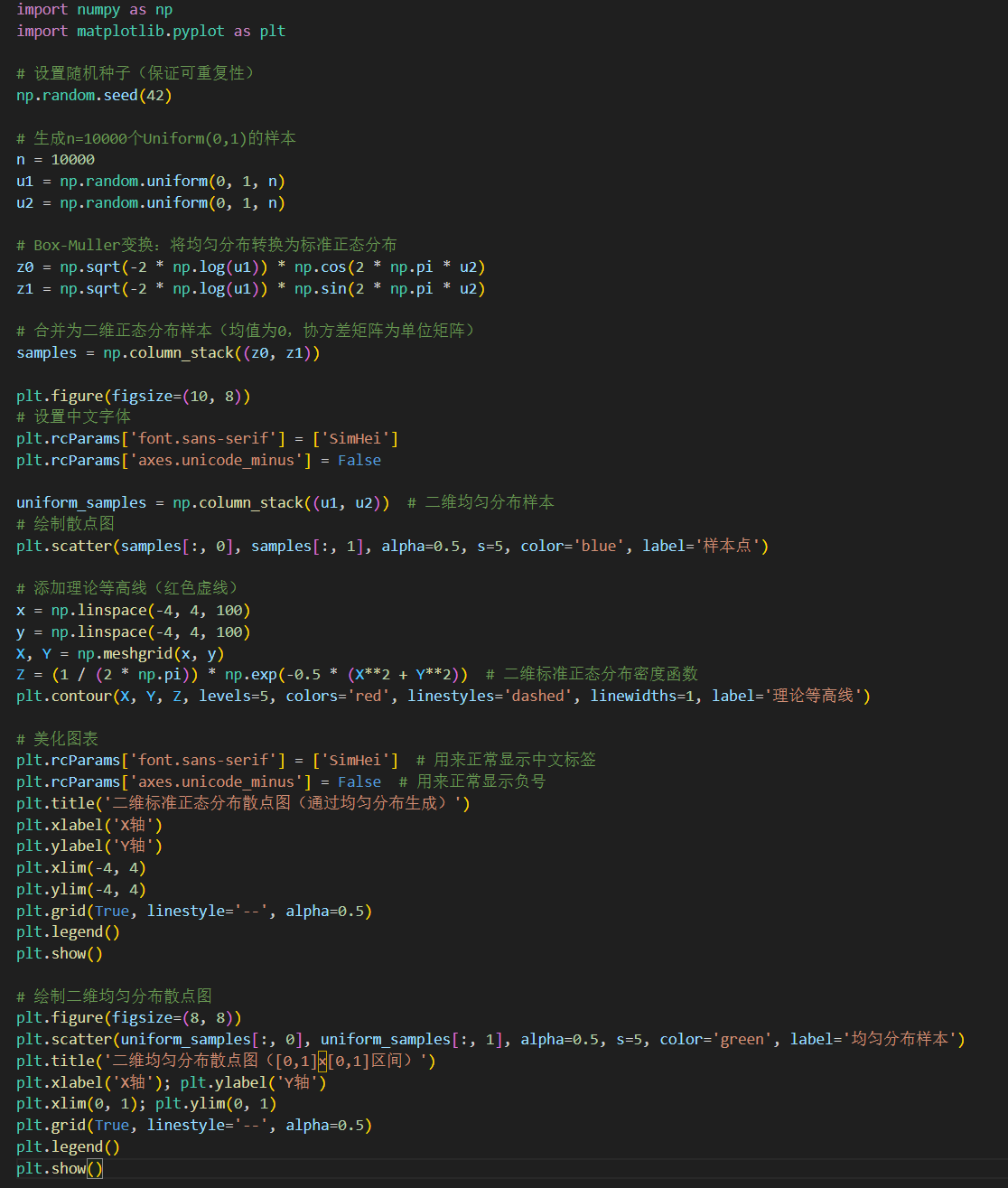
上式最右边的这个形式就是蒙特卡罗方法的一般形式。

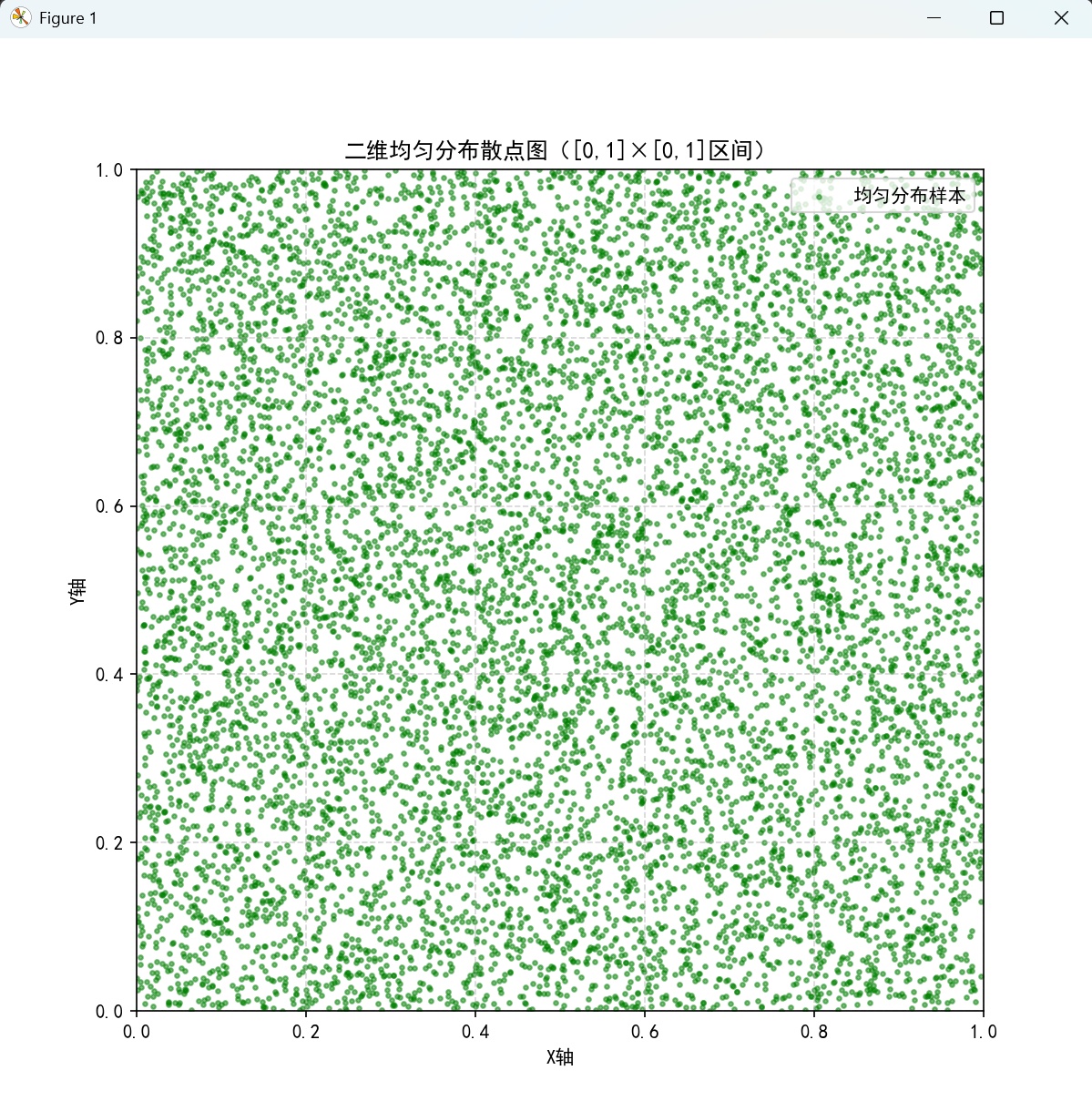
（3）概率分布采样及接受-拒绝采样

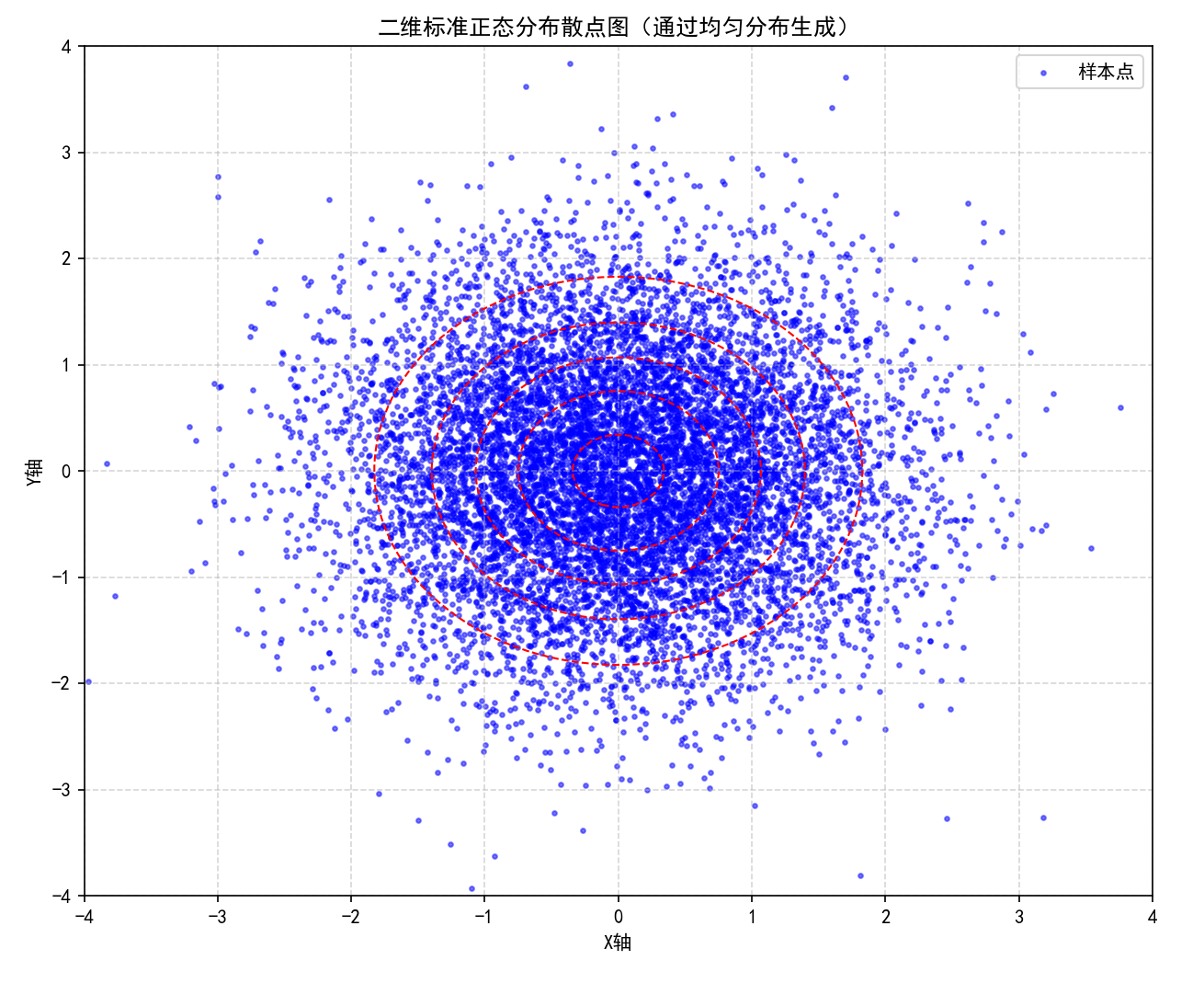
我们知道，如果求出了x的概率分布，我们可以基于概率分布去采样n个x的样本集。那么如何基于概率分布去采样这n个x的样本集？

常见的概率分布，比如t分布，F分布，Beta分布，Gamma分布等，无论是离散的分布还是连续的分布，它们的样本都可以通过uniform(0,1)先生成均匀随机样本，再通过对应公式转换而得。

Example：基于正态分布采样10000个x的样本集

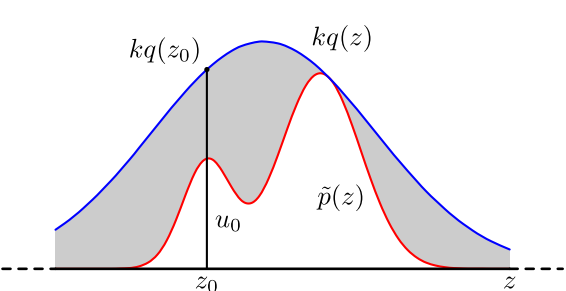






不过很多时候，我们的x的概率分布不是常见的分布，这意味着我们没法方便的得到这些非常见的概率分布的样本集。这时我们引入接受-拒绝采样：

设定一个方便采样的常用概率分布函数 q(x)，以及一个常量 k，使得 p(x)总在 kq(x)的下方。



整个过程中，我们通过一系列的接受拒绝决策来达到用q(x)模拟p(x)概率分布的目的。

但是此方法在高维当中难以实现，所以学要借助马尔科夫链来解决。

2、Markov Chain

（1）定义：假设某一时刻状态转移的概率只依赖于它的前一个状态。

Example：

共有三种状态：A, B和C。每一个状态都以一定的概率转化到下一个状态。



这样我们得到了马尔科夫链模型的状态转移矩阵为：



通过矩阵相乘的方法，可以得到某初始状态运动n次后状态的概率

（2）**状态转移矩阵的收敛性**

假设A,B,C的概率分布为[0.3,0.4,0.3],初始状态为t0，将其带入这个状态转移矩阵计算t1,t2,t3...的状态：

import numpy as np

matrix = np.matrix([[0.9,0.075,0.025],[0.15,0.8,0.05],[0.25,0.25,0.5]], dtype=float)

vector1 = np.matrix([[0.3,0.4,0.3]], dtype=float)

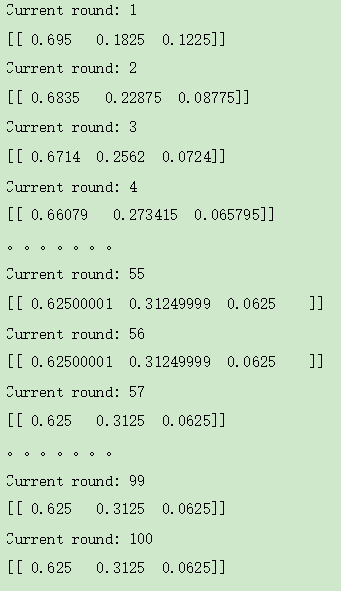
for i in range(100):

vector1 = vector1\*matrix

print "Current round:" , i+1

print vector1

部分输出如下：



后面经过测试发现，尽管采用了不同初始概率分布，最终状态的概率分布趋于同一个稳定的概率分布[0.625   0.3125  0.0625]，也就是说马尔科夫链模型的状态转移矩阵收敛到的稳定概率分布与我们的初始状态概率分布无关。

（3）周期和非周期马尔科夫链

周期马尔科夫链：

如果存在一个正整数 d>1，使得从任意状态 i 出发，只有经过 d 的倍数步才能返回到状态 i，那么称状态 i 是周期为 d 的。

* **特点**：
  + 周期性意味着返回到某个状态的步数总是某个固定数的倍数。
  + 周期马尔科夫链可能不会收敛到一个唯一的平稳分布，或者可能根本不收敛。

非周期马尔科夫链：

如果马尔科夫链的所有状态都不是周期的，即从任意状态出发，可以在任意步数后返回到该状态，那么称这个马尔科夫链是非周期的。

* **特点**：
  + 非周期性意味着返回到某个状态的步数没有固定的周期限制。
  + 非周期马尔科夫链更容易分析，因为它们通常具有更好的收敛性质。

（4）马尔科夫链的细致平稳条件

非周期马尔科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：



对于任意两个状态 i 和 j，从状态 i 转移到状态 j 的概率流必须等于从状态 j 转移到状态 i 的概率流。

如果假定我们可以得到我们需要采样样本的平稳分布所对应的马尔科夫链状态转移矩阵，那么我们就可以用马尔科夫链采样得到我们需要的样本集，进而进行蒙特卡罗模拟。

**MCMC及M-H采样**

在实际问题中，我们很难直接找到对应的马尔科夫链状态转移矩阵P，不是所有的马尔科夫链都有平稳分布，只有满足细致平稳条件才行。我们希望从目标分布 π(x) 中采样，但直接从这个分布中采样可能很困难。因此，我们使用一个易于采样的提议分布 Q 来辅助采样（类似蒙特卡罗）。然而，直接使用 Q 可能不满足细致平衡条件，即：



**MCMC初步：**

我们可以对上式做一个改造，使细致平稳条件成立。方法是引入一个α(i,j),使上式可以取等号：



α(i,j) 满足下两式：



通过引入 α(i,j)，我们可以构造一个新的状态转移矩阵 P，它满足细致平衡条件：



α(i,j)一般称之为接受率，取值在[0,1]之间。

由于α(xt,x∗)可能非常的小，比如0.1，导致我们大部分的采样值都被拒绝转移，采样效率很低，计算量大，因此引入M-H采样。

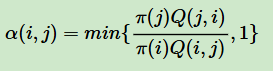
M-H采样：

这个算法首先由Metropolis提出，被Hastings改进，因此被称之为Metropolis-Hastings采样或M-H采样。

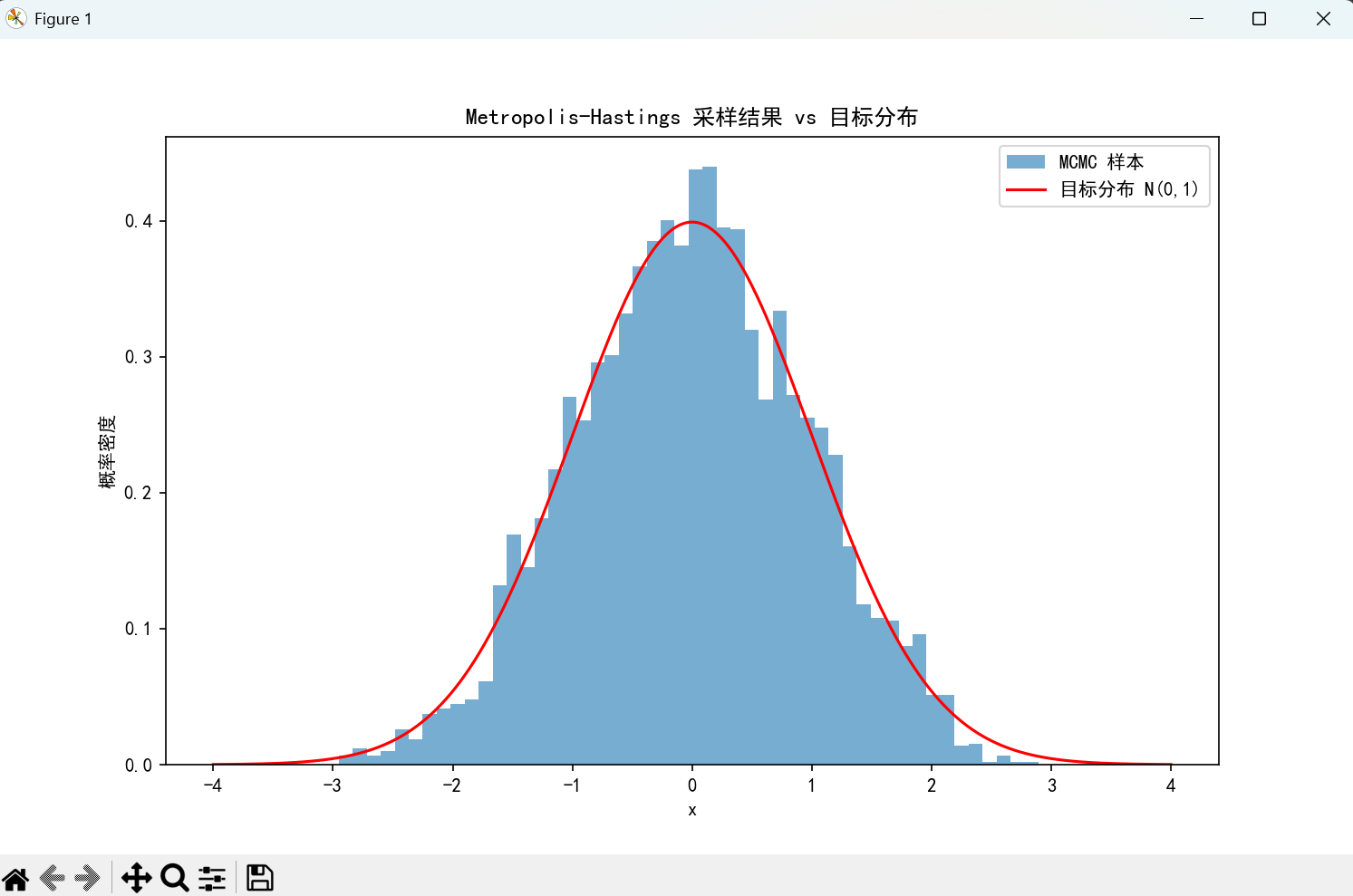
由于α(i,j)太小了，比如为0.1，而α(j,i)为0.2。即：

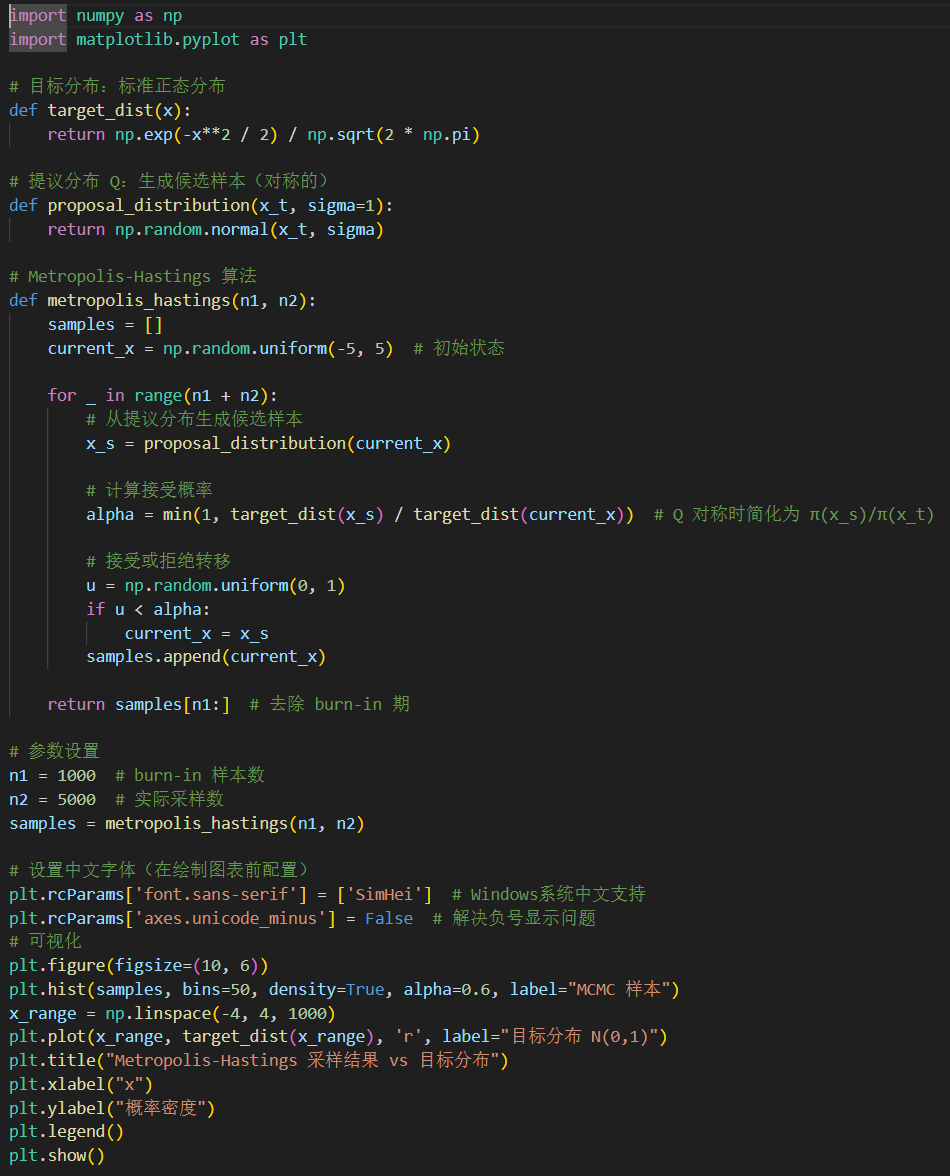


如果两边同时扩大5倍，细致平稳条件仍然满足，这样我们的接受率可以做如下改进，即：



（使用 min 函数取这个比率和 1 中的较小值。这是因为接受概率 α(i,j) 不能大于 1，否则会有超过 100% 的概率接受候选样本）

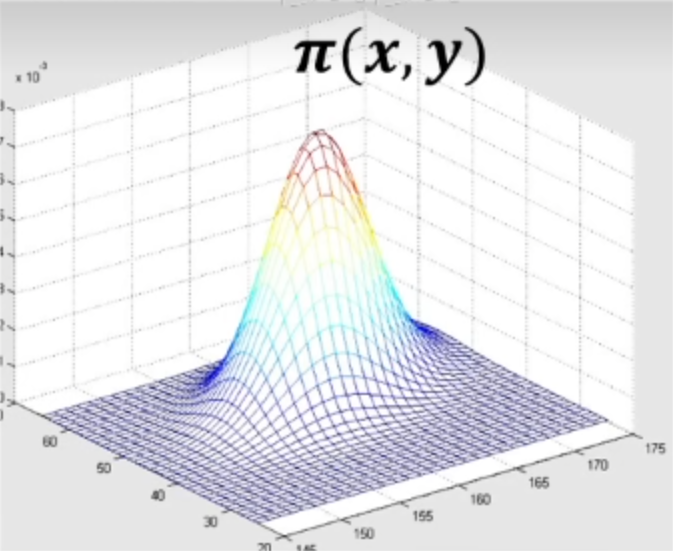
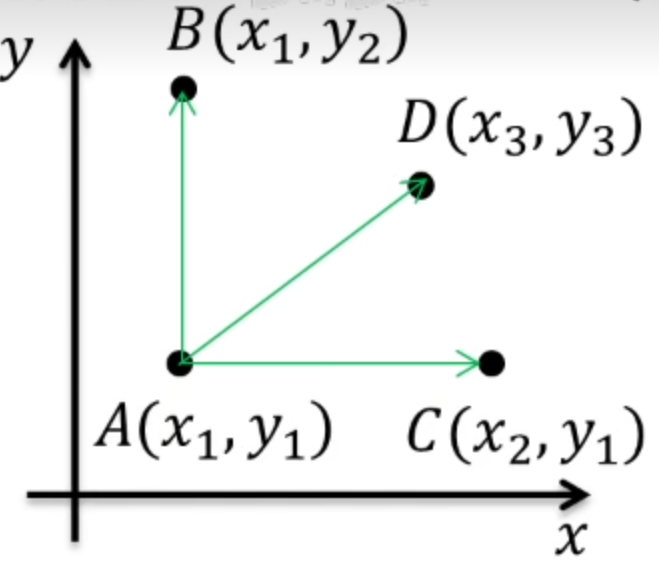




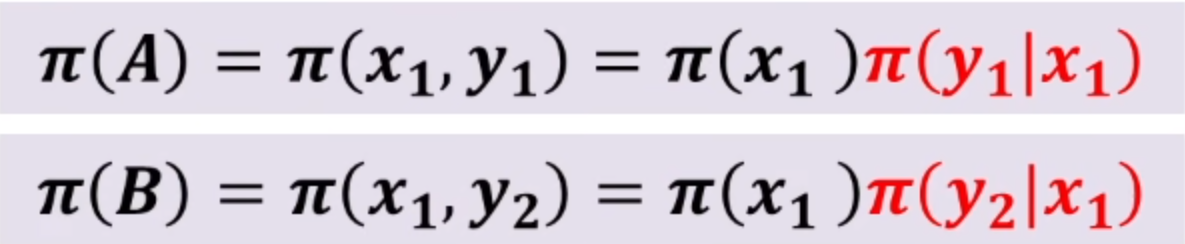
但是M-H采样有两个缺点：一是需要计算接受率，在高维时计算量大。因此引入Gibbs采样。

**Gibbs采样**

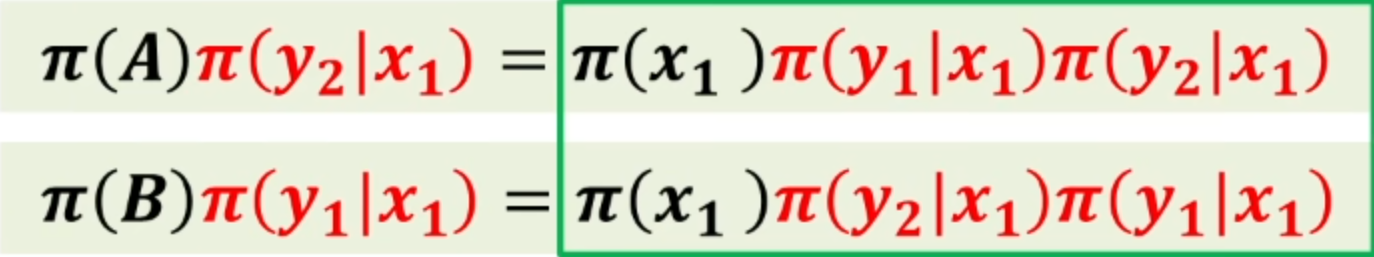
在M-H采样中我们通过引入接受率使细致平稳条件满足。现在我们换一个思路。

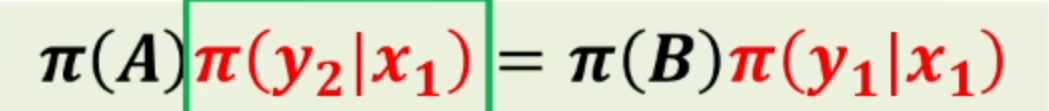
从二维的数据分布开始，假设π(x,y)是一个二维联合数据分布，观察第一个特征维度相同的两个点A,B，容易发现下面两式成立：



再做个变换：



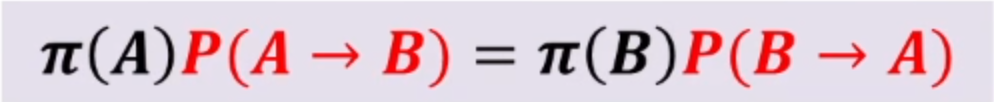
我们发现：



类似于细致平稳条件：

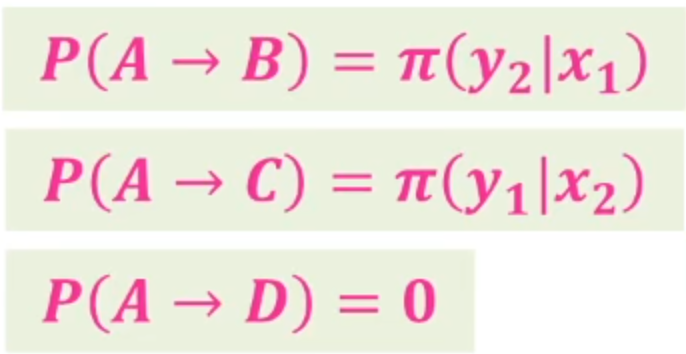


这样我们就得到状态转移概率：



如果用条件概率分布作为马尔科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。

基于上面的发现，我们可以这样构造分布π(x,y)的马尔可夫链对应的状态转移矩阵P：



（只允许在平行坐标轴方向上采样）