**受限Boltzmann机**

### 马尔科夫链

1. 随机过程

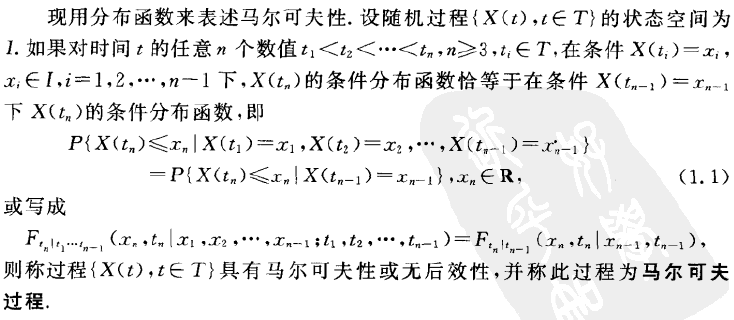
设T是一无限实数集，我们把依赖于参数的一族（无限多个）随机变量称为**随机过程**，记为，这里对每一个 ，X(t)是一随机变量，T叫做**参数集**。我们常把t看做时间，称X(t)为时刻t时过程的状态，而X(t1) = x（实数）说成是t = t1时过程处于状态x。对于一切，X(t)所有可能取的一切值的全体称为随机过程的**状态空间**。

随机过程还可依时间（参数）是连续或者离散进行分类。当时间集T是有限或无限区间时，称为**连续参数随机过程**（以下如无特别说明，“随机过程”总是指连续参数而言的）。如果T是离散集合，例如 ,则称为**离散参数随机过程**或**随机序列**，此时常记为

1. 马尔科夫性

系统将来所处的状态只跟现在所处的状态有关系，跟现在之前的状态无关（将来只依赖于现在，而不依赖于过去）。我们称具有这种性质的随机过程叫做具有**马尔科夫性**。

1. 马尔科夫过程



1. 马尔科夫链

时间和状态都是离散的马尔科夫过程称为**马尔科夫链**，简称马氏链。

1. 转移概率

由转移概率组成的矩阵P(m,m+n) = (Pij(m,m+n))称为马氏链的**转移概率矩阵**。

注：P(m,m+n)便是从第m时刻到第m+n时刻的转移概率。

当转移概率Pij(m,m+n)与m和m+n的时刻无关，只与i，j及时间间距n有关时，将其记做Pij(n),即 ,并称此转移概率具有**平稳性**，同时也称此链是**齐次的**或**时齐的**。

1. 多步转移概率矩阵
2. C-K方程（求多步转移概率矩阵）：P(n) = Pn(1)
3. 一般，设齐次马氏链的状态空间为I，若对于所有ai,aj∈I，转移概率Pij(n)存在极限

或

则称此链具有**遍历性**。又若 ，则同时称为链的**极限分布**。

1. 设齐次马氏链的状态空间为,P是它的一步转移概率矩阵，如果存在正整数m，使对任意的ai,aj∈I，都有

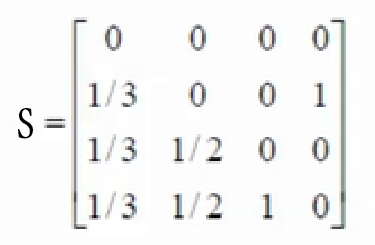
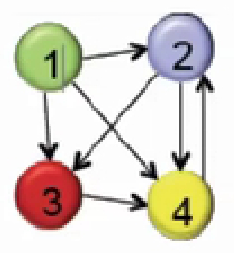
则此链具有遍历性，且有极限分布 **,** 它是方程组

的满足条件

的唯一解。

1. 应用（Page Rank）

* 这是Google最核心的算法，用于给每个网页价值评分。
* 图中四个节点代表四个网页，每两个节点间的连接代表网页间的互相引用。列出一步概率转移矩阵S。
* 为使S具有遍历性，对原有的S进行修正，引入阻尼因子α（一般为代表一个概率，是一个小于1的数，表示的是用户留在当前页面的倾向程度），n为节点总数，U为全1的n\*n矩阵。



### Gibbs抽样（吉布斯抽样）

1. Box-Muller变换（用均匀随机数产生服从标准正态分布的随机数）

如果随机变量U1，U2独立且,则

独立且服从标准正态分布。

1. Markov Chain Monte Carlo（马尔科夫链蒙特卡罗方法）
2. 细致平稳条件

如果非周期马氏链的转移矩阵P和分布满足：

则是马氏链的平稳分布，上式被称为**细致平稳条件**。

1. MCMC采样算法
2. 初始化马氏链初始状态
3. 对,循环以下过程进行采样

* 第t个时刻马氏链状态为,采样【q为随机的转移矩阵】

注：表示从状态的转移概率转移矩阵。

* 从均匀分布采样u ~ Uniform[0,1]
* 如果 ,则接受转移 ,即

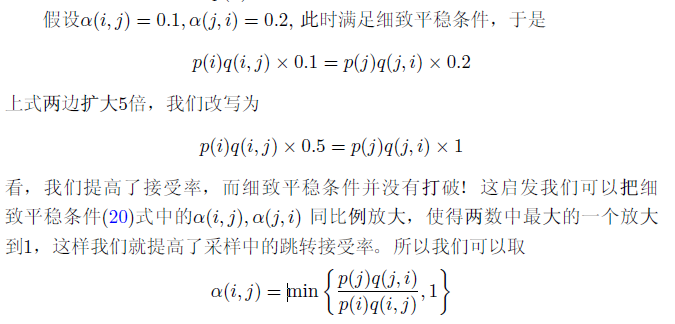
[证明及说明见链接文件](接受率证明.pdf)

* 否则不接受转移，即

1. 一般，在某一步后马氏链便开始收敛。则在这一步之后的抽样就相当于是p的一个抽样。
2. Metropolis Hastings采样算法

为访问MCMC采样算法中过小，取

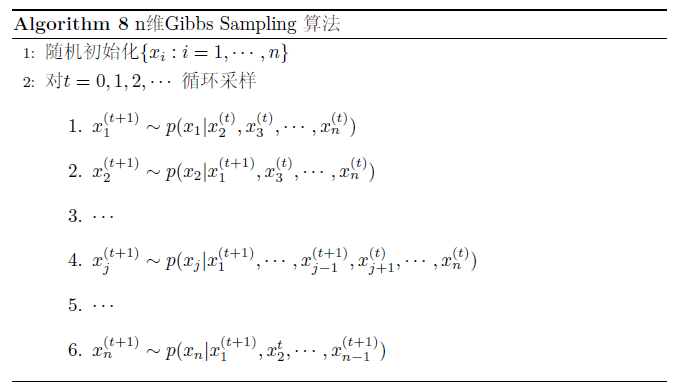
证：



1. Gibbs抽样（高位连续抽样降维为一维抽样）
2. 二维[[证明]](Gibbs抽样证明.pdf)
3. 随机初始化
4. 对,循环以下过程进行采样

* 【此时已变为一维，可通过条件概率公式及边缘概率密度函数计算】

1. 一般，在某一步后马氏链便开始收敛。则在这一步之后的抽样就相当于是p的一个抽样。
2. n维



1. 为了得到近似独立的采样，也可以在采样阶段设置每隔L次迭代采样一样。

### 三、受限Boltzmann机

1. 拓扑结构

* 权值对称
* 同层节点间没有连接
* 节点取值可以不只是0和1，可以是softmax单元、高斯单元、泊松单元等等
* vi代表可见层中不同的神经元的取值；hj代表隐藏层中不同的神经元的取值，ai为可见层的偏置值，bj为隐藏层的偏置值。



1. 算法：

能量函数：

上式中， 是RBM的参数。

在热平衡状态之下，v和h的概率会满足如下玻尔兹曼分布：

其中， 为归一化因子。

1. 计算输入的某个样本v的概率（相当于已知联合概率求边缘概率）

上式无法计算，而由于同一层间节点独立，故可拆解为如下诸多项的积

其中， 为sigmoid激活函数。

由于RBM的结构是堆成的，当给定隐单元的状态时，各可见单元的激活状态之间也是相互独立的，即第i个可见单元的激活概率为

**训练目标：**使似然函数取最大值（即在参数确定的情况下，重现学习样本的概率），因为乘积较难求，故取对数如下：

对比散度算法（CD算法）推导见引文[2](P4,5,6,7)

引文[3]得出Gibbs抽样时，只要迭代一步便可以得出结论。

1. 运行
2. 从可见层输入原样本，通过CD算法计算出权值和偏置值。
3. 将待测试样本从可见层输入，计算如上公式，后抽签判断，若小于该P便激活为1，否则为0；确定隐藏层后反向计算，计算输入层数据，如此往复迭代。
4. 最终达到一种热平衡状态，对比此时的可见层数据和预先输入的样本。

## 参考文献

[1] Rickjin . [*LDA数学八卦*](LDA数学八卦.pdf) . version 1.0 .

[2] 张春霞, 姬楠楠，王冠伟. [*受限波尔兹曼机简介*](受限玻尔兹曼机简介.pdf)

[3] Hinton. [*Training Products of Experts by Minimizing Contrastive Divergence*](Training%20Products%20of%20Experts.pdf)