统计学习读书报告

09017244 郑健雄

1. 自己提出的问题的理解：
2. 使用蒙特卡罗积分法计算积分是否有误差估计方法，也就是说给定一个误差限，能否大体确定随机取样次数达到多少后就能保证计算结果的积分误差符合要求。

讨论后的理解：

处理这个问题，最直观的方法就是多次抽样，然后记录结果的变化，如果在某个数量级达到稳定，即可以认为其符合要求。误差与随机次数之间的关系是负相关的。

除此之外，可以使用递归分层抽样和重要性抽样等方法加速这种估计。因为这种估计是统计学意义上的，所以要考虑的其实除了精确性还有结果正确的可能性，所以只能通过设定误差，计算结果符合的概率(比如99.9%)来尽可能逼近。

1. P362的转移核概念中，x是否也可以换成一个状态空间的子集，来表示从一个状态空间到某个状态或者某个状态空间的概率分布？

讨论后的理解：

应该是可以的，如果是范围上的转移核其实可以看作是无穷个状态的转移概率的积分，所以无论是源头还是目标设置为状态空间都是可以的，得出来的是一个总和。

1. 马尔科夫链在时刻t的状态分布由t-1时刻的状态分布和转移概率分布决定，那么(19.14)如何理解？其是否与状态只受前一个状态影响有矛盾？

讨论后的理解：

如果单纯从概率上看的话，任何状态的概率分布是由初始概率和转移状态决定的，但是这只是从概率上看，但是实际决定t步的不是概率本身，而是t-1步的状态，一旦前一步的状态确定了的话，后面的内容实际上只会受前一部分的制约。

1. 别人提出的问题的理解
2. 不可约的意思是不是经历一定的时间后状态就不会再发生改变？因为严格按照书本的定义例19.5中应该是不可约的，如P（x1=1|x0=2）=0.5>0，就不符合定义了。

自己的理解:

这个问题我认为是因为原始定义中规定是任意一个状态出发都可以到达任意状态，也就是说有一个状态不符合都是可约的，P（x1=1|x0=2）=0.5>0只是一个情况，并不是说其他状态也满足这个定义，如果考虑状态3的话就是一个反例，也就是可约的

1. 如何理解概率密度函数复杂，无法直接抽样？

自己的理解：

直接抽样法有一个应用的前提，就是说目标分布存在反函数并且这种反函数比较好表示出来或者说是进行离散抽样。直接抽样要做的事情就是求出反函数并且进行0-1分布进行抽样，然后去找随机数对应的原函数值，作为抽样值。而复杂的情况下，不一定存在反函数并且不一定可以很好的表示，所以很难进行直接抽样。

1. 如果马尔可夫链可逆，是否可以理解为平稳分布在初值不为平稳分布时不能达到只能逼近？

自己的理解：

马尔可夫链可逆其实说的事情是稳定状态本身具有特殊的性质，而不是说无法达到这个稳定状态。马尔可夫链可逆代表的是稳定状态可以满足平衡方程，是一个新的性质。但是可逆马尔可夫链一定有一个稳定状态，所以其首先满足了稳定状态的求解条件，也就是说可以通过极限逼近的方式得到需要的稳定状态。

1. 马尔可夫链的非周期性如何判断？只有列出步数列然后人工判别的方法吗？

自己的理解：

判断一个节点是不是非周期的一个条件是判断该节点是不是存在自环，若有自环，必然是非周期的。第二个充分条件是与非周期节点互通的节点是非周期的。可以通过这些条件来判断周期性。

1. 读书计划

1、本周完成的内容章节：复习19章，阅读完Speech and Language Processing第五章上半部分。

2、下周计划：复习20章，阅读完第五章内容。

四、读书摘要及理解

1. 蒙特卡罗法：有两类重要的随机抽样算法，拉斯维加斯算法和蒙特卡罗算法。蒙特卡罗算法的核心是随机抽样，用样本估计总体。一般常用的蒙特卡罗法有直接抽样，接受-拒绝抽样法，重要性抽样法。直接抽样法适合于可以求解反函数的简单概率分布，对于较为复杂的抽样，可以用接受-拒绝法，但是该方法可能会因为实际分布占比过低而导致抽样效率过低。

2. 应用：该算法最直接的应用是统计学中的期望估计，通过多次抽样取平均值的方式，可以近似估计总体的期望，类似的思路可以用于估计方差等其他参数。另一个应用是用算法估计函数的积分，其思路是用期望与积分的关系来表示积分，然后通过近似估计期望的方法来估计积分。

3. 马尔可夫链：马尔可夫链可以理解为一个随机过程，该过程有多个状态，按时间先后分布，而每个状态仅仅取决于其前一个状态以及转移概率分布，与其他状态无关。离散马尔可夫链的转移概率可以用转移概率矩阵表示，同时设定初始状态分布之后就可以知道接下来的发展趋势。连续状态马尔可夫链则是从离散状态上得出的，因为状态连续，所以很难定义概率转移矩阵，所以其定义了转移核来表征转移概率。但是其思路和离散情况一致。

4. 马尔可夫链性质：马尔可夫链的性质有不可约，非周期以及正常返。这三个性质可以用来判别马尔可夫链。同时可以用遍历定理求解平稳分布。此外，有一些特殊的马尔可夫链的平稳分布满足细致平衡方程，从而可以向过去与未来发展，属于可逆马尔可夫链，是马尔可夫链的特例。

思考：

马尔可夫链是一种重要的模型，其有着广泛的用途，在NLP领域最著名的模型就是n-gram模型，也就是说某个位置上的词语仅仅取决于其前面n个单词，此外该模型还可以进行很多预测性的工作。将马尔可夫链与蒙特卡罗法结合起来，从而进行一些复杂的抽样工作，这是传统方法不能完成的。