

吸收态马尔可夫链及其在高校学生 学业管理模型中的应用

张梅荣

(北京印刷学院 基础部, 北京 102600)

摘要: 马尔可夫链是一类重要且得到广泛应用的随机过程,对马尔可夫链的基础知识进行了介绍,着重研究吸收态马尔可夫链的转移概率矩阵的构成和基本矩阵的含义,结合高校学生学业管理实例,利用吸收态马尔可夫链和 Excel 中的 MINVERSE 函数和 MMULT 函数给出预测结果,为高校学生管理提供一种新的研究视角。

关键词: 吸收态马尔可夫链; 转移概率矩阵; 过渡态; 吸收态; 基本矩阵

中图分类号: O211.6

文献标识码: A

文章编号: 1004-8626(2011)04-0076-03

如教育、市场营销、医疗服务、金融和工业生产等,其重要特征为具备“无后效性”或“无记忆性”,即系统将来的状态只与它现在时刻的状态有关,而与以往状态无关。在数学上,用如下的定义描述“无后效性”^[1]:

定义 1 如果对于 $t=0,1,2,\dots$ 和所有的状态都有

$$P(X_{t+1} = i_{t+1} | X_t = i_t, X_{t-1} = i_{t-1}, \dots, X_1 = i_1, X_0 = i_0) = P(X_{t+1} = i_{t+1} | X_t = i_t) \quad (1)$$

1.2 转移概率和转移概率矩阵

我们假设对于所有的状态 i 和 j , 以及所有的 $t, P(X_{t+1}=j|X_t=i)$ 只与状态 i 和 j 有关, 而与 t 无关, 此时称马尔可夫链为齐次马尔可夫链, 简称马尔可夫链, 并记

$$P(X_{t+1} = j | X_t = i) = p_{ij} \quad (2)$$

其中, p_{ij} 为假设系统在时间 t 的状态为 i , 那么它在时间 $t+1$ 时状态将为 j 的概率。 p_{ij} 的值通常被称为马尔可夫链的一步转移概率, 简称转移概率。

在应用中, 我们通常将转移概率 p_{ij} 排成一个 $s \times s$ 的矩阵, 令

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & \vdots & \vdots \\ p_{s1} & p_{s2} & \cdots & p_{ss} \end{pmatrix},$$

称 P 为转移概率矩阵, 简称转移矩阵。

除了一步转移概率, 通常还需要研究以下问题, 如果马尔可夫链在时间 t 的状态为 i , 那么 k 个周期之后马尔可夫链的状态为 j 的概率是多少? 引入如下定义:

定义 2 (k 步转移概率) 称条件概率

$$p_{ij}^{(k)} = P(X_{t+k} = j | X_t = i), \quad i, j = 1, 2, \dots, s, t \geq 0, k \geq 1 \quad (3)$$

Markov Chain with Absorbing State and Application of Learning Management Model in College Students

ZHANG Mei-rong

(Beijing Institute of Graphic Communication, Beijing 102600, China)

Abstract: Markov chain is a special kind of stochastic process which is widely used. Firstly, the paper introduces some important concepts of Markov chain. Secondly, it primarily discusses structure of transition probability matrix and meaning of fundamental matrix. Thirdly, a learning management model in college students is given to illustrate the forecasting result which is based on Markov chain with absorbing state and some functions of Excel, it provides an new perspective in management to college students.

Key words: Markov chain with absorbing state; transition probability matrix; transition state; absorbing state; fundamental matrix

1 吸收态马尔可夫链基础知识

1.1 马尔可夫链

马尔可夫链是一类比较常用的随机变量和时间都是离散的随机过程, 它广泛应用于许多领域,

为马尔可夫链的 k 步转移概率,相应地称 $P^{(k)} = (p_{ij}^{(k)})$ 为 k 步转移矩阵。

显然, k 步转移概率 $p_{ij}^{(k)}$ 指的就是系统从状态 i 经过 k 步后转移到 j 的概率。可以证明, p_{ij} 和 $p_{ij}^{(k)}$ 存在如下关系^[2]:

$p_{ij}^{(k)}$ = 转移概率矩阵 P 的 k 次幂 P^k 中的第 i 行第 j 列元素。

1.3 吸收态

如果马尔可夫链上的两状态可以相互转移,则称两状态是连通的。如果状态空间中的任意两状态都是连通的,则称此状态空间是连通状态空间。根据连通的概念,马尔可夫链的状态空间可以分为过渡类和封闭类两类,每个状态唯一地属于其中一类。若连通状态空间的任意状态可以到达连通空间以外的状态,而外面的状态不可转入其内,此状态空间称为过渡类,属于过渡类的状态称为过渡态。若连通状态空间内的任意状态都不可能到达状态空间外的任一状态,而外面的状态可以转入其内,此状态空间称为封闭类。所谓“吸收状态”就是状态独自形成一个闭集,一个状态是吸收态的充要条件是使其一步“返回”概率为 1。含有吸收态的马尔可夫链称为吸收态马尔可夫链。

2 吸收态马尔可夫链的分析模型

对于吸收态的马尔可夫链,我们通常关心以下问题:① 如果链以给定的过渡态开始,那么在到达吸收态之前,我们将进入各种状态的预期次数是多少?在吸收发生之前,我们应该在给定过渡态中度过多少个周期?② 如果链以给定过渡态开始,那么我们在各种吸收状态中结束的概率是多少?

要解决以上问题,我们需要引入几个概念。

2.1 转移概率矩阵的建立

设具有吸收态的马尔可夫链的状态空间有 s 个元素,其中过渡态有 $s-m$ 个,吸收态有 m 个,那么,其转移概率矩阵可以表示为如下形式:

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}$$

其中, Q 是过渡态到过渡态的状态转移矩阵,为 $(s-m) \times (s-m)$ 方阵; R 是过渡态到吸收态的状态转移矩阵,为 $(s-m) \times m$ 矩阵; 0 是包含全部 0 元素的 $m \times (s-m)$ 矩阵,这表明系统不能从吸收态到达过渡态; I 是 $m \times m$ 单位矩阵,它表明吸收态构成的集合是一个闭集。

2.2 基本矩阵

设 $q_{ij}^{(k)}$ 表示在被吸收态吸收之前,从过渡态 i 经过 k 步到达过渡态 j 的概率, $Q(K)$ 表示过渡态间的经过 k 步的转移概率矩阵,则与 1.2 中 k 步转移概率 $p_{ij}^{(k)}$ 的表达式相类似,我们有:

$$Q(K) = (q_{ij}^{(k)}) = Q^k \quad (k = 0, 1, 2, \dots)$$

设 m_{ij} 表示过渡态 i 在被吸收之前转移过渡态 j 的平均转移次数,矩阵 M 表示过渡态间的平均转移次数,则可以证明:

$$M = (m_{ij}) = \left(\sum_{k=0}^{\infty} q_{ij}^{(k)} \right) = \left(\sum_{k=0}^{\infty} Q^k \right) = (I - Q)^{-1}$$

矩阵 M 称为吸收态马尔可夫链的基本矩阵。

设 b_{ij} 表示由过渡态 i 开始被吸收态 j 吸收的概率,矩阵 B 表示由过渡态最终被吸收态吸收的概率矩阵,则可以证明^[3]:

$$B = (b_{ij}) = M \cdot R = (I - Q)^{-1} \cdot R$$

3 高校学生学业管理模型的建立及其求解

某高校的招生办公室将该校学生在大学的求学路径模拟为吸收态马尔可夫链,其具体数据如图 1 所示。

学校在每个秋季学期开始时观察学生的状态。例如,如果一名学生在这个学期开始是 2 年级学生,那么在下个秋季学期开始时,他将成为 3 年级学生的概率为 85%,仍然为 2 年级学生的概率为 10%,退学的概率为 5% (假设学生一旦退学,就不能再入学了)。下面我们讨论以下问题:

1) 如果某学生作为新生进入该高校,那么他

	新生	2 年级学生	3 年级学生	4 年级学生	退学	毕业
新生	0.10	0.80	0	0	0.10	0
2 年级学生	0	0.10	0.85	0	0.05	0
3 年级学生	0	0	0.15	0.80	0.05	0
4 年级学生	0	0	0	0.10	0.05	0.85
退学	0	0	0	0	1	0
毕业	0	0	0	0	0	1

图 1 某高校学生求学路径模拟图

将在大学里度过多少年?

2) 新生毕业的概率是多少?

根据所给的资料,我们可以得出该马尔可夫链的状态空间包含6个状态,分别为新生、2年级学生、3年级学生、4年级学生、退学和毕业,其中新生、2年级学生、3年级学生和4年级学生为过渡态,退学和毕业为吸收态,且:

$$Q = \begin{pmatrix} 0.10 & 0.80 & 0 & 0 \\ 0 & 0.10 & 0.85 & 0 \\ 0 & 0 & 0.15 & 0.80 \\ 0 & 0 & 0 & 0.10 \end{pmatrix}$$

$$R = \begin{pmatrix} 0.10 & 0 \\ 0.05 & 0 \\ 0.05 & 0 \\ 0.05 & 0.85 \end{pmatrix}$$

于是,

$$I - Q = \begin{pmatrix} 0.90 & -0.80 & 0 & 0 \\ 0 & 0.90 & -0.85 & 0 \\ 0 & 0 & 0.85 & -0.80 \\ 0 & 0 & 0 & 0.90 \end{pmatrix}$$

由于 $I-Q$ 方阵比较复杂,且元素中小数居多,我们利用 Excel 中的 MINVERSE 函数来求解 $(I-Q)^{-1}$,结果如下^[4]:

$$M = (I - Q)^{-1} = \begin{pmatrix} 1.111 & 0.988 & 0.988 & 0.878 \\ 0 & 1.111 & 1.111 & 0.988 \\ 0 & 0 & 1.176 & 1.046 \\ 0 & 0 & 0 & 1.111 \end{pmatrix}$$

在 B 矩阵的计算中要用到 $(I-Q)^{-1}$ 与 R 的乘积,同样,我们利用 Excel 中的 MMULT 函数来求解,结果如下:

$$B = (I - Q)^{-1}R = \begin{pmatrix} 0.254 & 0.746 \\ 0.160 & 0.840 \\ 0.111 & 0.889 \\ 0.056 & 0.944 \end{pmatrix}$$

矩阵 M 第一行的4个元素表明:新生作为新生在校的预期学习时间为1.111学年,作为2年级学生和3年级学生在校的预期学习时间均为0.988学年,作为4年级学生在校的预期学习时间为0.878学年,于是问题1)得到解决,即

新生在校的总预期时间为 $1.111 + 0.988 + 0.988 + 0.878 = 3.965$ 学年。

类似地,我们可得出2年级学生、3年级学生和4年级学生在校的预期时间。

由第2部分的分析可知,矩阵 B 第1列的4个元素分别表示新生、2年级学生、3年级学生和4年级学生退学的概率,第2列的4个元素分别表示新生、2年级学生、3年级学生和4年级学生毕业的概率。所以,问题2)的答案为:新生毕业的概率是0.746。

此外,吸收态马尔可夫链也可用于研究银行贷款回收问题、人才留退问题、应收账款问题和部门间费用分配等问题,实际应用价值较高。

参考文献:

- [1] 温斯顿. 运筹学—概率模型应用范例与解法 [M]. 4版. 李乃, 崔群法, 林细财, 等, 译. 北京: 清华大学出版社, 2006.
- [2] 张波, 张景肖. 应用随机过程 [M]. 北京: 清华大学出版社, 2006.
- [3] 倪杰. 吸收态马尔可夫链及其应用 [J]. 商业研究, 2000(7): 75-76.
- [4] 唐冬梅, 郑晗. 应用 Excel 在平差计算中求逆矩阵 [J]. 铁道勘察, 2011, 37(1): 8-9.

(责任编辑: 周宇)

印刷设备及器材5月份数据进口好于出口

据中国印刷及设备器材工业协会统计,2011年5月,我国印刷设备、器材进出口总额为4.6052亿美元,较上月4.3155亿美元增长6.71%;与2010年同期的4.1448亿美元相比增长11.11%。1月~5月,印刷设备、器材累计进出口金额为20.9944亿美元,较2010年同期的18.4252亿美元增长13.94%。统计数据显示,5月印刷设备、器材进口金额为3.0143亿美元,同上月相比增长12.47%,同2010年5月相比增长14.70%。其中,印刷设备进口金额为2.2250亿美元,和2010年同期相比增长21.99%。同上月相比,印刷设备中印前设备、印后设备处于下滑状态,辅机、零件上升幅度为6.81%;但印刷机有较大幅度上升,达28.66%,又因其所占比重较大,故带动整个印刷设备有较大幅度上扬。印刷器材5月进口金额为0.7893亿美元,较上月下降7.96%。与2010年5月相比下滑1.84%。数据显示,5月出口金额为1.5908亿美元,同上月相比下降2.73%,同2010年5月相比上升4.89%。其中,印刷设备出口金额为1.0305亿美元,较上月上升4.58%,和2010年同期相比增长7.74%。和上月相比,仅辅机、零件处下滑态势,其他3项均处上升状态。印刷器材5月出口金额为0.5603亿美元,较上月下滑13.79%,和2010年5月相比仅增长0.01%。

(来源: <http://www.keyin.cn/news/sczc/201107/21-538151.shtml>)