lmpr使用马尔可夫链建模学生留存率与毕业 s

作者：

Mason N. Tedeschi[1], Tiana M. Hose[2], Emily K. Mehlman[3], Scott Franklin[4], Tony E. Wong[2]*

所属机构：

[1] 佛罗里达新学院，美国佛罗里达州萨拉索塔 34243[2] 罗切斯特理工学院数学科学学院，美国纽约州罗切斯特 14623[3] 罗切斯特理工学院理学院，美国纽约州罗切斯特 14623[4] 罗切斯特理工学院物理与天文学学院，美国纽约州罗切斯特 14623 * 通讯作者：aewsma@rit.edu

摘要：

毕业率是衡量学业干预长期成效的关键指标。然而，使用传统方法估算代表性不足的学生的毕业率面临诸多挑战，包括样本量天然较小以及数据需求量大。本文表明，马尔可夫模型能够提高对代表性不足的少数族裔学生和第一代大学生毕业率估算结果的置信度，并减少偏差。我们以学习助理项目为例，展示马尔可夫模型在评估项目成效方面的优势。我们发现，在基础科学课程中配备学习助理与六年毕业率提高 9% 相关。代表性不足的少数族裔学生（21%）和第一代大学生（18%）的增幅更大。结果表明，学习助理能够提高整体毕业率，并改善代表性不足学生毕业率方面的不平等状况。

关键词：不确定性、学习助理、留存率、毕业率、马尔可夫、纵向

1 引言

大学决策者会投资教育项目，以支持不同背景学生的成功。然而，由于项目实施以及项目成果衡量的复杂性，评估项目成效并非易事。管理者通常使用单一的学生成功指标来衡量项目成效，例如课程不及格率（通常指 D、F 成绩或退课；DFW），并据此决定资金支持或项目持续时间。尽管这一策略能够呈现项目成效的一个重要方面，但分析仅限于学生学业历程中的某个时间点，因此可能会忽略教育项目可能影响的其他重要学生成功成果，尤其是更长时间跨度内的留存率和毕业率。因此，需要采用统计方法，将对较长期（即“后续”）学生成功的衡量纳入项目评估指标。

此外，如果教育项目的范围较小（例如，仅在一门课程的单个班级中实施），或者管理者关注该项目对具有边缘化身份的学生（例如理科专业女性或代表性不足的少数族裔学生）的影响，那么小样本量会给项目评估带来挑战。样本量越小，对不及格、退课或撤课（DFW）或留校率等学生成功指标的置信度就越低。因此，有必要采用相应方法，在样本量较小时，对项目效能指标作出更准确的估计。

当项目管理者和管理者将毕业率作为衡量项目成功与否的指标时，当前参与所研究特定项目的学生，与其数据被纳入毕业率计算的学生之间存在固有差异。具体而言，目前就读于某个学术项目的学生中，只有极少数

会被纳入该项目毕业率的“当前”估计值。这些学生中的大多数已经离开该项目，无论是否获得学位（或其他代表成功完成项目的标志）。因此，制定毕业率的“在线”估计值是有益的——也就是说，制定一种可定期更新、[1] 并纳入所有学生队列数据（包括仍在项目中就读的学生）的毕业率估计值。从项目评估的角度来看，这种毕业率在线估计值可以说更为合适，因为它利用了有关学生留存和持续就读的最新信息。另一方面，传统的六年毕业率（SYGR）估计完全依赖至少六年前的数据。如果五年前，所研究学术项目的一年级留存率有了显著提升，那么次年估算出的六年毕业率也会因此提高（其他条件不变）。这凸显了在线估计的优势：它能够纳入所有可用的逐年转移信息。

由于马尔可夫链用于模拟系统在不同状态之间的转移，因此它为表示学生在不同学业状态（例如大学一年级、大学二年级）之间逐年转移提供了自然的数学结构。（一阶）马尔可夫模型指的是：在给定系统当前状态的情况下，系统从一个状态转移到另一个状态的概率有条件地独立于系统此前的所有状态。先前的研究利用马尔可夫链估算大学生的留存率和持续就读情况 [1–3]。Bou mi and Vela [1] 还允许状态自反转移，即转移到其自身；也就是说，学生可以在多个时间点保持在同一学业阶段。这种“多层吸收马尔可夫链”允许学生按顺序多次处于某一状态（例如，学生在大二状态停留超过一个学年），也允许学生非连续地多次处于某一状态（例如，学生退学多个非连续学期后重新注册）。这里，“吸收”一词指的是，一旦学生进入毕业状态或辍学状态，就假定他们会

保持在该状态。因此，按照马尔可夫链的术语，这些系统状态称为吸收态。相比之下，每个学年都是暂态，因为学生会经过这些状态，最终到达一个永久结果（即两个吸收态之一）。然而，Boumi and Vela [1] 引入自反转移后，会导致基于马尔可夫方法的 SYGR 估计产生低估偏差。作者提出了一种校正方法来处理这一偏差；但为避免这个问题，我们将研究范围限定为学生入学后六个日历年内的毕业情况。

在此，我们通过将马尔可夫链模型的表现与传统毕业率计算进行比较，评估该模型能否缩小 SYGR 估计值的不确定范围。我们使用来自美国纽约州罗切斯特市罗切斯特理工学院（RIT）这所大型私立大学的八年数据计算 SYGR 估计值，并通过构建自助法置信区间来刻画不确定性。具体而言，我们假设，在马尔可夫估计中使用更多数据，将比传统计算得出更可靠的 SYGR 估计值。利用更多数据的能力，使在线毕业率估计适用于学生人数较少的群体。我们通过将这项分析应用于一个本地化学术项目——学习助理（LA）项目，以及该项目内更小的人口统计学亚组，展示了这种分析的作用。不过，我们强调，在线毕业率估计相较于传统计算并无缺点。在线估计会利用与传统估计相同的所有信息，还会纳入尚未离校学生的逐年转移信息。因此，在线估计也适用于其他情境。我们利用 RIT 的真实数据，评估使用马尔可夫模型估计学生人数较少的边缘化群体（代表性不足的少数族裔和第一代大学生）的 SYGR 所带来的益处。最后，我们以 RIT 的 LA 项目为案例，评估学生接受学习助理支持对高退课率科学课程学生 SYGR 改善情况的影响。

具体而言，我们考察曾在一门高退课率科学课程中获得 LA 支持的学生，其 SYGR 是否有所提高。

2 方法

2.1 数据与背景

我们分析了 RIT 学生的学业进展。RIT 是美国所有私立大学中科学、技术、工程和数学专业毕业生的第二大培养机构。RIT 是一所以本科教育为主的院校，约有 15,000 名本科生和 3,000 名研究生。学生群体丰富多元，其中包括 2,500 多名国际学生，以及 2,900 名非裔美国人、拉丁裔美国人和美洲原住民（AALANA）学生。RIT 设有国家聋人技术学院，超过 1,500 名聋人或听障学生与听力正常的同学一起上课。本研究的数据涵盖 RIT 的全部学生，来源于罗切斯特理工学院机构数据办公室。数据包括过往和后续的学业记录、每门课程开设时学生所修的项目（专业），以及学生毕业时的毕业年份和所获学位。

我们通过运用马尔可夫分析评估一个本地化项目——RIT 的学习助理（LA）项目，展示了马尔可夫分析的效力。在该项目中，RIT 聘用本科生学习助理（LAs），协助教师在课程中采用循证的、以学生为中心的学习实践。学习助理通过与教师共同开展备课会以及参加教学法课程，接受教学知识和教学内容知识方面的培训[4,5]，从而能够在课堂上成为有效的近龄同伴教育者。科学、技术、工程和数学领域的教师运用学习助理实施循证教学策略，这些策略已被证明能够提升学习成效[6,7] 并降低 DFW比率[7–10]。一项[11] 涵盖39项同行评审研究的范围综述发现，LA支持课程与高阶认知技能的提升[12] 以及改善

概念学习[6], 相关，而这一结果并不取决于某一套特定的课程材料（例如，[13–15]）。LA支持课程的 DFW比率低于未获支持的课程[7,9,10], ，且对代表性不足的人口群体中的学生影响更大[8]。重要的是，以往研究尚未深入考察 LA支持与毕业率、留校率等长期指标之间的关系，也无法直接比较同期获得支持和未获得支持的课程。在 RIT，LA项目仅在特定课程的部分教学班中实施，而马尔可夫分析使研究者能够比较学习助理支持的学生与未获支持教学班中“对照组”学生的表现。由此，RIT 的 LA项目自然形成了这样一种情境：针对小数据情境改进项目效能指标，有助于决策者更有把握地估计学生成功情况。

2.2 六年毕业率的传统估算

传统的毕业率估算首先需要确定感兴趣的学生队列。假设 N_{start} 是该队列中最初入学的学生人数。在这些学生中，令 N_{deg} 表示六年内获得学位的学生人数。那么，$GR_{tr} = N_{deg}/N_{start}$ 就是对这一目标队列的 SYGR 的估算。采用这种传统方法的一个主要局限是：当目标队列由代表性不足的学生群体构成，或是接受学业干预的学生中相对较小的子群体时，N_{start} 往往会非常少。

2.3 基于马尔可夫链的六年毕业率估算

2.3.1 模型

与传统的 SYGR 计算不同，我们采用马尔可夫模型，通过利用学生在给定初始年级状态（例如，大一学生或大二学生）的情况下逐年继续就读的概率来估算 SYGR。

定义 S = {1, 2, 3, 4, 5, 6, D, G} 为马尔可夫链模型的状态空间，其中整数 1、2、……、6 表示学生在 RIT 注册就读的年数（例如，处于状态 2 的学生通常是大学二年级学生）；D 表示已辍学且不再攻读学位的学生，G 表示已获得学位的学生。状态 D 和 G 均为吸收态，其余状态均为暂态。

假设 N_i 表示一年开始时处于状态 i 的学生人数，N_j 表示这些学生中到该年年底处于状态 j 的人数。那么，从状态 i 转移到状态 j 的转移概率为

$$P_{ij} = P(X_{t+1} = j \mid X_t = i) = N_j/N_i. \qquad \text{(Eq. 1)}$$

根据定义，如果 j 不属于 {i+1, D, G}，则该概率为 0。此外，根据定义，$P(X_t+1 = j \mid X_t= D) = 0$ 如果 j 不等于 D，同理，$P(X_{t+1} = j \mid X_t =G) = 0$ 如果 j 不等于 G。因此，转移概率矩阵的整体结构是稀疏的（表 1）：

表 ***1.*** 状态空间 *{1, 2, 3, 4, 5, 6*（大学就读年级）*, D*（辍学）*, G*（毕业生）*}* 的转移概率矩阵的一般结构，用于表示系统从某个特定初始状态（行）出发后转移到最终状态（列）的条件概率。

	1	2	3	4	5	6	D	G
1	0	P_{12}	0	0	0	0	P_{1D}	P_{1G}
2	0	0	P_{23}	0	0	0	P_{2D}	P_{2G}
3	0	0	0	P_{34}	0	0	P_{3D}	P_{3G}
4	0	0	0	0	P_{45}	0	P_{4D}	P_{4G}
5	0	0	0	0	0	P_{56}	P_{5D}	P_{5G}
6	0	0	0	0	0	0	P_{6D}	P_{6G}
D	0	0	0	0	0	0	1	0
G	0	0	0	0	0	0	0	1

学生在入学后六年内到达毕业状态（G）的六步转移概率（从状态1开始）为 $P^6_{1G} = P(X_{t+6} = G \mid X_t = 1)$。这些概率可用于估算 SYGR：$GR_{ma} = P^6_{1G}$，即转移概率矩阵 P 的六次幂中第1行、第G列（第八列）的值。

马尔可夫建模方法的一项关键优势是能够合并来自多个学生队列的信息。例如，设 A 和 B 表示两个感兴趣的队列。设 $N_{A,\,i}$ 和 $N_{B,\,i}$ 分别表示在一学年开始时处于状态 i 的队列 A 和 B 中的学生人数。同样，设 $N_{A,\,j}$ 和 $N_{B,\,j}$ 分别表示到该学年结束时处于状态 j 的学生人数。那么，从状态 i 转移到状态 j 的转移概率的合并估计为

$$P_{AB,\,ij} = (N_{A,\,j} + N_{B,\,j})/(N_{A,\,i} + N_{B,\,i})。 \quad \text{(Eq. 2)}$$

使用公式（2）可使马尔可夫模型纳入最新数据。例如，截至本文撰写时，2021年秋季入学的学生在大学只度过了完整的一年。采用传统方法，至少还要五年才能得到包含这些学生的 SYGR 估计值。不过，这些学生的数据可以通过影响 P_{12}, P_{1D} 和 P_{1G} 的估计值，纳入用于计算 SYGR 的马尔可夫模型。对于这种“部分数据”队列，这些学生不会影响后续任何转移概率估计值。

2.3.2 验证

我们首先进行一项“阳性对照”实验，以证明简约形式的马尔可夫模型能够得出与传统 SYGR 计算相符的结果。对于分别在 2013年秋季、2014年秋季和2015年秋季入学的各队列，我们仅使用相应队列的数据构建简约形式的马尔可夫模型，以构建每个马尔可夫转移概率矩阵（公式 1）。这样，这些模型采用的信息与传统方法针对同一队列估算 SYGR 时所用的信息相同。我们将这些简约形式马尔可夫模型的毕业率估计值与传统毕业率估计值进行比较，以此验证马尔可夫模型能够忠实地复现传统基线。

与这个阳性对照相比，马尔可夫模型的完整形式涉及一个完整数据队列和五个部分数据队列。例如，2021年秋季的项目经理可以使用2015年秋季队列的完整六年数据，按照传统方法估算 SYGR。也可以使用2016、2017、2018、2019 和 2020 年秋季队列的信息，采用马尔可夫链方法估算 SYGR。此外，马尔可夫链模型提供了一种自然结构（公式 2），可将所有可用信息合并为一个 SYGR 估算值，同时纳入2013和2014年秋季队列的数据。这构成了 SYGR 的“完整”马尔可夫模型。

2.4 自助法不确定性表征

为了检验我们的假设——马尔可夫模型比传统模型更能得出置信度更高的毕业率估计值——我们通过自助重采样构建置信区间，以量化 SYGR 估计值的不确定性。我们分别对传统模型和马尔可夫链模型进行此操作。从原始的整体学生数据集中有放回地重采样行，抽取的行数与原始数据集的行数相同。对于每个

重采样数据集，我们使用传统方法（或马尔可夫方法）计算 SYGR。我们使用由 1,000 个重采样数据集组成的集合，并将计算所得 SYGR 集合中 2.5% 至 97.5% 的范围估计为 95% 置信区间。采用不同规模重采样集合的实验表明，该集合规模足够大，因此我们的结果不会受到这一选择的影响。

2.5 学生子集

我们通过考察本科生学习助教支持的课程与没有此类 LA 支持的课程中，学生的 SYGR 有何不同，来展示马尔可夫模型在研究规模相对较小的学生队列时的优势。此外，我们还考察 LA 支持对代表性不足的少数族裔学生和第一代大学生可能产生的不同影响。这些群体（获得 LA 支持的学生、代表性不足的少数族裔学生和第一代大学生）本身的规模都相对较小，因此它们的交集为验证马尔可夫模型在计算 SYGR 方面的优势提供了充足的试验场（例如，获得 LA 支持的第一代大学生）。对于这些针对较小学生子群体的实验，我们不展示使用传统方法计算 SYGR 的结果，因为基准比较实验（第 2.3.2 节）将证明马尔可夫模型估算 SYGR 的准确性。

为控制学生专业领域的多样性，以及纳入 LA 支持的课程种类，我们在 LA 支持组中仅纳入那些在 RIT理学院高流失率课程中接受 LA 支持的学生。这些课程包括普通物理和数学课程，例如微积分预备课程和微积分（见补充材料）。我们注意到，还有其他高流失率课程，但其中纳入学习助理的班级很少或没有。在这些实验中，我们将数据集限定为 2013年秋季至2021年夏季期间至少修读过上述一门课程的学生。

在上述至少一门课程中接受过 LA 支持的学生，从其首次获得 LA 支持的学年起纳入基于马尔可夫的计算。这样，LA 支持学生的马尔可夫转移矩阵中的转移概率应仅代表可能受到 LA 支持这一“处理”影响的转移。相比之下，“对照”组包括所有未在高流失率理科课程中接触过 LA 支持的学生（见上文）。每学期的学生人数和班级数，以及接受 LA 支持与未接受 LA 支持的学生人数细分情况，详见补充材料。

我们注意到，根据 LA状态（有 LA支持或没有）对学生进行区分，并考察各组之间毕业率的差值，可以确立相关或关联关系，但无法证明因果关系。当然，所有此类处理组与对照组比较实验（例如疫苗试验）都具有这一性质。现有大量文献记录了 LA支持对不同院校类型和学科学生的益处——包括在高阶认知技能、概念学习和问题解决技能方面可迁移且持久的提升（见第 1 节）——这为提出如下假设提供了充分理由：学生接受 LA支持应当与其在六年内毕业的可能性存在因果关系。

3 结果

3.1 与传统毕业率计算进行基准比较

仅使用拥有完整六年数据的队列，通过马尔可夫方法计算 SYGR 时，简约形式马尔可夫模型得出的 GR_{ma} 估计值与传统毕业率估计值 GR_{tr} 高度吻合（表 2）。对于2013年秋季，两个中位数以及2.5%至97.5%自助法置信区间的上下限（和区间宽度）均相差不超过一个百分点，且结果一致，

2014年秋季和2015年秋季队列也是如此；将这三个队列的数据合并后（使用公式 2），结果同样如此。这些结果有力地证明，马尔可夫模型是估计 SYGR 的有效方法。在本文余下部分中，马尔可夫模型使用在理论上可于计算传统 SYGR 估计值时获得的全部学生数据。例如，2013年秋季队列的马尔可夫方法 SYGR 将纳入2014年秋季队列五年的数据、2015年秋季队列四年的数据，依此类推。“全部合并”模型纳入数据集中2013年秋季至2021年夏季的所有可用学生数据（2021年夏季是数据集中的最后一个学期）。我们使用“百分点”来表示 SYGR 的范围和绝对差值，以免与 SYGR 的相对变化或相对差异混淆。

表 *2*。使用传统计算和简化形式（仅完整数据队列）的马尔可夫模型估算的六年毕业率（%）。数值四舍五入到最接近的百分点，因此报告的*95%*置信区间宽度可能看起来并不等于第*97.5*百分位数与第*2.5*百分位数之差。

队列	方法	第 2.5 百分位数	中位数	第 97.5 百分位数	95%置信区间宽度
2013年秋季	传统	66	71	75	8
	马尔可夫	66	71	75	9
2014年秋季	传统	66	71	75	9
	马尔可夫	67	71	75	9
2015年秋季	传统	69	73	77	9
	马尔可夫	69	73	77	8
全部合并	传统	69	72	74	5
	马尔可夫	69	72	74	5

3.2 不确定性量化

与传统方法相比，马尔可夫模型对 SYGR 的估计置信度显著更高（图 1）。对于2013—2015年秋季入学队列，采用马尔可夫方法后，95%置信区间宽度比传统计算所得宽度缩小了33%至44%。重要的是，两种方法得出的所有 SYGR 中位数估计值相差均不超过1个百分点：2013年秋季两种方法均为71%；2014年秋季两种方法均为71%；2015年秋季传统方法为73%，马尔可夫模型则为72%（表3）。使用所有可用数据时，95%置信区间宽度的缩减幅度略低（相较于传统计算缩小34%），但仍相当显著。这是因为使用所有可用数据时，部分数据队列（2016年秋季至今）所提供的信息量，相对于完整队列信息（2013—2015年秋季入学队列）较少。

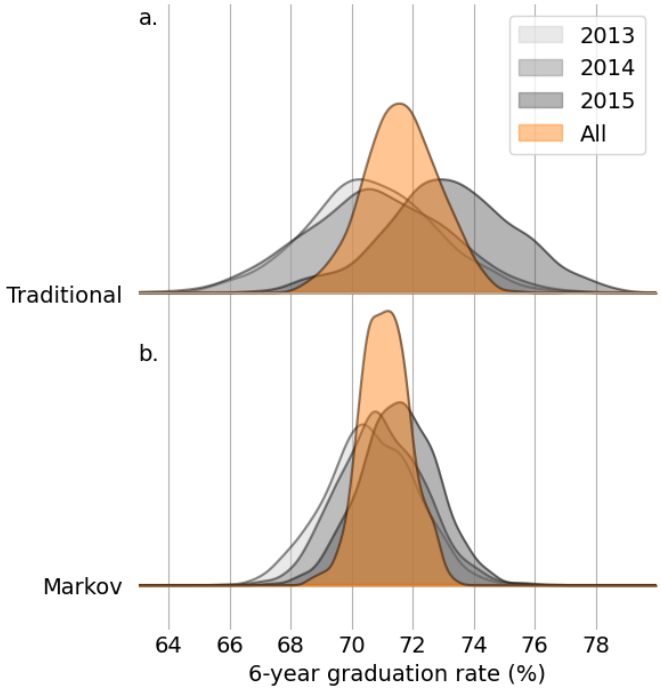

图*1*　　*.* 使用马尔可夫模型估计 *SYGR*（下图），所得置信区间更窄，为 *95%* 置信区间，而使用传统方法时则不然（上图）。图中显示的是核

*2013*年秋季（浅灰色）、*2014*年秋季（灰色）和*2015*年秋季（深灰色）队列，以及这三个队列合并（橙色）的六年毕业率密度估计。

表***3***。传统计算和完整形式马尔可夫模型对六年毕业率（%）的估计。数值已四舍五入至最接近的百分点，因此所报告的*95%*置信区间宽度可能看起来不等于第*97.5*和第*2.5*百分位数之间的差值。

队列	方法	第2.5百分位数	中位数	第97.5百分位数	95%置信区间宽度
2013年秋季	传统	66	71	75	8
	马尔可夫	68	71	73	6
2014年秋季	传统	66	71	75	9
	马尔可夫	68	71	74	6
2015年秋季	传统	69	73	77	9
	马尔可夫	69	72	74	5
全部合并	传统	69	72	74	5
	马尔可夫	70	71	73	3

3.3 学生的小型子集

正如预期，人数较少的学生子群体（AALANA学生和第一代大学生）的置信区间宽度明显大于总体学生群体（表4）。这是因为这些代表性不足群体的样本量本身较小。值得注意的是，在理学院专业的AALANA学生和第一代学生的SYGR基于马尔可夫方法的估计中，也观察到了置信区间宽度缩小的类似优势（表4）。

表*4.* 采用传统计算和全形式马尔可夫模型，对理科专业学生中*AALANA*学生（上组）和第一代学生（下组）的六年毕业率（%）进行估计。数字四舍五入至最接近的百分点，因此报告的*95%*置信区间宽度可能与第*97.5*和第*2.5*百分位数之差不完全相等。

	队列	方法	第2.5 百分位数	中位数	第97.5 百分位数	95%置信区间 宽度
AALANA	2013年秋季	传统	40	55	68	28
		马尔可夫	54	62	71	17
	2014年秋季	传统	48	59	70	22
		马尔可夫	57	64	73	17
	2015年秋季	传统	59	70	83	24
		马尔可夫	62	69	77	15
	All 合并	传统	55	62	69	14
		马尔可夫	61	65	70	9
第一-代	2013年秋季	传统	61	70	81	19
		马尔可夫	64	70	76	11
	2014年秋季	传统	57	67	77	20
		马尔可夫	61	68	76	15
	2015年秋季	传统	63	72	81	17
		马尔可夫	65	70	75	10
	All 合并	传统	64	70	75	11
		马尔可夫	67	70	73	7

与传统模型相比，马尔可夫方法使置信区间宽度缩小了25%至42%（图2a和图2c，以及表4）。值得注意的是，采用传统方法计算 SYGR 时，AALANA 理科专业学生的置信区间宽度始终超过20个百分点，而且中位数估计值存在显著的年际变化（2013年秋季为55%，2014年秋季为59%，2015年秋季为70%）。采用传统方法时出现的这种年际变异令人担忧，因为学生和学术项目管理者可能会依据 SYGR 信息做出决策，而这些信息每年可能有很大差异。例如，对于2013年秋季队列，AALANA 理科专业学生 SYGR 的95%置信区间为40%至68%，甚至未涵盖采用2015年秋季队列时的 SYGR 中位数（70%）。虽然这并不一定表明不同年份之间存在统计学上的显著差异，但这些结果凸显出，对于规模较小的学生子群体，采用传统方法时 SYGR 的最佳估计值会有显著差异。

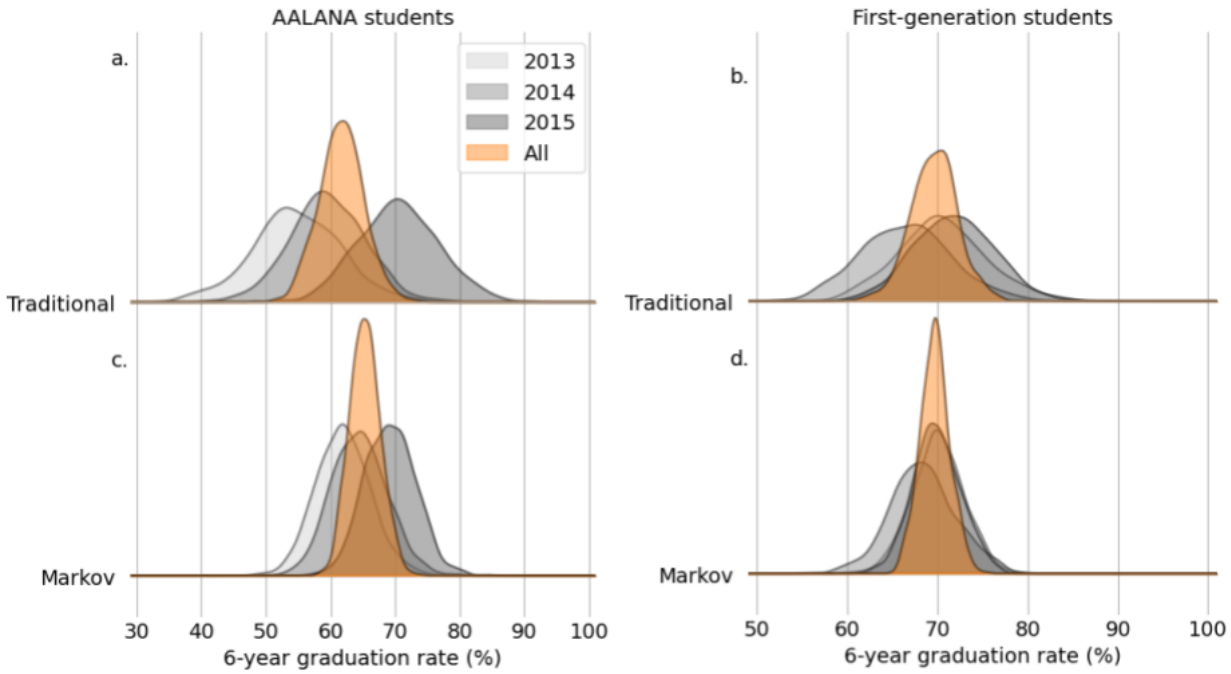

图 ***2.*** 与传统方法（上方图组）相比，使用马尔可夫模型（下方图组）估算 *AALANA*（左侧图组）和第一代理科专业学生（右侧图组）的 *SYGR*，可获得更窄的 *95%* 置信区间，并降低年度间的变异性。图中所示

这是*2013*年秋季（浅灰色）、*2014*年秋季（灰色）和*2015*年秋季（深灰色）队列，以及这三个队列合并后的数据（橙色）的六年毕业率核密度估计。

相比之下，使用马尔可夫模型可以随着新信息的获得将其纳入分析。因此，SYGR的估计值更好地围绕长期中位数分布：AALANA学生在2013年秋季为62%、2014年秋季为64%、2015年秋季为69%，总体为65%。此外，置信区间宽度始终低于20个百分点。将马尔可夫模型应用于第一代大学生理科专业学生这一子集时，也能看到类似的改善：置信区间更窄，估计的SYGR年际变异更低（图2b和d，以及表4）。

3.4 概念验证：评估学习助理支持的影响

现在，我们将重点关注这样一群学生：他们在RIT就读期间的某个阶段，至少修读过一门高流失率理科课程，并且已申报理科专业（见第2.5节）。我们将这群学生分为两组：“LA组”学生所修读的至少一门高流失率理科课程获得了本科生学习助理的支持；“无LA组”学生则修读了未获得学习助理支持的这些高流失率课程班级。本节呈现的所有结果均纳入了2013年秋季至2021年夏季期间所有可用的学生数据（即前文结果部分中的“全部合并”情形），并采用马尔可夫模型。

我们发现，在高流失率理科课程中接受学习助理支持，与未获得学习助理支持、选修这些课程其他班级的学生相比，SYGR提高了9个百分点（图3a）。学习助理支持的学生SYGR中位数为77%（95%置信区间为73–81%）。这一比例显著高于非LA组，其SYGR中位数为68%（95%置信区间为66–70%）。

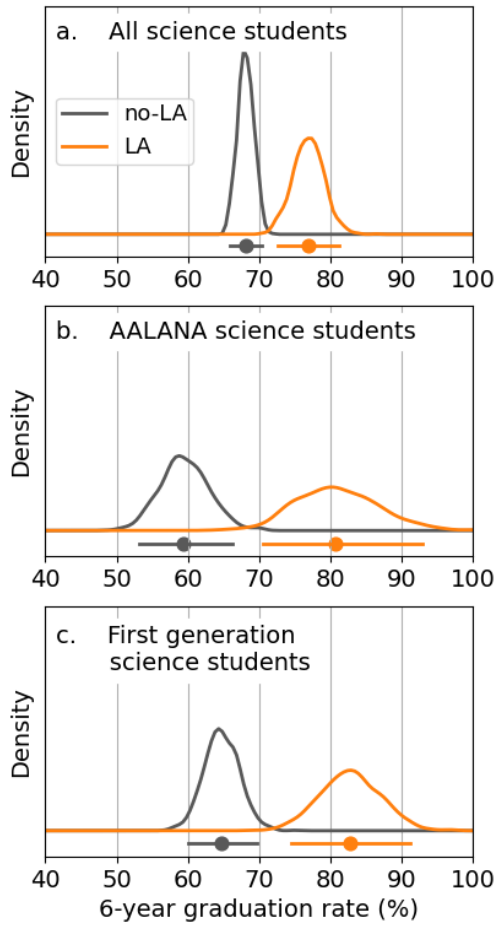

图 ***3.*** 学习助理支持与理科专业学生总体六年毕业率提高 *9* 个百分点相关（*a*）。*AALANA* 理科专业学生的提升幅度更大（*b*；*SYGR* 提高 *21* 个百分点），第一代大学生中的理科专业学生提升幅度也更大（*c*；*SYGR* 提高 *18* 个百分点）。

对于代表性不足群体的学生，这些增益更加显著。对于选修高流失率课程的AALANA理科专业学生，学习助理支持与SYGR提高21%相关。获得学习助理支持的AALANA学生SYGR中位数为81%（71–93%）；相比之下，未获得学习助理支持的AALANA学生SYGR中位数为59%（53–66%）。同样，对于理科专业的第一代大学生，无学习助理支持组的

中位数 SYGR 为 65%（60–70%），而 LA组的中位数 SYGR 为 83%（75–91%），提高了 18 个百分点。

4 讨论与结论

我们介绍了一种利用马尔可夫链估计大学生队列 SYGR 的方法。我们使用 RIT 的真实数据拟合该模型，并证明它能够忠实地复现采用传统方法得出的毕业率估计值。我们发现，与传统方法相比，马尔可夫模型得出的 SYGR 估计值置信区间更窄，年际变异也更低。我们认为，这是因为马尔可夫模型利用了所有可用数据，而不是仅使用至少有六年数据的队列数据。这一点对于学生人数较少的队列尤为重要；我们通过 AALANA学生、第一代大学生，以及接受特定学业干预的学生（在高流失率课程中获得 LA支持）展示了这一点。代表性不足群体的 SYGR 置信区间通常较宽，这是由于样本量较小。即使使用马尔可夫模型并纳入所有可用数据，对于在课程中获得 LA支持的代表性不足群体，SYGR 的置信区间仍可能超过 20 个百分点。在这些情况下，采用传统方法得出的置信区间过宽，失去了实际价值。因此，SYGR 马尔可夫模型在评估教育项目成效方面尤其有价值，特别是对于代表性不足群体的学生。

毕业率常用于评估教育项目的成效，但这些评估通常不会讨论或纳入毕业率中的不确定性。在这项研究中，我们证明，马尔可夫模型可以随着学生数据的不断产生，将全部学生数据纳入分析，从而得出更窄且波动更小的置信区间。这对于学术机构和项目而言十分重要，使其能够尽可能及时地掌握有关项目成效和学生成果的信息

。这项研究为解决一个问题指明了方向：大学一旦公布毕业率信息，这些信息就会立即过时。实际上，参与该毕业率计算的学生几乎都已不在那所大学就读。

通过展示马尔可夫模型在学生小群体中的效用，我们发现，在高流失率基础科学课程中，学习助理支持与学习助理支持的学生 SYGR 提高 9 个百分点相关，相较之下，未接受学习助理支持的学生则没有这一提升。我们强调：i）学习助理组与非学习助理组学生的人口统计特征相似；ii）学生选课时并不知道哪些课程班级会有学习助理支持；iii）除学习助理支持外，课程中有学习助理支持的班级与没有学习助理支持的班级之间没有其他系统性差异（例如，各班使用相同的课程大纲）。我们承认，我们的方法无法直接控制个别教师或其教学选择可能造成的“教师效应”。然而，在本研究考察的高流失率科学导论课程中，与学习助理支持相关的 SYGR 提升结果逐年一致，且适用于所有教师和课程。此外，将学习助理融入课程有助于采用多种循证教学法[6,7], ，例如更多采用主动学习方法。鉴于主动学习已被证明具有益处[16–18], ，我们认为，研究结果源于学习助理支持使教师能够实施更多循证教学法，进而提升学生表现。

与 LA 支持相关的这些提升，在代表性不足群体学生中幅度显著更大。LA 支持与 AALANA 学生 SYGR 提高 21 个百分点相关，也与第一代大学生 SYGR 提高 18 个百分点相关。这些人数较少的学生子群体的 SYGR 估计值具有

较宽的置信区间，因此难以区分本科生学习助教等学业干预措施所产生的影响。不过，我们在此证明，马尔可夫方法能够有效约束人数少至几十人的子群体的 SYGR 估计值。除了 SYGR 的显著提升外，LA 支持还与学生在 RIT 就读期间逐年持续率的提高相关（见补充材料）。我们的分析表明，在高流失率科学导论课程中提供本科生学习助教支持，或许是缩小非代表性不足学生与代表性不足学生之间持续就读和毕业率差距的一条可行途径。

致谢

我们感谢罗切斯特理工学院科学与数学教育研究协作组织的成员与我们进行了富有成果的交流。本研究部分由美国国家科学基金会拨款资助，项目编号：DUE #1757477。本研究及纵向数据集的使用已获罗切斯特理工学院人类受试者研究办公室批准（HSRO #01011221；FWA # 00000731）。本研究附带的代码可从 https://github.com/tonyewong/LA_Markov_analysis 获取。原始数据包含学生隐私信息，不得共享。经罗切斯特理工学院机构审查委员会批准后，可直接向作者获取处理后的汇总数据。

参考文献

1. Boumi S, Vela AE. 使用定期更新的多层吸收马尔可夫链改进毕业率估算。《教育科学》. 2020;10: 377. doi:10.3390/educsci10120377 2. Nicholls MG. 使用吸收马尔可夫链评估博士和硕士学位候选人的进展及其流动的内在特征。《高等教育》. 2007;53: 769–790. doi:10.1007/s10734-005-5275-x 3. Witteveen D, Attewell P. 《大学完成学业之谜》：隐马尔可夫模型

方法 h. Res High Ed 2017年d 58 449 467 oi:10.1007/ 11162 016 9430 s -- -2 4. Etkina E. 教学内容知识与高中物理教师的培养。Phys Rev Spec Top - Phys Educ Res. 2010;6: 020110. doi:10.1103/PhysRevSTPER.6.020110 5. van Driel JH, Verloop N, de Vos W. 科学教师教学内容知识的发展。J Res Sci Teach. 1998;35: 673–695. doi:10.1002/(SICI)1098- 2736(199808)35:6<673::AID-TEA5>3.0.CO;2-J 6. Pollock SJ. 转变教学改革：大型讲授课程中的学习增益、学生态度及多名教师的影响。美国物理学会；2006年。第141–144页。7. Van Dusen B, Nissen J. 学习助理、通过物理入门课程与公平之间的关联：一项定量批判种族理论研究。Phys Rev Phys Educ Res. 2020;16: 010117. doi:10.1103/PhysRevPhysEducRes.16.010117 8. Alzen JL, Langdon L, Otero V. 学习助教模式与物理入门课程中的DFW比率。2017年。第36–39页。9. Alzen JL, Langdon LS, Otero VK. 对学习助教模式与STEM入门课程不及格率之间关系的逻辑回归研究。Int J STEM Educ. 2018;5: 1–12. 10. Close EW, Mailloux-Huberdeau J-M, Close HG, Donnelly D. 本科物理项目中检测改革影响的时间尺度特征。2018年。第88–91页。获取地址：https://www.per-central.org/items/detail.cfm?ID=1457511. Barrasso AP, Spilios KE. 评估学习助理模式影响的文献范围综述。Int J STEM Educ. 2021;8: 12. doi:10.1186/s40594-020-00267-8 12. Sellami N, Shaked S, Laski FA, Eagan KM, Sanders ER. 实施学习助理项目可提升学生在高阶评估中的表现。CBE—Life Sci Educ. 2017年;16: ar62。

13. Goertzen RM、Brewe E、Kramer LH、Wells L、Jones D。迈向变革：通过实施学习助教模式和开放源代码教程实现改革制度化。《物理评论：物理教育研究专题》。2011；7：020105。doi:10.1103/PhysRevSTPER.7.020105

14. Herrera X, Nissen JM, Van Dusen B. 协作学习环境中的学生成果。2018。可访问：https://www.per-central.org/items/detail.cfm?ID=14796

15. Van Dusen B, Nissen J. 大学物理学生学习公平性：一项批判性定量交叉性研究。J Res Sci Teach. 2020;57: 33–57. doi:10.1002/tea.21584

16. Deslauriers L, McCarty LS, Miller K, Callaghan K, Kestin G. 在课堂中积极参与时，衡量实际学习与学习感受。Proc Natl Acad Sci. 2019;116: 19251–19257. doi:10.1073/pnas.1821936116

17. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, 等。主动学习可提升学生在科学、工程和数学领域的表现。Proc Natl Acad Sci. 2014;111: 8410–8415. doi:10.1073/pnas.1319030111

18. Theobald EJ、Hill MJ、Tran E、Agrawal S、Arroyo EN、Behling S 等。主动学习缩小了本科科学、技术、工程和数学领域代表性不足的学生之间的成就差距。《美国国家科学院院刊》。2020；117：6476–6483。doi:10.1073/pnas.1916903117

补充材料

随附于 Tedeschi 等人撰写的《使用马尔可夫链改进学生留校率和毕业率》。

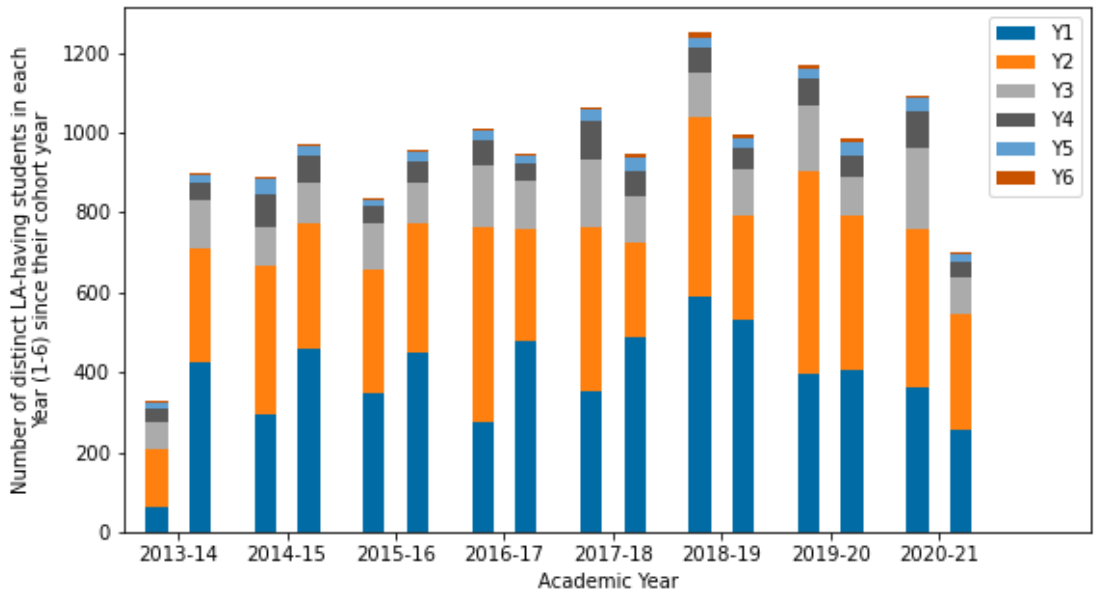

图 ***S1.*** 学习助理每学期持续影响 *800* 多名学生，其中约 *20%* 的学生就读三年级或更高年级。每对堆叠柱形图中，左侧和右侧的柱分别代表该学年的秋季学期和春季学期。

表 ***S1***。 *2013*年秋季至*2021* 年春季期间，理学院每门高不及格、退课或撤课课程中学习助理支持的学生和无 *LA* 支持的学生人数，以及课程班级数量。数据项为学生人数，括号内为课程班级数量。

		Fa 2013	Sp 2014	Fa 2014	Sp 2015	Fa 2015	Sp 2016	Fa 2016	Sp 2017	Fa 2017	Sp 2018	Fa 2018	Sp 2019	Fa 2019	Sp 2020	Fa 2020	Sp 2021	Total
College Physics I (PHYS 111)	**LA**	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	79 (1)	123 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	202 (2)
	no-LA	395 (6)	436 (4)	308 (6)	416 (5)	265 (5)	411 (5)	164 (3)	218 (3)	267 (4)	355 (4)	256 (5)	387 (4)	277 (4)	385 (4)	249 (3)	403 (4)	5192 (69)
College Physics II (PHYS 112)	**LA**	0 (0)	122 (1)	0 (0)	0 (0)	0 (0)	0 (0)	114 (1)	102 (1)	0 (0)	0 (0)	0 (0)	99 (1)	0 (0)	0 (0)	64 (1)	0 (0)	501 (5)
	no-LA	137 (2)	137 (2)	175 (2)	167 (2)	151 (2)	165 (2)	30 (1)	42 (1)	132 (2)	146 (2)	126 (2)	61 (1)	121 (2)	138 (2)	83 (1)	146 (1)	1957 (27)
University Physics I (PHYS 211)	**LA**	84 (2)	166 (4)	42 (1)	42 (1)	42 (1)	123 (3)	123 (3)	207 (5)	210 (5)	210 (5)	149 (4)	123 (3)	166 (4)	83 (2)	129 (4)	144 (3)	2043 (50)
	no-LA	213 (6)	182 (5)	134 (4)	304 (8)	175 (5)	219 (6)	93 (4)	189 (7)	28 (2)	214 (7)	50 (2)	264 (7)	75 (2)	259 (7)	70 (2)	250 (8)	2719 (82)
University Physics IA (PHYS 211A)	**LA**	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	42 (1)	0 (0)	0 (0)	0 (0)	16 (1)	64 (2)	17 (1)	0 (0)	35 (1)	82 (2)	256 (8)
	no-LA	114 (3)	179 (5)	111 (3)	189 (5)	116 (4)	258 (7)	102 (3)	151 (5)	112 (3)	110 (4)	22 (1)	42 (1)	35 (1)	88 (3)	0 (0)	0 (0)	1629 (48)
University Physics II (PHYS 212)	**LA**	65 (2)	81 (2)	82 (2)	121 (3)	69 (2)	167 (4)	90 (2)	80 (2)	25 (1)	76 (2)	115 (3)	129 (4)	81 (2)	190 (5)	90 (3)	83 (2)	1544 (41)
	no-LA	261 (8)	326 (10)	333 (9)	153 (5)	301 (9)	171 (6)	418 (12)	312 (10)	420 (13)	260 (8)	294 (9)	144 (4)	279 (8)	118 (4)	261 (7)	246 (7)	4297 (129)
Precalculus (MATH 111)	**LA**	0 (0)	0 (0)	36 (1)	0 (0)	78 (2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	114 (3)
	no-LA	440 (13)	216 (6)	425 (12)	131 (4)	426 (12)	174 (5)	238 (8)	102 (4)	218 (7)	100 (4)	211 (6)	85 (3)	409 (14)	97 (3)	519 (14)	103 (3)	3894 (118)
Calculus A (MATH 171)	**LA**	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	268 (7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	268 (7)
	no-LA	348 (9)	279 (7)	372 (9)	301 (8)	282 (8)	275 (8)	244 (7)	143 (4)	248 (7)	132 (4)	0 (0)	151 (4)	333 (9)	197 (6)	332 (9)	248 (6)	3885 (105)
Calculus II (MATH 182A)	**LA**	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	29 (1)	0 (0)	0 (0)	29 (1)
	no-LA	0 (0)	165 (5)	0 (0)	202 (6)	35 (1)	369 (11)	70 (2)	269 (8)	66 (2)	314 (8)	77 (2)	303 (9)	77 (2)	228 (6)	103 (3)	289 (7)	2567 (72)
Elements of Multivariable Calculus and Differential Equations (MATH 211)	**LA**	0 (0)	0 (0)	44 (1)	110 (3)	144 (4)	132 (3)	89 (3)	119 (3)	138 (4)	110 (3)	146 (4)	39 (1)	0 (0)	102 (3)	0 (0)	0 (0)	1173 (32)
	no-LA	137 (4)	77 (4)	85 (3)	0 (0)	0 (0)	0 (0)	34 (1)	0 (0)	0 (0)	0 (0)	0 (0)	70 (2)	156 (4)	0 (0)	154 (4)	117 (3)	830 (25)

表 ***S2.*** 将自助法集成的规模增大后，六年毕业率（%）置信区间的估计值与正文中的结果（*1,000* 次重复抽样）相比并不敏感。此处以*2013*年秋季队列的传统计算为例。

集成 size	第2.5 百分位数	中位数	第97.5 百分位数
1000	66.4	70.6	74.8
2000	66.2	70.5	75.0
4000	66.2	70.5	75.0
8000	66.2	70.6	75.2

表 ***S3***。 学习助理（*LA*）支持与理科专业学生逐年持续率（%）的提高相关。持续率根据所有可用数据（*2013*年秋季至*2021*年夏季）使用马尔可夫转移概率进行估算。数值四舍五入至最接近的百分点。

转移	无学习助理	LA	差值（学习助理 - 无学习助理）
Y1→Y2	87	95	+9
Y2→Y3	91	93	+2
Y3→Y4	91	93	+3

表 ***S4.*** 学习助理（*LA*）的支持与 *AALANA* 理科专业学生逐年持续率（%）的提高相关。持续率根据所有可用数据（*2013*年秋季至*2021*年夏季）估算为马尔可夫转移概率。数值四舍五入至最接近的百分点。

转移	无学习助理	LA	差值（学习助理 - 无学习助理）
Y1→Y2	84	92	+9
Y2→Y3	88	92	+4
Y3→Y4	89	94	+5

表 ***S5.*** 学习助理（*LA*）支持与第一代大学生理科专业学生逐年持续率（%）的提高相关。持续率根据所有可用数据（*2013*年秋季 *- 2021*年夏季）估算为马尔可夫转移概率。数值四舍五入至最接近的百分点。

转移	**无学习助理**	**LA**	**差值（学习助理 - 无学习助理）**
Y1→Y2	86	96	+9
Y2→Y3	90	93	+3
Y3→Y4	90	91	+1