Advances in Education 教育进展, 2025, 15(1), 341-347
Published Online January 2025 in Hans. https://www.hanspub.org/journal/ae
https://doi.org/10.12677/ae.2025.151051

Hans汉斯

基于试验设计和大数据子抽样技术的“概率论与数理统计”课程教学探索

李文龙

北京交通大学数学与统计学院，北京

收稿日期：2024年12月7日；录用日期：2025年1月8日；发布日期：2025年1月15日

摘　要

本教学探索聚焦于“概率论与数理统计”课程，深入剖析课程教学现状及痛点问题。将试验设计方法融入教学过程，阐述其与概率论及数理统计知识的紧密联系，涵盖试验设计基本原则、常用方法及其在课程中的应用实例。同时引入大数据子抽样技术，详细说明其原理、方法步骤，并结合课程知识点通过实例展示其作用。经教学实践检验，这些举措有效提升学生学习效果、实践能力与创新思维，为课程教学改革提供有益参考。

关键词

大数据，“概率论与数理统计”，试验设计，教学探索

Exploration on the Teaching of “Probability Theory and Mathematical Statistics” Based on Experiment Design and Big Data Subsampling Technology

Wenlong Li

School of Mathematics and Statistics, Beijing Jiaotong University, Beijing

Received: Dec. 7th, 2024; accepted: Jan. 8th, 2025; published: Jan. 15th, 2025

Abstract

This teaching exploration focuses on the course “Probability Theory and Mathematical Statistics”, and deeply analyzes the current situation and pain points of the course teaching. This paper integrates

文章引用: 李文龙. 基于试验设计和大数据子抽样技术的“概率论与数理统计”课程教学探索[J]. 教育进展, 2025, 15(1): 341-347. DOI: 10.12677/ae.2025.151051

the experimental design method into the teaching process, explains its close connection with probability theory and mathematical statistics, and covers the basic principles of experimental design, common methods, and their application examples in the course. At the same time, the big data subsampling technology is introduced, its principles, methods, and steps are explained in detail, and its role is demonstrated through examples in combination with the course knowledge points. Through the test of teaching practice, these measures can effectively improve students' learning effectiveness, practical ability, and innovative thinking and provide a useful reference for curriculum and teaching reform.

Keywords

Big Data, "Probability Theory and Mathematical Statistics", Experiment Design, Teaching Exploration

1. 引言

"概率论与数理统计"作为高校理工及经管类本科专业的关键基础课程，在当今大数据时代发挥着举足轻重的作用[1]。它不仅是现代科学和技术发展的基石，为众多领域如金融风险管理、医学统计分析、工程和物理学应用、机器学习和人工智能、生态学和环境科学等提供了不可或缺的理论和方法，还能帮助学生提升逻辑思维、科学推理和实证分析能力，为其职业发展和学术研究奠定坚实基础。然而，当前课程教学面临诸多挑战，如理论抽象、教学方法传统、实践环节薄弱以及教学评价单一等，亟待创新教学模式以适应时代需求。

在当今大数据时代，为契合时代需求、增强学生数据分析能力以及助力学生运用概率统计思想方法有效解决各类随机现象的相关问题，概率统计课程的教学改革与创新势在必行。国内外众多研究已从多维度对概率统计课程教学改革予以探讨。在情境化教学领域，王泽龙等[2]从课程属性、所处的智慧教学环境以及教学理念等多个角度剖析了推行情境化教学模式的必然性；在教学模式领域，金今姬和高彦伟[3]深入研究了问题驱动式学习模式于概率统计教学中的应用方式；教学方法方面，Kazak 和 Pratt [4]探讨了概率教学中综合建模方法的关键意义与面临的挑战；肖敏等[5]主张教学应重视知识背景与实际需求，借助信息化教学手段与案例教学法优化教学效果；在具体教学主题层面，肖进胜等[6]针对现代数理统计中的假设检验教学过程展开研讨，强调原假设设计方法，并以实例阐述依据"小概率事件原理"设计假设的方案；课程思政方面，王贶和朱靖红[7]以全概率公式和贝叶斯公式为范例，探索教学中开展课程思政的素材与方式。基于已有研究成果，本文结合试验设计(参考[8]-[10])的特点以及大数据子抽样技术(参考[11] [12])，着重阐述其与概率论及数理统计知识的紧密联系，涵盖试验设计基本原则、常用方法及其在课程中的应用实例。同时引入大数据子抽样技术，详细说明其原理、方法步骤，并结合课程知识点通过实例展示其作用。经教学实践检验，这些举措有效提升学生学习效果、实践能力与创新思维，为课程教学改革提供有益参考。

2. 课程教学现状与痛点分析

2.1. 理论抽象，学生理解困难

课程中的概念和理论较为抽象，如大数定律、中心极限定理等，学生往往难以直观理解其内涵和应

用场景。传统教学模式下，教师主要通过理论推导和公式讲解进行授课，缺乏生动形象的实例辅助，导致学生学习积极性不高，学习效果不佳。例如，在讲解大数定律时，学生难以理解为何随着试验次数的增加，随机事件的频率会逐渐稳定于其概率。

2.2. 教学方法传统，缺乏创新

教学过程中，部分教师仍采用“板书 + 讲解”或“PPT + 讲解”的单一教学方式，教学手段陈旧。这种传统教学方法注重知识的灌输，忽视了学生的主体地位，缺乏与学生的互动交流，难以激发学生的学习兴趣和主动性。例如，在讲解概率分布时，只是简单地罗列各种分布的公式和性质，没有引导学生深入理解其实际意义和应用场景。

2.3. 实践环节薄弱，应用能力不足

课程的实践教学环节相对薄弱，实验课程设置较少且内容简单，与实际应用场景脱节。学生缺乏运用所学知识解决实际问题的机会，导致其在面对实际问题时，无法有效地运用概率论与数理统计方法进行分析和解决，实践应用能力亟待提高。例如，在进行统计推断实验时，学生只是按照给定的步骤进行计算，没有真正理解如何根据实际问题设计实验、收集数据和分析结果。

2.4. 教学评价单一，考核不够全面

课程的教学评价主要依赖于期末考试成绩，平时成绩占比较小且考核方式单一，通常仅包括考勤和作业。这种评价方式过于注重结果，忽视了学生的学习过程和能力提升，不能全面、准确地反映学生的学习效果和综合素质。例如，学生在课堂上的参与度、小组合作能力以及实践项目中的表现等未得到充分体现。

3. 教学方法创新：试验设计的融入

3.1. 试验设计的基本原则

3.1.1. 重复性原则

在相同条件下进行多次独立重复试验，以增加试验结果的可靠性和稳定性。这一原则与概率论中的大数定律密切相关。大数定律表明，随着试验次数的增加，事件发生的频率会趋近于其概率。例如，在抛硬币试验中，重复抛硬币多次，正面朝上的频率会逐渐接近 0.5。通过多次重复试验，学生可以更直观地理解大数定律的内涵，同时也能体会到试验设计中重复性原则的重要性。

3.1.2. 随机性原则

试验中涉及的因素水平组合应随机分配给试验对象，确保每个对象都有同等机会接受各种处理。这有助于消除潜在的系统误差，使试验结果更具客观性。在概率论中，随机变量的随机性是其重要特征之一。例如，在抽样调查中，随机抽取样本可以保证样本的代表性，从而基于样本对总体进行推断。随机分配处理因素水平组合的过程，类似于从总体中随机抽取样本，使每个处理水平都有相同的机会被选中，进而保证试验结果不受人为因素的干扰。

3.1.3. 区组化原则

将试验对象按某些特征或条件划分为不同区组，使区组内的对象具有较高的同质性，区组间则存在较大差异。在分析试验数据时，可将区组效应从总变异中分离出来，提高试验的精度和灵敏度。在方差分析中，区组因素可以作为一个额外的变量来解释部分变异。例如，在比较不同教学方法对学生成绩的影响时，如果考虑到学生的基础水平可能存在差异，可以将学生按照基础水平相近的原则划分为不同区

组，然后在每个区组内实施不同的教学方法。这样可以减少学生基础水平差异对教学方法效果评估的影响，更准确地判断教学方法的优劣。

3.2. 常用的试验设计方法

3.2.1. 完全随机设计

将试验对象完全随机地分配到各个处理组中，每个对象接受何种处理完全随机决定。该设计简单易行，适用于因素水平较少、试验对象同质性较高的情况。从概率论的角度来看，每个对象被分配到各个处理组的概率相等，这符合随机事件的概率定义。例如，在比较不同药物对某种疾病的治疗效果时，假设有三种药物 A、B、C，将患者完全随机地分为三组，分别接受三种药物治疗。在这个过程中，每个患者被分配到任何一组的概率都是 1/3。通过这样的设计，可以减少其他因素对治疗效果评估的干扰，更准确地判断药物的疗效差异。

3.2.2. 随机区组设计

先将试验对象按某些特征或条件划分为若干区组，然后在每个区组内将对象随机分配到各个处理组。这种设计在考虑处理因素的同时，还能控制区组因素对试验结果的影响，提高试验效率。在数理统计中，随机区组设计可以看作是一种特殊的方差分析模型。例如，在农业试验中，要比较不同肥料对农作物产量的影响，同时考虑不同地块的土壤肥力差异。可以将土壤肥力相近的地块划分为一个区组，然后在每个区组内随机分配不同的肥料处理。这样，在分析数据时，可以将土壤肥力的差异从总变异中分离出来，更精确地评估肥料对产量的影响。

3.2.3. 拉丁方设计

当试验涉及三个因素，且每个因素的水平数相等时，可采用拉丁方设计。该设计能同时控制两个区组因素的影响，使试验的精度更高。在概率论与数理统计中，拉丁方设计可以用于研究多个因素之间的交互作用。例如，在研究不同品种的种子、不同施肥量和不同灌溉方式对农作物产量的影响时，如果种子、施肥量和灌溉方式都有三个水平，可以采用拉丁方设计安排试验。通过这种设计，可以分析每个因素的主效应以及因素之间的交互作用，为农业生产提供更科学的决策依据。

3.3. 试验设计在课程教学中的应用实例

3.3.1. 案例一：产品质量检验

假设某工厂生产一种产品，有 A、B、C 三种生产工艺，现要检验不同工艺对产品质量的影响。可采用完全随机设计，从生产线上随机抽取一定数量的产品，分别采用三种工艺进行生产，然后对产品的质量指标进行测量和分析。从概率论的角度来看，随机抽取产品可以保证样本具有代表性，使得基于样本的统计推断能够推广到整个生产过程。设产品质量指标为随机变量 X，其分布可能受到生产工艺的影响。通过计算不同工艺下产品质量指标的均值、方差等统计量，并进行假设检验(如检验三种工艺下产品质量指标的均值是否相等)，可以判断不同工艺对产品质量是否有显著影响。如果假设检验的结果拒绝原假设，说明至少有一种工艺与其他工艺在产品质量上存在显著差异，从而为企业改进生产工艺提供依据。

3.3.2. 案例二：学生学习效果研究

为研究不同学习环境(安静、嘈杂)和学习时间(上午、下午、晚上)对学生学习效果的影响，可采用拉丁方设计。选择若干名学生作为试验对象，将他们随机分配到不同的学习环境和学习时间组合中，进行一段时间的学习后，对学生的学习成绩进行测试。在这个案例中，学习环境和学习时间是两个区组因素，学生的学习成绩是响应变量。利用方差分析等方法分析学习环境、学习时间及其交互作用对学习效果的

影响。从概率论的角度来看，学生的学习成绩可以看作是一个随机变量，其分布受到学习环境和学习时间等因素的影响。通过拉丁方设计，可以有效地控制这两个因素的影响，更准确地评估它们对学习成绩的主效应和交互效应。例如，如果发现学习环境和学习时间存在显著的交互作用，那么意味着不同学习环境下，学习时间对学习效果的影响可能不同，这为学生合理安排学习时间和选择学习环境提供了科学依据。

4. 大数据子抽样技术的引入

4.1. 大数据子抽样技术的原理

在大数据时代，数据量呈爆炸式增长，直接处理海量数据往往面临计算资源和时间的限制。大数据子抽样技术基于统计学原理，从海量数据中抽取具有代表性的子样本，通过对子样本的分析来推断总体特征。其核心思想是在保证一定精度的前提下，降低数据规模，提高分析效率。例如，利用随机抽样、分层抽样等方法，从大数据集中抽取部分样本，使子样本能够反映总体的分布特征和统计规律。这与概率论中的抽样分布理论密切相关。根据中心极限定理，当样本量足够大时，样本均值的抽样分布近似服从正态分布。通过合理的抽样方法，可以在不损失太多信息的情况下，用子样本的统计量来估计总体参数。

4.2. 大数据子抽样技术的方法步骤

4.2.1. 确定抽样目标和总体

明确研究目的，确定需要分析的总体数据。例如，研究某地区居民的消费行为，总体即为该地区所有居民的消费数据。这一步骤需要对研究问题有清晰的理解，明确要研究的总体范围和特征，为后续抽样提供准确的方向。

4.2.2. 选择抽样方法

根据总体特征和研究目的，选择合适的抽样方法。如总体数据具有明显分层结构，可采用分层抽样；若总体数据分布较为均匀，可采用简单随机抽样。在选择抽样方法时，需要考虑数据的特点、研究问题的性质以及对抽样误差的要求等因素。例如，在分析某电商平台用户的购买行为时，如果已知用户按照年龄、性别等因素具有明显的分层特征，那么采用分层抽样可以提高样本的代表性。

4.2.3. 确定样本容量

综合考虑总体规模、数据变异程度、允许误差和置信水平等因素，运用统计学公式计算所需的样本容量。例如，在进行总体均值估计时，可根据公式$n = z^2\sigma^2/E^2$ (其中n为样本容量，z为对应置信水平的标准正态分布分位数，σ为总体标准差，E为允许误差)计算样本量。确定合适的样本容量是保证抽样精度和效率的关键，样本容量过小可能导致抽样误差过大，无法准确推断总体特征；样本容量过大则会增加不必要的计算成本和时间。

4.2.4. 确定样本容量抽取子样本

按照选定的抽样方法和确定的样本容量，从总体数据中抽取子样本。在抽取过程中，要确保抽样的随机性和独立性，避免抽样偏差。例如，在进行简单随机抽样时，可以使用随机数生成器来确定抽取的样本个体。对于分层抽样，要在每一层内进行独立的随机抽样，以保证各层样本的随机性和代表性。

4.2.5. 确定样本容量对子样本进行分析和推断

对抽取的子样本进行统计分析，计算相关统计量，并根据样本统计量对总体特征进行推断和估计。例如，计算子样本的均值、方差、比例等统计指标，并构建置信区间估计总体参数。通过对子样本的分

析，可以在一定置信水平下对总体特征进行推断，如总体均值、总体比例等。同时，还可以进行假设检验等统计推断，以验证关于总体的某些假设是否成立。

4.3. 大数据子抽样技术在课程教学中的应用实例

4.3.1. 案例一：网络用户行为分析

某互联网公司拥有海量用户行为数据，要分析用户的点击行为特征。若直接对所有数据进行分析，计算成本高昂且耗时。可采用分层抽样，根据用户的地域、年龄、性别等特征将总体数据分为若干层，然后从每层中随机抽取一定数量的用户数据作为子样本。从概率论的角度来看，不同地域、年龄、性别的用户可能具有不同的点击行为模式，分层抽样可以保证每个子群体都有足够的代表性。对该子样本进行分析，如计算不同类型用户的点击频率、平均停留时间等统计量，进而推断总体用户的行为模式，为公司优化网站设计和营销策略提供依据。例如，通过分析发现某地区年轻女性用户的平均点击频率较高，且在特定页面的停留时间较长，公司可以针对这一用户群体优化相关页面的内容和布局，提高用户体验和转化率。

4.3.2. 案例二：市场调研数据分析

在进行市场调研时，收集了大量消费者对某产品的评价数据。为快速了解消费者的总体满意度，可先采用简单随机抽样抽取一部分评价数据作为子样本。在这个过程中，每个评价数据被抽中的概率相等，符合简单随机抽样的原理。对该子样本进行情感分析，统计正面评价、负面评价和中性评价的比例，并计算平均满意度得分。通过对子样本的分析结果，在一定置信水平下估计总体消费者的满意度情况，为企业产品改进和市场决策提供参考。

5. 教学实践与效果评估

5.1. 教学实践过程

1) 在课程教学中，将试验设计和大数据子抽样技术融入到相关知识点的讲解中。例如，在讲解抽样分布时，引入大数据子抽样技术，让学生通过实际操作从大数据集中抽取子样本，计算样本统计量并观察其分布规律，加深对抽样分布理论的理解。

2) 安排课程实验项目，要求学生运用试验设计方法设计实验方案，并结合大数据子抽样技术对实验数据进行收集和分析。例如，在市场调研实验中，学生分组设计调查方案，确定抽样方法和样本容量，采集数据后利用所学技术进行处理和分析，最后撰写实验报告。

3) 鼓励学生参与实际项目或竞赛，如大学生数学建模竞赛等。在竞赛过程中，引导学生运用试验设计优化模型构建过程，利用大数据子抽样技术处理海量数据，提高模型的准确性和实用性。

5.2. 教学效果评估

5.2.1. 学生成绩分析

对比改革前后学生的课程考试成绩，发现学生在涉及试验设计和大数据应用的题目上得分明显提高，整体成绩呈上升趋势。这表明学生对相关知识的掌握和应用能力得到了有效提升。

5.2.2. 学生实践能力评价

通过学生在课程实验和实际项目中的表现进行评价。观察到学生能够熟练运用试验设计方法设计合理的实验方案，准确采集和分析数据，并能根据结果提出有针对性的建议。在大数据子抽样技术的应用方面，学生能够根据实际问题选择合适的抽样方法和样本容量，有效处理和分析大数据，实践能力得到显著增强。

5.2.3. 学生学习兴趣和创新思维调查

采用问卷调查和课堂互动的方式了解学生的学习兴趣和创新思维变化。结果显示，大部分学生对融入试验设计和大数据子抽样技术的教学内容表现出浓厚兴趣，课堂参与度明显提高。在解决实际问题过程中，学生能够主动思考，提出创新性的解决方案，创新思维得到了有效培养。

6. 结论与展望

6.1. 教学探索成果总结

通过将试验设计和大数据子抽样技术融入“概率论与数理统计”课程教学，有效解决了课程教学中的部分痛点问题。试验设计方法的应用使学生更深入理解概率统计知识在实际中的应用，提高了学生的实践能力和解决问题的能力；大数据子抽样技术的引入让学生适应大数据时代的数据处理需求，培养了学生的数据分析思维和创新能力。教学实践结果表明，这些教学方法的创新显著提升了学生的学习效果和综合素质。

6.2. 对未来教学的展望

未来教学中，可进一步优化试验设计和大数据子抽样技术的教学内容和方法。例如，引入更多实际案例和前沿研究成果，让学生接触到更复杂、更具挑战性的问题，拓宽学生的视野；加强与其他学科的交叉融合，培养学生的综合应用能力；利用在线教学平台和虚拟实验室等技术手段，为学生提供更加丰富的学习资源和实践环境，提升教学质量和效果。同时，持续关注行业发展动态，及时更新教学内容，使学生所学知识与实际应用紧密结合，为学生未来的职业发展和学术研究奠定更坚实的基础。

基金项目

信息与计算科学一流重点专业建设(356651535416)；北京交通大学人才基金项目资助(2023XKRC052)；国家自然科学基金青年基金项目(NSFC12201042)。

参考文献

[1] 耿直. 大数据时代统计学面临的机遇与挑战[J]. 统计研究, 2014, 31(1): 5-9.

[2] 王泽龙, 刘吉英, 余奇. 概率论与数理统计课程情境化教学模式探索与实践[J]. 高教学刊, 2024, 10(31): 121-124.

[3] 金今姬, 高彦伟. PBL 教学模式在“概率论与数理统计”教学中的应用[J]. 长春师范大学学报, 2023, 42(10): 152-157.

[4] Kazak, S. and Pratt, D. (2021) Developing the Role of Modelling in the Teaching and Learning of Probability. *Research in Mathematics Education*, **23**, 113-133. https://doi.org/10.1080/14794802.2020.1802328

[5] 肖敏, 徐静, 唐叶云. 《概率论与数理统计》教学改革思考[J]. 教育进展, 2021, 11(4): 1090-1094.

[6] 肖进胜, 杨力衡, 丁玲, 张海剑. 现代数理统计中假设检验的教学探讨[J]. 高教学刊, 2024, 10(8): 117-120.

[7] 王贶, 朱靖红. “概率论与数理统计”课程思政教学研究——以全概率公式和贝叶斯公式为例[J]. 辽宁工业大学学报(社会科学版), 2024, 26(2): 133-135.

[8] 方开泰, 刘民千, 周永道. 试验设计与建模[M]. 第 2 版. 北京: 高等教育出版社, 2024.

[9] 方开泰, 刘民千, 覃红, 周永道. 均匀试验设计的理论和应用[M]. 北京: 科学出版社, 2019.

[10] Fang, K.T., Liu, M.Q., Qin, H. and Zhou, Y.D. (2018) Theory and Application of Uniform Experimental Designs. Springer and Science Press.

[11] Wang, H., Yang, M. and Stufken, J. (2018) Information-Based Optimal Subdata Selection for Big Data Linear Regression. *Journal of the American Statistical Association*, **114**, 393-405. https://doi.org/10.1080/01621459.2017.1408468

[12] Wang, H., Zhu, R. and Ma, P. (2018) Optimal Subsampling for Large Sample Logistic Regression. *Journal of the American Statistical Association*, **113**, 829-844. https://doi.org/10.1080/01621459.2017.1292914