

大数据时代下概率论与数理统计课程教学探索

刘 然^{1*}, 李晨晖²

¹北京交通大学数学与统计学院, 北京

²清华大学政策研究室, 北京

收稿日期: 2024年6月12日; 录用日期: 2024年7月16日; 发布日期: 2024年7月23日

摘 要

概率论与数理统计课程作为一门研究随机现象模型和数据分析的大学数学公共基础课, 在培养学生数据素养和创新能力方面发挥着重要作用。为了适应大数据时代的要求, 本文从两个方面对概率论与数理统计课程教学的改进做了探讨。第一, 创新教学方法, 要充分使用可视化、系统化等方法, 生动自然地引入知识点, 做到理论联系实际, 利用大数据时代的新技术, 达到提高教学效果的目的。第二, 坚持立德树人, 强化课程思政。

关键词

大数据, 概率论与数理统计, 教学探索, 课程思政

Exploration on the Teaching of Probability and Statistics in the Era of Big Data

Ran Liu^{1*}, Chenhui Li²

¹School of Mathematics and Statistics, Beijing Jiaotong University, Beijing

²Policy Research Office, Tsinghua University, Beijing

Received: Jun. 12th, 2024; accepted: Jul. 16th, 2024; published: Jul. 23rd, 2024

Abstract

As a public basic course of college mathematics, Probability and Statistics which studies stochastic phenomenon model and data analysis plays an important role in cultivating students' data literacy and innovation ability. In order to meet the requirements of the era of big data, this paper discusses the teaching improvement of Probability and Statistics from two aspects. First, to innovate

*通讯作者。

teaching methods and improve teaching effect, we should make full use of visualization and systematization methods. We also need to introduce knowledge points vividly and naturally, integrate theory with practice and take advantage of new technologies in the era of big data. Second, adhere to the task on fostering integrity and promoting rounded development of people and strengthen ideological and political education.

Keywords

Big Data, Probability and Statistics, Teaching Exploration, Ideological and Political Education

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

21 世纪以来, 大数据技术及其相关联的云计算、人工智能、区块链等技术加速创新, 日益融入经济社会发展各领域全过程。数字化时代带来了统计数据的巨大变化, 大数据成为时代最鲜明的特征和标志之一[1]。面对数据的获取方式和呈现形式愈发多样、数据体量爆发式增长等特征, 经济社会发展各领域都对提升本领域数据分析能力提出了迫切需求。特别是随着人工智能的兴起, 如何从大数据中寻找有价值的信息, 对数据进行正确的统计分析进而厘清庞杂数据之间的相互关系, 日益成为不同学科背景下的学生适应大数据时代发展的基本科学素养和技能。在此背景下, 概率论与数理统计(以下简称概率统计)课程作为一门研究随机现象模型和数据分析的大学数学公共基础课, 蕴含着数据分析的数学基础与基本方法, 也为学生学习后续高级数据分析方法打下了基础, 正在发挥越来越重要的作用[2]。

为了适应大数据时代的要求, 提高学生数据分析能力, 使其更好地利用概率统计的思想方法解决各种随机现象相关的问题, 需要对概率统计课程的教学做出与时俱进的改革和创新。国内外已有大量研究从多个方面对概率统计课程的教学改革做了探讨, 例如, 在教学模式方面, 金今姬和高彦伟[3]探讨了问题驱动式学习模式在概率统计教学中的应用; 在教学方法方面, Kazak 和 Pratt [4]讨论了概率教学中综合建模方法的重要作用和面临的挑战, 肖敏等[5]建议在教学中重视知识背景和实际需求, 采用信息化教学手段和案例教学法改进教学; 在具体教学主题方面, 肖进胜等[6]研讨了现代数理统计中假设检验的教学过程, 提出要强化原假设设计方法, 并通过实例说明了按照“小概率事件原理”来设计假设的方案; 在课程思政方面, 王颀和朱靖红[7]以全概率公式和贝叶斯公式为例, 探讨了在教学中开展课程思政的材料和方式。本文在既往研究的基础上, 结合大数据时代的特点和要求, 对概率统计课程的教学方法创新和课程思政建设这两个方面做了探讨。

2. 与时俱进创新教学方法, 夯实概率统计知识基础

对于概率统计课程的授课教师来说, 如何通过得当的教学设计、先进的教学手段帮助学生准确理解课程中的概念和方法是十分重要的事情。我们在授课过程中就这个问题做了一些探索和创新, 概要总结为下述五点。

2.1 借助可视化助力知识点讲授

概率统计课程中知识点的难度很大程度上来源于其抽样性, 因而, 教师对应的授课策略应该是尽力

让抽象的概念和方法具象化。而采取图片、动画等可视化工具是实现具象化的有效方法。比如,在讲授区间估计的概念时,按照频率学派的观点,参数是一个没有随机性的数,但估计参数的置信区间却涉及到概率。如果教师仅仅从抽象概念出发解释这里的概率不是指随机的参数在两个具体的估计上下限之间的概率,而是指作为随机变量的估计上下限涵盖固定的参数的概率,学生们往往会感到不明所以,难以理解其实际意涵。这时,我们可以利用可视化手段,将参数的值画在数轴的一个固定位置上,通过计算机模拟等手段抽取多个样本,根据每个样本做出一个区间估计,再将这些区间一一表示在数轴上。这样,数轴上有一个固定位置的参数值的点,而因为样本的随机性每一个估计区间彼此不同,估计上下限的随机性和未知参数的非随机性就一目了然了[8]。又比如,二元正态分布是概率统计课程要学习的重要分布,但是,由于其概率密度函数参数较多、形式较为复杂,学生往往会迷惑于抽象的数学符号,不能直观地把握二元正态分布的特征,更不能理解二元正态分布和一元正态分布的关系。这时,我们可以绘制二元正态分布的三维密度函数图,让学生直观把握二元正态分布的形式特点,并根据投影、截面等方法从图像出发建立起二元正态分布和一元正态分布的联系,从而减轻学生的学习难度。

2.2. 系统化地讲授知识点

概率统计课程中知识点之间的相互关系是讲授过程中的一个难点,如果只是按照顺序逐一进行讲授,学生经常会混淆各种相关的概念和方法,难以掌握它们之间的关系。但如果我们在授课过程中有意识地在讲授过程中对知识点进行结构化,以系统化的方式进行讲授,勾勒出知识点之间内在的逻辑关系,就能够起到事半功倍的效果。

系统化地讲授知识点,首先要注意新的知识与学生已有知识之间的衔接,帮助学生建立新旧知识之间的联系。这样既可以降低学习难度,减轻学生的畏难情绪,又能够帮助学生加深对新知识的印象。比如,在讲授概率论基础时,样本空间、随机事件等概率论的内容其实都是用集合论的语言进行描述的,我们要向学生清楚解释这一点,点出随机事件之间的关系与运算其实就是集合之间的关系与运算。这样就把新的知识和学生学过的集合论的知识建立起了联系,大大减轻了学生的学习负担和畏难情绪。又比如在讲授连续型随机变量函数的分布运算时,分布计算需要用到微积分课程中学过的积分计算知识,这部分基础有所欠缺的同学可能由于没有掌握或者已经忘记积分的计算方法而跟不上进度。这时,我们可以拿出一定时间对相关计算做扼要复习,帮助学生衔接好新旧知识。有时也可以通过建立与其他课程知识的联系来帮助学生加深和拓展对所学知识的理解,比如在讲授回归模型时,可以与微积分课程中常微分方程的知识建立联系,简要介绍基于常微分方程的回归模型,拓宽学生的概念视野。

系统化地讲授知识点,还要注重知识之间的相互关系。思维导图、表格、二叉树等都是可用的系统化工具。比如,多元随机变量部分涉及大量新概念,是学生学习的一个难点。这时,我们可以将主要概念列为一个 3×3 的表格,如表1所示。其中,横向的三个词代表三个前缀:联合、边缘、条件,而纵向的三个词代表三个主要概念:分布函数、分布律、概率密度函数,表格三三组合出九个概念,而这九个概念正是学习的重点。我们在授课时只需指出,在学习一元随机变量时已经学过分布函数、分布律、概率密度函数这三个重要概念,其中分布函数是描述随机变量性质的通用概念,而分布律和概率密度函数分别是针对离散型和连续型随机变量的专门概念,多元随机变量也仍然要用这三个概念,只不过需要考虑同时关注多个随机变量、只关注一个随机变量、关注部分随机变量给定时另外的随机变量这三种情况,而这三种情况恰好对应了联合、边缘、条件这三个前缀。进一步,我们可以通过该表格方便地掌握概念之间的关系,例如分布函数和分布律、概率密度函数之间是求和-差分、积分-求导的关系,联合的概念除以边缘的概念能够得到条件的概念。这样,我们就深入浅出地把多元随机变量部分的繁复概念系统化为少数几个关键概念,大大简化了学习的过程,也加深了学生的印象。

Table 1. Summary of important concepts of multivariate random variables
表 1. 多元随机变量重要概念总结

	联合	边缘	条件
分布函数(通用)	联合分布函数	边缘分布函数	条件分布函数
分布律(离散型)	联合分布律	边缘分布律	条件分布律
概率密度函数(连续型)	联合概率密度函数	边缘概率密度函数	条件概率密度函数

2.3. 知识点引入要生动自然

在讲授概率统计课程的知识点时,简单地列举概念和性质难以消除学生初学时的“陌生感”和“生硬感”。这就需要注意讲授知识的顺序,发掘和利用生动活泼的事例自然地引入知识。比如,泊松分布是重要的离散型随机变量分布。但是由于其分布律形式较为复杂,直观意义不如之前学习的伯努利分布、二项分布明晰,学生掌握起来比较困难。如果我们按照常见的顺序,先一般地先给出泊松分布的分布律形式,再告诉学生其描述的对象,最后推导泊松定理,那么学生常会感到不明所以,不知道为什么可以把泊松分布的分布律形式和单位时间或空间上的计数过程联系起来,也往往会机械地把泊松定理看作是一种数学上的计算方法,无法理解泊松分布和二项分布的深刻联系。这时,我们可以改变讲授顺序,通过二项分布的极限计算引出泊松分布,由此自然地引入泊松分布和泊松定理。比如,我们可以从计算玻璃液体单位体积液体中的气泡数这一有趣例子出发来进行讲解,在一定条件下,这一概率可以用二项分布进行计算,并且随着液体中的气泡总数趋于无穷,该概率将收敛到泊松分布。这样,学生们就能够容易地理解为什么泊松分布能够描述单位时间或空间上的计数过程,并在推导过程中自然地掌握泊松定理[9]。

运用科学史素材是生动自然地引入知识点的有效策略。科学史上的故事既可以起到引人入胜的作用,又可以让学生对知识点有一个扼要直接的认识,更可以使学生了解知识的现实应用。比如,在讲解假设检验时,由于学生之前从未接触过假设检验这一全新的统计方法,一下子面对零假设、备择假设、显著性水平、两类错误等繁多概念时容易感到无所适从、难以掌握。这时,我们可以从统计学史上经典的女士品茶案例出发来切入。简要地说,该案例是现代统计学之父费舍尔运用假设检验方法解决问题的真实史事:在一次品茶会上,有一位女士声称能够分辨奶茶制作工艺是倒奶入茶还是倒茶入奶,为了鉴别女士说法的真伪,费舍尔用假设检验的观点,从女士不能分辨奶茶制作工艺的假设出发,通过实验收集女士分辨奶茶正确与否的数据来检验这个假设,最后否定了原来的假设[10]。这则科学史材料生动有趣、贴近生活,可以有效地引发学生的兴趣;不涉及复杂的数学计算,即使是数学基础较差的学生也可以理解,使学生集中注意力到假设检验的本质特征和关键步骤上;让学生认识到假设检验在生活中的应用。这样,我们就自然地引入了假设检验这一新的统计学方法,降低了学生的学习障碍。

2.4. 联系实际学以致用

在大数据时代的背景下,无论是在日常生活里还是专业领域中,处处可以见到概率统计知识的用武之地。但是,一些传统教材更多强调知识点的理论性,对概率统计知识的实际应用重视不足,影响了学生对概率统计知识的深入理解和应用能力。为弥补这一缺憾,我们广泛搜集日常生活和专业领域的概率统计应用案例用作教学,以此将知识点和现实实际相联系。这样可以引发学生兴趣,让学生了解概率统计的重要作用,进而加深学生对统计学方法的理解。

在讲授概率统计知识时,要善于利用新闻等时事材料,为学生提供理论联系实际的方法和路径。比

如,在讲授假设检验时,可以向学生展示我国新药研发临床试验成功的新闻,向学生说明评价新药的安全性和有效性,其方法本质上是通过试验收集数据,然后用试验组和对照组的数据做假设检验。这样既向学生展示了假设检验方法的巨大现实作用,也加深了学生对假设检验方法的理解。我们还可以选择较为简单的应用案例作为课堂分析材料,带领学生分析解决现实问题,比如,针对疫情防控中的核酸检测方法,可以引导学生从数学期望的角度出发,思考多合一混采检测技术相对于单采技术的优越性。

科研论文也是一种理论联系实际的有效工具。统计学方法在物理学、生物学等自然科学,经济学、社会学等社会科学乃至历史学等人文学科的研究中都有着广泛而重要的应用,许多论文的结论都是应用统计学方法分析数据得到的。在授课过程中,我们可以通过课堂讨论、课后小组探究等方式让学生学习一些利用统计学方法做出结论的论文。比如,我们可以选择权威期刊《经济研究》上刊载的“重点产业政策与制造业就业”一文作为材料,向学生讲授该文章在论证中运用的回归和假设检验等统计学方法,说明怎样通过统计学方法得到科学的结论[11]。这个过程既引发学生兴趣,又提供了额外的科学训练,更加深了学生对统计学方法的理解。

2.5. 结合大数据时代的新技术改进概率统计教学

大数据时代下,教育教学与大数据技术及其关联的人工智能、虚拟现实等先进信息技术深度融合逐渐成为高等教育的发展趋势,利用大量数据训练得到的生成式人工智能大模型展示出在生成文字、图像、代码、视频等方面的强大能力,给教育带来无穷想象。在概率统计教学中,要广泛了解、积极尝试使用大数据时代的新技术,更新教学手段,通过新技术对教学设计与教学内容、教学场景与教学资源、教学模式与学习方式、学情分析与教学评价进行改革创新,提高学生学习效率。例如,可以利用人工智能大模型智能学习概率统计教材、大纲、论文等各类教学数据来提取知识点,再结合学生个人作业等个性化数据,为每位学生建立个性化的概率统计知识图谱,帮助他们有效梳理所学知识;可以使用人工智能大模型建立概率统计 AI 助教,随时与学生互动,对他们的问题进行解答反馈,并借助互动数据对课堂教学质量做出评价,帮助教师发现并完善授课中的不足;可以借助人工智能实现快速蒙特卡洛编程,让编程知识基础不足的学生也能体会到统计模拟的作用,全面提升教学质量和学生体验。

3. 加强课程思政建设, 能力培养和价值塑造并重

教育的根本任务在于立德树人,这就要求我们把专业知识和思想政治教育相结合。课程思政不是简单的“课程”加“思政”,而是要深入挖掘课程中的思政元素,在专业课程中巧妙加入、有机融入思政内容,努力实现二者的水乳交融,做到“如盐化水”、润物无声。在概率统计课程上,我们探究数理专业知识背后的思想内涵,在统计学理论中挖掘哲理,引导学生用统计学的观点来认知世界,培养逻辑性思维与创造性想象的能力,同时使学生潜移默化中受到正确价值观的熏陶。下面,我们从三个角度扼要论述概率统计课程中可供挖掘的思政内容。

第一,概率统计知识深刻体现了辩证唯物主义世界观。比如,大数定律深刻地揭示出频率和概率的深刻联系,展现出唯物主义世界观中偶然性和必然性的辩证关系。中心极限定理则通过阐释众多变量的加和作用,深刻解释了现实世界中人群身高、体重等大量常见分布类似于正态分布的背后机理。

第二,概率统计知识的发现过程蕴含着求真精神。概率统计的学科发展史是人类认识世界、追求真理的历史,在概率统计学科发展过程中留名的重要学者,比如伯努利、高斯、贝叶斯、柯尔莫哥洛夫、费舍尔、高尔顿等数学家和统计学家,都展示出求真唯实的探索精神。一方面,要充分讲授概率论和统计学的发展历程,突出其中求真探索的过程。比如,讲授上文提到的女士品茶的例子,让学生体会现代统计学之父费舍尔是怎么样从一个有趣的日常辩论出发,求真务实,探索其中的科学道理,最后总结凝

练出假设检验的方法。又比如, 可以讲授高尔顿研究父代和子代身高提出回归分析的例子, 让学生理解高尔顿是怎么样从数据出发, 通过科学分析提出回归分析并发现回归现象的。另一方面, 可以引导学生通过做实际实验、思考科学问题、进行计算机模拟等方法, 自己发现概率统计知识, 在这个过程中领略科学求真的方法, 感受探索发现的乐趣。比如, 让学生做高尔顿板的实验, 引导他们发现随着实验次数的增多, 二项分布越来越近似于正态分布这一重要的概率论现象, 为接下来中心极限定理的讲授打好基础。

第三, 概率统计知识的重要应用生动展示着科学奉献精神。在授课过程中, 我们可以向学生讲授概率统计知识在大数据时代中的重要作用, 特别是在人工智能、新药研发等重要领域的应用, 带领学生体会所学知识的“大用”, 引导学生努力求学, 学以致用, 为人类进步、社会发展和国家富强贡献力量。

4. 结语

随机现象在现代生活中无处不在, 需要运用概率统计知识深入探讨。在大数据时代下, 概率统计知识在信息技术、航空航天、生物医药、人工智能等高科技领域都起到了重要作用, 推动人们认识现实世界背后的规律, 为正确决策提供定量依据。因此, 在教学活动中要加强概率统计课程建设, 改变传统的教学模式, 提升学生学习兴趣和学习自主性。在课程设计中, 结合实际应用生动自然地引入知识点, 通过可视化的形式对新知识点进行系统化讲授。同时要强化课程思政, 引导学生从被动的听课转变为主动的学习探究, 进而在能力、科学思维和价值观方面给学生们以更大的启迪。

基金项目

中央高校基本科研业务费基础研究项目(2022JBMC040); 北京交通大学校级教改项目(SQ20240292); 国家自然科学基金青年基金项目(NSFC12201031)。

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] 茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程[M]. 北京: 高等教育出版社, 2019.
- [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, 等. 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] 王颀, 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] 看见统计. 置信区间[EB/OL]. <https://seeing-theory.brown.edu/frequentist-inference/cn.html#section2>, 2024-03-08.
- [9] 李启寨, 孟珍. 概率论与数理统计[M]. 北京: 高等教育出版社, 2023.
- [10] 萨尔斯伯格. 女士品茶[M]. 刘清山, 译. 南昌: 江西人民出版社, 2016.
- [11] 王贤彬, 陈春秀. 重点产业政策与制造业就业[J]. 经济研究, 2023, 58(10): 34-54.