

现代统计学本科生培养的课程体系与路线图

统计学是一门与时俱进的应用学科，它的研究问题和研究手段是“常为新”的。统计学植根于早年的农牧业、后来生物学、以及当今的 IT 互联网等领域，这驱动和丰富了统计学的研究问题。而数学特别是概率论给统计带来基础支撑，使得从经验中获得的方法系统化、普适化，从而奠定了统计学的学科基础和体系。计算技术的迅速发展给统计学带来新的发展阶段，算法的作用日益突出，与机器学习交融贯通拓宽了统计学科边界。近年来，数据科学、人工智能以等领域强势崛起又给统计学带来新的挑战 and 机遇。

但是，我们同时应清醒认识到我国统计学本科生培养体系不能完全适应现代统计学发展以及国家、市场对人才的需求，特别是不能有效服务于创新、交叉人才的脱颖而出。存在专业细化、知识老化、教材陈旧等问题。比如，统计学核心课程有：概率论、数理统计、随机过程，多元回归分析、多元统计分析、抽样调查、试验设计、生存分析、时间序列、非参数统计、生物统计等等。传统课程多、细，且内容重叠。但算法或计算课程少，培养体系还没有完全适应计算机科学带来的红利。

统计学可以分为三个层面：应用、方法/算法和理论。因此，课程体系应该围绕这三个层次建设，服务不同人的不同需求。我个人认为本科生培养应该重点在于方法、计算、原理和基础理论，而弱化专门应用领域知识训练，这应留在研究生阶段。统计学是较为通识的金牌专业，不必要再设置二级本科专业，也不需要任何形容词来加持。

统计学课程建设总体思路应该以“深基础、宽领域、高视野”为宗旨，强调概率论、随机过程和数理统计等课程的基础理论，强调统计模型、机器学习和算法设计等培养建模、分析和计算能力的课程。基础理论课程建设关键在于课程内容和相应教材。针对学生不同层面的需求，概率论、随机过程和数理统计三门基础课北京大学分别设立了荣誉和普通两套培养体系。

在方法/算法层面，建议考虑设置三门宽口径专业课程。第一门是“统计学原理”或“统计思维”，该课程围绕估计、推理、学习、置信区间、假设检验讲授其中统计原理和思想，包括频率派和贝叶斯派，可以取材《all of statistics》(Larry Wasserman)中的第二部分，以及经验贝叶斯和信息论相关内容。第二门是“机器学习”或“统计模型与方法”，内容包括多元统计分析、多元回归分析、高维统计和贝叶斯分析等，可以取材《all of

statistics》中的第二部分、《Computer Age Statistical Inference》(Bradley Efron & Trevor Hastie)第三部分和《Information Theory, Inference, and Learning Algorithms》(David MacKay)部分内容。第三门课是“统计计算与随机模拟”，可以参考《Computational Statistics》(Geof Givens and Jennifer Hoeting)和《Simulation》(Sheldon Ross)。此外，如果有条件，建议开设一门高年级专业选修课《高等机器学习》来讲述深度学习、强化学习和因果学习等机器学习的较为前沿主题。

超越本专业，拓宽学生知识，特别是培养学生的计算能力，建议学生选修计算机课程“计算机系统导论”（参考《Computer Systems: A Programmer's Perspective》by Bryant & O'Hallaron）和“并行编程”（参考《Programming massively parallel processors》by David Kirk & Wen-mei Hwu）。另外，我注意到耶鲁大学统计与数据科学系 Daniel Spielman 教授最近开设了一门课“Computation and Optimization”（S&DS631）。这门课内容涵盖了数值线性代数、优化和自动微分等内容，对统计和数据科学专业学生非常有用，值得我们借鉴和引进。

数据科学与大数据技术作为统计学与计算机科学的交叉专业，其本科生课程体系，我设想设置六门课：“统计学大全”（参考 Larry Wasserman 书）、“计算与优化”（Daniel A Spielman, S&DS631）、“机器学习”（参考 David MacKay 书）、“算法导论”（参考 Thomas Cormen, Charles Leiserson, Ronald Rivest, and Clifford Stein 书）、“计算机系统导论”（参考 Randal Bryant & David O'Hallaron 书）、“并行编程”（参考 David Kirk & Wen-mei Hwu 书）。我相信这六门课同样适用于人工智能专业。

教育承载着个人、国家、民族的期望和希望。为师者或教育为政者需要倾听学生的声音，洞察学科的前沿，知晓市场的需求，并由此调整培养体系和课程内容，推陈出新。前面提到耶鲁大学统计与数据科学系主任 Daniel Spielman 教授是美国科学院院士、世界顶级的理论计算机和应用数学学家，两次哥德尔奖获得者，他对教学情有独钟，热衷开新课。我博士后合作导师加州大学伯克利分校 Michael Jordan 教授是美国科学院、工程院、科学艺术与科学院院士，其学术成就名满天下，但是最让我们这些学生感动的是他对教学和教育孜孜不倦的热爱。有境界则自成高格，自有名士，学科能常青在此。

附注：本文根据笔者 2021 年 5 月 22 日在中国人民大学数据科学专业教学研讨会的报告内容整理而成。文章表叙的仅为我个人观点，与他人和团体无关。

张志华

2021年6月16日星期三