

Probability and Statistics

概率统计

南京大学

高 尉



Course Information and Textbooks

- Instructor: 高尉
 - gaow@nju.edu.cn
 - gaow@lamda.nju.edu.cn
- Office: 国际学院A-305



概率论与数理统计

- 盛骤、谢式千等编
- 高等教育出版社

Probability and statistics

- Morris H. DeGroot and Mark J. Schervi
- 机械工业出版社



概率的萌芽：赌博

两人进行一场赌博，5局3胜，赌金为1000；假设当前比分为2 : 1，而比赛由某种原因不得不中止。

问题：最“公平合理”的奖金分配方式？

— 1509年意大利数学家/修道士帕西奥尼 (Luca Pacioli)

这个问题持续了150年左右



Luca Pacioli

概率的起源

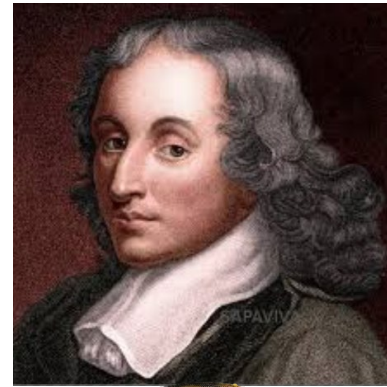
1650左右的法国

- 赌博流行且时尚,不受法律限制
- 赌博变得更加复杂, 风险增大
- 有必要通过数学方法来计算胜率



古典概型 (等可能概型)

- 有限个基本事件
- 每个基本事件发生的可能性相同



Blaise Pascal



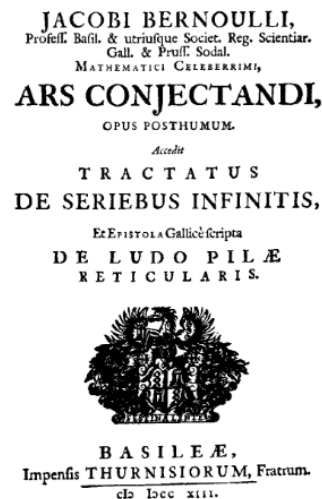
Pierre Fermat

概率论的形成和发展

《推想的艺术》 1713年

- 贝努利 (James Bernoulli)

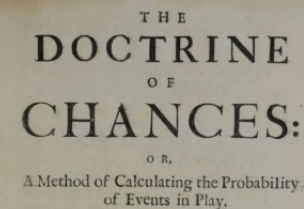
- 大数定律
- 频率稳定性理论化
- 特殊问题到一般理论



《机遇原理》 1781年

- 棣谟佛 (Abraham de Moivre)

- 概率乘法法则
- 正态分布律
- 为“中心极限定理”奠定了基础



概率论的形成和发展(续)

《偶然性的算术试验》 1706

-蒲丰(Comte de Buffon)

- 结合概率和几何，开始几何概率的研究
- 采取概率的方法来求圆周率的尝试

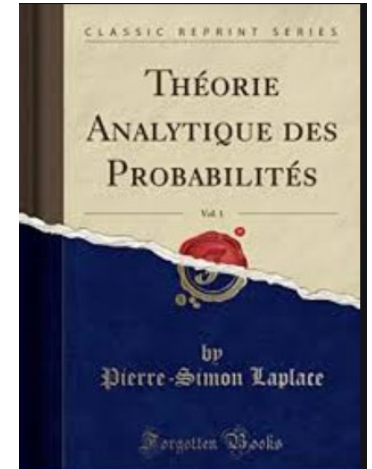


19世纪：概率论进一步发展-应用

Theorie Analytique des Probabilités

- Pierre-Simon Laplace

A mathematical theory of probability with an emphasis on scientific applications



高斯
Carl F. Gauss



麦克斯韦
James C. Maxwell



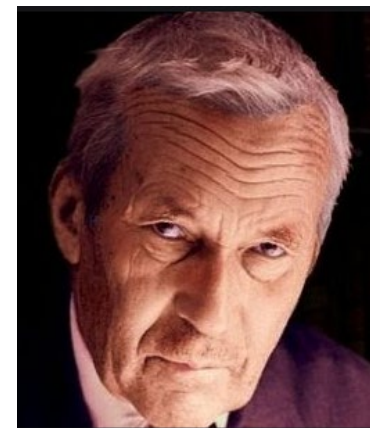
吉布斯
Josiah W. Gibbs

20世纪： 概率的公理化定义

Grundbegriffe der Wahrscheinlichkeitsrechnung 1933

- 柯尔莫哥洛夫(Andrey Kolmogorov)

建立概率公理化理论体系，利用基本性质来定义概率，可媲美于欧几里得几何公理化



概率公理化 (三点)

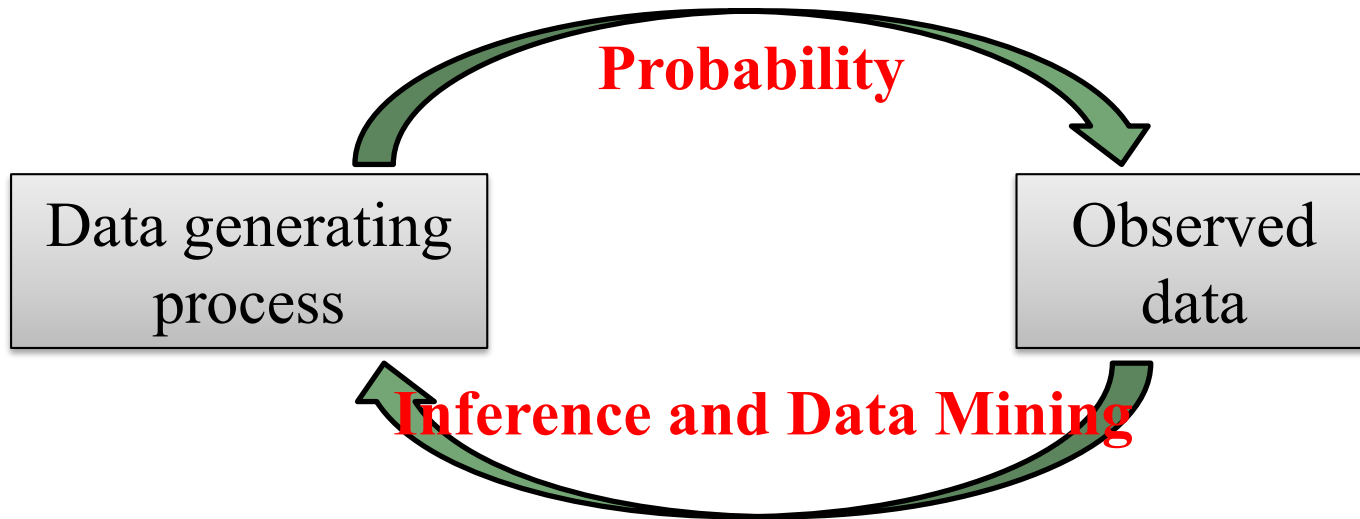
- 非负性
- 规范性
- 可列可加性

现代概率统计：测度论



概率与统计

- **概率**: 研究事件的不确定性, 在给定数据生成过程中观察、研究数据的性质, 强调**公理体系**、**推理**
- **统计**: 收集与分析数据, 根据观察的数据反思其数据生成过程, 强调“**归纳**” (与机器学习相关)



- 概率: 随机变量、分布、大数定律、概率不等式等
- 统计: 推断、参数估计、假设检验、模型、算法

应用



顶级国际期刊与会议

- 概率统计及其相关的顶级国际期刊
 - Annals of Statistics
 - Journal of the American Statistical Association (JASA)
 - Annals of Probability
 - [Journal of Machine Learning Research](#)
- 国际会议
 - ICML: International Conference on Machine Learning
 - COLT: Annual Conference on Learning Theory
 - STOC: ACM Symposium on Theory of Computing
 - FOCS: IEEE Symposium on Foundations of Computer Science

Researcher work (ICML/COLT/STOC 2020)

- DeltaGrad: Rapid retraining of machine learning models
- on the noisy gradient descent that generalizes as SGD
- club: a contrastive log-ratio upper bound of mutual information
- optimistic bounds for multi-output prediction
- on the generalization effects of linear transformations in data augmentation
- uniform convergence of rank-weighted learning
- online learning with imperfect hints
- the implicit and explicit regularization effects of dropout
- Single Point Transductive Prediction
- Sample Amplification: Increasing Dataset Size even when Learning is Impossible
- In Defense of Uniform Convergence: Generalization via Derandomization with an Application to Interpolating Predictors
- Deep k-NN for Noisy Labels
- Deep Streaming Label Learning
- Fine-Grained Analysis of Stability and Generalization for Stochastic Gradient Descent
- A Distributional Framework for Data Valuation
- Class-Weighted Classification: Trade-offs and Robust Approaches
- Extreme Multi-label Classification from Aggregated Labels
- structured prediction with partial labelling through the infimum loss
- Optimality and Approximation with Policy Gradient Methods in Markov Decision Processes
- Learning Mixtures of Linear Regressions in Subexponential Time via Fourier Moments
- Efficiently Learning Structured Distributions from Untrusted Batches

考核方式与先修课程

- 数学分析
 - 高等代数
 - 计算机编程
-
- Home work: 20% (每周1次, 下一周上课之前交)
 - Mid-Term exam: 20%
 - Final exam: 60%