

# Probability and Statistics

## 概率统计

南京大学

高 尉



# Course Information and Textbooks

- Instructor: 高尉
  - [gaow@nju.edu.cn](mailto:gaow@nju.edu.cn)
  - [gaow@lamda.nju.edu.cn](mailto:gaow@lamda.nju.edu.cn)
- Office: 计算机系楼919



## Probability and statistics

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



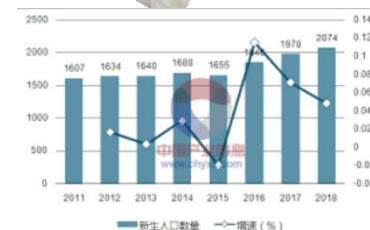
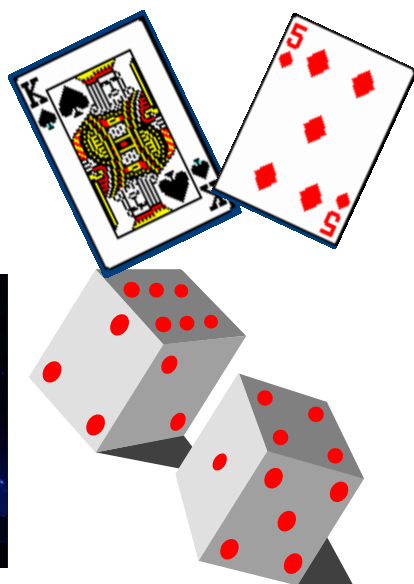
## 概率论与数理统计

- 盛骤、谢式千等编
- 高等教育出版社

# 我们生活的世界充满了随机现象

我们无时无刻不面对具有不确定性现象，即随机现象

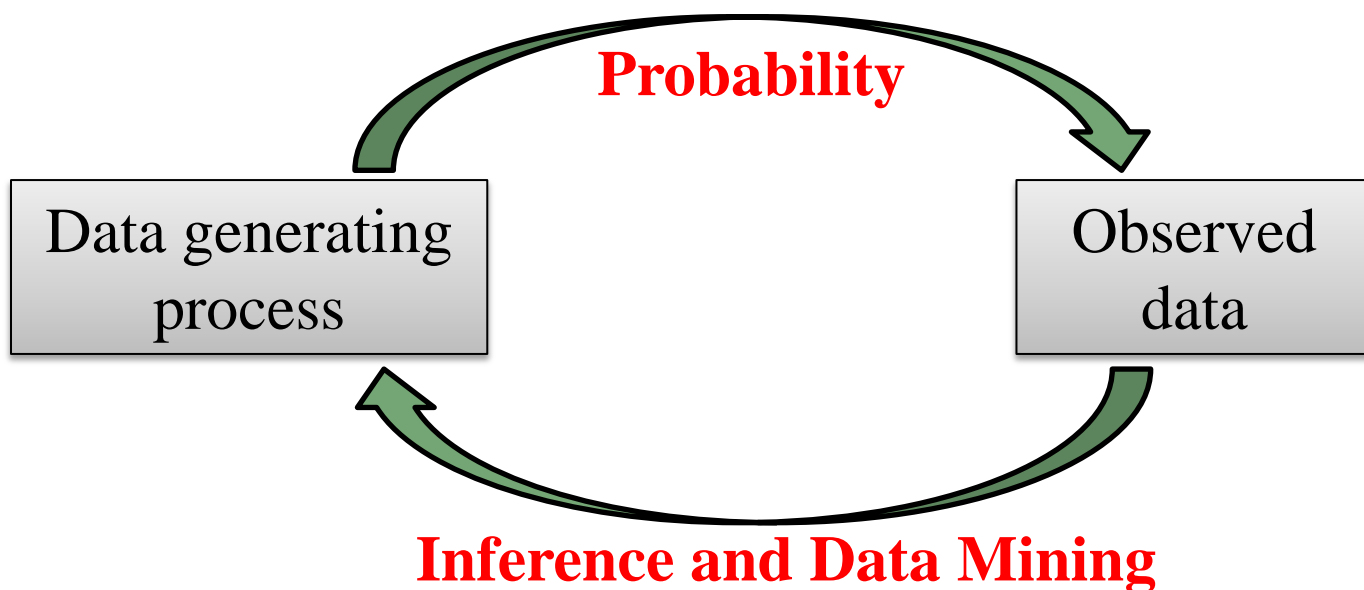
- 机会游戏：硬币、掷骰子和摸扑克、彩票 ...
- 社会现象：婴儿的诞生、流星殒落、蝴蝶效应 ...



**拉普拉斯：“生活中最重要的问题，绝大多数实质上是概率问题”**

# 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）

# 概率与统计（续）

---

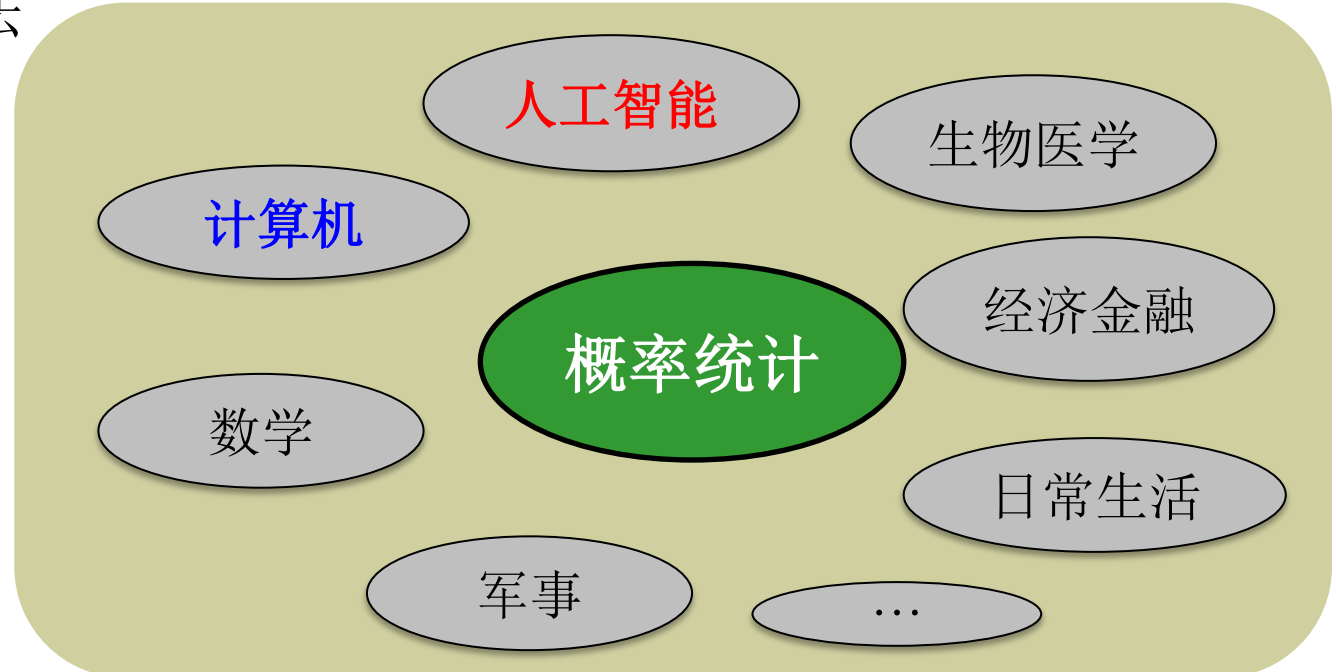
## □ 概率

- 随机变量、分布、大数定律、概率不等式等
- 随机过程

## □ 统计

- 推理：参数估计、假设检验
- 模型、算法

## Applications



## 16世纪：赌博与概率

---

数学家修道士帕西奥尼 (Luca Pacioli)  
大约于1509年提出了一个难题：

甲、乙两人进行一场赌博，5局3胜，  
赌金为100；假设当前比分为2 : 1，而  
比赛由某种原因不得不中止。



Luca Pacioli

最“公平合理”的奖金分配方式如何？

这个问题持续了150年左右



# 17世纪：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



赌徒兼数学家德米尔(C. De Mere)引起了极大兴趣, 向帕斯卡(Blaise Pascal)请教该问题

帕斯卡与费马(Pierre Fermat)通信讨论此问题, 独立迅速解决该问题。



Blaise Pascal



Pierre Fermat

## 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

# 18世纪：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

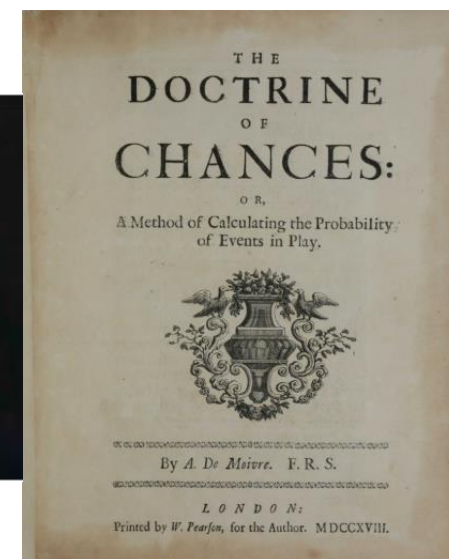


James Bernoulli

## 《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





# 18世纪：概率论的形成和发展

---

## 《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

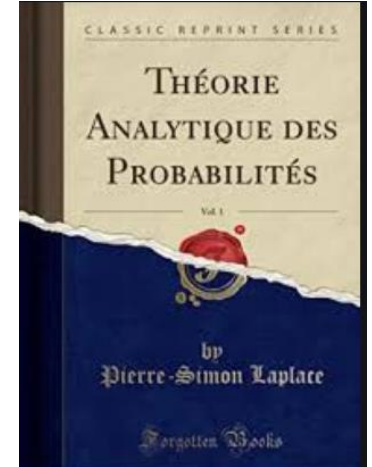


# 19世纪：概率论进一步发展-应用

## 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

- Andrey Kolmogorov

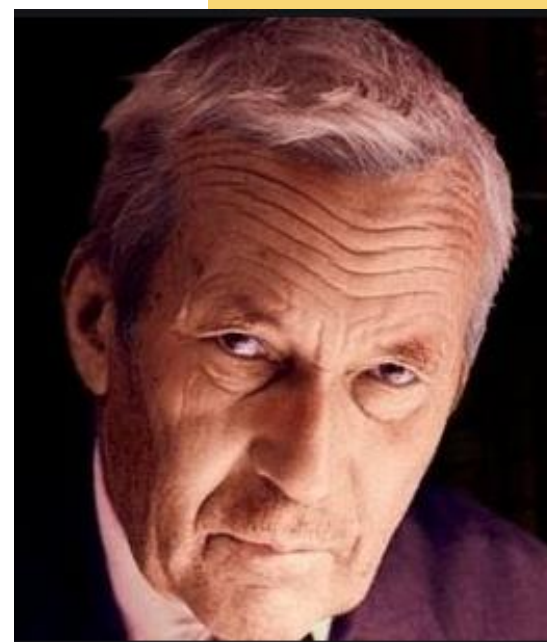
建立概率公理化理论体系，利用基本性质来定义概率，可媲美于欧几里得几何公理化

A. Kolmogoroff  
Grundbegriffe  
der  
Wahrscheinlichkeits-  
rechnung

### 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



Andrey Kolmogorov

# Contents

---

- 概率基本概念 (Counting)
- 条件概率与独立性
- 离散性随机变量及其分布
- 连续性随机变量及其分布
- 联合分布
- 期望、方差、相关性
- 大数定律与中心极限定律

# Contents

---

- 概率不等式 (Concentration)
- 蒙特卡洛采样 (Sampling、simulation)
- Random number
- 随机过程
- 样本与抽样分布
- 参数估计
- 假设检验
- 方差分析

# 考核方式与先修课程

---

- 数学分析
- 高等代数

- Home work: 20% (4-6次作业)
- Mid-Term exam: 20%
- Final exam: 60%



# Researcher work (ICML/COLT/STOC 2019)

---

- Non-Gaussian Component Analysis using Entropy Methods
- Performance of Johnson–Lindenstrauss Transform for k-Means and k-Medians Clustering
- Regression from Dependent Observations
- Approximation Algorithms for Distributionally-Robust Stochastic Optimization with Black-Box Distributions
- How Hard is Robust Mean Estimation
- Faster Algorithms for High-Dimensional Robust Covariance Estimation
- Fast Mean Estimation with Sub-Gaussian Rates
- High probability generalization bounds for uniformly stable algorithms with nearly optimal rate
- A Theory of Selective Prediction
- Combining Online Learning Guarantees
- Estimation of smooth densities in Wasserstein distance
- On the Computational Power of Online Gradient Descent
- On Mean Estimation for General Norms with Statistical Queries
- Lipschitz Adaptivity with Multiple Learning Rates in Online Learning
- Learning to Prune: Speeding up Repeated Computations
- Learning from Weakly Dependent data under Dobrushin's condition
- Is your function low-dimensional?
- Inference under Information Constraints: Lower Bounds from Chi-Square Contraction
- Parameter-Free Online Convex Optimization with Sub-Exponential Noise
- The Optimal Approximation Factor in Density Estimation
- The implicit bias of gradient descent on nonseparable data

# Researcher work (ICML/COLT/STOC 2019)

---

- The Complexity of Making the Gradient Small in Stochastic Convex Optimization
- Testing Mixtures of Discrete Distributions
- Space lower bounds for linear prediction in the streaming model
- Sharp Analysis for Nonconvex SGD Escaping from Saddle Points
- When can unlabeled data improve the learning rate?
- A Better k-means++ Algorithm via Local Search
- A Convergence Theory for Deep Learning via Over-Parameterization
- Adversarial Online Learning with noise
- Analyzing and Improving Representations with the Soft Nearest Neighbor Loss
- Are Generative Classifiers More Robust to Adversarial Attacks
- AUC<sub>μ</sub> A Performance Metric for Multi-Class Machine Learning Models
- Bayesian leave-one-out cross-validation for large data
- Better generalization with less data using robust gradient descent
- Boosted Density Estimation Remastered
- Bridging Theory and Algorithm for Domain Adaptation
- Concentration Inequalities for Conditional Value at Risk
- Fast Rates for a kNN Classifier Robust to Unknown Asymmetric Label Noise
- Faster Stochastic Alternating Direction Method of Multipliers for Nonconvex Optimization
- Geometric Losses for Distributional Learning
- Improving Adversarial Robustness via Promoting Ensemble Diversity
- Improving Model Selection by Employing the Test Data
- Learning from a Learner
- On Symmetric Losses for Learning from Corrupted Labels