SI140A: Probability & Statistics for EECS

Ziyu Shao

School of Information Science and Technology
ShanghaiTech University

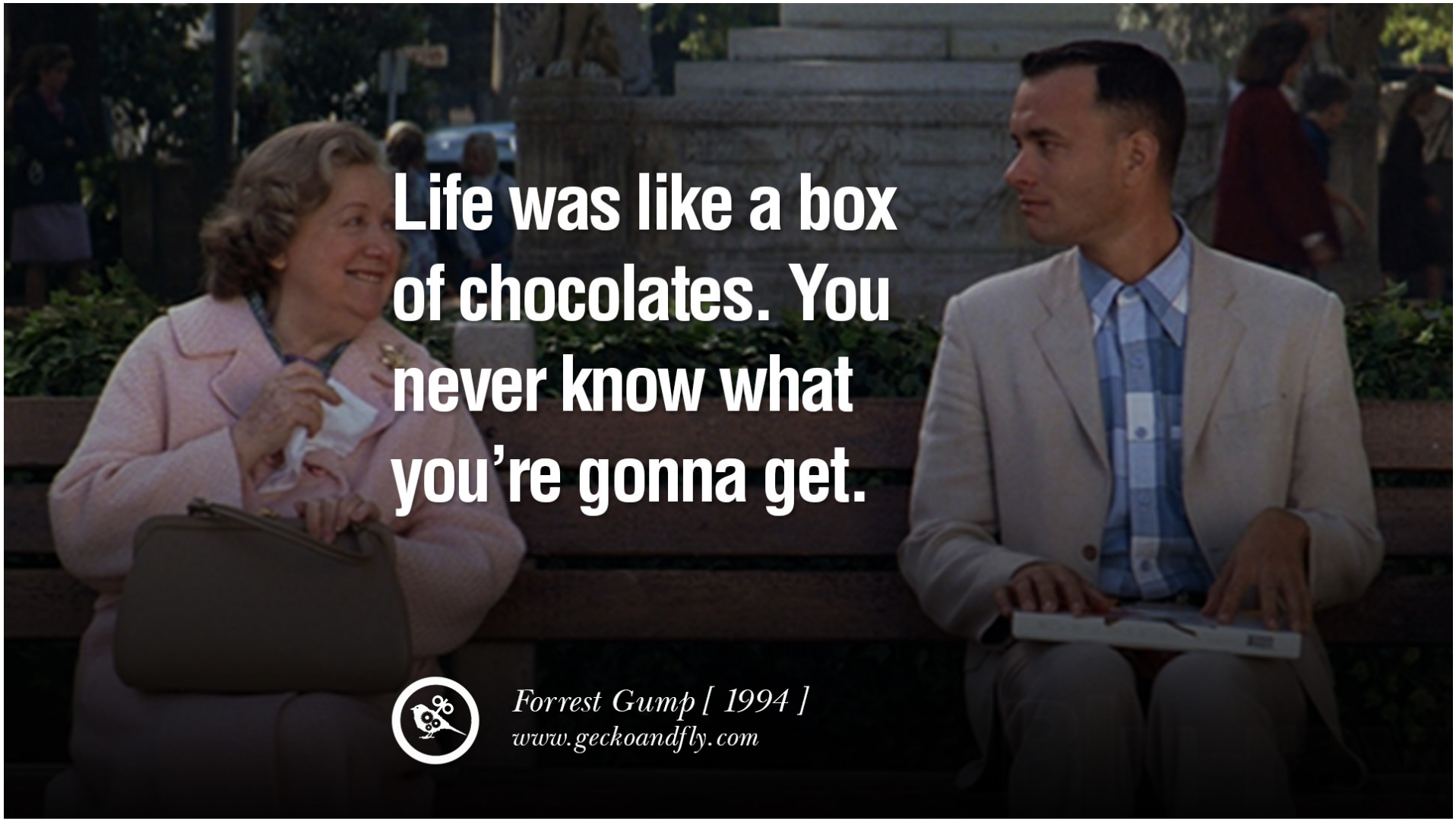
February 27, 2024

Outline

Outline

The Dark Forest in Random Kingdom

Life is Random



Life is Random

- **Cicero**:"Probability is the very guide of life".
- **Aristotle**: "The probable is what usually happens".
- **Baruch de Spinoza**: "In practical life we are compelled to follow what is most probable; in speculative thought we are compelled to follow truth".

Life is Random

- **Pierre-Simon Laplace**: "Probability theory is nothing but common sense reduced to calculation".
- **James Clerk Maxwell**:" Therefore the true logic for this world is the calculus of Probabilities".
- **Sherlock Holmes**:" When you have eliminated the impossible, whatever remains, however improbable, must be the truth".

Random Related Description in Chinese Words

- 机会
- 机遇
- 命运
- 奇迹
- 缘分
- 无常
- 投机
- 冒险
- 孤掷一注

Even Great Scientists Failed

- Isaac Newton: "I can calculate the motion of heavenly bodies, but not the madness of people."

Even Great Scientists Failed

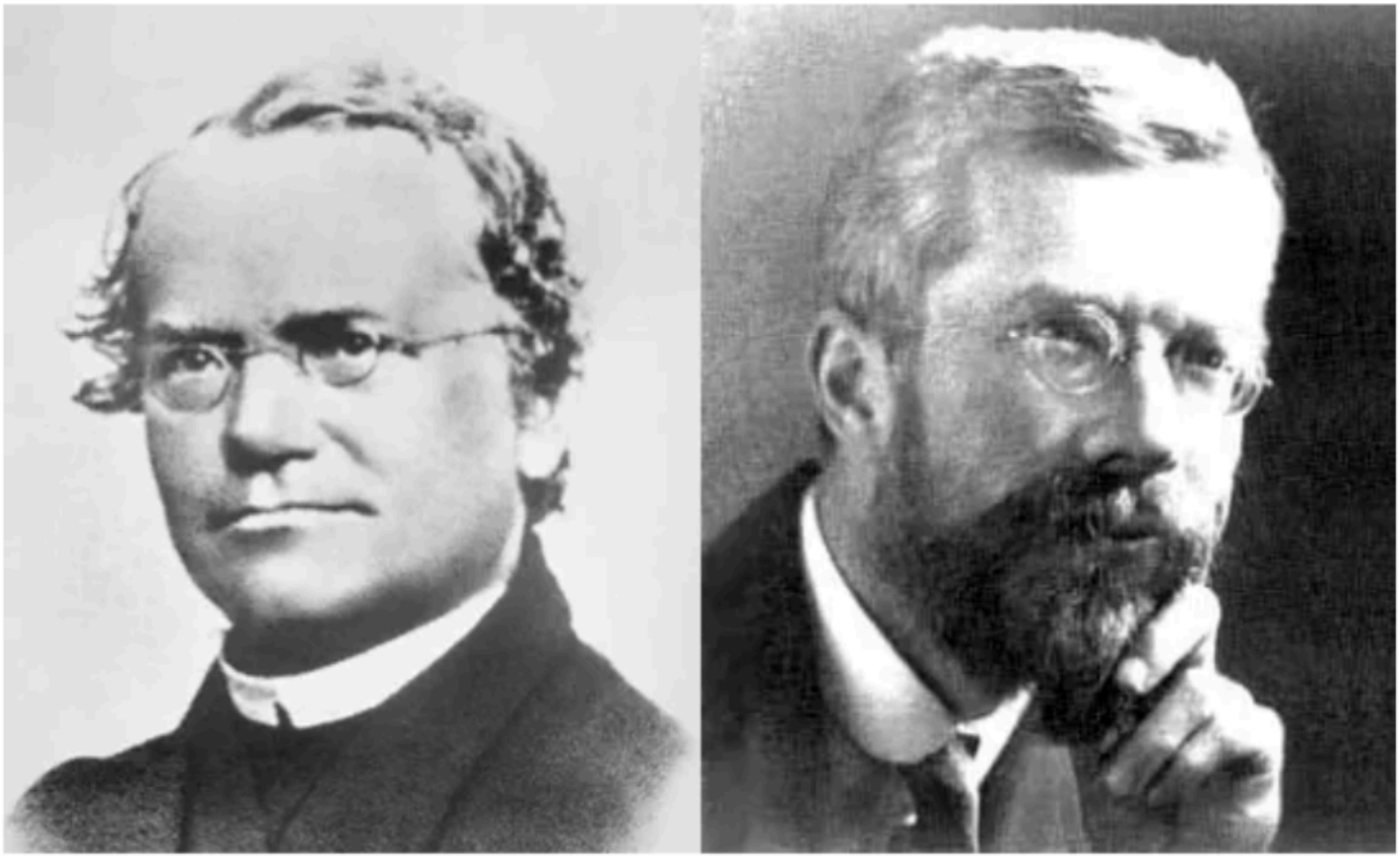
- Albert Einstein: “God does not play dice with the universe.”

Sometimes Mysterious

- Schrodinger's Cat: "遇事不决，量子力学；解释不通，穿越时空."

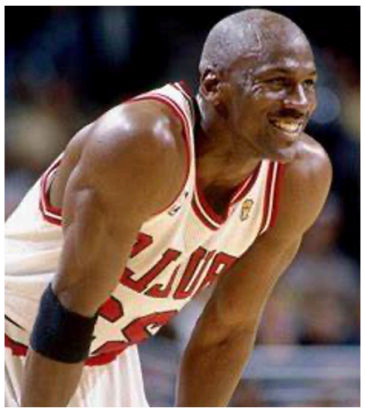
Sometimes Controversial

- Mendel-Fisher Controversy: Cherry Picking?



Sometimes Controversial

Case 3:Who is the greatest NBA player?



VS

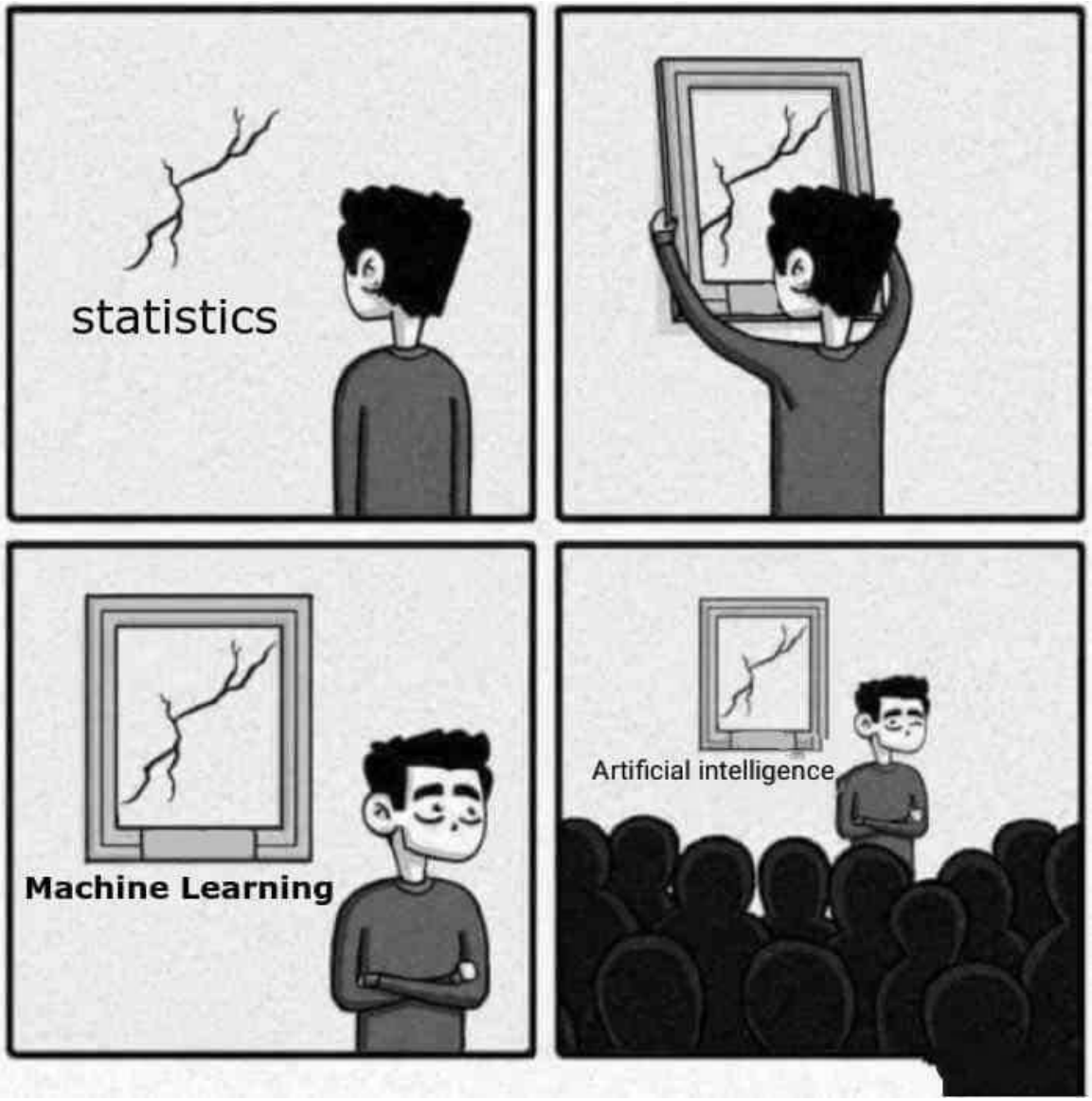
篮球得分	乔丹	詹姆斯
二分球	12192/24537 = 49.7%	12424/24654 = 50.4%
三分球	581/1778 = 32.7%	1860/5409 = 34.4%
总计	12773/26315 = 48.5%	14284/30063 = 47.5%

Sometimes Reinventing the Wheel

Reduction of random noise has many names in different fields.

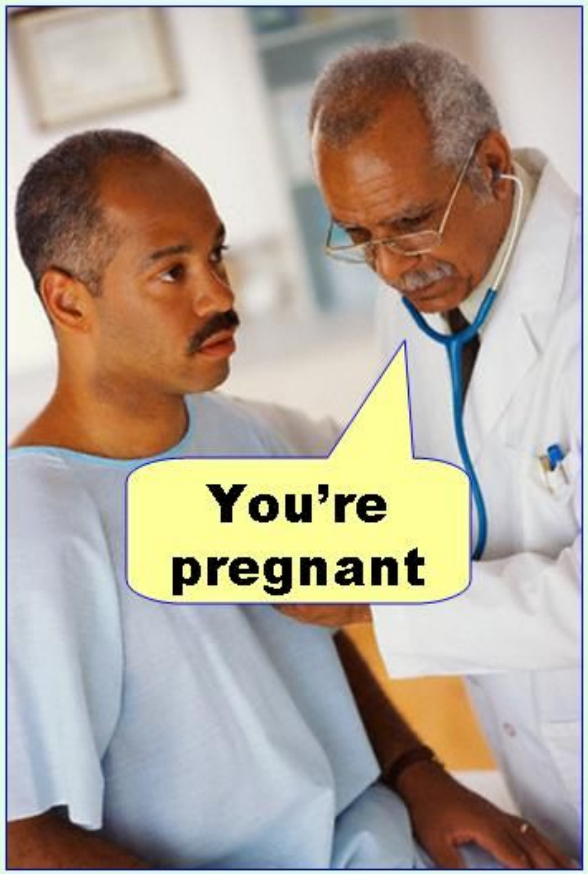
- Terminology in EECS: signal processing (filter)
- Terminology in Probability & Statistics: estimation
- Terminology in ML & AI: prediction

Sometimes Reinventing the Wheel

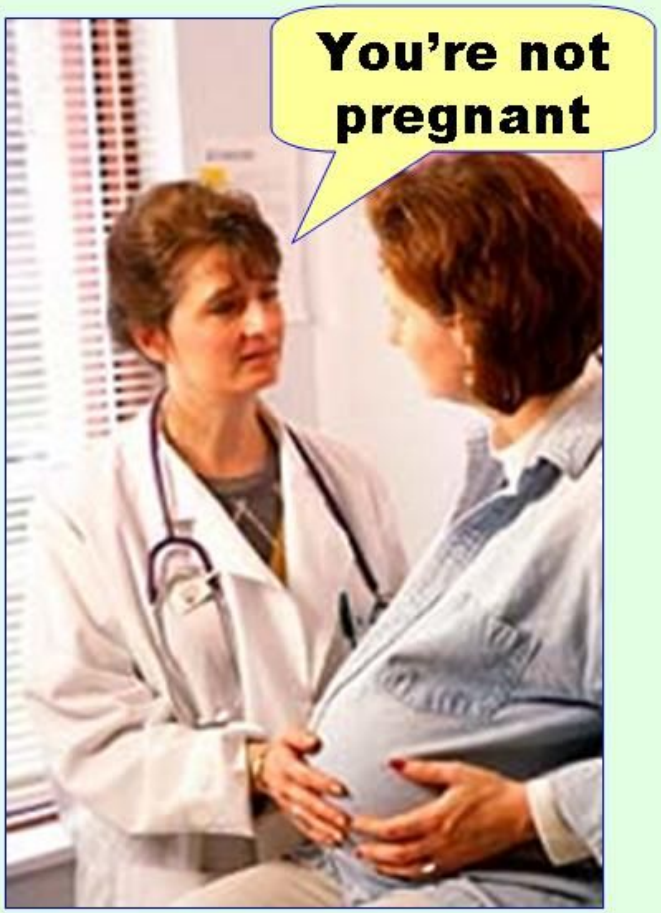


Sometimes Making Mistakes

Type I error
(false positive)



Type II error
(false negative)

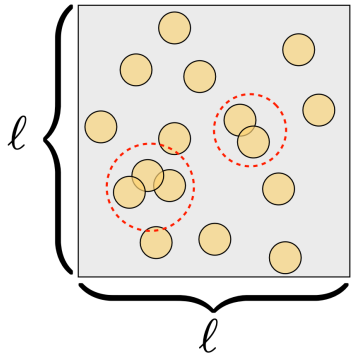


Sometimes Challenging

- John von Neumann (1951): given a biased coin with unknown probability of landing heads (unknown p), how can you generate unbiased (fair) coin-flips? ($f(p) = 0.5$)
- How can you generate coin-flips with any probability of landing heads? ($0 < f(p) < 1$)

Sometimes Challenging

Pennies on a Carpet
(hard spheres in 2D square)



- Drop n pennies on a square-shape carpet at random. What is the probability that no two pennies will overlap?

- In 1-dimension (n needles on a line segment):

$$\begin{cases} \left(\frac{\ell - nd}{\ell - d}\right)^n & \text{if } \ell \geq nd \\ 0 & \text{otherwise} \end{cases}$$

- In 2-dimension: Nothing is known about this problem. as of 1979–98.

硬球(hard spheres)模型：This problem is one of the most important problems of statistical mechanics. If we could answer it we would know, for example, why water boils at 100°C, on the basis of purely atomic computations.

M. Jerrum, H. Guo (2021): 高效数值求解hard spheres模型的蒙特卡罗算法

Gian-Carlo Rota teaching 18.313

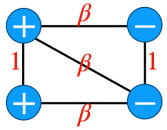
Sometimes Challenging

Phase Transition of Ferromagnetism

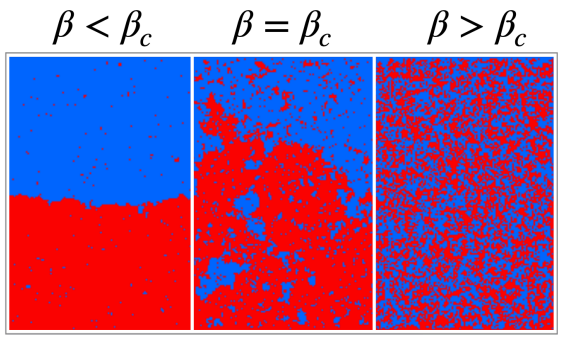
(critical behavior of Ising model)

- Ising model (Lenz 1920):
- Given a graph $G(V,E)$ and $\beta \in (0,1)$, each $\sigma \in \{-1,+1\}^V$ is assigned a weight:

$$w(\sigma) = \beta^{\sum_{uv\in E} |\sigma_u - \sigma_v|/2}$$



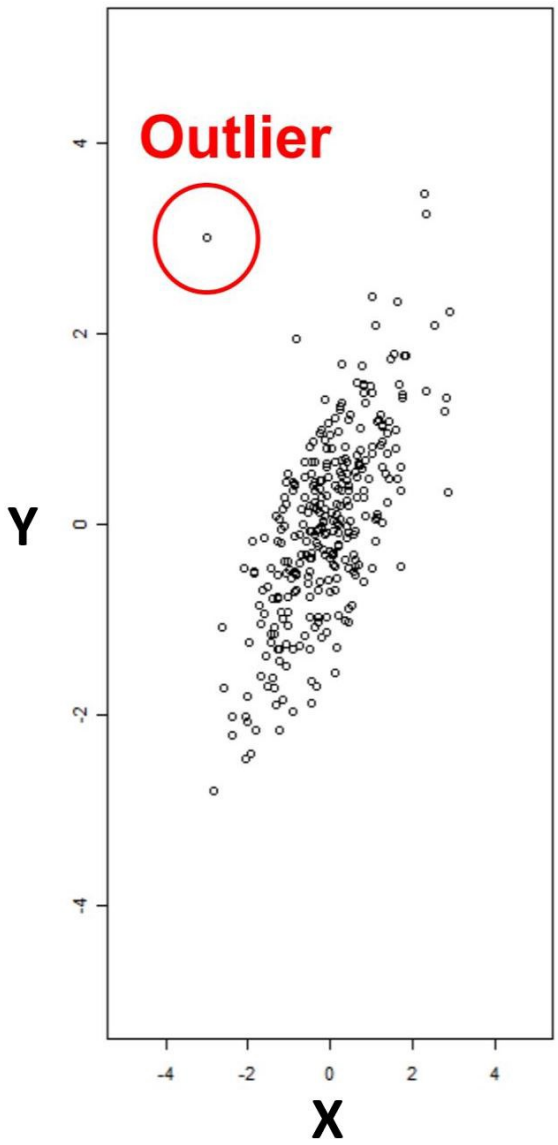
- Generate a random $\sigma \in \{+1,-1\}^V$ with probability $\propto w(\sigma)$.
- "Resolve" the model:
  - Ising (1D), Onsager (2D)
  - Jerrum-Sinclair '89: Monte Carlo algorithm

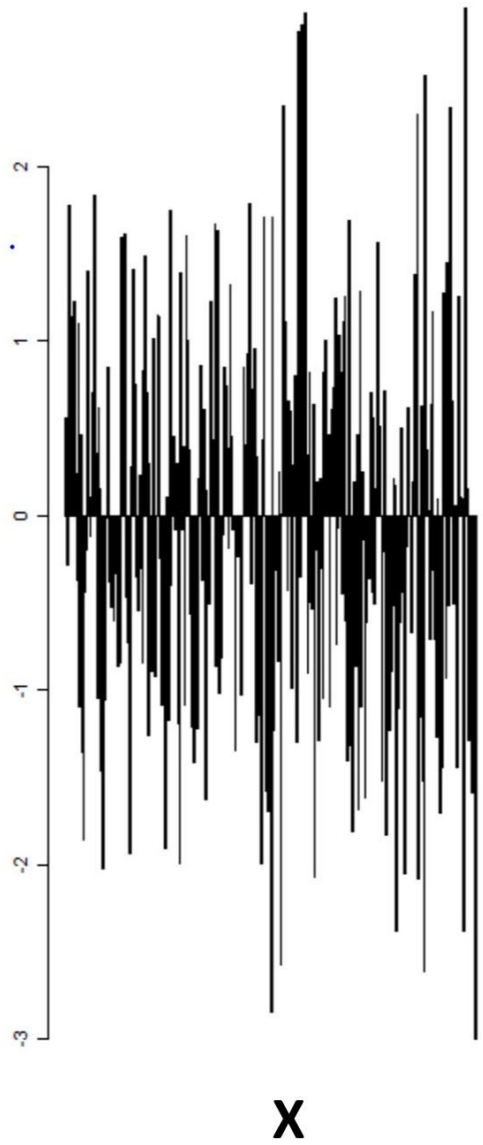


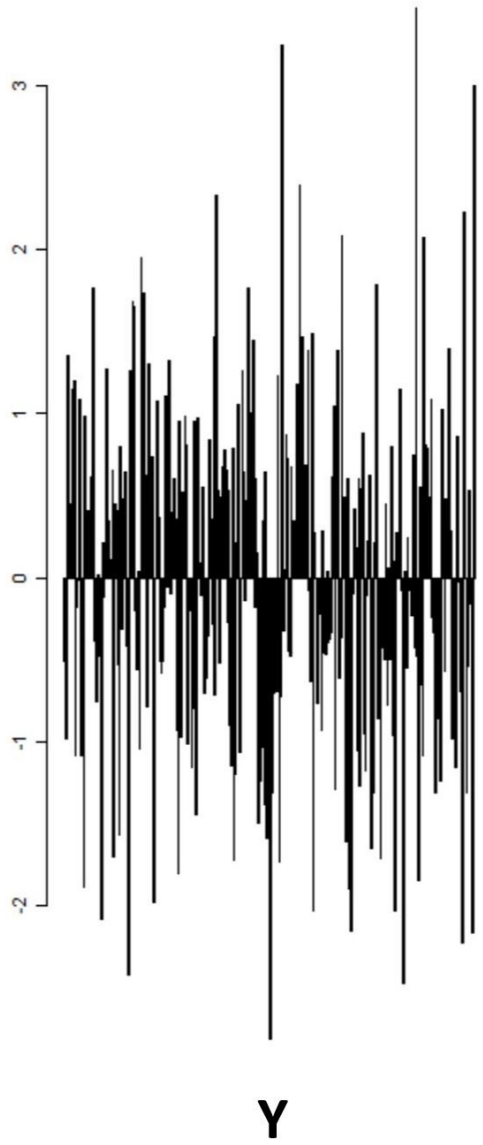
Ising (1924), Onsager (1944), Lee-Yang (1952)

Kaufman, Onsager, Yang, Kac, Ward, Potts, Montroll, Hurst, Green, Kasteleyn, McCoy, Wu, Vdovichenko, Fisher, Baxter, ...

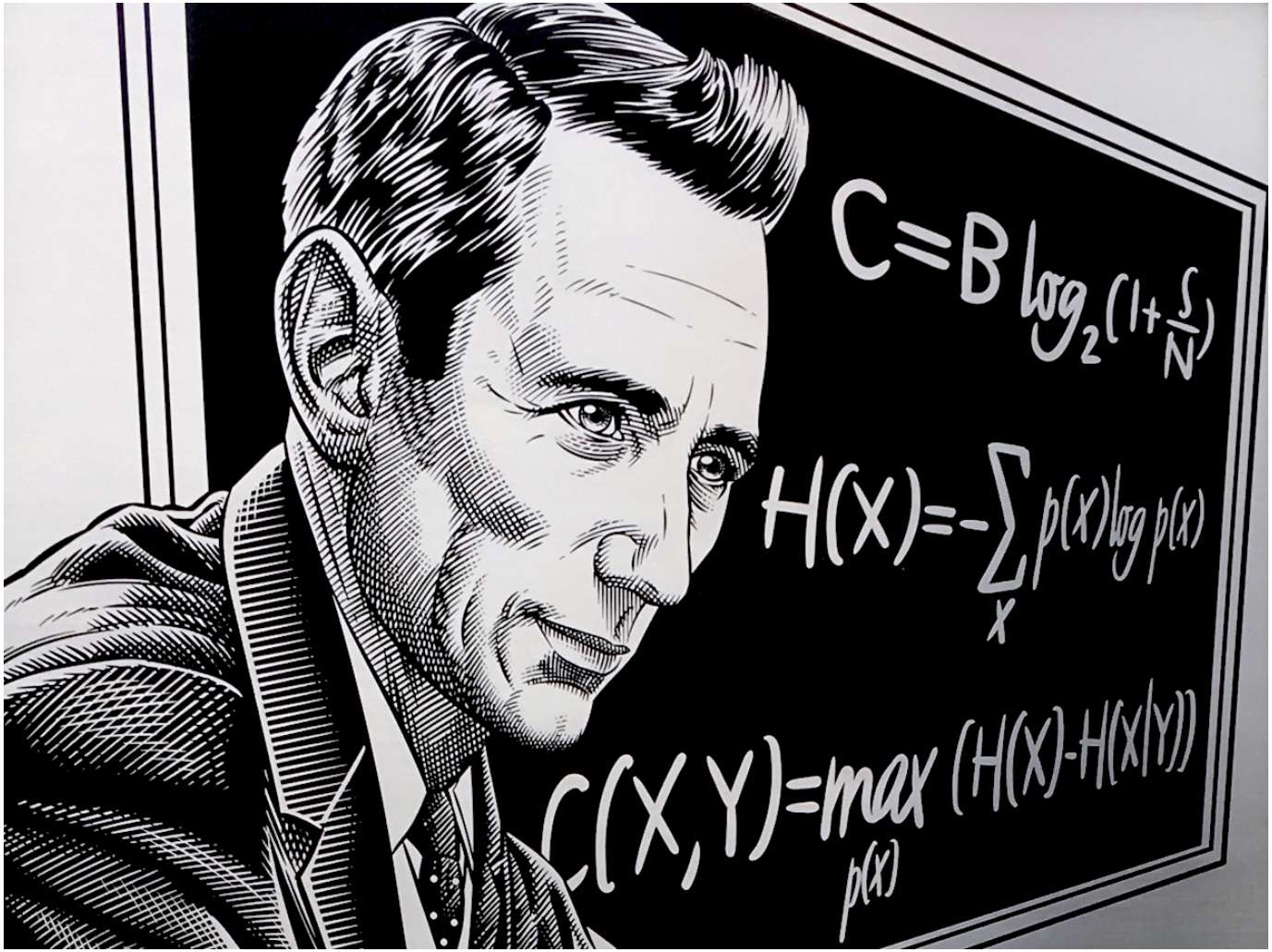
But Always Powerful: Anomality Detection



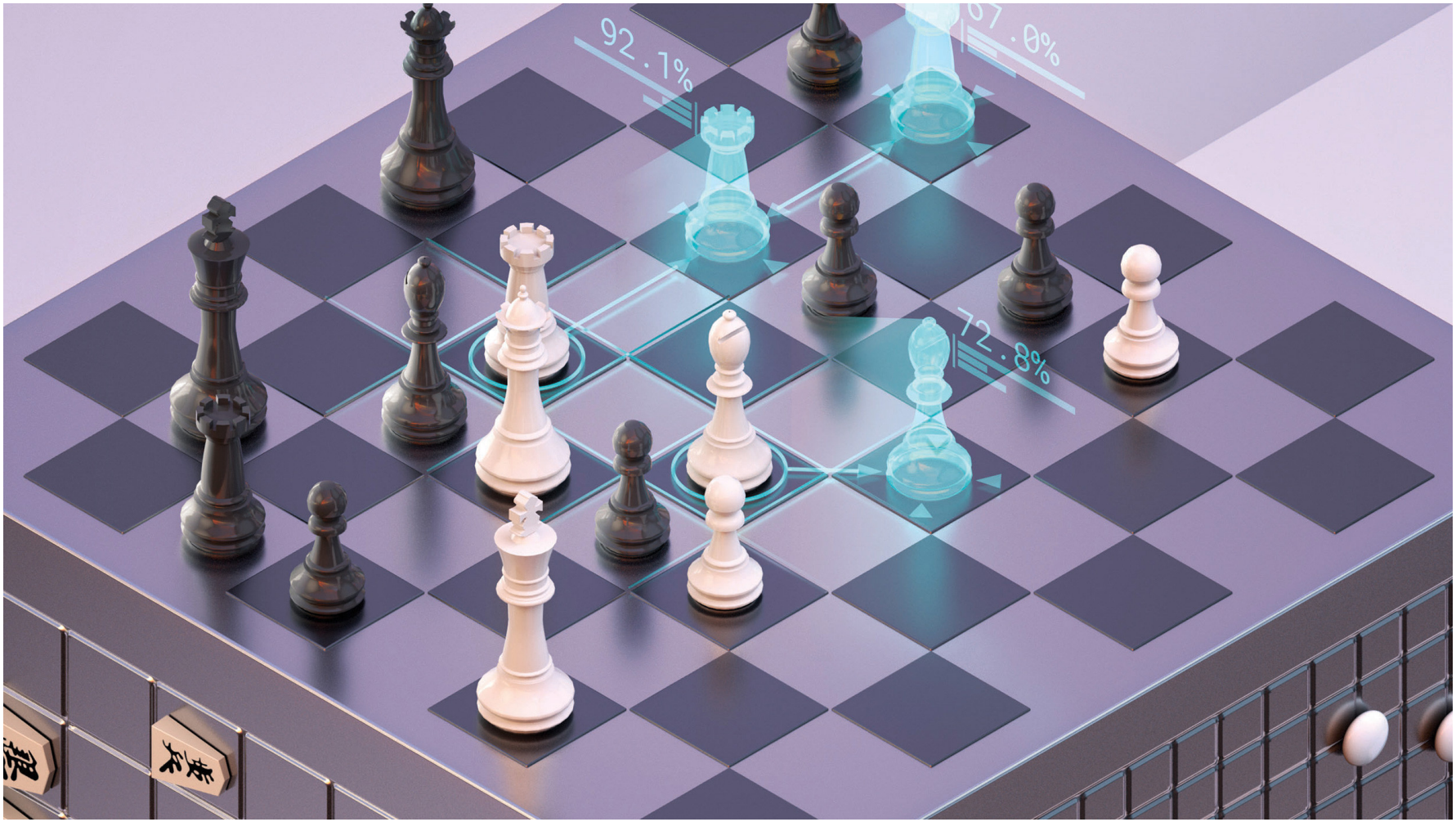




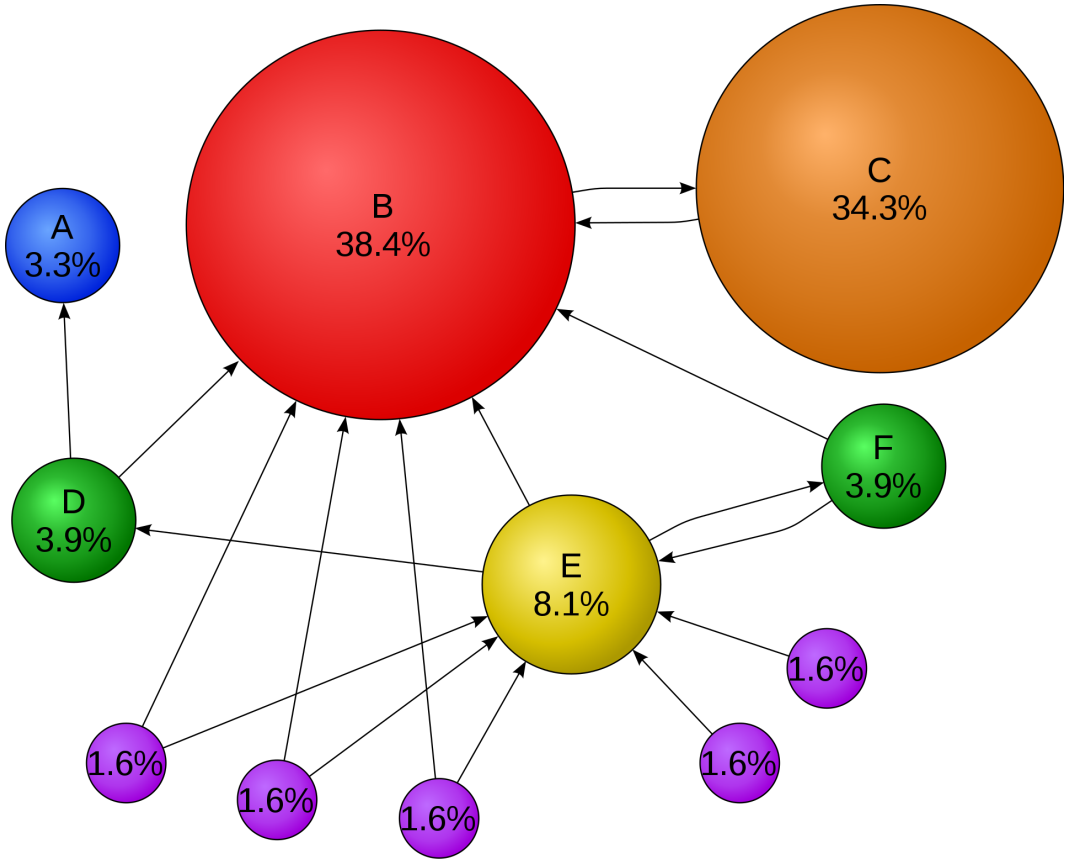
But Always Powerful: Information Theory



But Always Powerful: AlphaZero



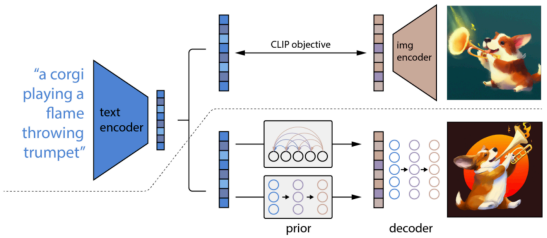
But Always Powerful: Google PageRank

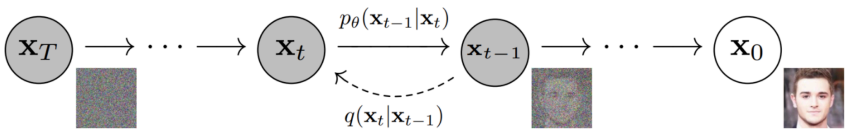


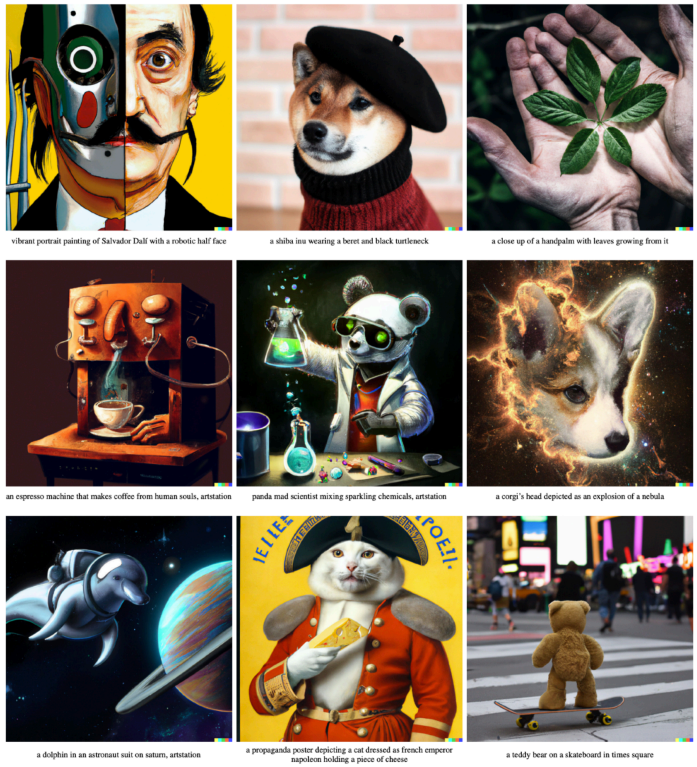
But Always Powerful: Diffusion Model

deep learning model to generate images from description

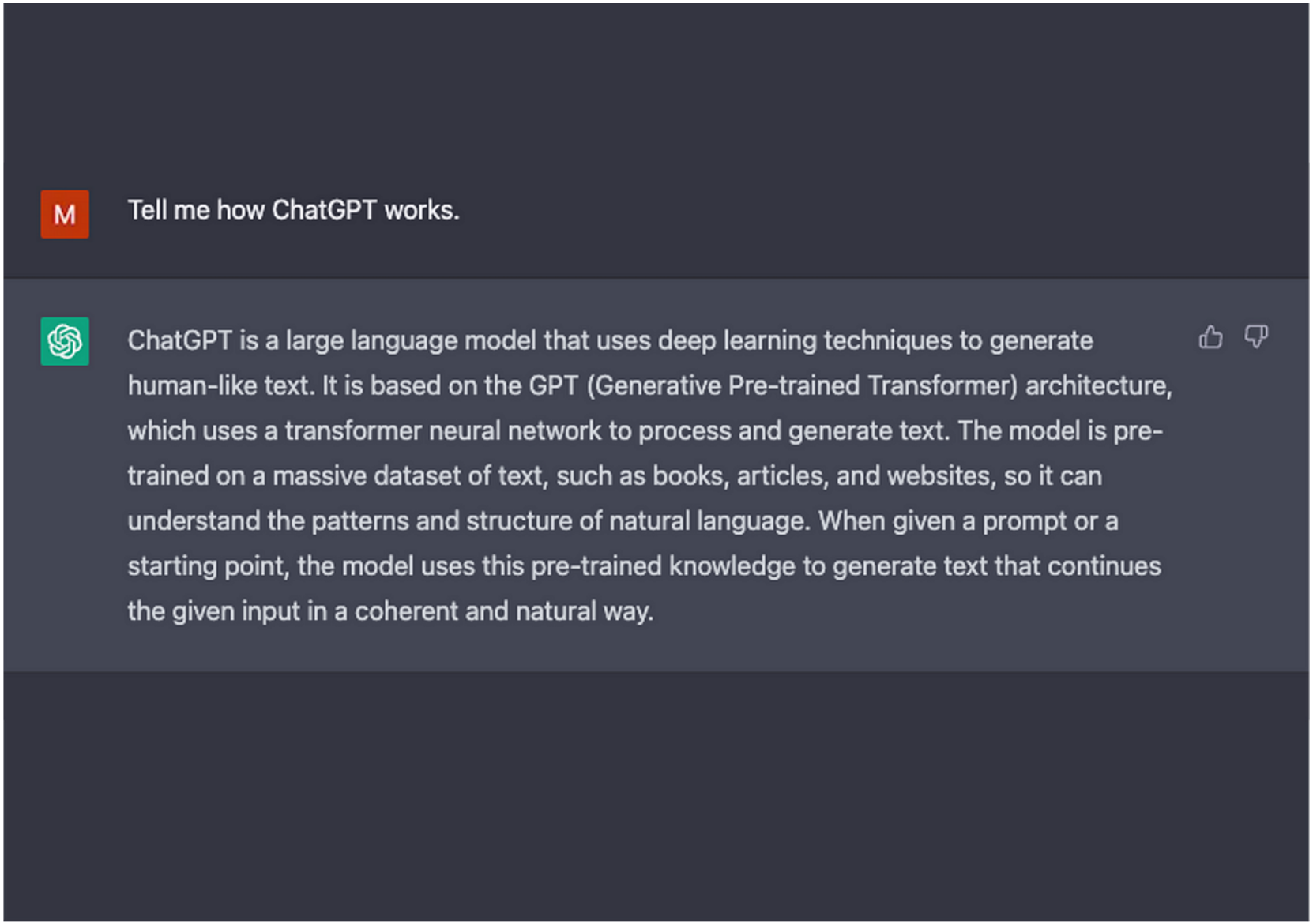
- DALL-E 2 is a deep learning model developed by OpenAI to generate digital images from natural language descriptions
- It uses ***diffusion process*** (a random process) to generate image.







But Always Powerful: ChatGPT

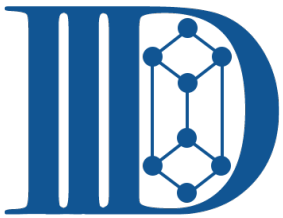


You Need Shepherd!

TA Teams

- 王上上
- 张天翼
- 周守琛

Professor Ziyu Shao: 邵子瑜



Laboratory for Intelligence Information and Decision

智 能 信 息 与 决 策 实 验 室

https://faculty.sist.shanghaitech.edu.cn/faculty/shaozy

Outline

Time & Venue

- **Time**: every Tuesday (10:15-11:55am) and Thursday (10:15-11:55am)
- **Location**: Room 201, Teaching Center
- **Course Forum in Piazza**:
  `https://piazza.com/shanghaitech.edu.cn/spring2024/si140a`

Course Grade

- *Homework* 20%: 5 or 6 problems per week
- *Behavior* 5%: miscellaneous behaviors
- *Final Project* 15%
- *Midterm Exam* 30%
- *Final Exam* 30%
- *Overall Winners*: lunch
  - Gauss Award
  - Markov Award
  - Bernoulli Award
  - Laplace Award
  - Fisher Award

Homework Policy

- Write the homework with LaTeX and output it as a PDF file.
- Write the programming assignment with Python and submit source codes.
- **Later** Homework receives no credit.
- You are allowed to discuss with others and use any references, but if you do so please list your collaborators and cite your references for each question.
- Not writing your own solutions or not listing your collaborators or not citing your references may be considered plagiarism.

Project Policy

- Project can be done by a team with ≤ 3 students.
- Teams with four and more students are NOT allowed.
- Your team is required to use Python for the programming part.
- Your team needs to submit the iPython Notebook including your source codes, simulation results, analysis, discussions, tables, figures, etc.
- **Later** Project receives no credit.

Project Topics

Your team select one and only one topic from the project list:

- Reverse engineering the mechanism of grab red envelope in Wechat and design the optimal strategy to grab red envelope. (微信抢红包机制分析以及抢红包最优策略设计).
- Evaluation and discussion of multi-armed bandit algorithms (多臂老虎机算法评估与讨论).
- Simulate and evaluate phase transition process(相变过程的仿真与评估).
- More topics (incoming)

LaTeX Setup

- LaTeX Package (Unix, Windows): TeXLive (*https://tug.org/texlive/*)
- LaTeX Package (MacOs): MacTeX (*http://tug.org/mactex*)
- Popular LaTeX Editors
  - Texpad (MacOs): *https://www.texpad.com*
  - TeXstudio (Windows,MacOs): *http://www.texstudio.org*
  - LyX(Windows,MacOs): *https://www.lyx.org*
  - WinEdt (Windows): *http://www.winedt.com/*
- Online LaTeX Editor: Overleaf (*http://overleaf.com/*)

LaTeX Tutorials

- Wiki LaTeX Book: *http://en.wikibooks.org/wiki/LaTeX*
- Collection of links for LaTeX: *https://www.tug.org/begin.html*
- Forum: *https://tex.stackexchange.com/*

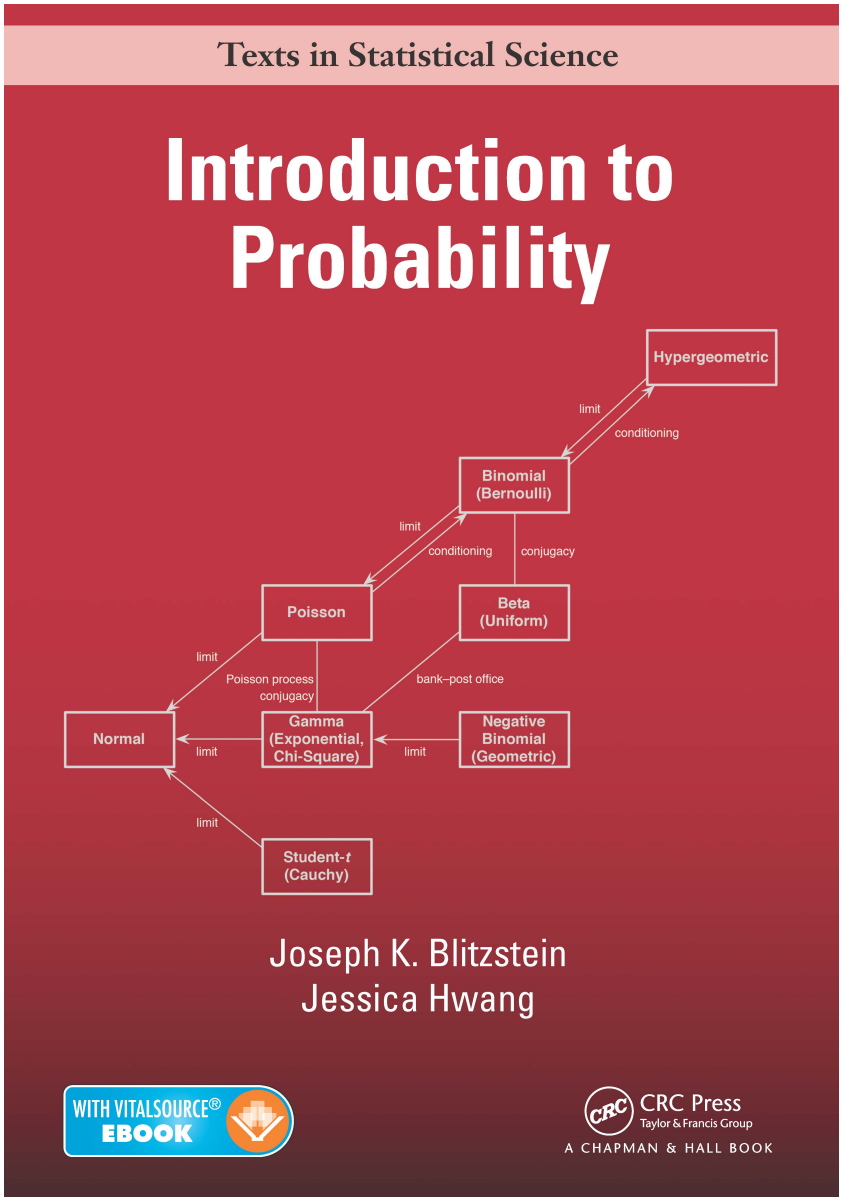
R: Optional

- Setup: *https://www.r-project.org/*
- Editor: *https://www.rstudio.com/*
- Data Visualization in R: *http://ggplot2.org/*
- Reference 1: *R in Action: Data analysis and graphics with R*, Robert I. Kabacoff, Manning Publications, 2015.
- Reference 2: *R for Data Science*, Hadley Wickham & Garrett Grolemund, O'Reilly Media, 2016.

Textbooks: Required

- *Introduction to Probability*, Joseph K. Blitzstein & Jessica Hwang, Chapman & Hall/CRC, 2014 or 2019. (**BH**)
- *Introduction to Probability (2nd Edition)*, Dimitris P. Bertsekas & John N. Tsitsiklis, Athena Scientific, 2008. (**BT**)

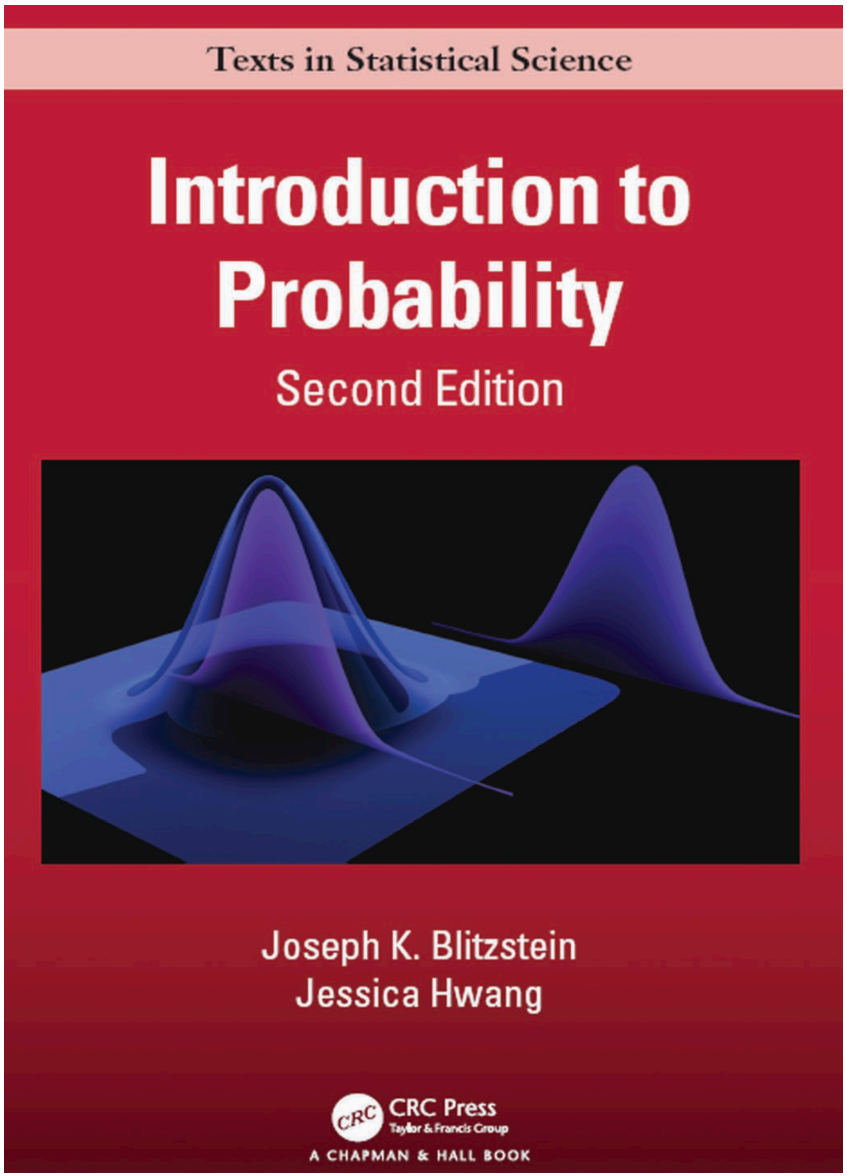
Required Textbook: BH



Joseph K. Blitzstein & Jessica Hwang

- Introduction to Probability
- Chapman & Hall/CRC, 2014.

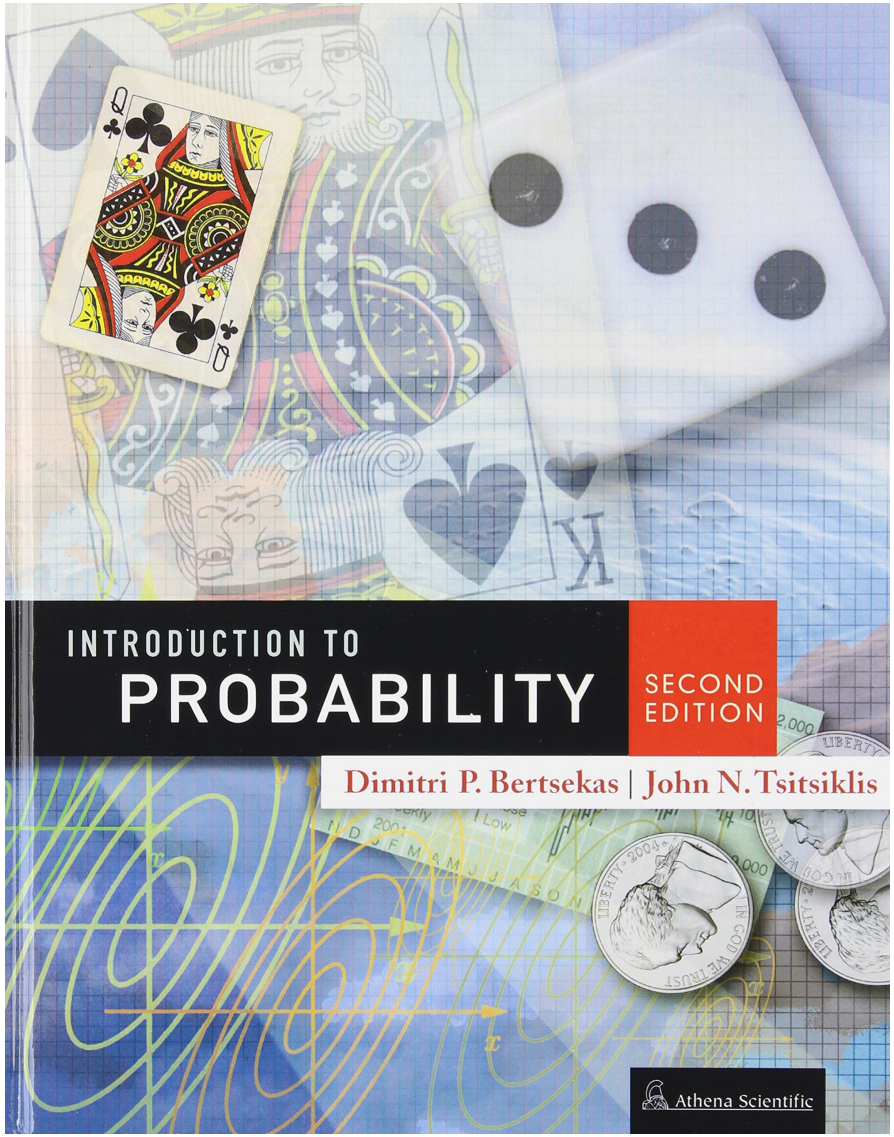
Required Textbook: BH



Joseph K. Blitzstein & Jessica Hwang

- Introduction to Probability (2nd Edition)
- Chapman & Hall/CRC, 2019.

Required Textbook: BT



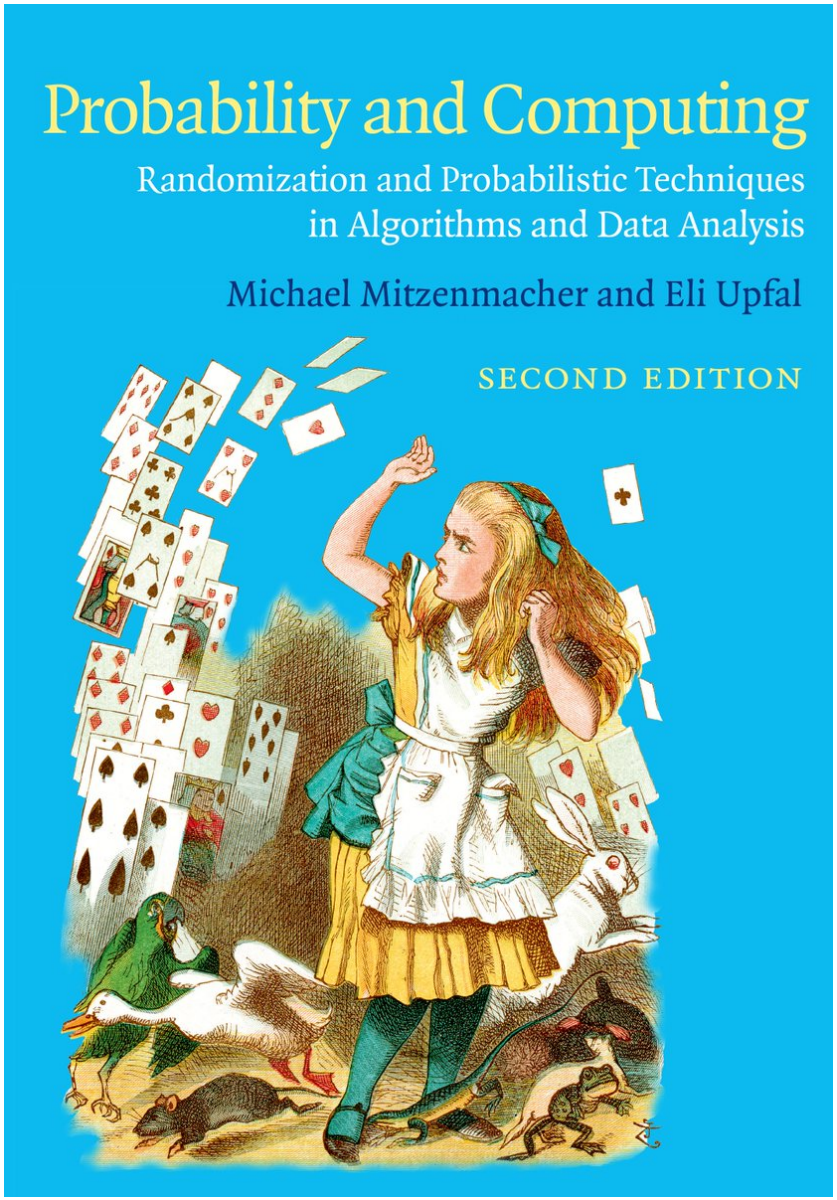
Dimitris P. Bertsekas & John N. Tsitsiklis

- Introduction to Probability (2nd Edition)
- Athena Scientific, 2008.

Recommended References: 学有余力的学生

- *Probability and Computing: Randomized Algorithms and Probabilistic Analysis (2nd edition)*, by Michael Mitzenmacher & Eli Upfal, Cambridge University Press, July 2017. (**MU**)
- *Probability and Random Processes (4th Edition)*, by Geoffrey R. Grimmett & David R. Stirzaker, Oxford University Press, September 2020. (**GS**)
- *All of Statistics: A Concise Course in Statistical Inference*, by Larry Wasserman, Springer Texts in Statistics, December 2010. (**LW**)
- *Statistical Inference*, by George Casella & Roger L. Berger, Duxbury Press, June 2001. (**CB**)

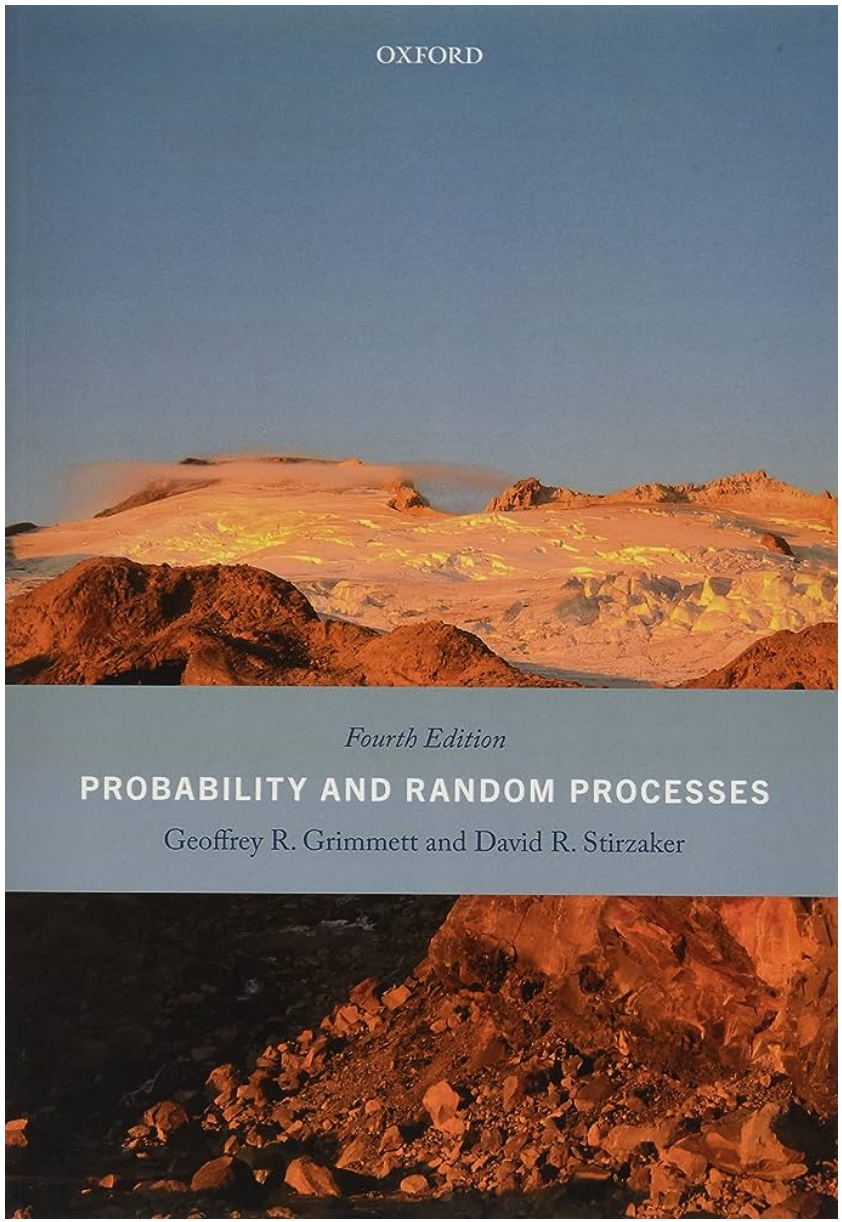
Reference: MU



Michael Mitzenmacher & Eli Upfal

- Probability and Computing (2^{nd} edition)
- Cambridge University Press, July 2017.

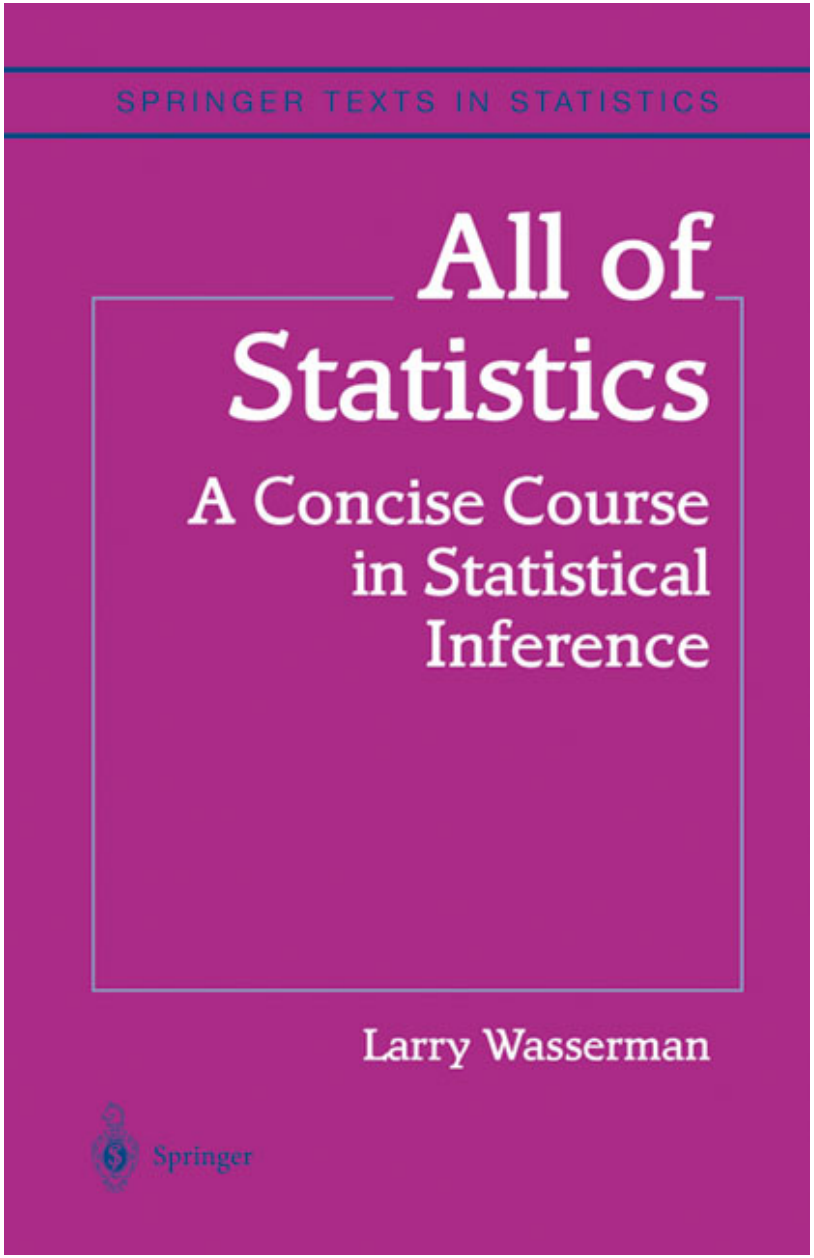
Reference: GS



Geoffrey R. Grimmett & David R. Stirzaker

- Probability and Random Processes (4^{th} edition)
- Oxford University Press, August 2020.

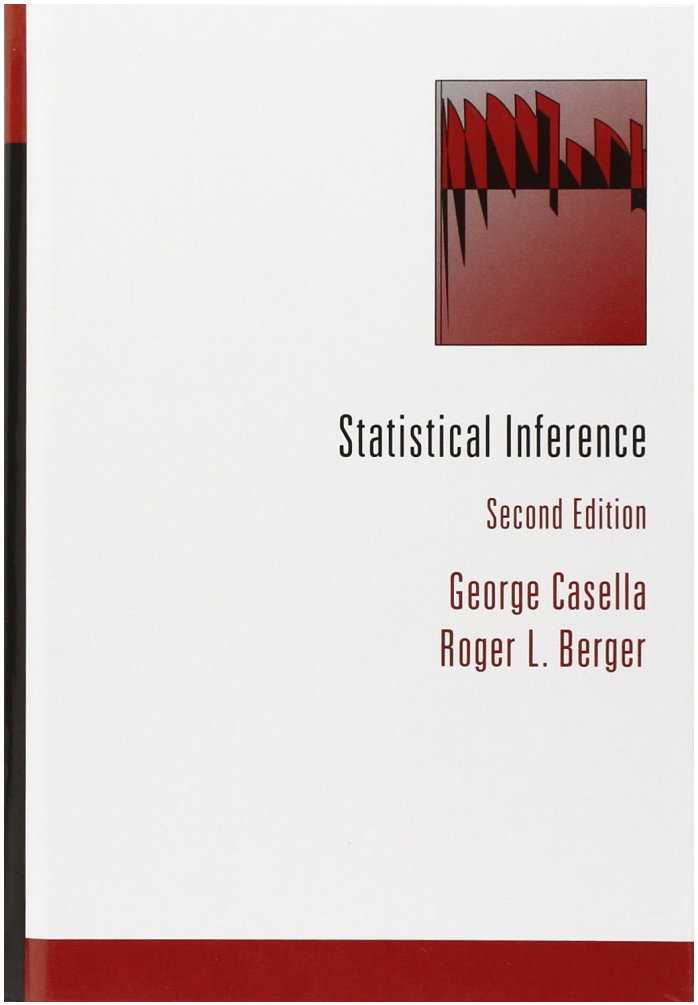
Reference: LW



Larry Wasserman

- All of Statistics: A Concise Course in Statistical Inference
- Springer, Dec. 2010.

Reference: CB



George Casella & Roger L. Berger

- Statistical Inference
- Duxbury Press, June 2001.

Recommended References: 基础薄弱的学生

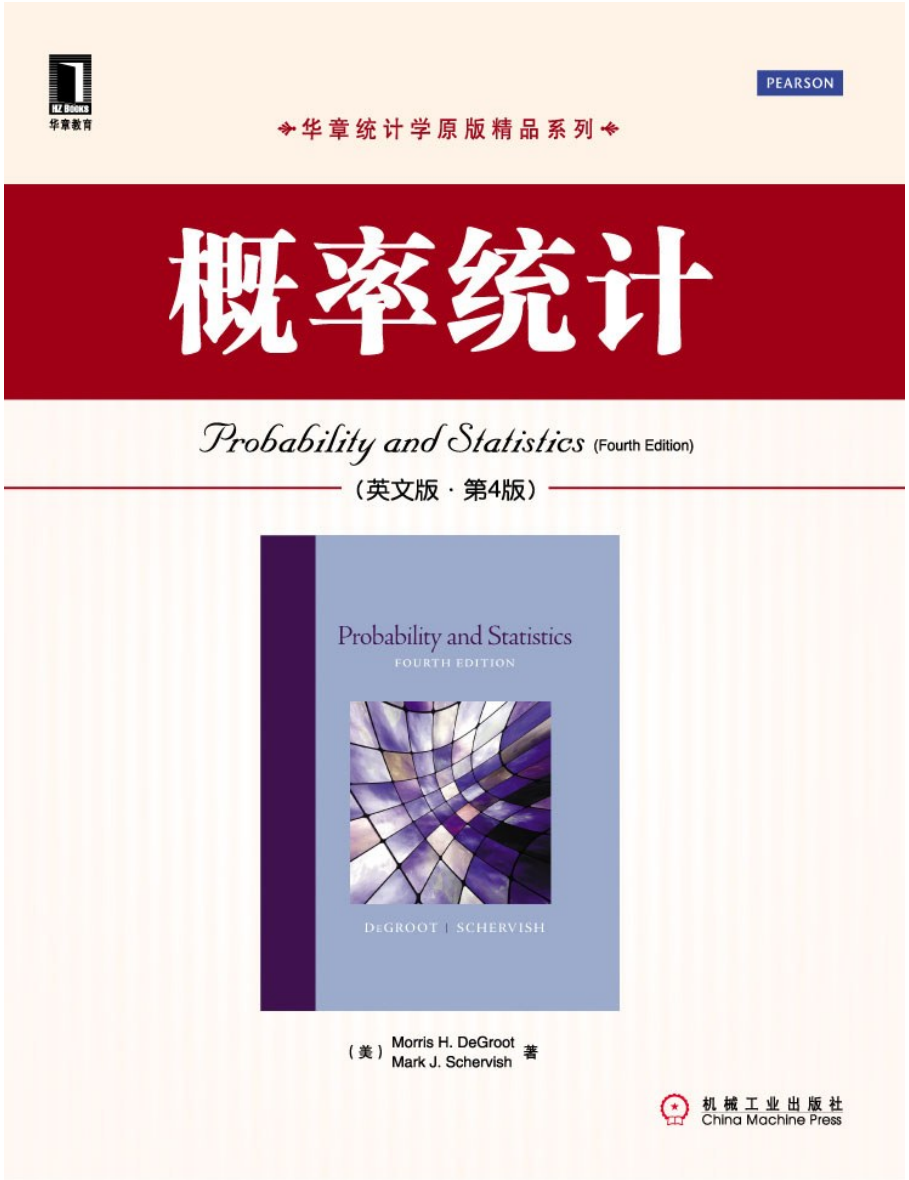
- 概率论与数理统计教程 *(第三版)*, 茆诗松、程依明、濮晓龙, 高等教育出版社, 2019 年 11 月. (**MCP**) (此外还有配第二版概率论与数理统计教程的习题解答)
- *Probability and Statistics (4th Edition)*, by Morris DeGroot & Mark Schervish, Pearson Press(机械工业出版社影印版) July 2012. (**DS**)

Reference: MCP

茆诗松、程依明、濮晓龙

- 概率论与数理统计教程 (第三版)
- 高等教育出版社, 2019 年 11 月.

Reference: DS

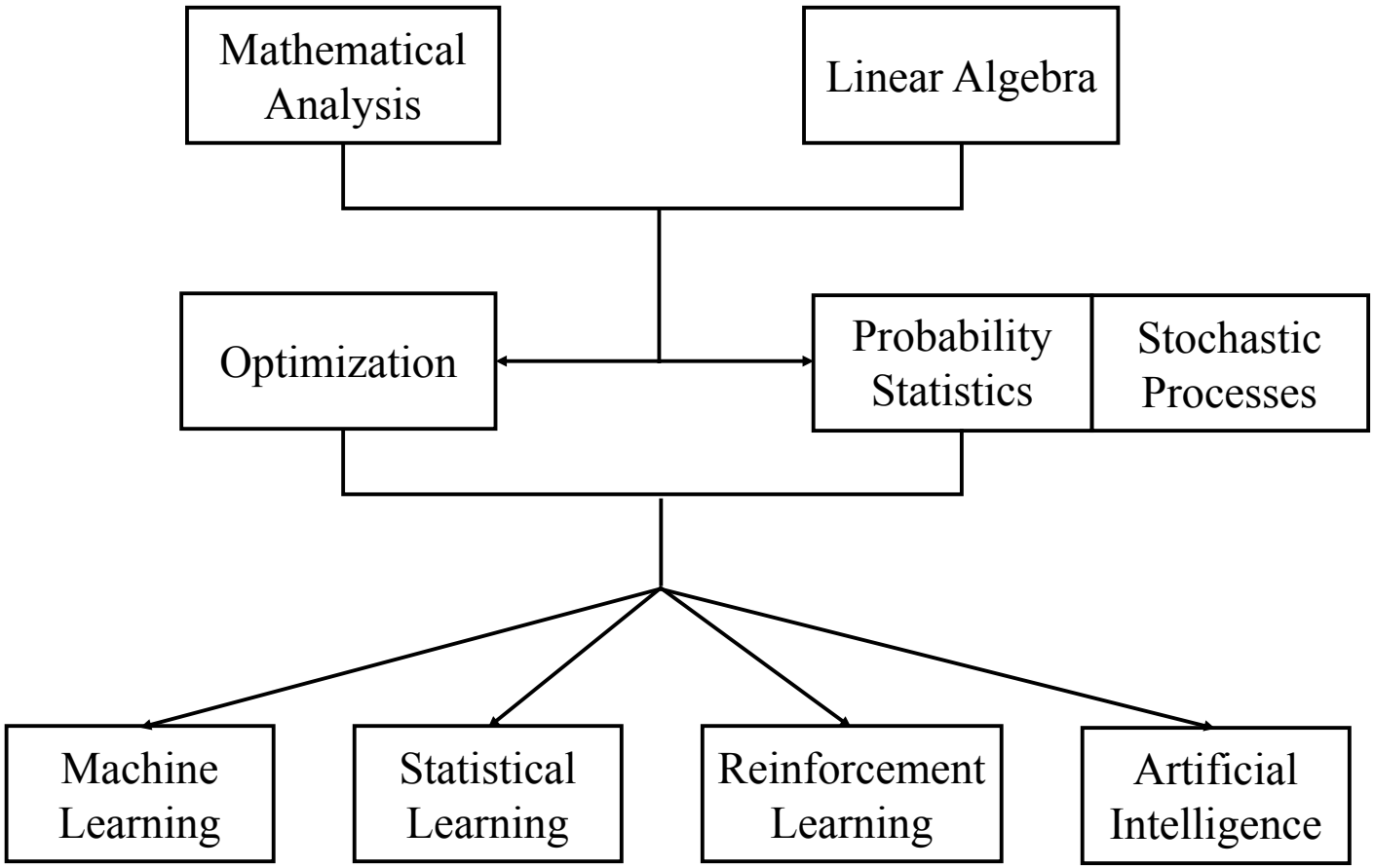


Morris DeGroot & Mark Schervish

- Probability and Statistics (4^{th} edition)
- Pearson Press, July 2012.

Outline

The Role of This Course



Course Objectives

- Integration of Probability, Statistics, Data Science, Random Algorithm
- **Conversation between mathematics and engineering**: theory, model, algorithm, practice
  - **Solid foundation of probability & statistics**: modeling, analysis, and theoretical proof
  - **Computational thinking**: algorithmic perspective and programming practice
  - **Application mentality**: many examples from EECS

Teaching Philosophy of This Course

Probability
Math

Statistics
Science

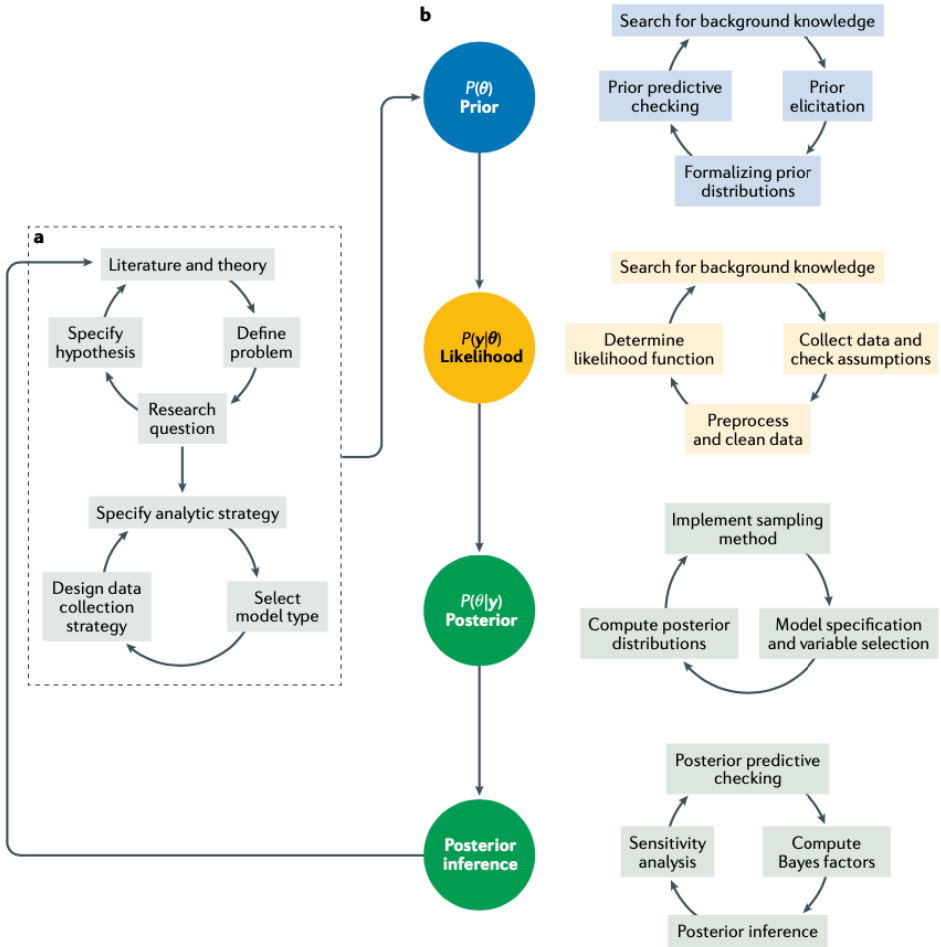
Monte Carlo
Computing

Course Focus

- **Motivation**: In the ocean of mathematical definitions, theorems, and equations, why should we spend our time on this particular topic but not another?
- **Intuition**: When going through the derivations, is there a physical intuition beyond those equations?
- **Implication**: After we have learned a topic, what new problems can we solve?

Theme

- One Core: Bayesian Inference
- Two Basic Tools: Conditional Probability & Conditional Expectation



Basic Contents

- Probability and Counting
- Conditional Probability
- Random Variables and Distributions
- Expectation
- Continuous Random Variables
- Joint Distributions
- Transformations
- Monte Carlo & Limit Theorems & Inequalities
- Statistical Inference & Conditional Expectation
- Markov Chain

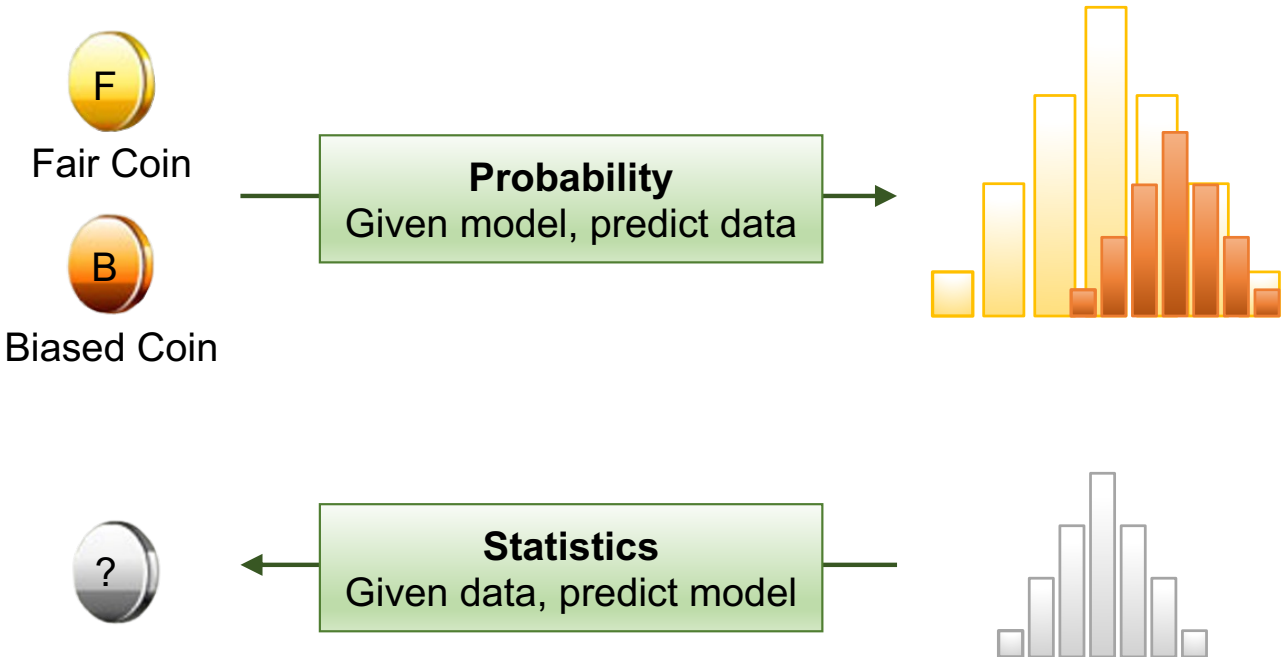
Outline

Probability vs. Statistics

Statistics: Given the information in your hand, what is in the pail?

Probability: Given the information in the pail, what is in your hand?

Probability vs. Statistics



Probability vs. Statistics

- **Focus of Probability**: given a data generating process (model), what are the properties of the outcomes?
- **Focus of Statistics**: the inverse of probability, i.e., given the outcomes, what can we say about the process that generated the data (model)?

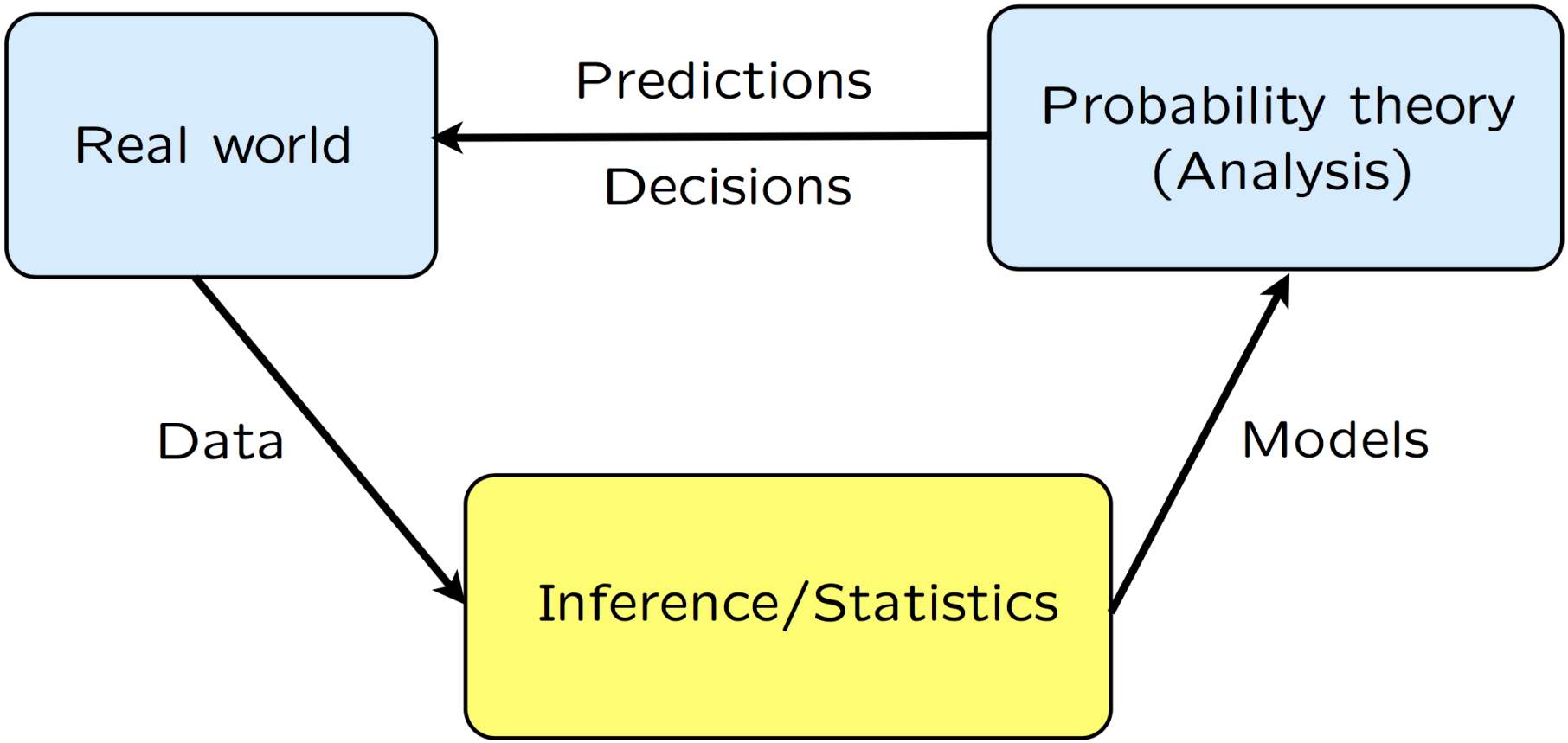
Probability & Statistics

Probability & Statistics

词以境界为最上。有境界自成高格，自有名句。有造境，有写境，此理想与现实二派之所分。

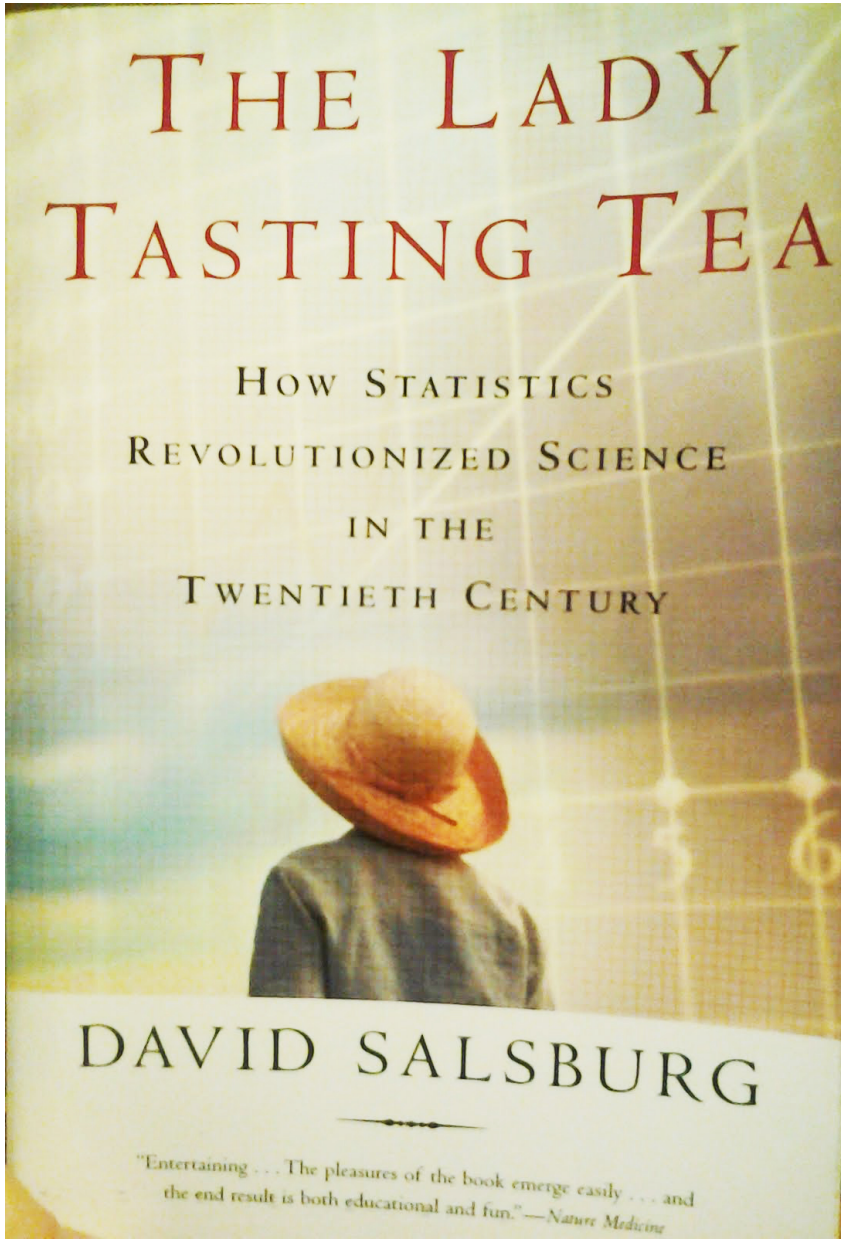
人间词话-王国维

A Framework to Quantify Uncertainty



Outline

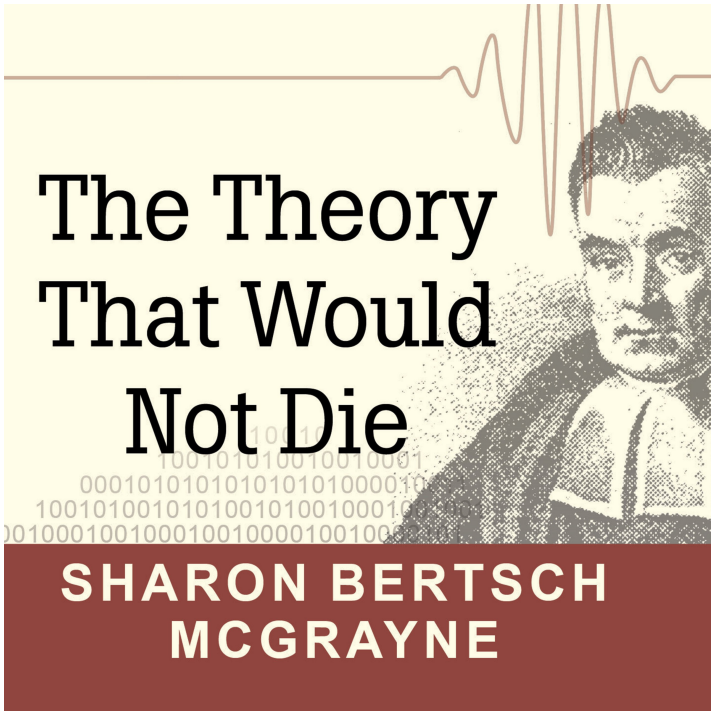
The Lady Tasting Tea



David Salsburg

- The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century
- Holt Paperbacks, May 2002.

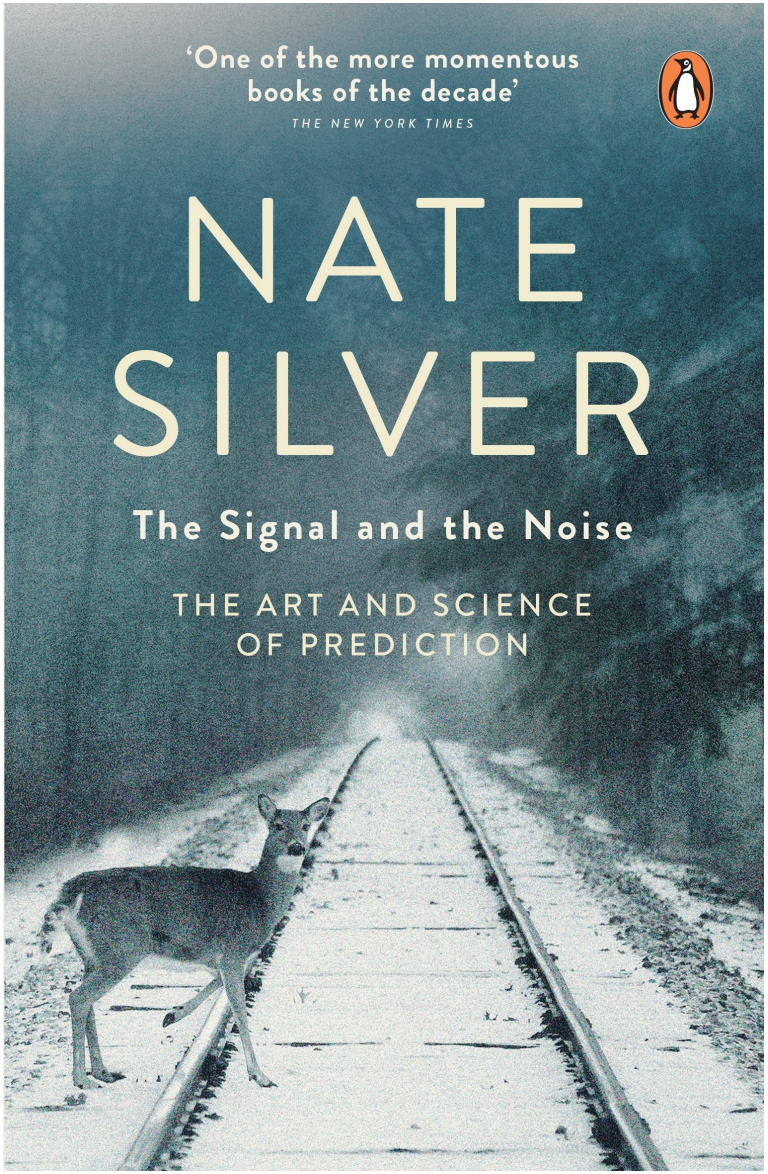
The Theory That Would Not Die



Sharon Bertsch McGrayne

- The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy
- Yale University Press, September 2012

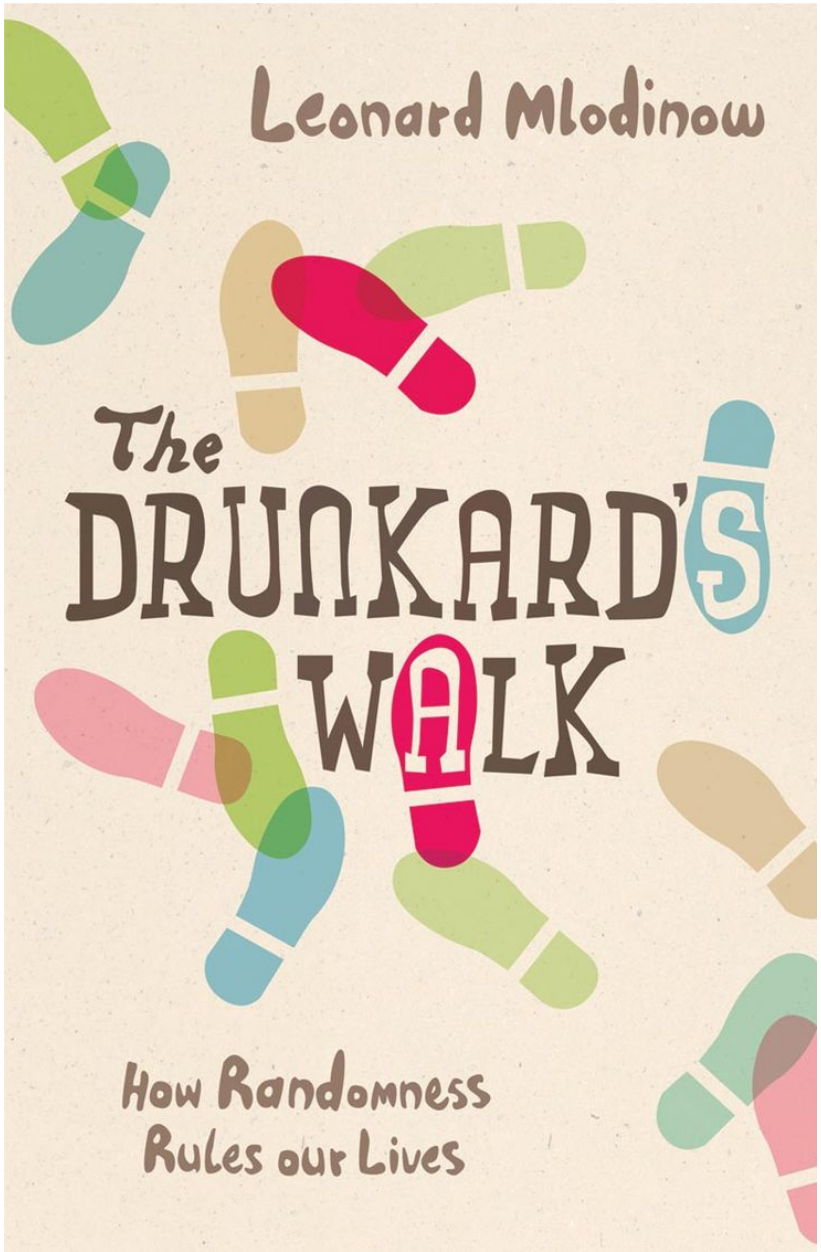
The Signal and the Noise



Nate Silver

- Penguin Books, February 2015.

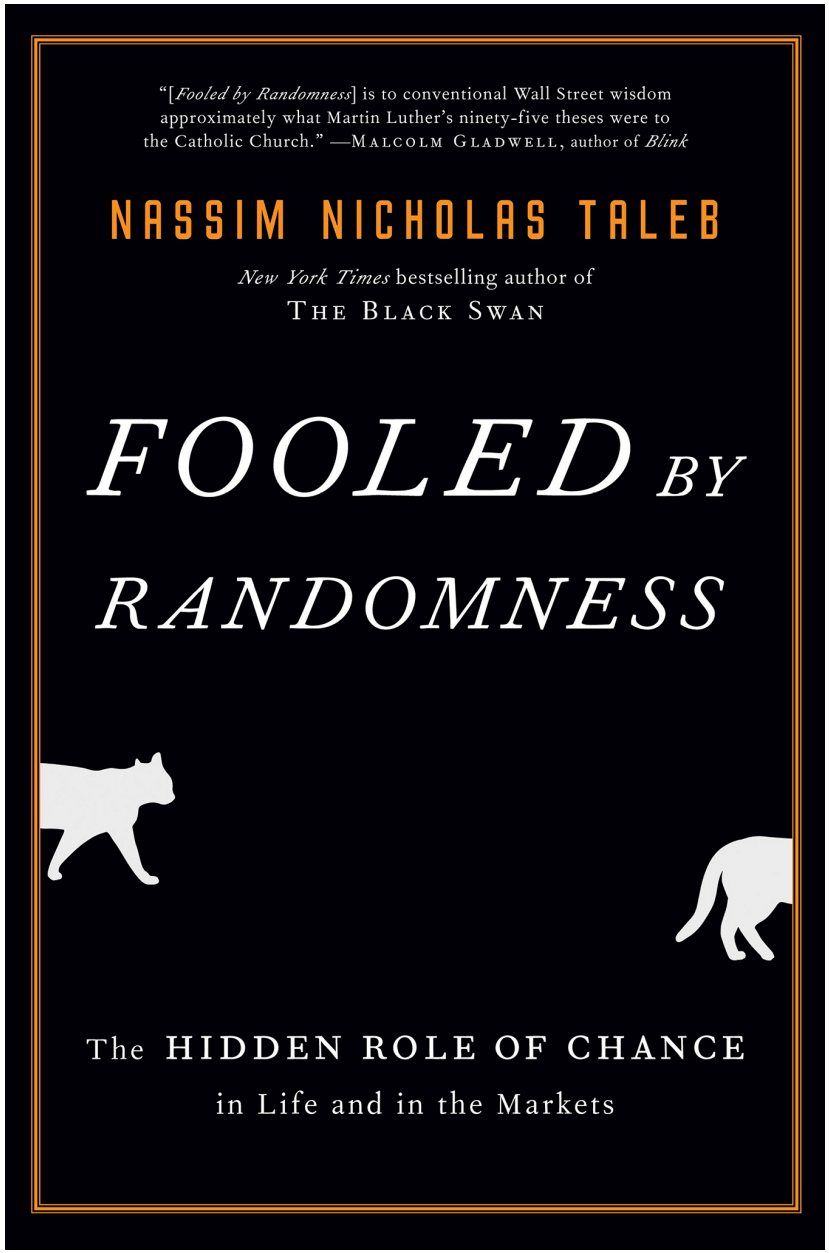
The Drunkard's Walk



Leonard Mlodinow

- The Drunkard's Walk: How Randomness Rules Our Lives.
- Vintage, May 2009.

Fooled by Randomness



Nassim Nicholas Taleb

- Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets

Let the Adventure Begin!