ORIE 4580/5580: Simulation Modeling and Analysis

ORIE 5581: Monte Carlo Simulation

Fall 2020

Sid Banerjee

School of ORIE, Cornell University

Clicker Question

Suppose we toss a fair coin 10 times. Which of the following is the 'most likely' observation:

(a) $H\ H\ H\ H\ H\ H\ H\ H\ H\ H$

(b) $H\ H\ H\ H\ T\ T\ T\ T\ T\ T$

(c) $T\ H\ T\ H\ T\ H\ T\ H\ T\ H$ 17%

(d) $T\ H\ T\ T\ T\ H\ T\ H\ H\ T$ 23%

(e) All seem the same to me. . . rest

Clicker Question: Solution

Suppose we toss a fair coin 10 times. Which of the following is the 'most likely' outcome:

$$\mathbb{P}[HTT\ldots H] = \mathbb{P}[H]\,\mathbb{P}[T]\ldots\mathbb{P}[H]$$

$$= 0.5 \times 0.5 \ldots 0.5$$

$$= (0.5)^{10}$$

Clicker Question

Suppose we toss a biased coin (with $\mathbb{P}[Heads] = 0.6$) 10 times. Which of the following is the 'most likely' outcome:

(a) $H\,H\,H\,H\,H\,H\,H\,H\,H\,H$ 81% - $\mathbb{P}[H \ldots H] = (0.6)^{10}$

(b) $H\,H\,H\,H\,T\,T\,T\,T\,T\,T$ $\mathbb{P}[seq] = (0.4)^6 \cdot (0.6)^{10}$

(c) $T\,H\,T\,T\,H\,T\,T\,H\,T\,H$

(d) $T\,H\,T\,T\,T\,H\,T\,H\,H\,T$

(e) All seem the same to me... 9%

Clicker Question: Solution

Suppose we toss a biased coin (with $\mathbb{P}[Heads] = 0.6$) 10 times.

Which of the following is the 'most likely' outcome:

Simulation = Probability + Computing

Why Study Simulation?

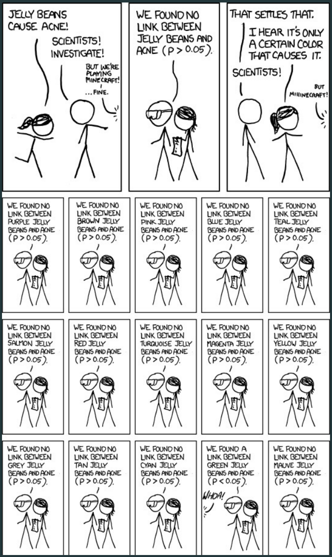
- Develop probabilistic thinking
- Understand how to summarize and present data
- Learn about stochastic models for complex systems DES
- Tool for fast computation and algorithms Monte Carlo

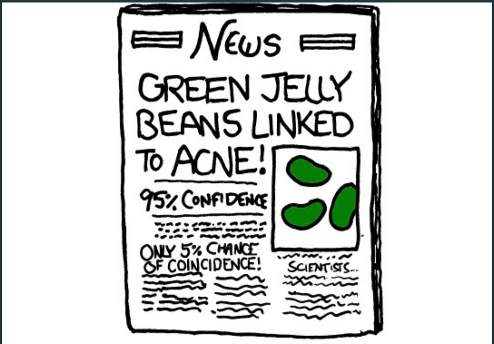
Applications

Probability in computing has five major applications

- **Numerical Computation**: Monte Carlo algorithms for scientific computing + Data mining/ML
- **Algorithms for massive data**: Sampling, sketching, streaming data, random-walk network algorithms, graphical models, etc.
- **Cryptography and Privacy**: Sampling, sketching, streaming data, random-walk network algorithms, graphical models, etc.
- Risk Analysis: Quantifying and hedging against random 'shocks' in daily life
- Counterfactual ('What-if') Analysis: Understanding and optimizing complex systems in-vitro

Understanding randomness and sampling





What are the fundamental misunderstandings here?

Risk Analysis: Choosing Projects

What are the fundamental misunderstandings here?

- The project has a 70% chance of success. Why not try?
- Why not increase the chance of a success by doing 10?

Risk Analysis: Playing Blackjack

The book Bringing Down the House (made into the movie 21) showed how a team from MIT used card counting (a technique which lets you know when the odds are slightly in your favor) to win big in blackjack in Las Vegas.

- How do slightly favorable odds get turned into big winnings?
- What can go wrong (besides getting caught)?
- Why wouldnt you bet your entire retirement savings or college tuition on one hand?

Counterfactual Analysis

COVID-19 Mathematical Modeling for Cornell's Fall Semester

PhD Students: J. Massey Cashore, Ning Duan, Alyf Janmohamed, Jiayue Wan, Yujia Zhang
Faculty: Shane Henderson, David Shmoys, Peter Frazier*

June 15, 2020

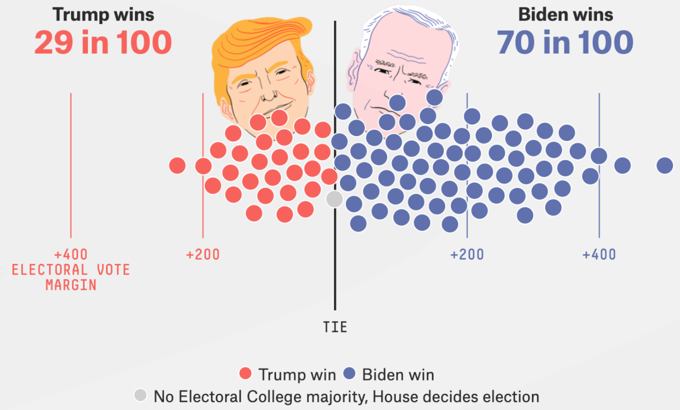
Executive Summary:

- Initial modeling results suggest that a combination of contact tracing, asymptomatic surveillance, and low initial prevalence (supported through testing students prior to, and upon, returning to campus) can achieve meaningful control over outbreaks on Cornell's Ithaca campus in the fall semester if asymptomatic surveillance is sufficiently frequent and if we have sufficient quarantine capacity. This would dovetail with a complementary effort at Cornell to reduce transmissions through housing policy, class organization, and regulations on social gatherings.

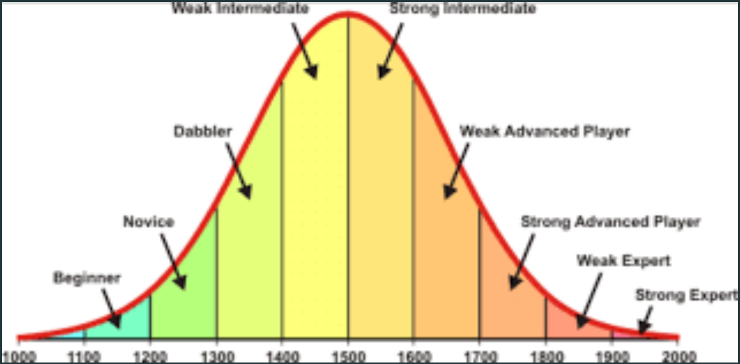
- We use our model to predict outcomes for a full return of students, faculty and staff in the fall semester over a 16 week time period, with cases imported from returning students and from Tompkins county, counterbalanced by aggressive asymptomatic surveillance where every member of the campus community is tested every 5 days. The course of the epidemic is random and we directly model that randomness. Accordingly, our model produces a range of potential futures. In the median random potential future, under our nominal set of parameters, 3.6% of the campus population (1254 people) become infected, and 0.047% of the campus population (16 people) require hospitalization. The 90% quantile rises to 4.02% infected and 0.051% requiring hospitalization. Of the 1254 infections in the median outcome, 570 are due to direct outside infections and ensuing additional infections prior to isolation, while 31 (0.09% of the campus population) are infected before arrival to campus but missed in the test-on-return protocol. There are an additional set of people infected before arrival, found through test-on-

Prediction

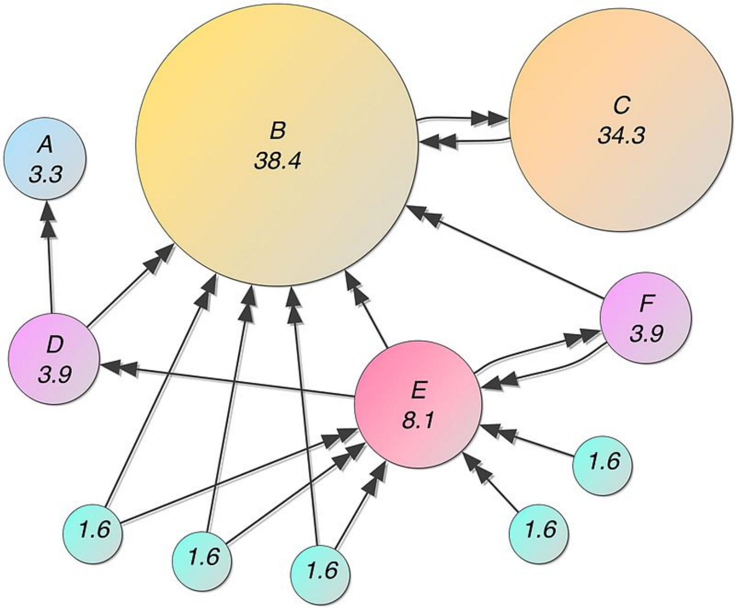
We simulate the election 40,000 times to see who wins most often. The sample of 100 outcomes below gives you a good idea of the range of scenarios our model thinks is possible.



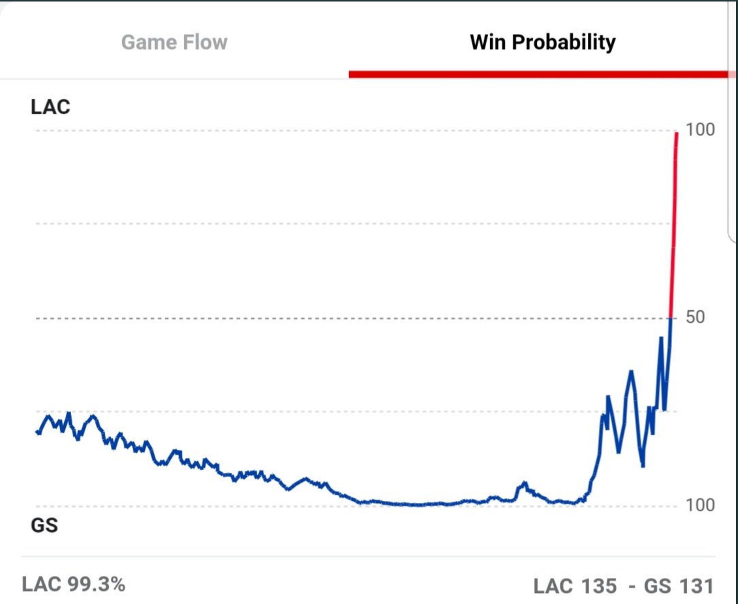
Quantifying Quality



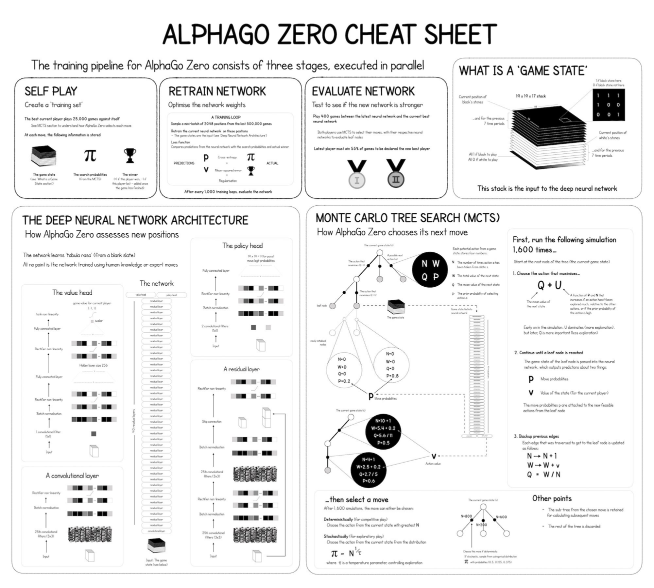
Quantifying Utility



Surprise!


Game Flow
Win Probability
LAC
100
50
100
GS
LAC 99.3%
LAC 135 - GS 131

. . .and beyond: Optimization, control and RL



Logistics

Essential Course Information

- *Instructor*

  Sid Banerjee <sbanerjee@cornell.edu>

- *Teaching Assistants*

  Ruifan Yang <ry298@cornell.edu>
  Shuwen Lu <sl3243@cornell.edu>
  Yuxuan Liu <yl2999@cornell.edu

  Brittany Stenekes <bss99@cornell.edu>
  Gary Li <hl2362@cornell.edu>

  Office hours: See Piazza

Essential Course Information (contd.)

- **Lectures**

  Course Number: ORIE 4580

  Class time: TR 11:30am-12:25pm

  Class location: Statler Auditorium 185

- **Recitation Sessions**

  1. Wednesday 4:15pm-6:10pm, 453 Rhodes
  2. Wednesday 7:30pm-9:25pm, online only
  3. Friday 9:10am-11:05am, 571 Rhodes
  4. Friday 4:15pm-6:10pm, 453 Rhodes
  5. Monday 11:30-1:25, 571 Rhodes Hall

- **Course Communication**:

  Piazza: `https://piazza.com/class?nid=kec1z24huh637p`

  Website: `https://people.orie.cornell.edu/sbanerjee/courses/orie4580f20/`

Course Resources

- **Course notes**: Slides with blanks uploaded on Piazza before class – with annotations after class
- **iClicker** or iClicker App: Required - see webpage for details
- **Software**
  - All coding in Python + Jupyter notebook
  - Try Google Colab for online collaborative coding
  - Intro to Simio in second half of course

Homeworks

Weekly homework assignments (5 for ORIE 5581, 10 for 4580/5580)

- solutions must be submitted online on CMSx
  Students must *typeset* all solutions
- Homeworks due on Wednesday at 12pm (noon)
- **Collaboration**: Can do homeworks in pairs. Pairs must submit a single solution with both names and netids on the solution.
- **Late submissions and drops**:
  4 late days for 4580/5580, and 2 late days for 5581 across all homeworks (at most 2 late days per hw).
  ORIE 4580/5580 students can drop 2 lowest homework grade.
  (ORIE 5581 students can drop *one lowest grade*).

Grading

Prelim: October 22, 7.30pm

(as of now) **no (semi)-final exam**

- online exam format TBA
- (tentative) grading scheme

Component	4580/5580	5581
Clicker responses	4	4
Homeworks	45	55
Project	20	-
Prelim (Oct 22)	30	40
Course Eval	1	-

- Clicker response: Points are for participation, not correctness

Course Outline

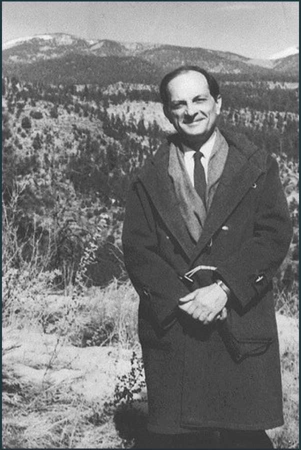
Topics cover simulation analysis, modeling and optimization

Review of probability	1-2
Basic Monte Carlo Simulation	1-2
Generating random variables	2
Input modeling	3
Variance reduction and importance sampling	2-3
Intro to Markov Chains	1
Markov chain Monte Carlo	1
Intro to Discrete-event system simulation	1
Markovian models and queues	2
Modeling in Simio	1
Output analysis	2
Comparing systems; ranking and selection	1-2
Simulation and dynamic programing	2

Some history and demos

Birth of Monte Carlo Simulation: Ulam in Hospital

"...in 1946 as I was convalescing from an illness and playing solitaires ... I thought what are the chances that a Canfield solitaire laid out with 52 cards will come out successfully? After spending a lot of time trying to estimate them by pure combinatorial calculations, I wondered whether a more practical method than "abstract thinking" might not be to lay it out say one hundred times and simply observe and count the number of successful plays. This was already possible to envisage with the beginning of the new era of fast computers, and I immediately thought of problems of neutron diffusion and other questions of mathematical physics..."

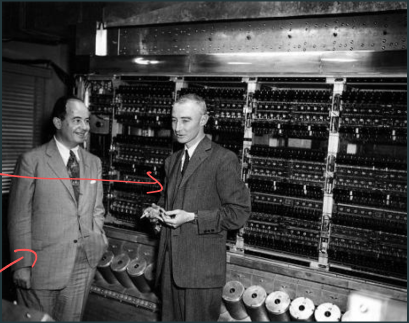


Stanislaw Ulam

Monte Carlo Simulation at Los Alamos

Progress in Monte Carlo simulation was driven in the early years by the development of the MANIAC computer at Los Alamos National Labs by John von Neumann and Nicholas Metropolis.

Ro



JVN

John von Neumann

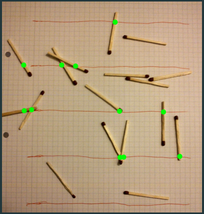
Nicholas Metropolis

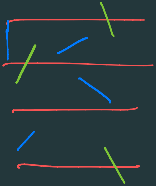
An Even Older Simulation!

Although Ulam first formalized Monte Carlo simulation, similar ideas had been used even before. The most famous early example of simulation is the 18th century Buffon's Needle Problem

- Throw matches of length 1 on horizontal grid with lines at distance 1
- Compute $\hat{X} = 2 \times D/C$, where D =number dropped, and C =number which touch line

N, C_N





Comte de Buffon

See Numberphile Video for more details

$$\frac{2N}{C_N} = \frac{14}{3}$$