S&DS 241 Lecture 1

Introduction

Instructor

Prof. Yihong Wu, yihong.wu@yale.edu

- Lectures MW 9-1015am. Davies Auditorium
- Office hours M 3-4pm. Room 235 Dunham Lab

Staff

- Teaching Fellows
  - Alexandra Djorno alexandra.djorno@yale.edu
  - Sky Lee sikai.lee@yale.edu
  - Jack Ma jack.ma.rm2545@yale.edu
  - Ashley Oaks ashley.oaks@yale.edu
- Undergraduate Learning Assistants
  - Carly Benson
  - Zach Brown
  - Philos Kim
  - Matthew Li
  - Jackson Pullman
  - Joshua Rothbaum
  - Ben Wonderlin
  - Rock Zhu
- Course manager
  - Leona Wang leona.wang@yale.edu
  - **Contact Leona for all logistic questions**

Syllabus

Please go over the syllabus on Canvas!

Class format

- Class meets in person
- Lectures will be recorded and posted on canvas. Office hours will not be recorded.
- TF/ULA office hours
  - Location/Time: TBA
  - Before shopping period ends: zoom

Grading

- Weekly Homework: 30%.
- Midterm: 30%.
- Final Exam: 40%.

Homeworks

- Typically posted every Wed, **due next Wed** 6pm ET
- First homework: to be posted Sep 9 next Wed
- Gradescope for online submission
- Discussions are allowed, but you are expected to write your own solutions independently, clearly justifying each step
- Regrade requests must be accompanied by a formal, written explanation – see Canvas syllabus.
- No late homework allowed except for medical reasons, but the lowest score of your problem sets will be dropped.

Exams

- Midterm:
  - Oct 12 Wed in class
  - One letter-size (8.5x11") sheet of notes allowed
- Final:
  - Dec 20 Monday 2-5pm
  - Two letter-size (8.5x11") sheets of notes allowed
- Electronic devices (calculators, phone, laptop, desktop, . . .) are NOT allowed.

Email Guidelines

- For logistic questions or private matters related to the course, contact course manager Leona.
- Ask questions during office hours.
- Ask questions on Piazza.

Overlapping class

- SDS 240/540: similar in scope, but 241 offers a more in-depth treatment. 240 doesn't count toward the SDS major, but does count toward the undergraduate data science certificate.
- SDS 238/538: covers probability theory and Bayesian statistics, with a substantial component on computing and data analysis (whereas 241/541 focus on probability theory, with no component on coding or data analysis). There's a helpful discussion of 238 vs 241 near the end of the S&DS Major FAQs
- Taking both 238 and 241 for credit is allowed.
- Taking both 240 and 241 for credit is NOT allowed.
- Any of the three courses (238/538, 240/540, or 241/541) will cover the probability background needed for SDS 242/542 (Theory of Statistics).

Course materials

- Lecture notes: Canvas
- **Textbook:** *Introduction to Probability* (Chapman & Hall/CRC Texts in Statistical Science) by Joseph K. Blitzstein and Jessica Hwang. (Free online.)
  - Will use 2nd ed by default. But 1st works too.
- **Additional reference**: *Introduction to Probability* by Charles M. Grinstead and J. Laurie Snell. (Free online.)

Course materials

- Lecture notes: Canvas
- **Textbook:** *Introduction to Probability* (Chapman & Hall/CRC Texts in Statistical Science) by Joseph K. Blitzstein and Jessica Hwang. (Free online.)
  - Will use 2nd ed by default. But 1st works too.
- **Additional reference**: *Introduction to Probability* by Charles M. Grinstead and J. Laurie Snell. (Free online.)
- Questions?

What's this course about?

Probability theory

- Historically the study of probability started in analyzing various gambling schemes in 16-17th century Europe
  - coin, dice, poker
  - lottery
  - stock market

Cardano

Fermat

Pascal

- Formalized as a mathematical subject to model chance or randomness
- Foundations of many scientific disciplines

Dice

- The history of probability begins with an exchange of letters from Pascal to Fermat, on 29 July 1654, in which he solved two problems posed to him by gambler Chevalier de Méré:

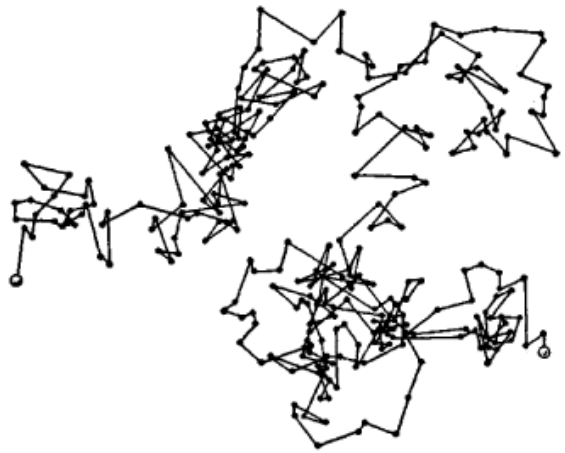
Which is more probable?

1. getting at least a six in 4 tosses of a die, or
2. getting at least a double-six in 24 tosses of a pair of dice

https://www.york.ac.uk/depts/maths/histstat/pascal.pdf

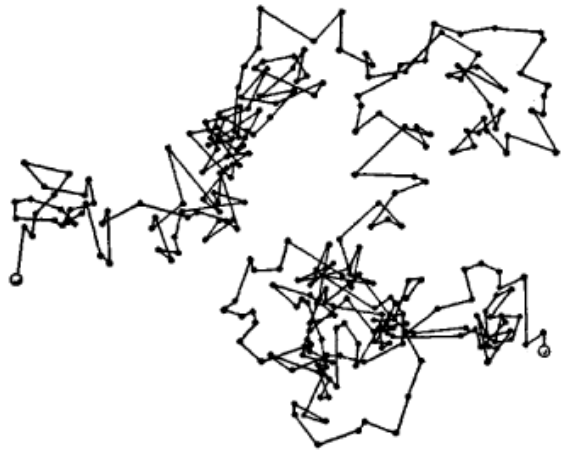
Physics

- Movement of a particle in a random environment:

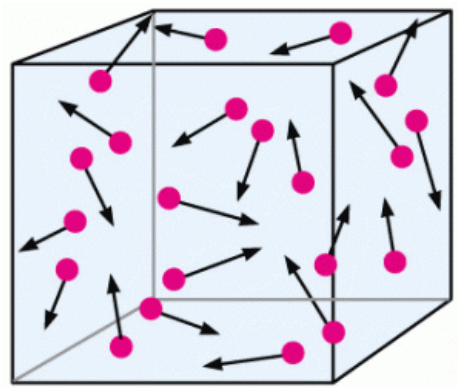


Physics

- Movement of a particle in a random environment:

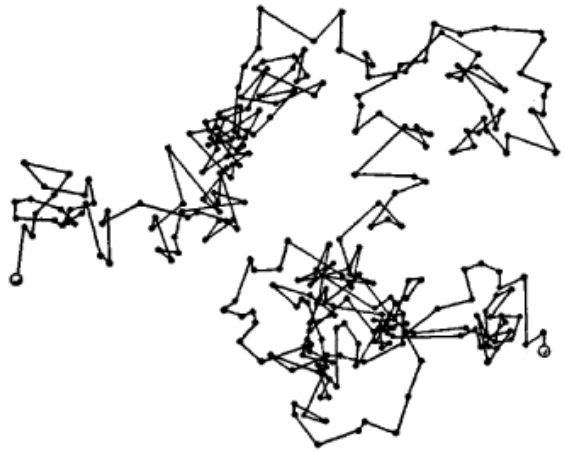


- statistical physics: macroscopic properties of complex systems

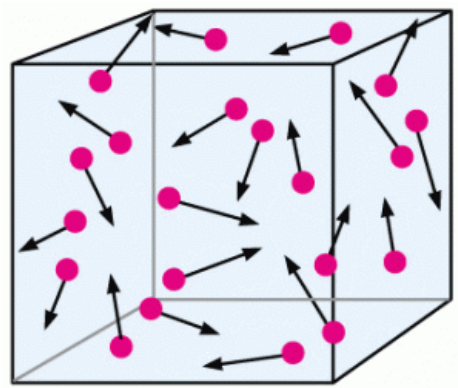


Physics

- Movement of a particle in a random environment:



- statistical physics: macroscopic properties of complex systems



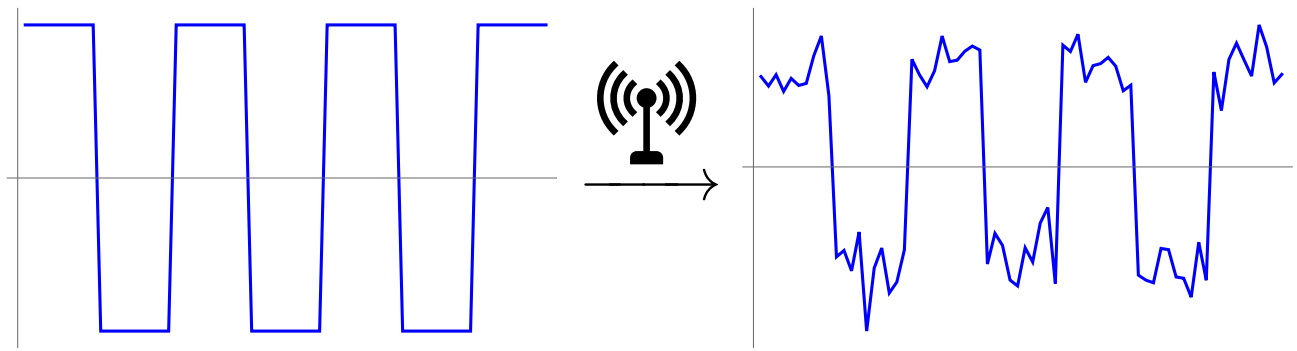
- quantum mechanics: probabilistic nature of physical phenomena at atomic scales

Natural/Social sciences

- Spread of epidemics/propagation of rumors
- Extinction of species/family names
- ...

How to contend with randomness/uncertainty

Wireless communication



Communication systems are engineered based on probabilistic models to combat noise or faulty operations.

Statistical inference

- Decision making based on data in the presence of noise

Overbooking

Overselling is a common practice in the travel and lodging industry, in which it is expected that some buyer will cancel or not show up

PBS NEWSHOUR

SUBSCRI

MAKING SEN$E

How Delta masters the game of overbooking flights

BY MAX NESTERAK *December 29, 2015 at 1:49 PM EDT*

MAX NESTERAK

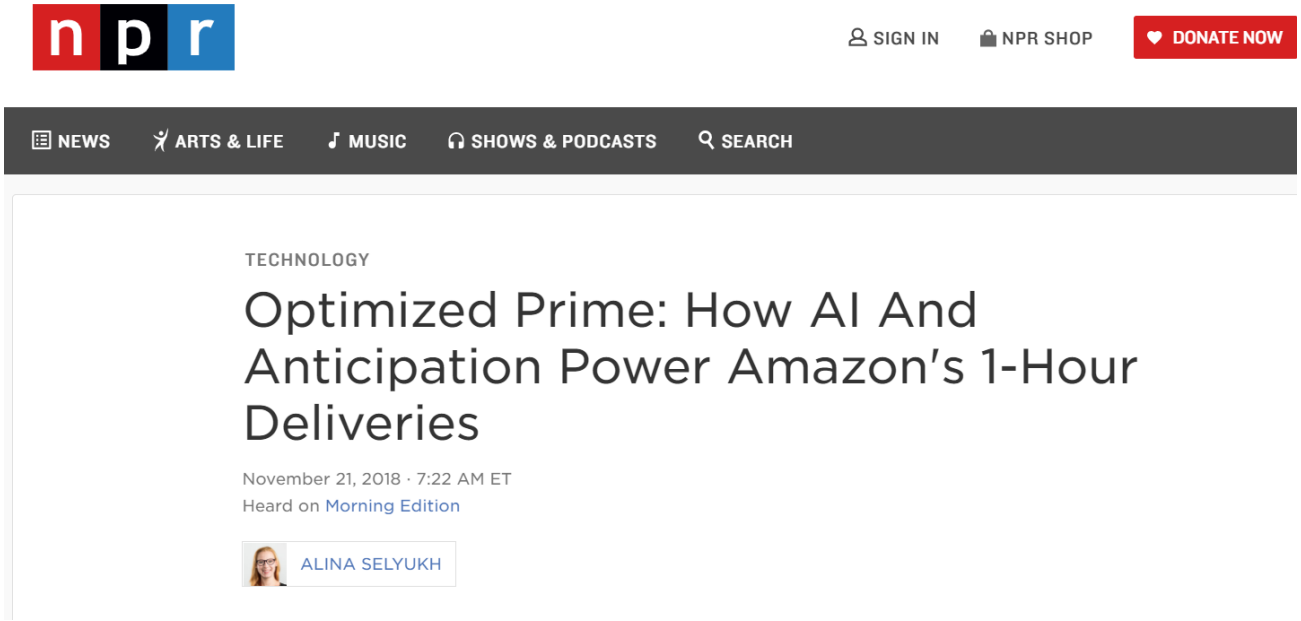
FOLLOW

3180

EMAIL

How one airline is changing the game when it comes to overbooked flights. Photo by Jeff Haynes/Reuters

Predictive analytics



npr

SIGN IN NPR SHOP DONATE NOW

NEWS ARTS & LIFE MUSIC SHOWS & PODCASTS SEARCH

TECHNOLOGY

Optimized Prime: How AI And Anticipation Power Amazon's 1-Hour Deliveries

November 21, 2018 · 7:22 AM ET
Heard on Morning Edition

ALINA SELYUKH

the warehouses to the actual delivery to your doorstep. In corporate lingo, Amazon calls that the "first mile," "middle mile" and "last mile."

Forecasting for the "first mile"

In 2013, Amazon got a patent for so-called "anticipatory shipping." The idea was to get your order as close as possible to your address before you actually click buy.

Since then, Amazon has built a massive warehousing footprint around the country. And it's been adding smaller warehouses closer to city centers where Prime Now promotes super-fast delivery options. It's also using Whole Foods locations for faster access to groceries and basic pickups.

Amazon regularly tests what new products people might want with extra-fast shipping,

Randomness as a great resource

How to sauté onions?

How to sauté onions?

Two ways:

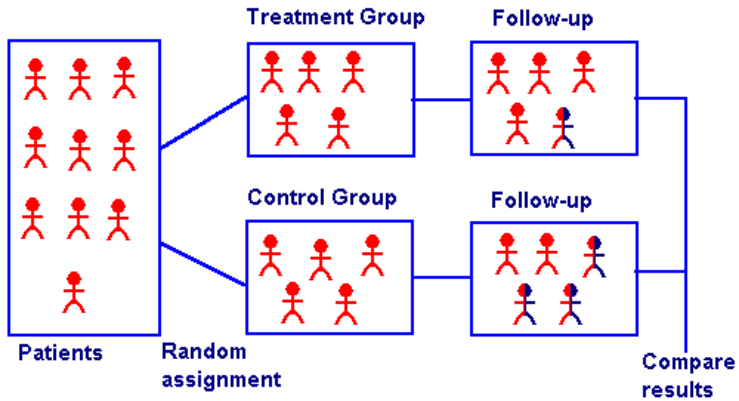
1. Carefully brown each side evenly; or

How to sauté onions?

Two ways:

1. Carefully brown each side evenly; or
2. Stir randomly!

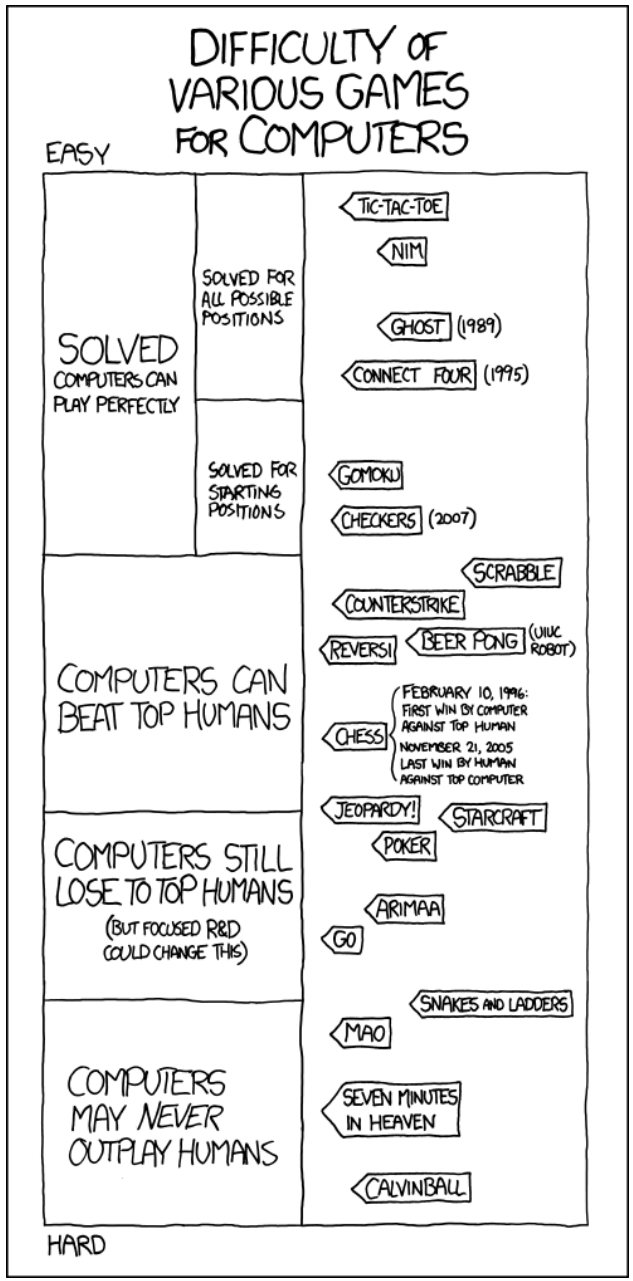
Randomized controlled trial



The method of choice to test the efficacy of a new drug or medical procedure in clinical trial:

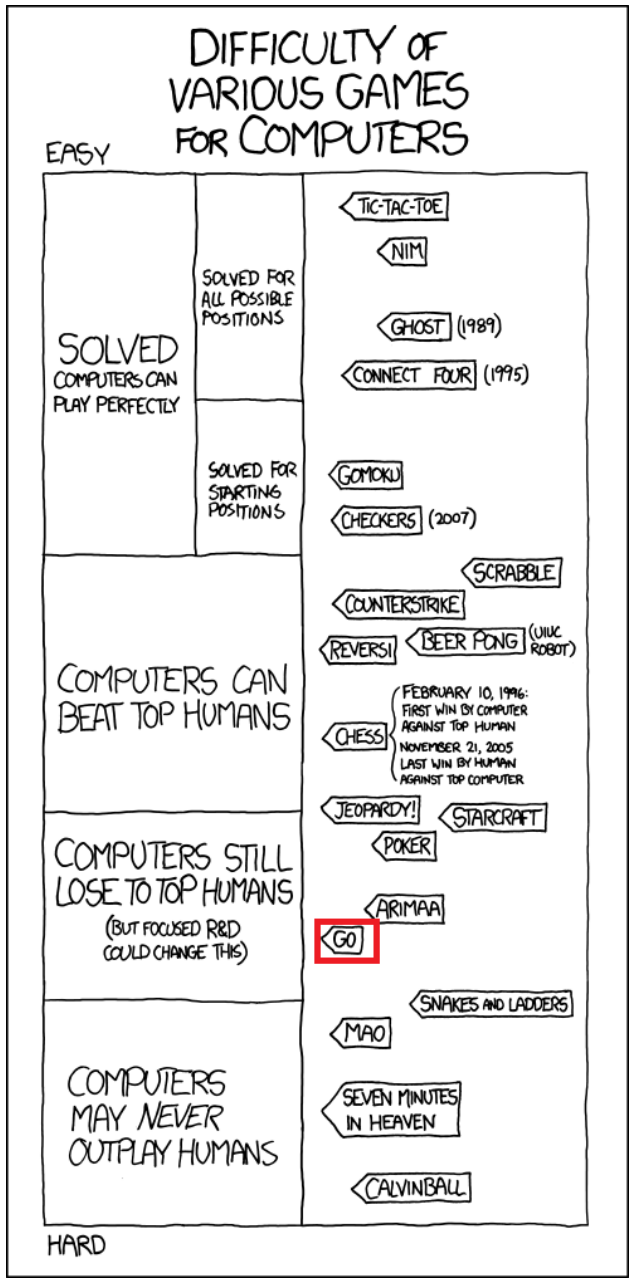
- Participants in the trial are randomly assigned to groups to receive either the treatment or placebo
- Scientifically sound, eliminate bias, ...

Game AI



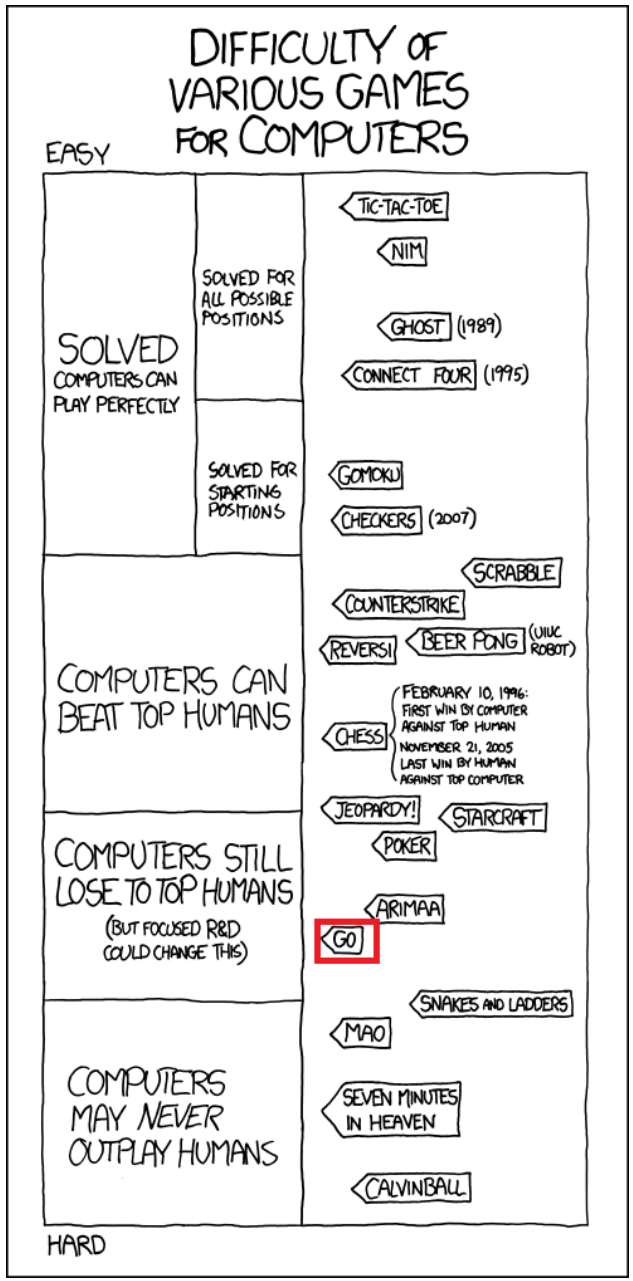
https://xkcd.com/1002/ Jan 2012

Game AI



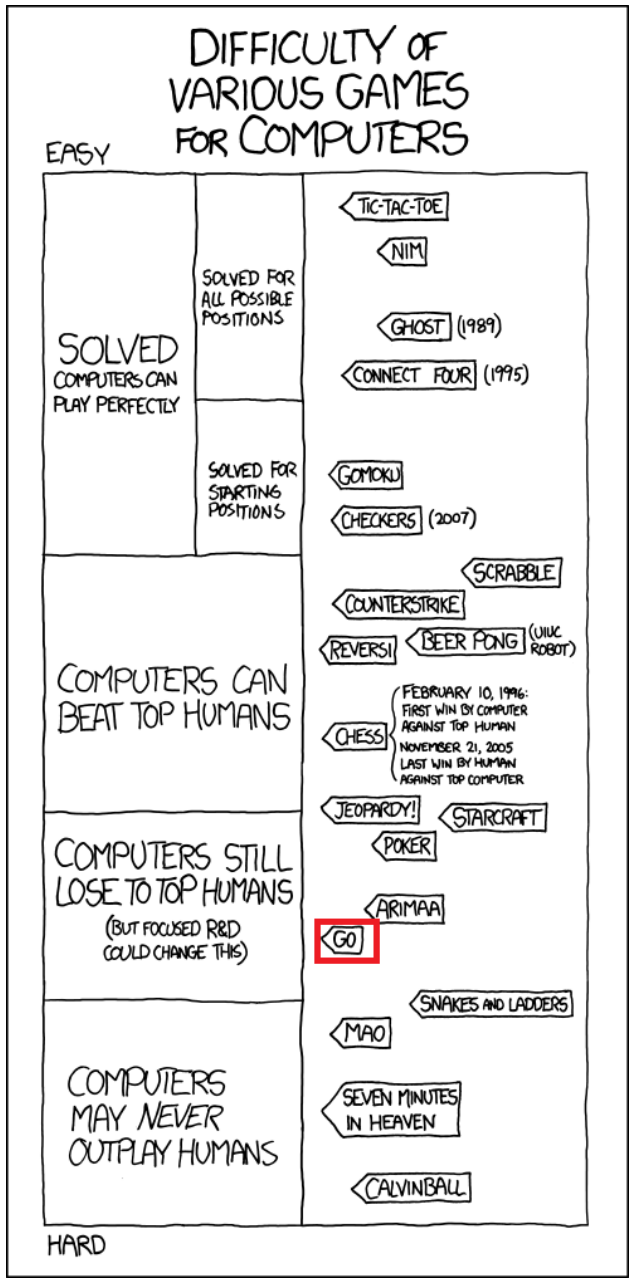
https://xkcd.com/1002/ Jan 2012

Game AI



https://xkcd.com/1002/ Jan 2012

Game AI



https://xkcd.com/1002/ Jan 2012

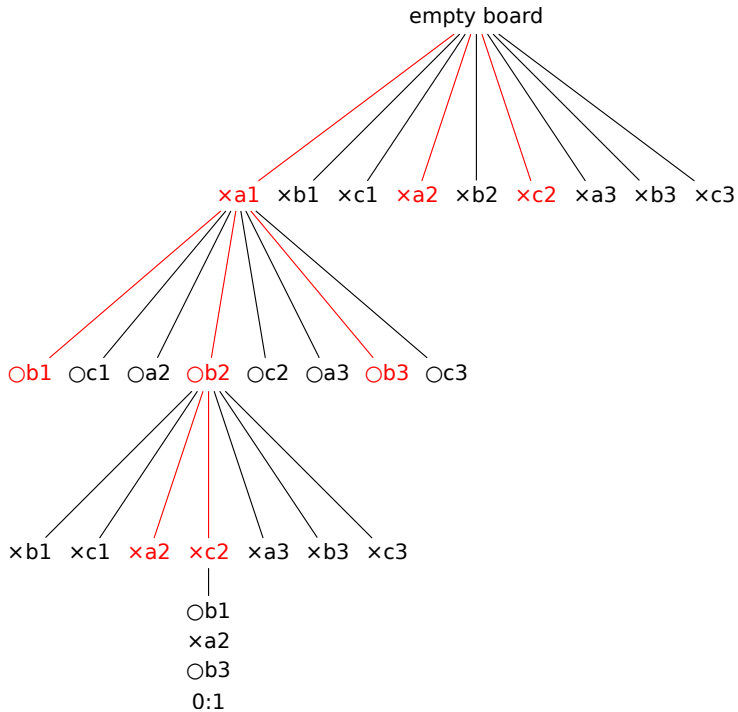
Go

Two-player deterministic sequential game

Why is go much harder than chess?

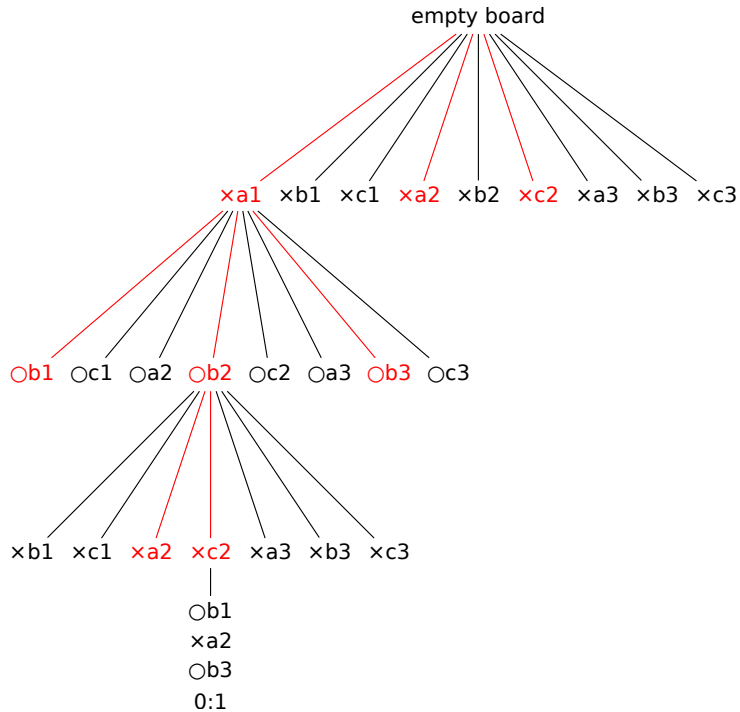
The vast strategy space!

Game tree



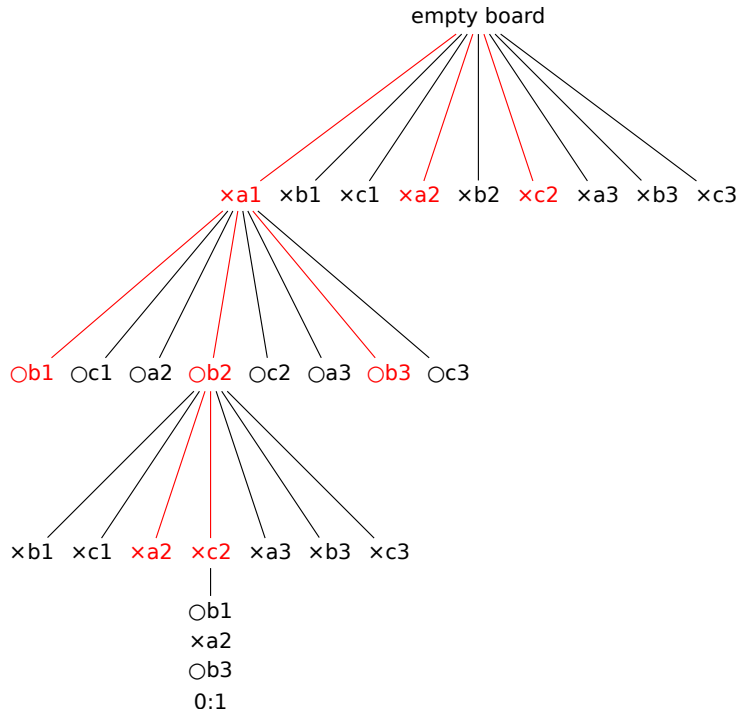
- Start: $19 \times 19 = 361$ possible moves

Game tree



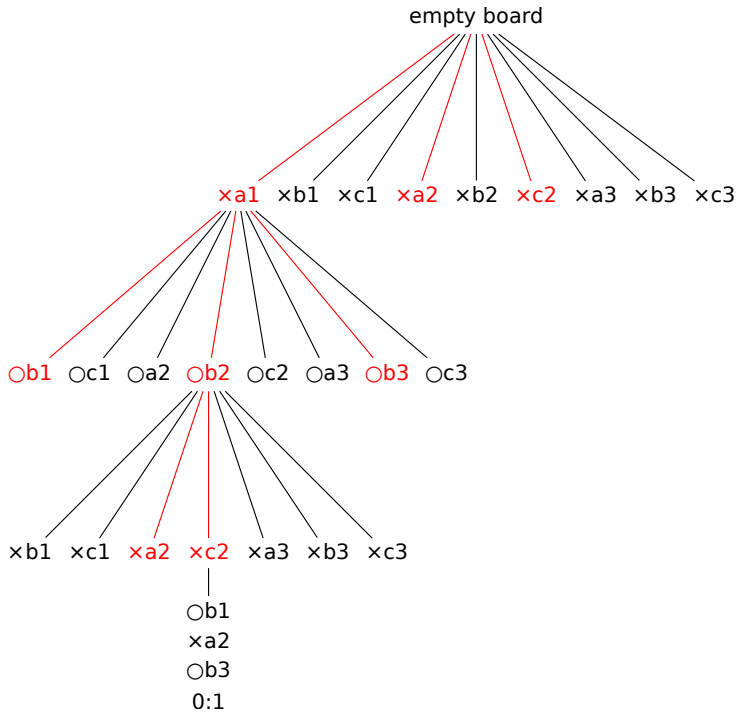
- Start: $19 \times 19 = 361$ possible moves
- Branching factor (average number of moves) of Go: 250 (Chess: 35)

Game tree



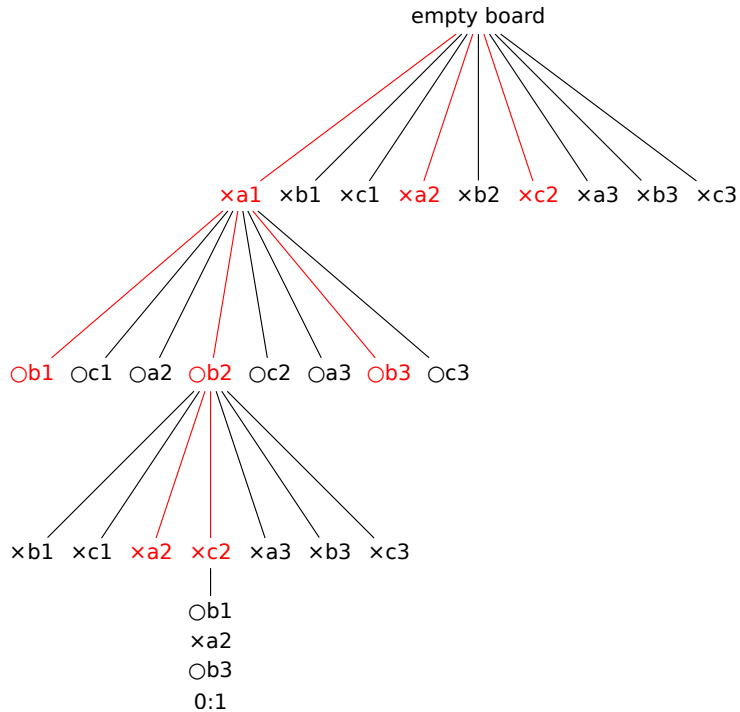
- Start: $19 \times 19 = 361$ possible moves
- Branching factor (average number of moves) of Go: 250 (Chess: 35)
- Length: 200 turns (Chess: 40)

Game tree



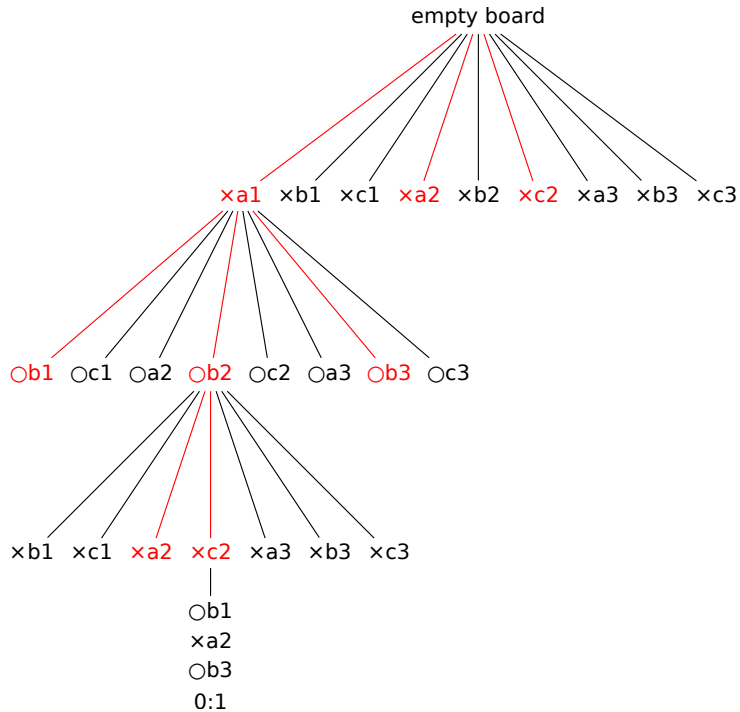
- Start: $19 \times 19 = 361$ possible moves
- Branching factor (average number of moves) of Go: 250 (Chess: 35)
- Length: 200 turns (Chess: 40)
- To anticipate next three moves of opponent using the supercomputer Deep Blue (200 million positions/sec):

Game tree



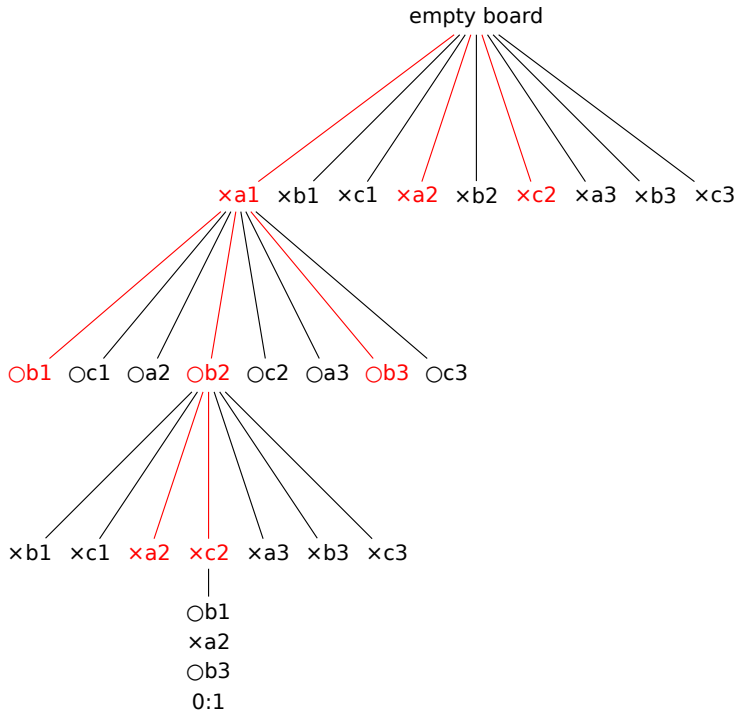
- Start: $19 \times 19 = 361$ possible moves
- Branching factor (average number of moves) of Go: 250 (Chess: 35)
- Length: 200 turns (Chess: 40)
- To anticipate next three moves of opponent using the supercomputer Deep Blue (200 million positions/sec):
  - Chess: $35^6/(200 \text{ million}) \approx 9$ sec

Game tree



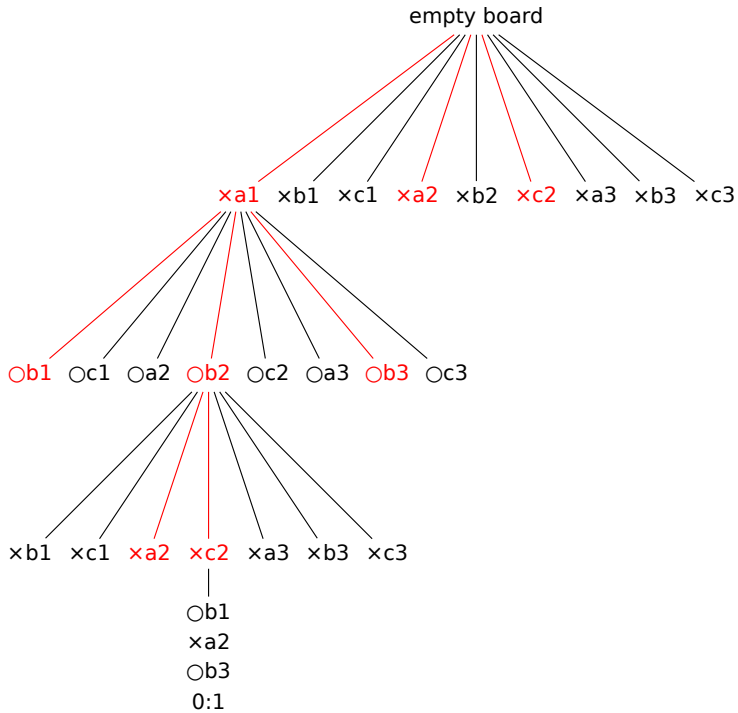
- Start: $19 \times 19 = 361$ possible moves
- Branching factor (average number of moves) of Go: 250 (Chess: 35)
- Length: 200 turns (Chess: 40)
- To anticipate next three moves of opponent using the supercomputer Deep Blue (200 million positions/sec):
  - Chess: $35^6/(200 \text{ million}) \approx 9$ sec
  - Go: $250^6/(200 \text{ million}) \approx 2$ weeks

Complexity of Go



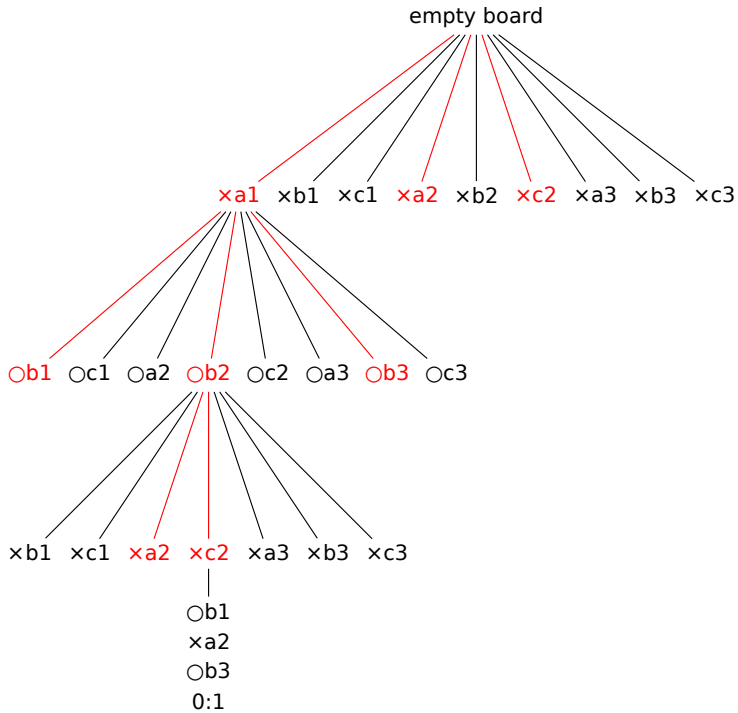
- 10^{360} possibilities

Complexity of Go



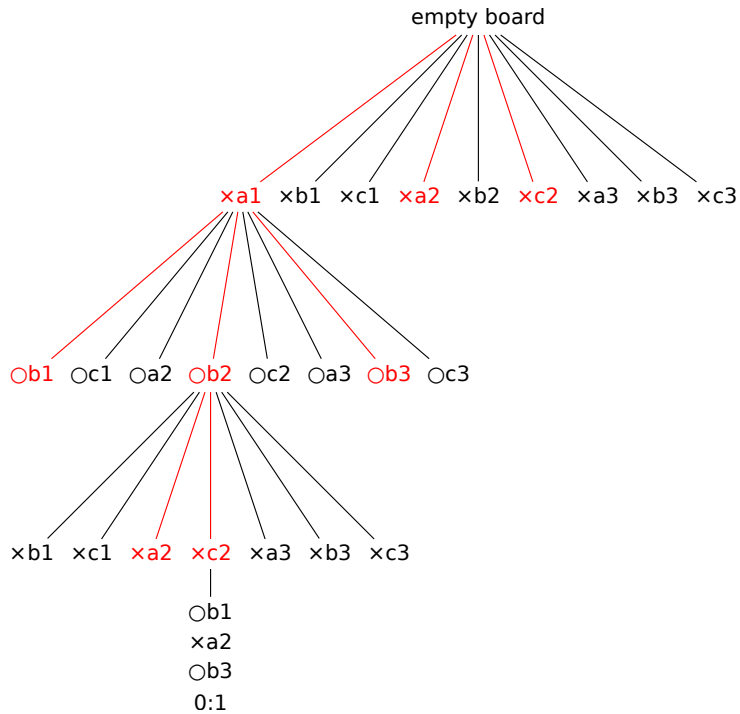
- 10^{360} possibilities
- Impossible to search for the best move

Complexity of Go



- 10^{360} possibilities
- Impossible to search for the best move
- Ingenious idea: do it randomly!

Monte-Carlo tree search



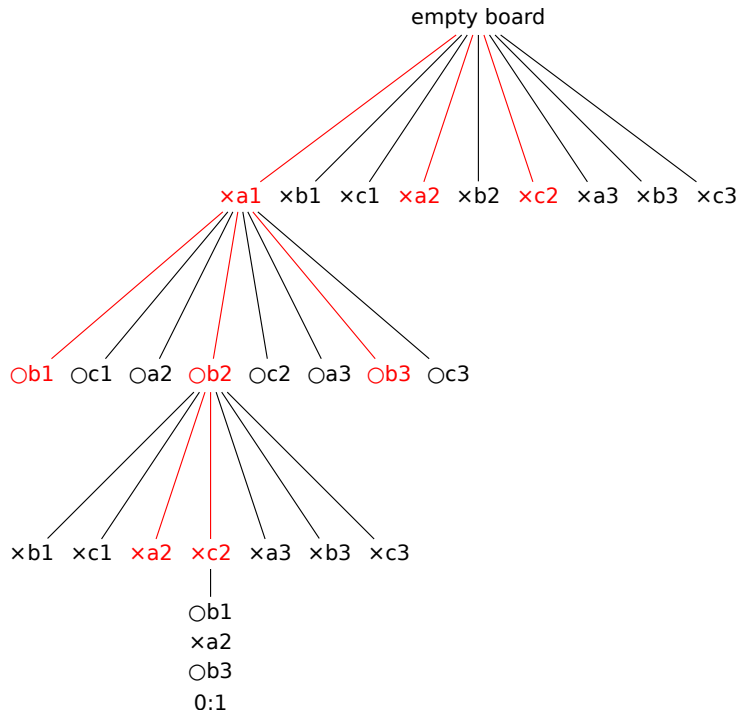
- From a given position, play 10000 games completely at random.

---

http://www.wired.com/2014/05/the-world-of-computer-go/

http://www.scientificamerican.com/article/how-the-computer-beat-the-go-master/

Monte-Carlo tree search



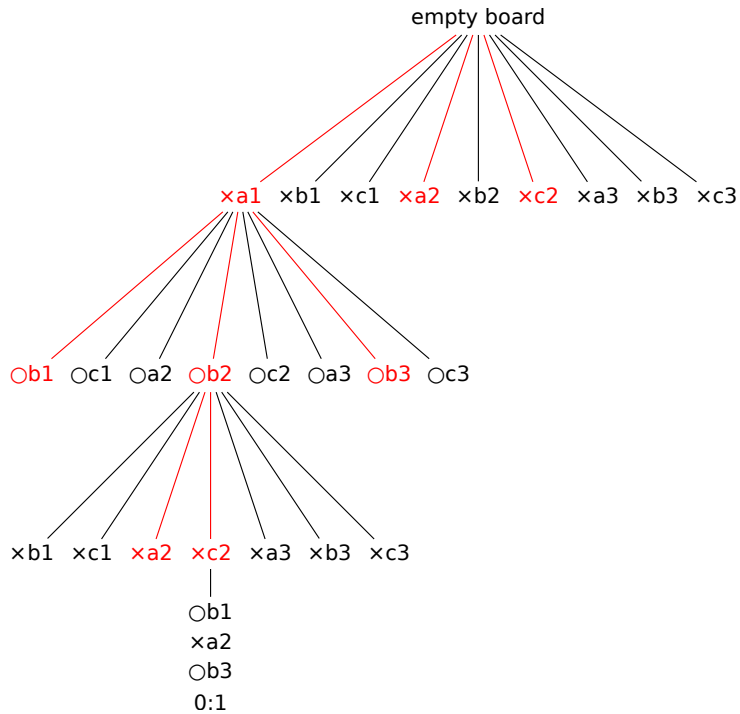
- From a given position, play 10000 games completely at random.
- Choose the most promising move that leads to the most wins.

---

http://www.wired.com/2014/05/the-world-of-computer-go/

http://www.scientificamerican.com/article/how-the-computer-beat-the-go-master/

Monte-Carlo tree search

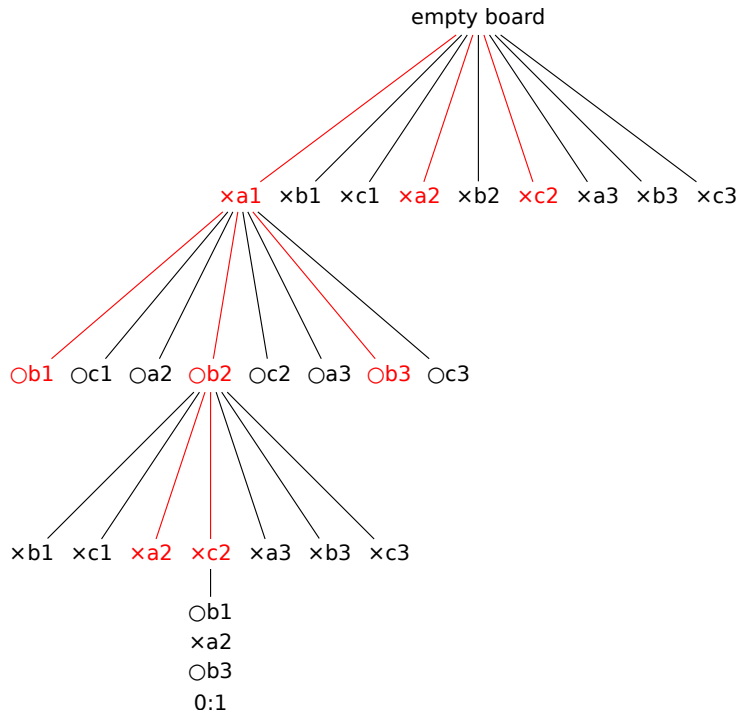


- From a given position, play 10000 games completely at random.
- Choose the most promising move that leads to the most wins.
- '2006: can beat pros in blitz games or with handicap

`http://www.wired.com/2014/05/the-world-of-computer-go/`

`http://www.scientificamerican.com/article/how-the-computer-beat-the-go-master/`

Monte-Carlo tree search



- From a given position, play 10000 games completely at random.
- Choose the most promising move that leads to the most wins.
- '2006: can beat pros in blitz games or with handicap
- '2018: can beat world champion (combined with deep learning)

`http://www.wired.com/2014/05/the-world-of-computer-go/`
`http://www.scientificamerican.com/article/how-the-computer-beat-the-go-master/`

List of topics

Topics

- Probability axioms. Events and random variables. Expectation. Independence.
- Conditional probability. Law of total probability. Bayes formula.
- Discrete random variables.
- Continuous random variables.
- Multiple random variables.

Topics

- Transformation of distributions.
- Sum of independent random variables. Covariance and correlation.
- Law of large numbers.
- Central Limit Theorem (CLT): Binomial (De Moivre-Laplace).
- Generating Functions. General CLT.
- Branching process. Markov Chains.

Objective of this class

- Develop probablistic intuitions
- Both qualitative and quantitative reasoning skills

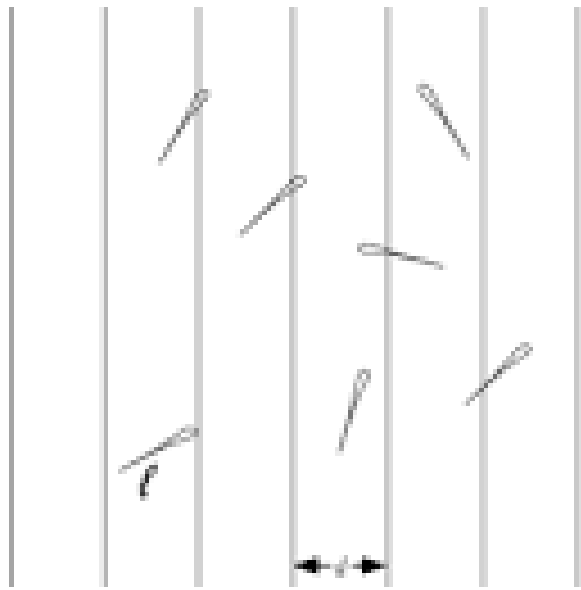
Some examples

Gambler's ruin

You enter a casino with \$1 in your pocket. You stake \$$a$. You either win \$$a$ with probability 0.5001 or lose \$$a$ with probability 0.4999. If the casino lets you play long enough and allows you to stake arbitrarily small amount, can you eventually own the casino?

Buffon's needle

Suppose we have a floor made of parallel strips of wood, each of unit width, and we drop a needle of unit length onto the floor. What is the probability that the needle will touch a line between two strips? Historically this experiment were used to approximate π.



Hypothesis testing: Is the coin biased?

Toss a coin 2000 times. Assume that the chance to get a head is p in each toss. You observe 1100 heads. Are you confident to say that $p \neq 1/2$?

Hypothesis testing: Who will win the election?

Poll 2000 people uniformly at random from a population: 1100 people say they will vote for Candidate A and 900 for Candidate B. How confident are you about that Candidate A will win at the election?