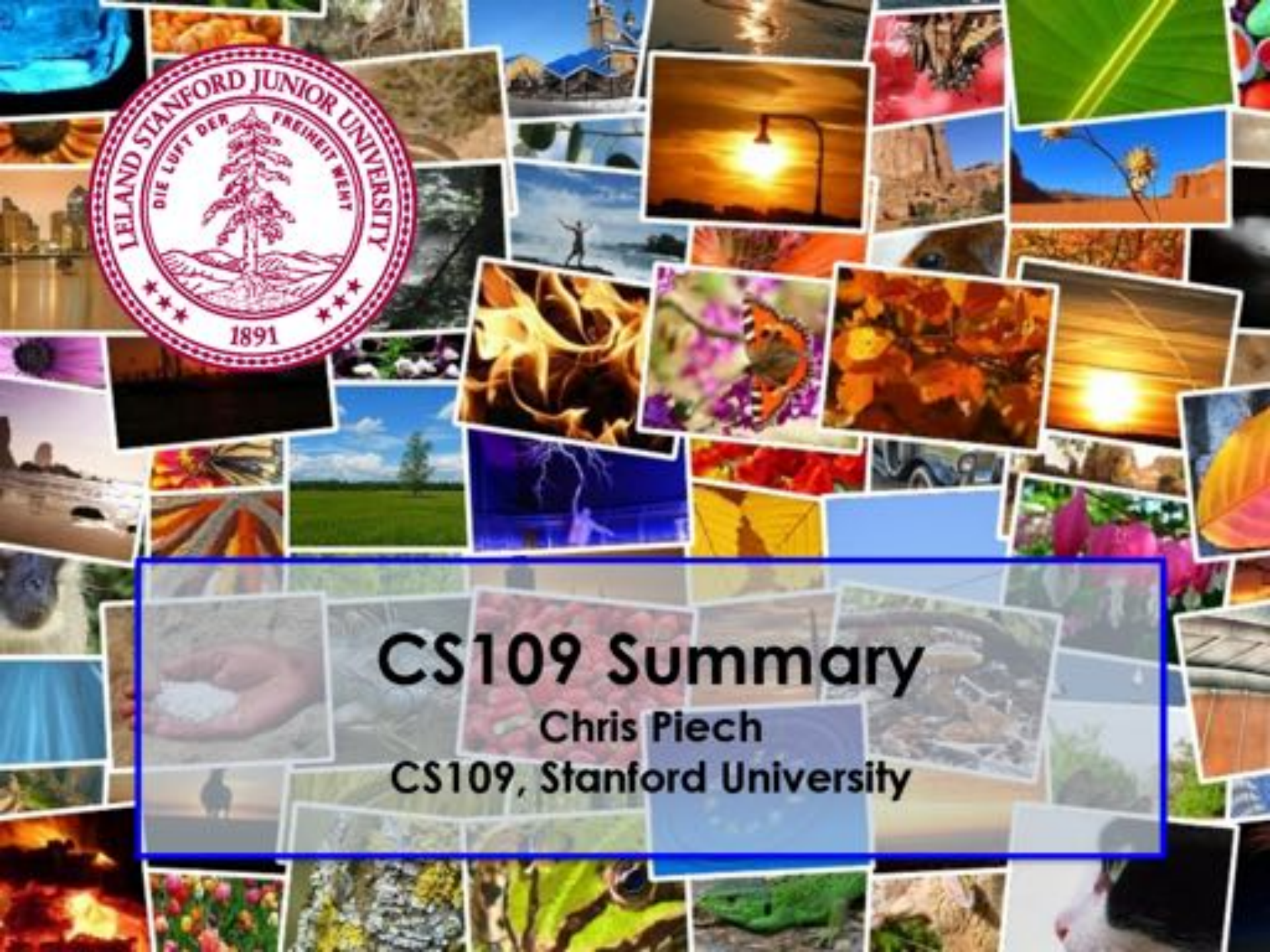


CS109 Summary

Chris Piech
CS109, Stanford University

Counting Rules

Counting operations on n objects

Sort (order matters) (perms)

- Distinct: $n!$
- Some Distinct: $\frac{n!}{n_1!n_2!\ldots}$

Choose k (combinations)

- Distinct: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Put in r buckets

- Distinct: r^n
- None Distinct: $\frac{(n+r-1)!}{n!(r-1)!}$

What is a Probability?

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

Sources of Probability

1. Experimentation

2. Dataset

3. Analytic Solution

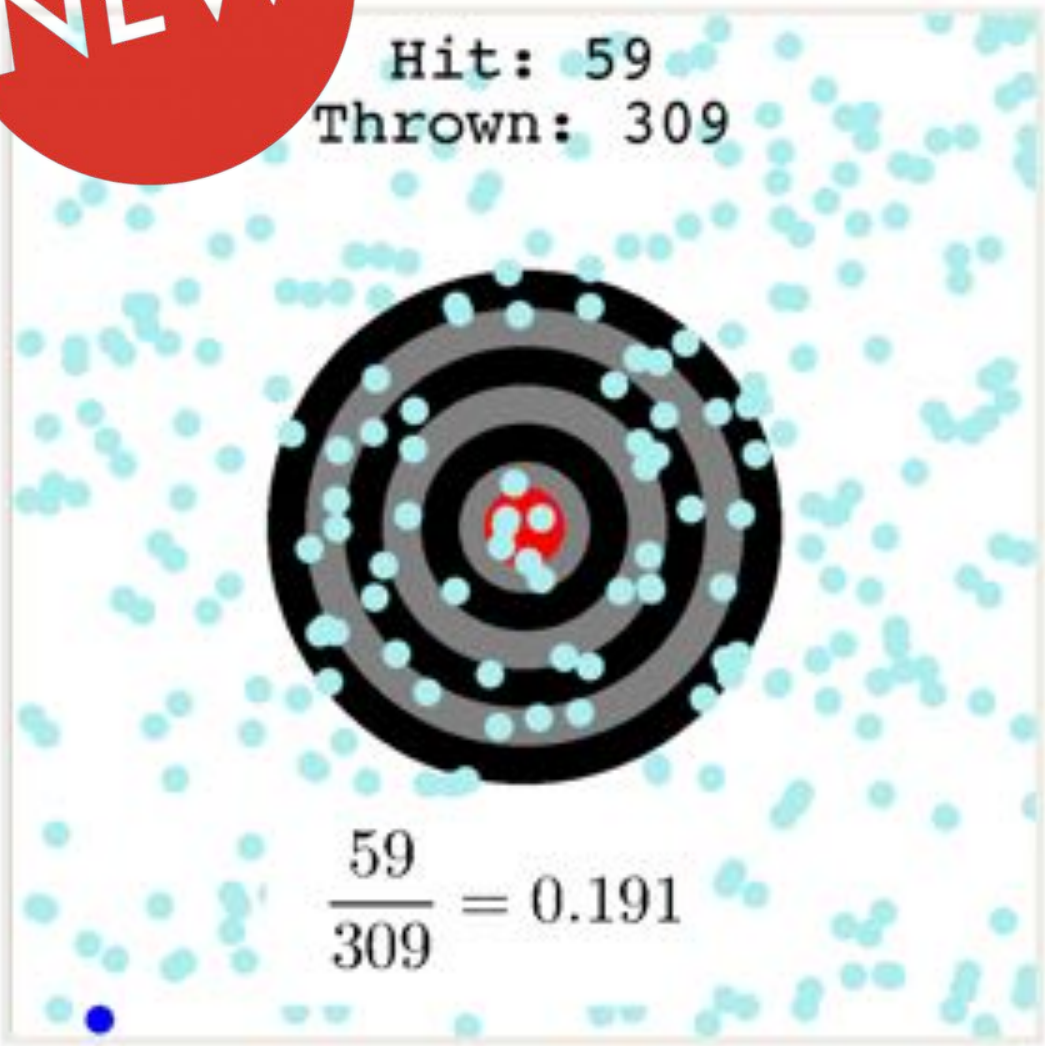
4. Expert Opinion


IMAGINE THAT YOU'RE DRAWING AT RANDOM FROM AN URN CONTAINING FIFTEEN BALLS— SIX RED AND NINE BLACK.
OK. I REACH IN AND... ...MY GRANDFATHER'S ASHES?!? OH GOD!
I... WHAT?
WHY WOULD YOU DO THIS TO ME?!?

Third Class with Coding!

Target Revisited

NEW



Screen size = 800x800
Radius of target = 200

The dart is equally likely to land anywhere on the screen.

What is the probability of hitting the target?

$$|S| = 800^2$$

$$|E| = \pi 200^2$$

$$p(E) = \frac{\pi \cdot 200^2}{800^2} \approx 0.1963$$

Sending Bit Strings

- Bit string with m 0's and n 1's sent on network
  - All distinct arrangements of bits equally likely
  - E = first bit received is a 1
  - F = k of first r bits received are 1's

P(E|F)?

*Think of the bits as distinct so that all outcomes are equally likely

Everything in the world is either

a potato or not a potato.

$$P(X) + P(X^C) = 1$$

WHEN YOU MEET YOUR BEST FRIEND

Somewhere you didn't expect to.

Trailing the dovetail shuffle to it's lair – Persi Diaconosis

Netflix and Learn

What is the probability that a user will watch Life is Beautiful?

$$P(E)$$

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \approx \frac{\#\text{people who watched movie}}{\#\text{people on Netflix}}$$

$$P(E) = 10{,}234{,}231 \,/\, 50{,}923{,}123 = 0.20$$

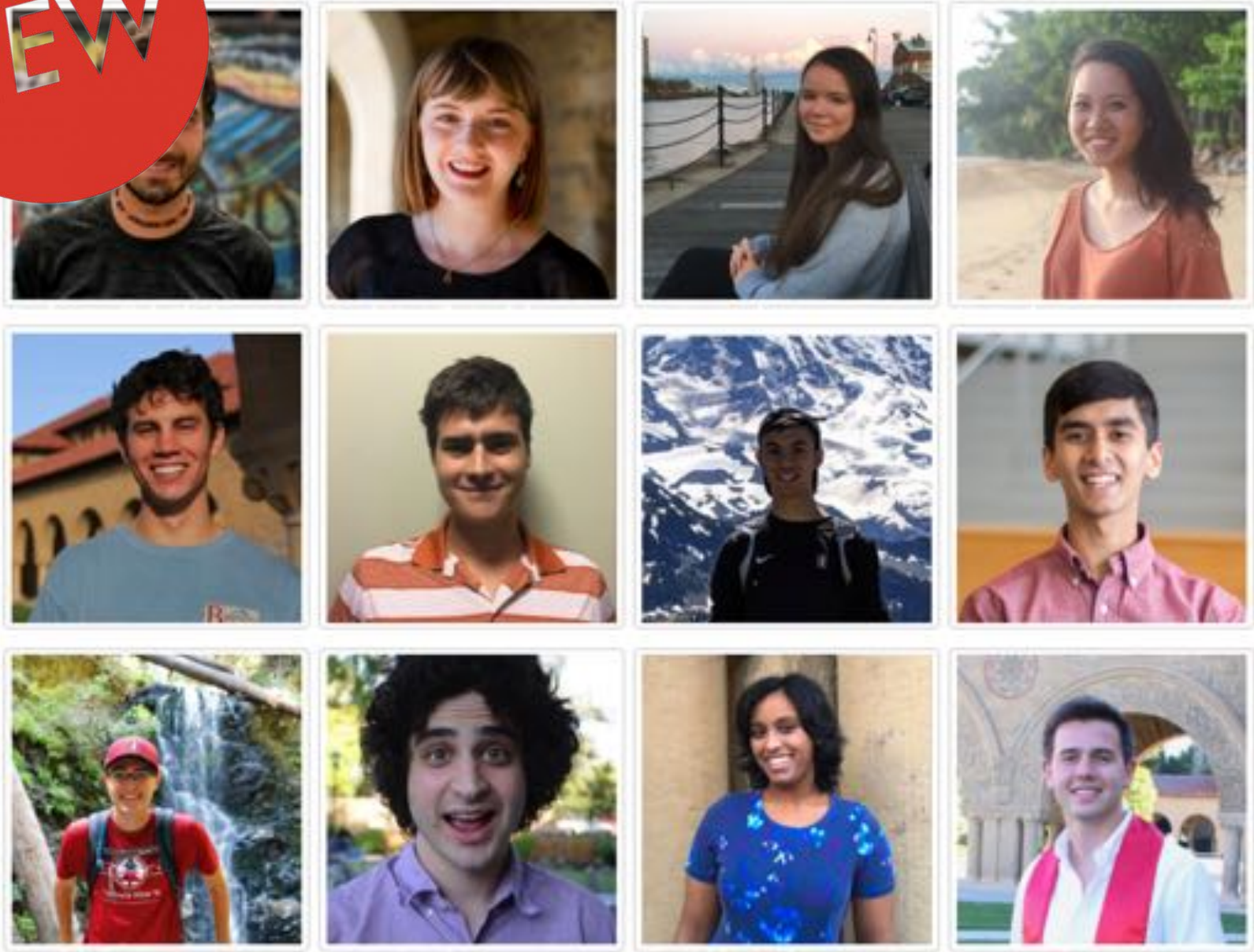
Let's Make a Deal

- Game show with 3 doors: A, B, and C

  - Behind one door is prize (equally likely to be any door)
  - Behind other two doors is nothing
  - We choose a door
  - Then host opens 1 of other 2 doors, revealing nothing
  - We are given option to change to other door
- Should we?
  - Note: If we don't switch, P(win) = 1/3 (random)

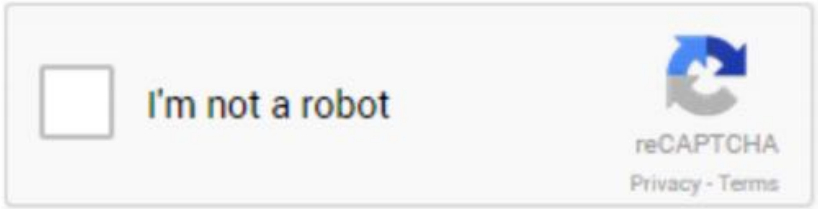
Second Ever Sections

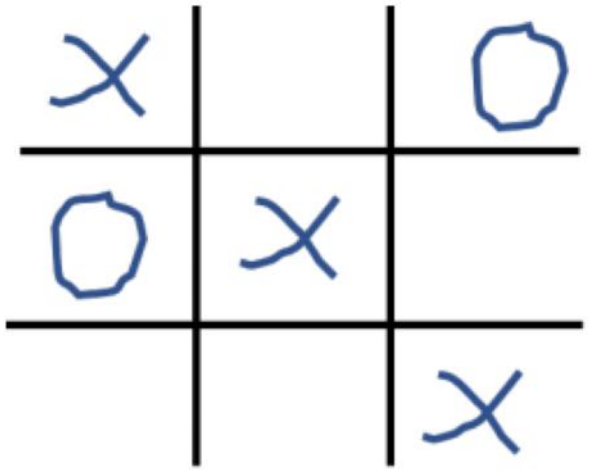
NEW



Second Ever Sections

NEW





Zika Test

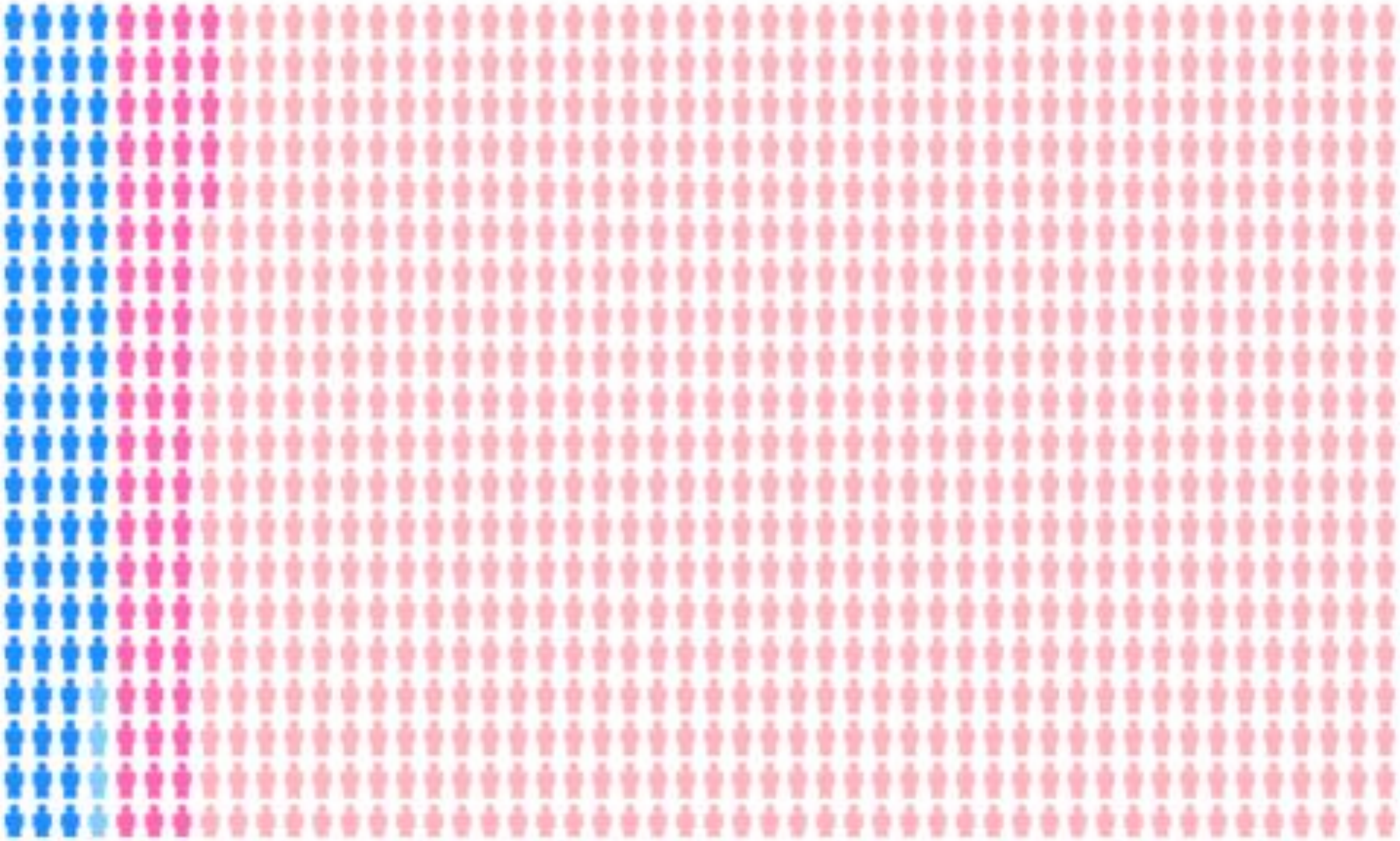
Positive Zika.

What is the probability of zika?

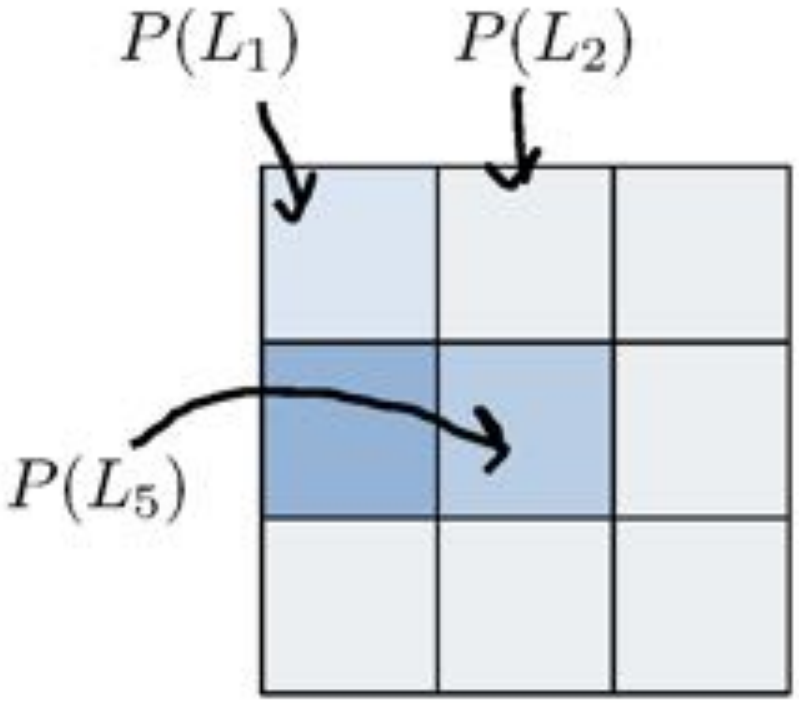
- *0.1% of people have zika*
- *90% positive rate for people with zika*
- *7% positive rate for people without zika*

The right answer is 1%

Bayes Theorem Intuition



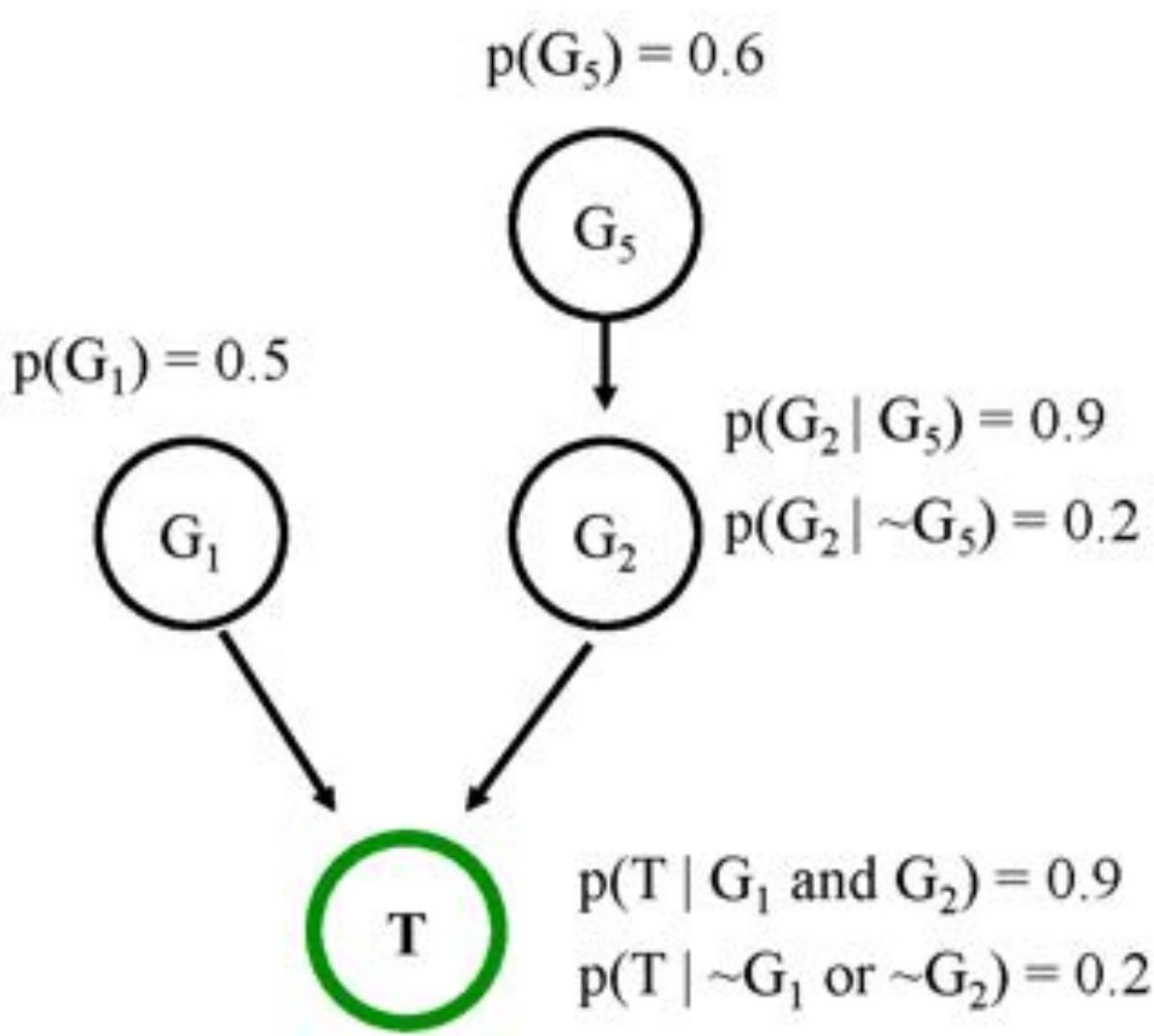
Update Belief



Before Observation

Recall our Ebola Bats

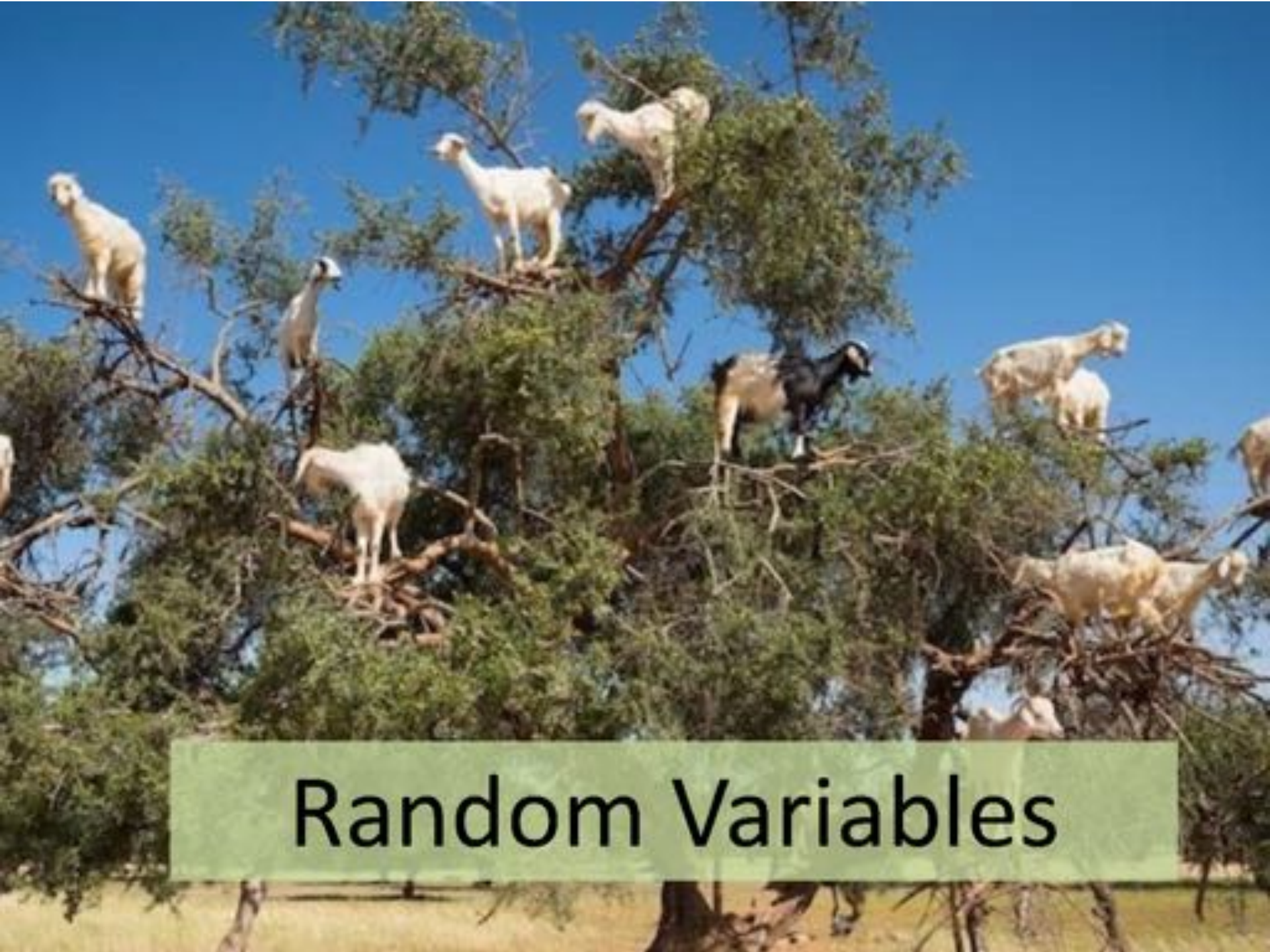
Discovered Pattern



These genes don't impact T

G4

G3

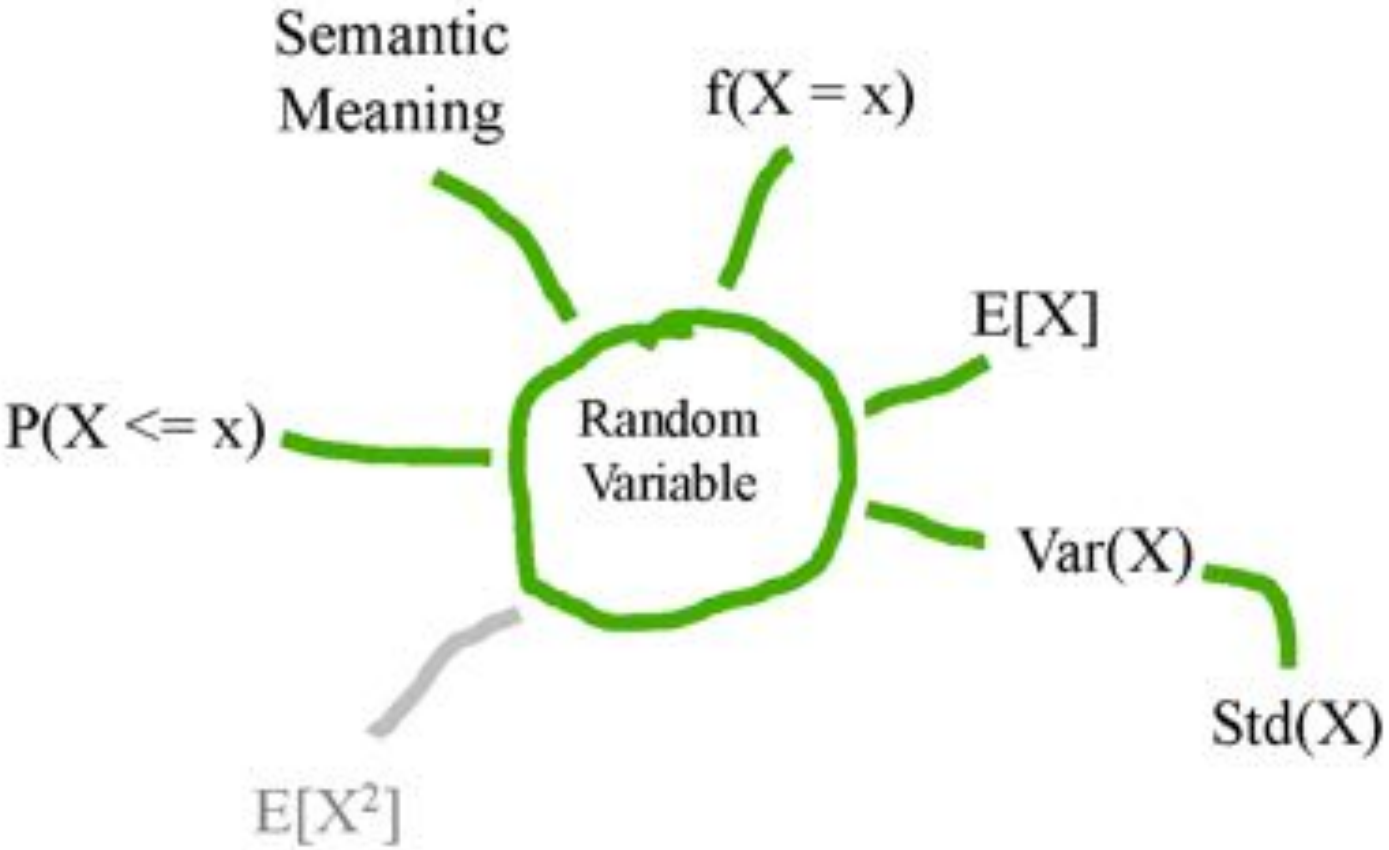

Random Variables

$$P(Y = k)$$

This is a function

For example Y is the number of heads in 5 coin flips

Fundamental Properties



Expectation

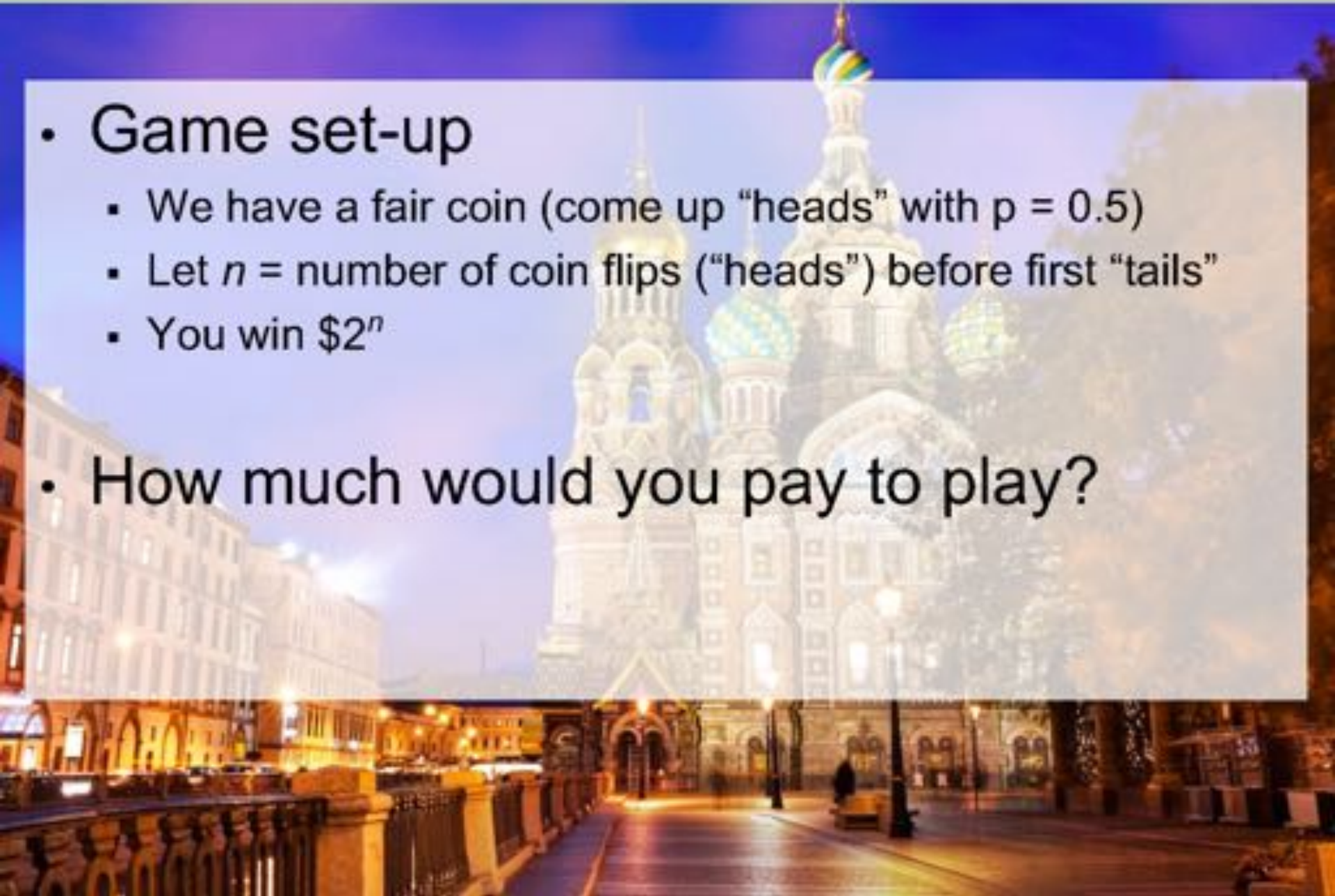
Big deal lemma: first stated without proof

$$E[X + Y] = E[X] + E[Y]$$

Generalized: $$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

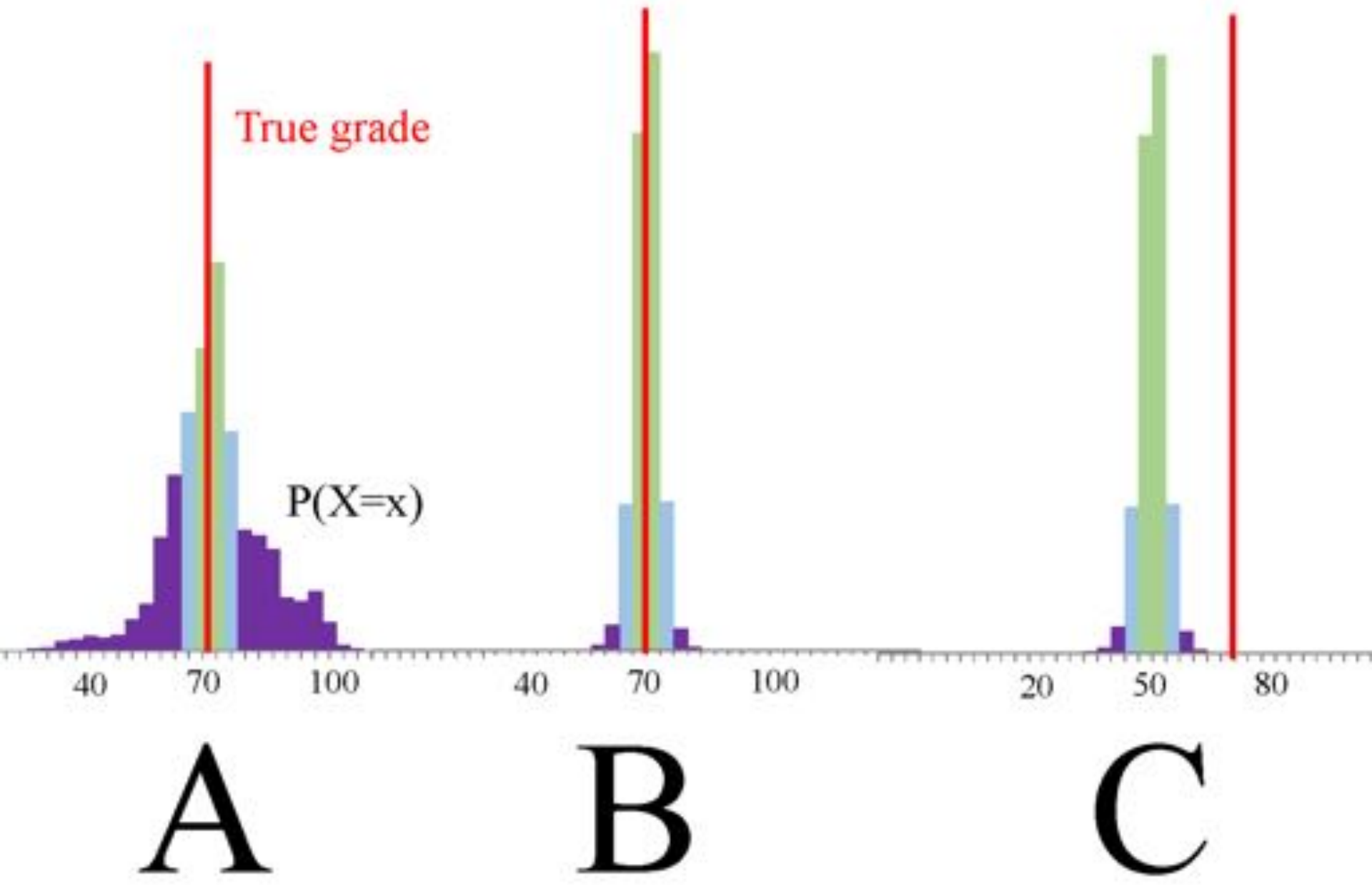
Holds regardless of dependency between X_i's

St Petersburg

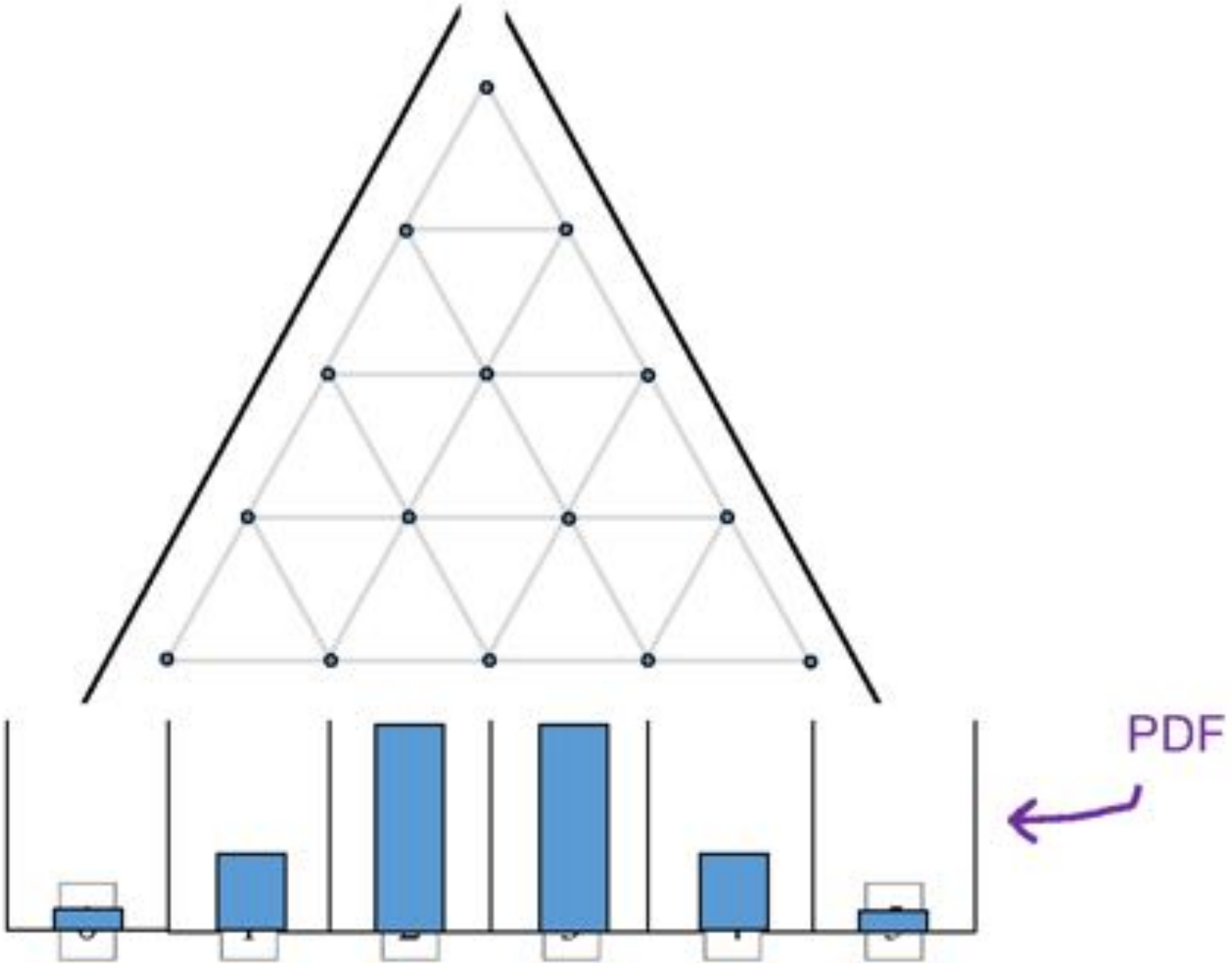


- Game set-up
  - We have a fair coin (come up "heads" with p = 0.5)
  - Let *n* = number of coin flips ("heads") before first "tails"
  - You win 2^n
- How much would you pay to play?

X is the score a peer grader gives to an assignment submission



Binomial




1001

Poisson

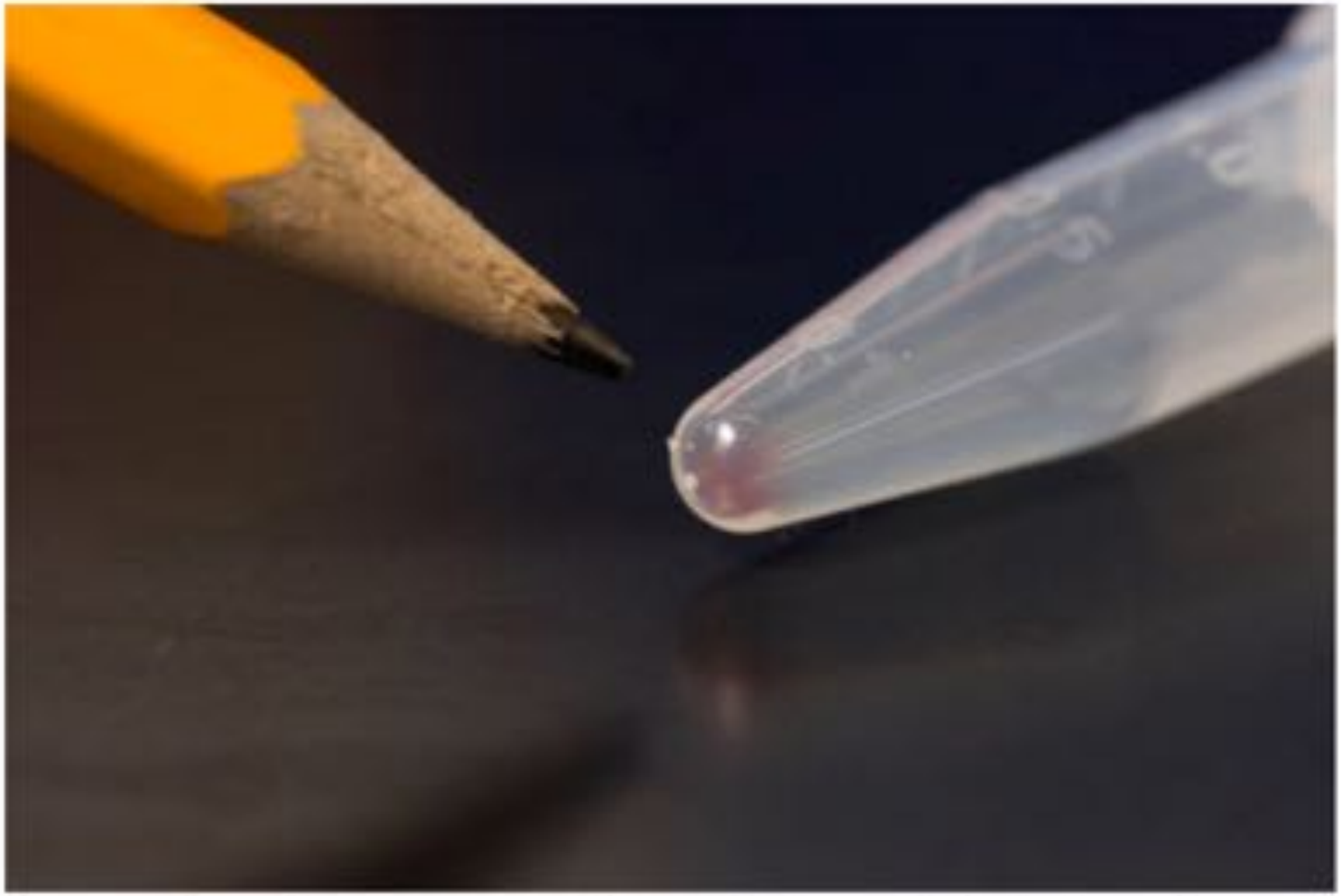

The Castro Theatre
THE CASTRO
MISSION DISTRICT
19th St
Atlas Cafe
20th St
Guerrero St
Alabama
Castro St
22nd St
On average λ = 5 requests per minute
Terra Mia Ceramic Studio
26th St
Diamond St
The Royal Cuckoo
Valley St
Folsom St
1591 Treat Avenue
Napoleon St
Toland St
Billy Goat Hill
BERNAL HEIGHTS
Lowe's Home Improvement
Lorraine Ave
431
55
25th St
Carolina St

Geometric

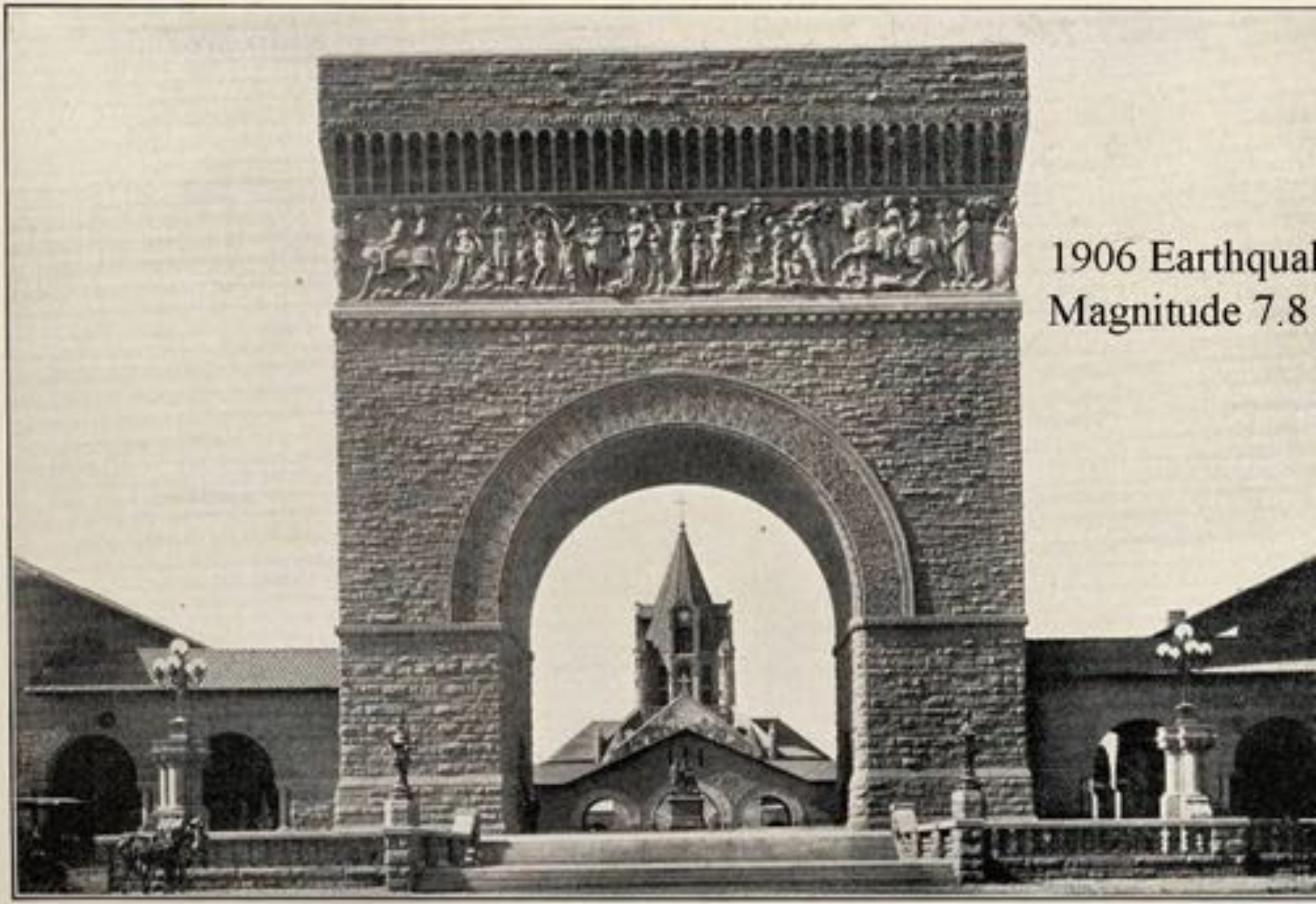
Sequence 1:

TTHHTHTTHTTTHTTTHTTTHTTHTHHTH
HTTHHTTHHHTHHTHTTHTHTTHHTHHHT
HTHHTHHHTHTHTTHTTHHTHTHTHTTHH
TTHHTHTTHTHTHTHTHTHTHHHTHTHTH
TTHHTHTHTHTHHHTTHHTHTTTHTHHTH

Storing Data on DNA



All the movies, images, emails and other digital data from more than 600 smartphones (10,000 gigabytes) can be stored in the faint pink smear of DNA at the end of this test tube.



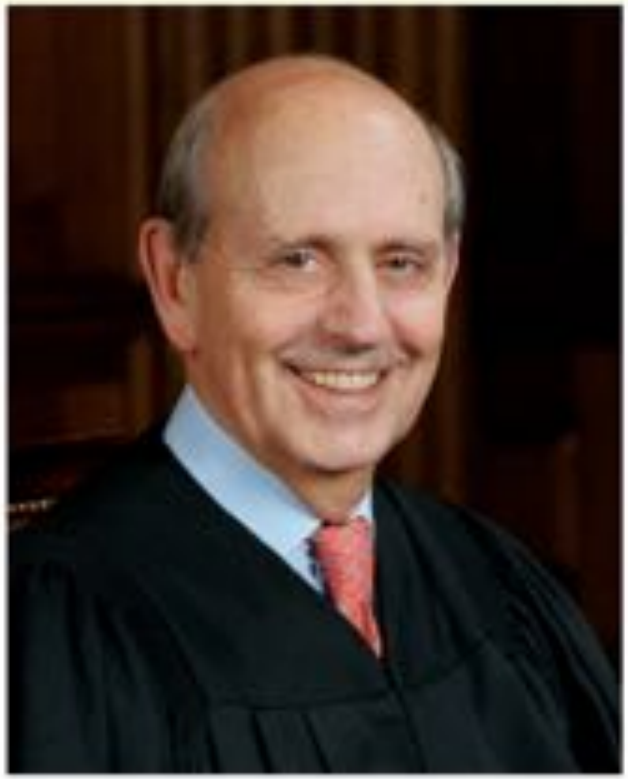
ILL. No. 65. MEMORIAL ARCH, WITH CHURCH IN BACKGROUND, STANFORD UNIVERSITY, SHOWING TYPES OF CARVED WO
WITH THE SANDSTONE.


Probability for Extreme Weather?

Bit Coin Mining

You "mine a bitcoin" if, for given data D, you find a number N such that Hash(D, N) produces a string that starts with g zeroes.

Representative Juries



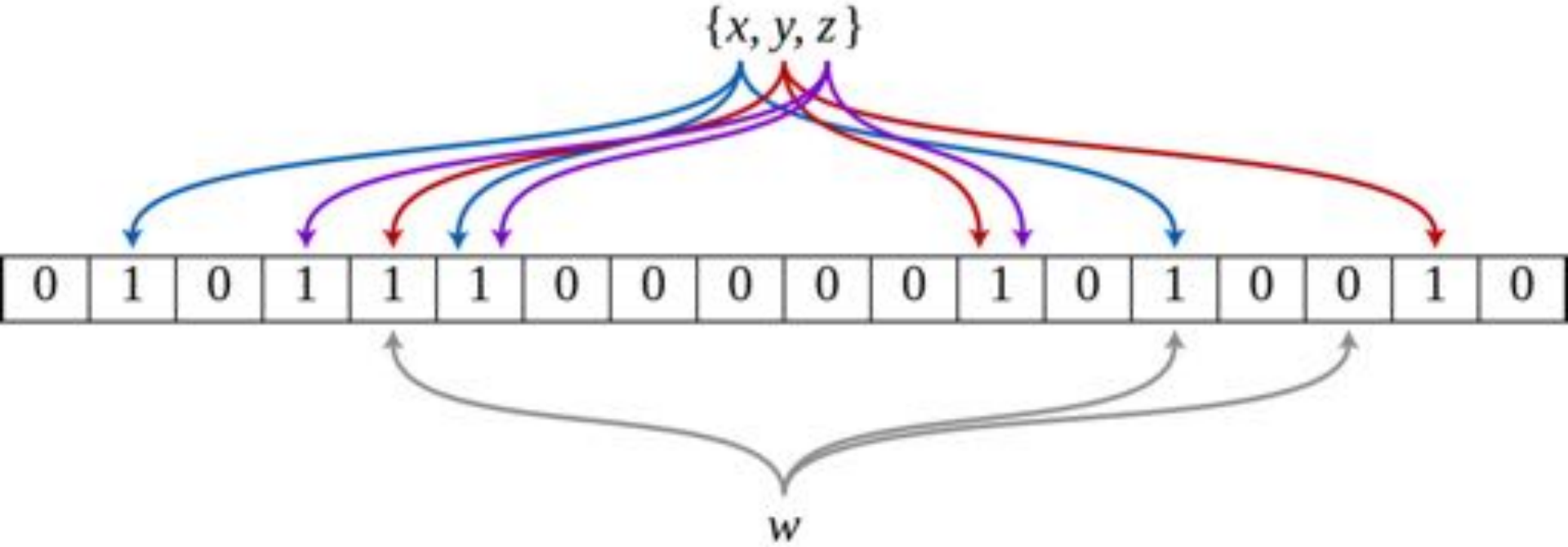
Simulate

Simulation:

Dating at Stanford

Each person you date has a 0.2 probability of being someone you spend your life with. What is the average number of people one will date? What is the standard deviation?

Bloom Filter

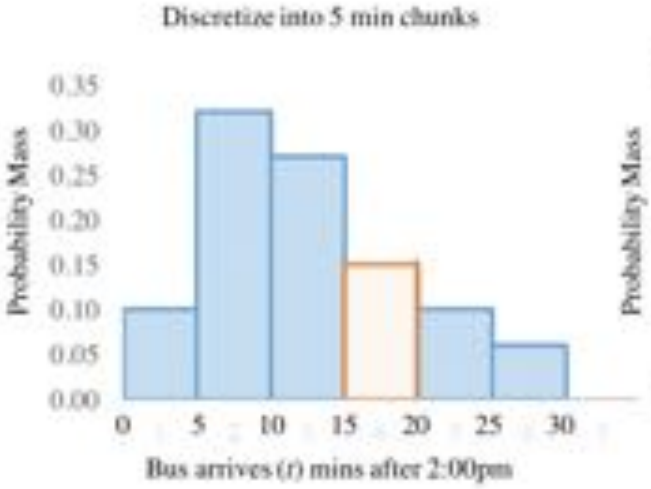


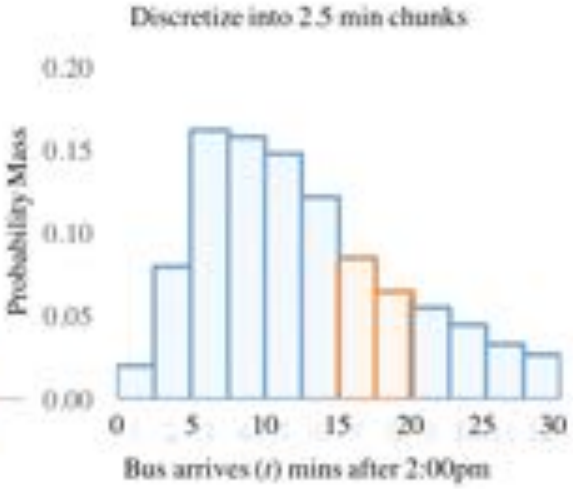
random()?

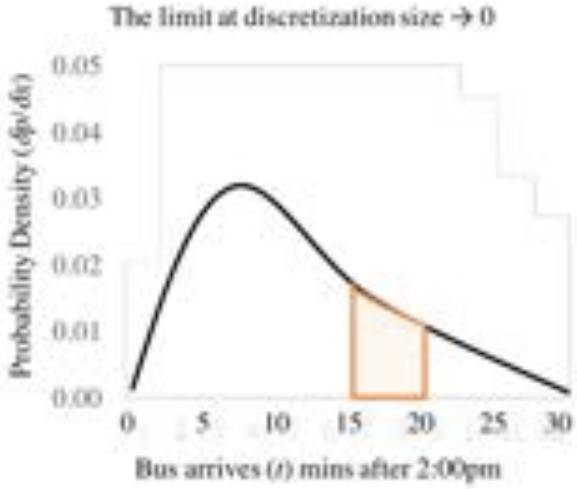
Riding the Marguerite

You are running to the bus stop. You don't know exactly when the bus arrives. You arrive at 2:20pm.

What is P(wait < 5 min)?







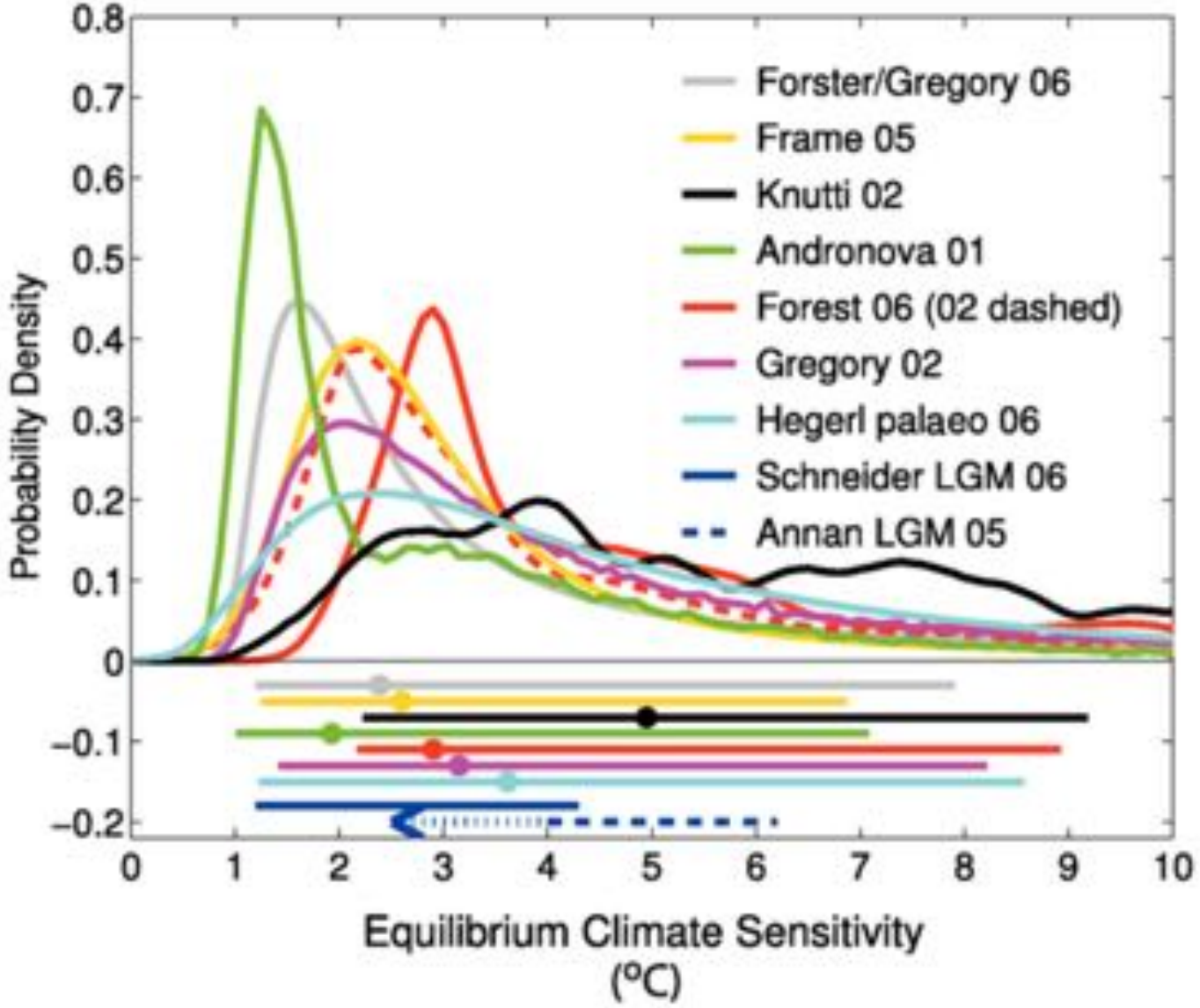
Integrals

*loving, not scary

What do you get if you integrate over a probability *density* function?

A probability!

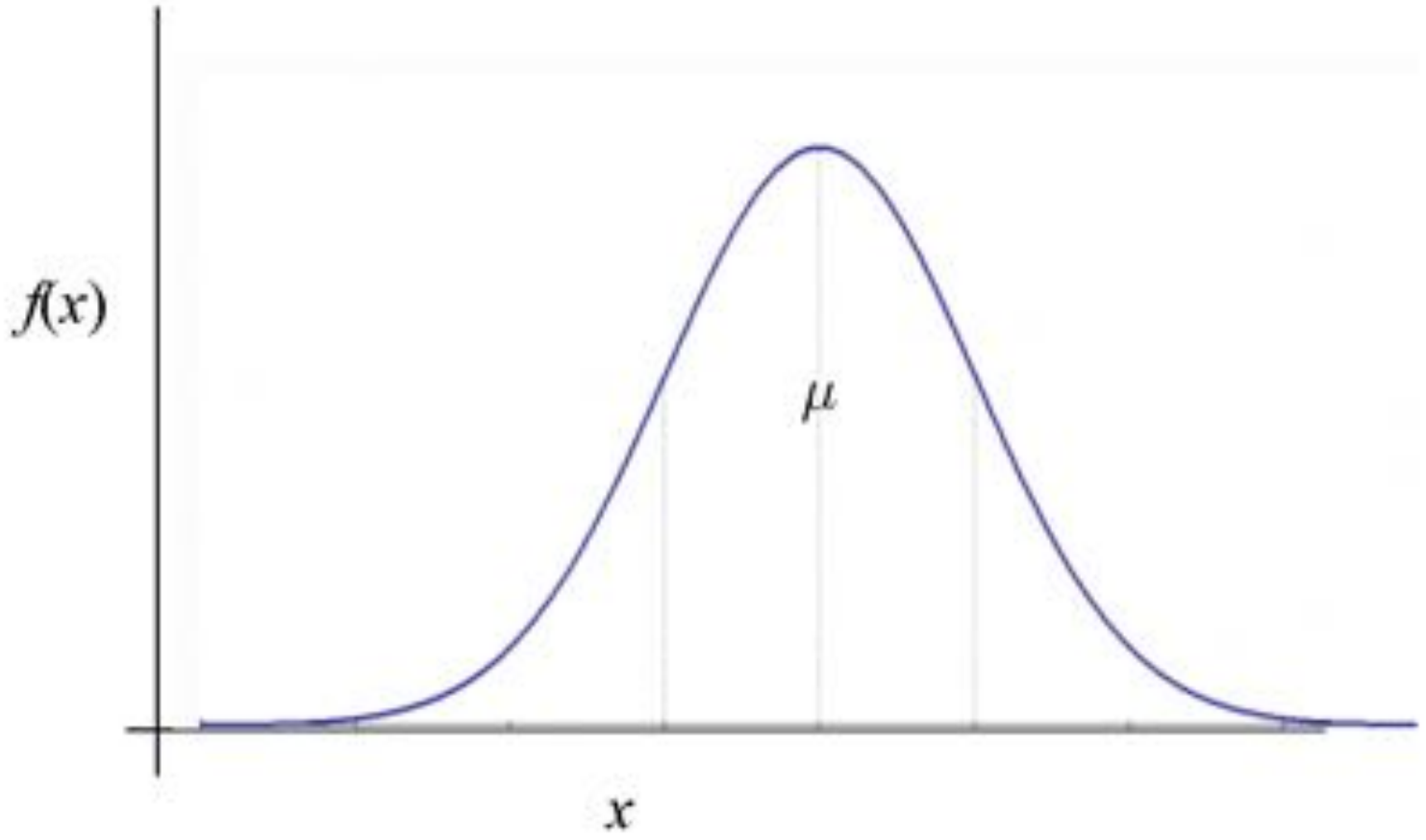
Climate Sensitivity



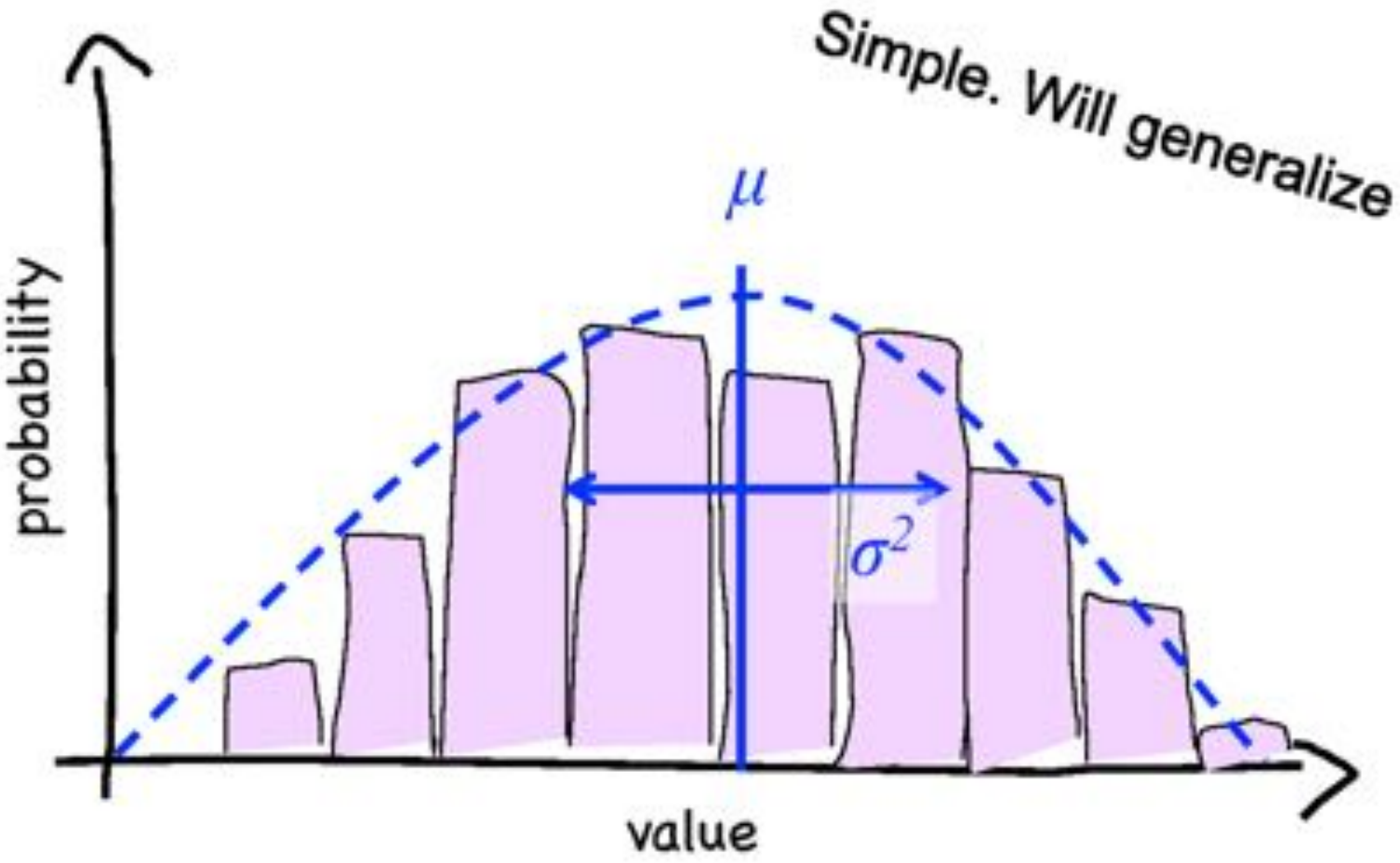
Probability Density Function

$$\mathcal{N}(\mu, \sigma^2)$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



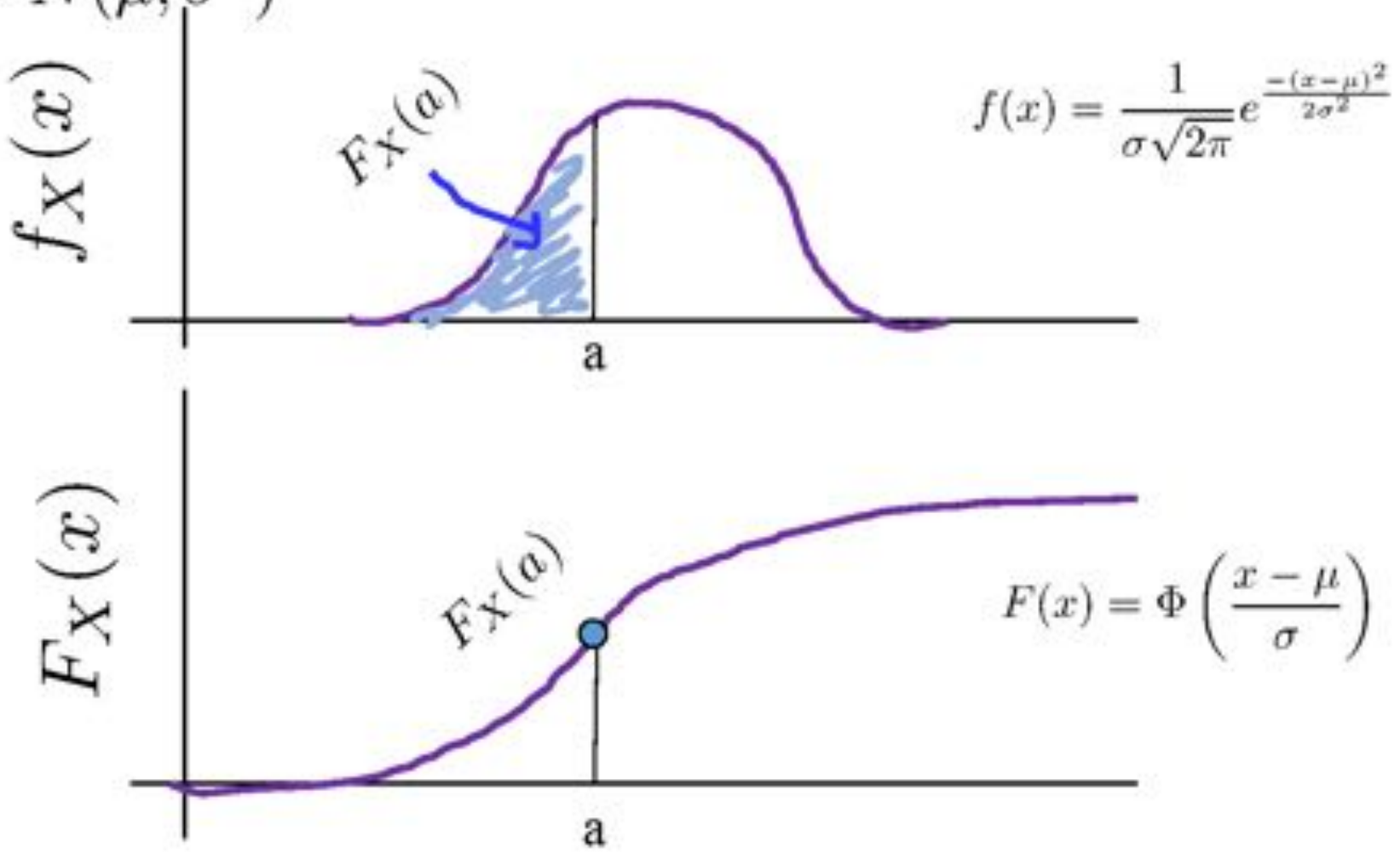
Simplicity is Humble



* A Gaussian maximizes entropy for a given mean and variance

PDF and CDF of a Normal

$X \sim N(\mu, \sigma^2)$



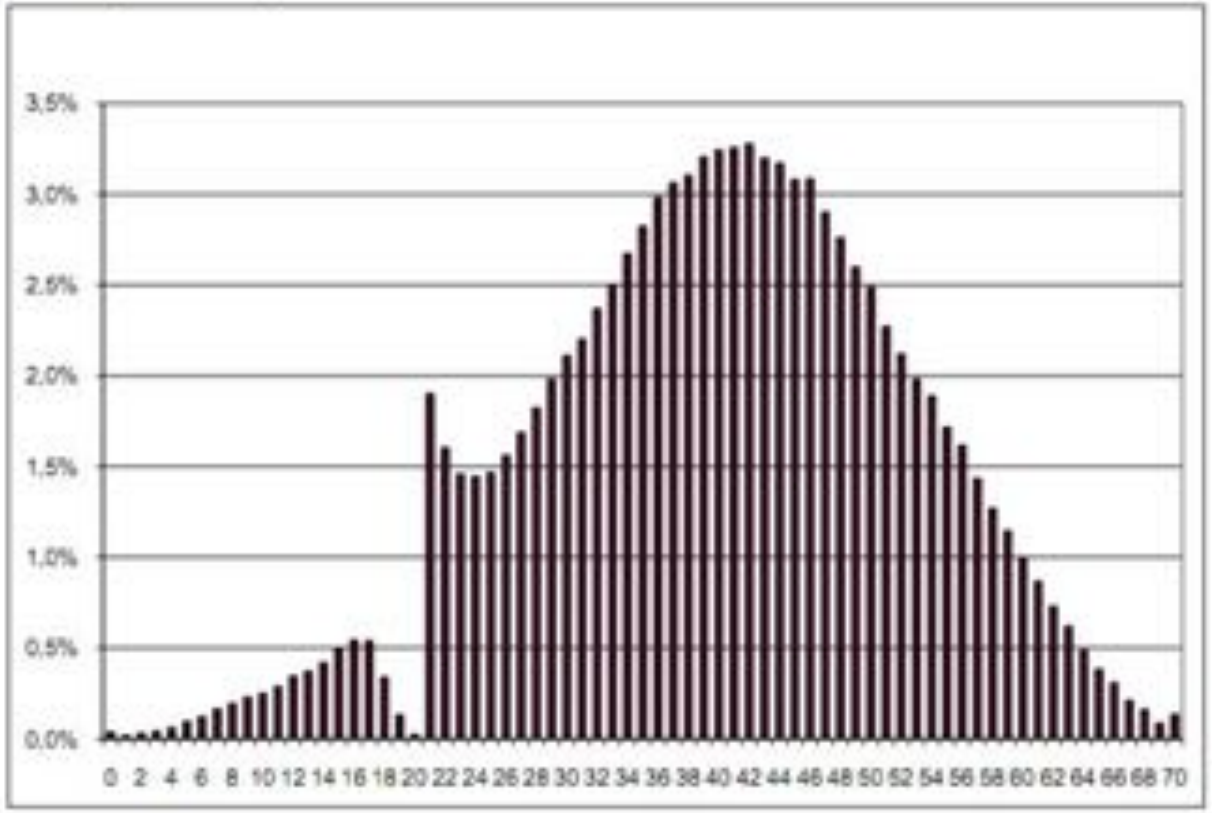
A CDF is the integral from –infinity to x of the PDF

Altruism?

Scores for a standardized test that students in Poland are required to pass before moving on in school

See if you can guess the minimum score to pass the test.

2.1. Poziom podstawowy



Wykres 1. Rozkład wyników na poziomie podstawowym

http://freakonomics.com/2011/07/07/another-case-of-teacher-cheating-or-is-it-just-altruism/comment-page-2/

Will the Warriors Win?

What is the probability that the Warriors beat the Blazers?

How do you model zero sum games?

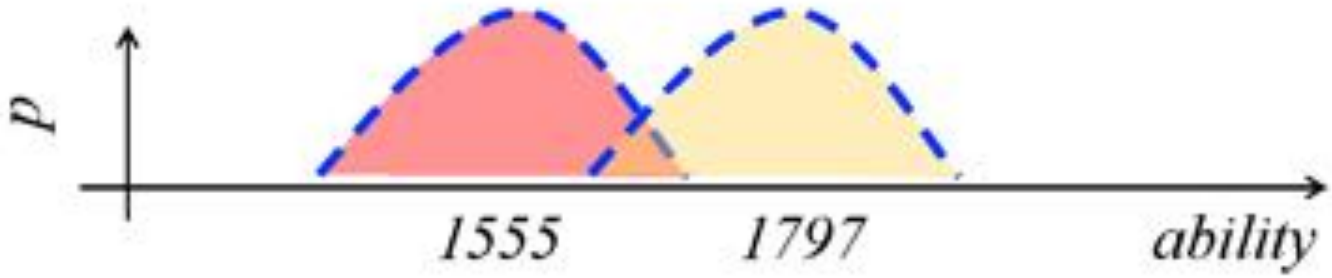
ELO Ratings

How it works:

- Each team has an "ELO" score S, calculated based on their past performance.
- Each game, the team has ability $A \sim N(S, 200^2)$
- The team with the higher sampled ability wins.

Arpad Elo

$A_B \sim \mathcal{N}(1555, 200^2)$ $A_W \sim \mathcal{N}(1797, 200^2)$



$$P(\text{Warriors win}) = P(A_W > A_B)$$

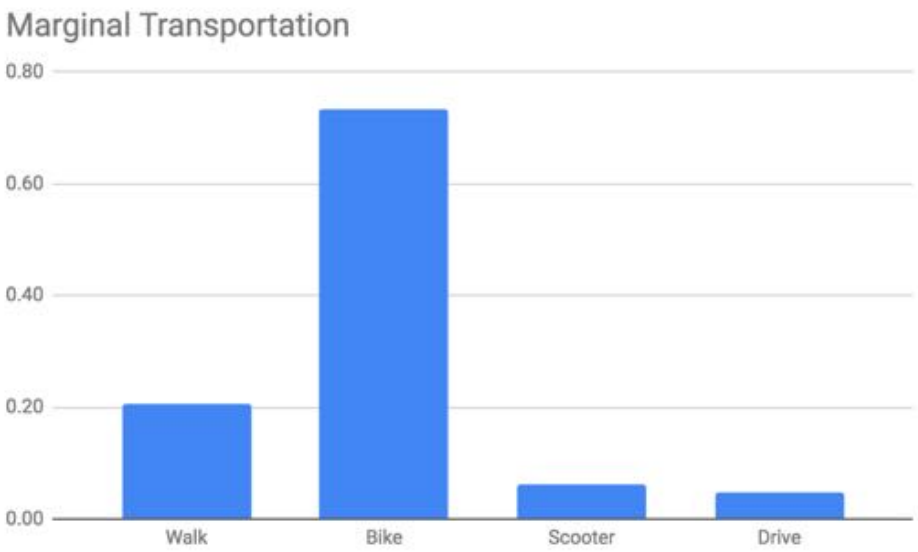

Joint Distributions

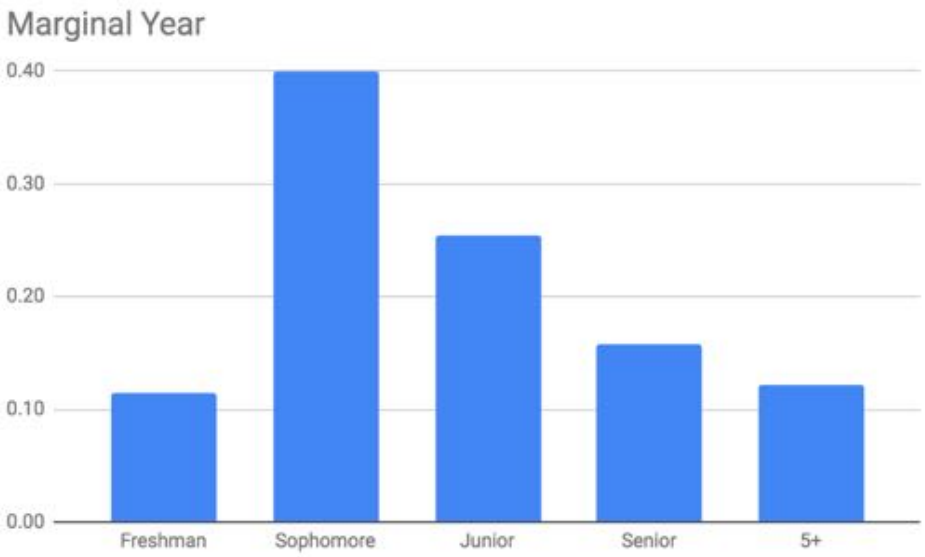
Joint Distributions

Go to this URL: https://goo.gl/Jh3Eu4

Joint Probability Table

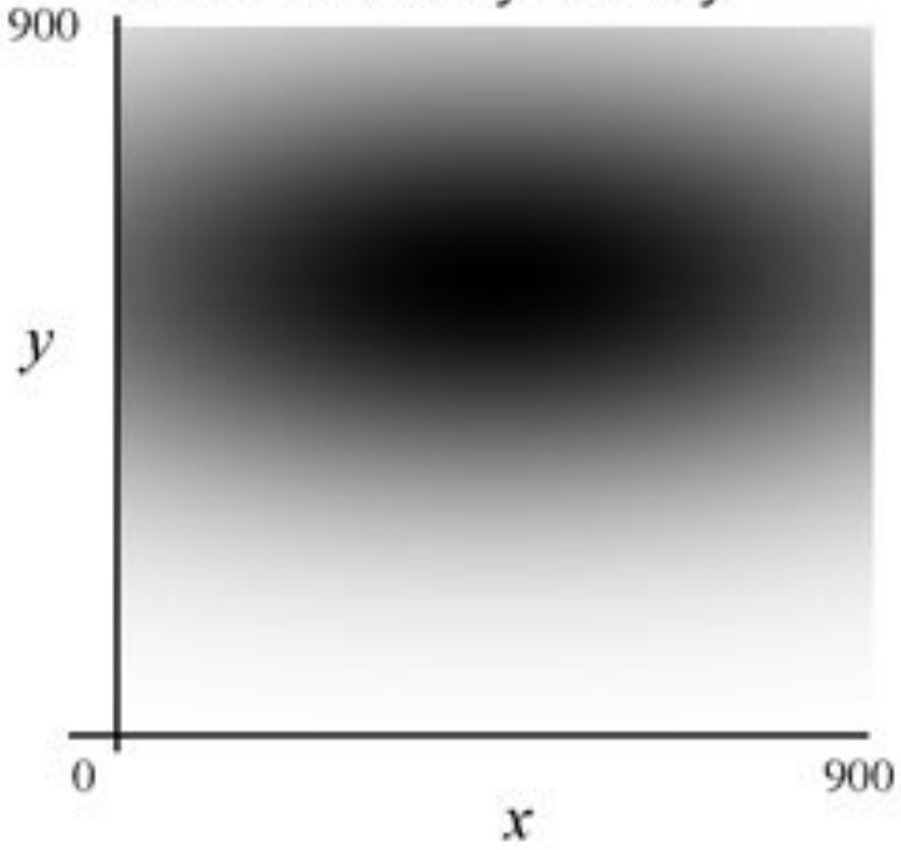
	Walk	Bike	Scooter	Drive	**Marginal Year**
Freshman	0.04	0.04	0.01	0.03	0.12
Sophomore	0.03	0.34	0.03	0.00	0.40
Junior	0.04	0.21	0.01	0.00	0.25
Senior	0.07	0.08	0.01	0.00	0.16
5+	0.04	0.07	0.00	0.02	0.12
Marginal Mode	0.21	0.73	0.06	0.05	





Joint Dart Distribution

Dart Probability Density



Dart Results

Joint Dart Distribution

rt Results

P(hit within R pixels of center)?

Dart x location

Multinomial

Example document:
"Pay for Viagra with a credit-card. Viagra is great. So are credit-cards. Risk free Viagra. Click for free."
$n = 18$

It's a Multinomial!

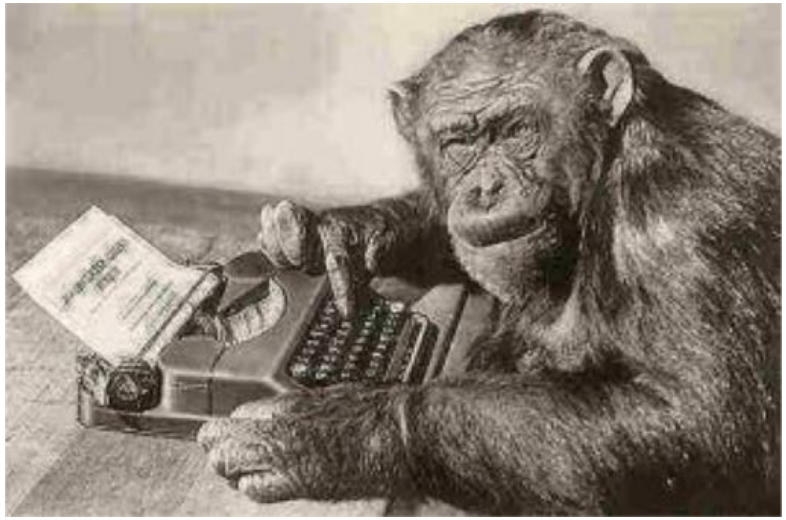
$$P\left(\begin{array}{l}\text{Viagra} = 2\\ \text{Free} = 2\\ \text{Risk} = 1\\ \text{Credit-card: } 2\\ \ldots\\ \text{For} = 2\end{array} \middle| \text{spam}\right) = \frac{n!}{2!2!\ldots 2!} p_{\text{viagra}}^2 p_{\text{free}}^2 \cdots p_{\text{for}}^2$$

Probability of seeing this document | spam

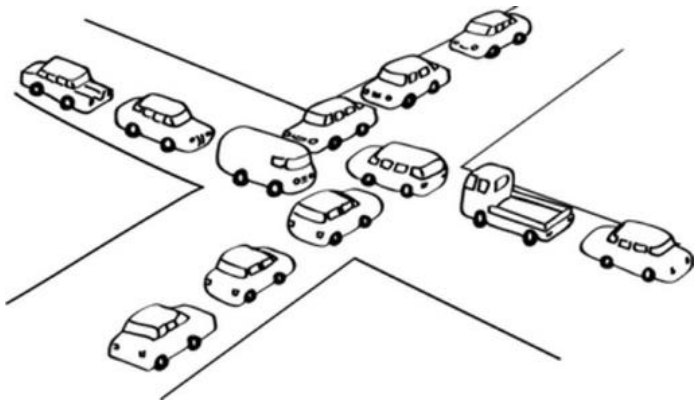
The probability of a word in spam email being viagra

Midterm (part 1)

1



2



3

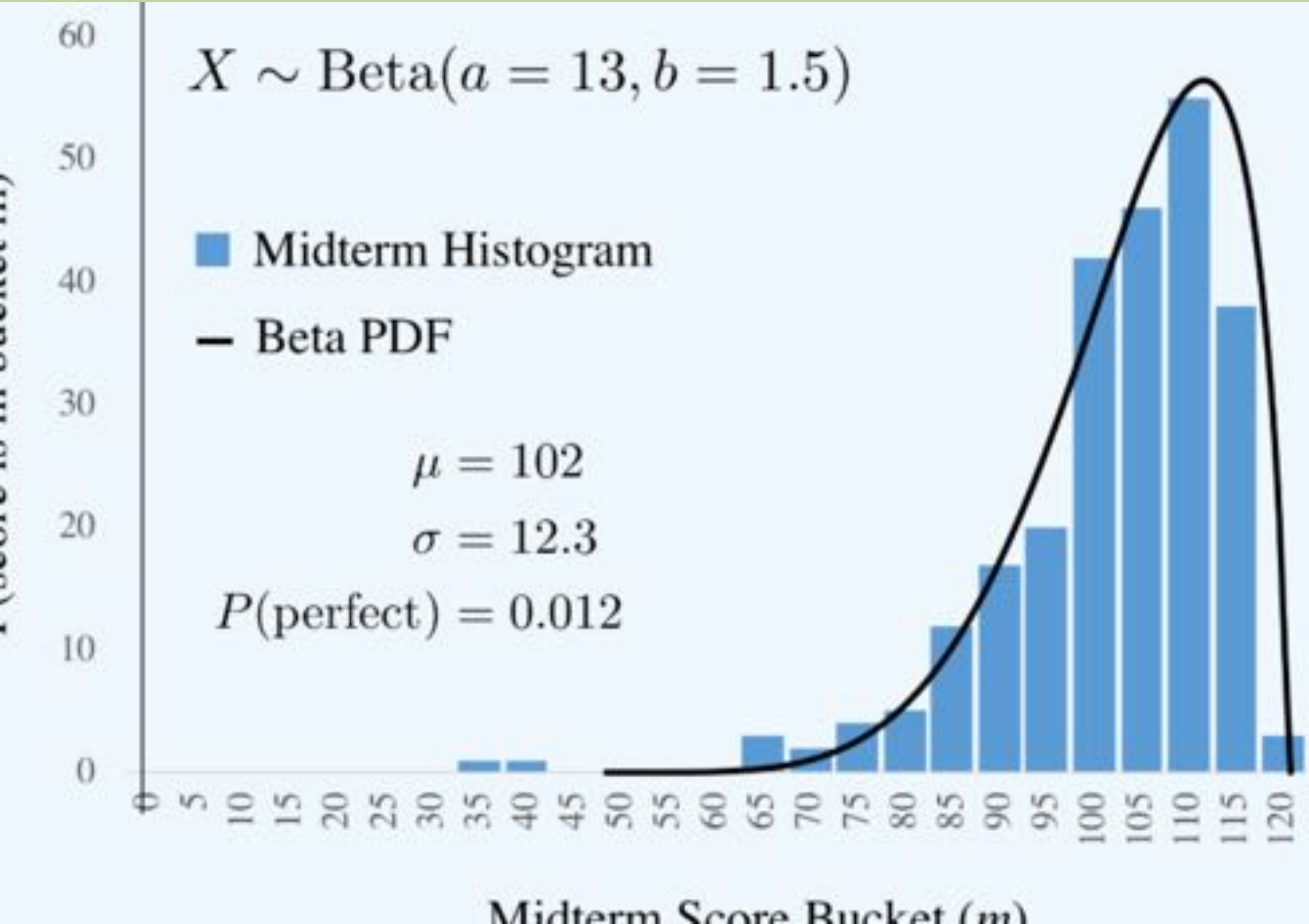
Midterm (part 2)

4

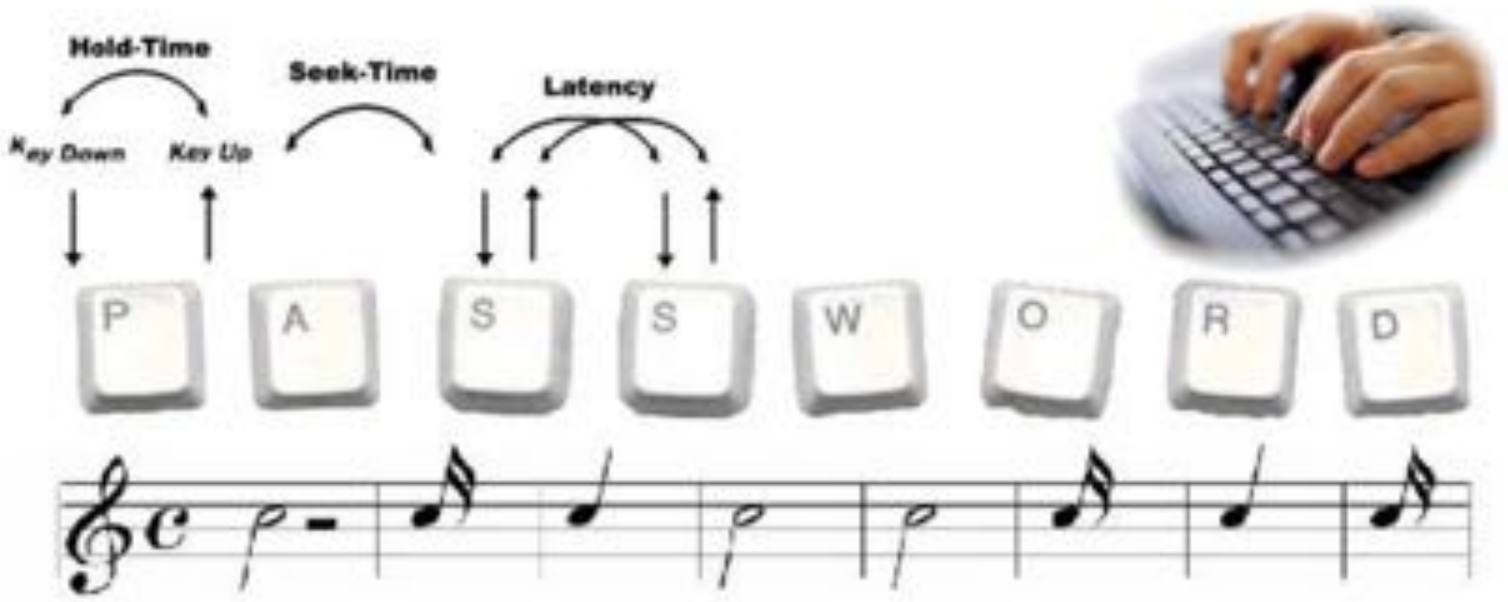
5

6

Midterm (part 3)

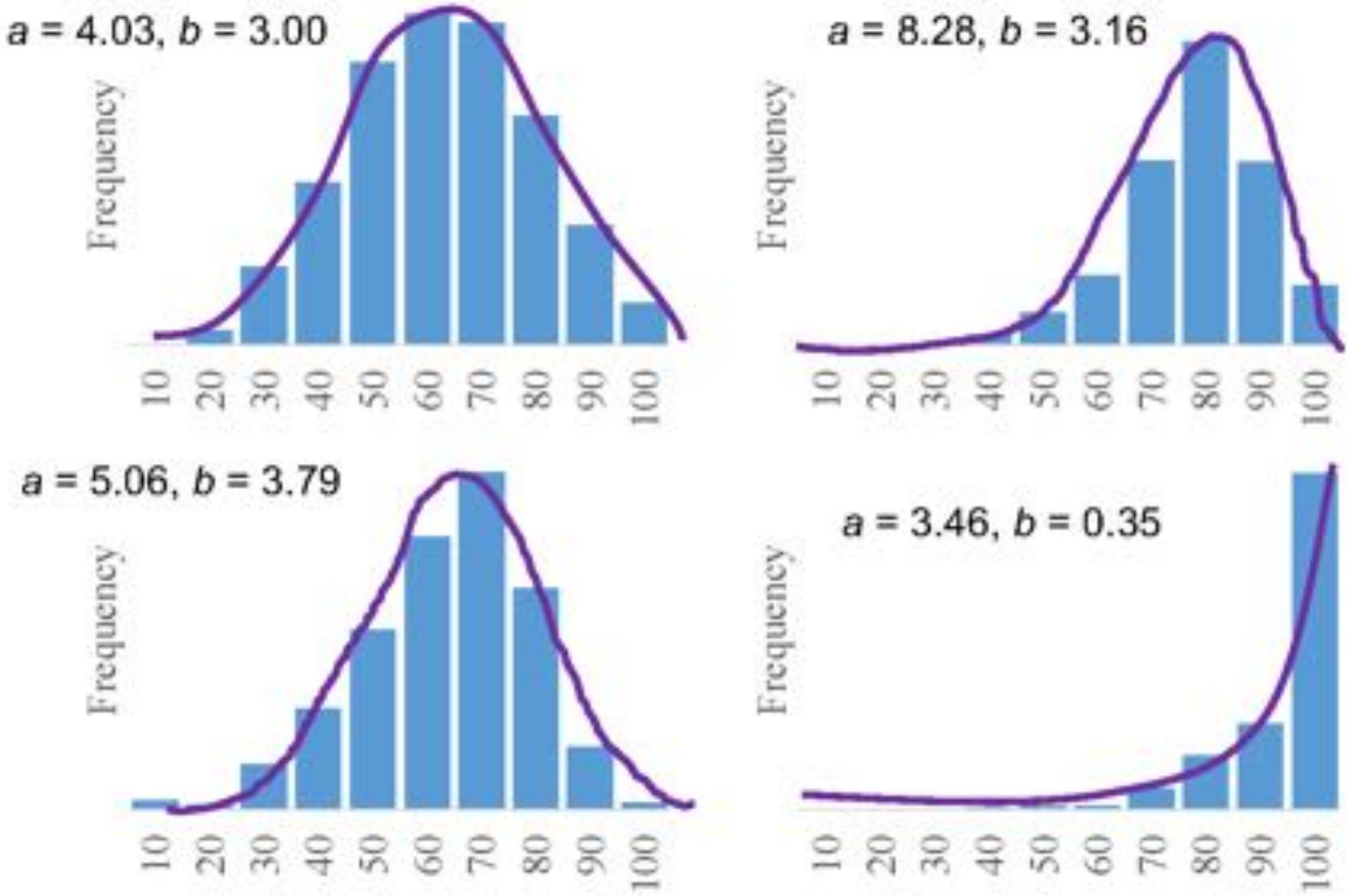


Biometric Keystroke



Enchanted Die

Assignment Grades



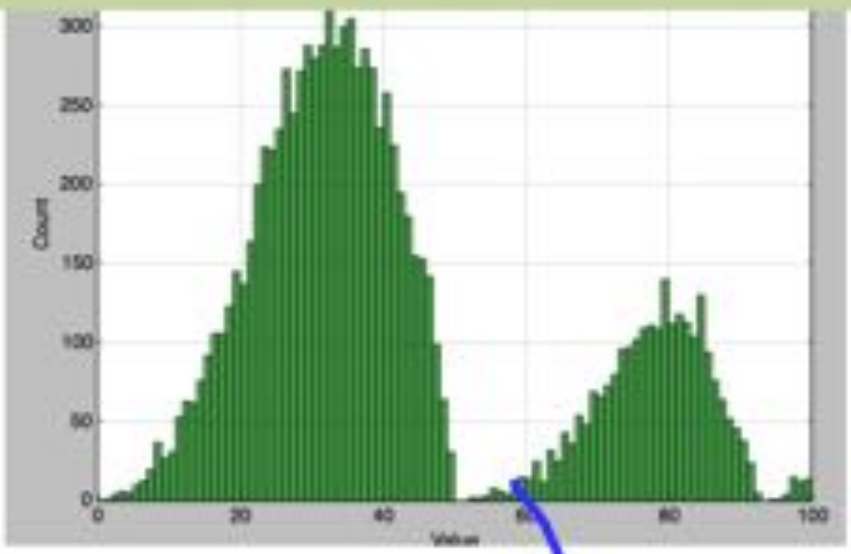
We have 2055 assignment distributions from grade scope

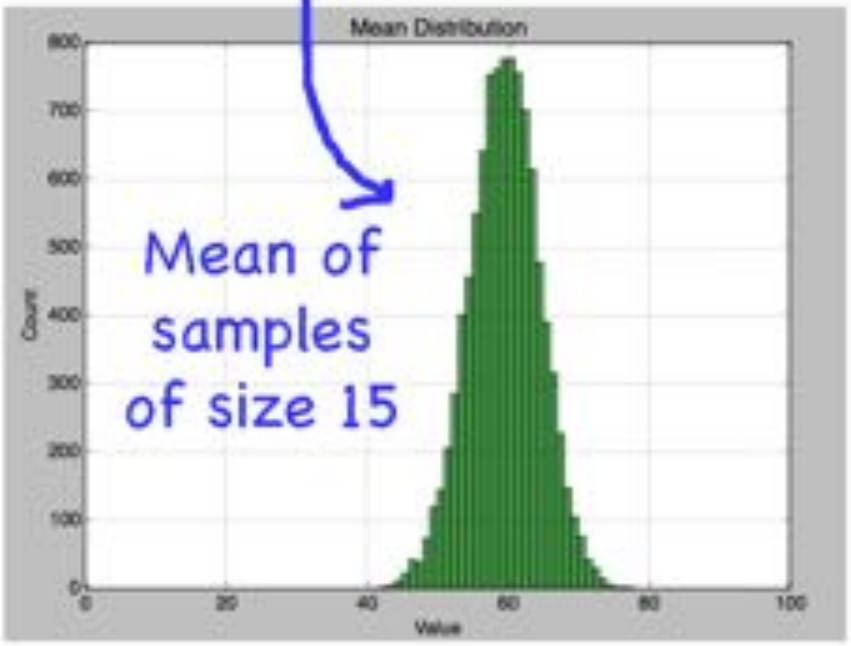
Mystery: Why is Binomial Normal?

Mystery: Why is the sum of IID uniforms normal?

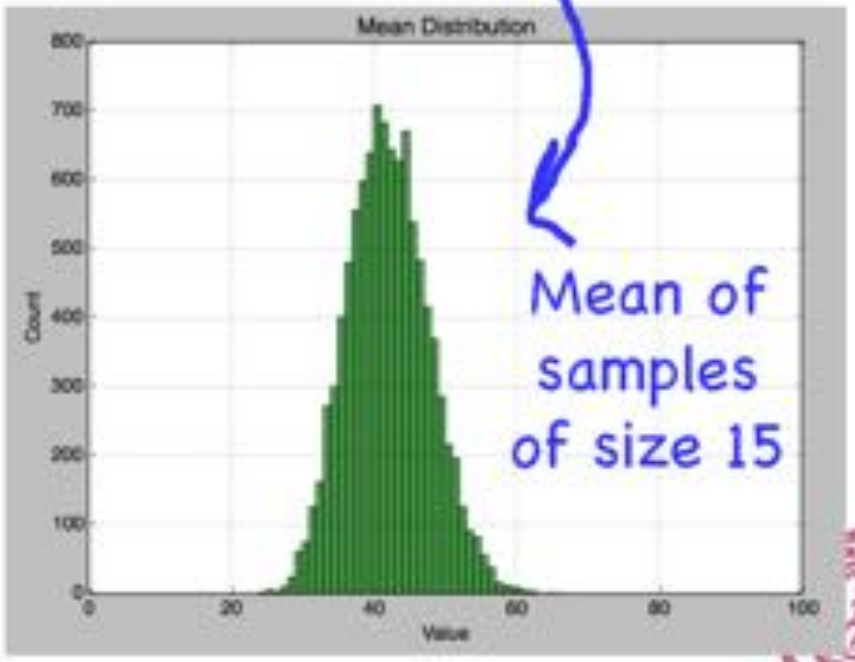
Mystery: Why is the mean of IID vars normal?

C.L.T. Explains This



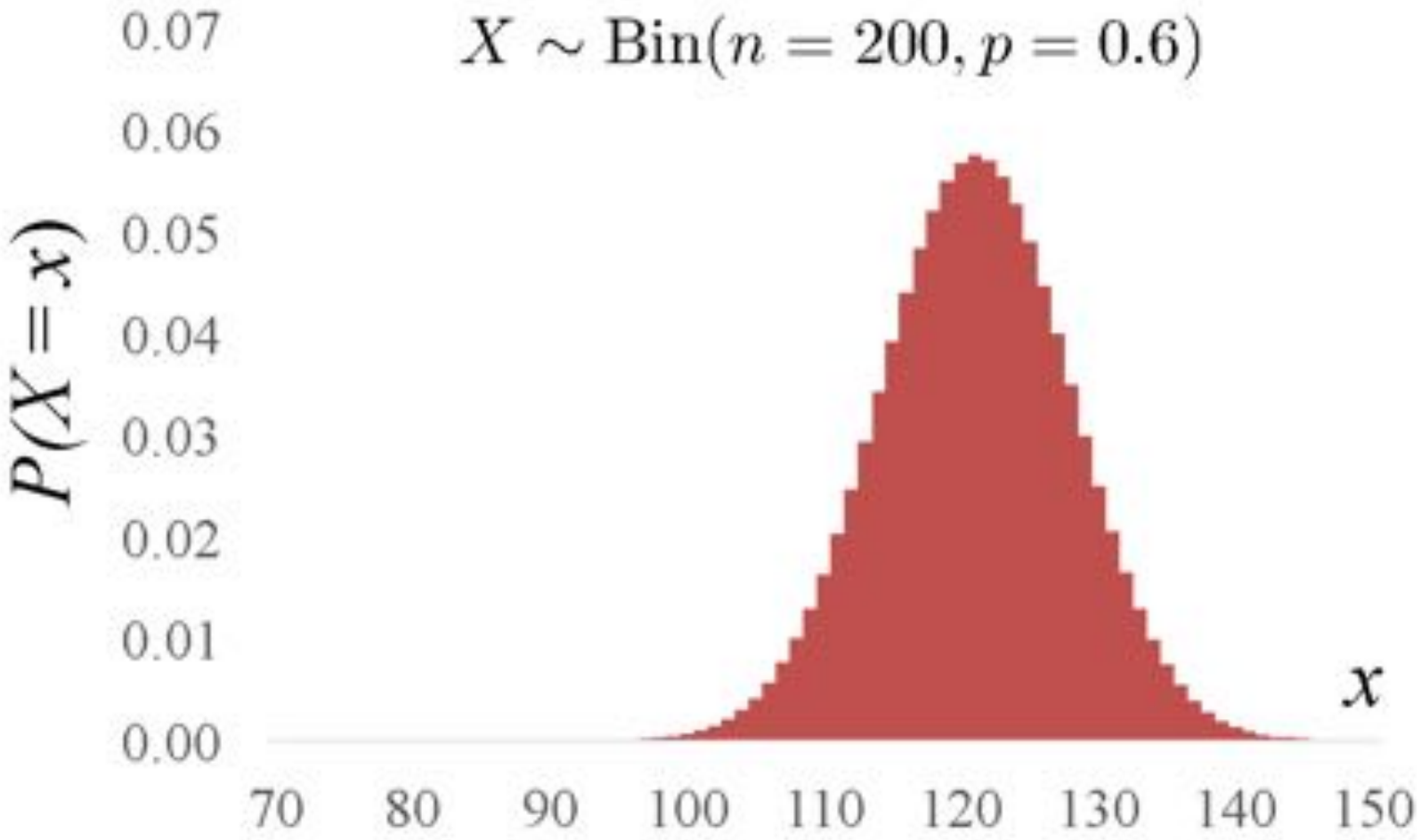


Mean of samples of size 15



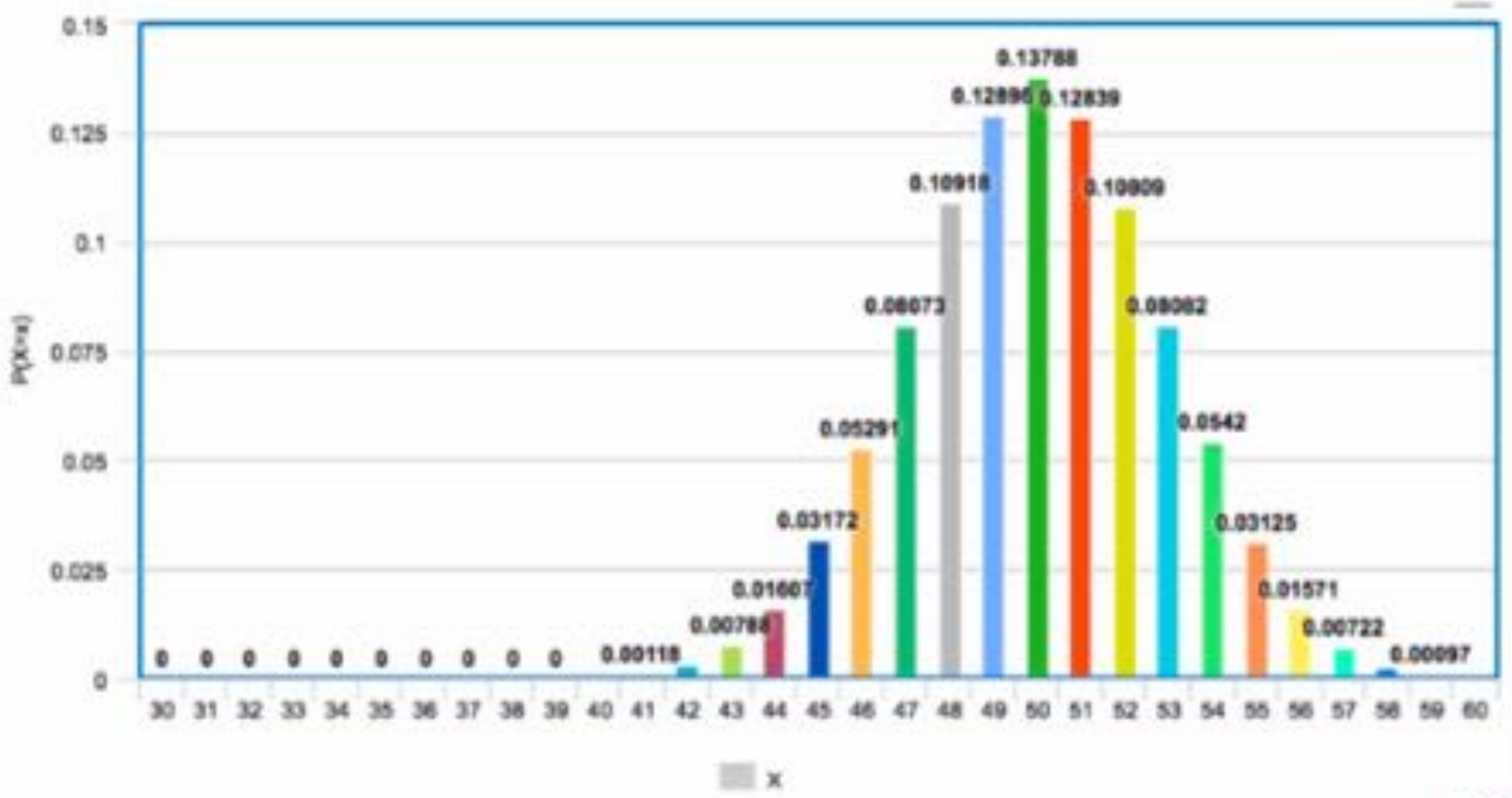
Mean of samples of size 15

C.L.T. Explains This



C.L.T. Explains This

Problem set 5: What is the sum of IID uniforms?



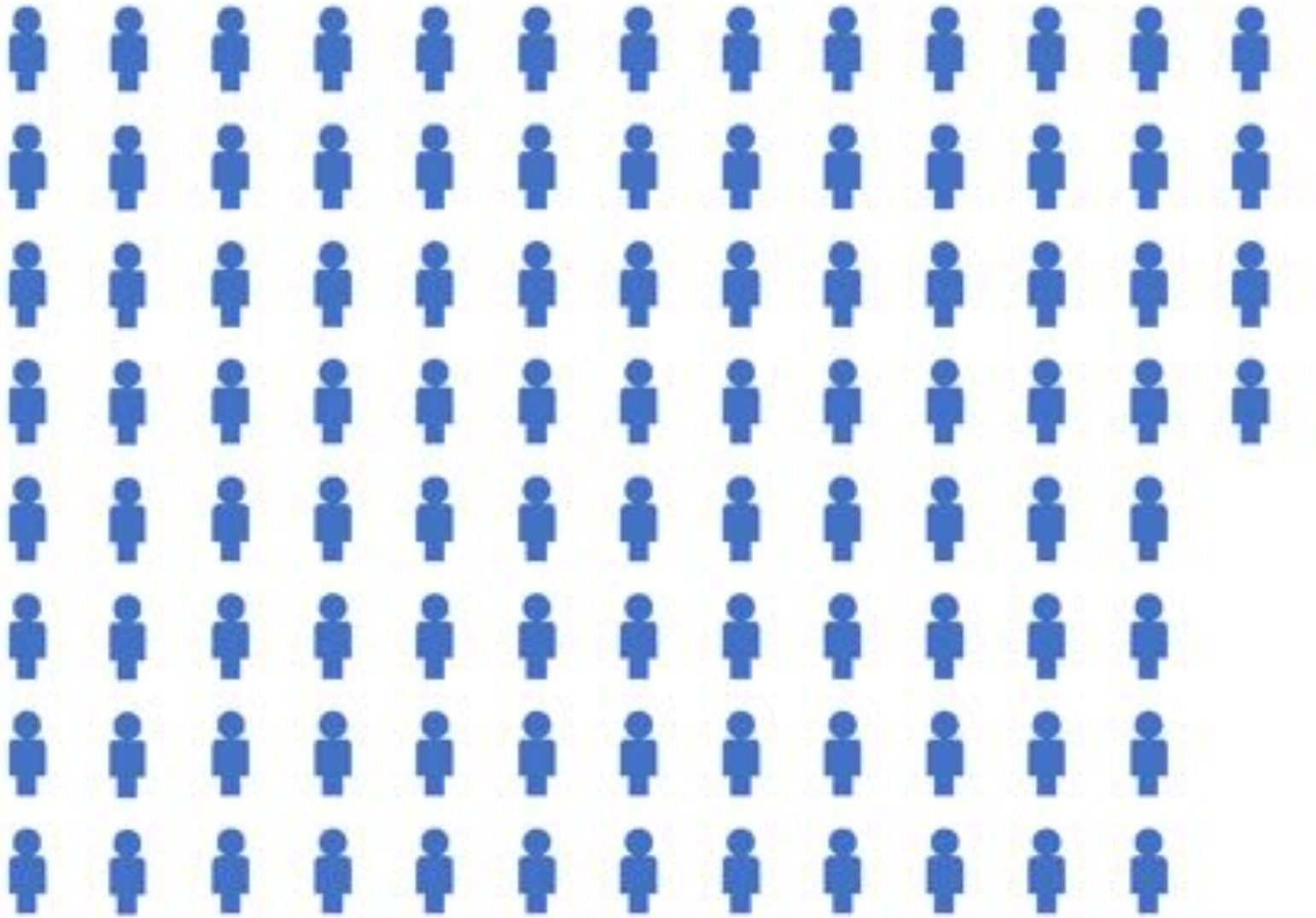
Machine Learning Example

- You want to know the true mean and variance of happiness in Buthan
  - But you can't ask everyone.
  - Randomly sample 200 people.
  - Your data looks like this:

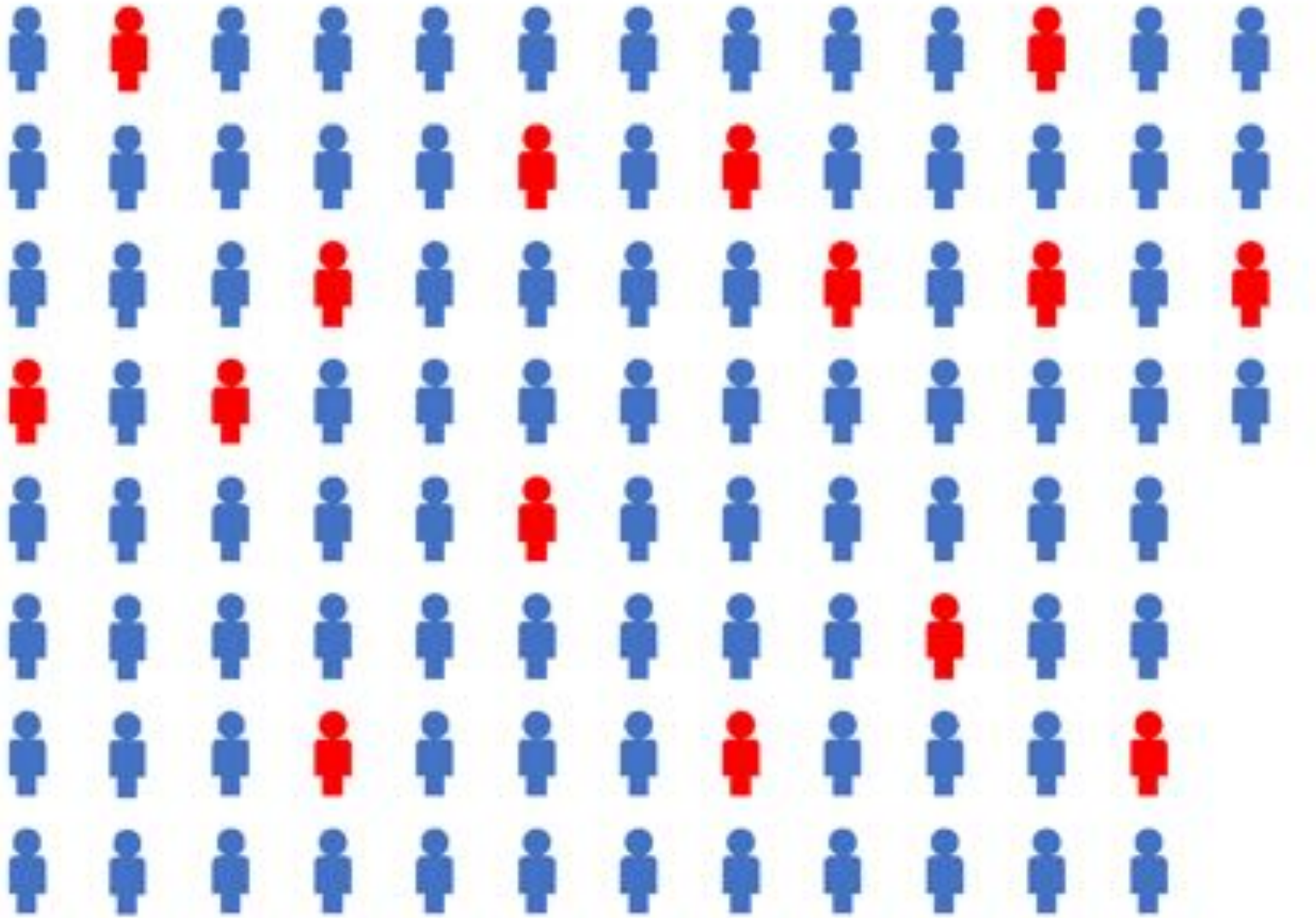
Happiness = {72, 85, 79, 91, 68, … , 71}

  - The mean of all of those numbers is 83. Is that the true average happiness of Bhutanese people?

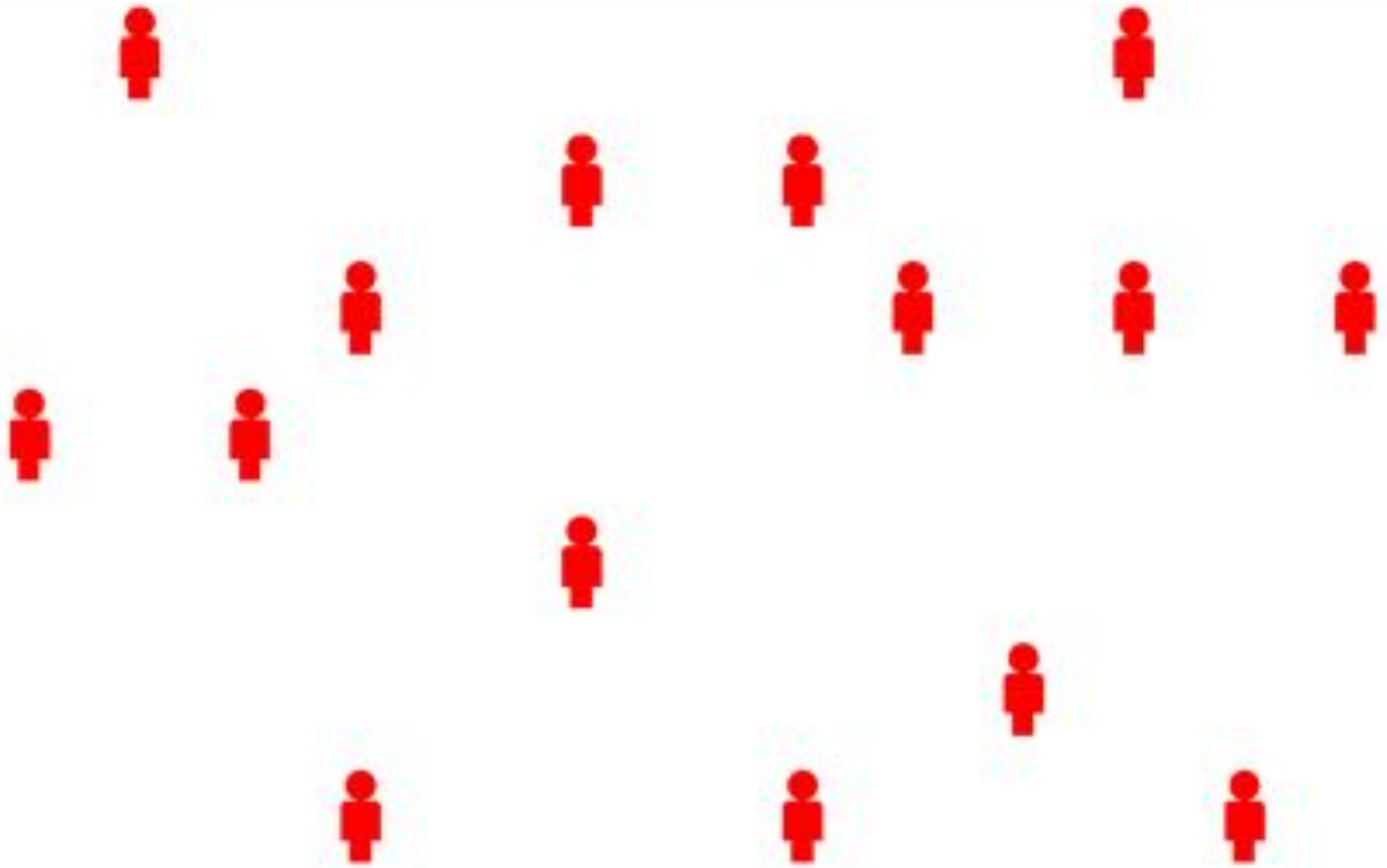
Population



Sample



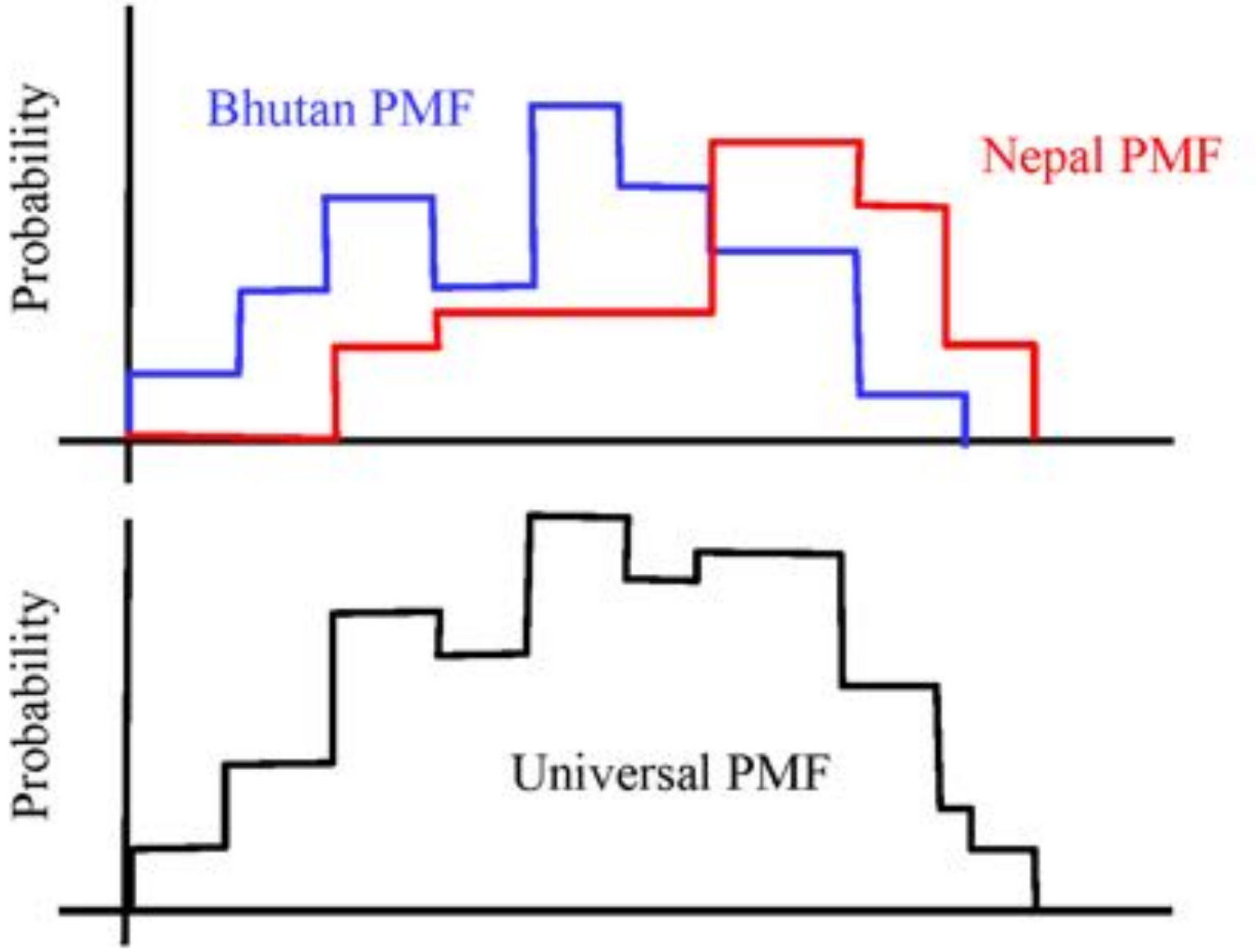
Sample



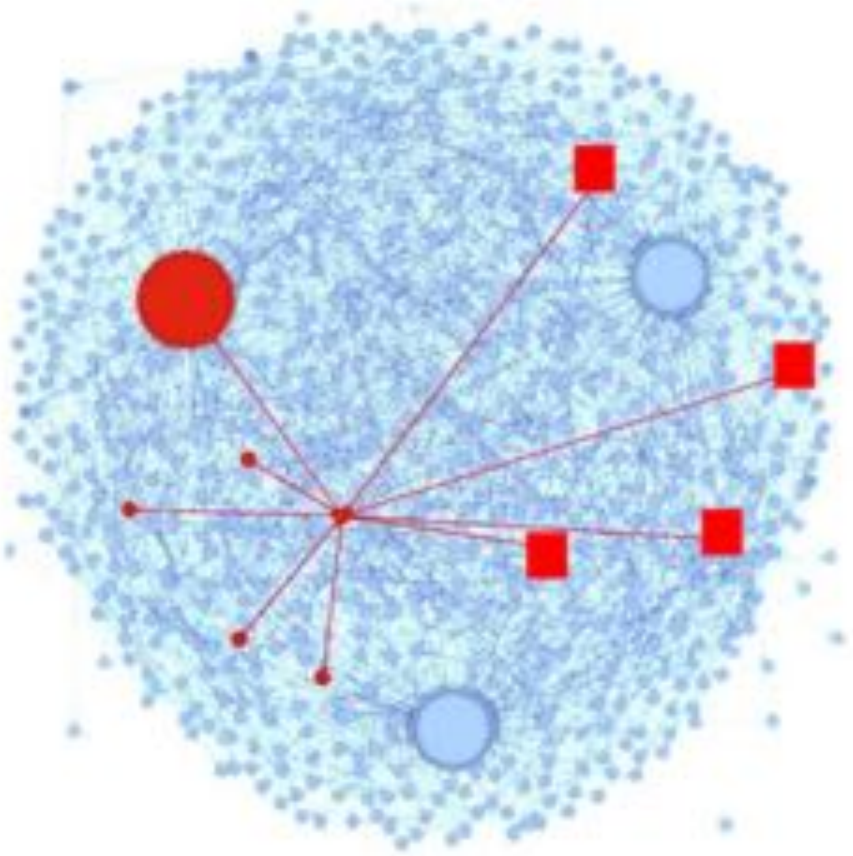
Collect one (or more) numbers from each person


Bootstrap

Universal Sample



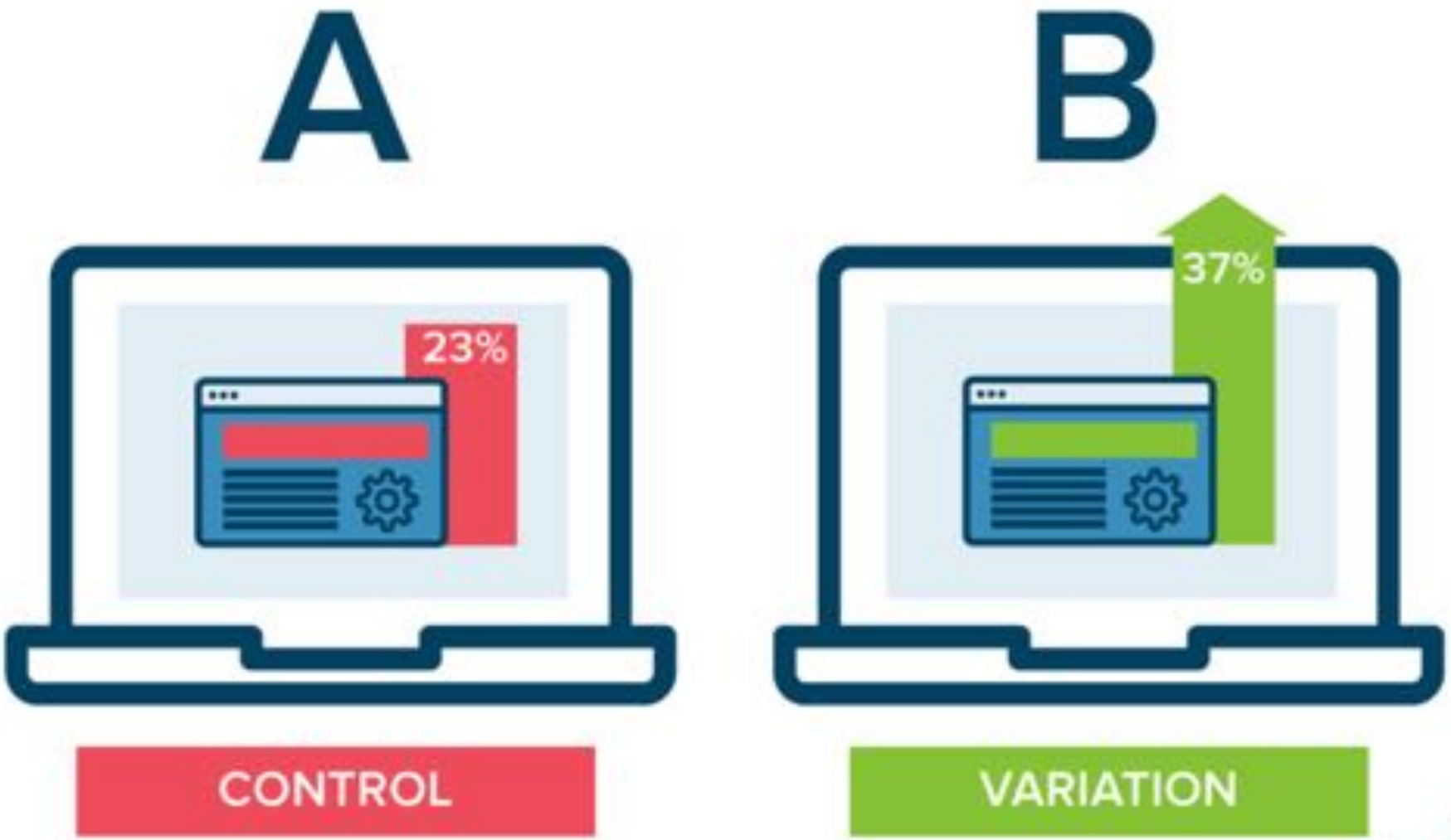
Peer Grading



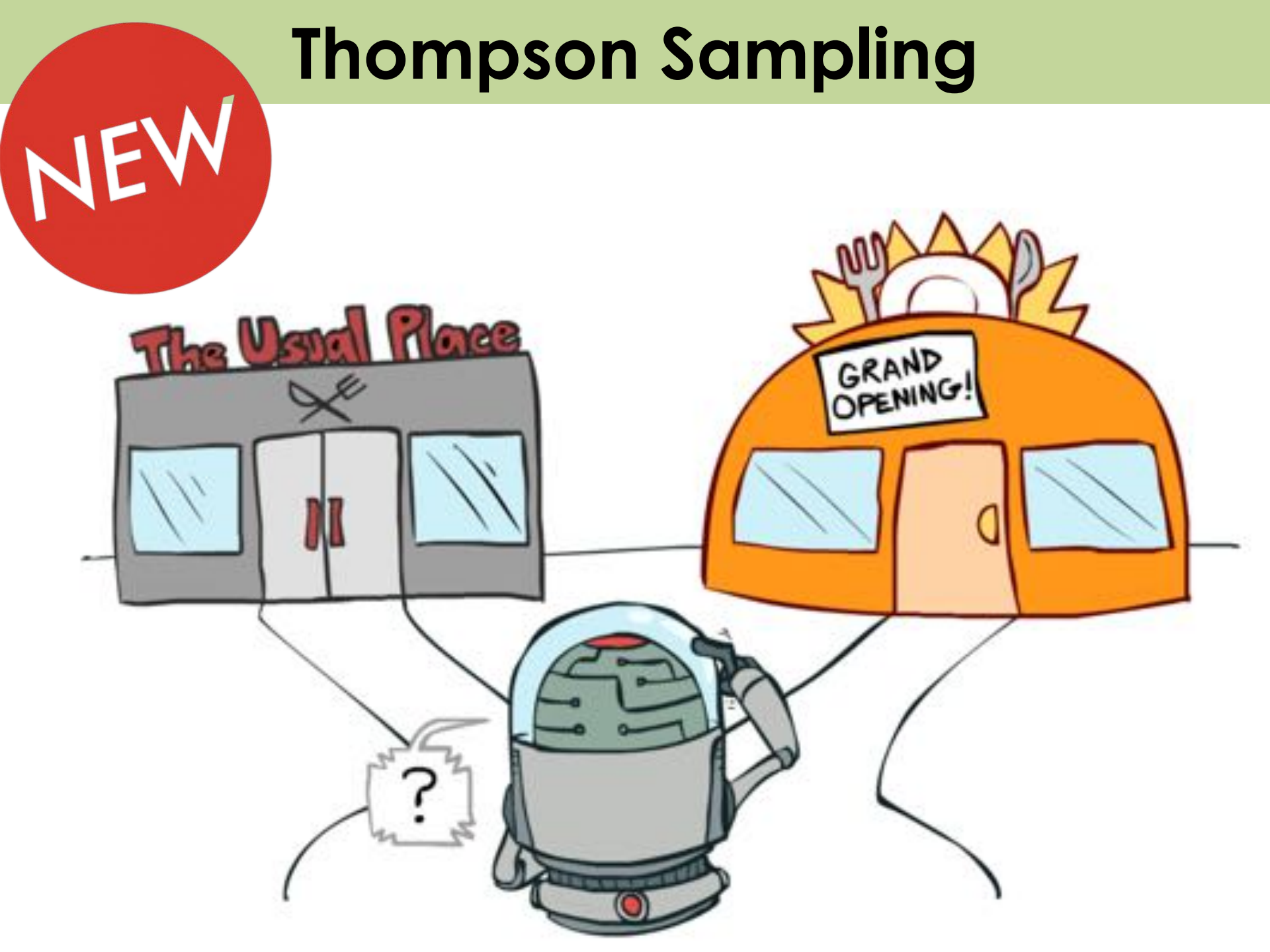
Peer Grading on Coursera HCI.

31,067 peer grades for 3,607 students.

A/B Testing

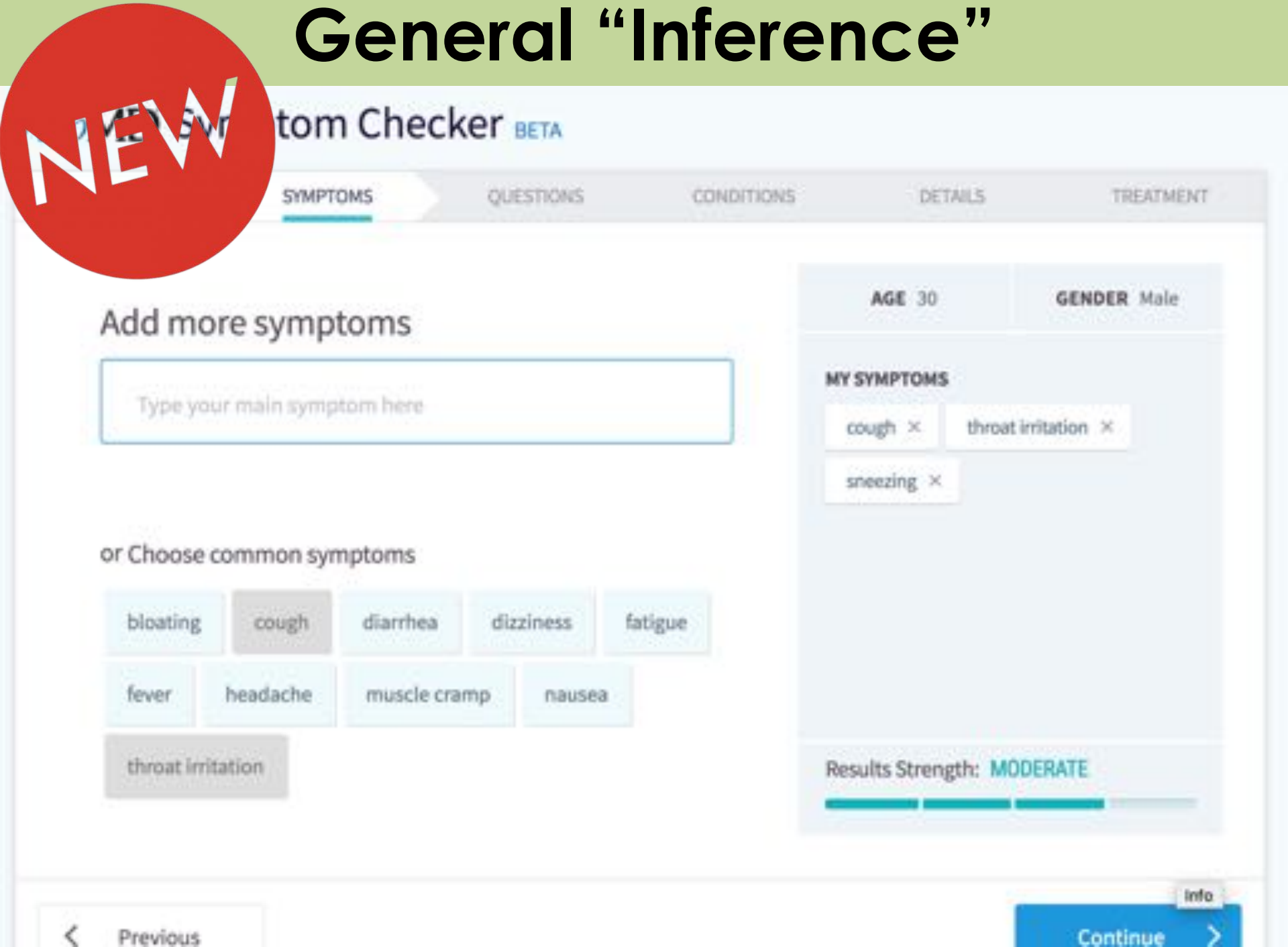


Thompson Sampling

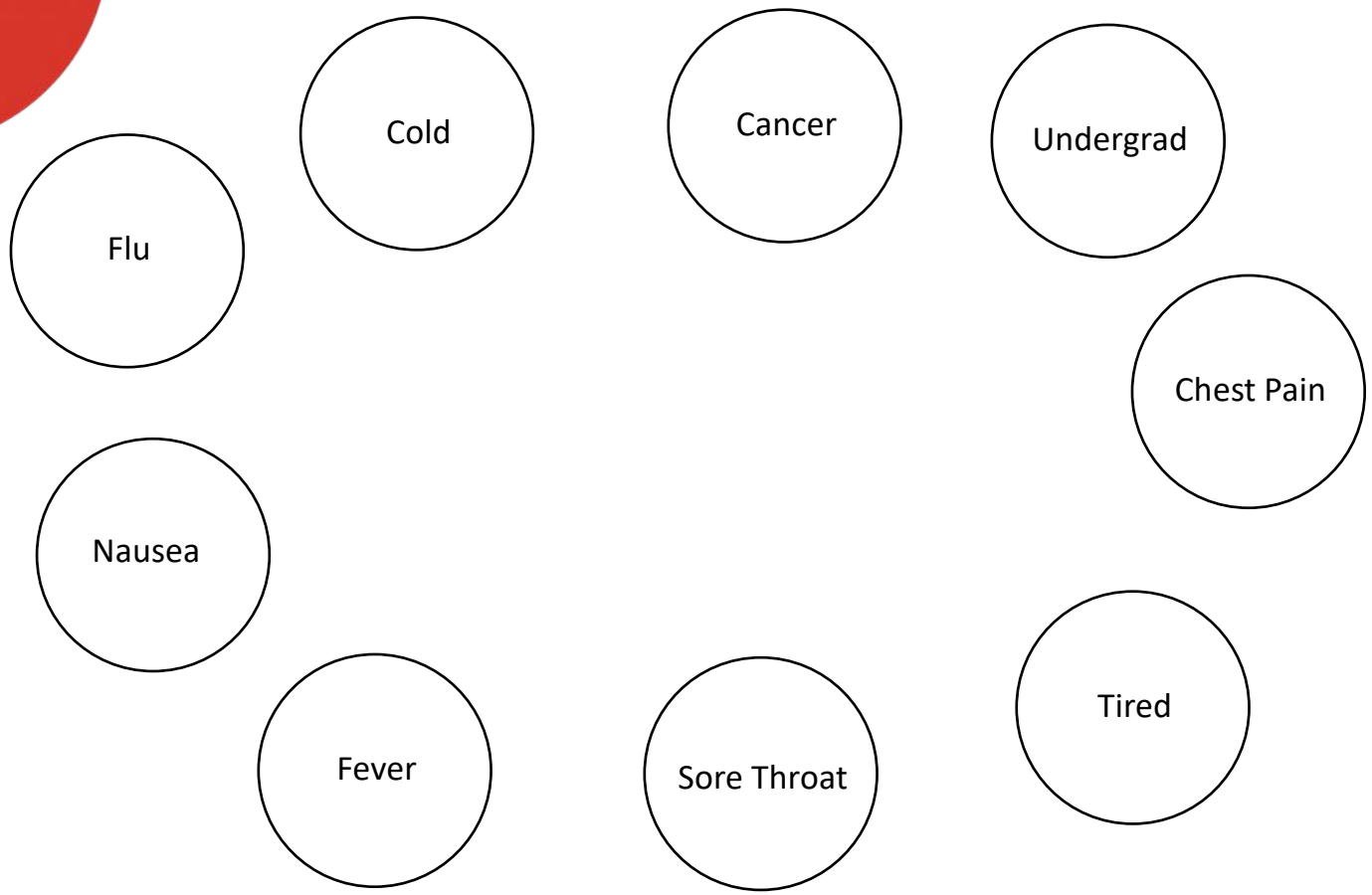


General “Inference”

General “Inference”

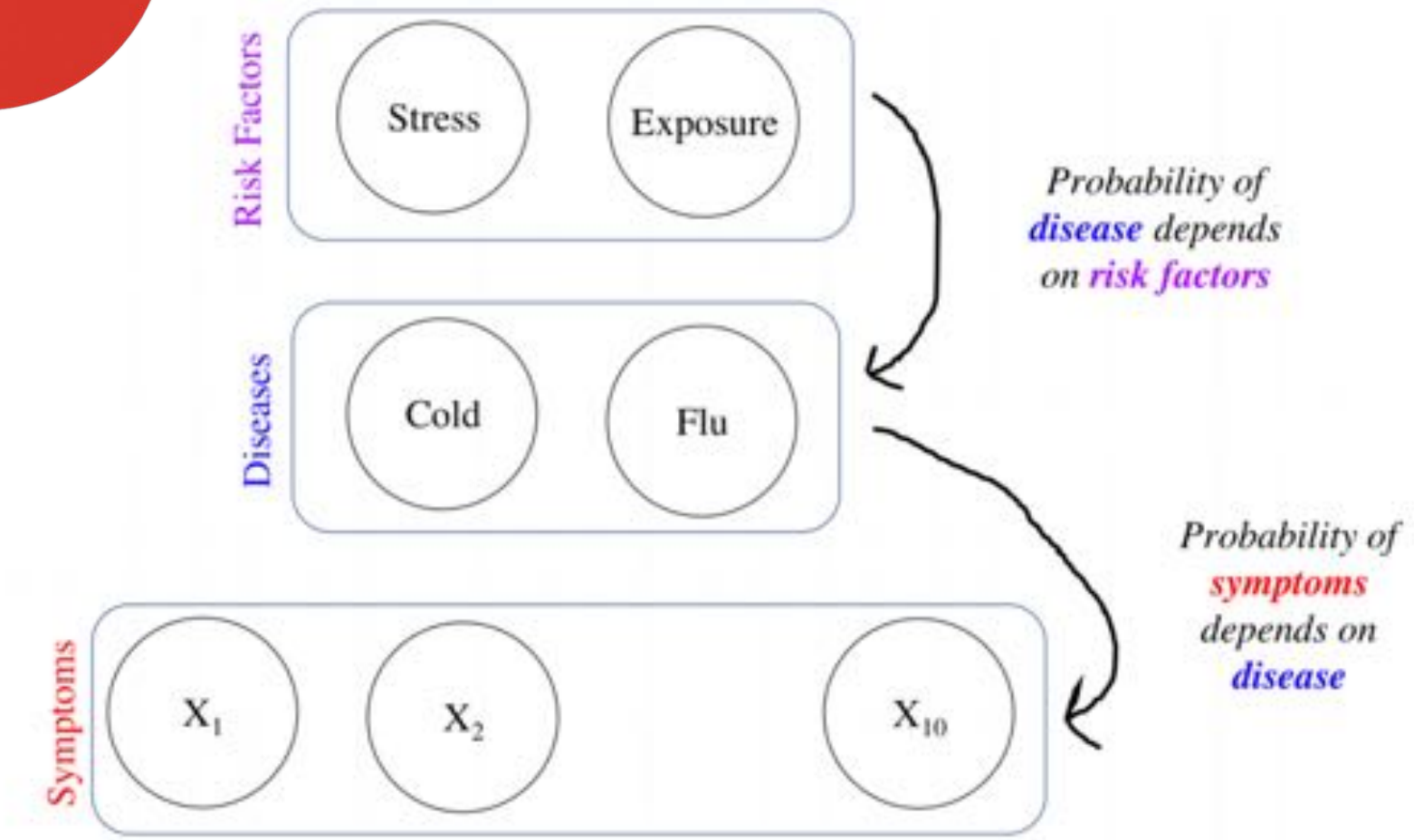


Lots of Random Vars?



Bayes Nets!

NEW



Alg #1: Joint Sampling


```
5  # Program: Joint Sa
6  # ------------------
7  # we can answer any
8  # with multivariate
9  # where conditioned
10 def main():
11     obs = getObserv
12     print 'Observat
13
14     samples = sampl
15     prob = probFluG
16     print 'Pr(Flu)
```

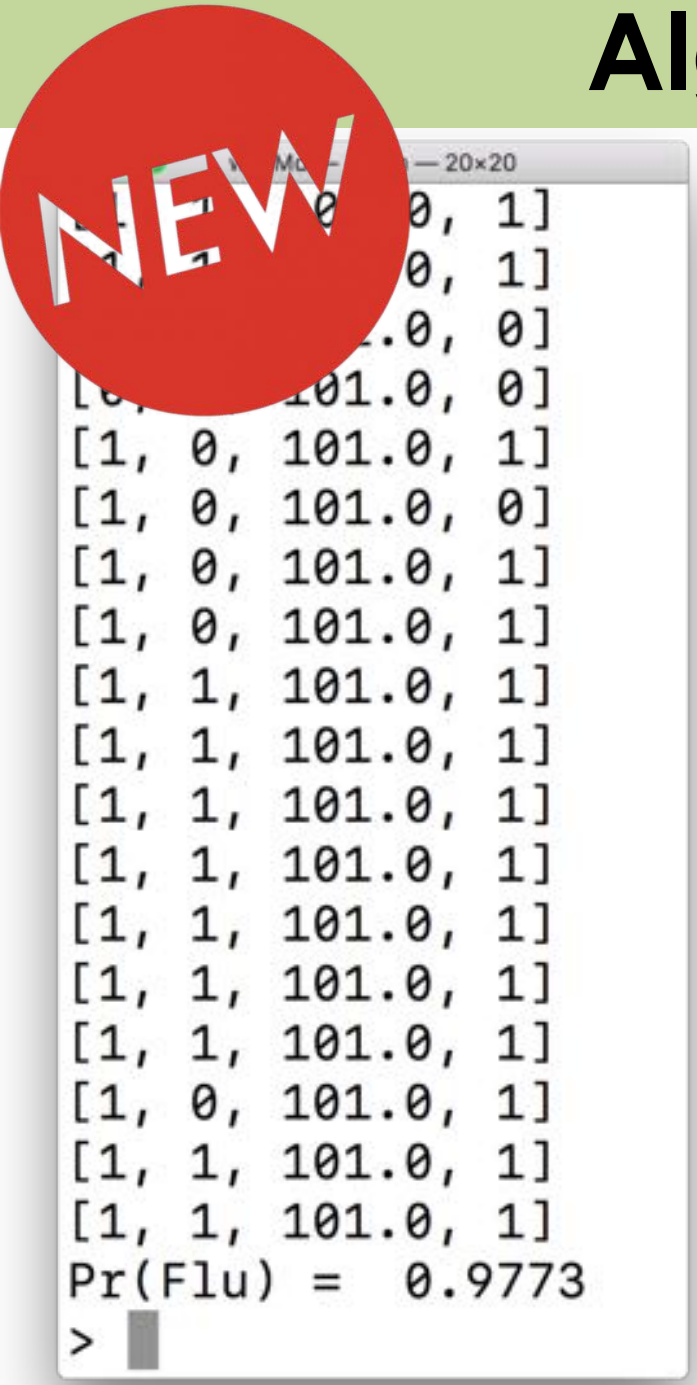
webMd — -bash — 38×22

```
[0, 0, 0, 0]
[0, 1, 0, 1]
[1, 0, 1, 0]
[1, 1, 1, 1]
[0, 1, 0, 1]
[0, 1, 0, 0]
[0, 0, 0, 0]
[0, 1, 1, 1]
[0, 1, 0, 0]
[0, 1, 0, 1]
[0, 1, 0, 0]
[0, 1, 0, 1]
[0, 1, 0, 1]
[0, 0, 0, 0]
[1, 1, 1, 1]
[0, 0, 0, 0]
[0, 0, 0, 0]
[1, 1, 1, 1]
[0, 1, 0, 0]
Observation =  [None, None, None, 1]
Pr(Flu | Obs) =  0.140635888502
>
```

Each one of these is one posterior sample:

[Flu, Ugrad, Fever, Tired]

Alg #2: MCMC

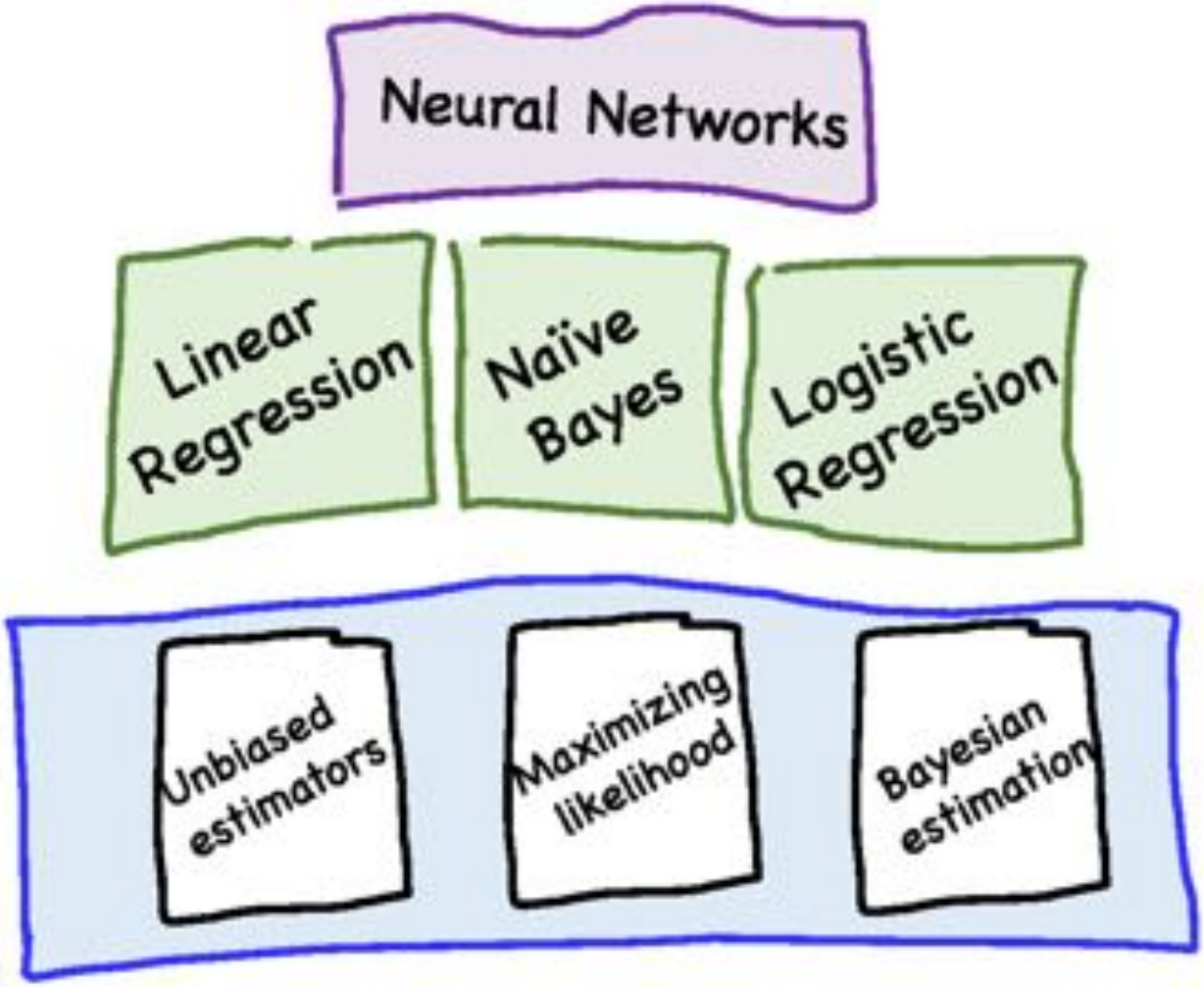


MCMC is a way to sample with conditioned variables fixed

Each one of these is one joint sample:

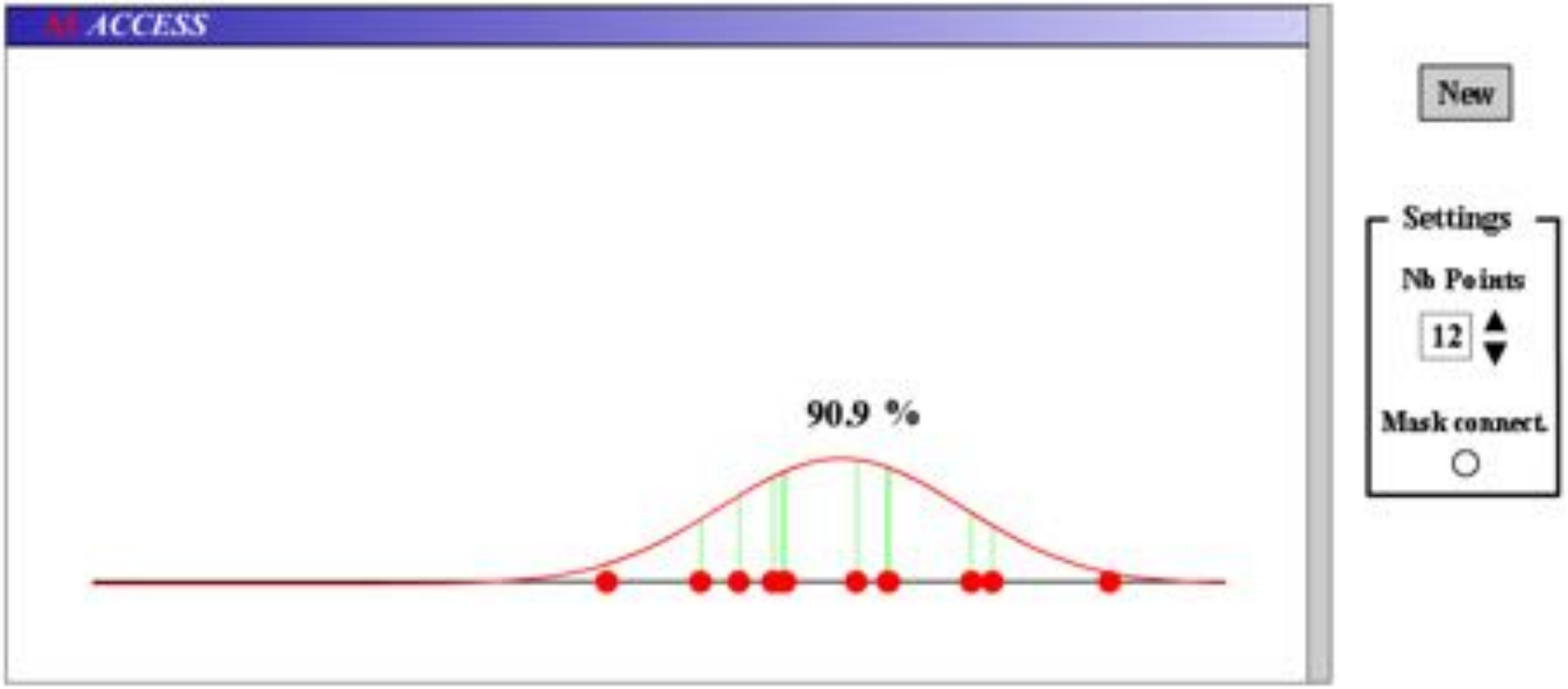
[Flu, Undergrad, Fever, Tired]

Towards Machine Learning

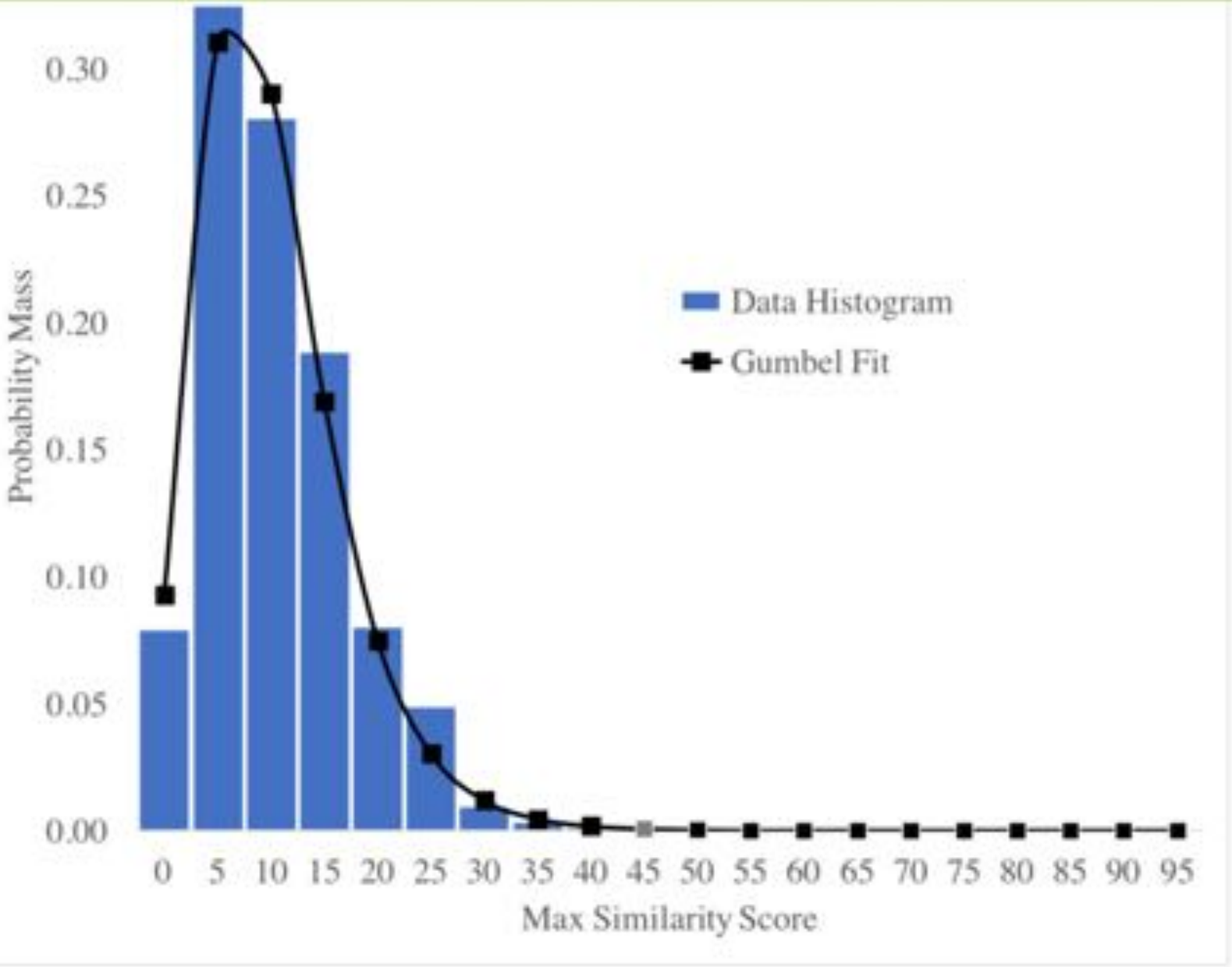


MLE: Likelihood of Data

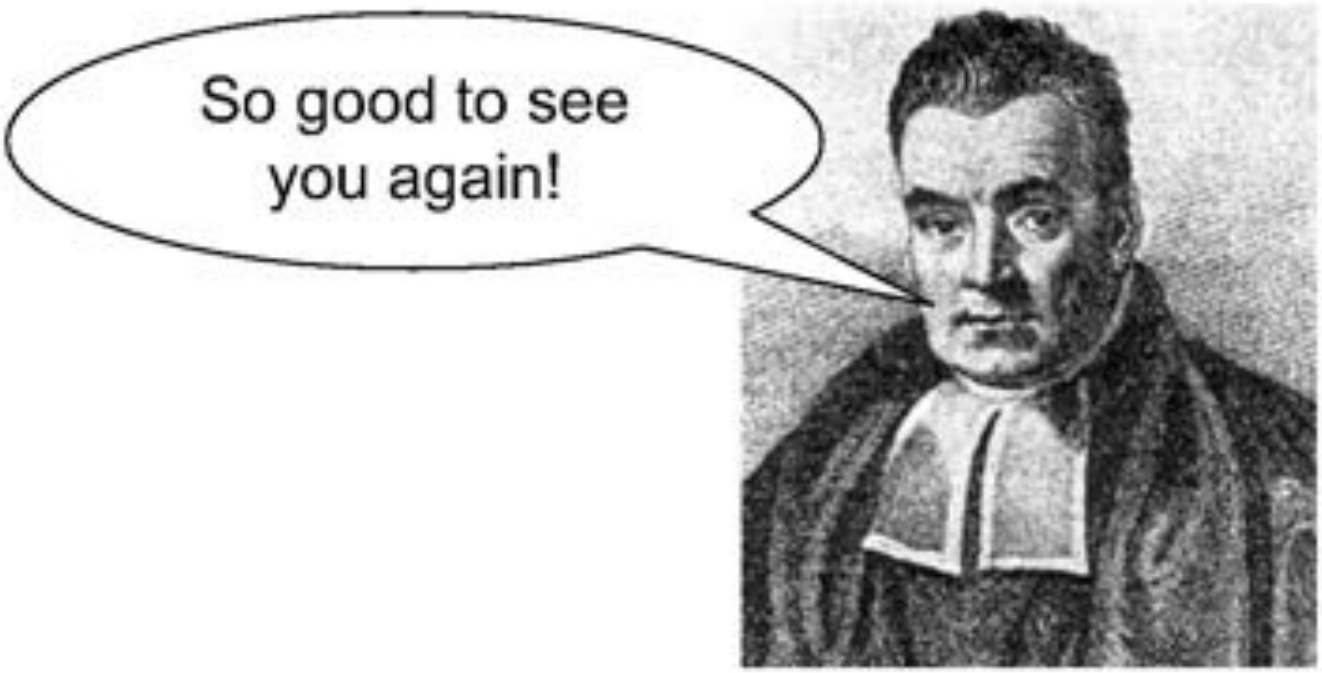
Likelihood of Data from a Normal



Gumbel Fit


0.30
0.25
0.20
0.15
0.10
0.05
0.00
Probability Mass
Data Histogram
Gumbel Fit
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Max Similarity Score

MAP: Most Probable Parameter



Machine Learning

Heart

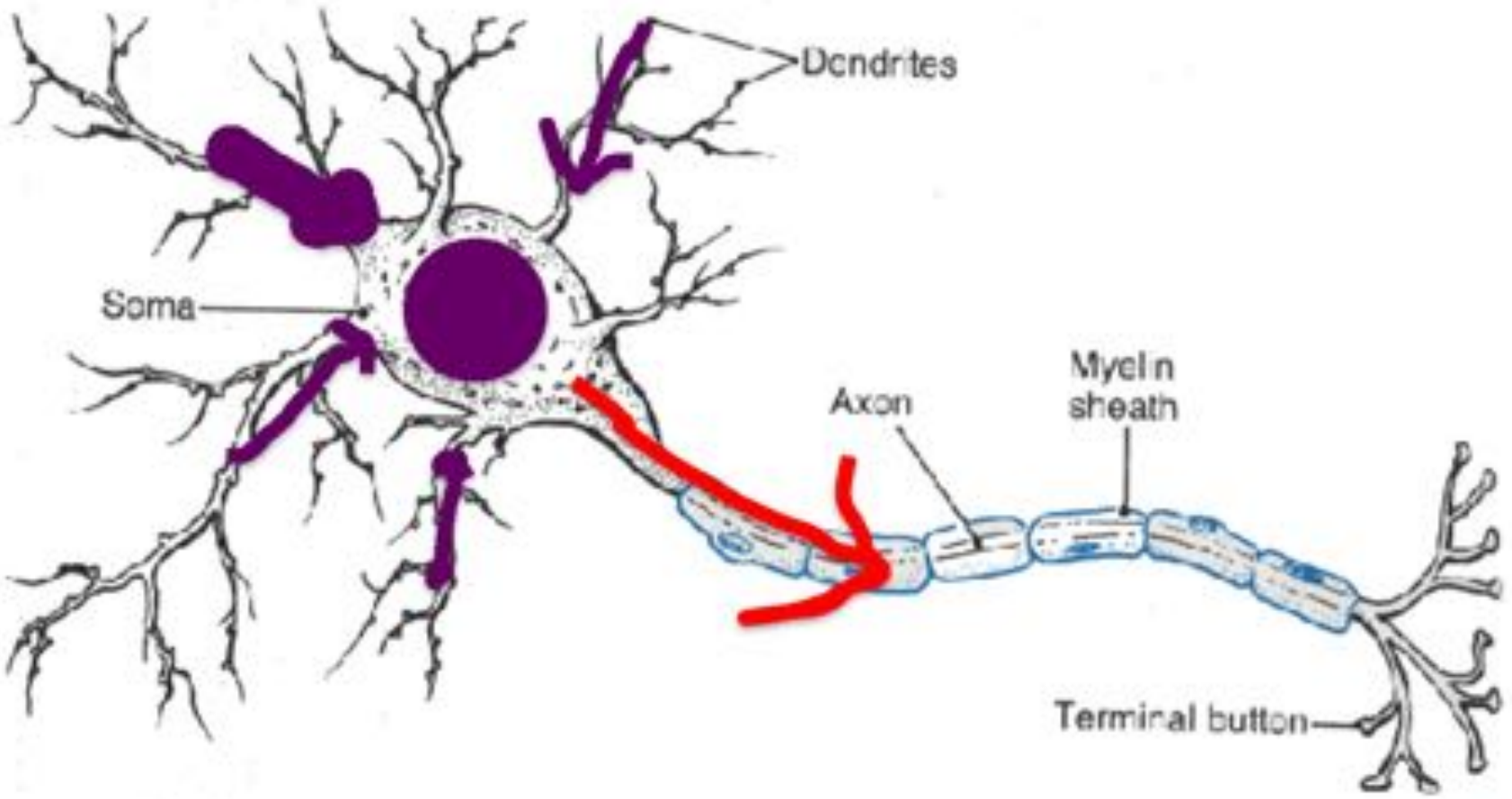
Ancestry

23andMe

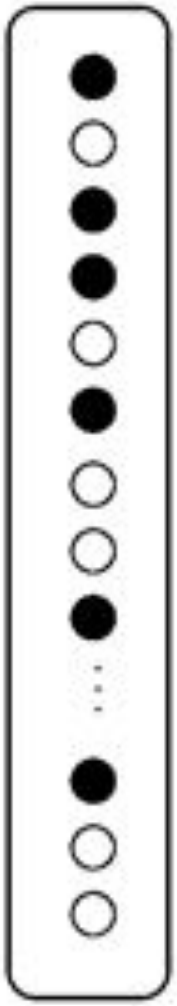
Netflix

NETFLIX

Logistic Regression



Deep Learning



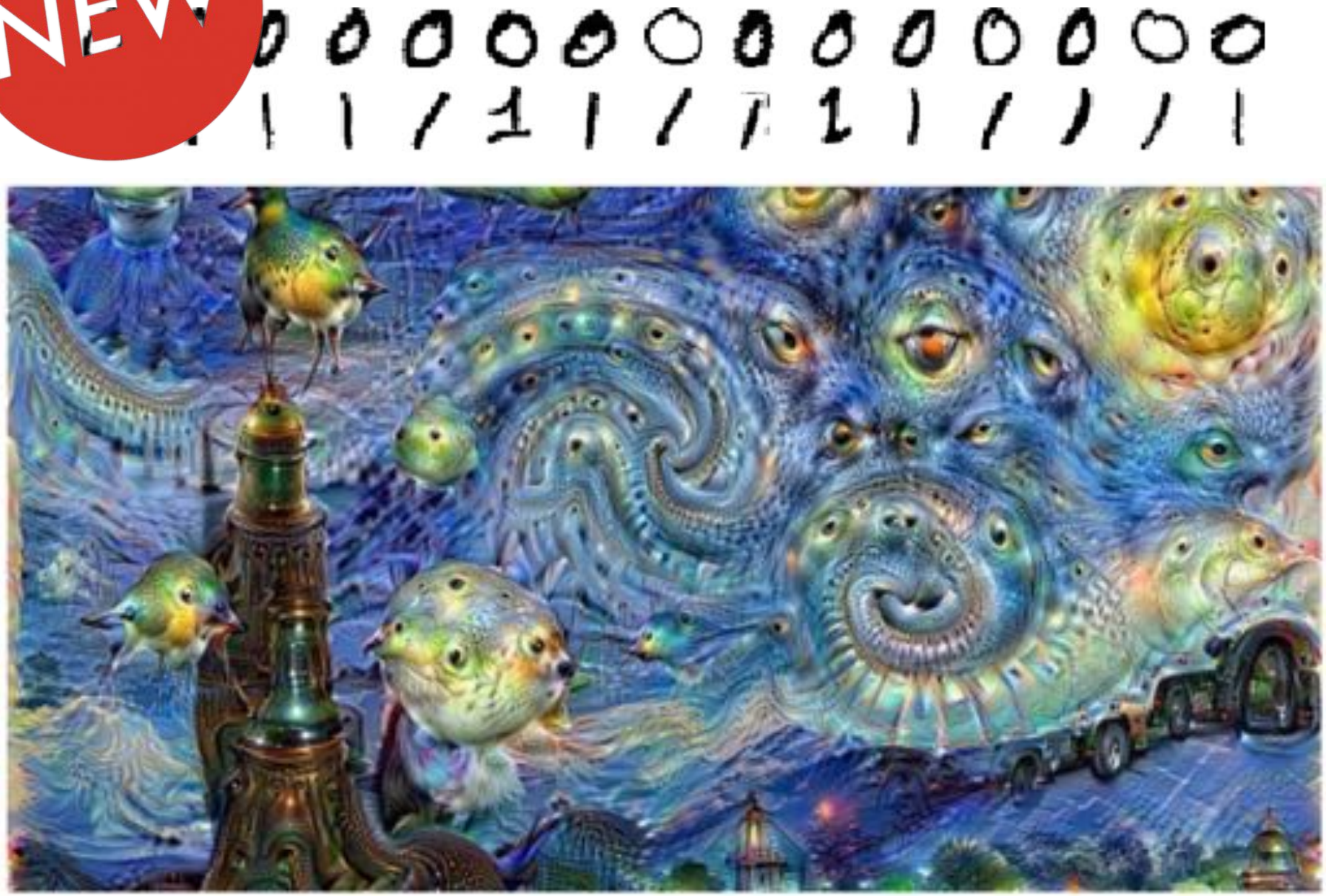
$$LL(\theta) = y \log \hat{y} + (1 - y) \log[1 - \hat{y}]$$

$$\hat{y} = \sigma\left(\sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})}\right)$$

$$\mathbf{h}_j = \sigma\left(\sum_{i=0}^{m_x} \mathbf{x}_i \theta_{i,j}^{(h)}\right)$$

Do it in Section!

NEW



By the numbers

~600 Fruit

~ 30 Major Keys

Naïve Bayes Assumption:

$$P(\mathbf{x}|y) = \prod_i P(x_i|y)$$

1 Contest

Thomas Bayes

- Rev. Thomas Bayes (1702 –1761) was a British mathematician and Presbyterian minister

- He looked remarkably similar to Charlie Sheen
  - But that's not important right now...

Jacob Bernoulli

- Jacob Bernoulli (1654-1705), also known as "James", was a Swiss mathematician

- One of many mathematicians in Bernoulli family
- The Bernoulli Random Variable is named for him
- He is my *academic* great^{12}-grandfather
- Same eyes as Ice Cube

Simeon-Denis Poisson

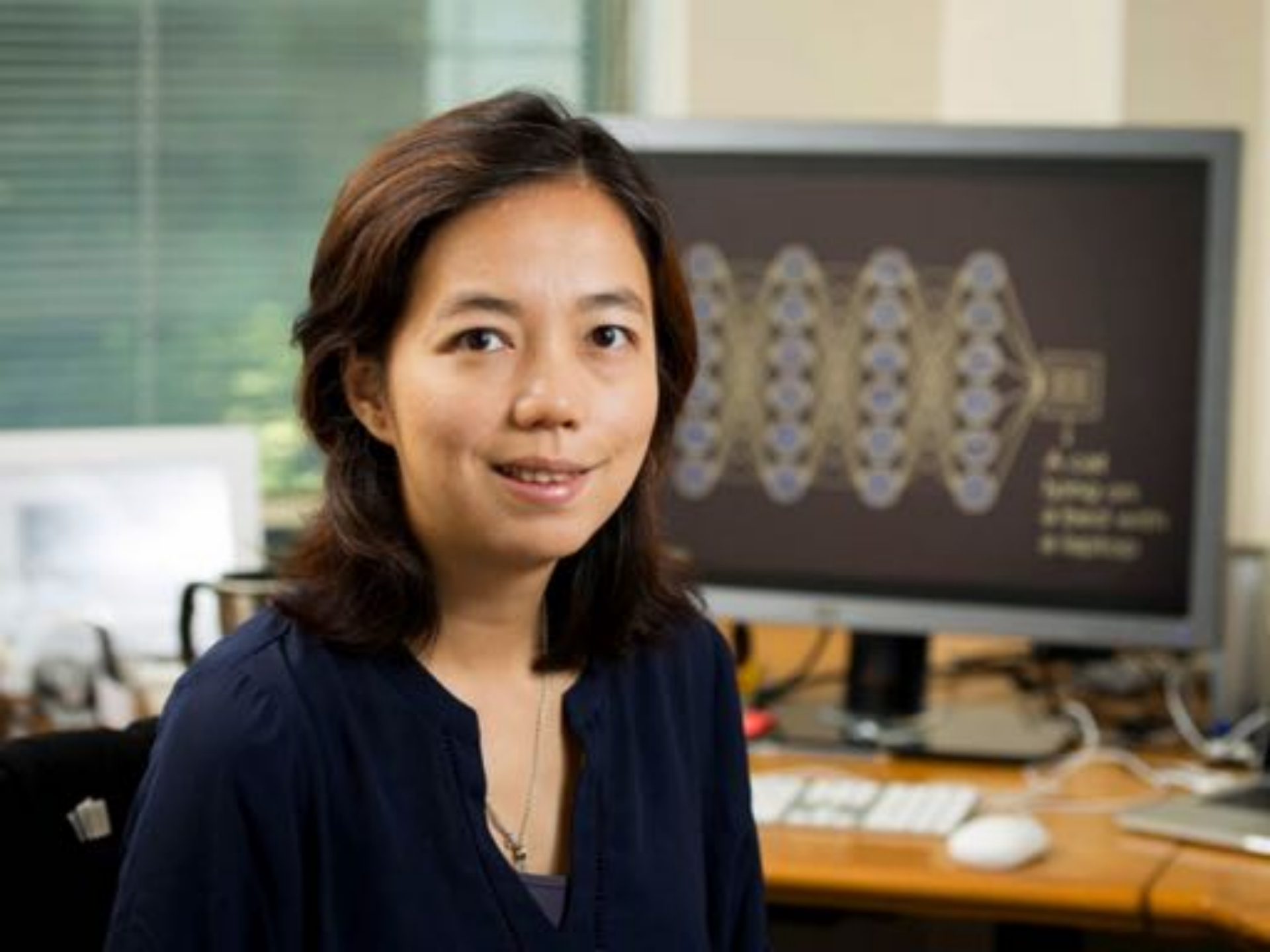
- Simeon-Denis Poisson (1781-1840) was a prolific French mathematician

- Published his first paper at 18, became professor at 21, and published over 300 papers in his life
  - He reportedly said *"Life is good for only two things, discovering mathematics and teaching mathematics."*
- I'm going with French Martin Freeman

Carl Friedrich Gauss

- Carl Friedrich Gauss (1777-1855) was a remarkably influential German mathematician

- Started doing groundbreaking math as teenager
  - Did not invent Normal distribution, but popularized it
- He looked like Martin Sheen
  - Who is, of course, Charlie Sheen's father



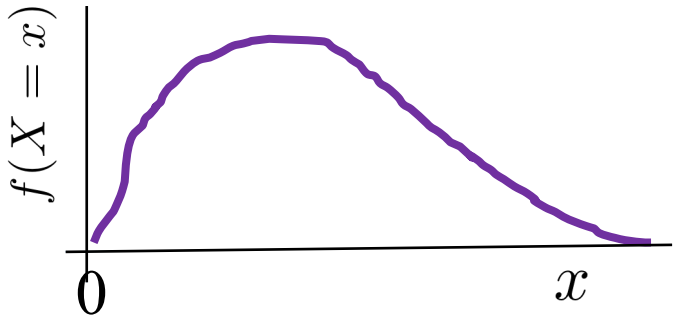
Proximal Concepts

Bounds: Markov's Inequality

- Say X is a **non-negative** random variable

$$P(X \geq a) \leq \frac{E[X]}{a}, \quad \text{for all } a > 0$$

- Proof:
  - I = 1 if X ≥ a, 0 otherwise
  - Since $X \geq 0, \quad I \leq \frac{X}{a}$
  - Taking expectations:



$$E[I] = P(X \geq a) \leq E\left[\frac{X}{a}\right] = \frac{E[X]}{a}$$

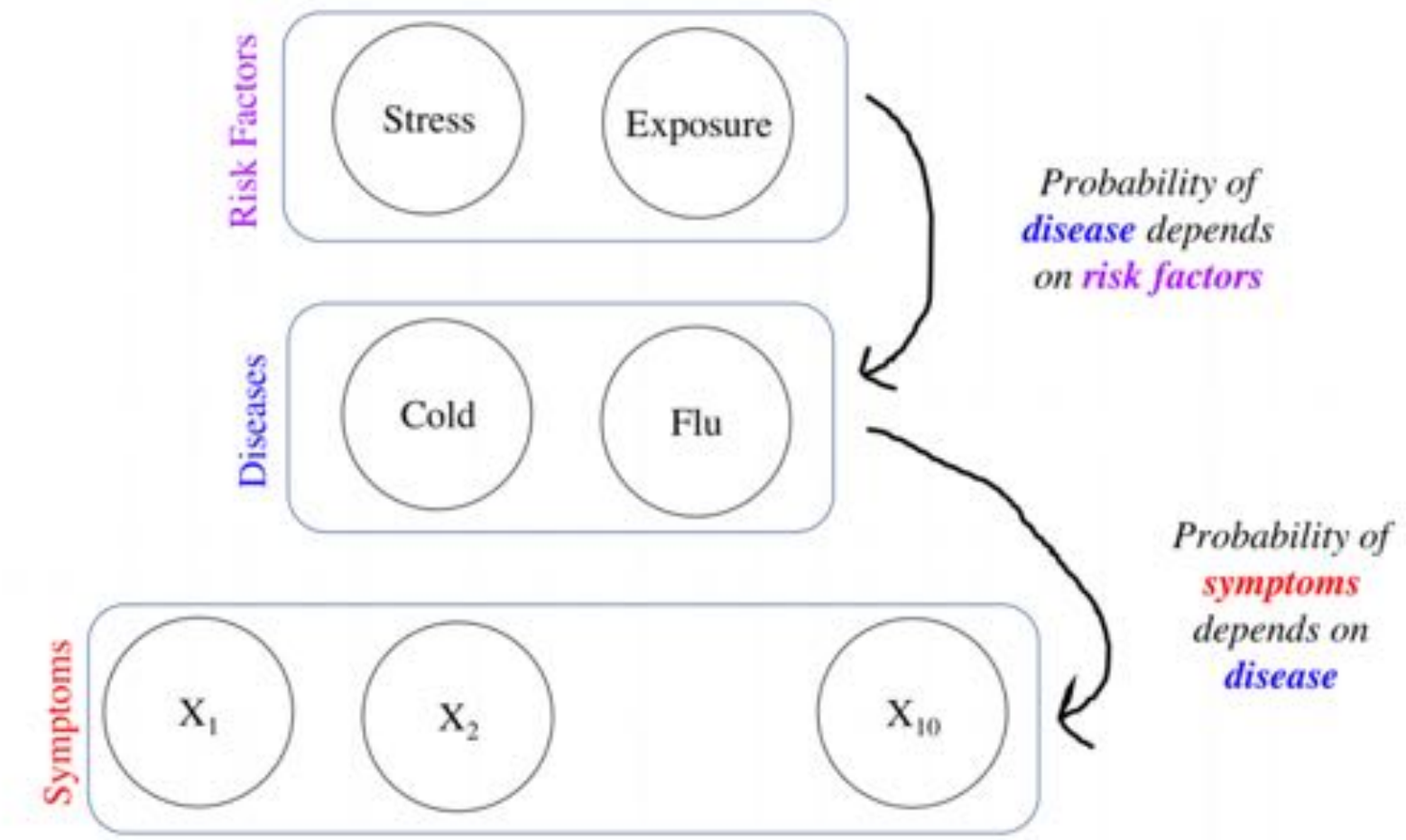
Markov and the Midterm

- Statistics from CS109 midterm
  - X = midterm score
  - Using sample mean $\overline{X}$ = 102.0 ≈ E[X]
  - What is P(X ≥ 110)?

$$P(X \geq 110) \leq \frac{E[X]}{110} = \frac{102}{110} = 0.93$$

  - Markov bound: ≤ 93% of class scored 110 or greater
  - In fact, 15.1% of class scored 110 or greater
    - Markov inequality can be a very loose bound
    - But, it made no assumption at all about form of distribution!

Learn Bayes Nets *Params*?



* That is what we did with Naïve Bayes

Learn Bayes Nets *Structure*?

* That is what we did with Ebola Bats!

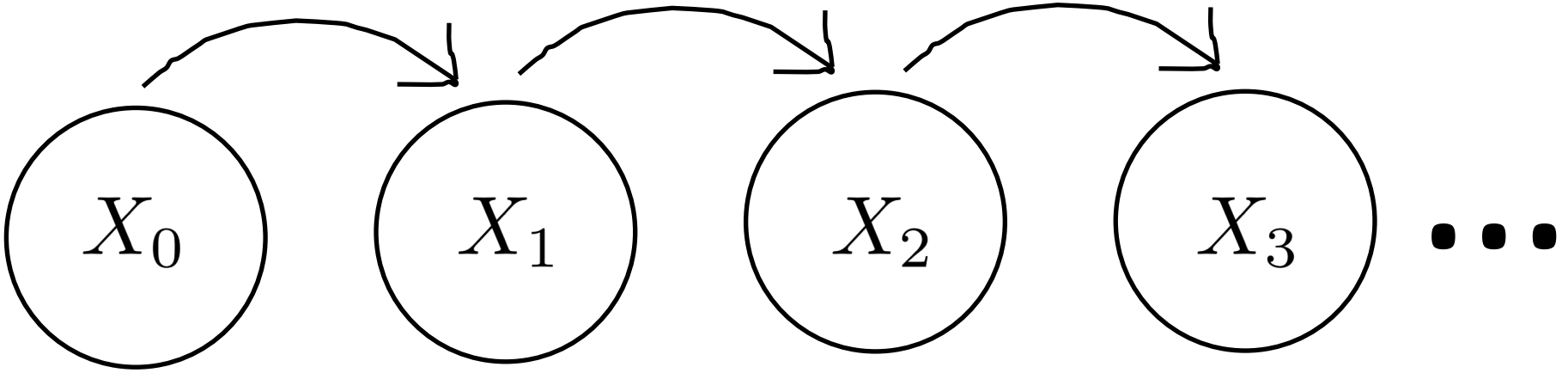
Missing Data?

Missing Not at Random: You collect data on whether or not people intend to vote for Ayesha, a candidate in an upcoming election. You send an electronic poll to 100 randomly chosen people. You assume all 100 responses are IID.

User Response	Count
Responded that they will vote for Ayesha	40
Responded that they will **not** vote for Ayesha	45
Did not respond	15

* Scratched the surface in section

Temporal Patterns?



* Special type of Bayesian Network called a **Markov Network**

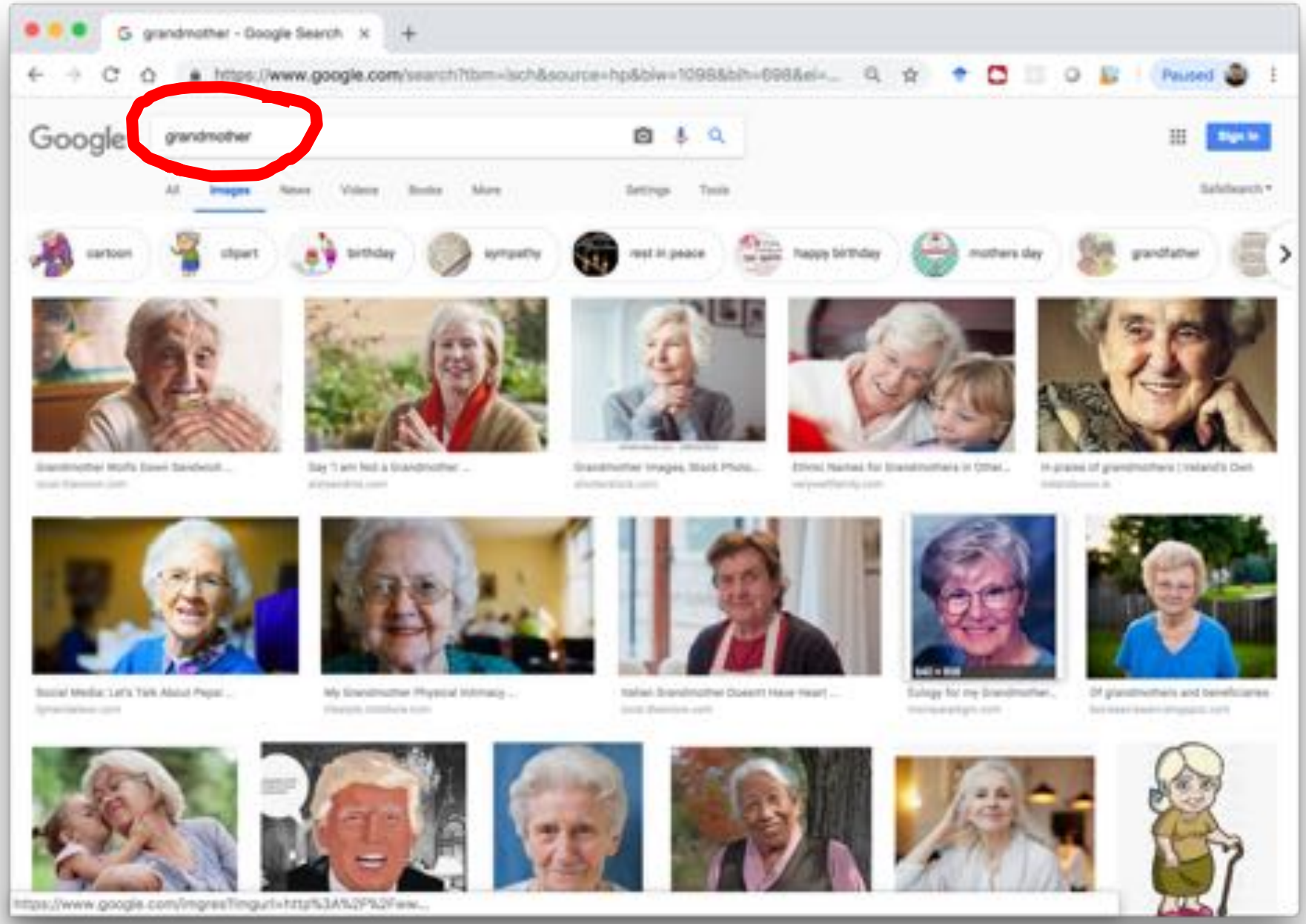
Ethics and AI

Ethics and Datasets?

Sometimes machine learning feels universally unbiased.

We can even call some estimators “unbiased”

Google/Nikon/HP had biased datasets


grandmother - Google Search
grandmother
Google
Sign In
All
Images
News
Videos
Books
More
Settings
Tools
SafeSearch
cartoon
clipart
birthday
sympathy
rest in peace
happy birthday
mothers day
grandfather

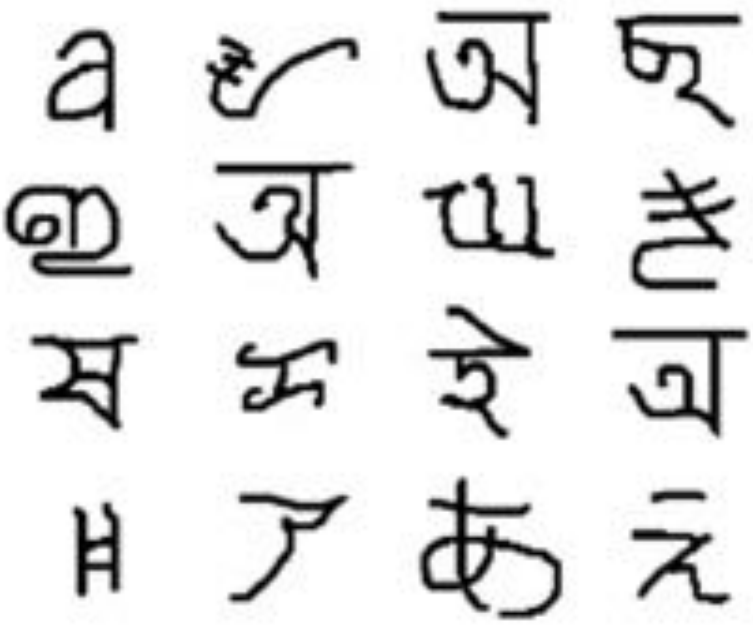
Much more to Ethics + AI

Open Problems

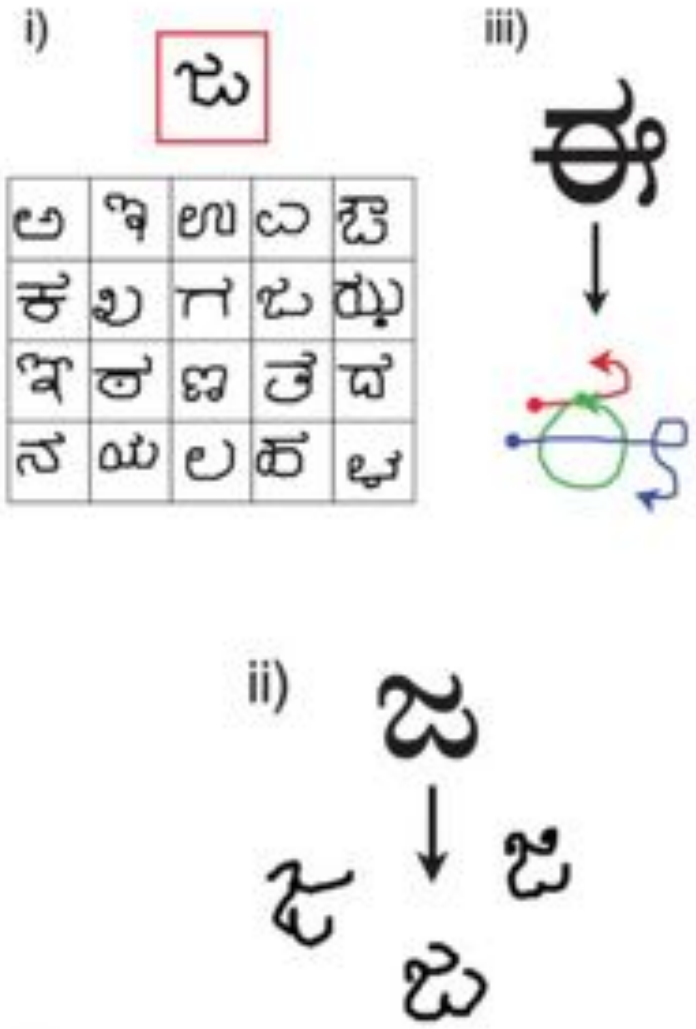
One Shot Learning

Single training example:

Test set:

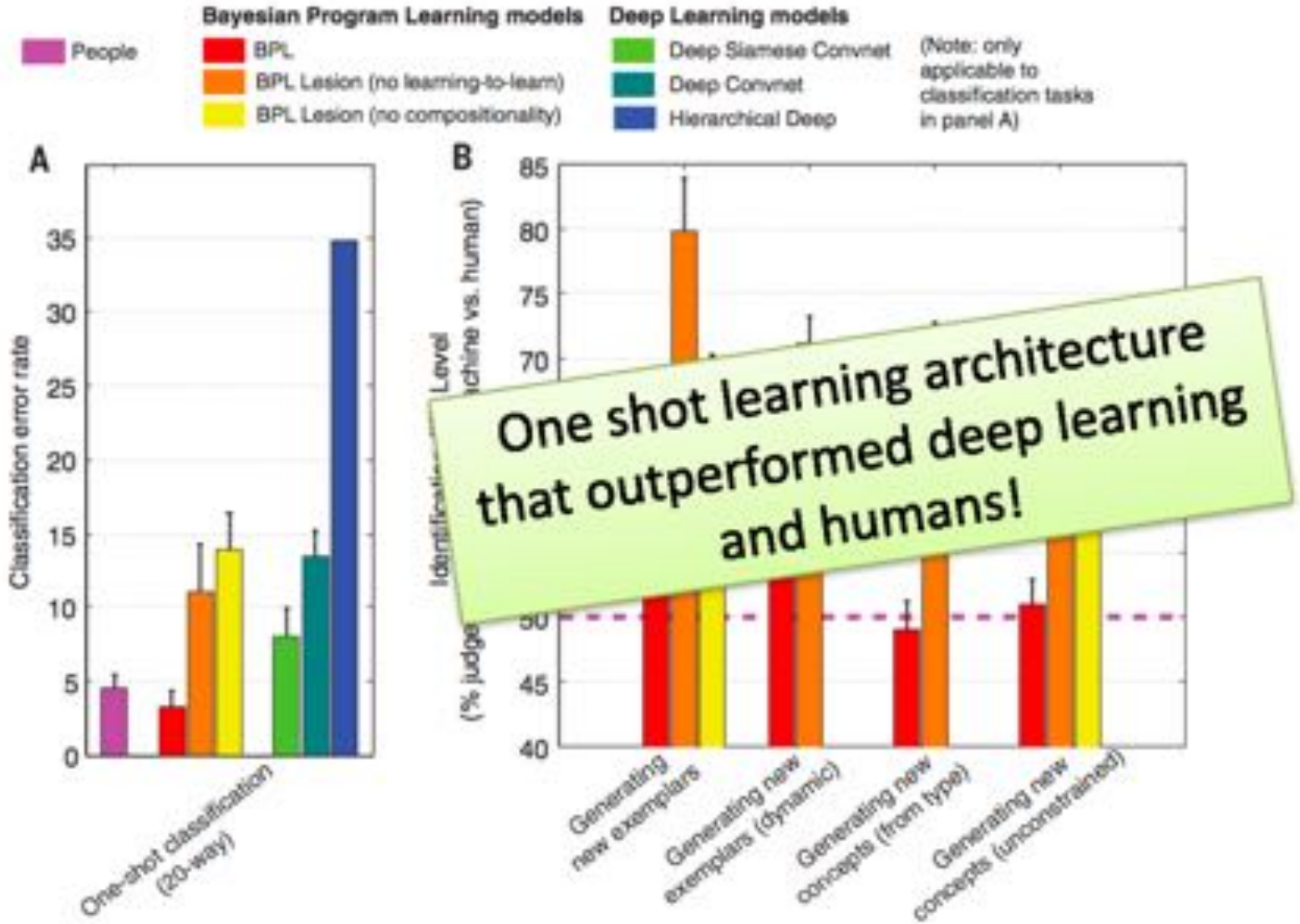


Bayesian Program Learning



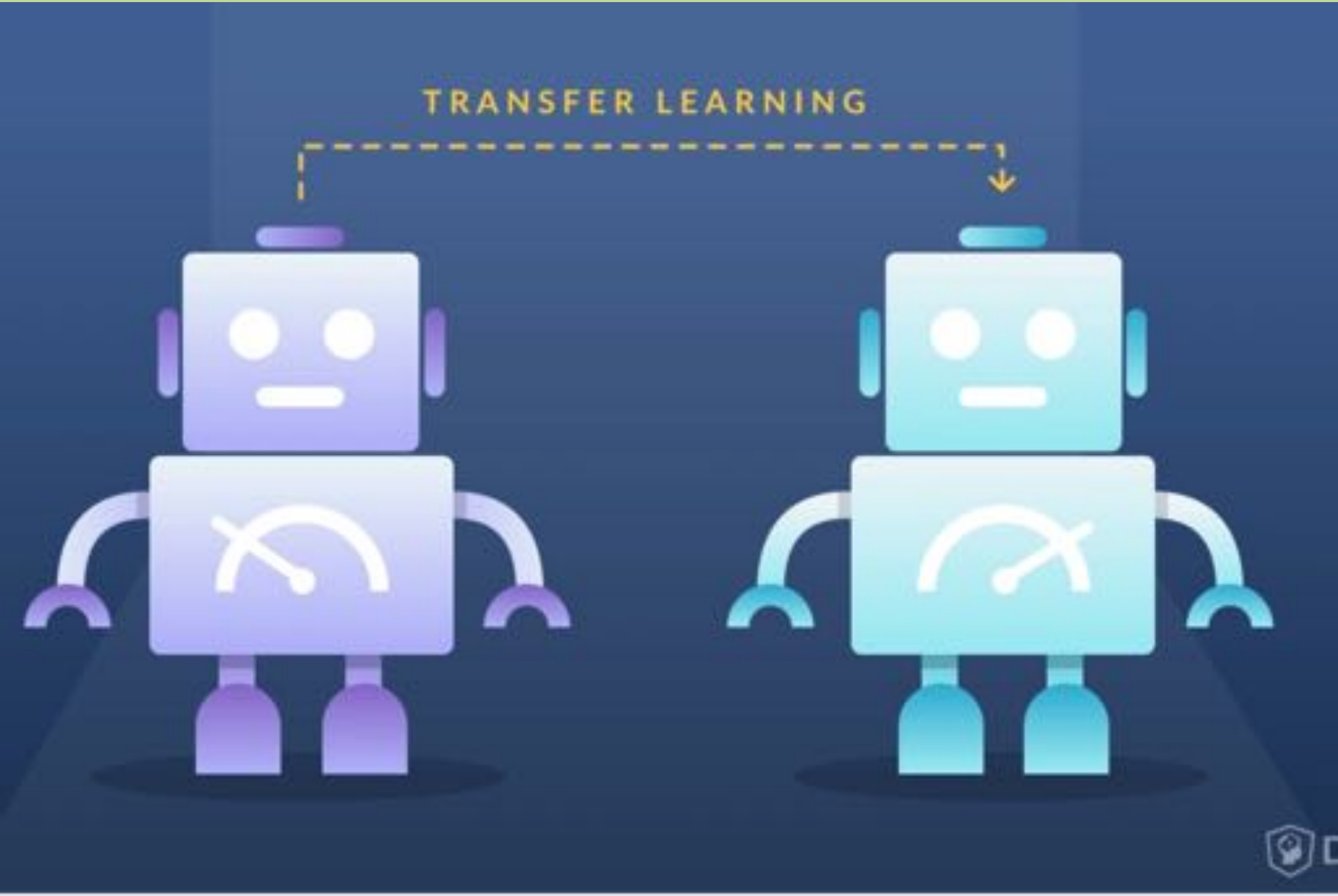
Lake et al. Human-level concept learning through probabilistic program induction

Bayesian Program Learning



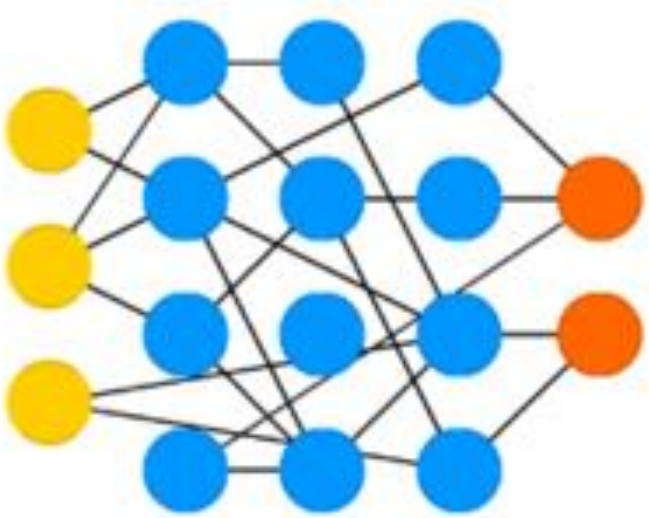
Lake et al. Human-level concept learning through probabilistic program induction

Transfer Learning

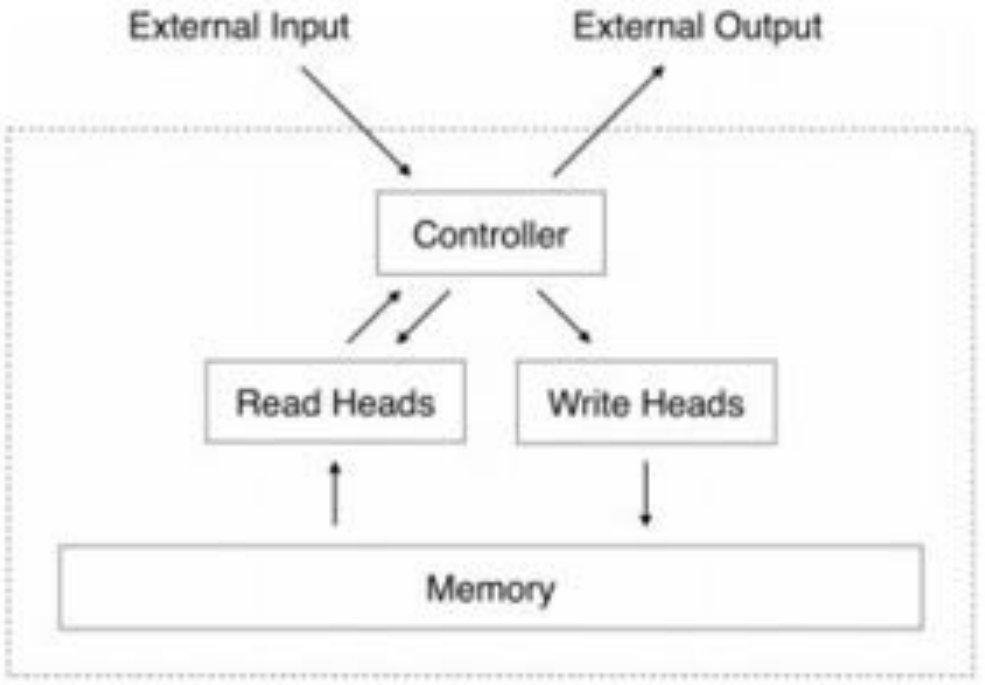


Neural Network *Structure*?

Neural Turing Machines

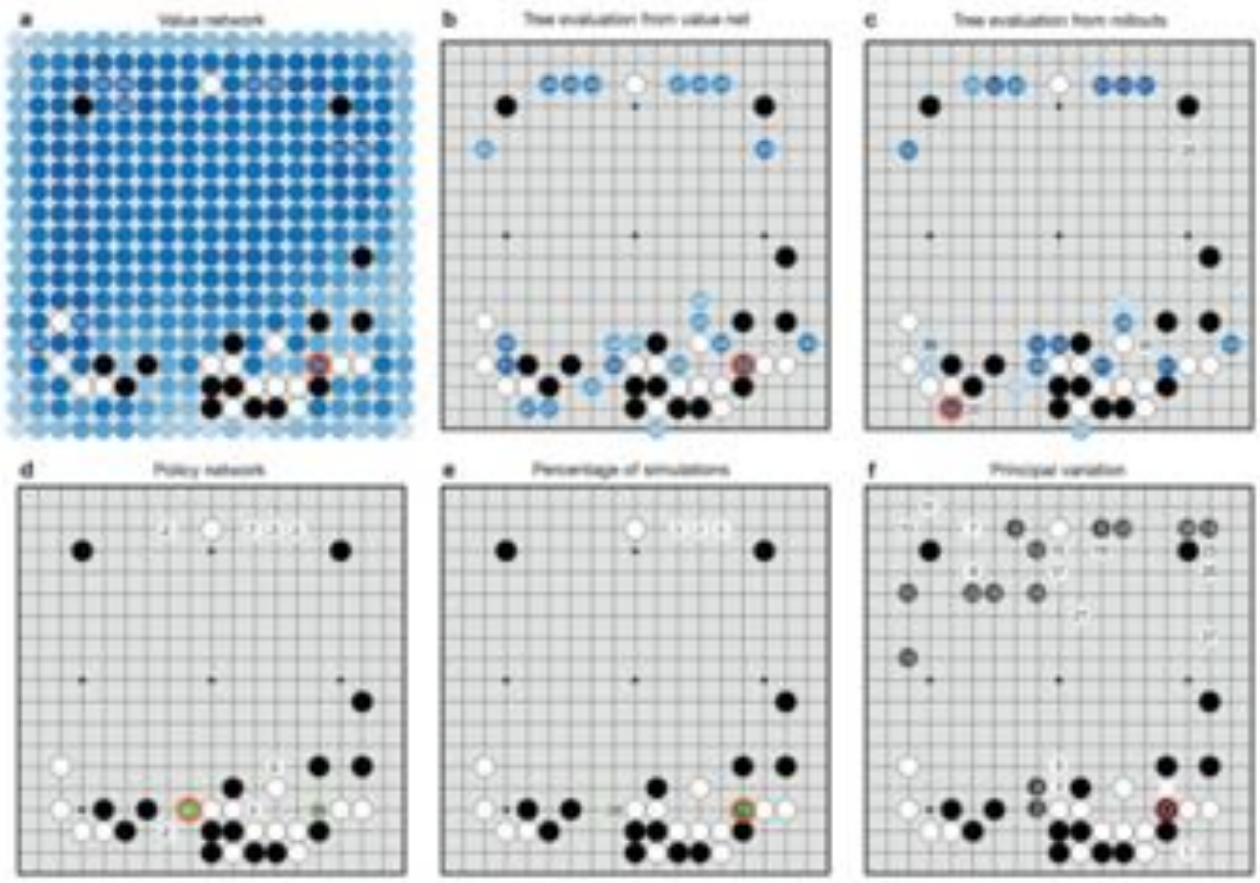


Neural Turing Machine (NTM)



Theoretical Deep Learning

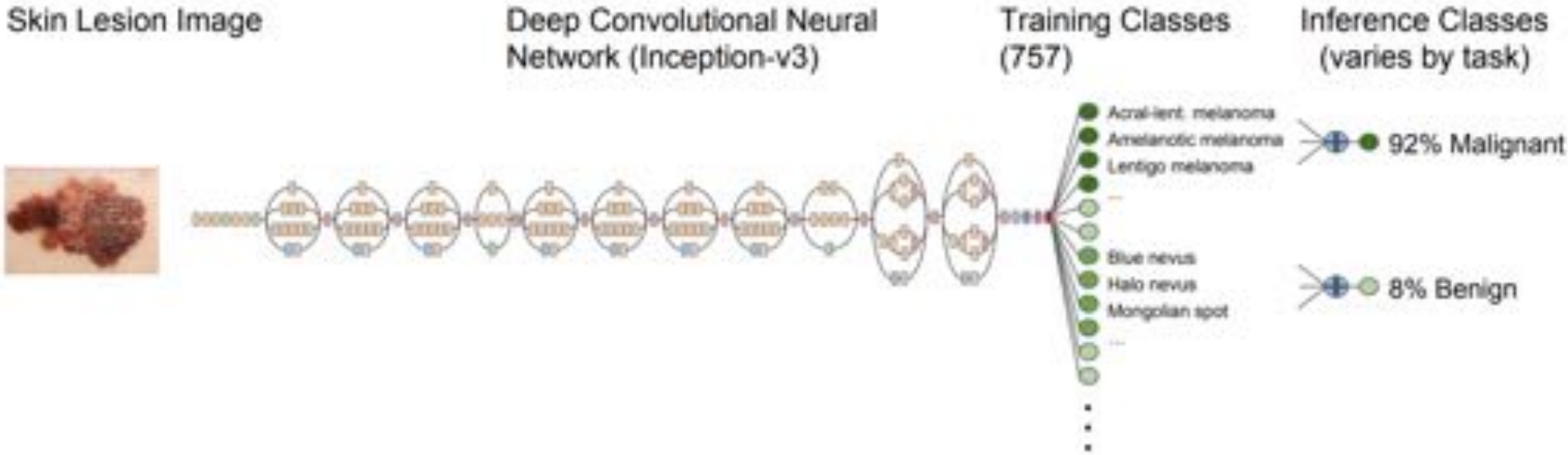
Sampling + Deep Learning!



AlphaGo

Natural Language

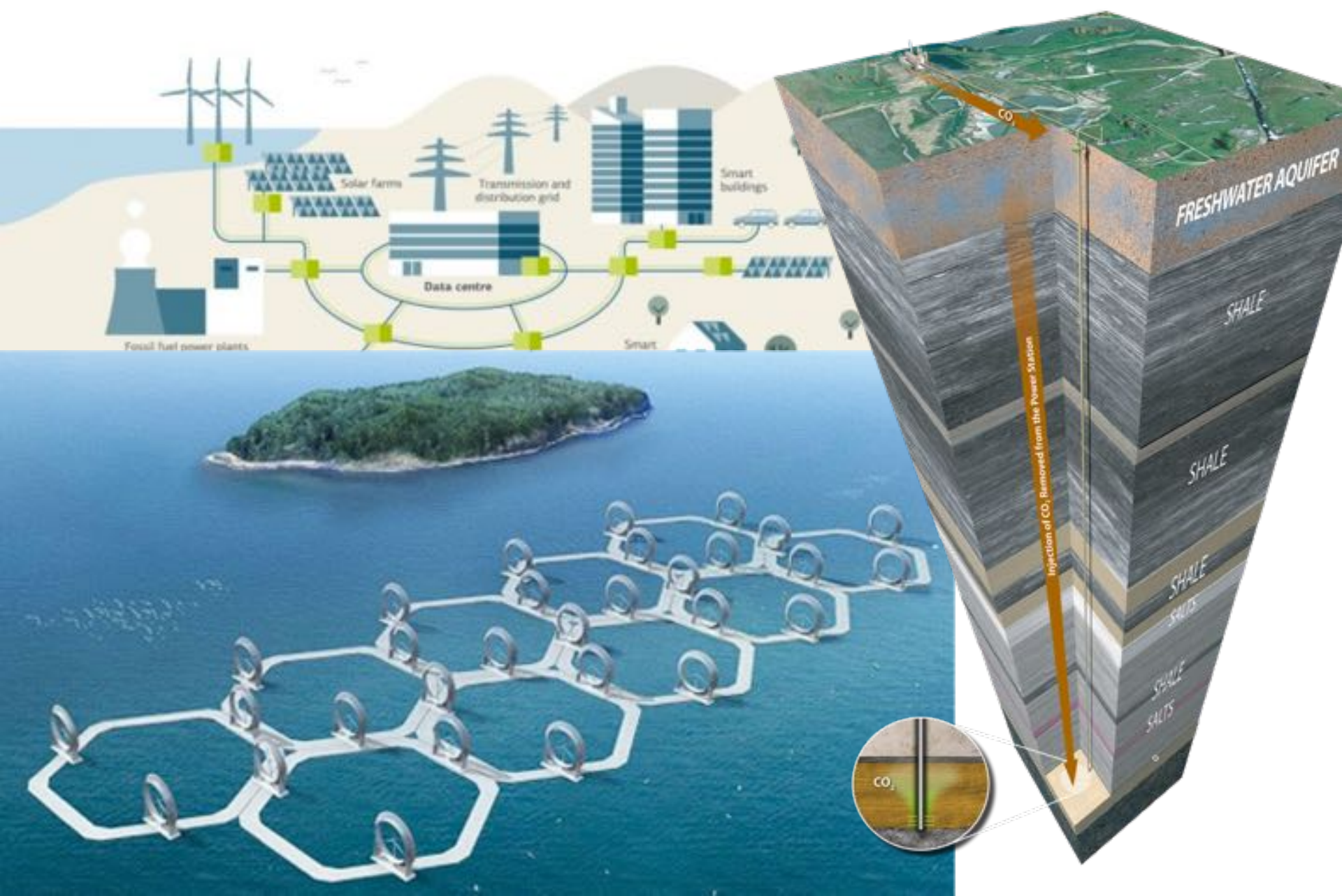
AI for Medicine



Esteva, Andre, et al. "Dermatologist-level classification of skin cancer with deep neural networks." *Nature* 542.7639 (2017): 115-118.

Climate Change?

Climate Change?



Honorable Mentions

Differential Privacy

Fairness and AI

General AI

Better Optimization

DeepLearning + X

Self Driving Cars

Understanding Video

After CS109

Theory

CS161 – Algorithmic analysis
Stats 217– Stochastic Processes
CS 238 – Decision Making Under Uncertainty
CS 228 – Probabilistic Graphical Models

AI

CS 221 – Intro to AI
CS 229 – Macine Learning
CS 230 – Deep Learning
CS 224N – Natural Language Processing
CS 234 – Reinforcement Learning

Applications

CS 279 – Bio Computation
Literally any class with numbers in it

Technology magnifies.
What do we want magnified?

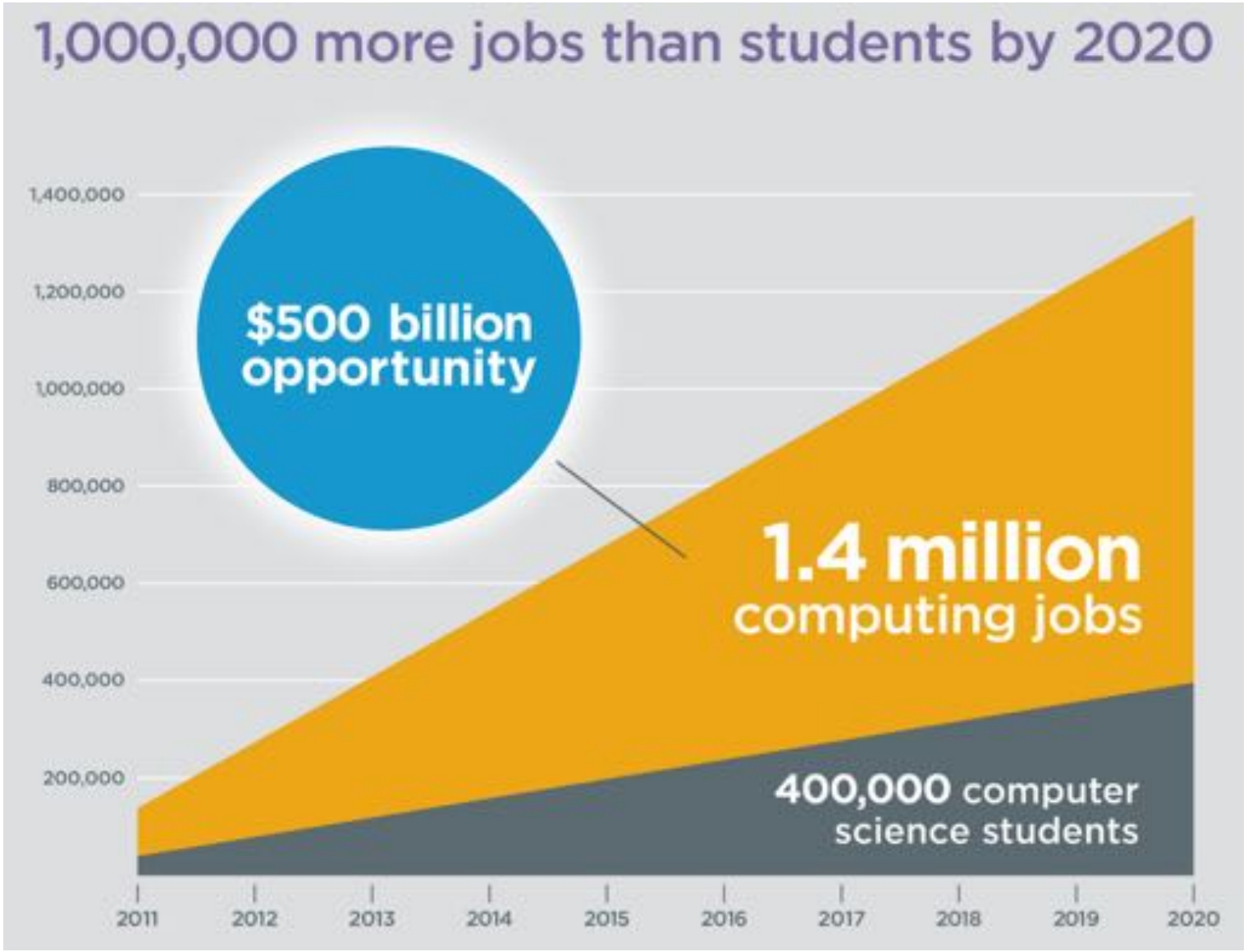
Why Study Probability + CS?

Interdisciplinary

Closest Thing To Magic

Now is the Time

Oh and Its Useful



Code.org

Everyone is Welcome

I guarantee the techniques will change…

You are close to the edge of human knowledge

(all of you)

The End