CS109 Montage

Chris Piech
CS109, Stanford University

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%

Counting Rules

Counting operations on *n* objects

Sort (order matters) (perms)

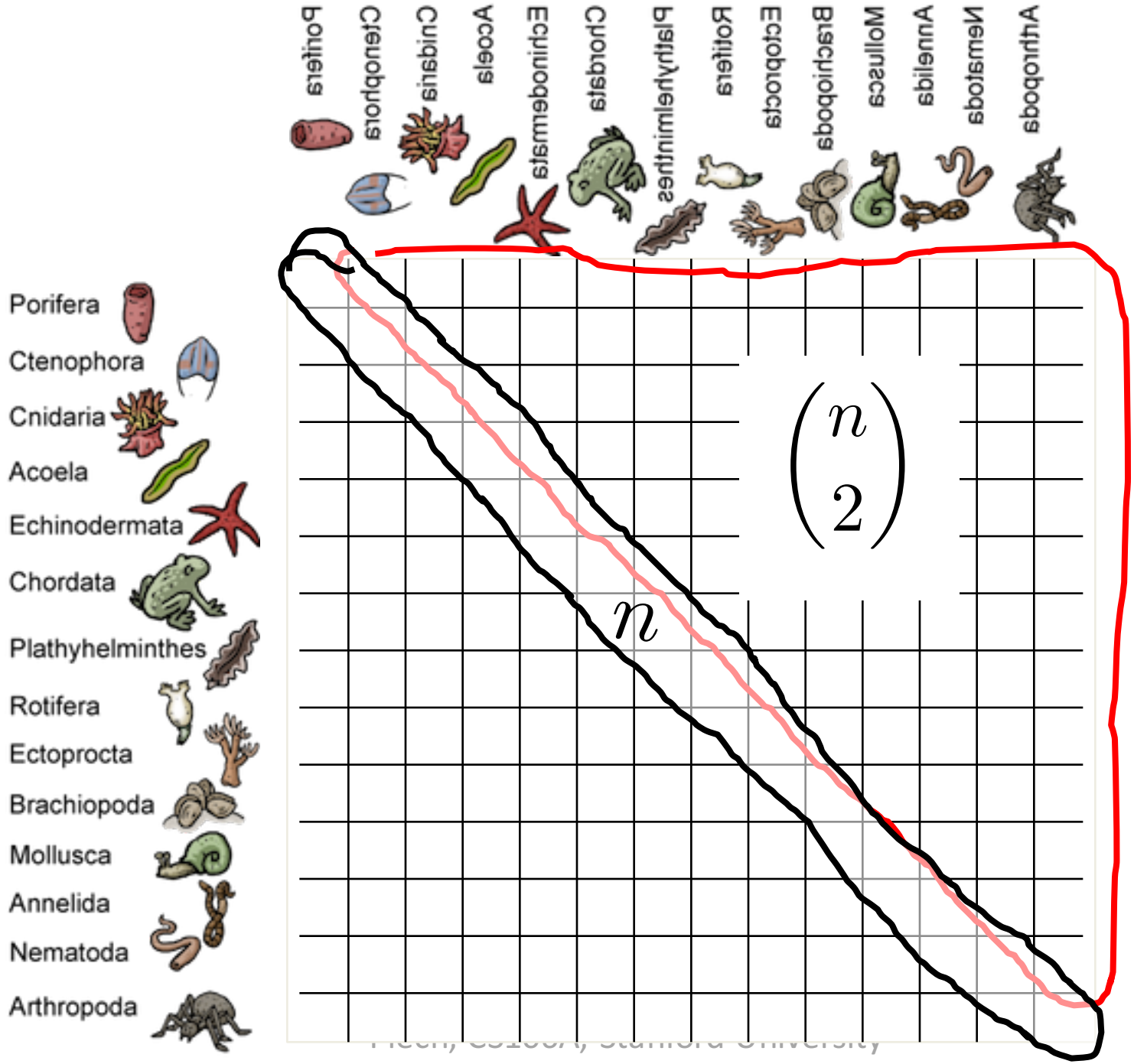
- Distinct: $n!$
- Some Distinct: $\frac{n!}{n_1!n_2!\ldots}$

Choose *k* (combinations)

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

Put in *r* buckets

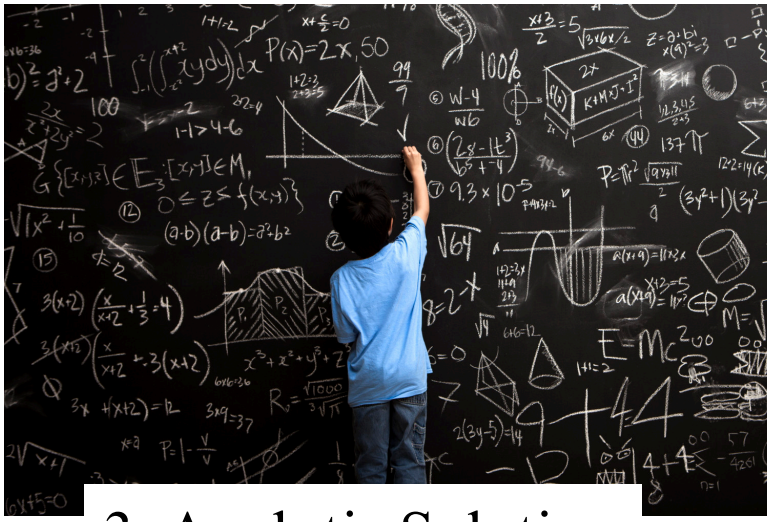
- Distinct: r^n
- None Distinct: $\frac{(n+r-1)!}{n!(r-1)!}$


Porifera
Ctenophora
Cnidaria
Acoela
Echinodermata
Chordata
Plathyhelminthes
Rotifera
Ectoprocta
Brachiopoda
Mollusca
Annelida
Nematoda
Arthropoda
$\binom{n}{2}$
n

Sources of Probability

1. Experimentation

2. Dataset



3. Analytic Solution

4. Expert Opinion

What is a Probability?

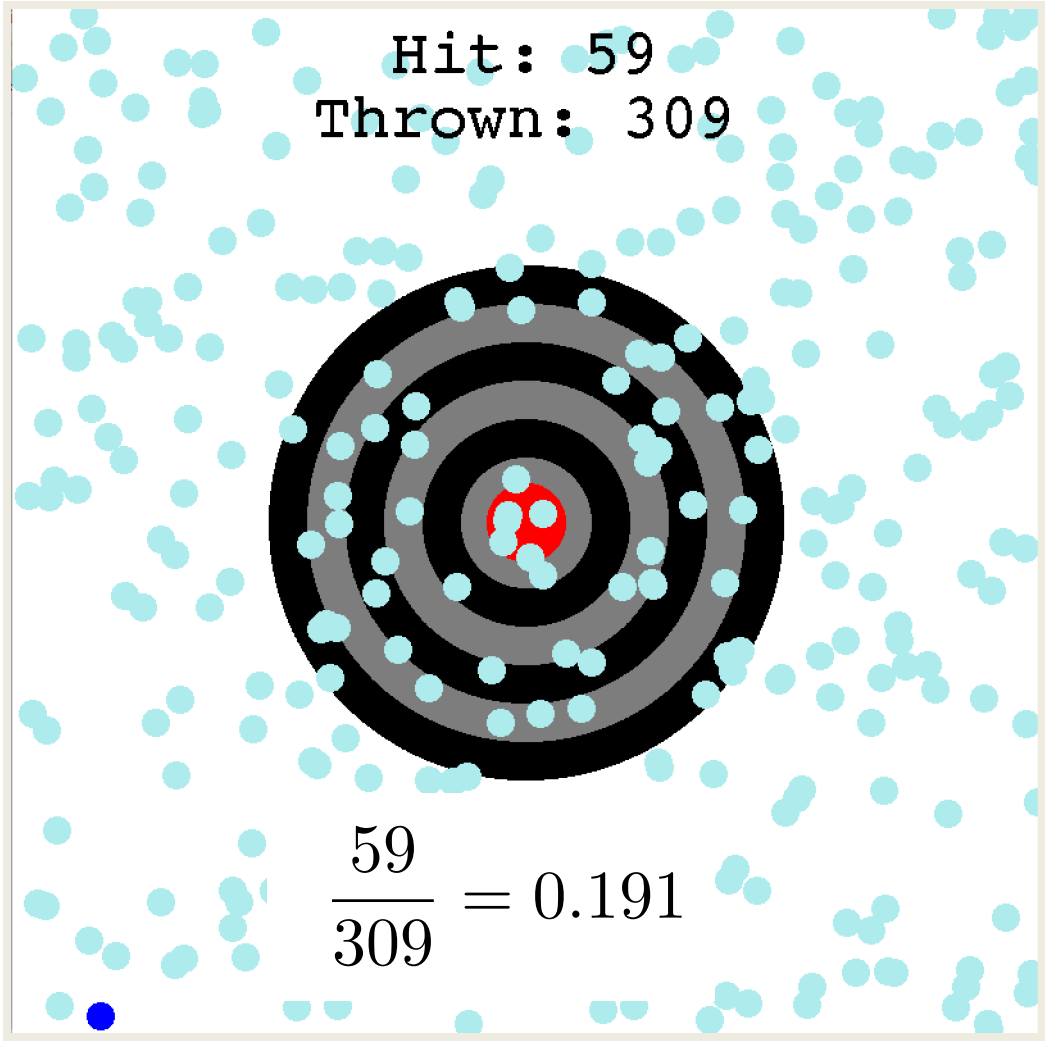
$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

Everything in the world is either

a potato or not a potato.

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

Target Revisited



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)?

Nancy Alisha Sophia Jun-Hyu Mike

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

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?

$$P(E)$$

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

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

Netflix and Learn

What is the probability that a user will watch Life is Beautiful, given they watched Amelie?

$$P(E|F)$$

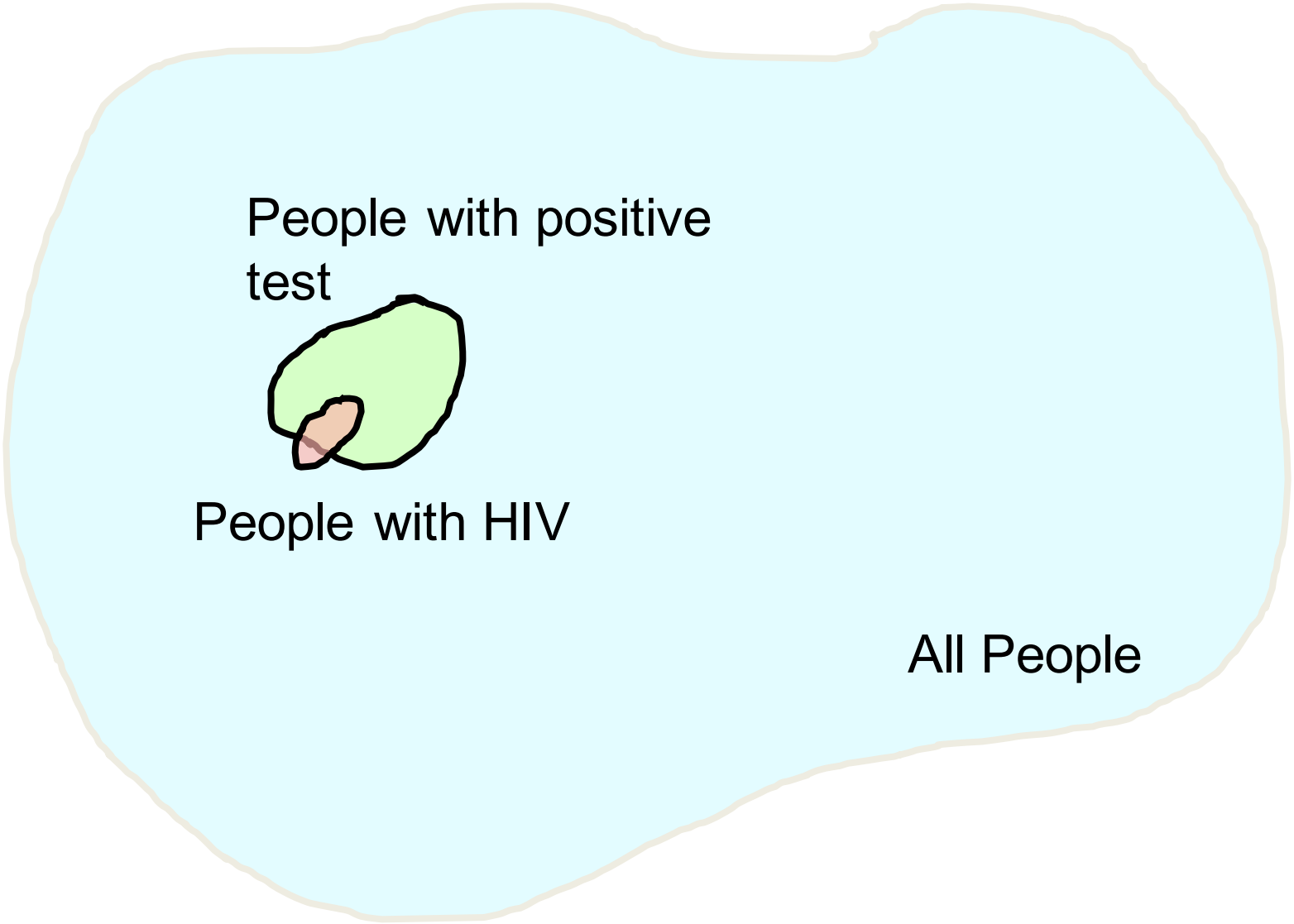
$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\#\text{people who watched both}}{\#\text{people who watched } F}$$

$$P(E|F) = 0.42$$

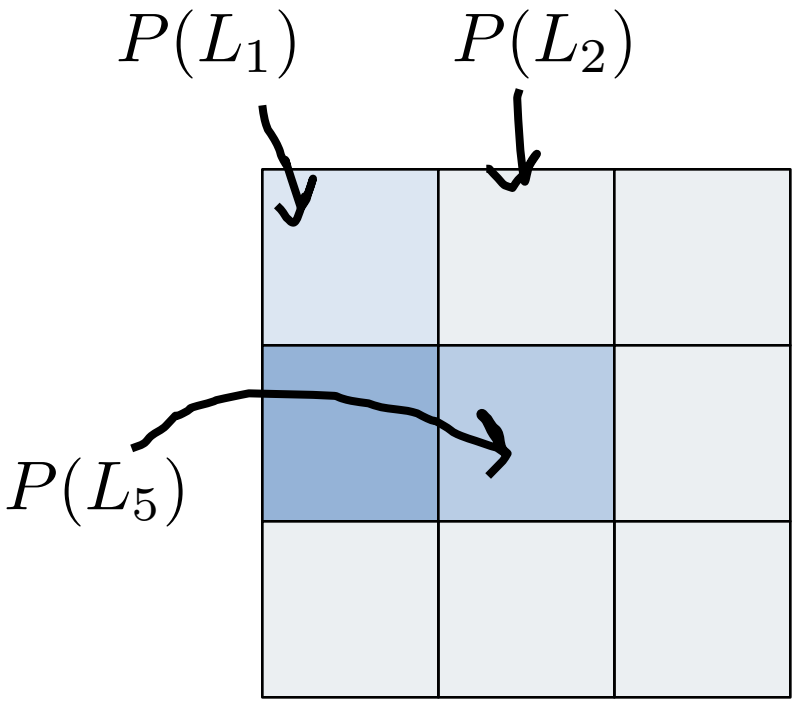
Bayes Theorem

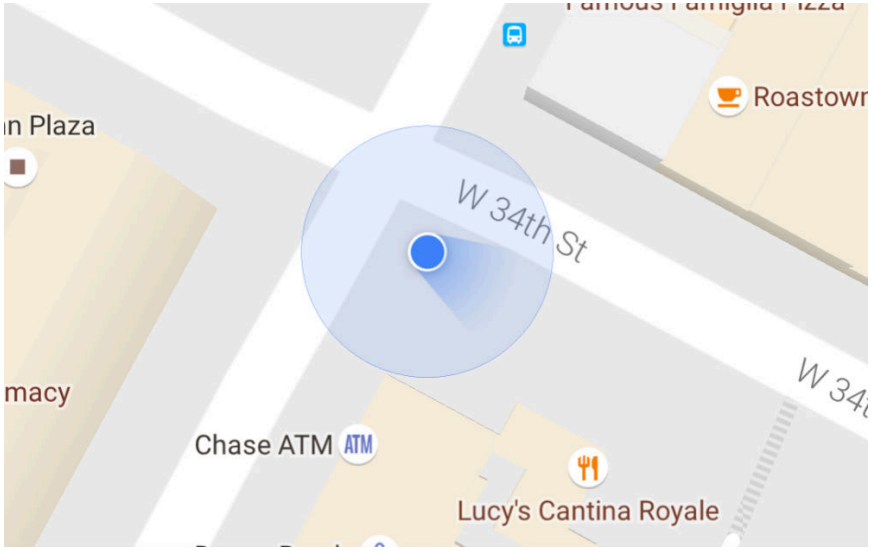
$$P(F|E) = \frac{P(E|F)P(F)}{P(E)}$$

Bayes Theorem Intuition



Update Belief





Before Observation

Let's Make a Deal

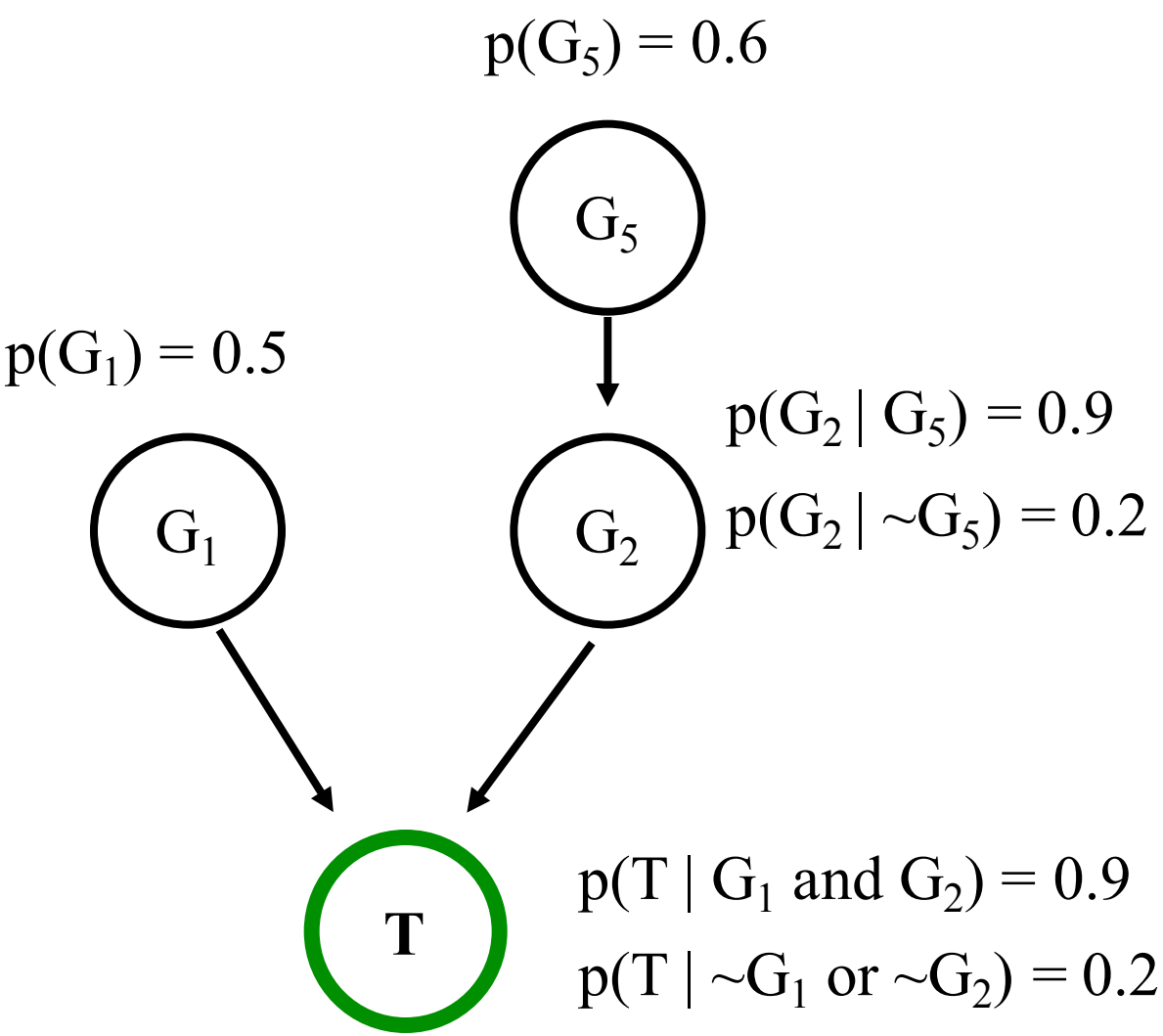
- Game show with 3 doors: A, B, and C

  - 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)

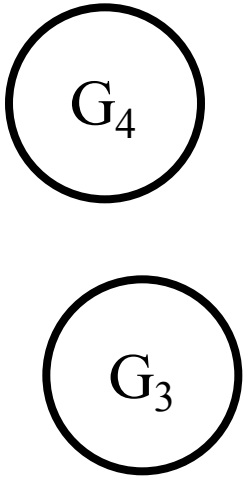
Independence

Recall our Ebola Bats

Discovered Pattern

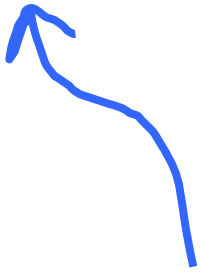


These genes don't impact T



Random Variables

$$P(Y = k)$$

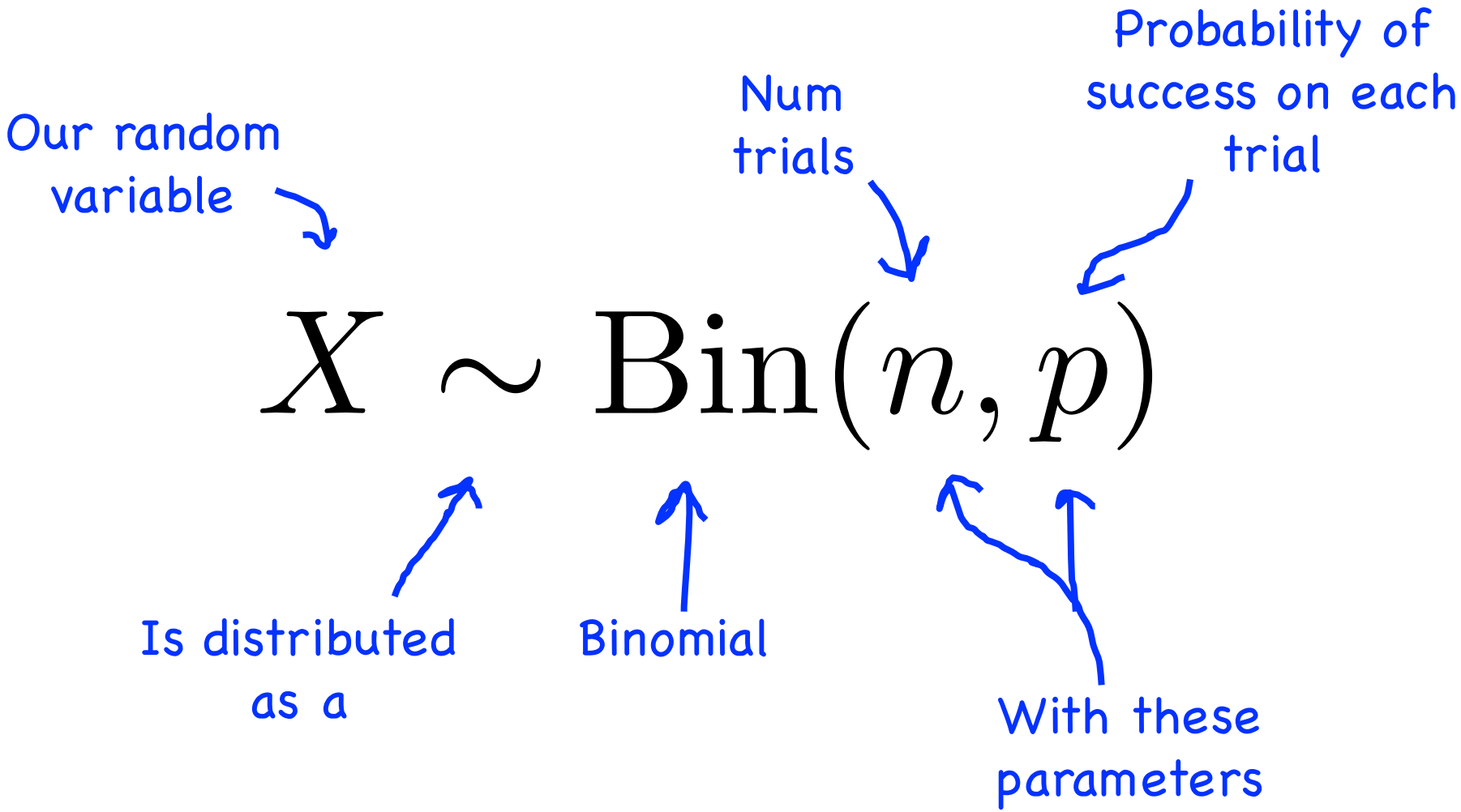


This is a function

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

Bernoulli

Binomial


Our random variable
Num trials
Probability of success on each trial
$X \sim \text{Bin}(n, p)$
Is distributed as a
Binomial
With these parameters

Poisson

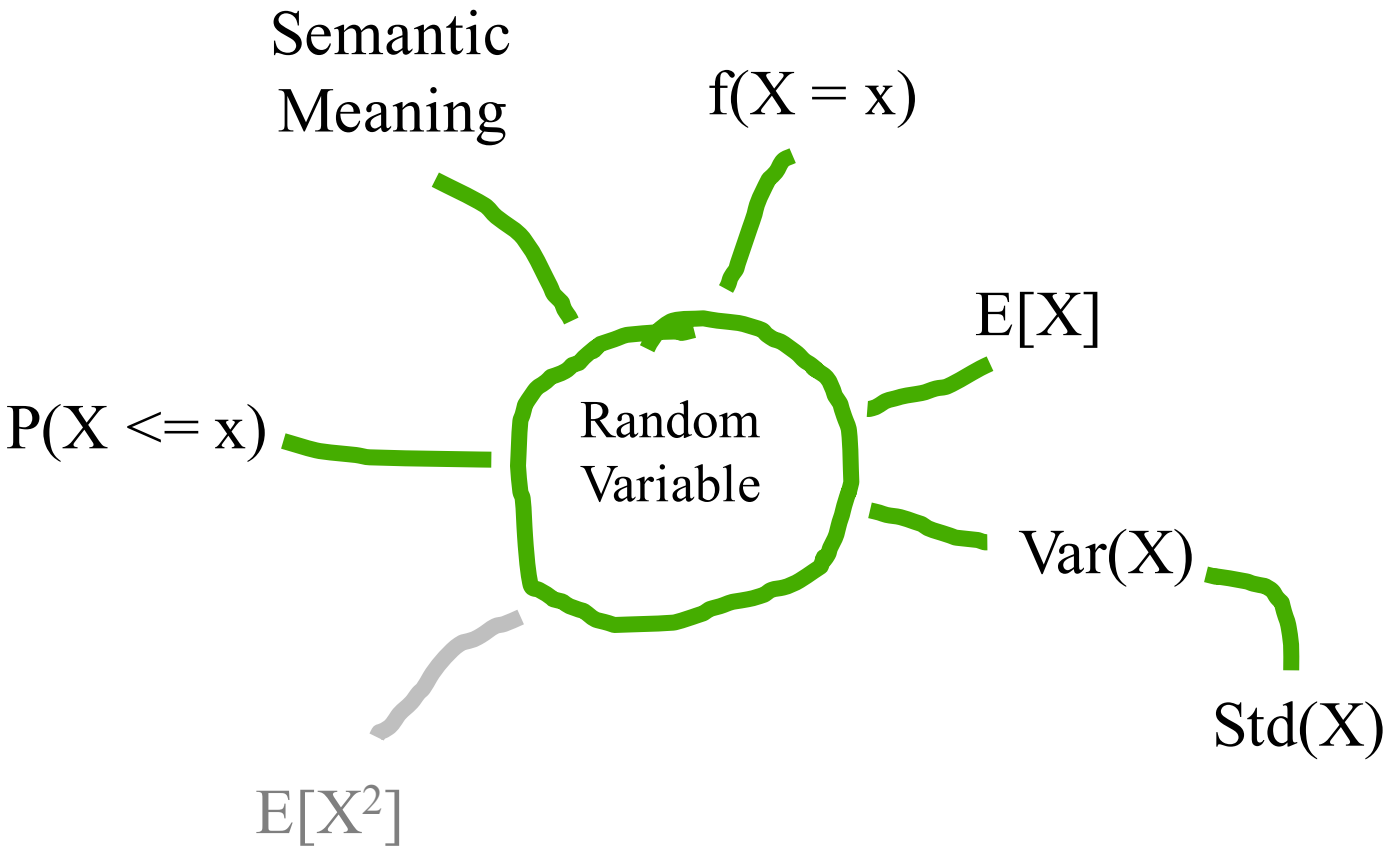

On average λ = 5 requests per minute
The Castro Theatre
THE CASTRO
MISSION DISTRICT
19th St
Atlas Cafe
20th St
Guerrero St
Alabama St
433A
Castro St
22nd St
The Spice Jar
Carolina St
De Haro St
25th St
26th St
Terra Mia Ceramic Studio
Cesar Chavez St
Folsom St
1591 Treat Avenue
Napoleon St
Toland St
Diamond St
The Royal Cuckoo
Valley St
55
Billy Goat Hill
BERNAL HEIGHTS
Lowe's Home Improvement
Cortland Ave
431

Geometric

Sequence 1:

TTHHTHTTHTTTHTTTHTTTHTTHTHHTH
HTTHHTTHHHTHHTHTTHTHTTHHTHHHT
HTHHTHHHTHTHTTHTTHHTHTHTHTTHH
TTHHTHTTHTHTHTHTHTHTHHHTHTHTH
TTHHTHTHTHTHHHTTHHTHTTTHTHHTH

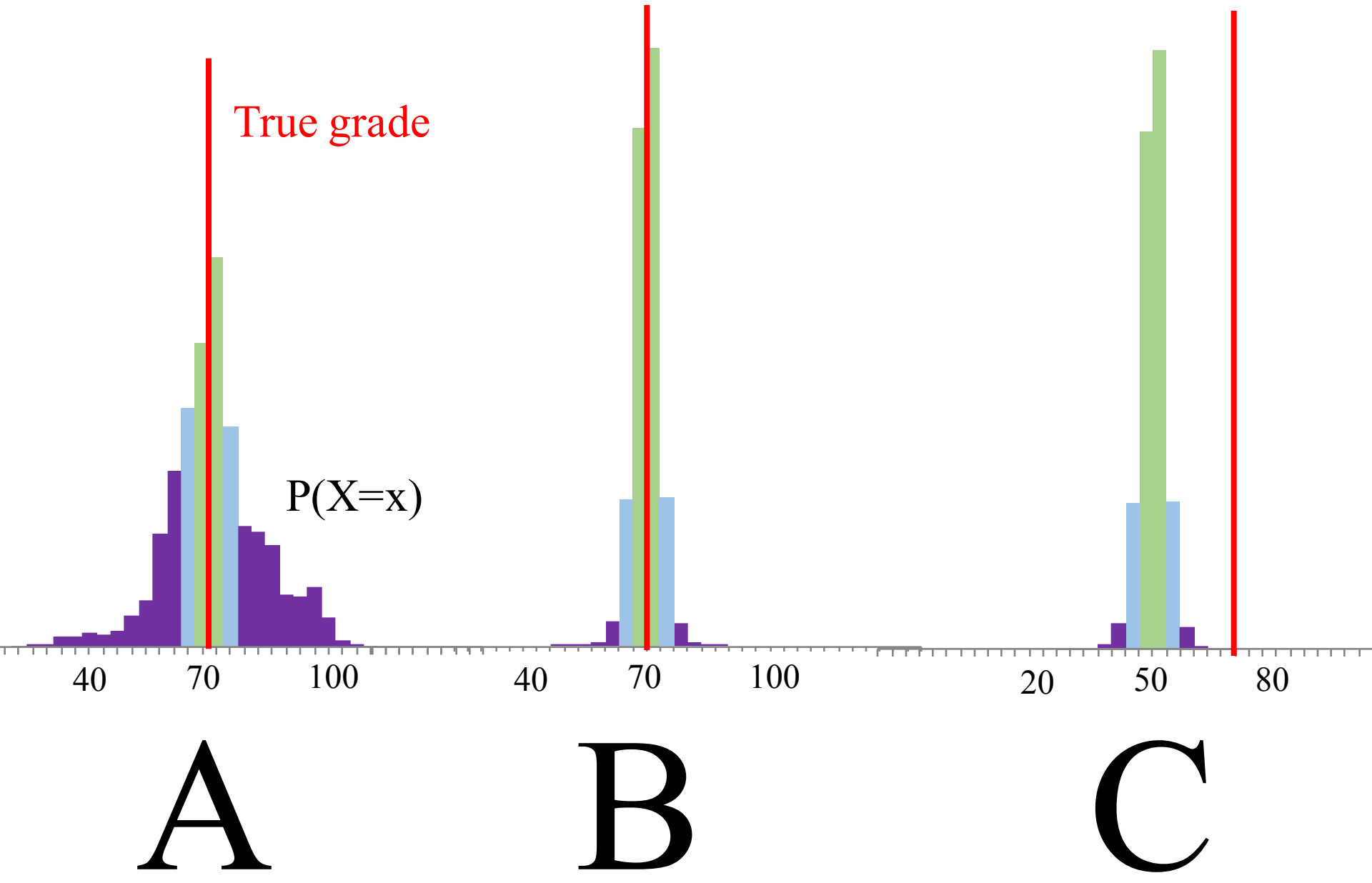
Fundamental Properties



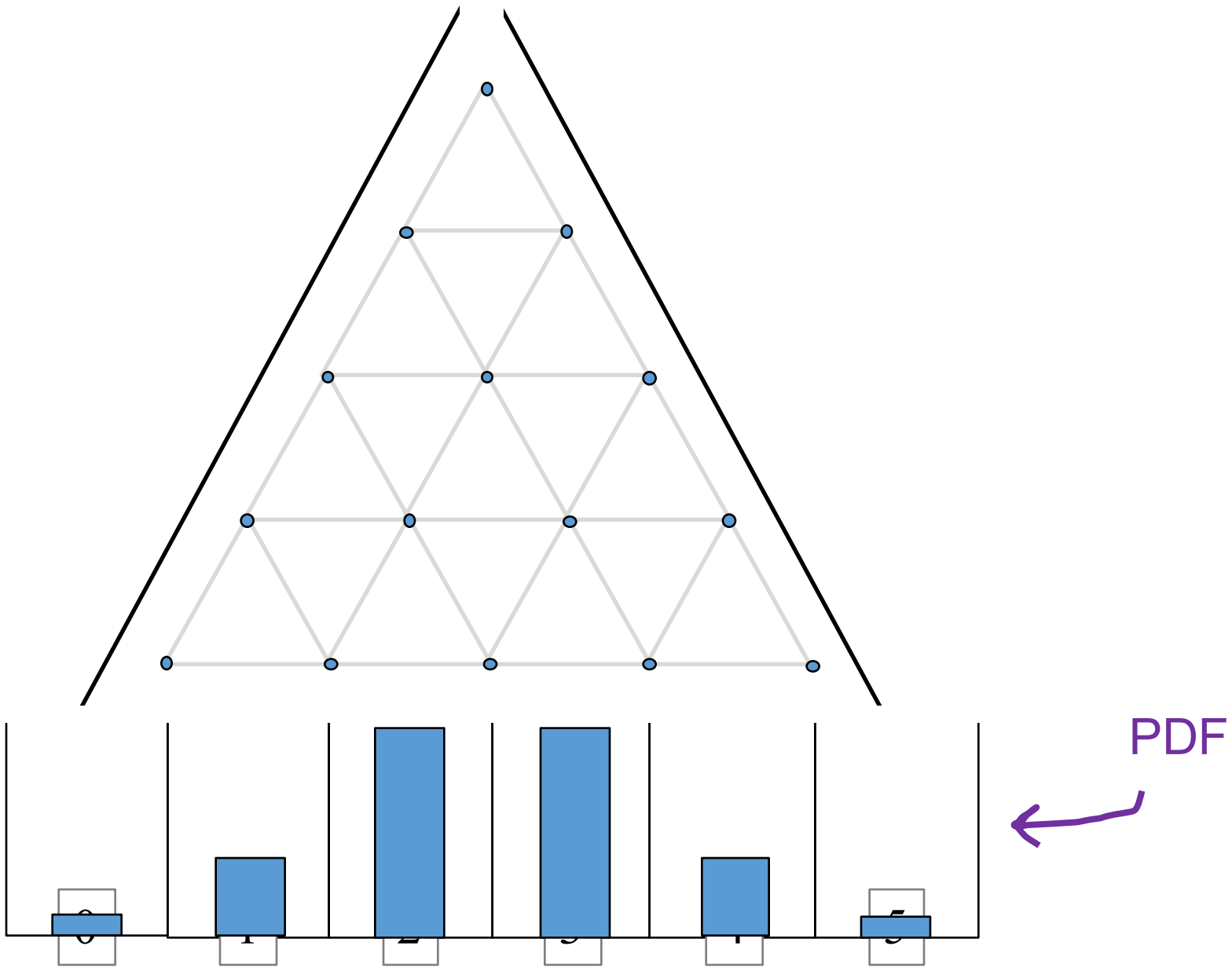
St Petersburg

- 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

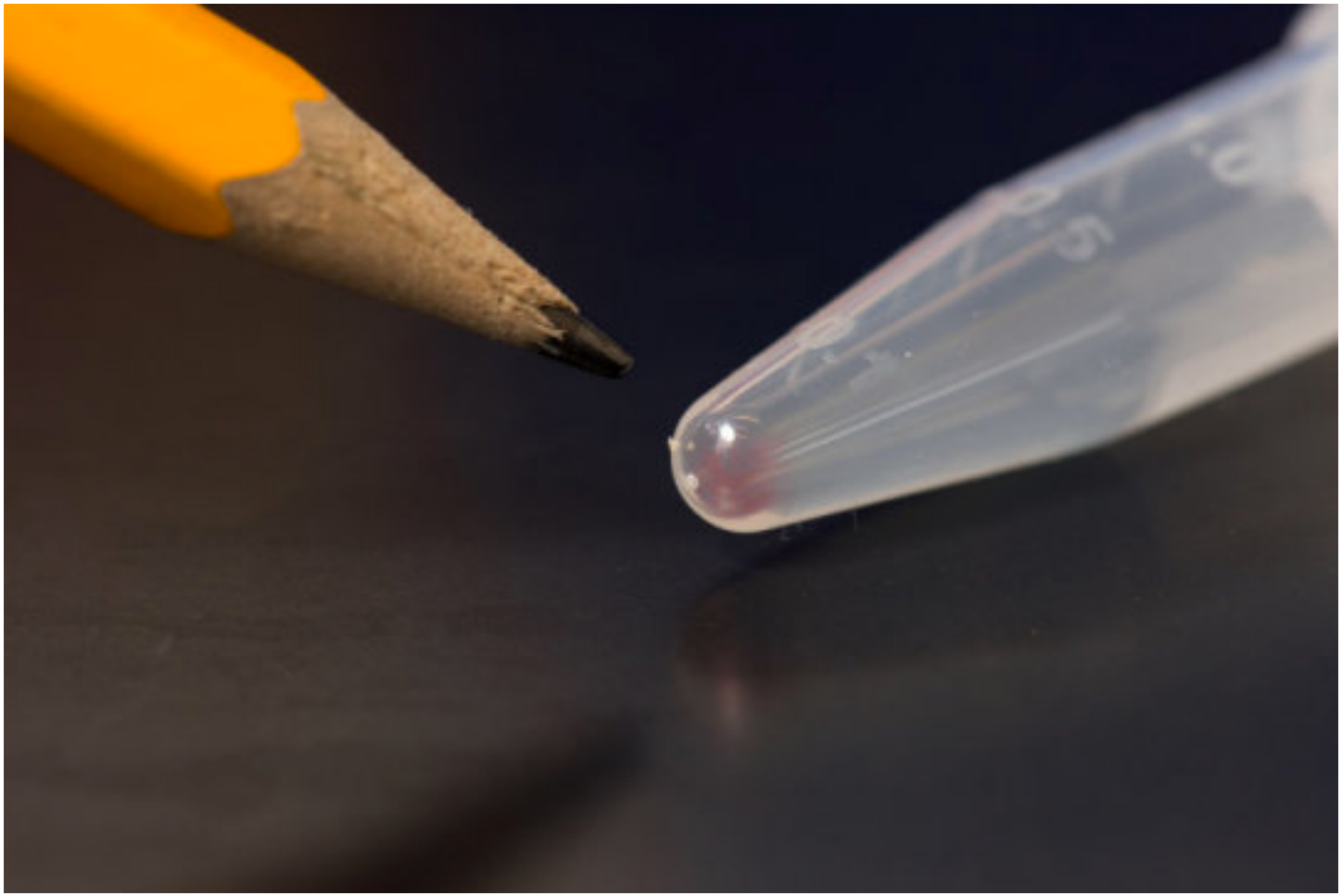


Binomial

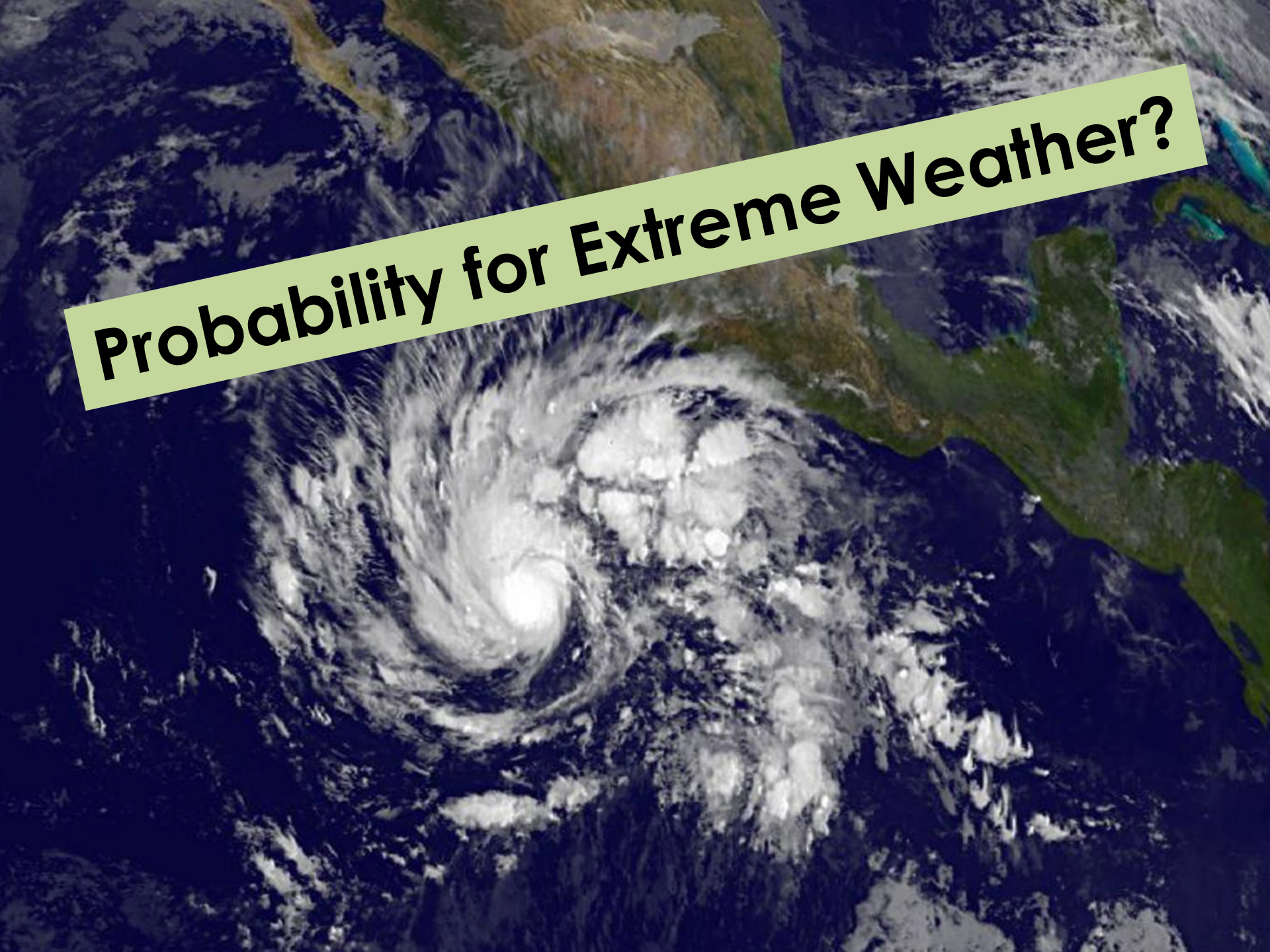



1001

Storing Data on DNA



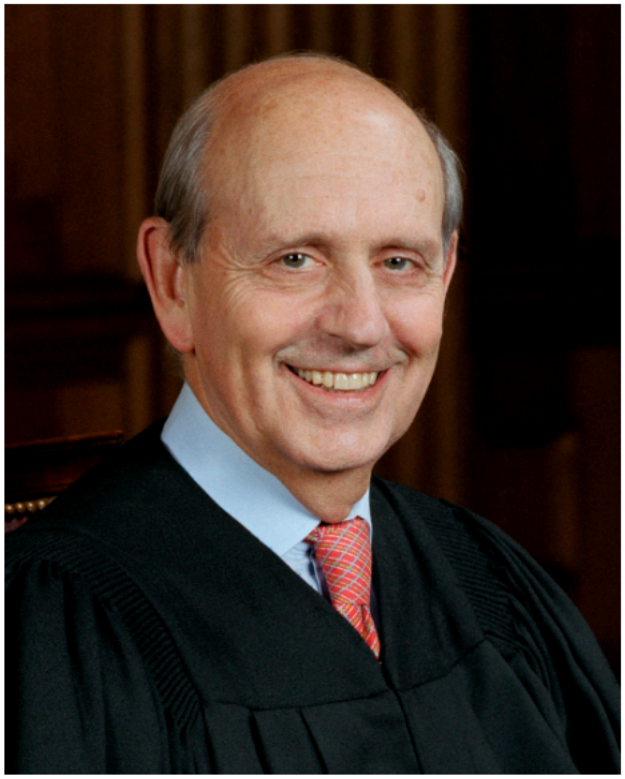
All the movies, images, emails and other digital data from more than 600 smartphones (10,000 gigabytes) can be stored in the faint pink smear of DNA at the end of this test tube.


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.

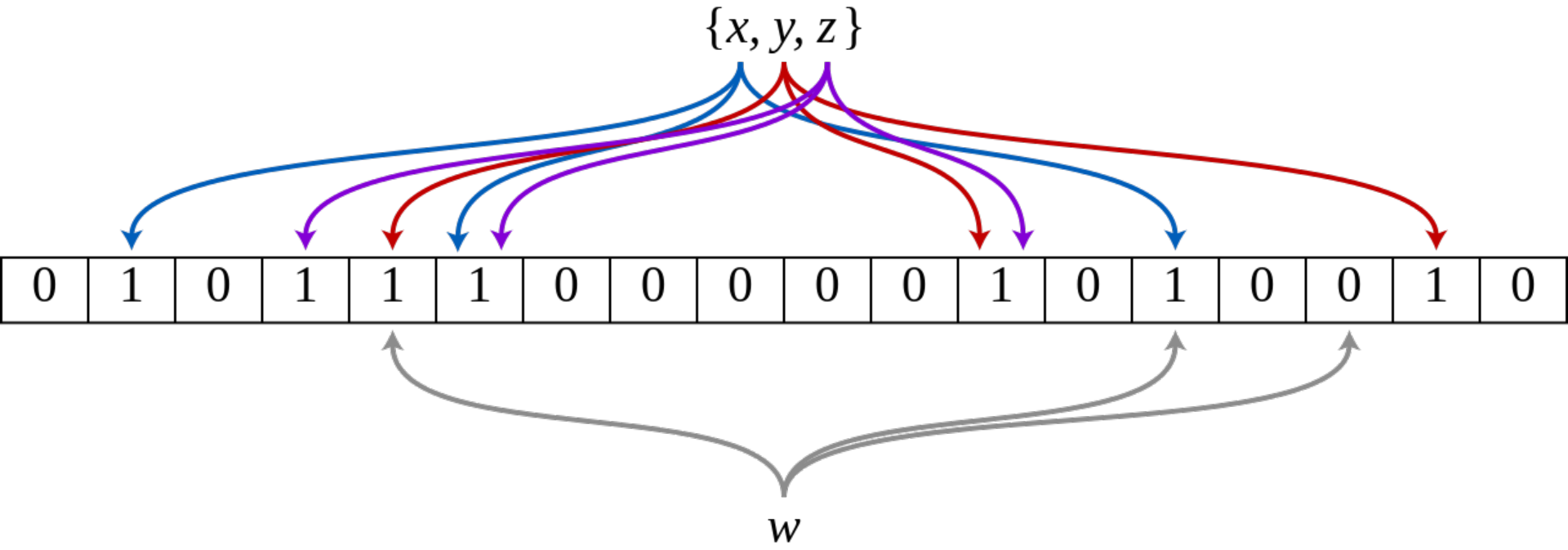
Representative Juries



Simulate

Simulation:

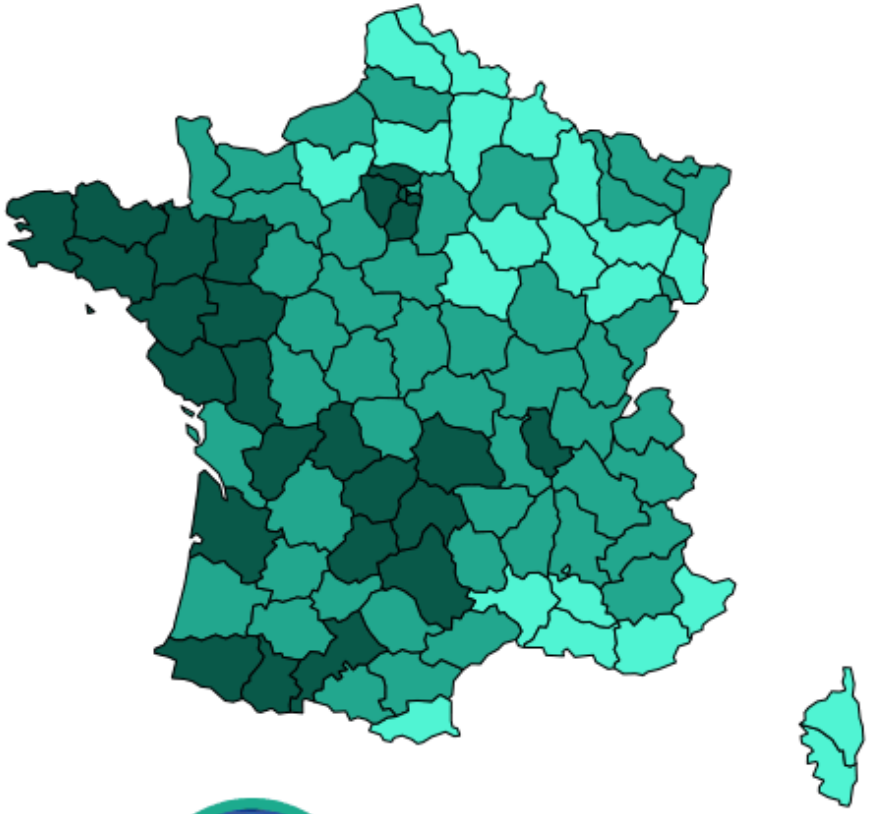
Bloom Filter



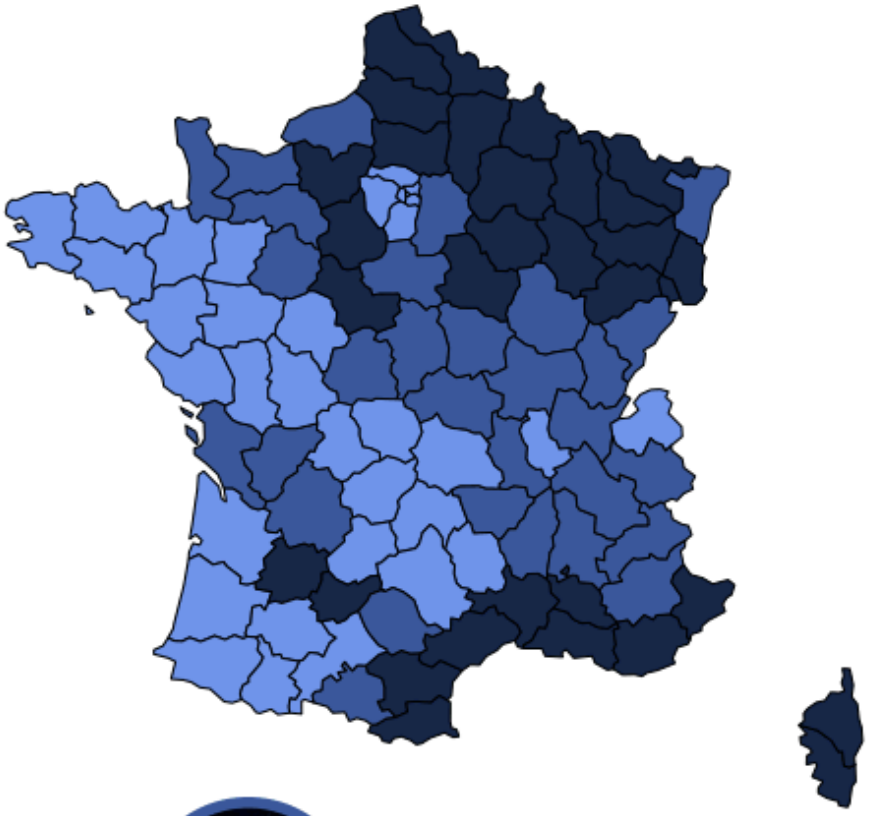
Election Prediction

Macron polled strongly in the west, while Le Pen fared better in the north-east and south coast

The two remaining presidential candidates' vote share mapped



Less than 20%
20-25%
More than 25%



Less than 20%
20-25%
More than 25%

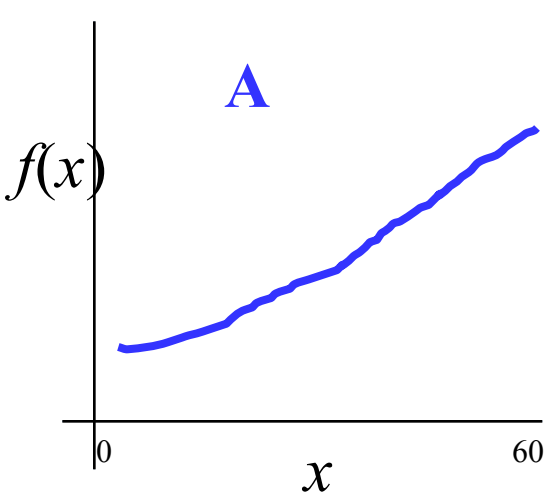
Source: French government

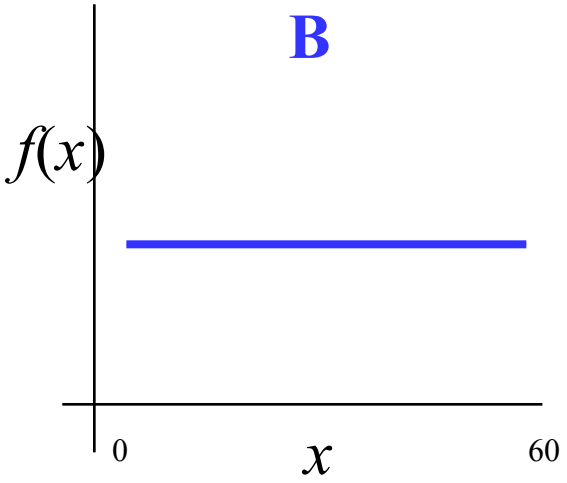
BBC

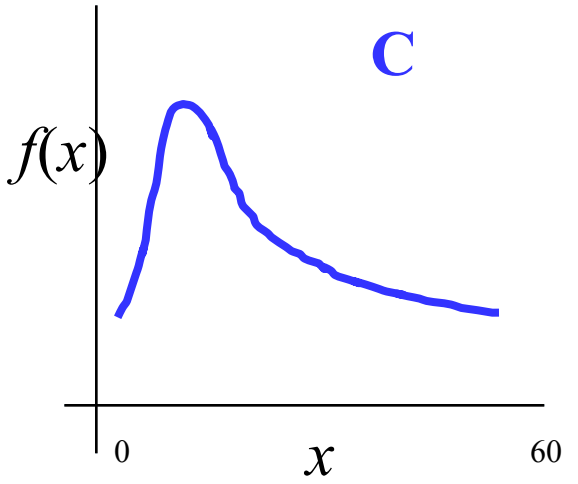
random()?

Probability Density Function

Probability density functions articulate relative belief.
Let X be a random variable which is the # of minutes after 2pm that you arrive at a bus station:







Which of these represent that you think your arrival is more likely to be close to 3

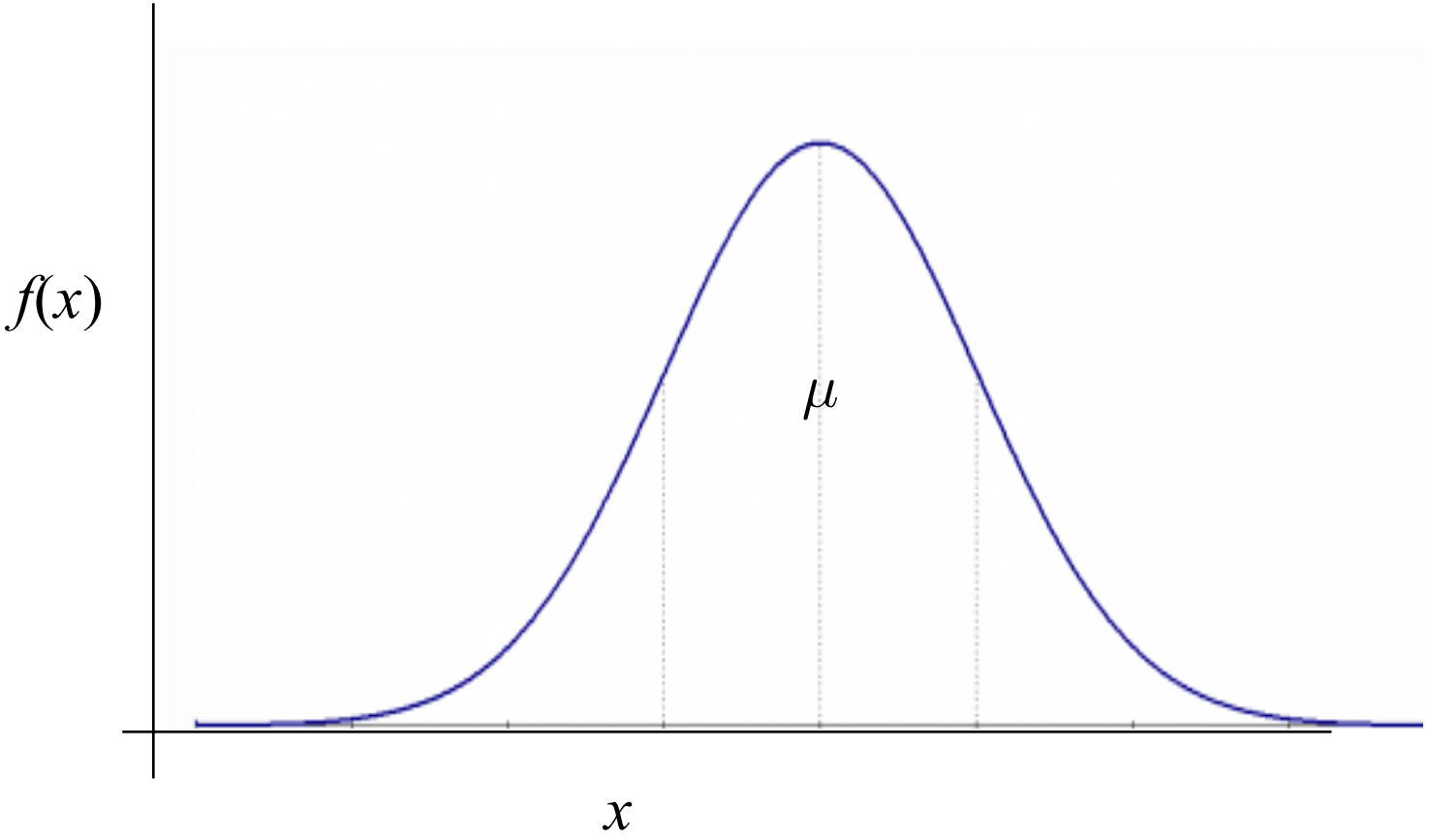
Integrals

*loving, not scary

Probability Density Function

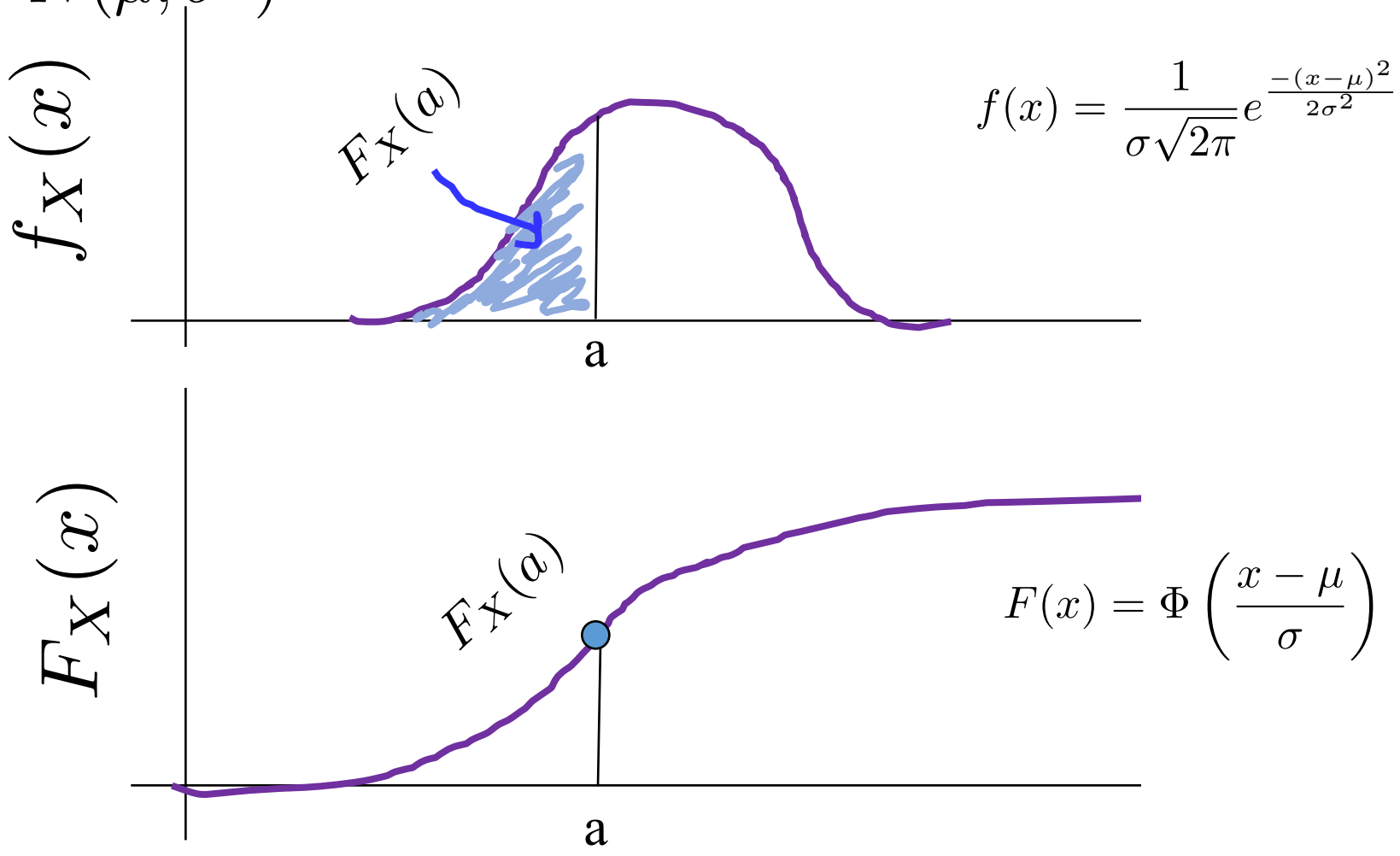
$$\mathcal{N}(\mu, \sigma^2)$$

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



PDF and CDF of a Normal

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



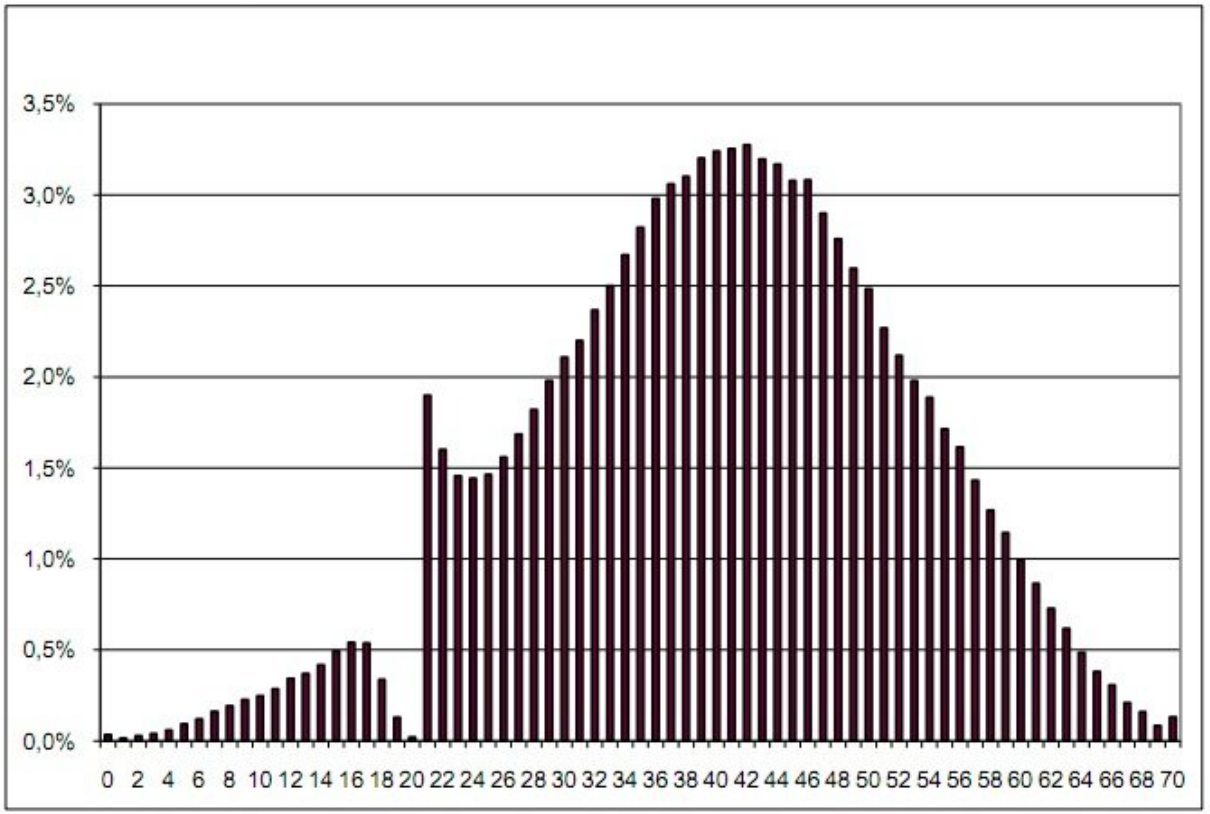
A CDF is the integral from –infinity to x of the PDF

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.

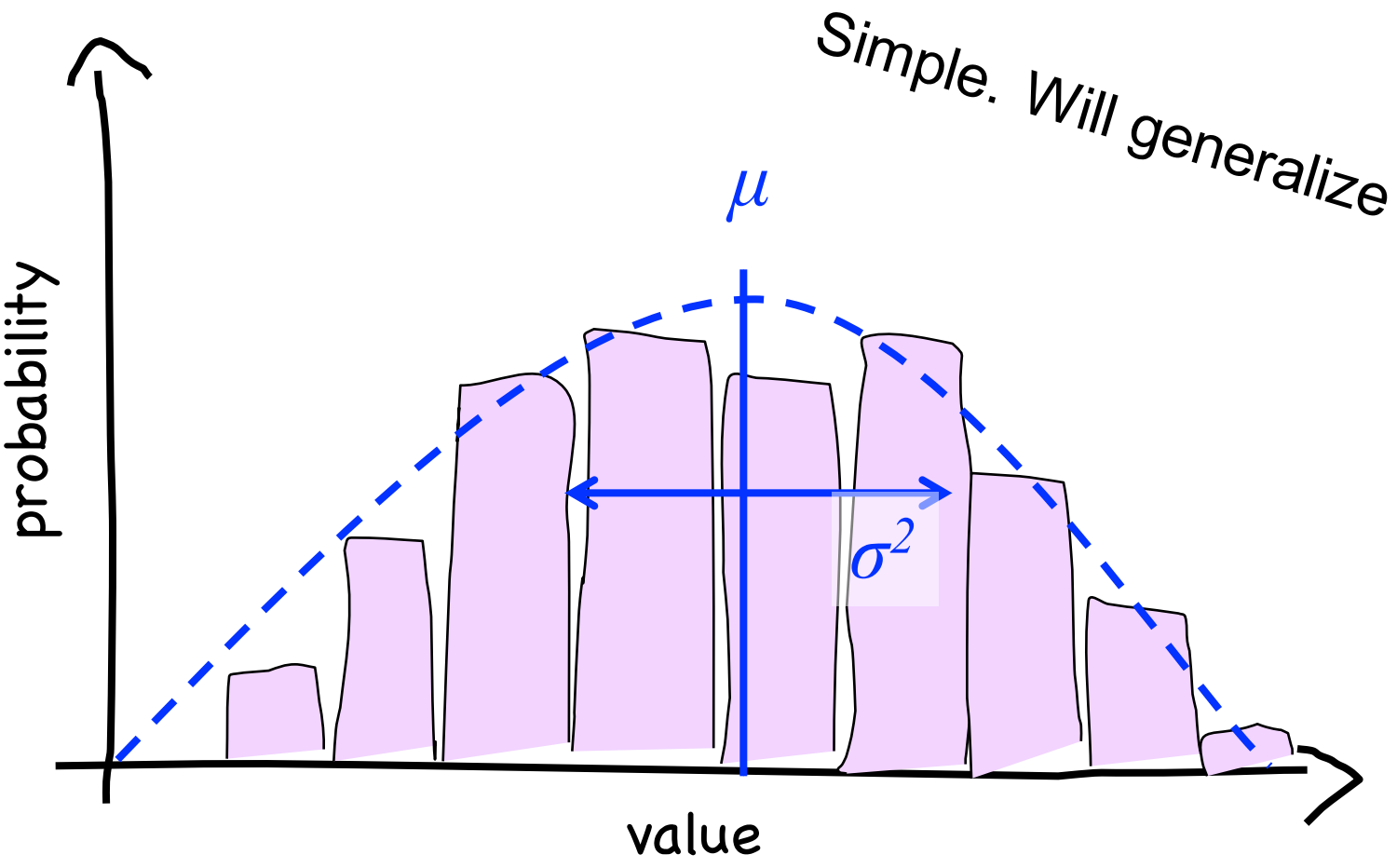
2.1. Poziom podstawowy



Wykres 1. Rozkład wyników na poziomie podstawowym

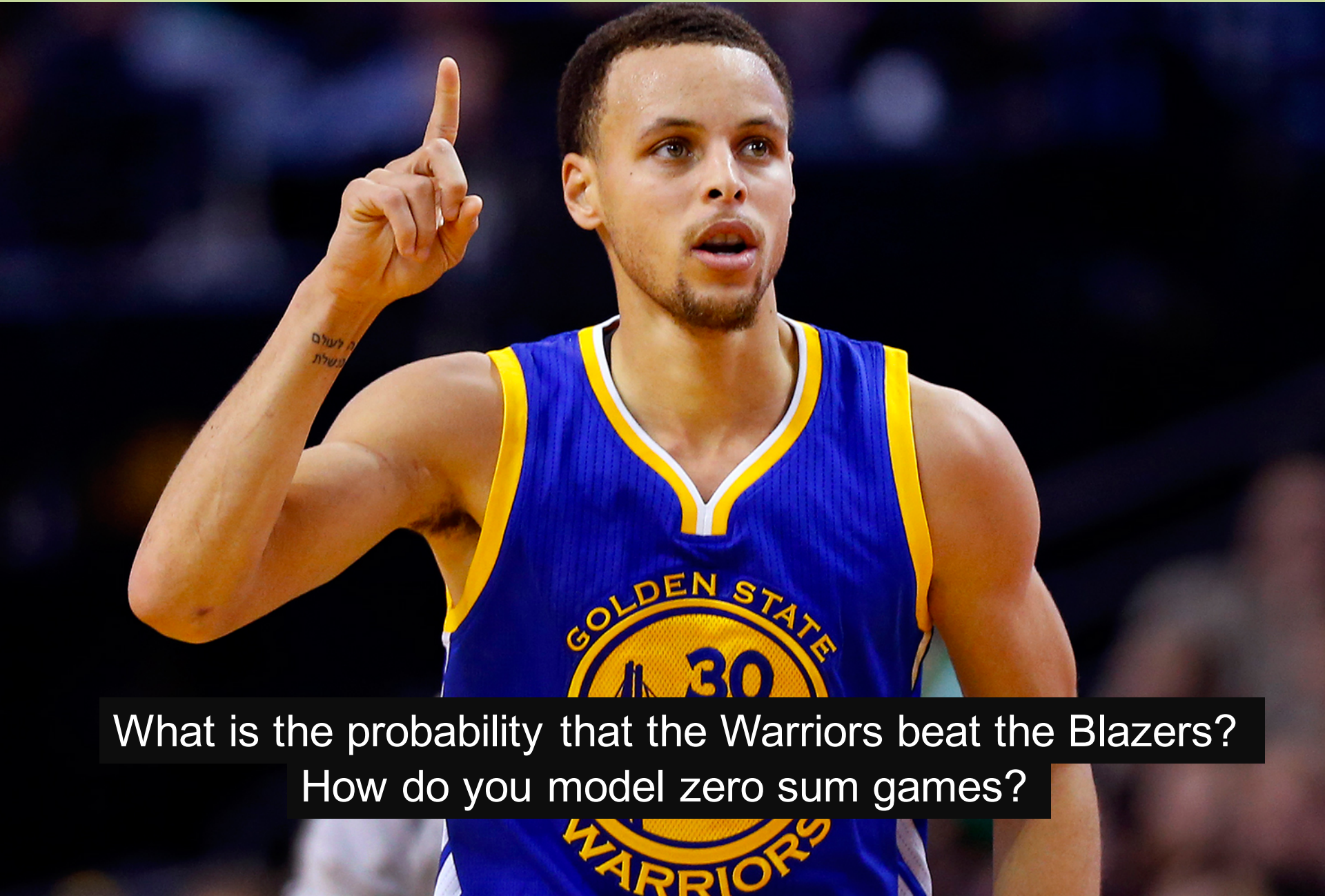
http://freakonomics.com/2011/07/07/another-case-of-teacher-cheating-or-is-it-just-altruism/comment-page-2/

Simplicity is Humble



* A Gaussian maximizes entropy for a given mean and variance

Will the Warriors Win?



What is the probability that the Warriors beat the Blazers?

How do you model zero sum games?

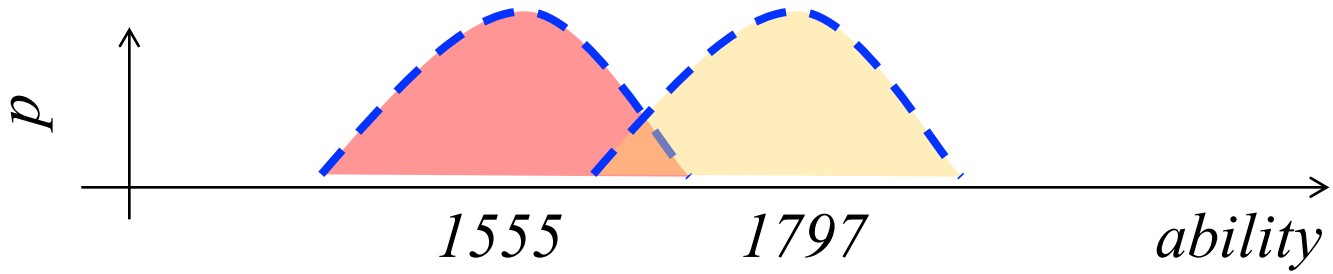
ELO Ratings

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

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



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

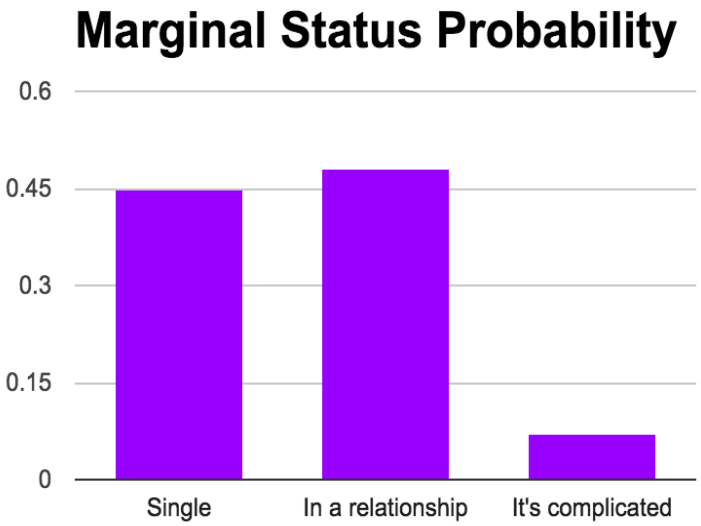
Joint Distributions

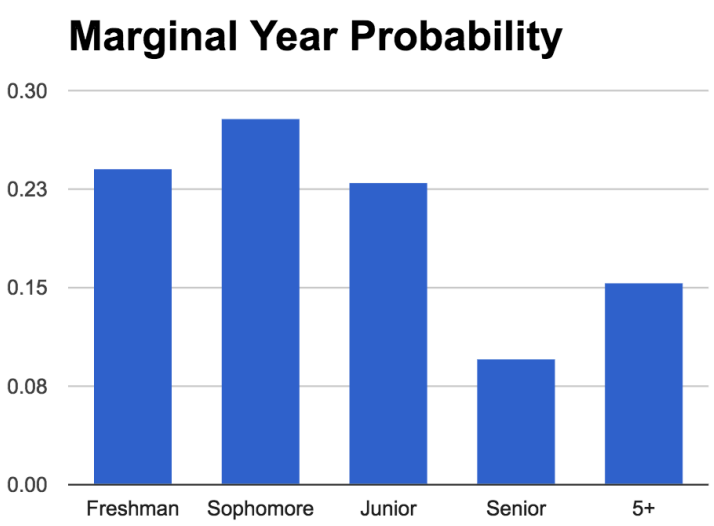
It's Complicated Demo

Go to this URL: **https://goo.gl/jCMY18**

Probability Table

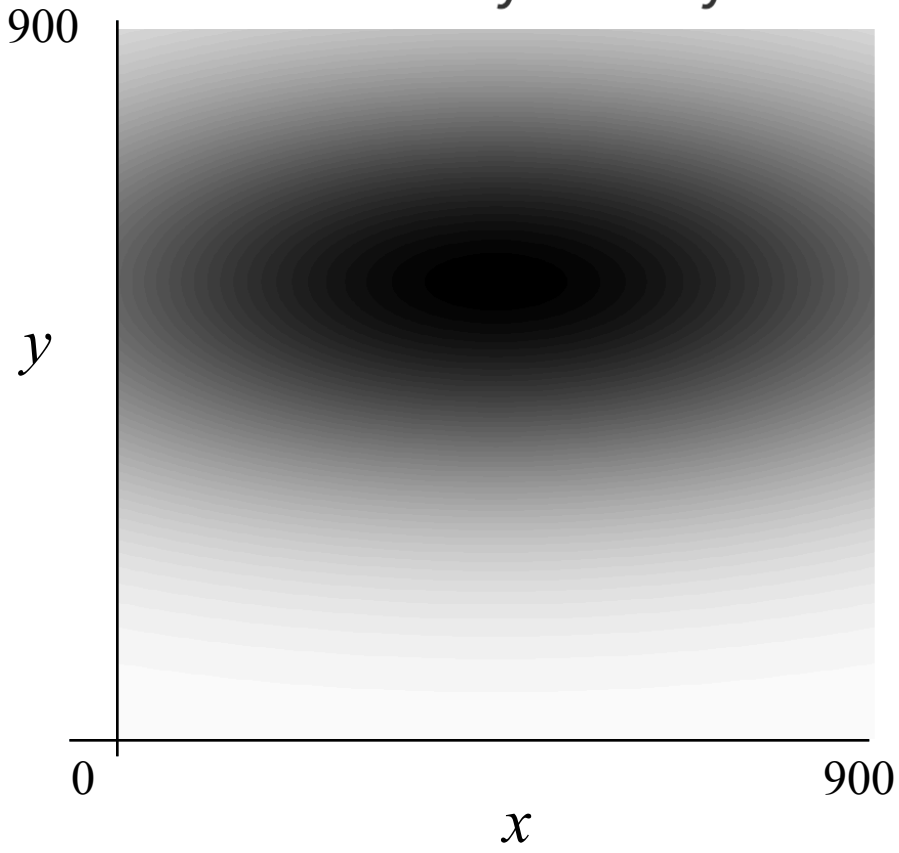
Joint Probability Table				
	Single	In a relationship	It's complicated	**Marginal Year**
Freshman	0.13	0.09	0.02	0.24
Sophomore	0.16	0.10	0.02	0.28
Junior	0.12	0.10	0.02	0.23
Senior	0.01	0.09	0.00	0.10
5+	0.03	0.12	0.01	0.15
Marginal Status	0.45	0.48	0.07	





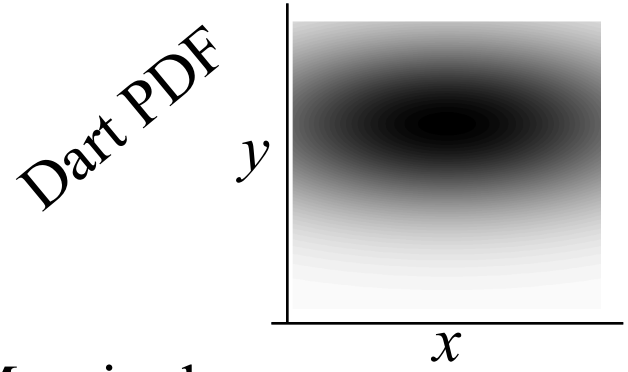
Joint Dart Distribution

Dart Probability Density

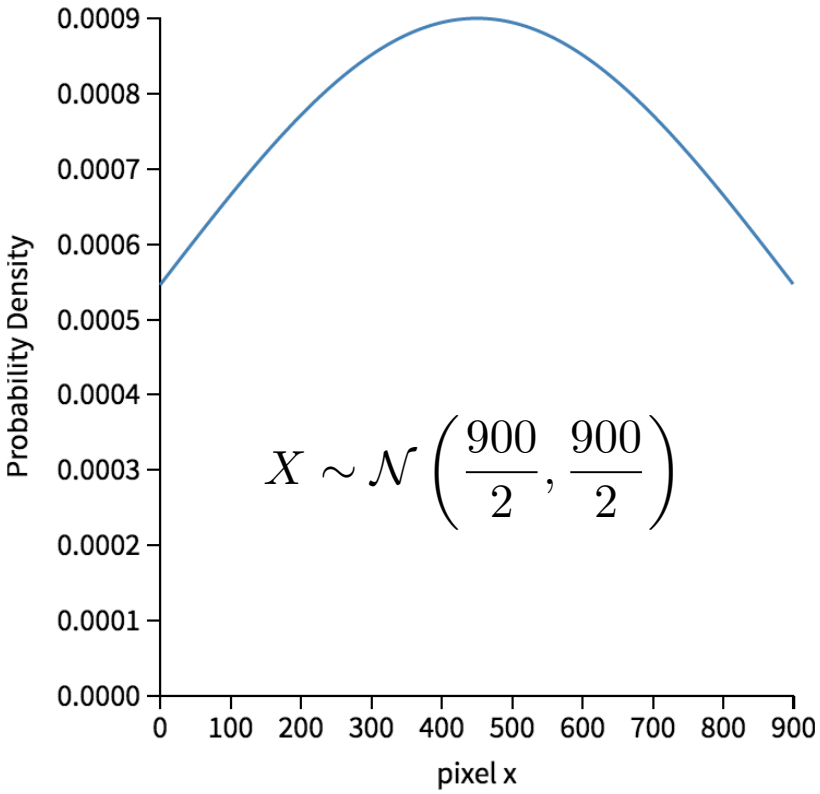


Dart Results

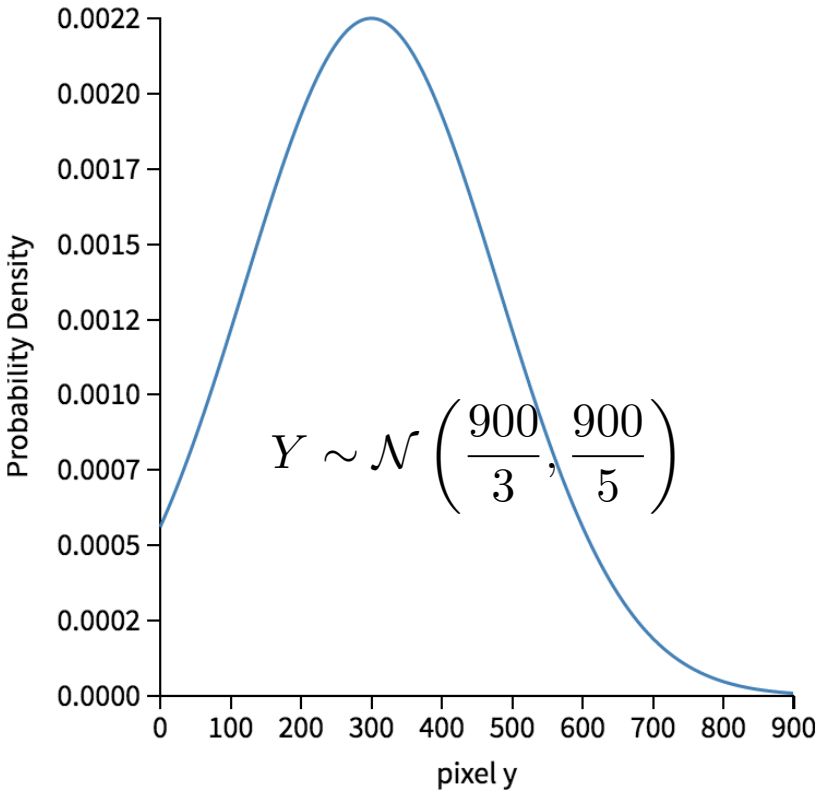
Darts!



X-Pixel Marginal



Y-Pixel Marginal



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

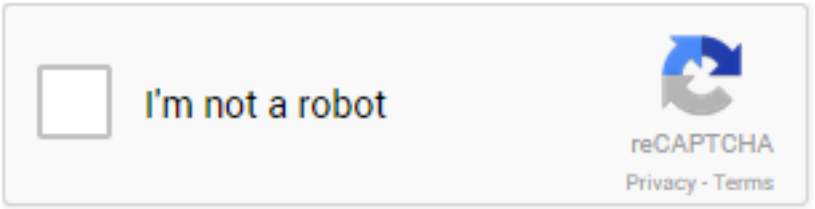
It's a Multinomial!

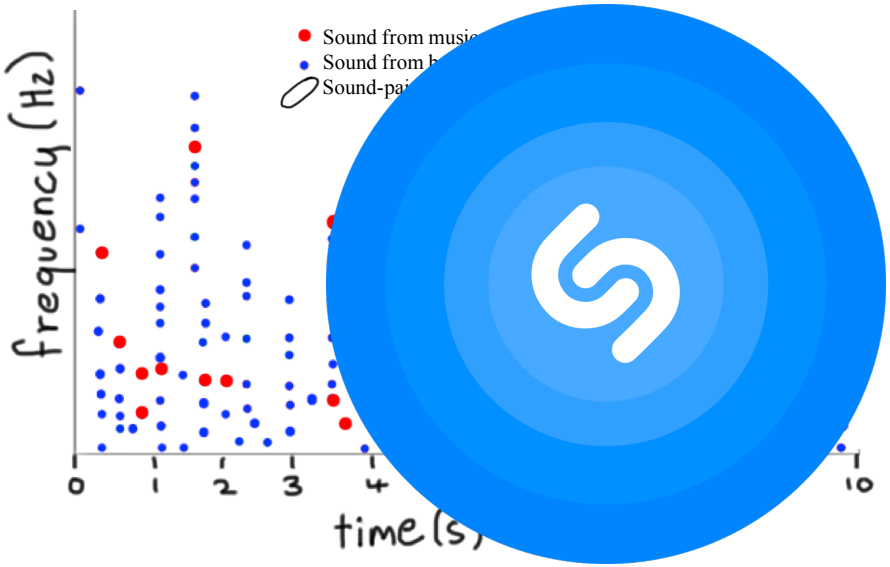
$$P\left(\begin{array}{l}\text{Viagra} = 2\\ \text{Free} = 2\\ \text{Risk} = 1\\ \text{Credit-card: } 2\\ \ldots\\ \text{For} = 2\end{array} \middle| \text{spam}\right) = \frac{n!}{2!2!\ldots 2!} p_{\text{viagra}}^2 p_{\text{free}}^2 \cdots p_{\text{for}}^2$$

Probability of seeing this document | spam

The probability of a word in spam email being viagra

Midterm = Education Time





Convolution

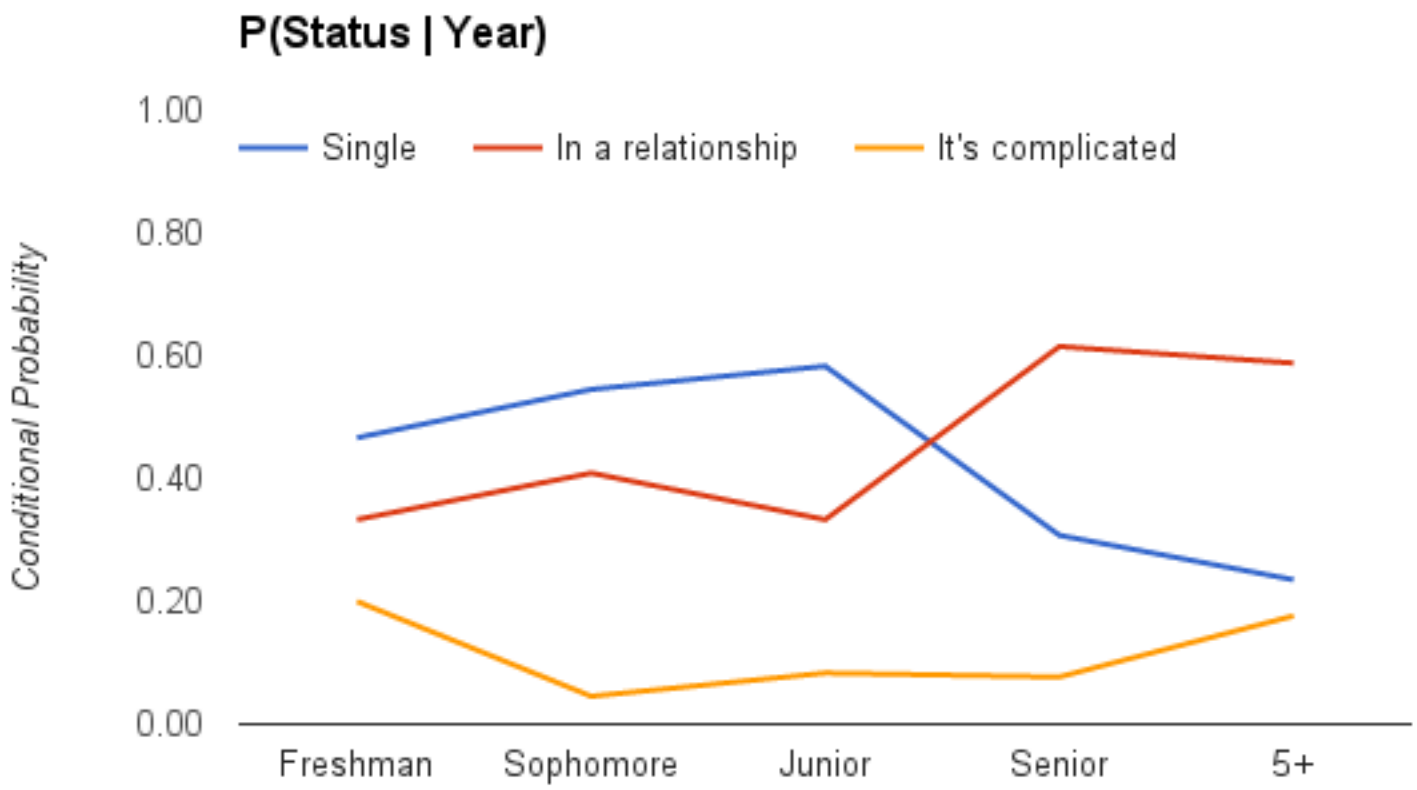
Big deal lemma: first stated without proof

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

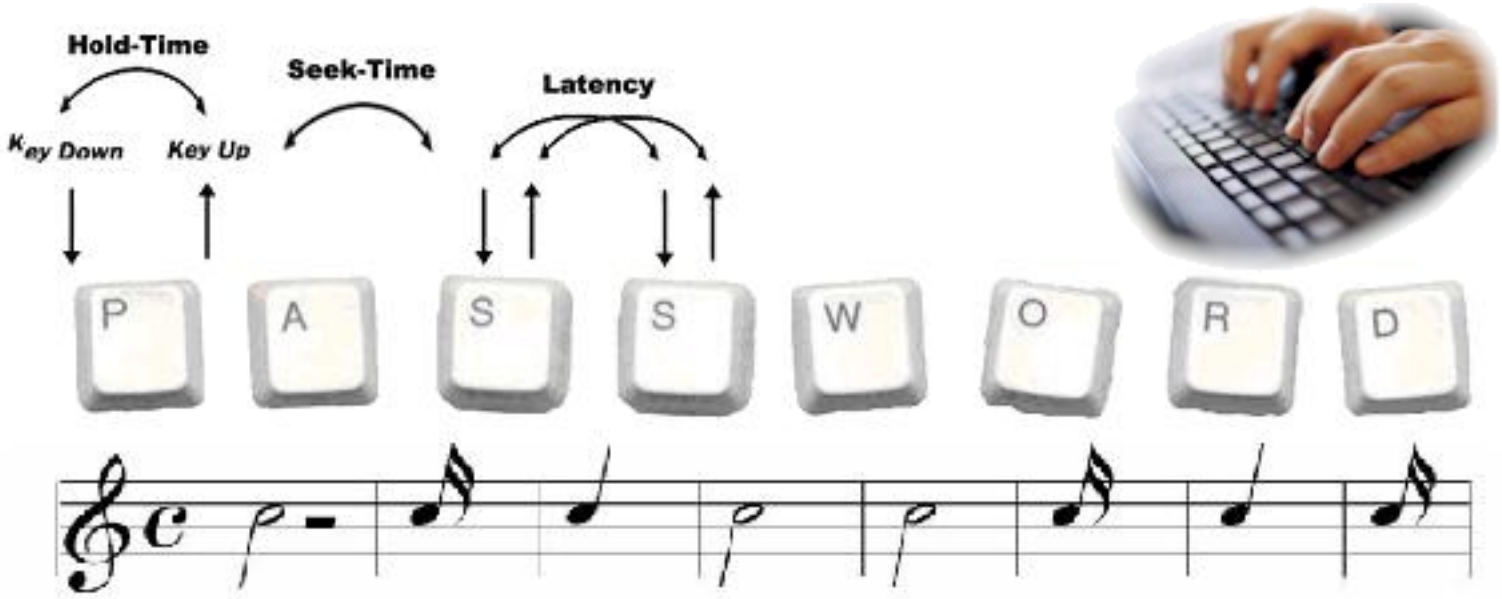
Generalized: $$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Holds regardless of dependency between X_i's

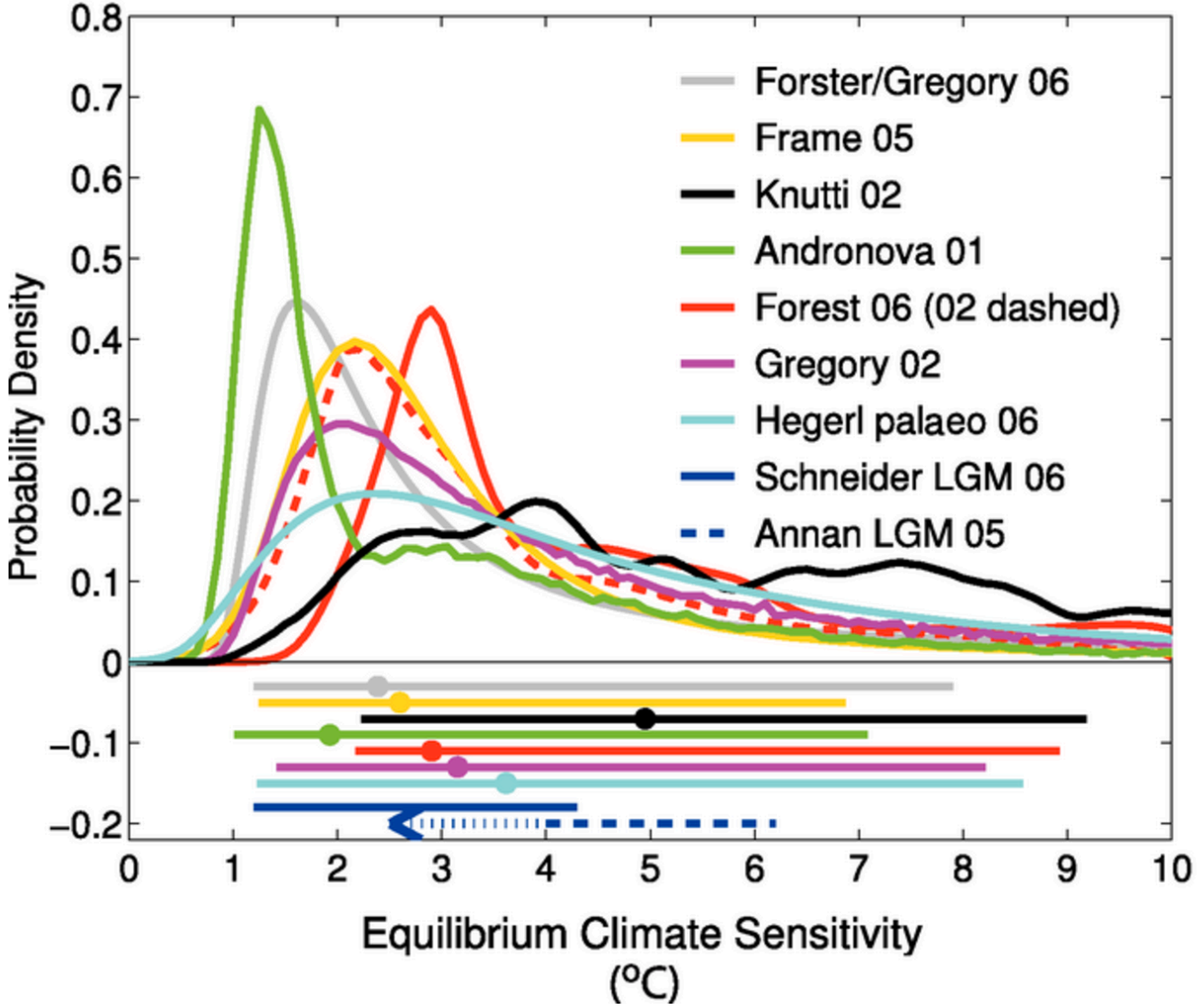
Relationship Status

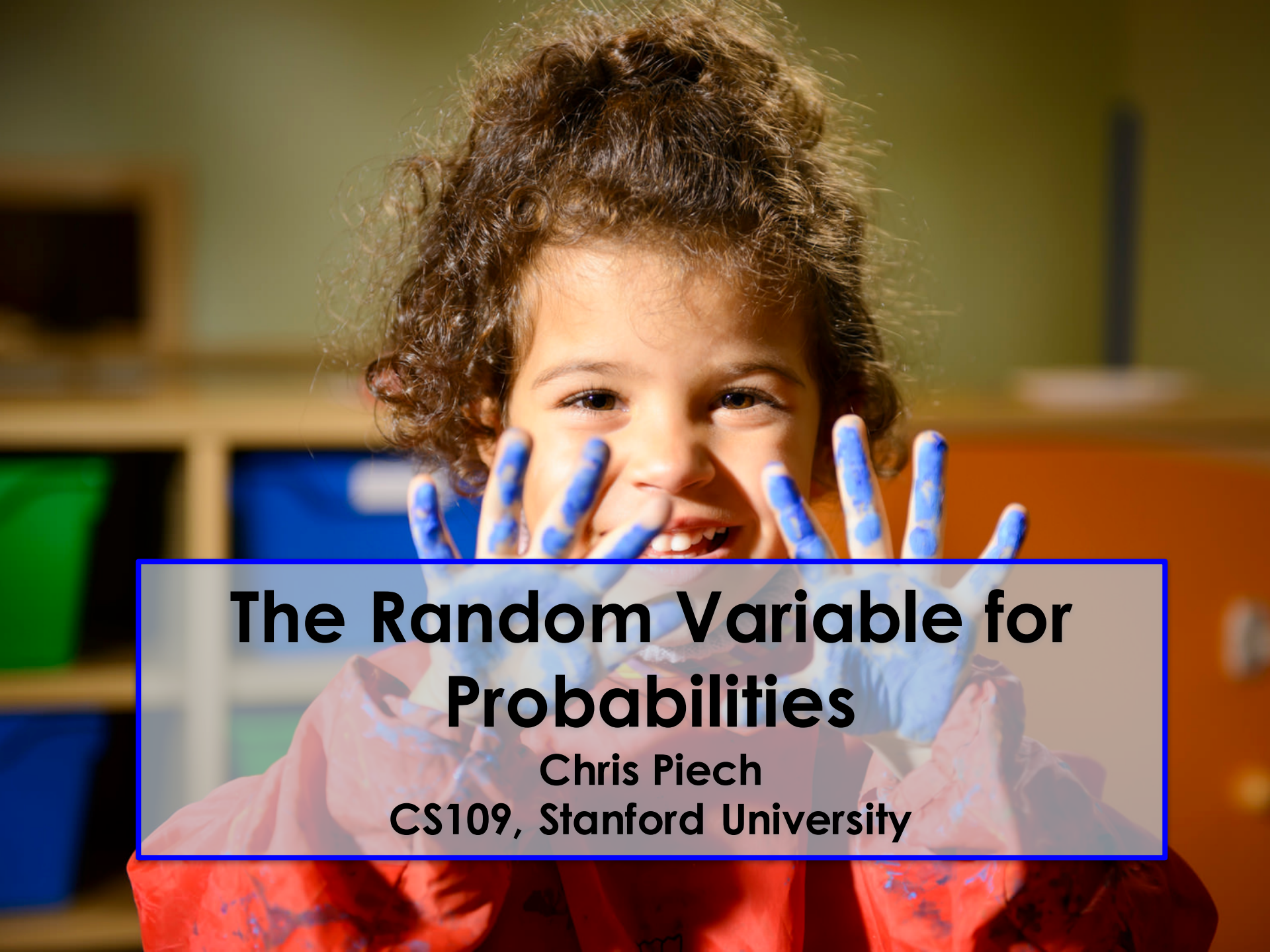


Biometric Keystroke



Climate Sensitivity





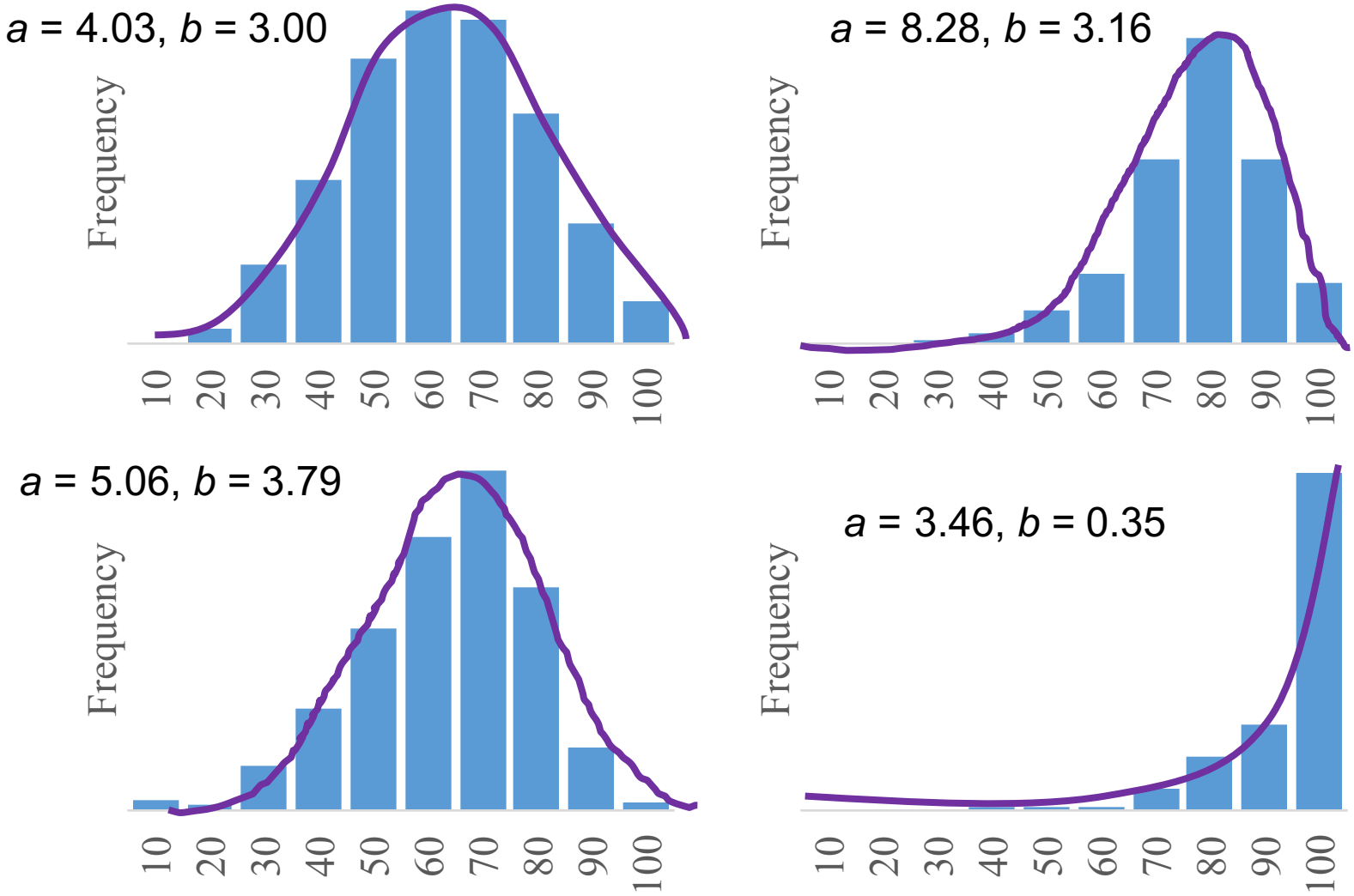
The Random Variable for Probabilities

Chris Piech
CS109, Stanford University

Flip a Coin With Unknown Probability

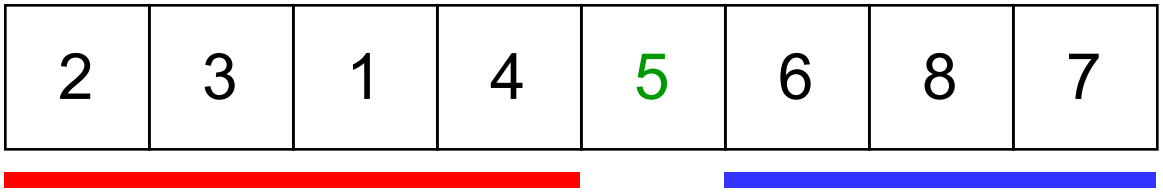
Demo

Assignment Grades



We have 2055 assignment distributions from grade scope

Recursive Insight



Partition array so:

- everything smaller than pivot is on left
- everything greater than or equal to pivot is on right
- pivot is in-between

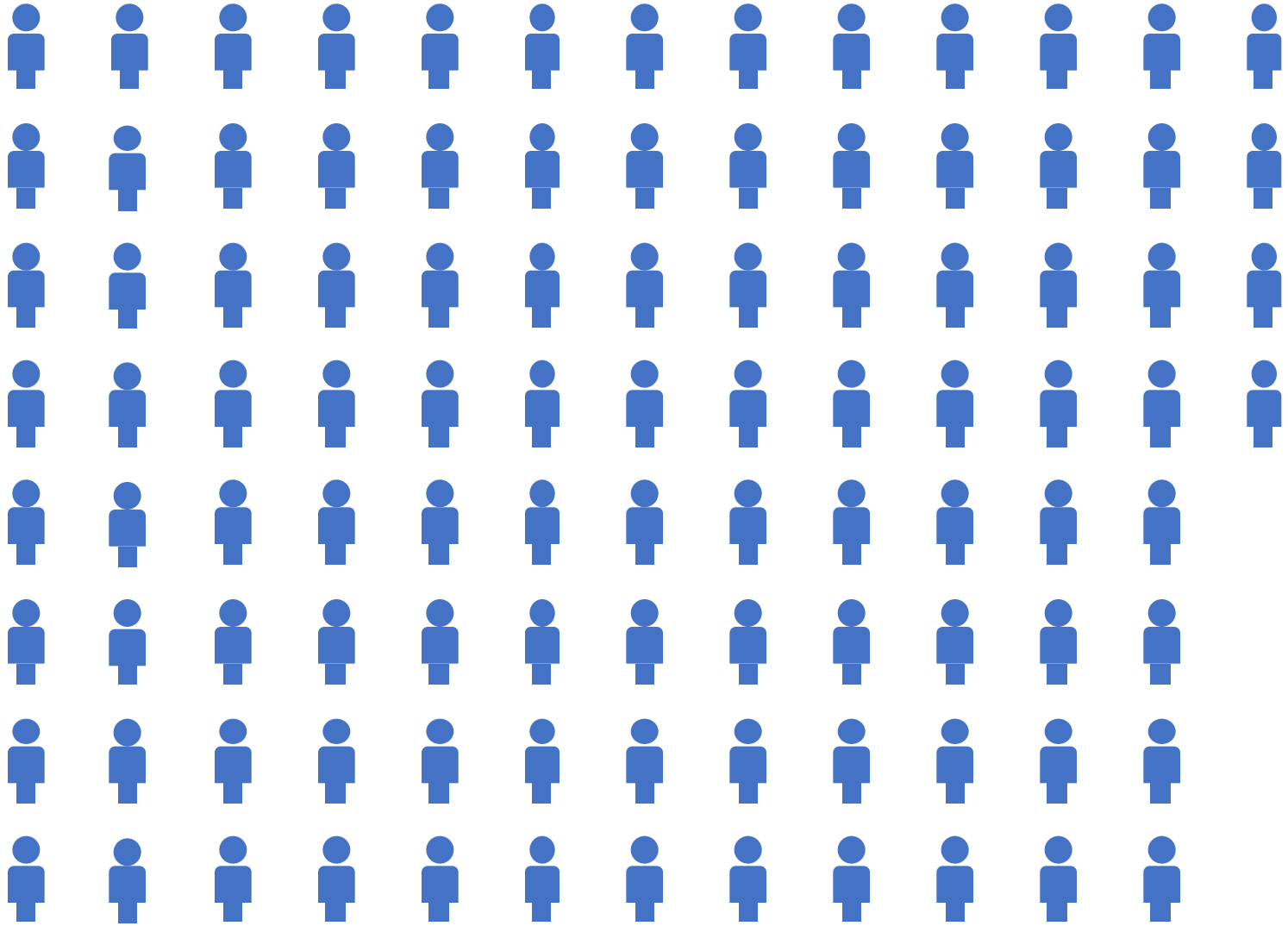
Machine Learning Example

- 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:

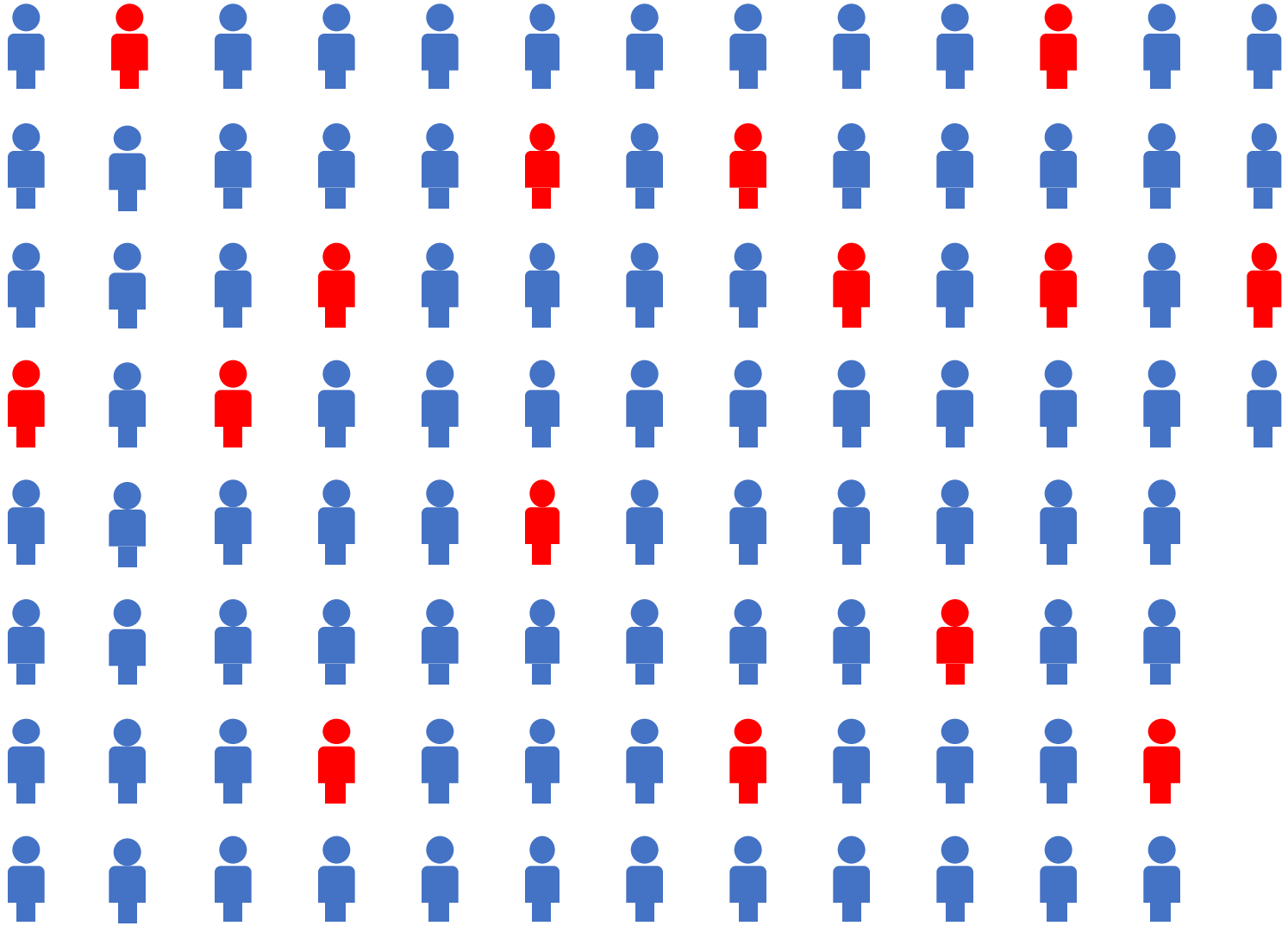
Happiness = {72, 85, 79, 91, 68, … , 71}

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

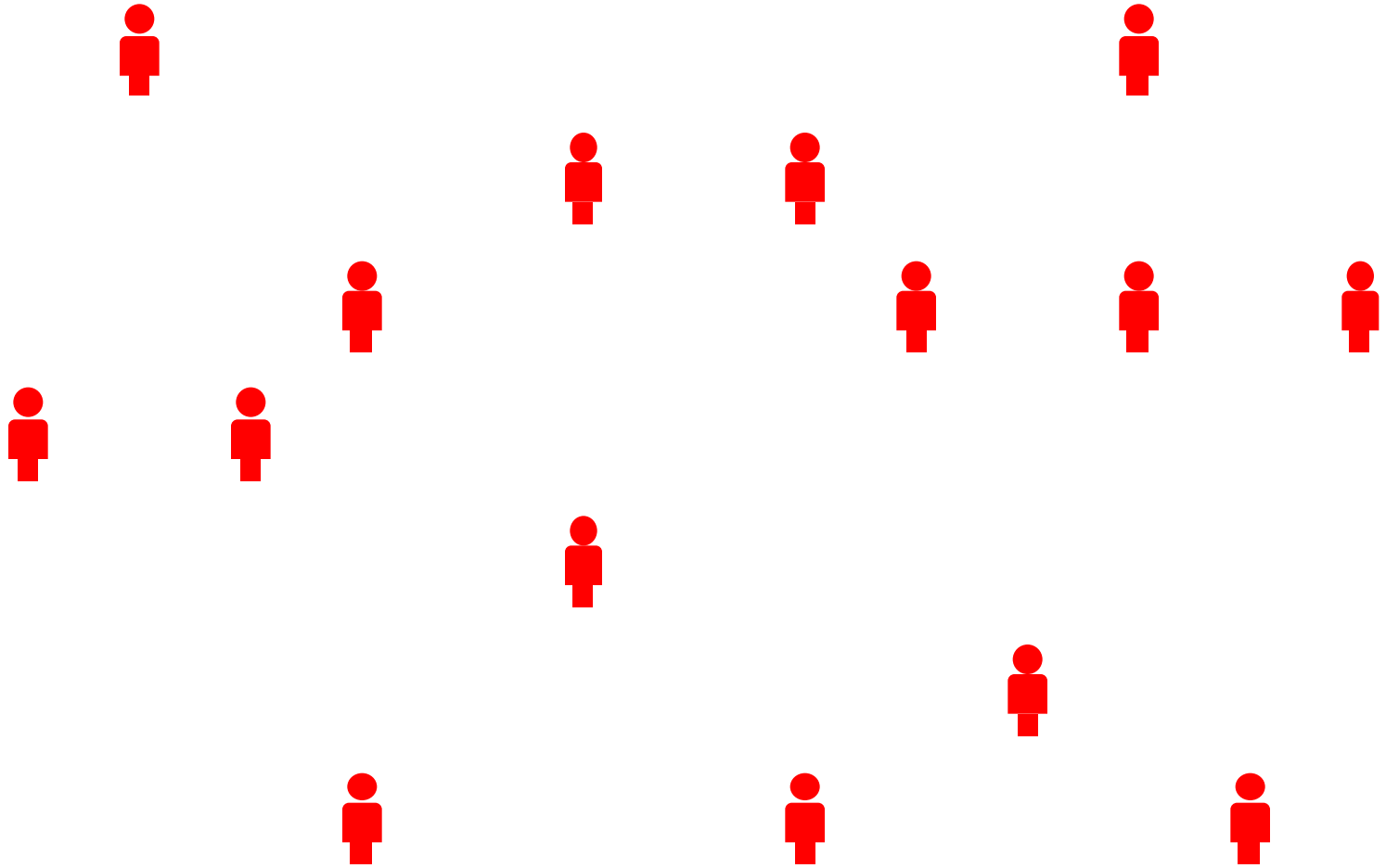
Population



Sample

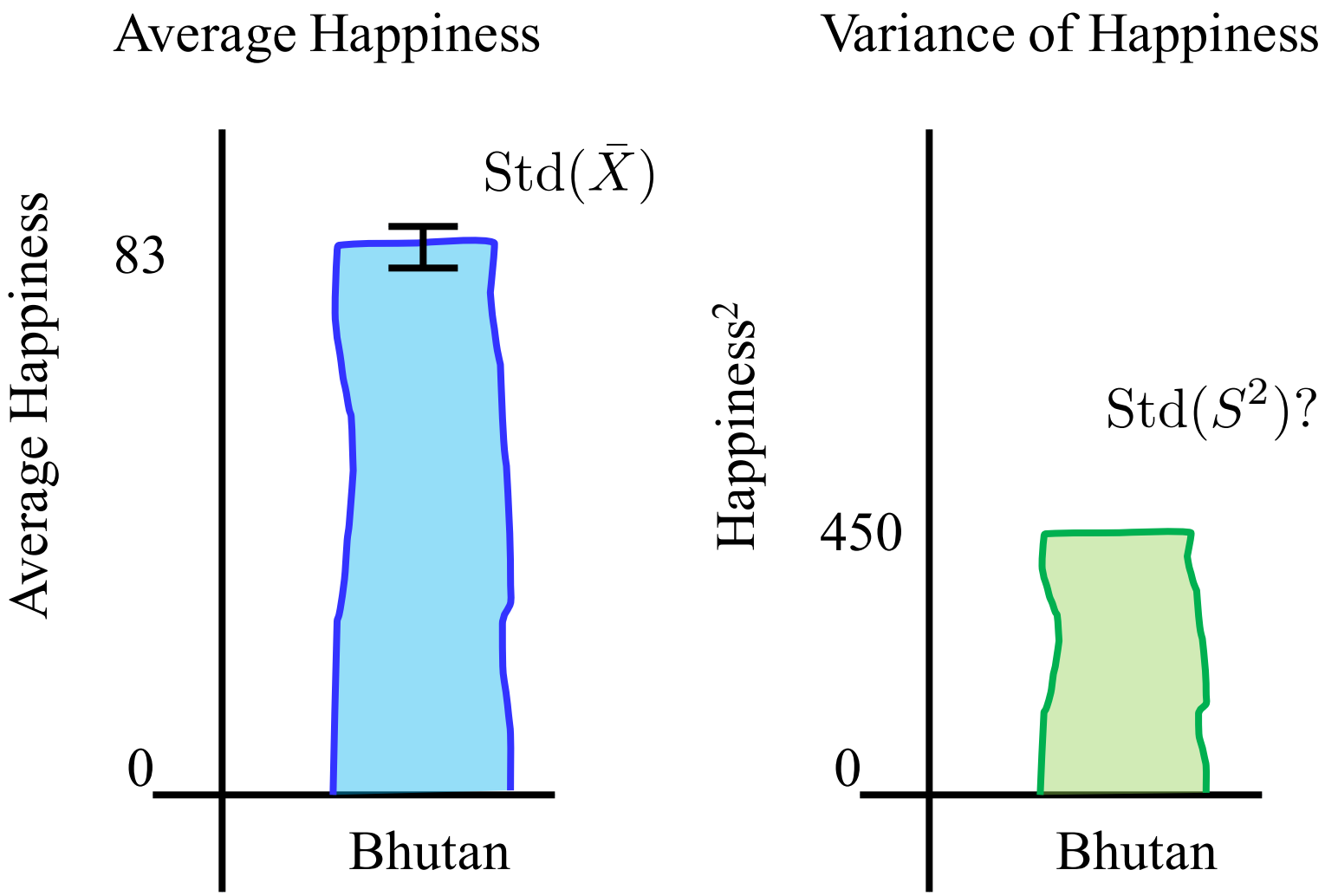


Sample



Collect one (or more) numbers from each person

Sample Mean



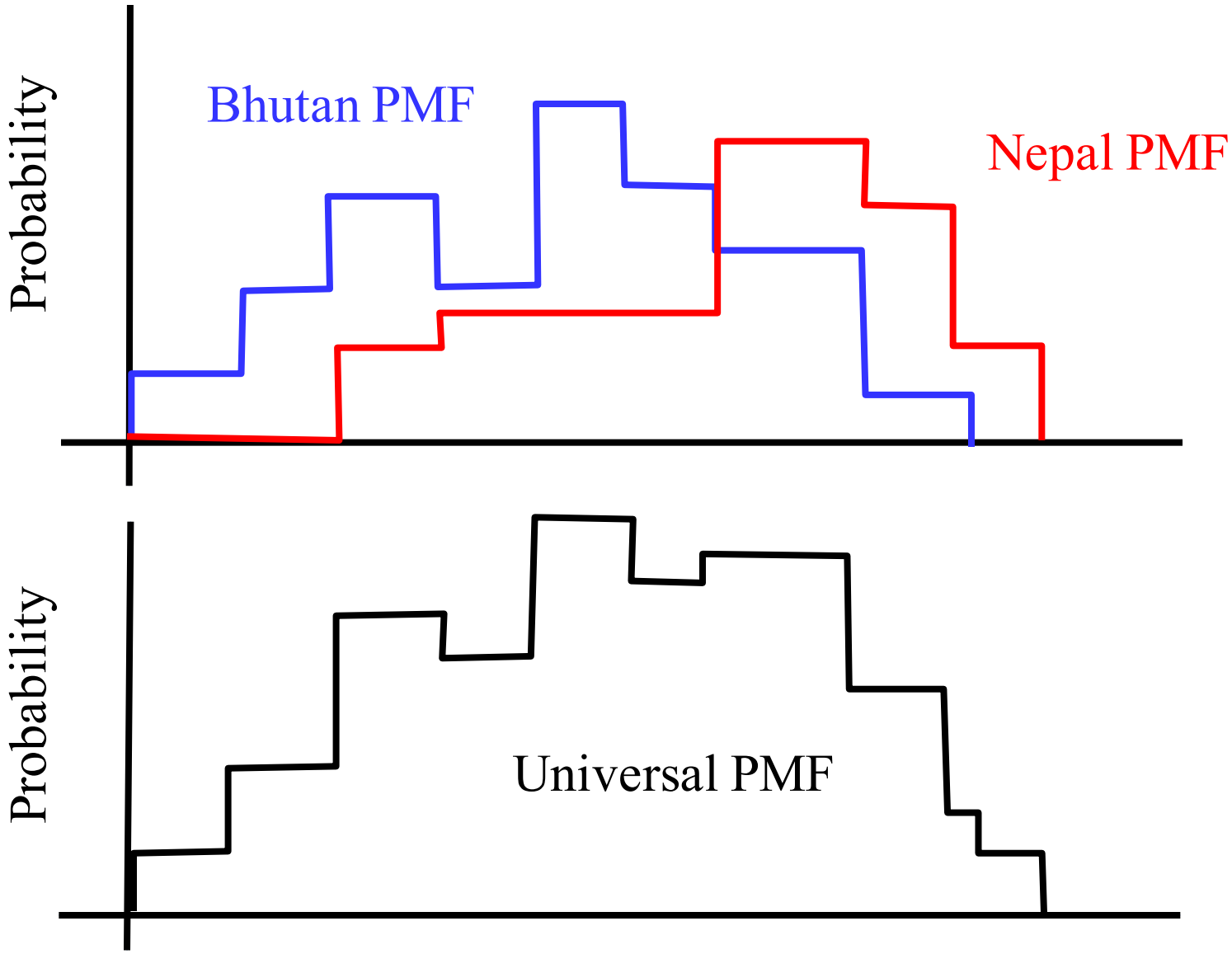
Claim: The average happiness of Bhutan is 83 ± 2

Bootstrap

Key Insight

You can estimate the PMF of the underlying distribution, from your sample.

Universal Sample



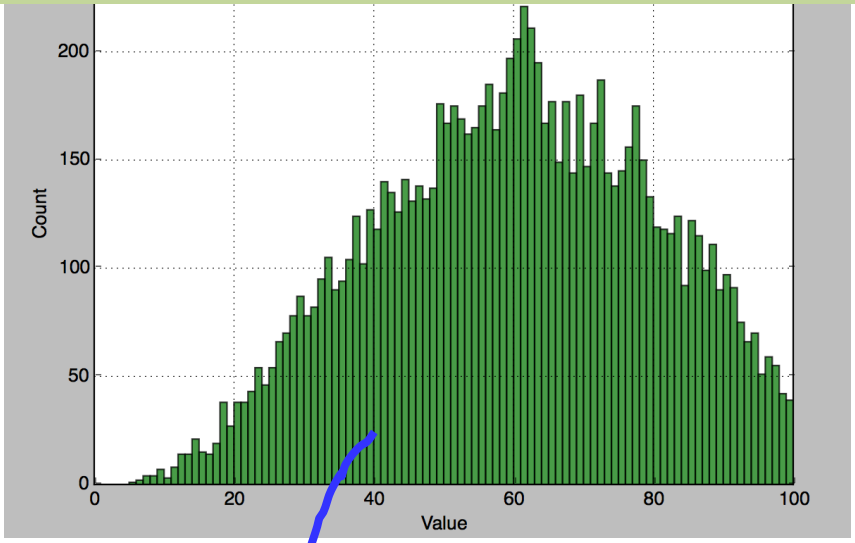
Inequalities

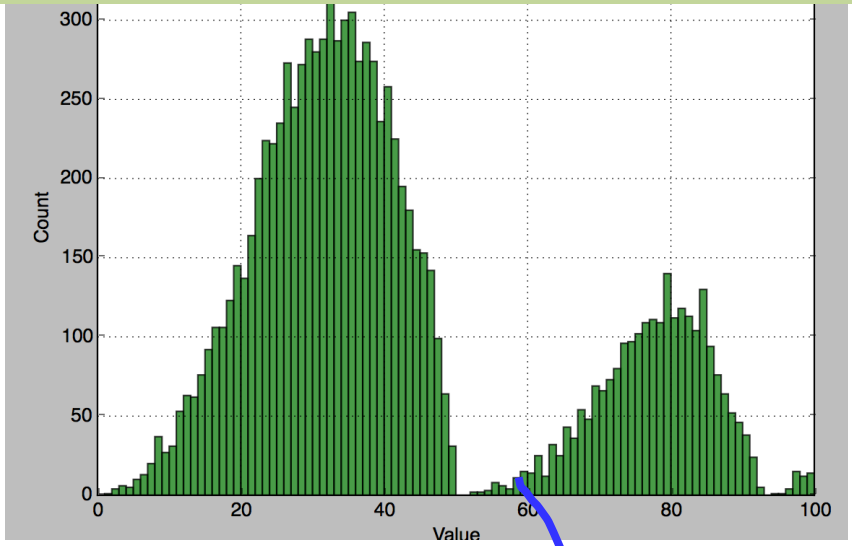
Mystery: Why is Binomial Normal?

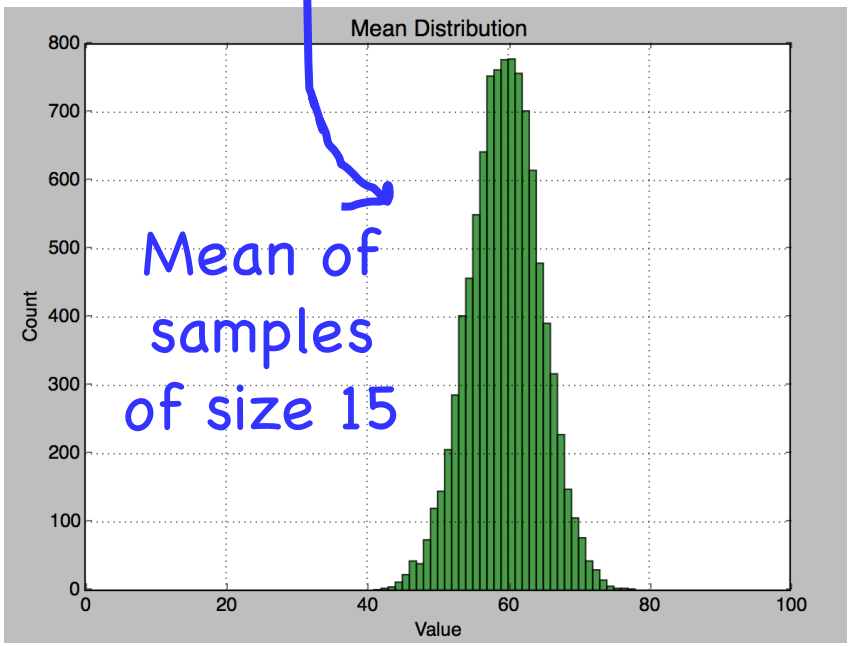
Mystery: Why is the sum of IID uniforms normal?

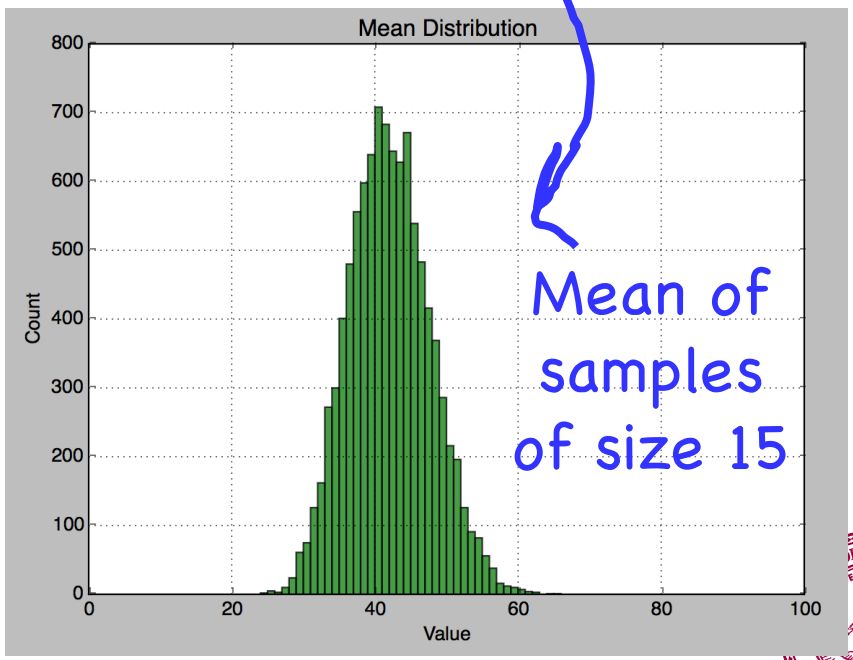
Mystery: Why is the mean of IID vars normal?

C.L.T. Explains This

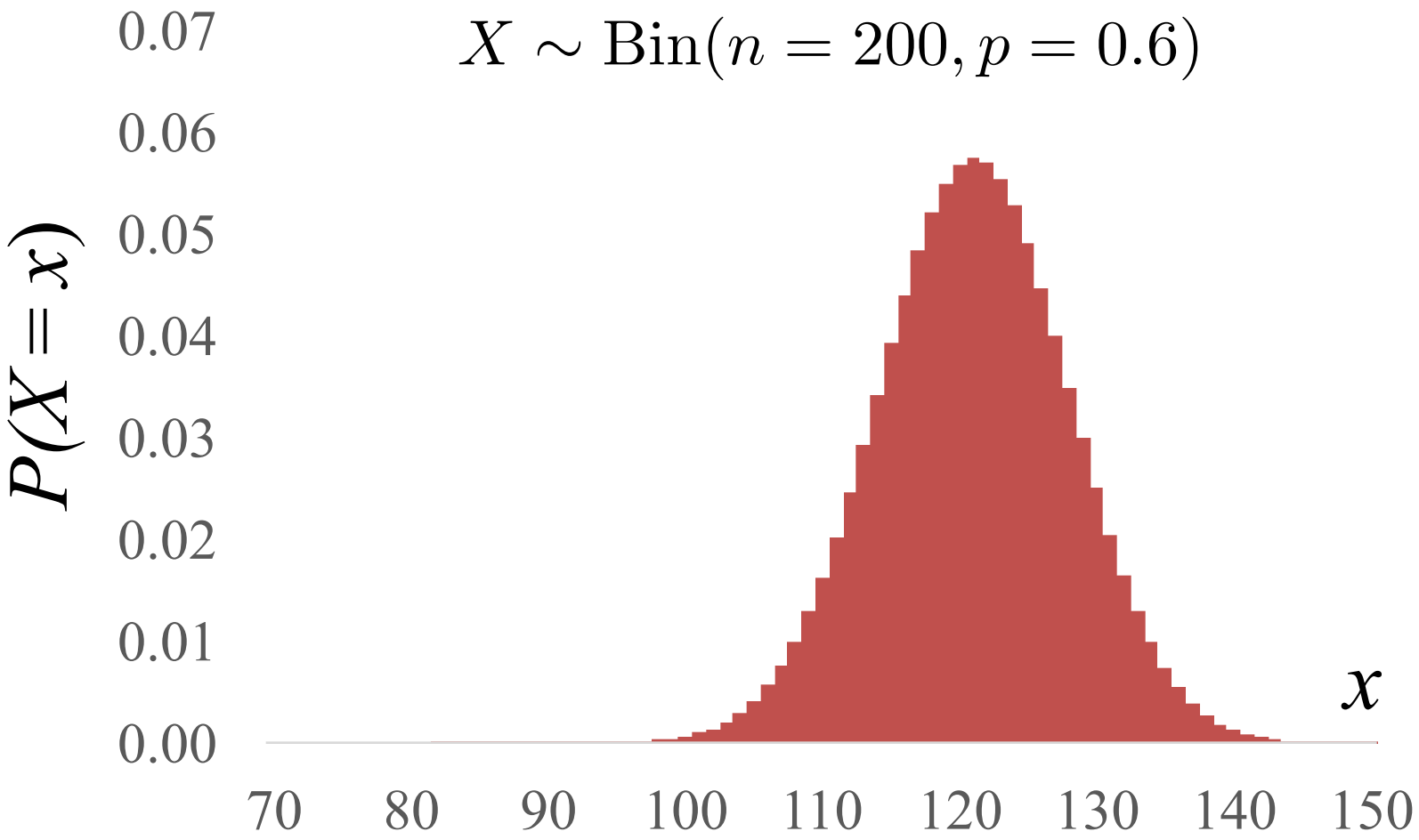






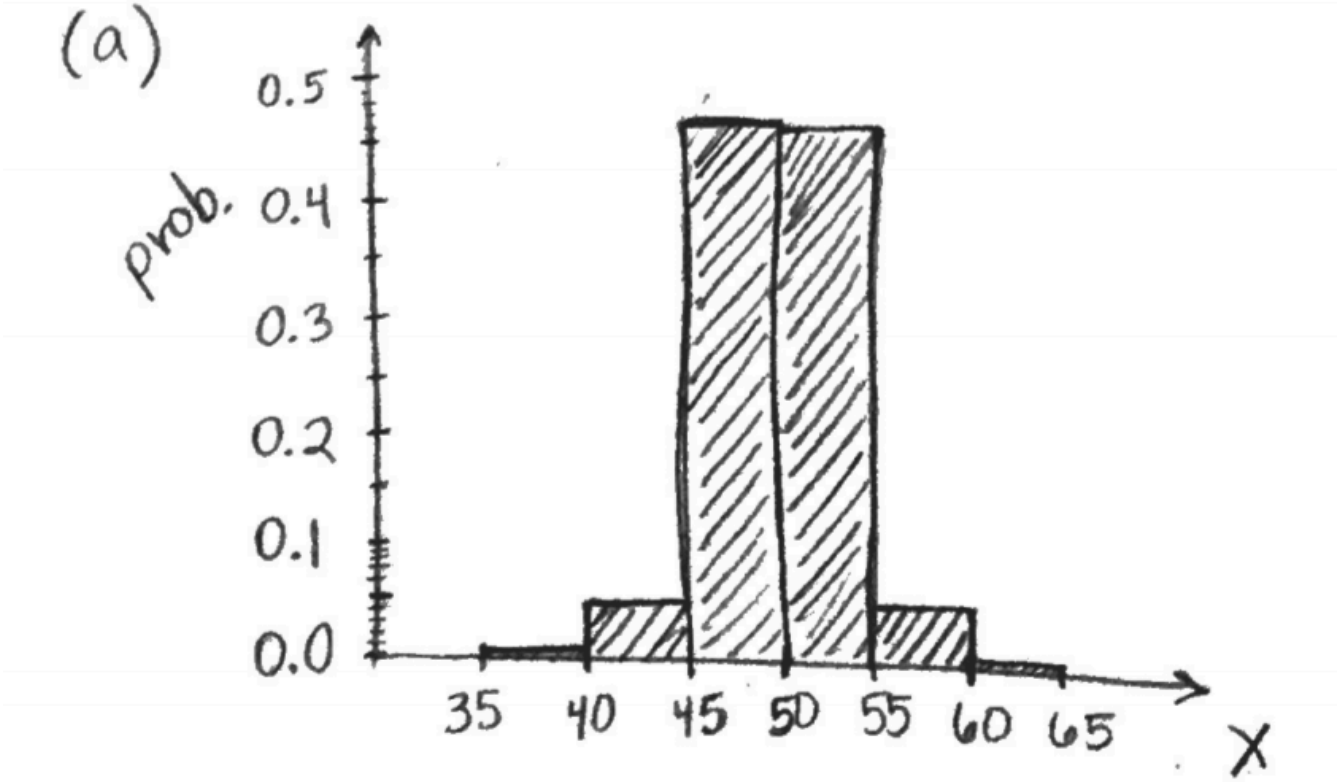


C.L.T. Explains This

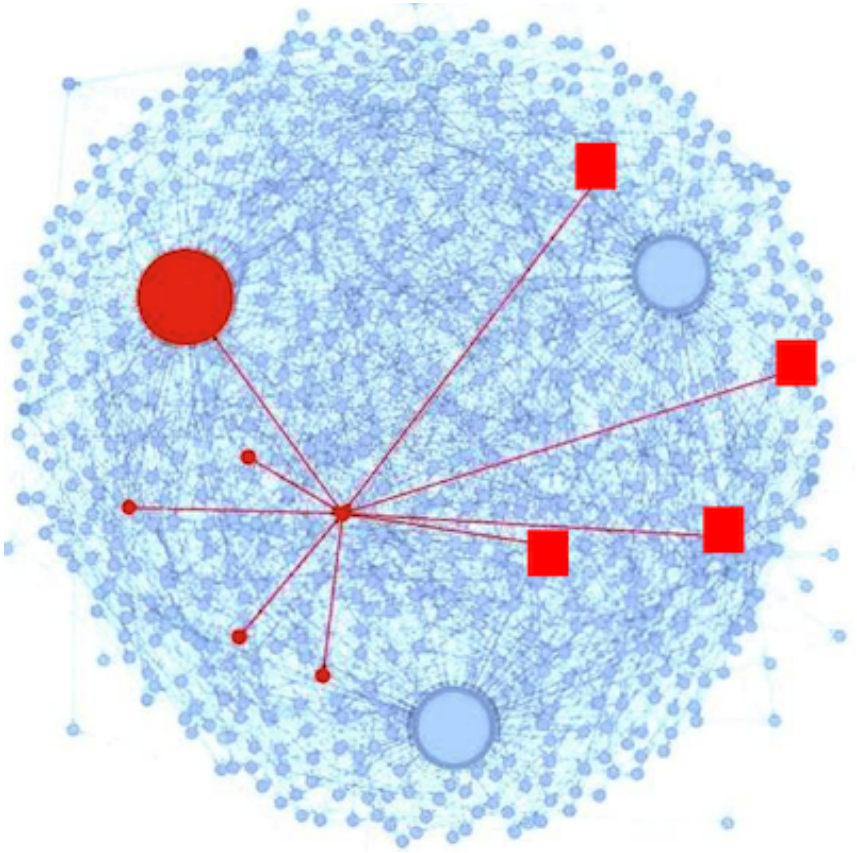


C.L.T. Explains This

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



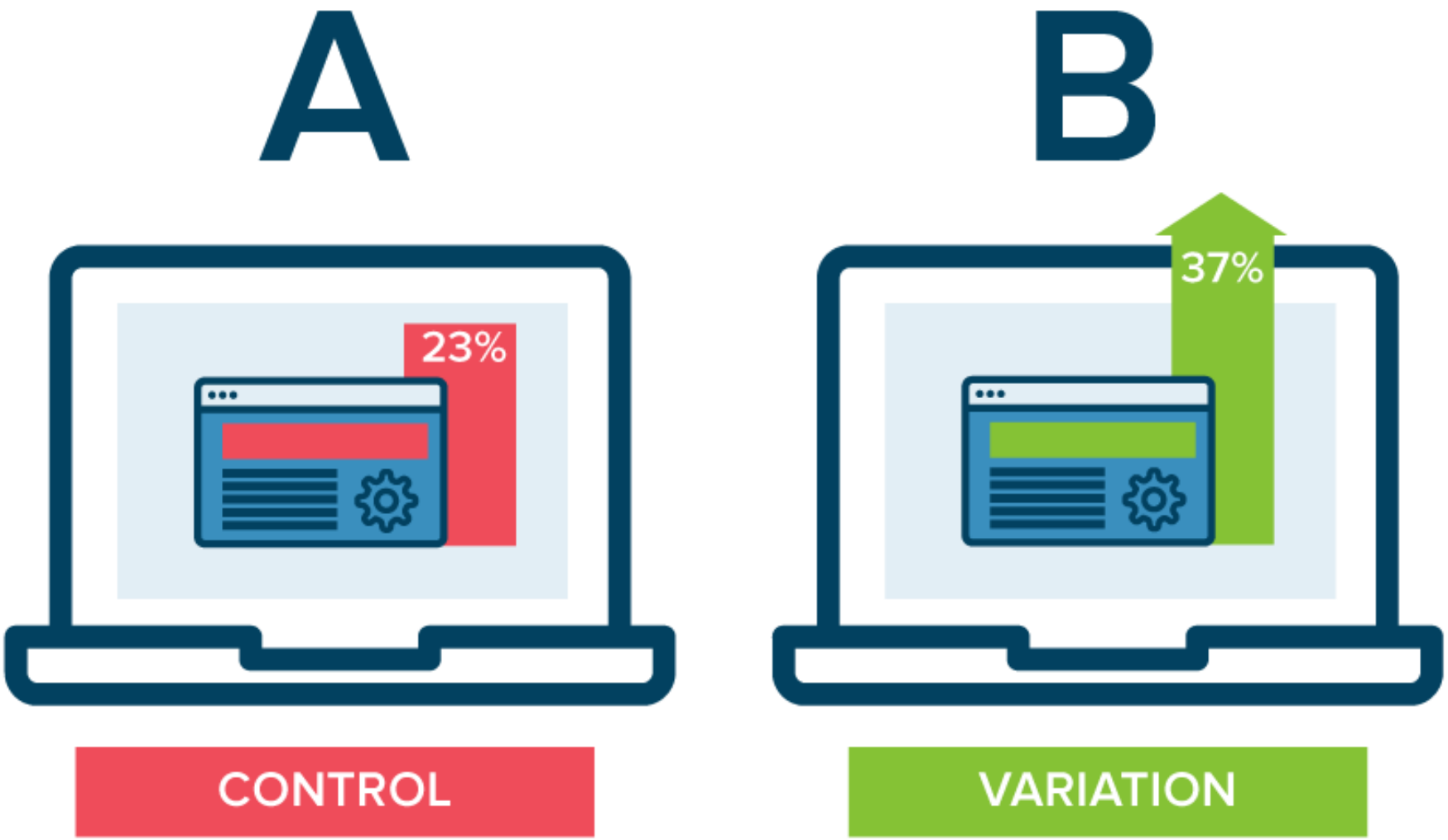
Peer Grading



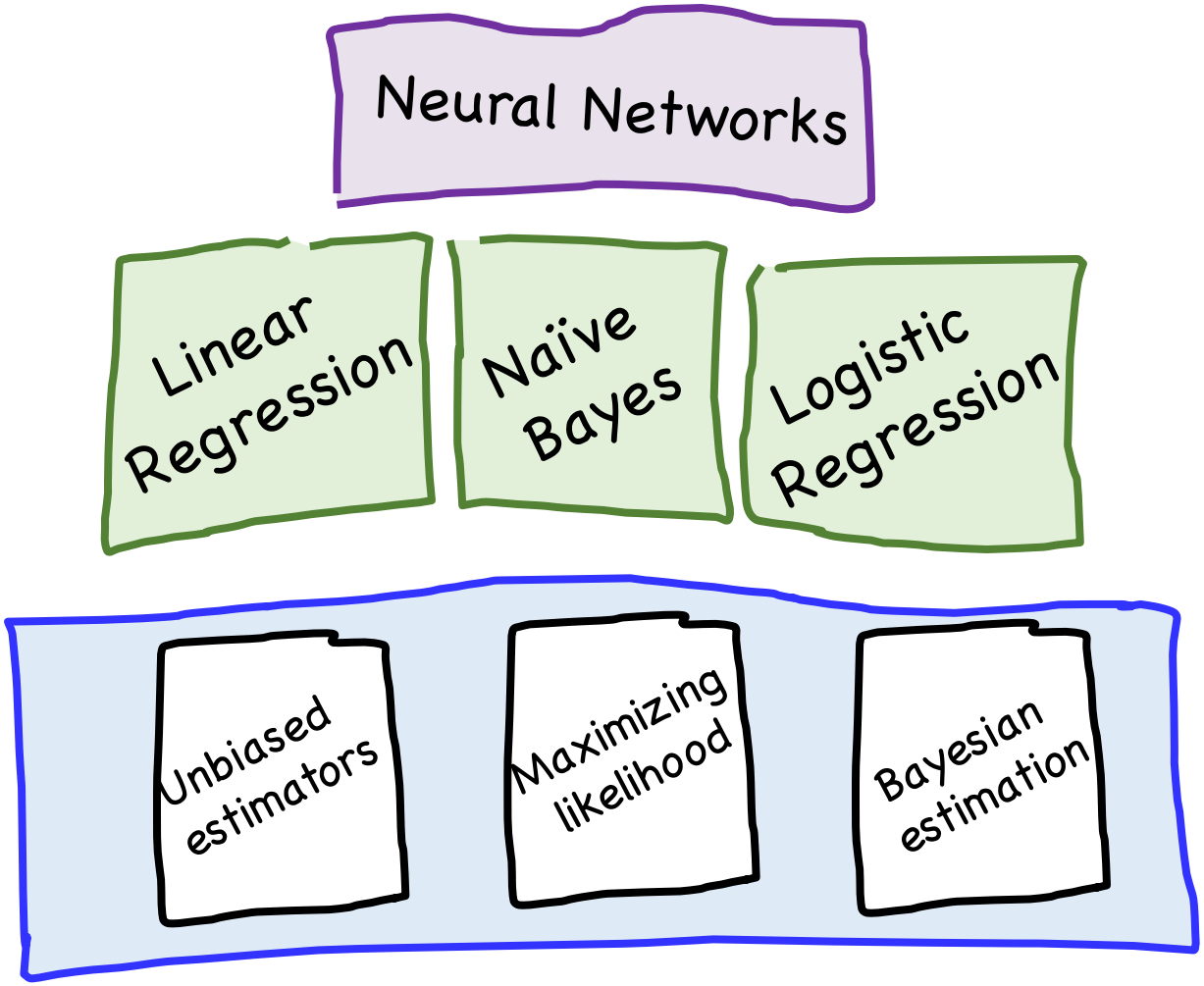
Peer Grading on Coursera HCI.

31,067 peer grades for 3,607 students.

A/B Testing

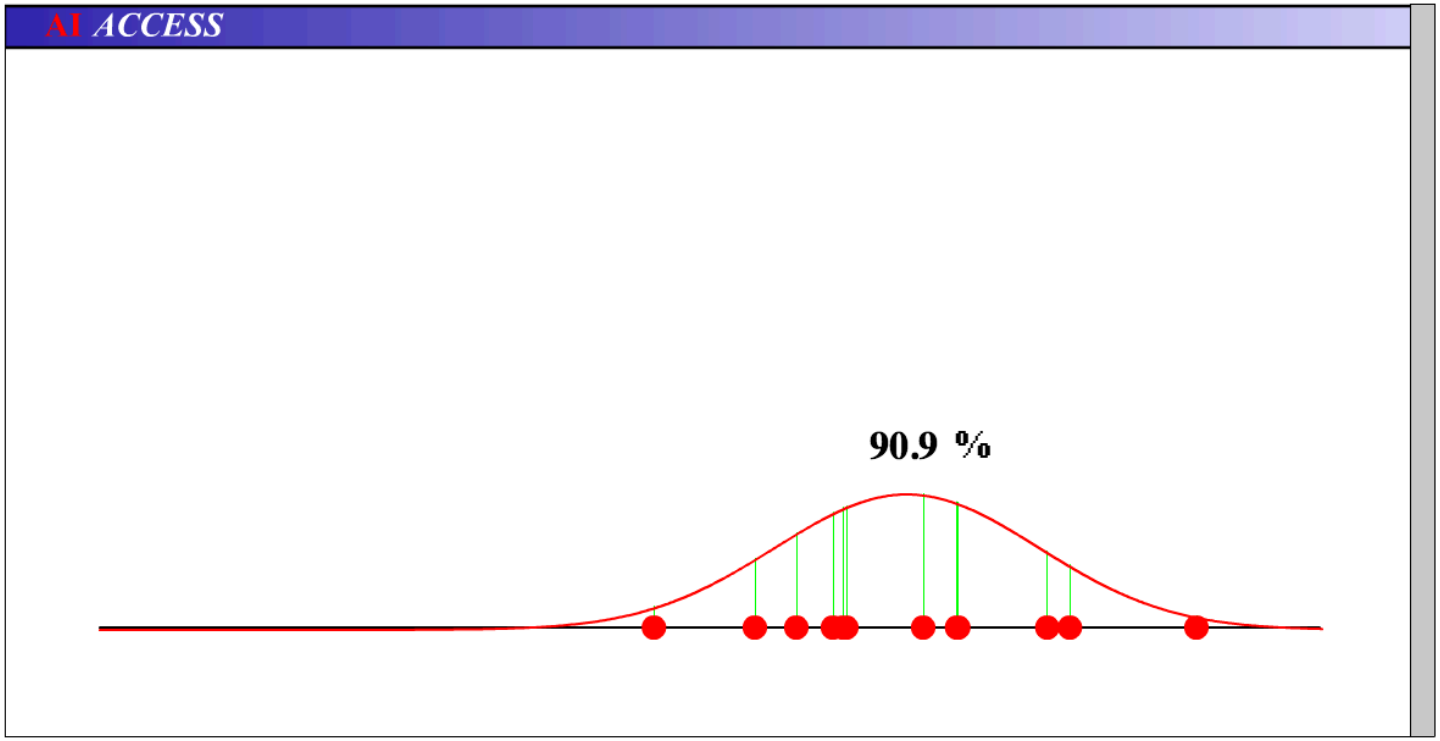


Towards Machine Learning



MLE: Likelihood of Data

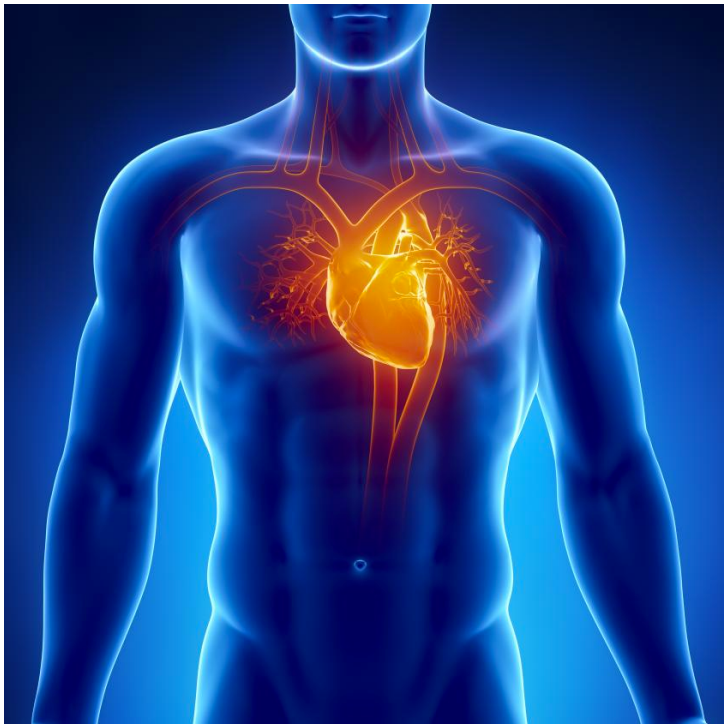
Likelihood of Data from a Normal



MAP: Most Probable Parameter

Machine Learning

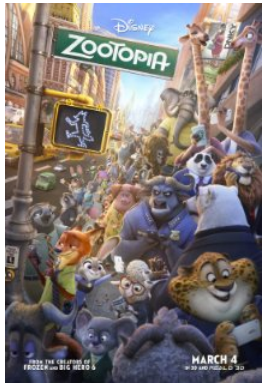
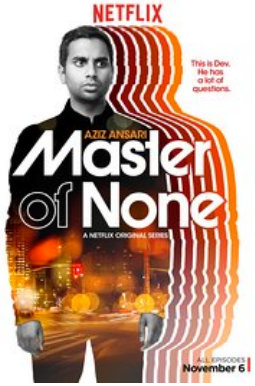
Heart



Ancestry

Netflix

Naïve Bayes

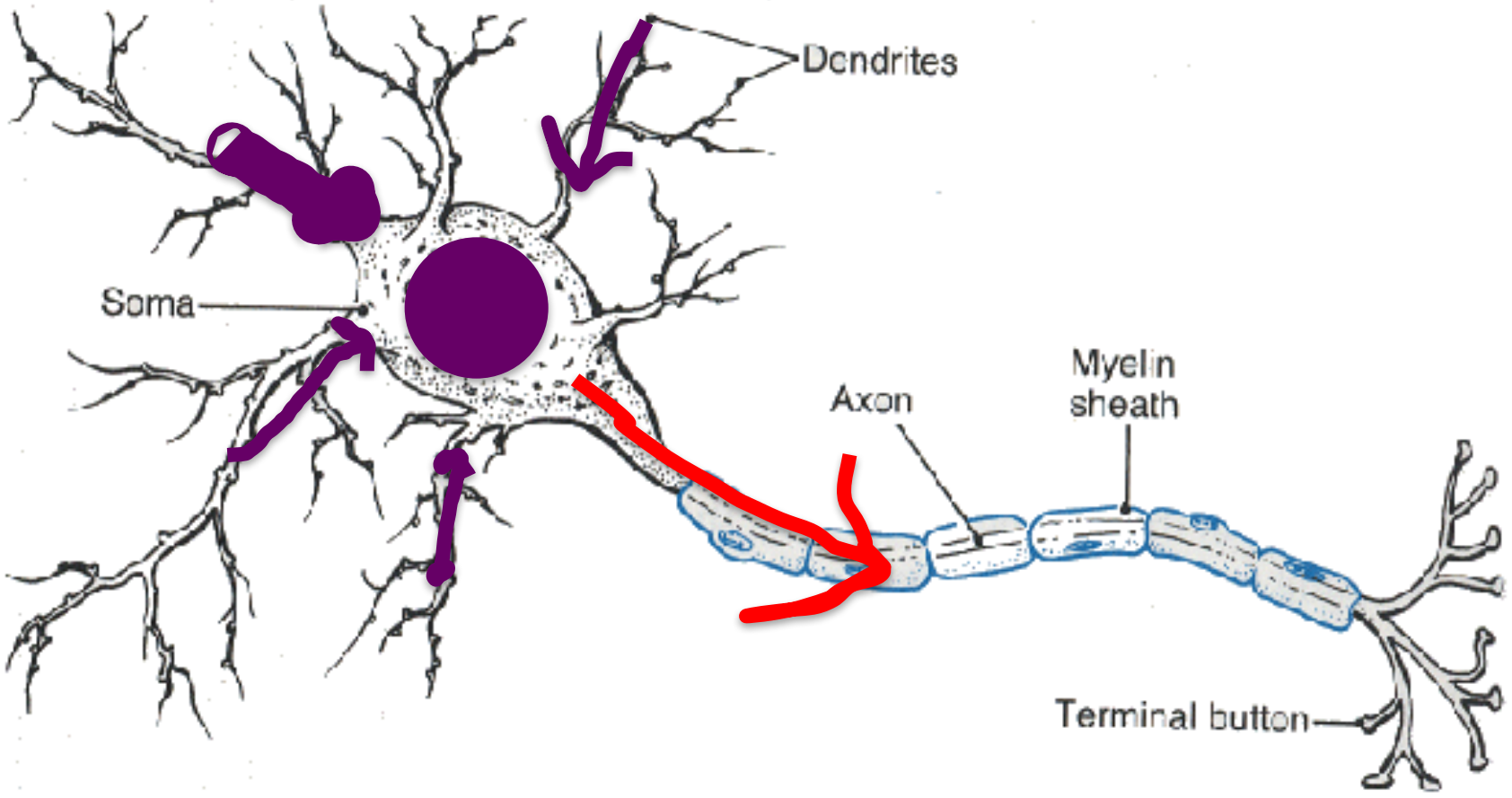
	x_1	x_2	x_3	y
				
User 1	1	0	1	1
User 2	1	0	1	0
				⋮
User n	0	1	1	1

Logistic Regression

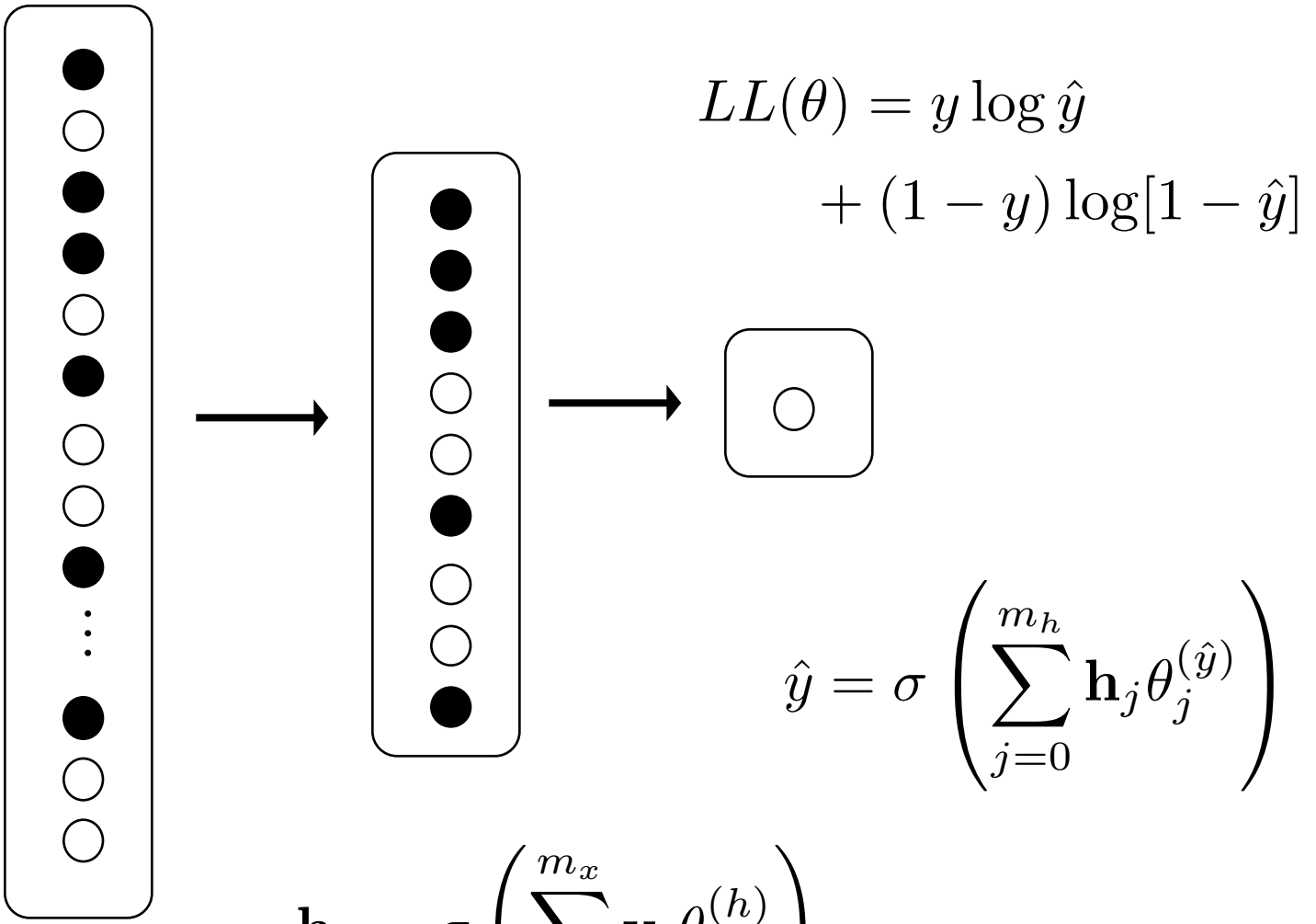
0.91

$$X$$

Logistic Regression



Deep Learning



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

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

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

By the numbers

500 to 1000 Bananas

~ 30 Major Keys

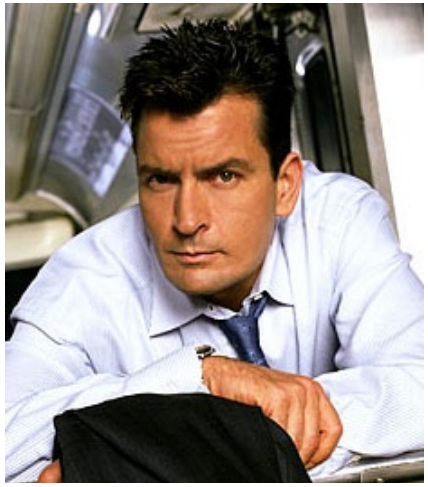
Naïve Bayes Assumption:

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

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

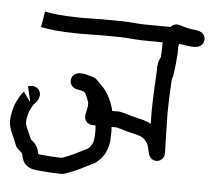
- Carl Friedrich Gauss (1777-1855) was a remarkably influential German mathematician

- 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

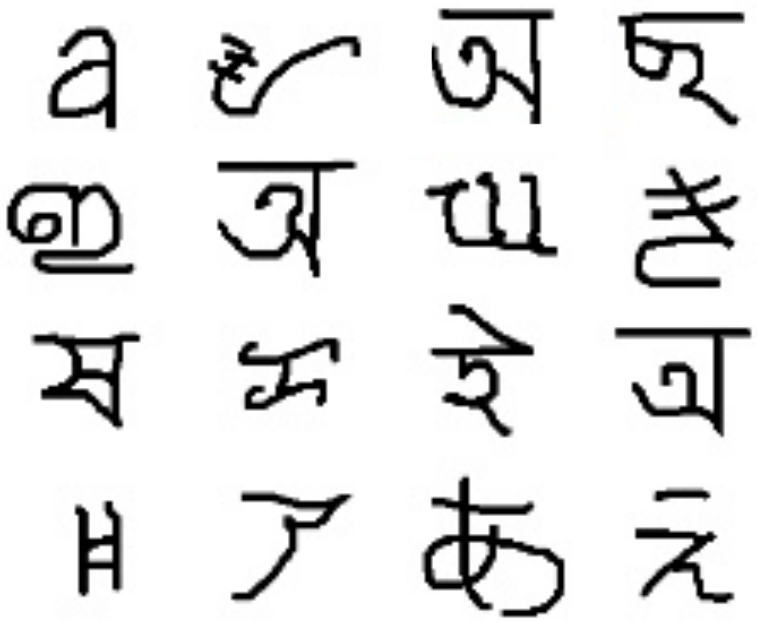
Open Problems

One Shot Learning

Single training example:



Test set:



Natural Language

Theoretical Deep Learning