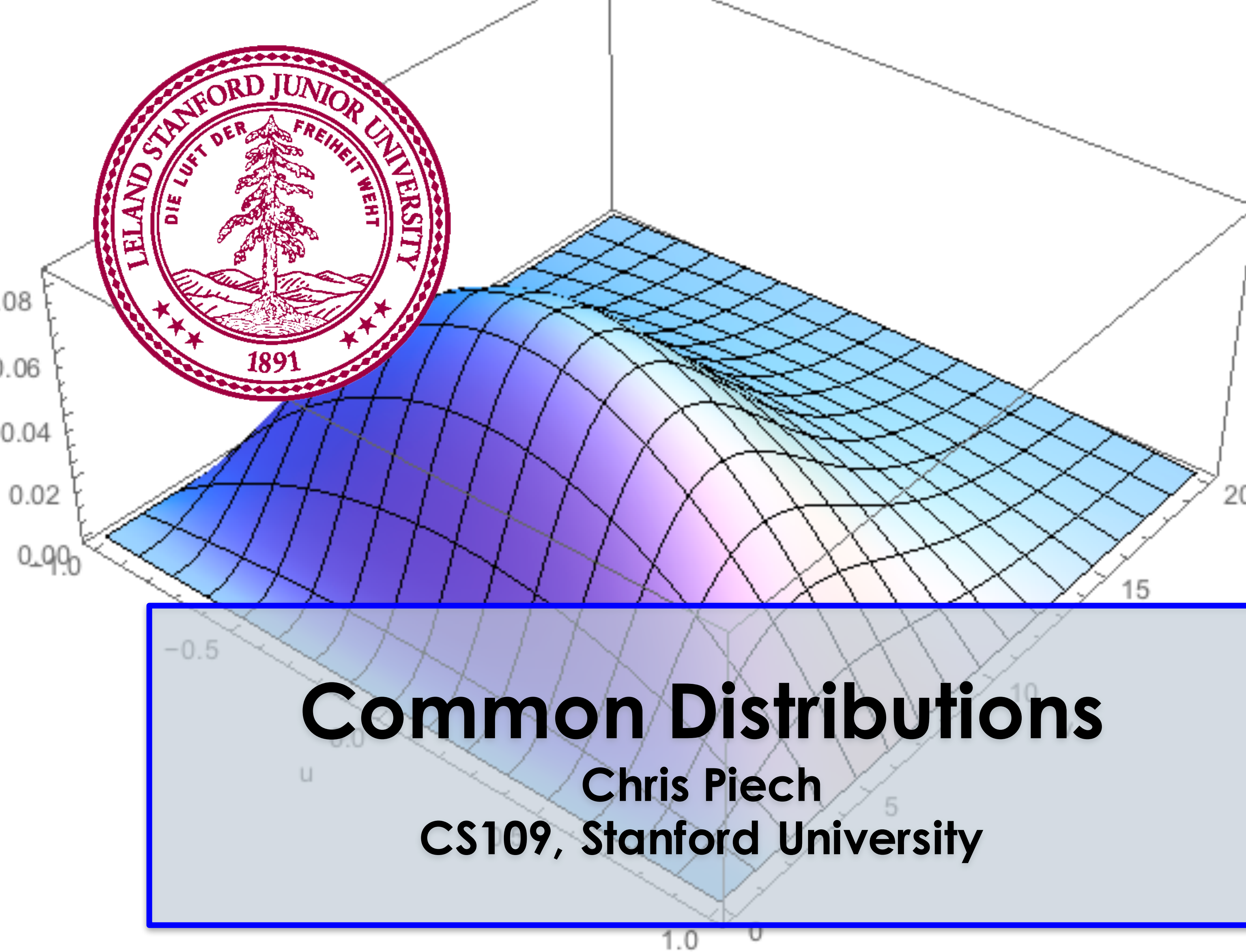


Common Distributions

Chris Piech
CS109, Stanford University

Binomial Random Variable

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

Binomial Random Variable

$$P(X = k) = \overbrace{\binom{n}{k} p^k (1-p)^{n-k}}^{\text{Probability Mass Function for a Binomial}}$$

Probability that our variable takes on the value k

Poisson Random Variable

- X is a **Poisson** Random Variable: the number of occurrences in a fixed interval of time.

$$X \sim \mathrm{Poi}(\lambda)$$

  - λ is the "rate"
  - X takes on values 0, 1, 2…
  - has distribution (PMF):

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

More?

Discrete Distributions

Don’t have to memorize all of the following distributions. We want you to get a sense of how random variables work.

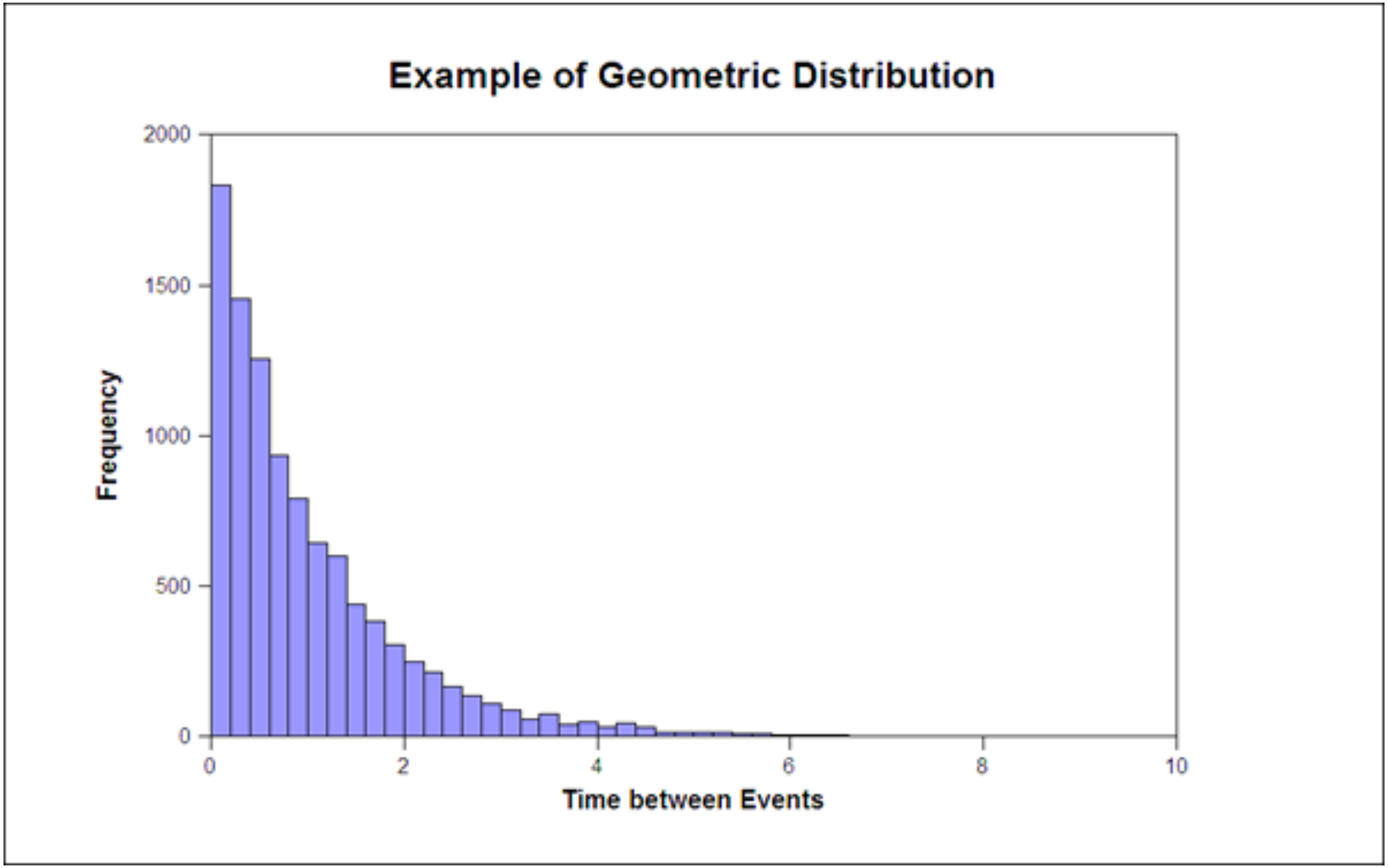
Geometric Random Variable

- X is **Geometric** Random Variable: $X \sim \text{Geo}(p)$
  - X is number of independent trials until first success
  - p is probability of success on each trial
  - X takes on values 1, 2, 3, …, with probability:

$$P(X = n) = (1 - p)^{n-1} p$$

  - $E[X] = 1/p$ $\quad\quad$ $\text{Var}(X) = (1 - p)/p^2$

- Examples:
  - Flipping a coin (P(heads) = p) until first heads appears
  - Urn with N black and M white balls. Draw balls (with replacement, $p = N/(N + M)$) until draw first black ball
  - Generate bits with P(bit = 1) = p until first 1 generated


Example of Geometric Distribution
2000
1500
1000
500
0
Frequency
0
2
4
6
8
10
Time between Events

Negative Binomial Random Variable

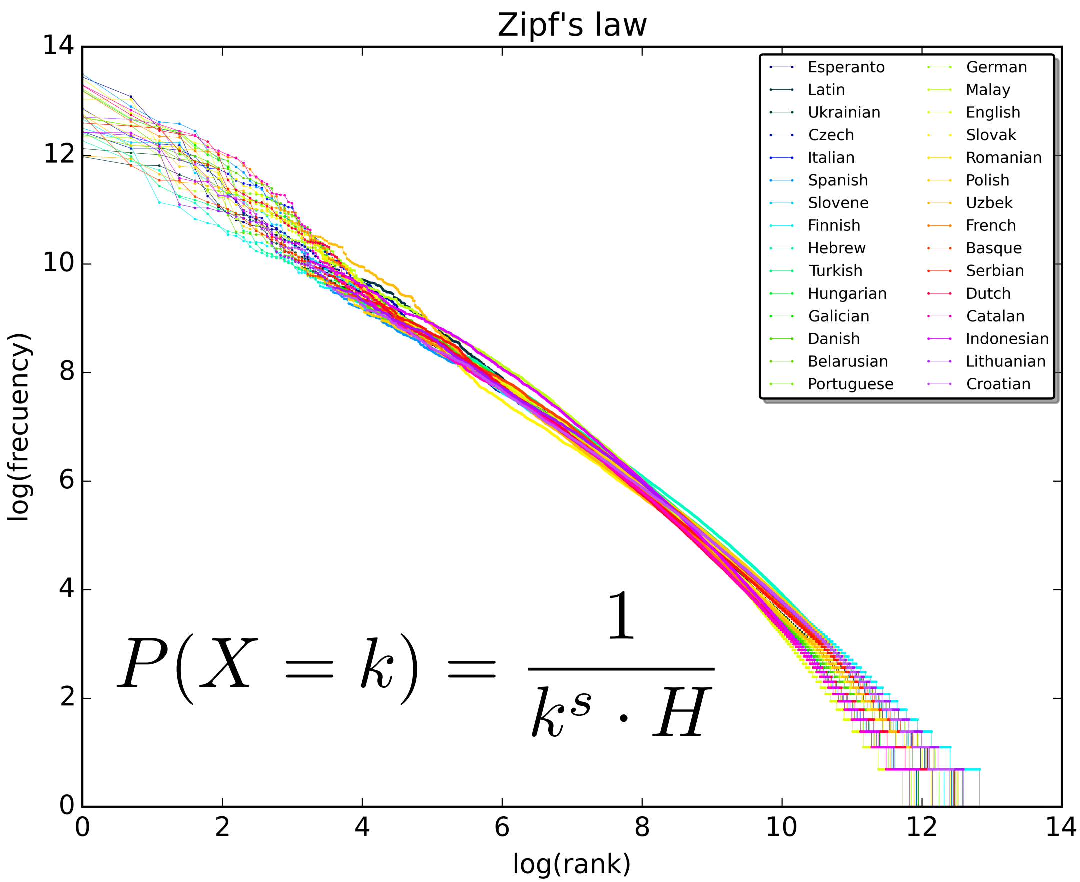
- X is **Negative Binomial** RV: X ~ NegBin(r, p)
  - X is number of independent trials until r successes
  - p is probability of success on each trial
  - X takes on values r, r + 1, r + 2…, with probability:

$$P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \text{ where } n = r, r+1, \ldots$$

  - E[X] = r/p Var(X) = $r(1 – p)/p^2$
- Note: Geo(p) ~ NegBin(1, p)
- Examples:
  - # of coin flips until r-th "heads" appears
  - # of strings to hash into table until bucket 1 has r entries

Zipf Random Variable

- X is **<u>Zipf</u>** RV: X ~ Zipf(s,N)
  - X is the popularity-rank index of a chosen element
  - S and N are properties of the language



Discrete Distributions

Bernoulli:

- indicator of coin flip X ~ Ber(p)

Binomial:

- # successes in n coin flips X ~ Bin(n, p)

Poisson:

- # successes in n coin flips X ~ Poi(λ)

Geometric:

- # coin flips until success X ~ Geo(p)

Negative Binomial:

- # trials until r successes X ~ NegBin(r, p)

Zipf:

- The popularity rank of a random word, from a natural language
- X ~ Zipf(s)

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.

(a) What is the probability that the first number you try will produce a bit string which starts with g zeroes (in other words you mine a bitcoin)?

(b) How many different numbers do you expect to have to try before you mine a bitcoin?

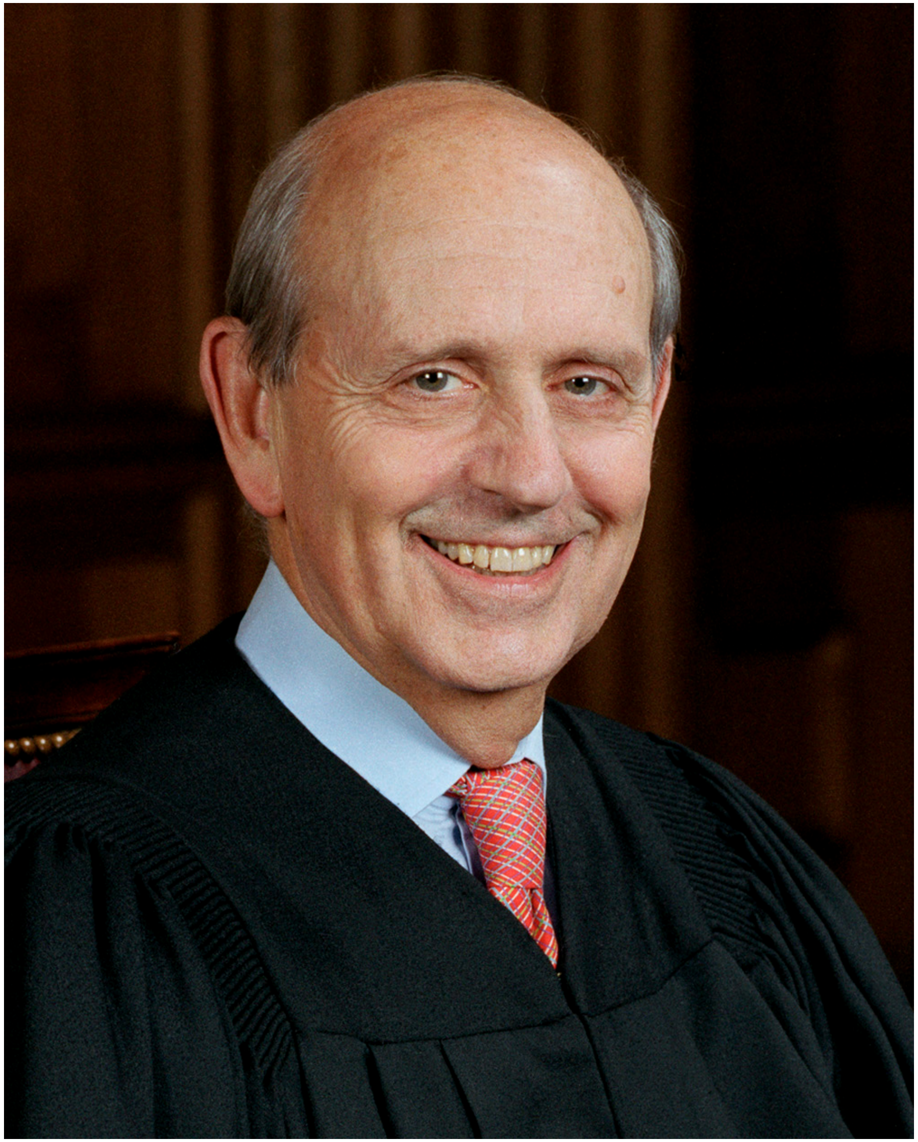
(c) Probability that it will take less than 10^3 tries to mine 5 bitcoins?

Dating

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 before finding a life mate? What is the standard deviation?

Equity in the Courts



Equity in the Courts

Supreme Court case: Berghuis v. Smith

If a group is underrepresented in a jury pool, how do you tell?

- Article by Erin Miller –January 22, 2010
- Thanks to (former CS109er) Josh Falk for this article

Justice Breyer [Stanford Alum] opened the questioning by invoking the binomial theorem. He hypothesized a scenario involving **“an urn with a thousand balls, and sixty are blue, and nine hundred forty are purple, and then you select them at random… twelve at a time.”** According to Justice Breyer and the binomial theorem, if the purple balls were under represented jurors then **“you would expect… something like <u>a third to a half</u> of juries would have at least one minority person”** on them.

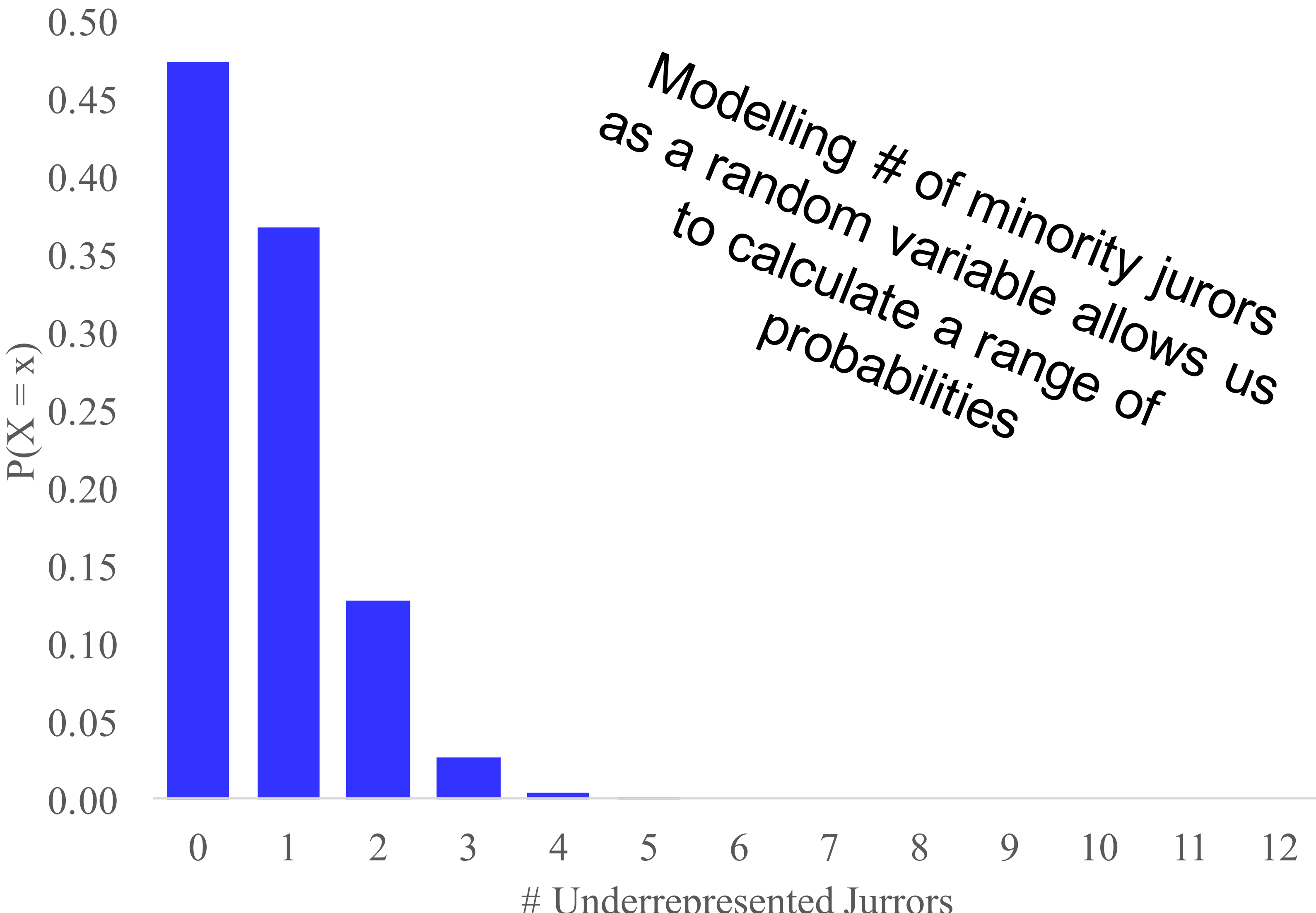
Justin Breyer Meets CS109

- Approximation using Binomial distribution
  - Assume P(blue ball) constant for every draw = 60/1000
  - X = # blue balls drawn. $X \sim \text{Bin}(12, 60/1000 = 0.06)$
  - $P(X \geq 1) = 1 - P(X = 0) \approx 1 - 0.4759 = 0.5240$

In Breyer's description, should actually expect just ***over half*** *of juries to have at least one black person on them*

Demo

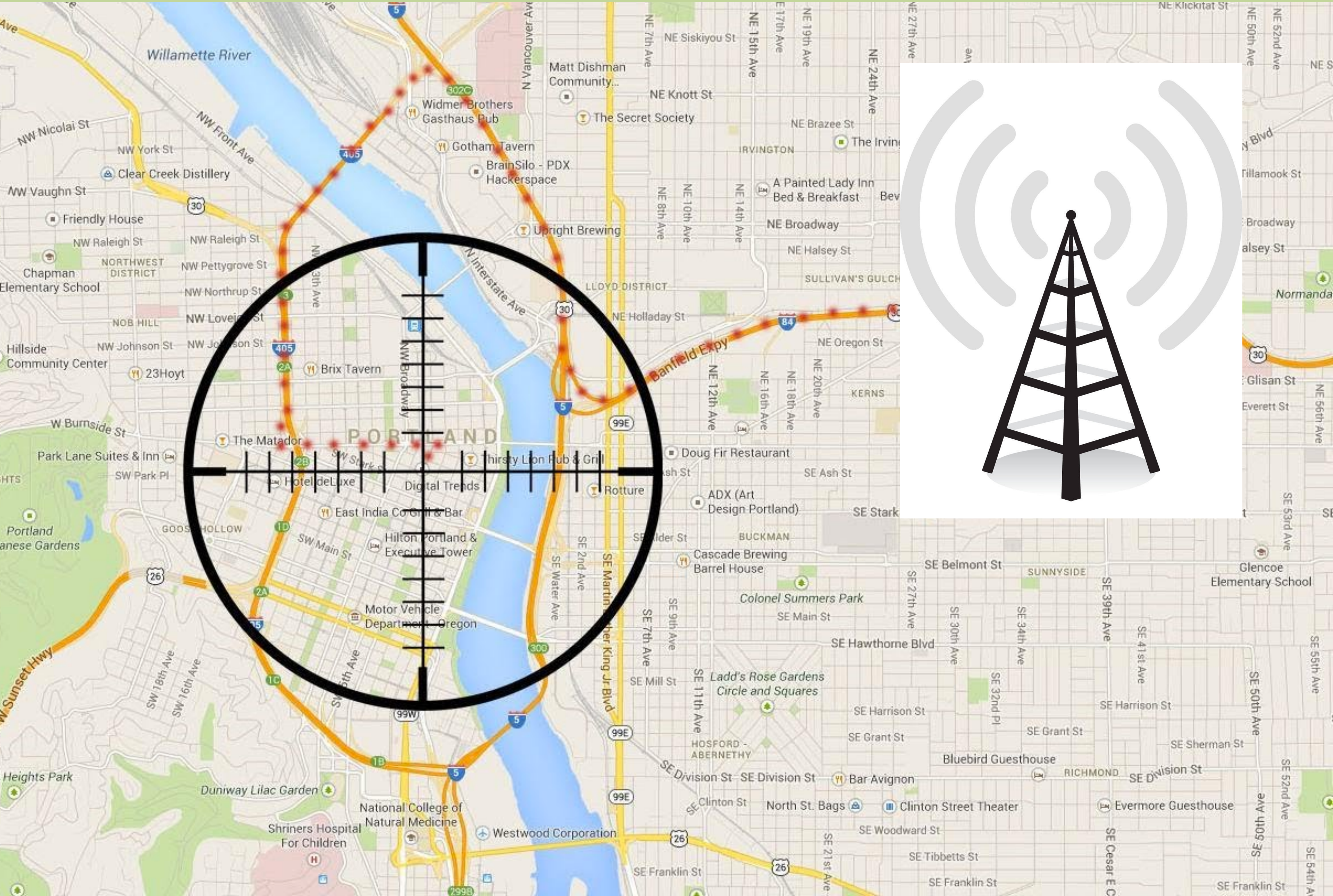
Underrepresented Juror PMF



Modelling # of minority jurors as a random variable allows us to calculate a range of probabilities

Big hole in our knowledge

Not all values are discrete



random()?

Riding the Marguerite

Riding the Marguerite

- Say the Marguerite bus stops at the Gates bldg. at 20 minute intervals (2:00, 2:20, etc.)
  - Passenger arrives at stop between 2-2:30pm
- P(Passenger waits < 5 minutes for bus)?

From Discrete to Continuous

- So far, all random variables we saw were *discrete*
  - Have finite or countably infinite values (e.g., integers)
  - Usually, values are binary or represent a count
- Now it's time for *continuous* random variables
  - Have (uncountably) infinite values (e.g., real numbers)
  - Usually represent measurements (arbitrary precision)
    - Height (centimeters), Weight (lbs.), Time (seconds), etc.
- Difference between how *many* and how *much*
- Generally, it means replace $\sum_{x=a}^{b} f(x)$ with $\int_{a}^{b} f(x)dx$

Integrals

*loving, not scary

Continuous Random Variables

- X is a **Continuous Random Variable** if there is function $f(x) \geq 0$ for $-\infty \leq x \leq \infty$, such that:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

- f is a Probability Density Function (PDF) if:

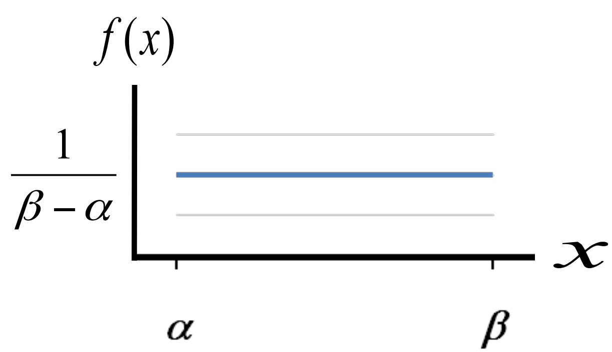
$$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x)dx = 1$$

Uniform Random Variable

- X is a **Uniform Random Variable**: $X \sim \text{Uni}(\alpha, \beta)$
  - Probability Density Function (PDF):

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \alpha \le x \le \beta \\ 0 & \text{otherwise} \end{cases}$$

    - Sometimes defined over range $\alpha < x < \beta$



  - $$P(a \le x \le b) = \int_a^b f(x)dx = \frac{b-a}{\beta - \alpha}$$ (for $\alpha \le a \le b \le \beta$)

Fun with the Uniform Distribution

- X ~ Uni(0, 20)

$$f(x) = \begin{cases} \frac{1}{20} & 0 \le x \le 20 \\ 0 & \text{otherwise} \end{cases}$$

  - P(X < 6)?

$$P(x < 6) = \int_0^6 \frac{1}{20}\, dx = \frac{6}{20}$$

  - P(4 < X < 17)?

$$P(4 < x < 17) = \int_4^{17} \frac{1}{20}\, dx = \frac{17}{20} - \frac{4}{20} = \frac{13}{20}$$

Riding the Marguerite

- Say the Marguerite bus stops at the Gates bldg. at 15 minute intervals (2:00, 2:15, 2:30, etc.)
  - Passenger arrives at stop uniformly between 2-2:30pm
  - X ~ Uni(0, 30)
- P(Passenger waits < 5 minutes for bus)?
  - Must arrive between 2:10-2:15pm or 2:25-2:30pm

$$P(10 < X < 15) + P(25 < x < 30) = \int_{10}^{15} \frac{1}{30}\, dx + \int_{25}^{30} \frac{1}{30}\, dx = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}$$

- P(Passenger waits > 14 minutes for bus)?
  - Must arrive between 2:00-2:01pm or 2:15-2:16pm

$$P(0 < X < 1) + P(15 < x < 16) = \int_{0}^{1} \frac{1}{30}\, dx + \int_{15}^{16} \frac{1}{30}\, dx = \frac{1}{30} + \frac{1}{30} = \frac{1}{15}$$