Algorithmic Analysis

CS109, Stanford University

Uncertainty Theory

- Beta Distributions
- Thompson Sampling
- Adding Random Vars
- Central Limit Theorem
- Sampling
- Bootstrapping
- Great Expectations

<review>

Simulations and Indicator Random Variables

CLT: $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Consider an event E. An Indicator Random Variable for E is defined as:

$$I_E = \begin{cases} 1 & \text{if } E \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

Fact 1: $\mathbb{E}[I_E] = P(E)$

$$\mathbb{E}[I_E] = 1 \cdot P(I_E = 1) + 0 \cdot P(I_E = 0)$$
$$= 1 \cdot P(E) + 0 \cdot P(E^C) = P(E)$$

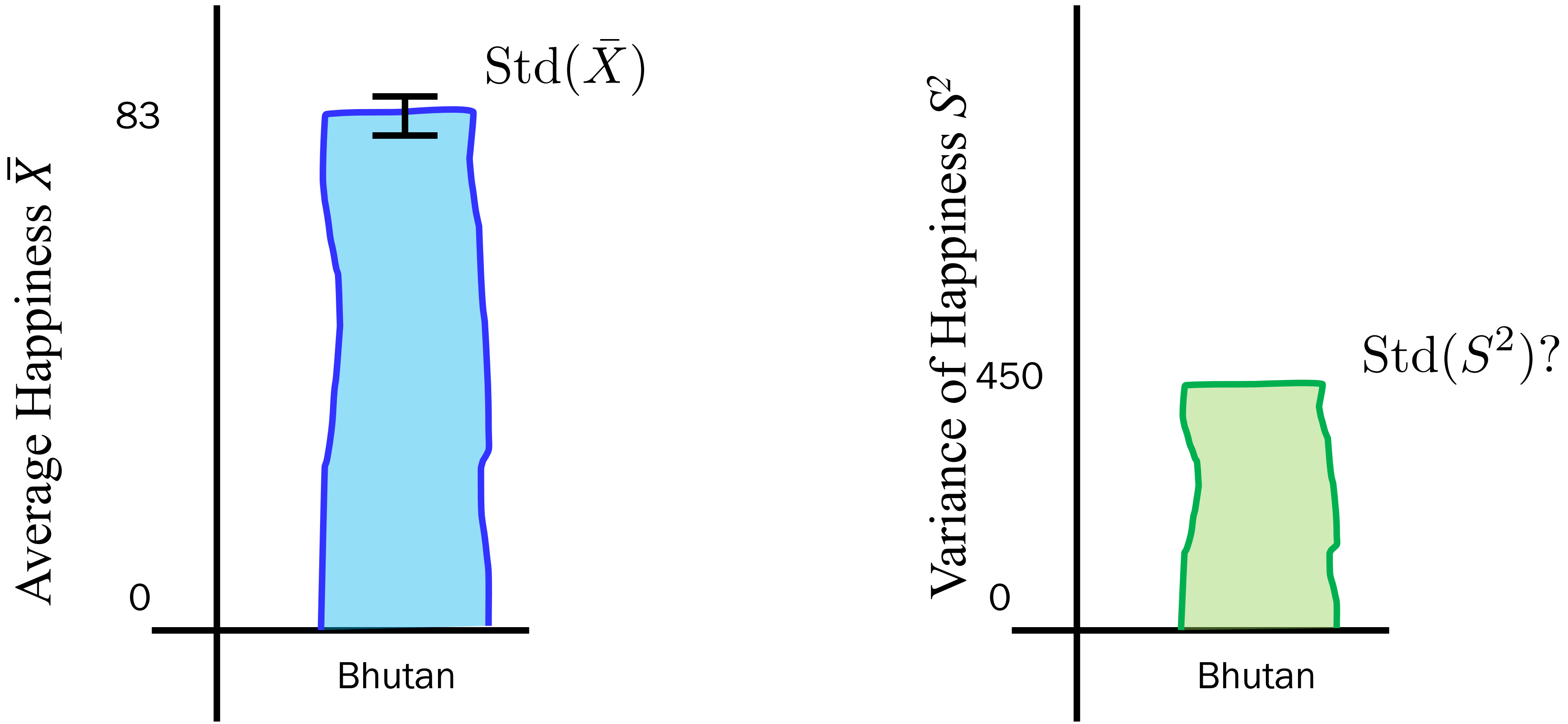
Fact 2: $Var(I_E) = P(E)\big(1 - P(E)\big) < 1$

Why Simulation Works

$$\frac{\#\text{ of times } E \text{ occurred}}{\#\text{ of Trials}} \sim \mathcal{N}\left(P(E), \frac{Var(I_E)}{n}\right)$$

→ Very strongly concentrated around P(E) as $n \to \infty$

Our Report to Bhutan Government

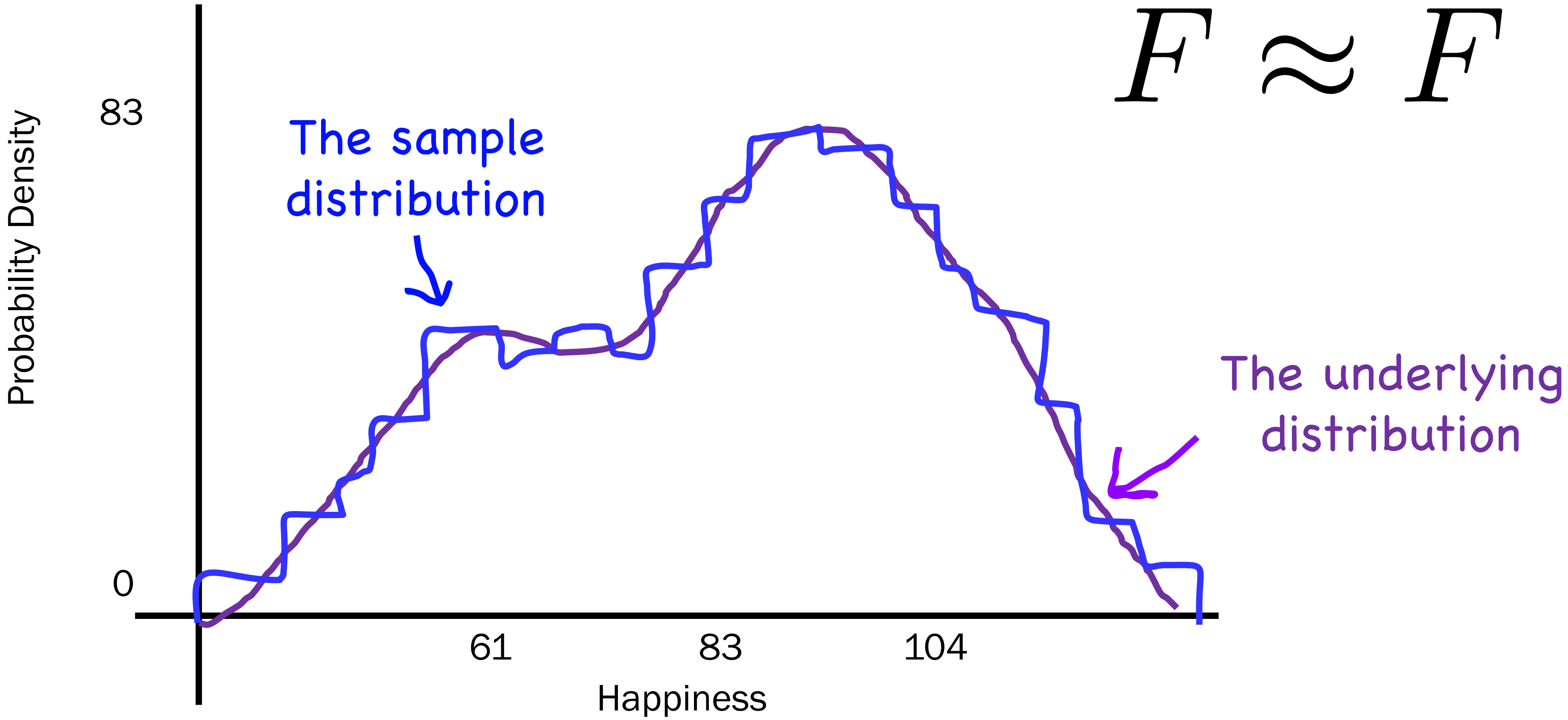


Claim: The average happiness of Bhutan is 83 ± 2

But Wait – What If You Actually Have a Good Estimate?

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

$$F \approx \hat{F}$$



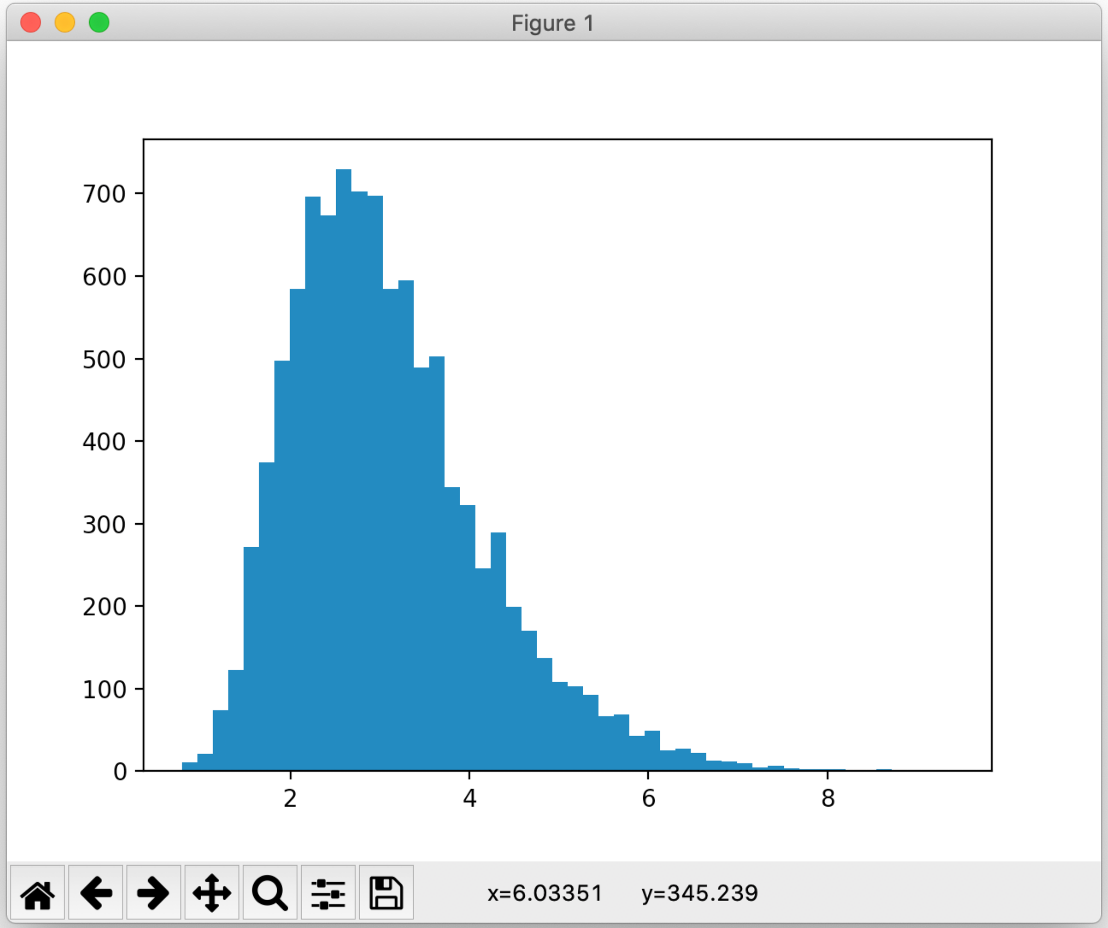
* This is just a histogram of your data!!

Bootstrapping in Practice

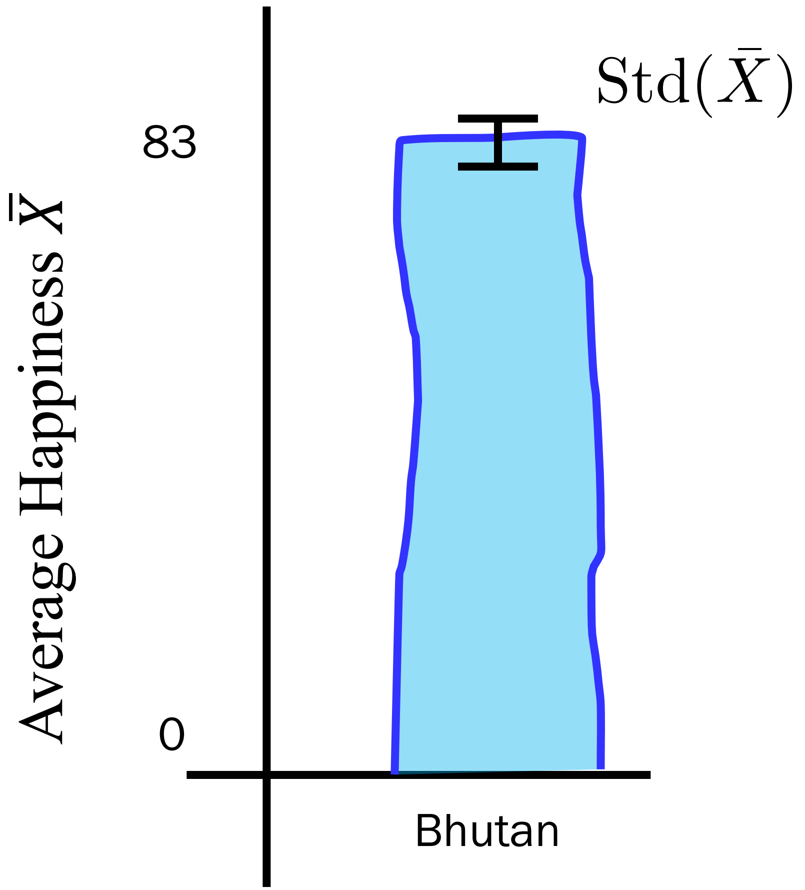
Bootstrap Algorithm (sample):

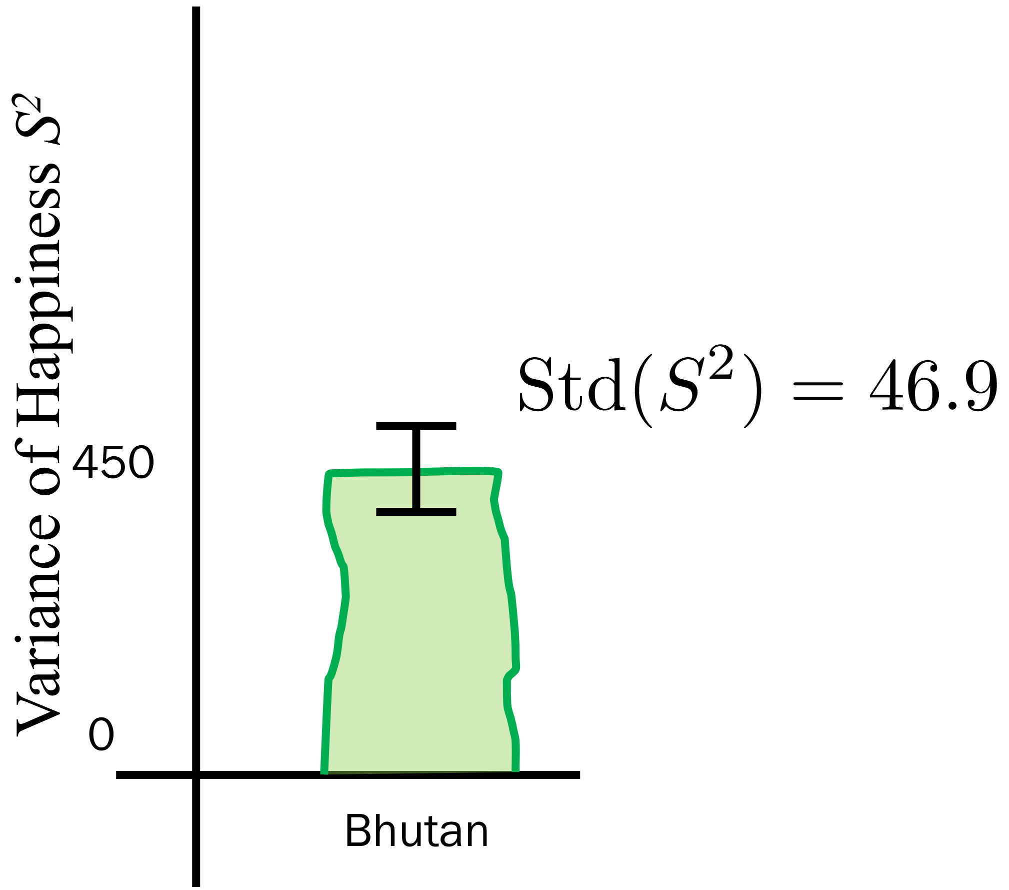
1. Repeat **10,000** times:
   a. **Choose len(sample) elems from sample, with replacement**
   b. Recalculate the stat on the resample
2. You now have a **distribution of your stat**

The Distribution of the Sampling Variance



Our Report to Bhutan Government





Claim: The average happiness of Bhutan is 83 ± 2

Bootstraping allows you to:

- Measure “accuracy” of **any statistic**
- Calculate **p values**
- By approximating the **distribution of statistic using computers**

A real difference?

18 students	Learning in Context A	Learning in Context B	23 students
	4.44	2.15	
	3.36	3.01	
	5.87	2.02	
	2.31	1.43	
	…	…	
	3.70	1.83	

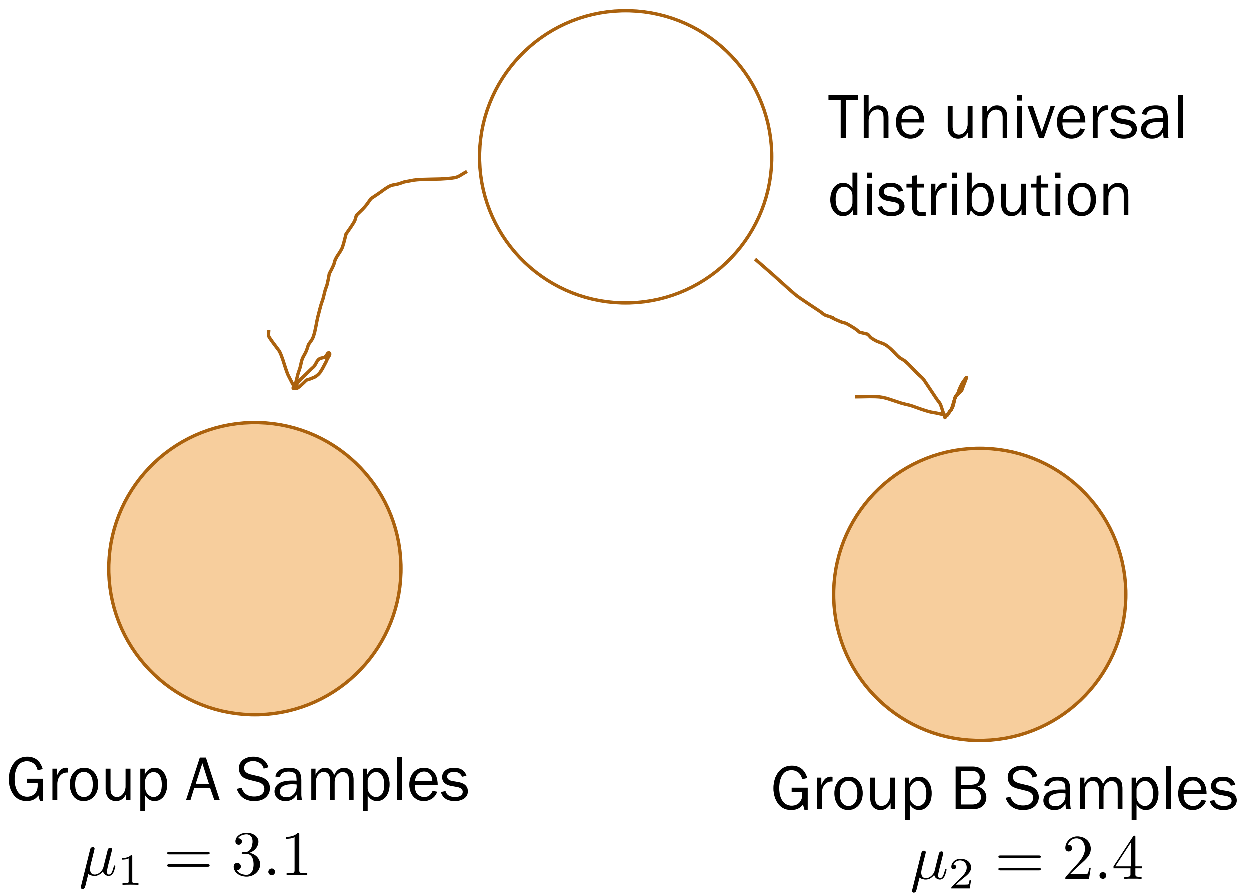
$\mu_1 = 3.1$ $\mu_2 = 2.4$

Claim: Group 1 and Group 2 are samples from different distributions with a 0.7 difference of means.

How confident are you in this claim?

The Null Hypothesis. How can we use bootstrapping here?

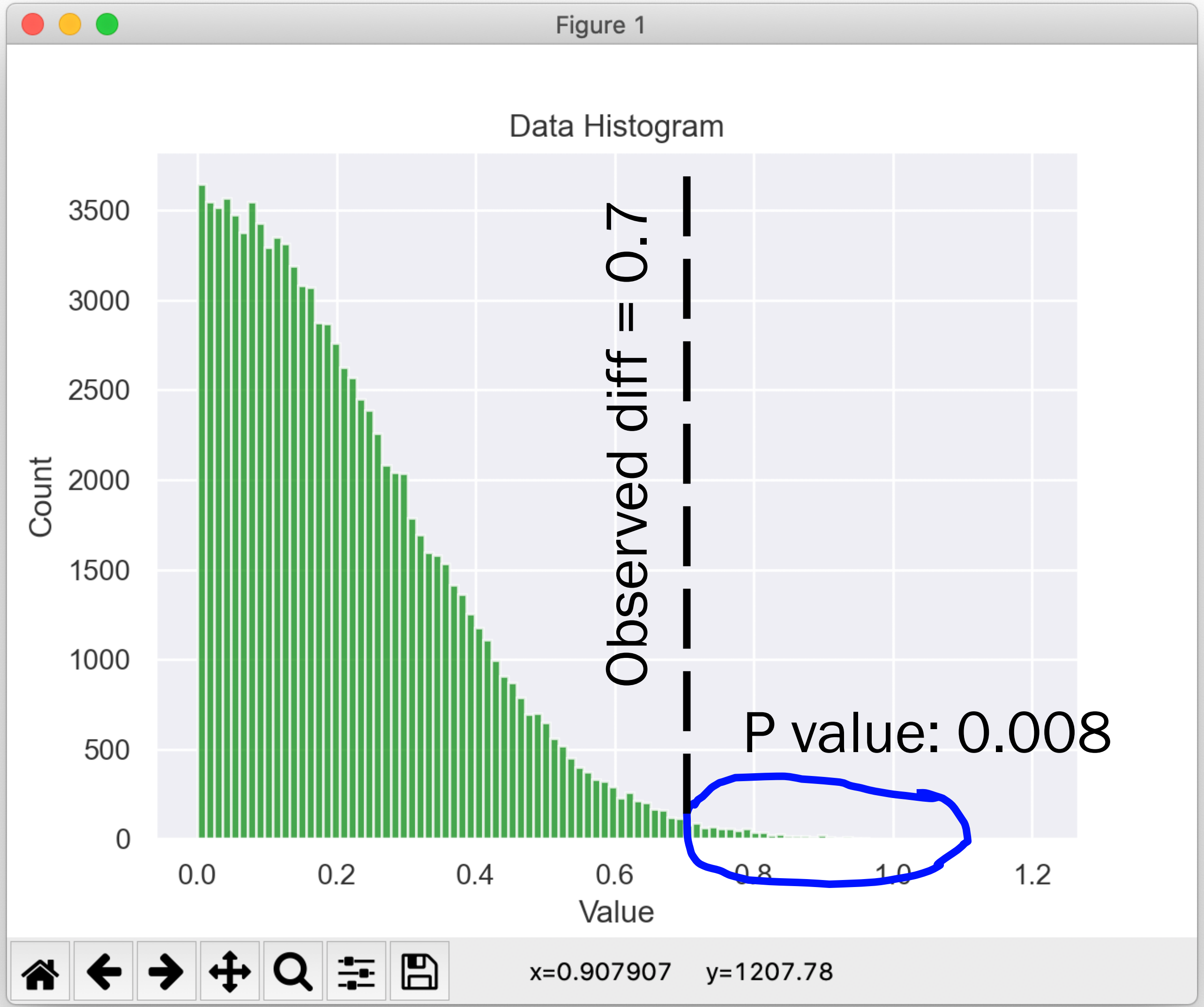
There is no difference between the two groups, so everyone is drawn from the same distribution. Any difference you observe is due to sampling error.



Distribution of Mean Diffs under Null Hypothesis

Q: But what distribution is this??

A: Folded Normal. Abs of a CLT process



End Review

Algorithmic Analysis

Expectation

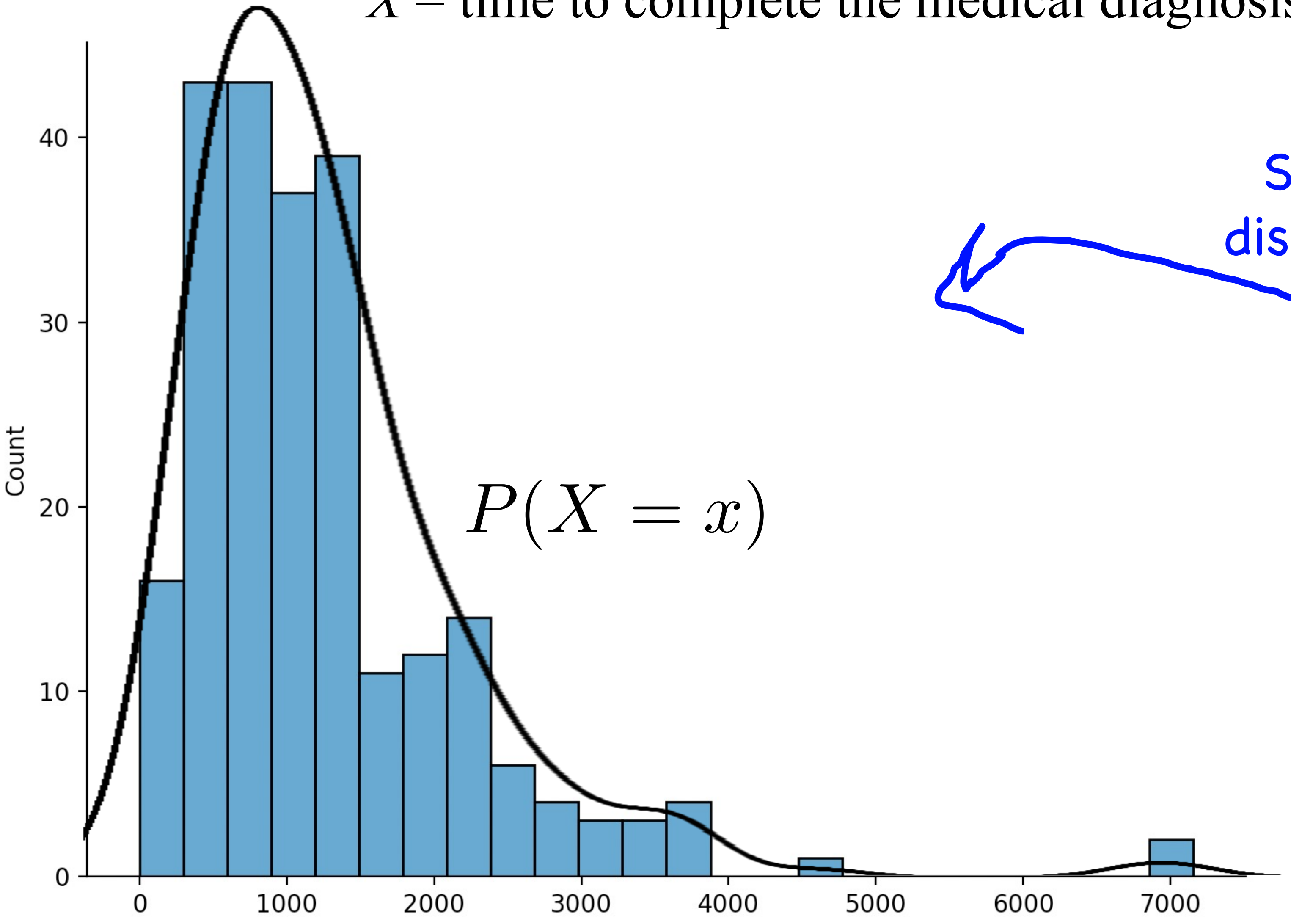
The probability
that X takes
on that value

$$E[X] = \sum_x x \cdot P(X = x)$$

All the values
that X can
take on

Limitation of Expectation

X = time to complete the medical diagnosis problem (in seconds)



$$P(X = x)$$

Summarize the distribution in one number

$$\text{E}[X] = 12600$$

Huge Advantage of Expectation: Concentration

Central Limit Theorem (Average)

Consider n independent and identically distributed (i.i.d) variables $X_1, X_2, \ldots, X_n$ with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$.

As $n \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

The **average** of the variables is normally distributed

Chris Piech, CS109 Stanford University 8

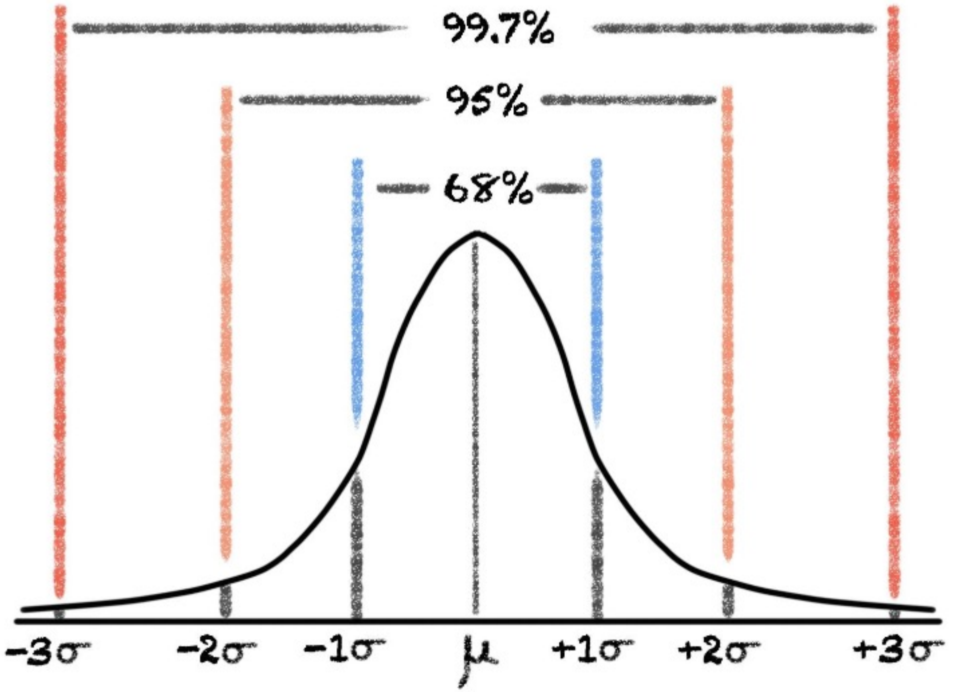
You can use the properties of Normal distribution to reason about sample means.

They center around the expectation!

"68-95-99.7 Rule" for Normal Distribution

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then:

$$P(\mu - 1\sigma \leq X \leq \mu + 1\sigma) \approx 68.27\%$$
$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 95.45\%$$
$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 99.73\%$$



Source: https://www.freecodecamp.org/news/normal-distribution-explained

(Weaker) Concentration for **any** distribution:

Chebyshev's Inequality: $P(|X - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2}$

Expectation of a Sum

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

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

Expectation of a Function

Law of unconscious statistician

$$E[g(X)] = \sum_x g(x) \cdot P(X = x)$$

So for example…

$$E[X^2] = \sum_x x^2 \cdot P(X = x)$$

Bool was Cool

Let E_1, E_2, ... E_n be events with indicator RVs X_i

- If event E_i occurs, then $X_i = 1$, else $X_i = 0$
- Recall $E[X_i] = P(E_i)$

- Why?

$$E[X_i] = 0 \cdot (1 - P(E_i)) + 1 \cdot P(E_i)$$

Bernoulli aka Indicator Random Variables were studied extensively by George Boole

Expectation of the Binomial

Let $Y \sim \text{Bin}(n, p)$

- n independent trials
- Let $X_i = 1$ if i-th trial is "success", 0 otherwise
- $X_i \sim \text{Ber}(p)$ $\quad E[X_i] = p$

$$Y = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

$$E[Y] = E[\sum_{i=1}^{n} X_i]$$

$$= \sum_{i=1}^{n} E[X_i]$$

$$= E[X_1] + E[X_2] + \ldots E[X_n]$$

$$= np$$

Expectation of the Negative Binomial

Let Y ~ NegBin(r, p)

- Recall Y is number of trials until r "successes"
- Let X_i = # of trials to get success after (*i* – 1)st success
- X_i ~ Geo(p) (i.e., Geometric RV)

$$Y = X_1 + X_2 + \cdots + X_r = \sum_{i=1}^{r} X_i \qquad E[X_i] = \frac{1}{p}$$

$$\begin{aligned} E[Y] &= E[\sum_{i=1}^{r} X_i] \\ &= \sum_{i=1}^{r} E[X_i] \\ &= E[X_1] + E[X_2] + \ldots E[X_r] \\ &= \frac{r}{p} \end{aligned}$$

Differential Privacy

Aims to provide means to
maximize the accuracy
of probabilistic queries while
minimizing the **probability** of
identifying its records.

Differential Privacy

100 independent values $X_1 \ldots X_{100}$ where $X_i \sim \text{Bern}(p)$

```
# Maximize accuracy, while preserving privacy.
def calculateYi(Xi):
    obfuscate = random()
    if obfuscate:
        return indicator(random())
    else:
        return Xi
```

`random()` returns `True` or `False` with equal likelihood

Differential Privacy

100 independent values X_1 ... X_{100} where $X_i \sim \text{Bern}(p)$

```
# Maximize accuracy, while preserving privacy.
def calculateYi(Xi):
   obfuscate = random()
   if obfuscate:
      return indicator(random())
   else:
      return Xi
```

`random()` returns `True` or `False` with equal likelihood

What is $E[Y_i]$?

$$E[Y_i] = P(Y_i = 1) = \frac{p}{2} + \frac{1}{4}$$

Differential Privacy

100 independent values $X_1 \ldots X_{100}$ where $X_i \sim \text{Bern}(p)$

```
# Maximize accuracy, while preserving privacy.
def calculateYi(Xi):
    obfuscate = random()
    if obfuscate:
        return indicator(random())
    else:
        return Xi
```

`random()` returns `True` or `False` with equal likelihood

Let $Z = \sum_{i=1}^{100} Y_i$

What is the E[Z]?

$$E[Z] = E\left[\sum_{i=1}^{100} Y_i\right] = \sum_{i=1}^{100} E[Y_i] = \sum_{i=1}^{100} \left(\frac{p}{2} + \frac{1}{4}\right) = 50p + 25$$

Differential Privacy

100 independent values $X_1 \ldots X_{100}$ where $X_i \sim \text{Bern}(p)$

```
# Maximize accuracy, while preserving privacy.
def calculateYi(Xi):
    obfuscate = random()
    if obfuscate:
        return indicator(random())
    else:
        return Xi
```

`random()` returns `True` or `False` with equal likelihood

Let $Z = \sum_{i=1}^{100} Y_i$ $E[Z] = 50p + 25$

How do you estimate *p*?

$$p \approx \frac{Z - 25}{50}$$

Challenge: What is the probability that our estimate is good?

Differential Privacy

Story which continues to unfold…

Generalization in Adaptive Data Analysis and Holdout Reuse*

Cynthia Dwork
Microsoft Research

Vitaly Feldman
IBM Almaden Research Center†

Moritz Hardt
Google Research

Toniann Pitassi
University of Toronto

Omer Reingold
Samsung Research America

Aaron Roth
University of Pennsylvania

Professor at Stanford

More Practice!

Computer Cluster Utilization

Computer cluster with *k* servers

- Requests independently go to server *i* with probability p_i
- Expected number of servers that received no request?
- Let event A_i = server *i* receives no requests
- Let Bernoulli B_i be an indicator for A_i
- X = # of events $A_1, A_2, \ldots A_k$ that occur
- Y = # servers that receive ≥ 1 request = *k* – X
- E[Y] after first *n* requests?
- Since requests independent:

$$X = \sum_{i=1}^{k} B_i$$

$$P(A_i) = (1 - p_i)^n$$

$$E[X] = E\left[\sum_{I=1}^{k} B_i\right] = \sum_{i=1}^{k} E[B_i] = \sum_{i=1}^{k} P(A_i) = \sum_{i=1}^{k} (1 - p_i)^n$$

$$E[Y] = k - E[X] = k - \sum_{i=1}^{k} (1 - p_i)^n$$

Amazon Monetized This

amazon web services™

* 52% of Amazons Profits

**More profitable than Amazon's North America commerce operations

http://www.zdnet.com/article/amazons-finds-its-profit-horse-in-aws-why-its-so-disruptive-to-its-old-guard/

Expectation of a sum is easy to calculate!

Can you express a RV as a sum of other RVs?

If you have a RV for a “count” try using Indicator RVs

Hash Tables (aka Toy Collection)

Consider a hash table with *n* buckets

- Each string equally likely to get hashed into any bucket
- Let X = # strings to hash until each bucket ≥ 1 string
- What is E[X]?
- Let X_i = # of trials to get success after *i*-th success
  - where "success" is hashing string to previously empty bucket
  - After *i* buckets have ≥ 1 string, probability of hashing a string to an empty bucket is $p = (n - i) / n$
  - $P(X_i = k) = \frac{n-i}{n}\left(\frac{i}{n}\right)^{k-1}$ equivalently: $X_i \sim \text{Geo}((n - i) / n)$
  - $E[X_i] = 1 / p = n / (n - i)$
- $X = X_0 + X_1 + \ldots + X_{n-1} \Rightarrow E[X] = E[X_0] + E[X_1] + \ldots + E[X_{n-1}]$

$$E[X] = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \ldots + \frac{n}{1} = n\left[\frac{1}{n} + \frac{1}{n-1} + \ldots + 1\right] = O(n \log n)$$

This is your final answer

Break

Conditional Expectation

Conditional Expectation

X and Y are jointly discrete random variables

- Recall conditional PMF of X given Y = y:

$$p_{X|Y}(x \mid y) = P(X = x \mid Y = y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

Define conditional expectation of X given Y = y:

$$E[X \mid Y = y] = \sum_x x P(X = x \mid Y = y) = \sum_x x\, p_{X|Y}(x \mid y)$$

Analogously, jointly continuous random variables:

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \qquad E[X \mid Y = y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x \mid y)\, dx$$

Conditional Expectation

$$\mathrm{E}[X|Y=y] = \sum_x x \cdot P(X=x|Y=y)$$

Roll two 6-sided dice D_1 and D_2

- X = value of $D_1 + D_2$ $\quad\quad$ Y = value of D_2
- What is E[X | Y = 6]?

$$E[X \mid Y=6] = \sum_x xP(X=x \mid Y=6)$$

$$= \left(\frac{1}{6}\right)(7+8+9+10+11+12) = \frac{57}{6} = 9.5$$

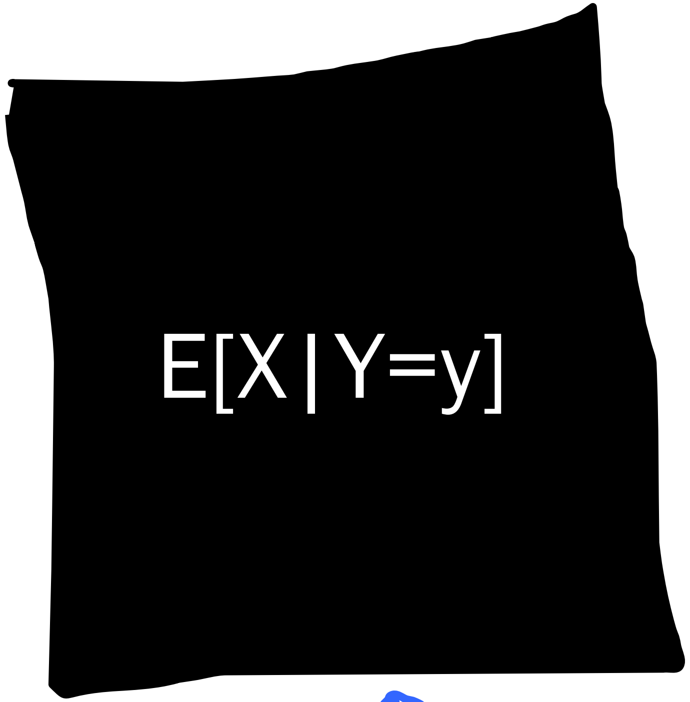
- Intuitively makes sense: 6 + E[value of D_1] = 6 + 3.5

Conditional Expectation

$$E[X|Y=y] = \sum_x x \cdot P(X=x|Y=y)$$

Define $g(y) = E[X \mid Y = y]$

This is just function of y

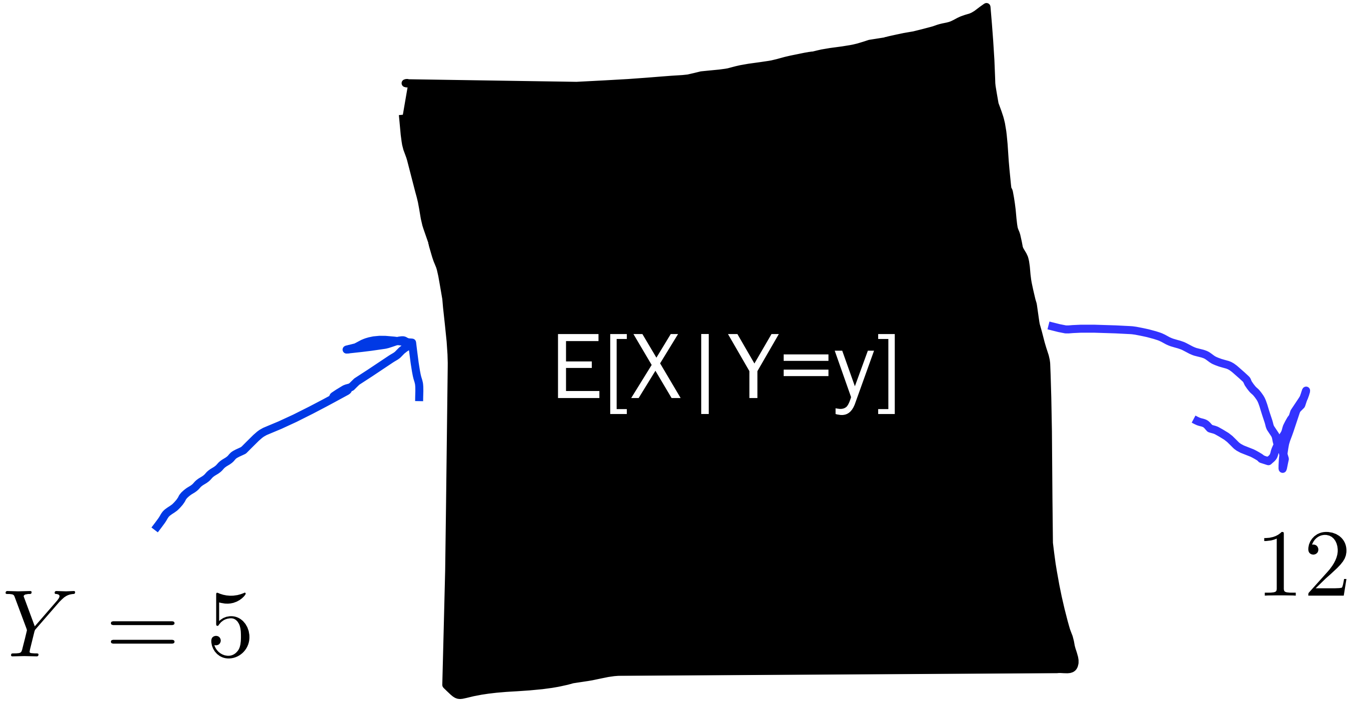


This is a function with Y as input

Conditional Expectation

$$E[X|Y = y] = \sum_x x \cdot P(X = x|Y = y)$$

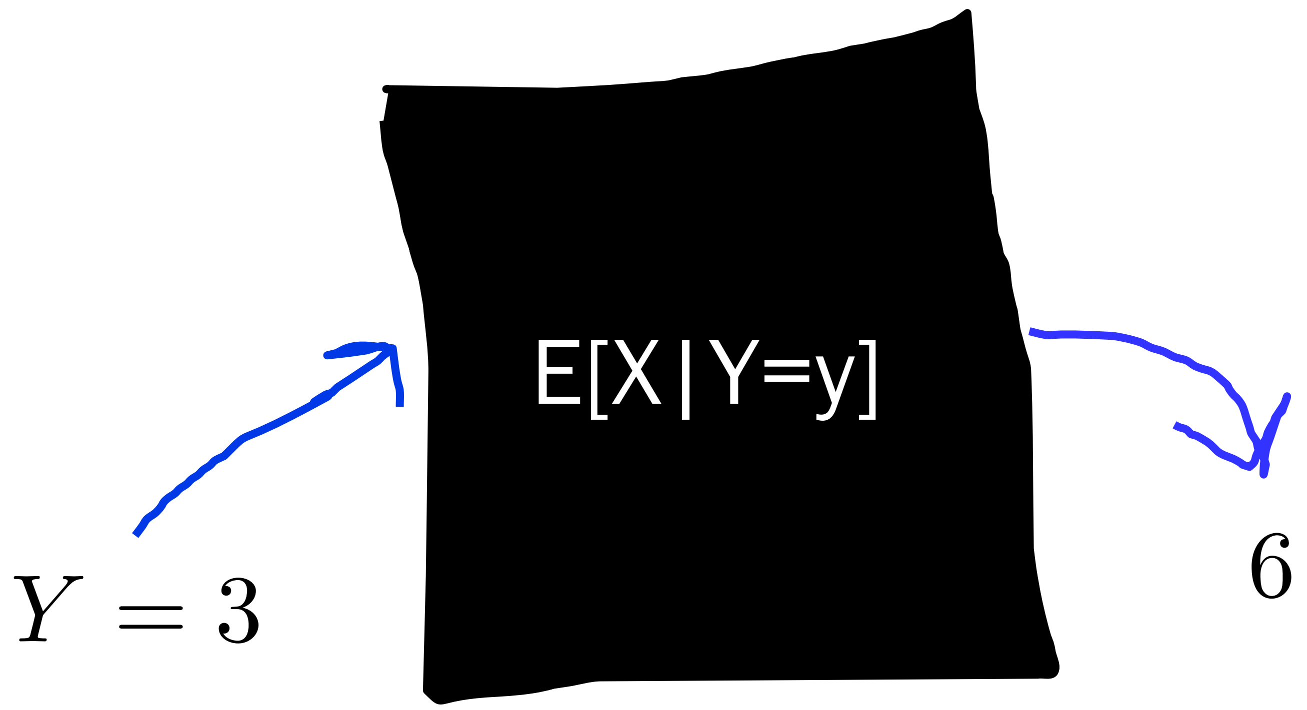
- Define $g(Y) = E[X \mid Y]$
- This is just function of Y



Conditional Expectation

$$E[X|Y = y] = \sum_x x \cdot P(X = x|Y = y)$$

- Define $g(Y) = E[X \mid Y]$
- This is just function of Y



Conditional Expectation

$$\mathrm{E}[X|Y=y] = \sum_x x \cdot P(X=x|Y=y)$$

This is a number:

$$E[X]$$

This is a function of y:

$$E[X|Y=y]$$

$$E[X=5]$$

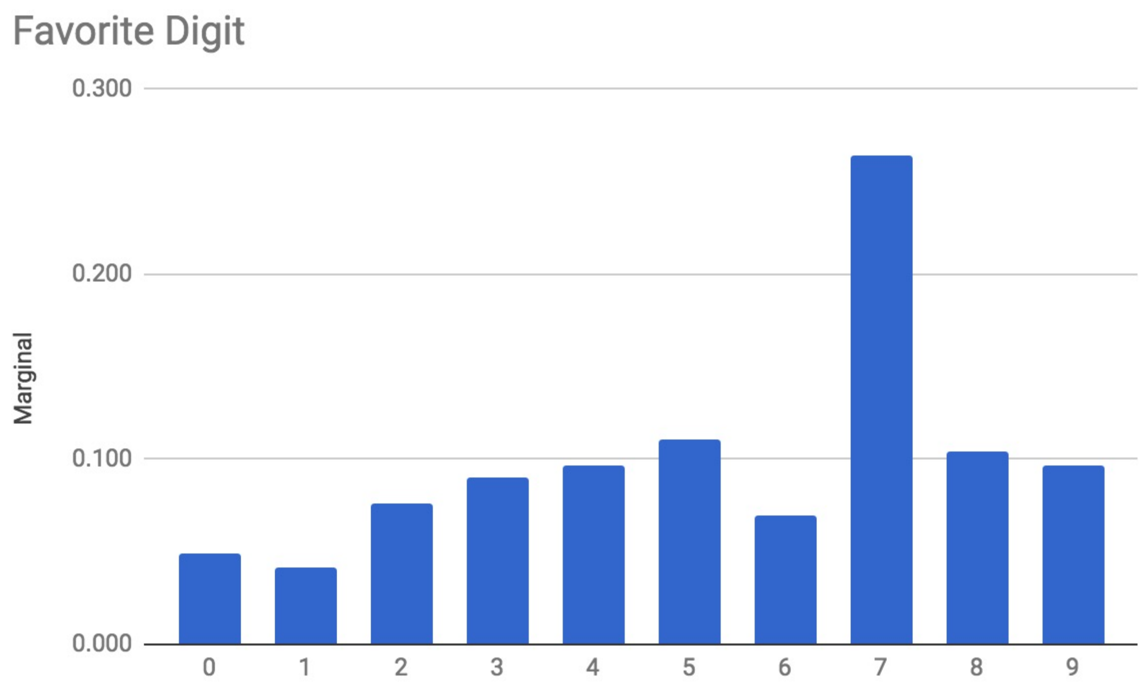
Doesn't make sense. Take expectation of random variables, not events

Expectation

$$E[X|Y=y] = \sum_x x \cdot P(X=x|Y=y)$$

X = favorite number

Y = year in school



E[X] = 0 * 0.05 + ... + 9 * 0.10 = 5.38

Conditional Expectation

$$\mathrm{E}[X|Y=y] = \sum_x x \cdot P(X=x|Y=y)$$

X = favorite number
Y = year in school

E[X | Y] ?

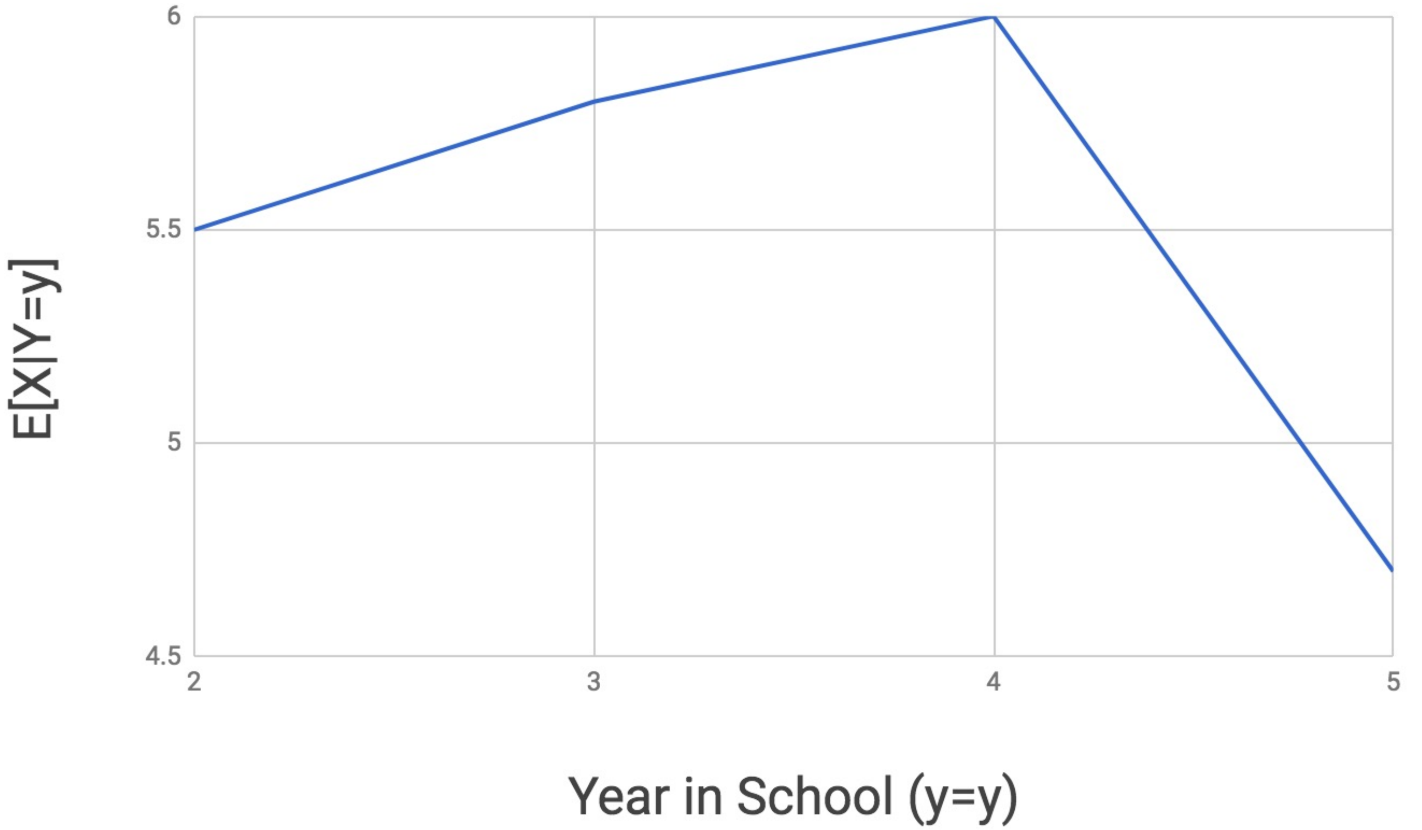
Year in school, Y = y	E[X | Y = y]
2	5.5
3	5.8
4	6.0
5	4.7

Conditional Expectation

$$E[X|Y=y] = \sum_x x \cdot P(X=x|Y=y)$$

X = favorite number
Y = year in school

E[X | Y] ?

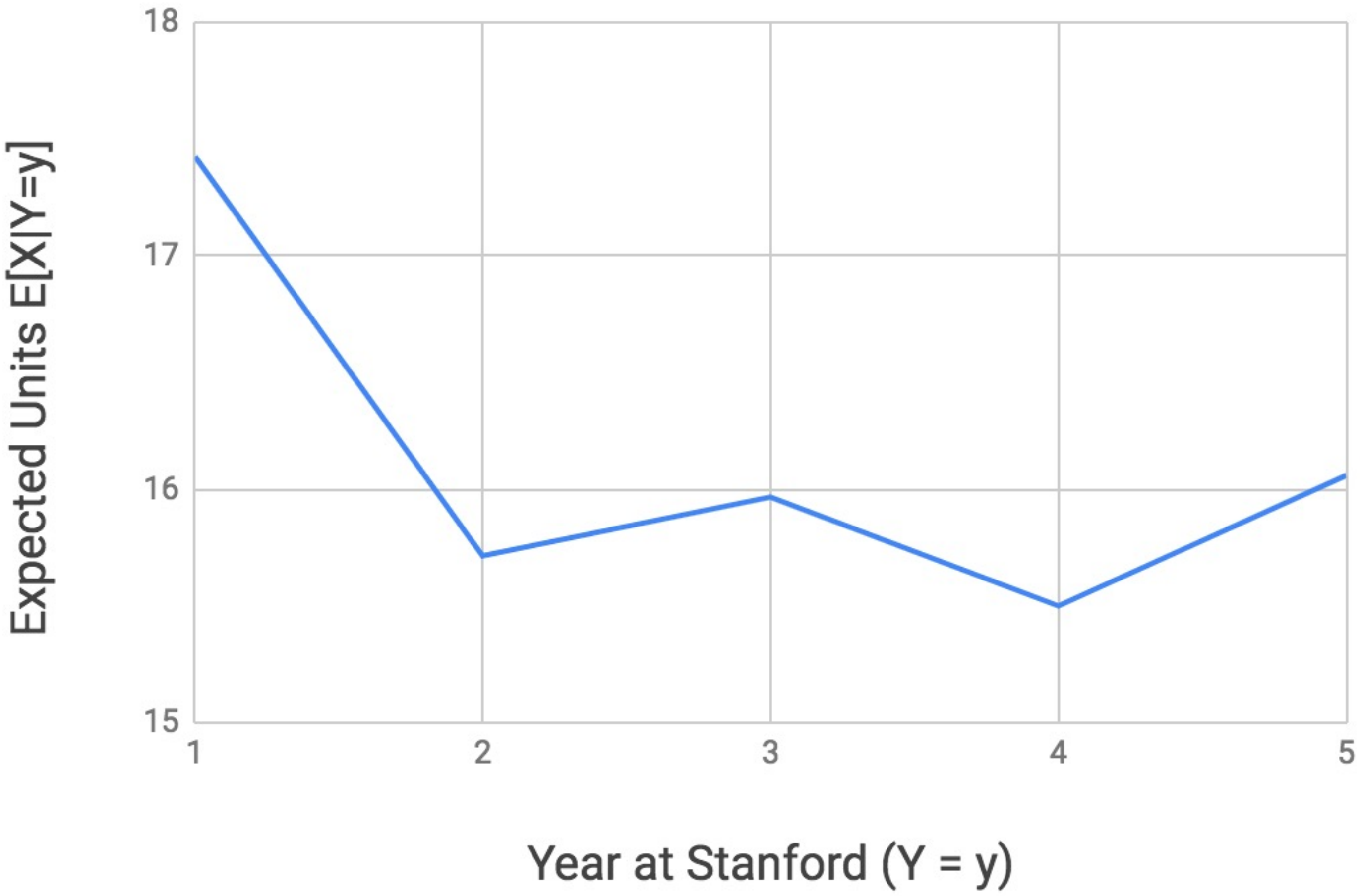


Conditional Expectation

$$E[X|Y = y] = \sum_x x \cdot P(X = x|Y = y)$$

X = units in fall quarter
Y = year in school

E[X | Y] ?



What is this???

$$E\big[E[X|Y]\big]$$

Function of Y

What is this???

$$E_Y\left[E_{X|Y}[X|Y]\right]$$

Function of Y

Law of Total Expectation

$$E[E[X|Y]] = E[X]$$

$$E[E[X|Y]] = \sum_y E[X|Y=y]P(Y=y)$$ g(Y) = E[X|Y]

$$= \sum_y \sum_x xP(X=x|Y=y)P(Y=y)$$ Def of E[X|Y]

$$= \sum_y \sum_x xP(X=x, Y=y)$$ Chain rule!

$$= \sum_x \sum_y xP(X=x, Y=y)$$ I switch the order of the sums

$$= \sum_x x \sum_y P(X=x, Y=y)$$ Move that x outside the y sum

$$= \sum_x xP(X=x)$$ Marginalization

$$= E[X]$$ Def of E[X]

Law of Total Expectation

For any random variable X and any discrete random variable Y

$$E[X] = \sum_y E[X|Y = y]P(Y = y)$$

Analyzing Recursive Code

```
int Recurse() {
    int x = randomInt(1, 3);  // Equally likely values

    if (x == 1) return 3;
    else if (x == 2) return (5 + Recurse());
    else return (7 + Recurse());
}
```

Let Y = value returned by `Recurse()`. What is E[Y]?

$$E[Y] = E[Y \mid X=1]P(X=1) + E[Y \mid X=2]P(X=2) + E[Y \mid X=3]P(X=3)$$

$$E[Y \mid X=1] = 3$$

$$E[Y \mid X=2] = E[5+Y] = 5 + E[Y]$$

$$E[Y \mid X=3] = E[7+Y] = 7 + E[Y]$$

$$E[Y] = 3(1/3) + (5+E[Y])(1/3) + (7+E[Y])(1/3) = (1/3)(15 + 2E[Y])$$

$$E[Y] = 15$$

Protip: do this in CS161

Uncertainty Theory

- Beta Distributions
- Thompson Sampling
- Adding Random Vars
- Central Limit Theorem
- Sampling
- Bootstrapping
- Great Expectations

Where are we in CS109?

On Friday...

Counting Theory | Core Probability | Random Variables | Probabilistic Models | Uncertainty Theory | Machine Learning