# Great Expectations

Chris Piech and Jerry Cain
CS109, Stanford University

# Quick slide reference

# Where are we now? A roadmap of CS109

Last Fri: Joint distributions

$p_{X,Y}(x, y)$

Wed: Statistics of multiple RVs!

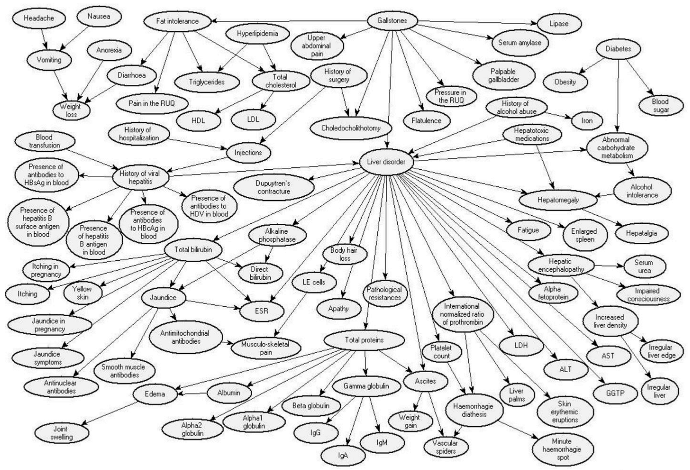
$\text{Var}(X + Y)$

$E[X + Y]$

$\text{Cov}(X, Y)$

$\rho(X, Y)$

Next Week: Modeling with Bayesian Networks



Today:
Conditional distributions

$p_{X|Y}(x|y)$

$E[X|Y]$

Time to kick it up a notch!

# Hiring and Engineer


[xkcd by Randall Munroe]

Your company has one job opening for a software engineer.

You have *n* candidates. But you have to say yes/no **immediately** after each interview!

What algorithm will maximize the probability that you hire the top candidate?

# Expectation of the 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

# Differential Privacy

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

Cynthia Dwork's celebrity lookalike is Cynthia Dwork.

# 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 $X_1 \dots X_{100}$ 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 $X_1$ ... $X_{100}$ where $X_i$ ~ 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 $X_1 \dots X_{100}$ 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 $Z = \sum_{i=1}^{100} Y_i$ $\quad E[Z] = 50p + 25$

How do you estimate *p*?

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

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

# Discrete conditional distributions

# Discrete conditional distributions

Recall the definition of the conditional probability of event $E$ given event $F$:

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

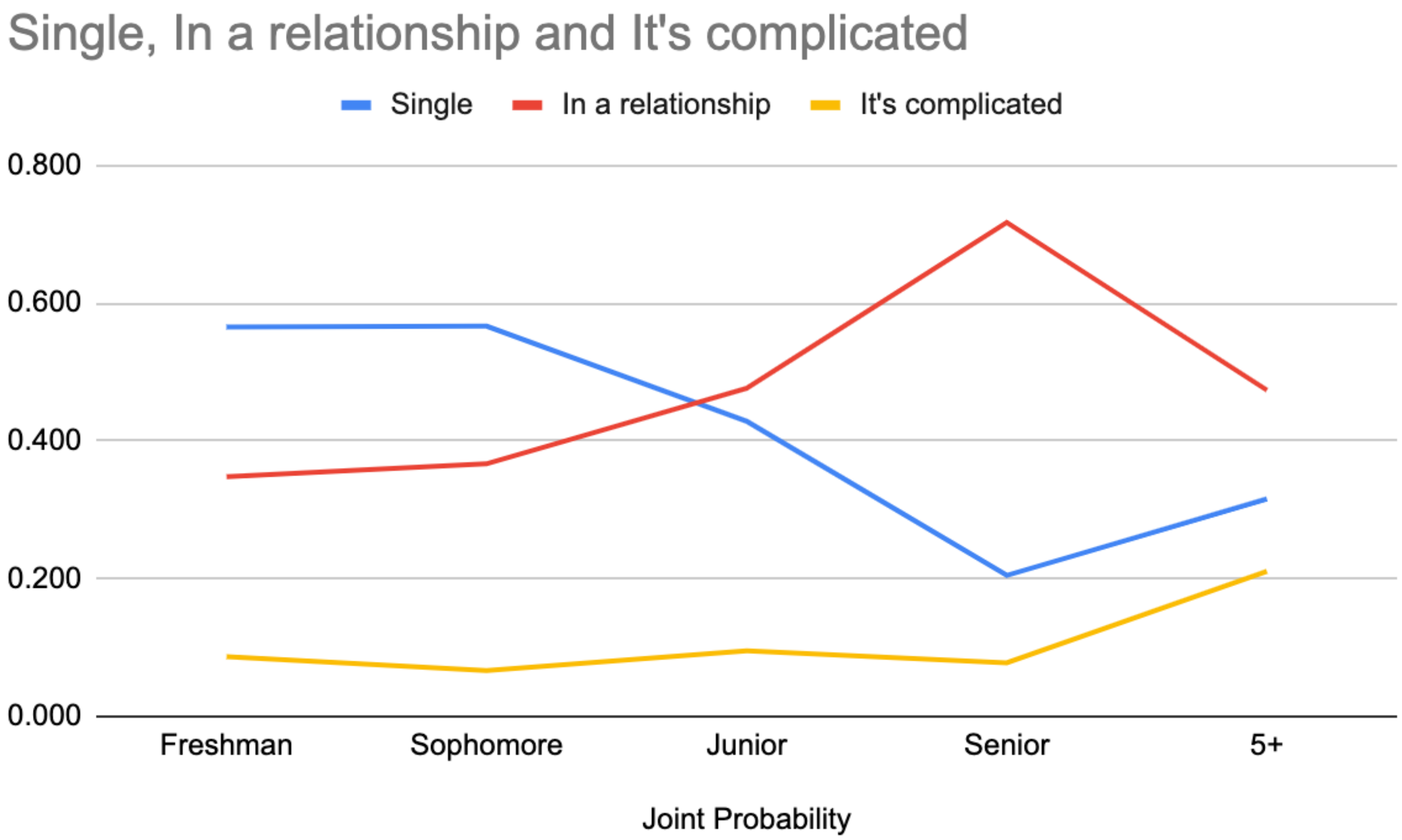
For discrete random variables $X$ and $Y$, the **conditional PMF** of $X$ given $Y$ is

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

Different notation, same idea:

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

# Dating at Stanford



# Discrete probabilities of CS109

Each student responds with:

Year $Y$

- 1: Frosh/Soph
- 2: Jr/Sr
- 3: Co-term/grad/NDO

Timezone $T$ (12pm California time in my timezone is):

- $-1$: AM
- 0: noon
- 1: PM

Joint PMF

| | $Y=1$ | $Y=2$ | $Y=3$ |
|---|---|---|---|
| $T=-1$ | .06 | .01 | .01 |
| $T=0$ | .29 | .14 | .09 |
| $T=1$ | .30 | .08 | .02 |

$P(Y=3, T=1)$

Joint PMFs sum to 1.

# Discrete probabilities of CS109

The below are **conditional probability tables** for conditional PMFs
(A) $P(Y = y|T = t)$ and (B) $P(T = t|Y = y)$.

1. Which is which?
2. What's the missing probability?

<u>Joint PMF</u>

| | $Y = 1$ | $Y = 2$ | $Y = 3$ |
|---|---|---|---|
| $T = -1$ | .06 | .01 | .01 |
| $T = 0$ | .29 | .14 | .09 |
| $T = 1$ | .30 | .08 | .02 |

| | $Y = 1$ | $Y = 2$ | $Y = 3$ |
|---|---|---|---|
| $T = -1$ | .09 | .04 | .08 |
| $T = 0$ | .45 | .61 | .75 |
| $T = 1$ | .46 | .35 | .17 |

| | $Y = 1$ | $Y = 2$ | $Y = 3$ |
|---|---|---|---|
| $T = -1$ | .75 | .125 | ? |
| $T = 0$ | .56 | .27 | .17 |
| $T = 1$ | .75 | .2 | .05 |

# Discrete probabilities of CS109

The below are **conditional probability tables** for conditional PMFs
(A) $P(Y = y|T = t)$ and (B) $P(T = t|Y = y)$.

1. Which is which?
2. What's the missing probability?

Joint PMF

| | $Y = 1$ | $Y = 2$ | $Y = 3$ |
|---|---|---|---|
| $T = -1$ | .06 | .01 | .01 |
| $T = 0$ | .29 | .14 | .09 |
| $T = 1$ | .30 | .08 | .02 |

(B) $P(T = t|Y = y)$

| | $Y = 1$ | $Y = 2$ | $Y = 3$ |
|---|---|---|---|
| $T = -1$ | .09 | .04 | .08 |
| $T = 0$ | .45 | .61 | .75 |
| $T = 1$ | .46 | .35 | .17 |

.30/(.06+.29+.30)

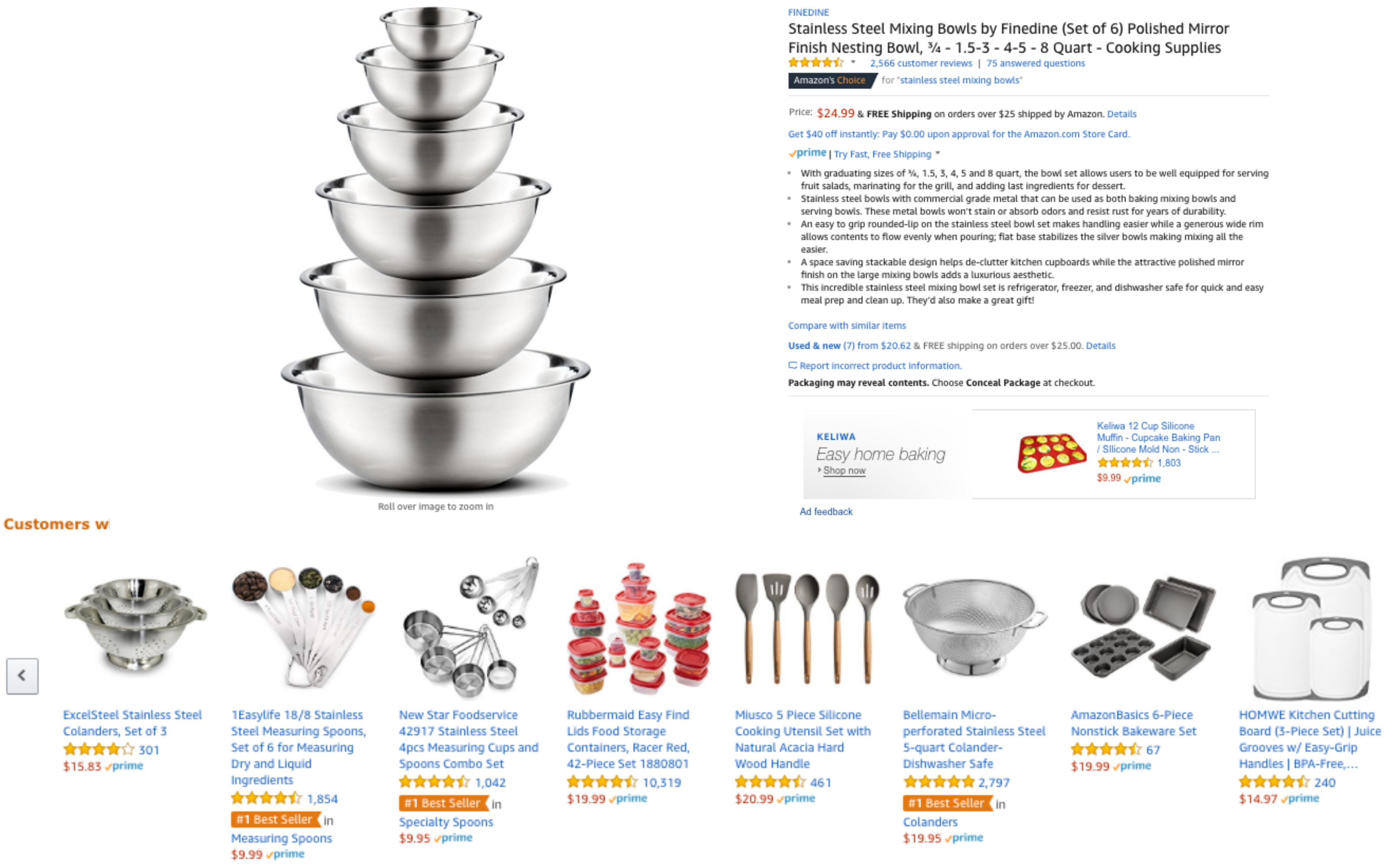
(A) $P(Y = y|T = t)$

| | $Y = 1$ | $Y = 2$ | $Y = 3$ |
|---|---|---|---|
| $T = -1$ | .75 | .125 | .125 |
| $T = 0$ | .56 | .27 | .17 |
| $T = 1$ | .75 | .2 | .05 |

1-.75-.125

Conditional PMFs also sum to 1 conditioned on different events!

# Extended to Amazon




P(bought item $X$ | bought item $Y$)

# Quick check

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

Number or function?

1. $P(X = 2|Y = 5)$
2. $P(X = x|Y = 5)$
3. $P(X = 2|Y = y)$
4. $P(X = x|Y = y)$

True or false?

5. $\sum_x P(X = x|Y = 5) = 1$
6. $\sum_y P(X = 2|Y = y) = 1$
7. $\sum_x \sum_y P(X = x|Y = y) = 1$
8. $\sum_x \left( \sum_y P(X = x|Y = y) P(Y = y) \right) = 1$

# Quick check

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

Number or function?

1. $P(X = 2|Y = 5)$

   number

2. $P(X = x|Y = 5)$

   1-D function

3. $P(X = 2|Y = y)$

   1-D function

4. $P(X = x|Y = y)$

   2-D function

True or false?

5. $\sum_x P(X = x|Y = 5) = 1$ true

6. $\sum_y P(X = 2|Y = y) = 1$ false

7. $\sum_x \sum_y P(X = x|Y = y) = 1$ false

8. $\sum_x \left( \sum_y P(X = x|Y = y) P(Y = y) \right) = 1$ true

# Independent RVs, defined another way

If $X$ and $Y$ are independent discrete random variables, then $\forall x, y$:

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

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

Note for conditional expectation, independent $X$ and $Y$ implies

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

# Conditional Expectation

# Conditional Expectation

Conditional Distributions Expectation

# Conditional expectation

Recall the the conditional PMF of $X$ given $Y = y$:

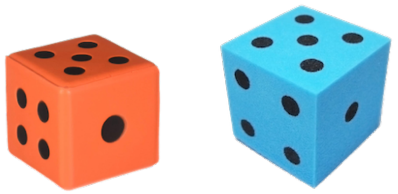
$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

The conditional expectation of $X$ given $Y = y$ is

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

# It's been so long, our dice friends

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



- Roll two 6-sided dice.
- Let roll 1 be $D_1$, roll 2 be $D_2$.
- Let $S$ = value of $D_1 + D_2$.

1. What is $E[S|D_2 = 6]$?

$$E[S|D_2 = 6] = \sum_{x=7}^{12} xP(S = x|D_2 = 6)$$

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

$$= \frac{57}{6} = 9.5$$

Intuitively: $6 + E[D_1] = 6 + 3.5 = 9.5$

Let's prove this!

# Properties of conditional expectation

1. LOTUS:

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

2. Sum of conditional expectation:

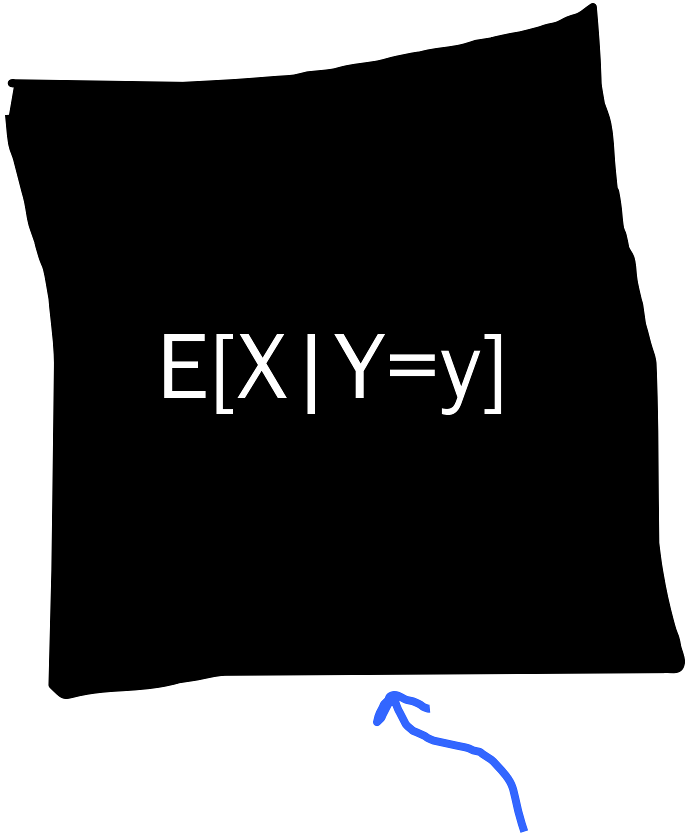
$$E\left[\sum_{i=1}^{n} X_i \mid Y = y\right] = \sum_{i=1}^{n} E[X_i|Y = y]$$

3. Law of total expectation

# Conditional Expectation Function

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

This is just function of $Y$

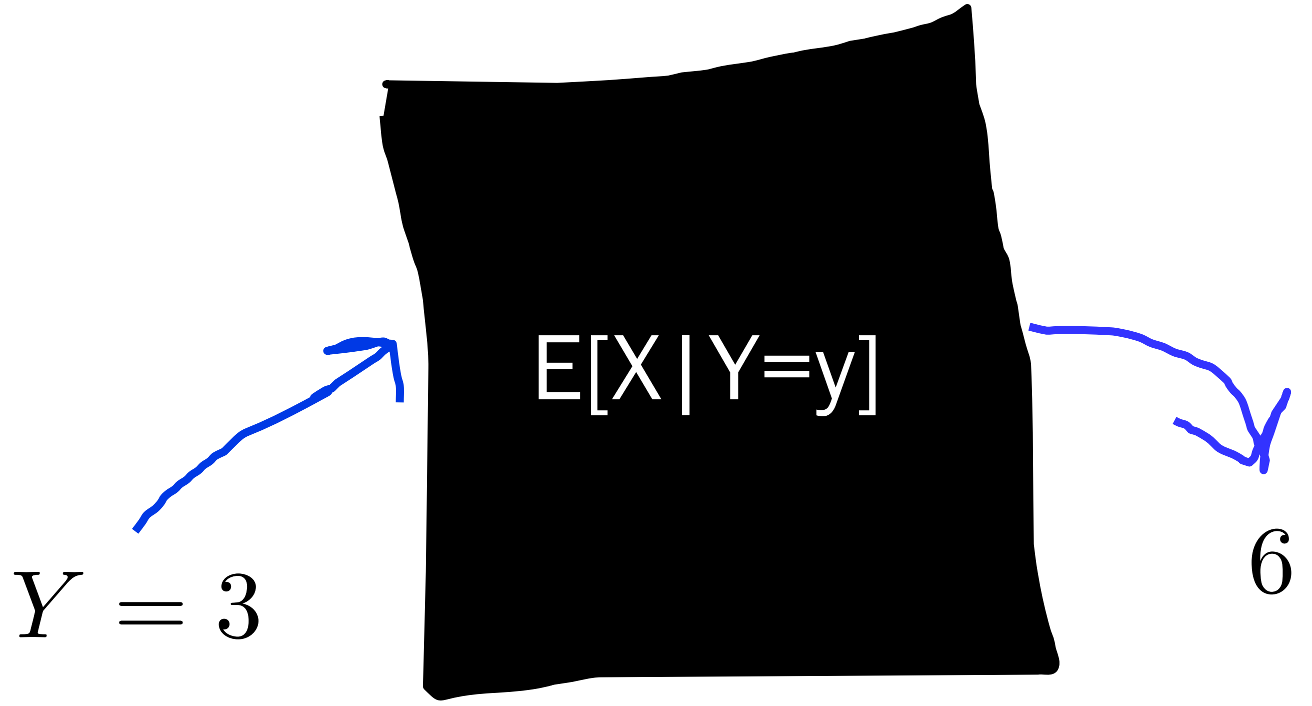



This is a function with Y as input

# Conditional Expectation Function

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

This is just function of Y



# Conditional Expectation Function

This is a number:

$$E[X]$$

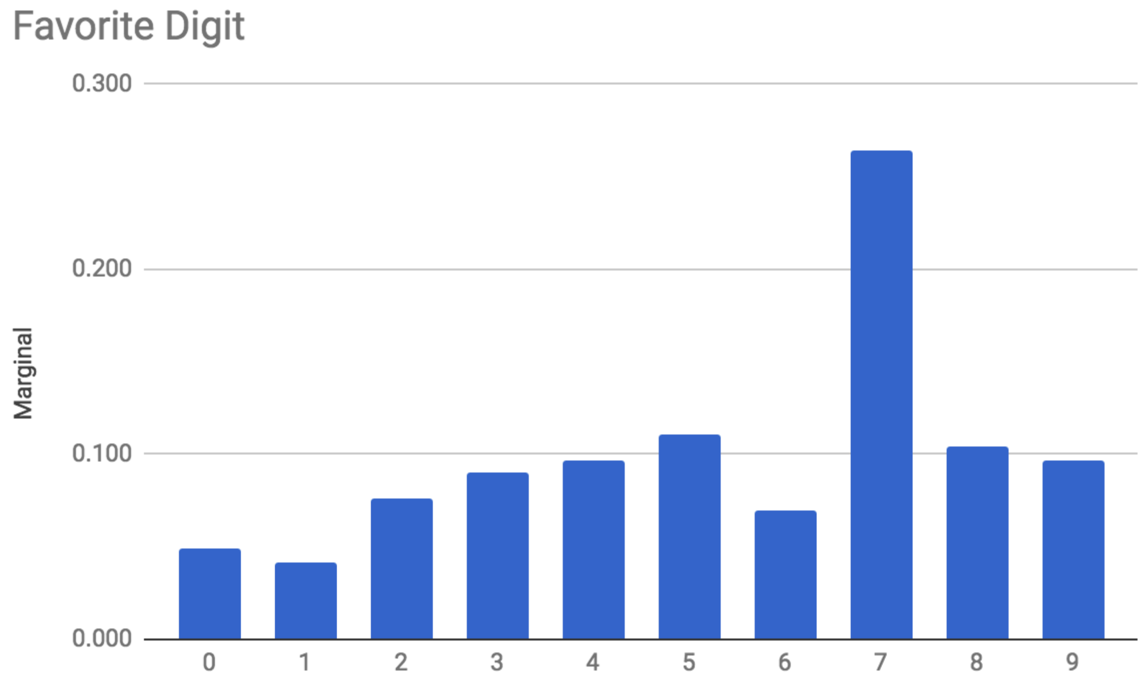
This is a function of y:

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

$E[X = 5]$ Doesn't make sense. Take expectation of random variables, not events

# Conditional Expectation Example

X = favorite number
Y = year in school




$E[X] = 0 * 0.05 + \ldots + 9 * 0.10 = 5.38$

# Conditional Expectation Example

X = favorite number
Y = year in school

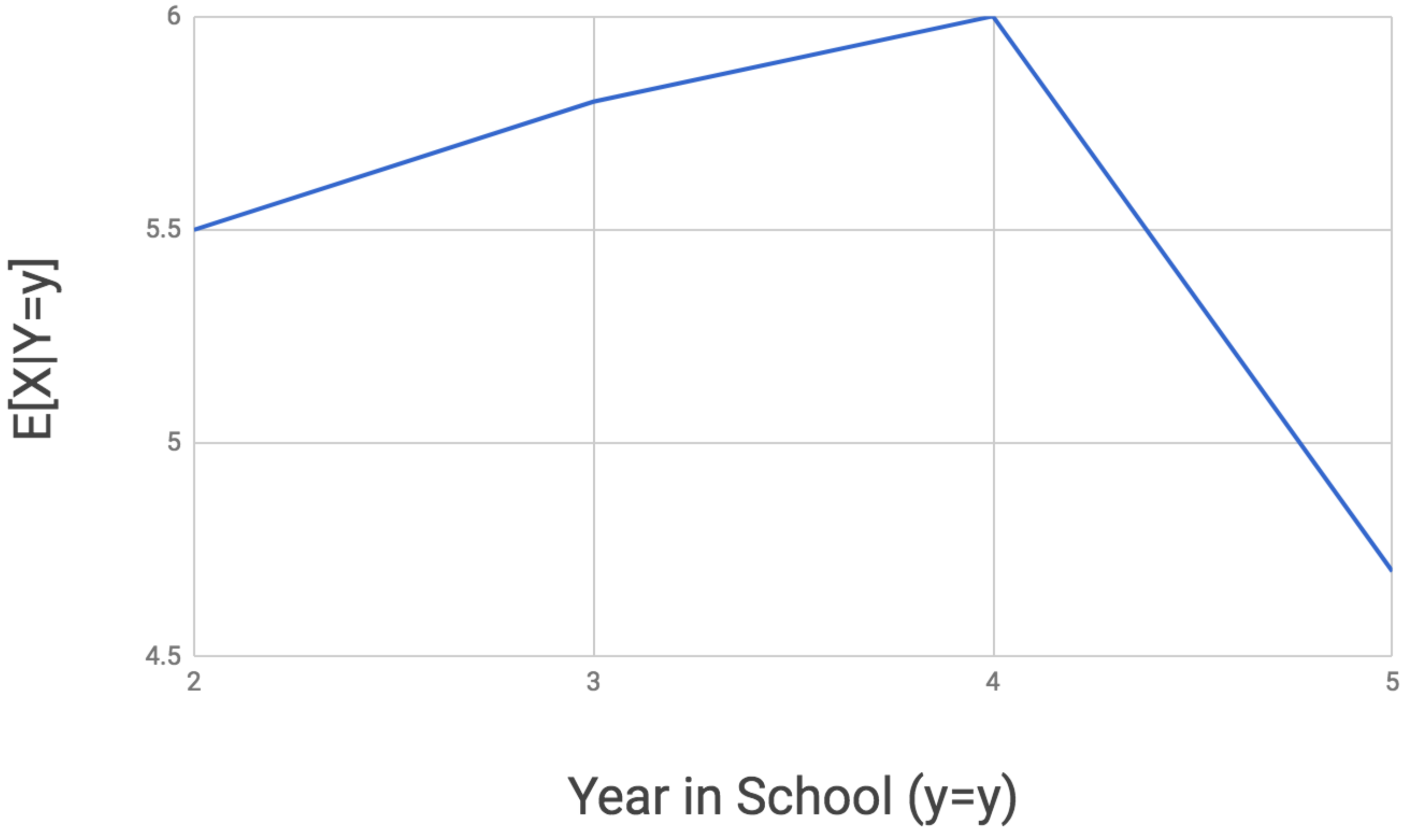
E[X | Y] ?

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

# Conditional Expectation Example

X = favorite number
Y = year in school

E[X | Y] ?

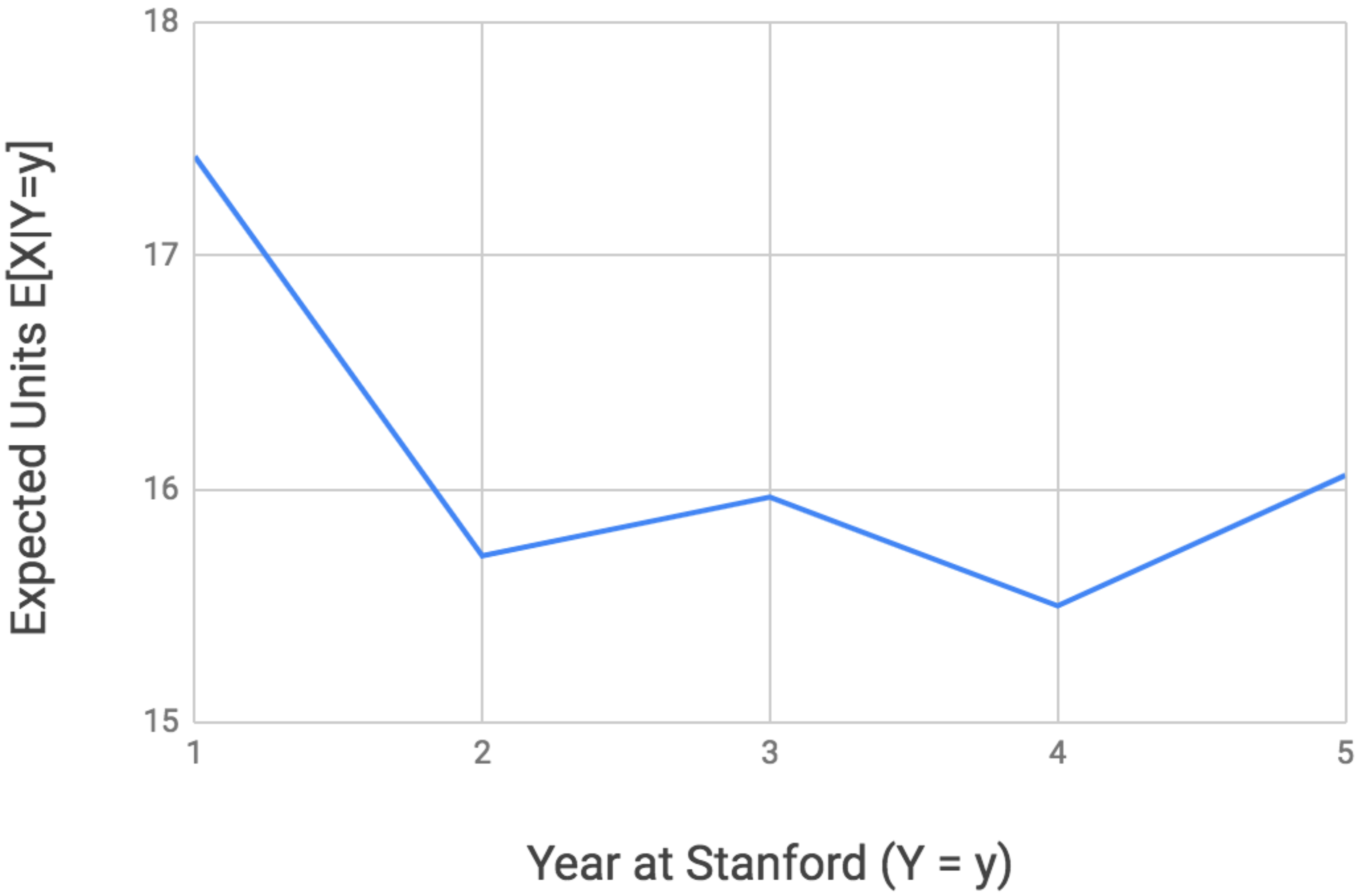


# Conditional Expectation Example

X = units in fall quarter

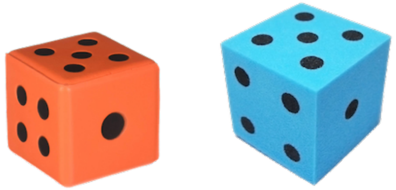
Y = year in school

E[X | Y] ?



# It's been so long, our dice friends

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

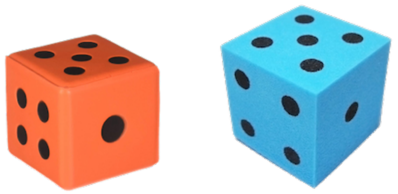


- Roll two 6-sided dice.
- Let roll 1 be $D_1$, roll 2 be $D_2$.
- Let $S$ = value of $D_1 + D_2$.

1. What is $E[S|D_2 = 6]$? $\frac{57}{6} = 9.5$
2. What is $E[S|D_2]$?
   A. A function of $S$
   B. A function of $D_2$
   C. A number
3. Give an expression for $E[S|D_2]$.

# It’s been so long, our dice friends

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



- Roll two 6-sided dice.
- Let roll 1 be $D_1$, roll 2 be $D_2$.
- Let $S$ = value of $D_1 + D_2$.

1. What is $E[S|D_2 = 6]$?

$$\frac{57}{6} = 9.5$$

2. What is $E[S|D_2]$?
   - A. A function of $S$
   - B. A function of $D_2$
   - C. A number
3. Give an expression for $E[S|D_2]$.

$$E[S|D_2 = d_2] = E[D_1 + d_2 | D_2 = d_2]$$

$$= \sum_{d_1} (d_1 + d_2) P(D_1 = d_1 | D_2 = d_2)$$

$$= \sum_{d_1} d_1 P(D_1 = d_1) + d_2 \sum_{d_1} P(D_1 = d_1)$$

($D_1 = d_1$, $D_2 = d_2$ independent events)

$$= E[D_1] + d_2 = 3.5 + d_2$$

$$E[S|D_2] = 3.5 + D_2$$

# Law of Total Expectation

# Properties of conditional expectation

1. LOTUS:

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

2. Linearity of conditional expectation:

$$E\left[\sum_{i=1}^{n} X_i \mid Y = y\right] = \sum_{i=1}^{n} E[X_i|Y=y]$$

3. Law of total expectation:

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

what?!

# Law of Total Expectation

$E[X]$ can be calculated using $E[X|Y]$

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

# Proof of Law of Total Expectation

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

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

# Another way to compute $E[X]$

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

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

If we only have a conditional PMF of $X$ on some discrete variable $Y$, we can compute $E[X]$ as follows:

1. Compute expectation of $X$ given some value of $Y = y$
2. Repeat step 1 for all values of $Y$
3. Compute a weighted sum (where weights are $P(Y = y)$)

```
def recurse():
  if (random.random() < 0.5):
      return 3
  else: return (2 + recurse())
```

Useful for analyzing recursive code!!

# Quick check

1. $E[X]$
2. $E[X, Y]$
3. $E[X + Y]$
4. $E[X|Y]$
5. $E[X|Y = 6]$
6. $E[X = 1]$

7.* $E[Y|X = x]$

A. value
B. one RV, function on $Y$
C. one RV, function on $X$
D. two RVs, function on $X$ and $Y$
E. doesn't make sense

# Analyzing recursive code

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

```
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let $Y$ = return value of `recurse()`.
What is $E[Y]$?

# Analyzing recursive code

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

```
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let $Y$ = return value of `recurse()`.

What is $E[Y]$?

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

$$E[Y|X=1] = 3$$

When $X = 1$, return 3.

# Analyzing recursive code

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

```
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let $Y$ = return value of `recurse()`.
What is $E[Y]$?

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

$$E[Y|X=1] = 3$$

What is $E[Y|X=2]$?

A. $E[5] + Y$
B. $E[Y+5] = 5 + E[Y]$
C. $5 + E[Y|X=2]$

(by yourself)

# Analyzing recursive code

If $Y$ discrete

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

```
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let $Y$ = return value of `recurse()`.

What is $E[Y]$?

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

$E[Y|X=1] = 3$

When $X = 2$, return $5 +$ a future return value of `recurse()`.

What is $E[Y|X=2]$?

A. $E[5] + Y$

B. $E[Y+5] = 5 + E[Y]$

C. $5 + E[Y|X=2]$

# Analyzing recursive code

If $Y$ discrete

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

```
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let $Y$ = return value of `recurse()`.
What is $E[Y]$?

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

$E[Y|X=1] = 3$

$E[Y|X=2] = E[5+Y]$

When $X = 3$, return $7 +$ a future return value of `recurse()`.

$$E[Y|X=3] = E[7+Y]$$

# Analyzing recursive code

If $Y$ discrete

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

```
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let $Y$ = return value of `recurse()`.
What is $E[Y]$?

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

$$E[Y|X=1] = 3 \qquad E[Y|X=2] = E[5+Y] \qquad E[Y|X=3] = E[7+Y]$$

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

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

$$E[Y] = 15$$

On your own: What is Var$(Y)$?

# Hiring and Engineer


Your company has one job opening for a software engineer.

You have *n* candidates. But you have to say yes/no **immediately** after each interview!

Proposed algorithm: reject the first *k* and accept the next one who is better than all of them.

What's the best value of *k*?

xkcd by Randall Munroe

# Hiring and Engineer

*n* candidates, must say yes/no **immediately** after each interview.
Reject the first *k,* accept the next who is better than all of them.
What's the best value of *k*?

**B**: event that you hire the best engineer
**X**: position of the best engineer on the interview schedule

---

What is the $P(B|X = i)$?

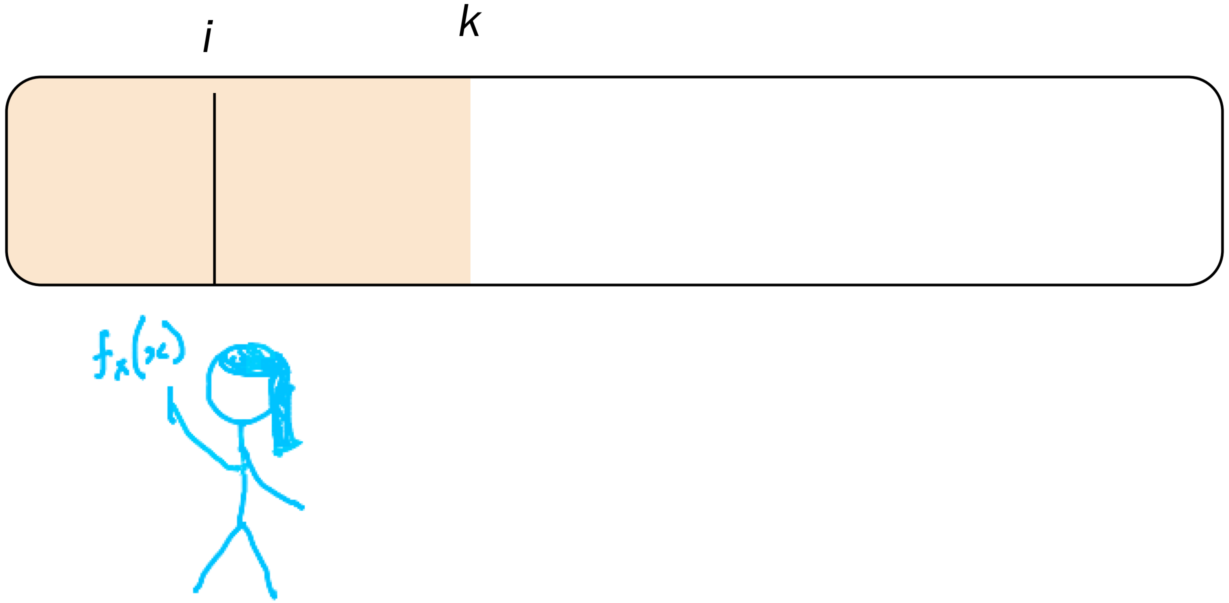
# Hiring and Engineer

*n* candidates, must say yes/no **immediately** after each interview.
Reject the first *k*, accept the next who is better than all of them.
What's the best value of *k*?

**B**: event that you hire the best engineer
**X**: position of the best engineer on the interview schedule

---

What is the $P(B|X = i)$?



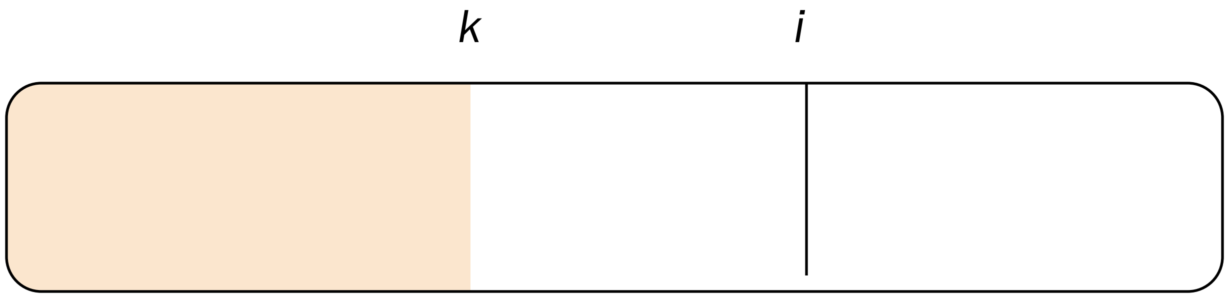
# Hiring and Engineer

*n* candidates, must say yes/no **immediately** after each interview.
Reject the first *k,* accept the next who is better than all of them.
What's the best value of *k*?

**B**: event that you hire the best engineer
**X**: position of the best engineer on the interview schedule

---

What is the $P(B|X = i)$?




Hint: where is the best among the first *i* – 1 candidates?

# Hiring and Engineer

*n* candidates, must say yes/no **immediately** after each interview.
Reject the first *k*, accept the next who is better than all of them.
What's the best value of *k*?

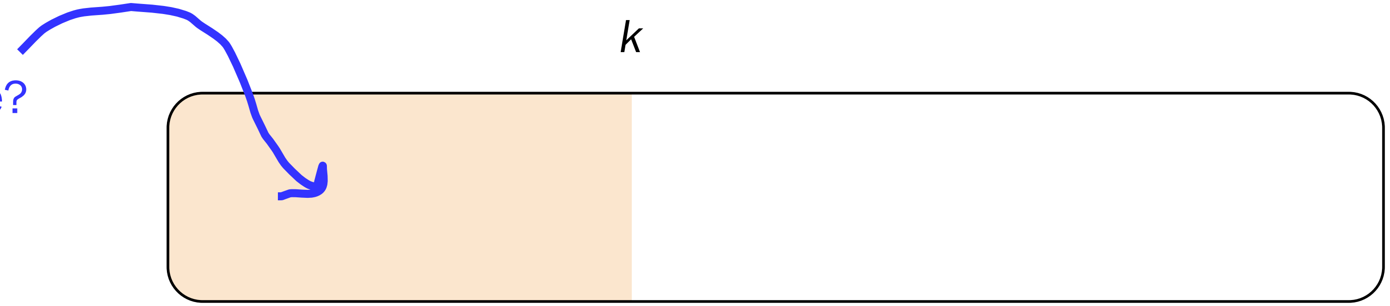
**B**: event that you hire the best engineer
**X**: position of the best engineer on the interview schedule

---

What is the $P(B|X = i)$?

*k*

Here?



Hint: where is the best among the first *i* – 1 candidates?

# Hiring and Engineer

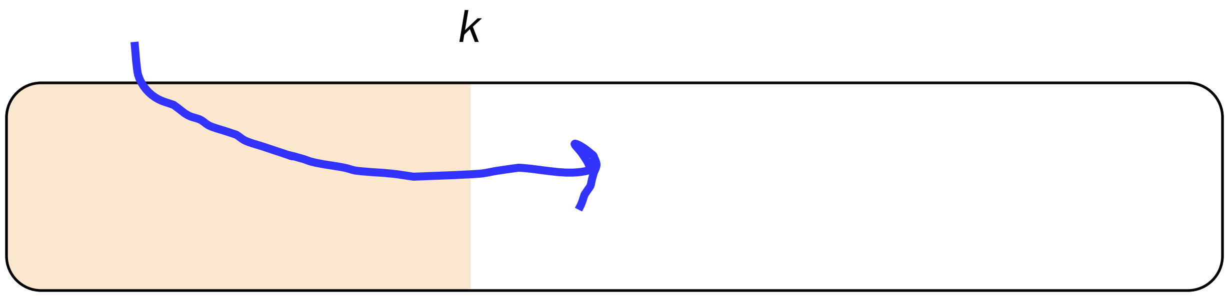
*n* candidates, must say yes/no **immediately** after each interview.
Reject the first *k*, accept the next who is better than all of them.
What's the best value of *k*?

**B**: event that you hire the best engineer
**X**: position of the best engineer on the interview schedule

---

Here?

What is the $P(B|X = i)$?



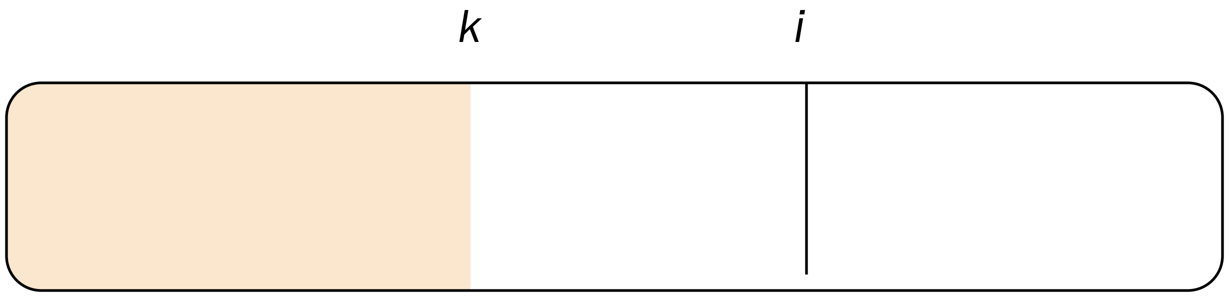

Hint: where is the best among the first *i* – 1 candidates?

# Hiring and Engineer

*n* candidates, must say yes/no **immediately** after each interview.
Reject the first *k*, accept the next who is better than all of them.
What's the best value of *k*?

**B**: event that you hire the best engineer
**X**: position of the best engineer on the interview schedule

$$P(B|X=i) = \frac{k}{i-1} \text{ if } i > k$$




Hint: where is the best among the first *i* – 1 candidates?

# Hiring and Engineer

*n* candidates, must say yes/no **immediately** after each interview.
Reject the first *k*, accept the next who is better than all of them.
What's the best value of *k*?

**B**: event that you hire the best engineer
**X**: position of the best engineer on the interview schedule

---

$$P_k(B) = \sum_{i=1}^{n} P_k(B|X=i)P(X=i)$$

By the law of total expectation

$$= \frac{1}{n}\sum_{i=1}^{n} P_k(B|X=i)$$

$$= \frac{1}{n}\sum_{i=k+1}^{n} \frac{k}{i-1}$$

since we know $P_k(Best|X=i)$

$$\approx \frac{1}{n}\int_{i=k+1}^{n} \frac{k}{i-1}di$$

By Riemann Sum approximation

$$= \frac{k}{n}\ln(i=1)\Big|_{k+1}^{n} = \frac{k}{n}\ln\frac{n-1}{k} \approx \frac{k}{n}\ln\frac{n}{k}$$

# Hiring and Engineer

*n* candidates, must say yes/no **immediately** after each interview.
Reject the first *k*, accept the next who is better than all of them.
What's the best value of *k*?

**B**: event that you hire the best engineer
**X**: position of the best engineer on the interview schedule

---

$$P_k(B) = \sum_{i=1}^{n} P_k(B|X=i)P(X=i)$$

By the law of total expectation

$$\approx \frac{k}{n} \ln \frac{n}{k}$$

Fun fact. Optimized when:

$$k = \frac{n}{e}$$