# CS 70 Su23: Lecture 20

Expectation II: Conditional expectation, Iterated expectation, Wald’s identity

Berkeley
UNIVERSITY OF CALIFORNIA

# A quick recap

- Joint distribution: defines the probability distribution of two (or more) random variables
  - example:

| Pr[X = a, Y = b] | a = 0 | a = 1 |
| --- | --- | --- |
| b = 0 | $\frac{1}{3}$ | $\frac{1}{6}$ |
| b = 1 | $\frac{1}{3}$ | $\frac{1}{6}$ |

- example:
  - Pr[X = 0]

    = Pr[X = 0, Y = 0] + Pr[X = 0, Y = 1]

    = $\frac{1}{3} + \frac{1}{3}$

    = $\frac{2}{3}$

- $\Pr[X = a] = \sum_b \Pr[X = a, Y = b]$
  - "Sum out" Y to get the *marginal distribution* of X
  - if this looks kind of like the Total Probability Rule, that's no coincidence!

Berkeley
UNIVERSITY OF CALIFORNIA

# It's like expectation, but… conditional

Conditional probability for joint distributions:

- $\Pr[X = x \mid Y = y] \quad \dfrac{\Pr[X = x, Y = y]}{\Pr[Y = y]}$

**Conditional expectation:**

- $E[X \mid Y = y] = \sum_x x \cdot \Pr[X = x \mid Y = y]$

Berkeley
UNIVERSITY OF CALIFORNIA

# Conditional Expectation

Some properties of conditional expectation:

- $E[X_1 + X_2 \mid Y] = E[X_1 \mid Y] + E[X_2 \mid Y]$ (additive)
- $E[aX + b \mid Y] = aE[X \mid Y] + b$ (linear)
- $E[f(X) \mid X] = f(X)$ (constant)
- $E[f(Y) \cdot X \mid Y] = f(Y) \cdot E[X \mid Y]$

# Conditional expectation (example)

| | x = 0 | x = 1 | | y = 0 | y = 1 |
|---|---|---|---|---|---|
| Pr[X = x \| Y = 0] | ⅓ | ⅔ | Pr[Y = y] | ¼ | ¾ |
| Pr[X = x \| Y = 1] | ½ | ½ | | | |

What is Pr[X]?

$$\Pr[X = x] = \sum_y \Pr[X = x, Y = y]$$

$$= \sum_y \Pr[X = x \mid Y = y] \cdot \Pr[Y = y]$$

$$\Pr[X = 0] = \Pr[X = 0 \mid Y = 0] \cdot \Pr[Y = 0] + \Pr[X = 0 \mid Y = 1] \cdot \Pr[Y = 1]$$

$$= \tfrac{1}{3} \cdot \tfrac{1}{4} + \tfrac{1}{2} \cdot \tfrac{3}{4} = 11/24$$

You've seen all of this before! We're just using RVs now instead of events

Berkeley UNIVERSITY OF CALIFORNIA

# Conditional expectation (example)

| | x = 0 | x = 1 | | y = 0 | y = 1 |
|---|---|---|---|---|---|
| Pr[X = x \| Y = 0] | ⅓ | ⅔ | Pr[Y = y] | ¼ | ¾ |
| Pr[X = x \| Y = 1] | ½ | ½ | | | |

What is E[X | Y = y]?

$E[X \mid Y = 0] = \sum_x x \cdot \Pr[X = x \mid Y = 0]$

$= 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}$

$E[X \mid Y = 1] = \sum_x x \cdot \Pr[X = x \mid Y = 1]$

$= 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$

| | |
|---|---|
| E[X \| Y = 0] | ⅔ |
| E[X \| Y = 1] | ½ |

Berkeley UNIVERSITY OF CALIFORNIA

# Iterated expectation

What is $E[X|Y]$?

It's a random variable! (for a given value y, specifies what $E[X \mid Y = y]$ is)

Because it's a random variable, we can compute its expectation

- $E[E[X \mid Y]] = \sum_y E[X \mid Y = y] \cdot \Pr[Y = y]$
    - $E[X \mid Y = y] = \sum_x x \cdot \Pr[X = x \mid Y = y]$
- $\sum_y (\sum_x x \cdot \Pr[X = x \mid Y = y]) \cdot \Pr[Y = y]$
- $\sum_y \sum_x x \cdot \Pr[X = x \mid Y = y] \cdot \Pr[Y = y]$
- $\sum_x x \cdot \sum_y \Pr[X = x \mid Y = y] \cdot \Pr[Y = y]$
- $\sum_x x \cdot \Pr[X = x]$
- $E[X]$

**Iterated expectation:**

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

also knowns as the Total Expectation Rule

# Summary

E[X] is a number

E[X|Y] is a random variable

E[E[X|Y]] is a number (specifically, it's E[X])

# Time for a bunch of examples

# Example 1: store items

A store has 3 aisles:

- the first aisle has 20 items with an average price of $80
- the second aisle has 50 items with an average price of $70
- the third aisle has 30 items with an average price of $60

We randomly pick an item from the store (100 total items). What is the expected price of the item?

# Example 1: store items

Let's define some random variables:

- X = the price of the item
- Y = the aisle the item is in

We are trying to find E[X].

What we know:

- Pr[Y = 1] = 1/5
- Pr[Y = 2] = 1/2
- Pr[Y = 3] = 3/10
- E[X | Y = 1] = 80
- E[X | Y = 2] = 70
- E[X | Y = 3] = 60

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

$= \sum_y E[X \mid Y = y] \cdot Pr[Y = y]$

$= 80 * 1/5 + 70 * 1/2 + 60 * 3/10$

$= 16 + 35 + 18$

$= 69$

# Example 2: more coins

Your friend now has a bag of coins:

- 3 of the coins flip heads with probability $\frac{1}{2}$
- 5 of the coins flip heads with probability $\frac{1}{5}$
- 2 of the coins flip heads with probability $\frac{3}{4}$

We randomly pick a coin from the bag and flip it 20 times. How many heads should we expect to get?

Berkeley
UNIVERSITY OF CALIFORNIA

# Example 2: more coins

Let's define some random variables:

- X = the number of heads
- Y = which coin we picked

We are trying to find E[X].

What we know:

- $X | Y = 1 \sim \text{Binomial}(20, \frac{1}{2})$
- $X | Y = 2 \sim \text{Binomial}(20, \frac{1}{5})$
- $X | Y = 3 \sim \text{Binomial}(20, \frac{3}{4})$

- $E[X | Y = 1] = 10$
- $E[X | Y = 2] = 4$
- $E[X | Y = 3] = 15$

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

$= \sum_y E[X \mid Y = y] \cdot \Pr[Y = y]$

$= 10 \cdot 3/10 + 4 \cdot 5/10 + 15 \cdot 2/10$

$= 3 + 2 + 3$

$= 8$

# Example 3: even more coins

Now let's say you pick a new coin each time you flip (you put the one you flipped back into the bag).

How many heads should you expect?

# Example 3: even more coins

Let's define some random variables:

- $X$ = the number of heads
- $Y$ = which coin we picked
- $X_i$ = indicator variable for $i^{th}$ flip
    - 1 if heads
    - 0 else
- $X = X_1 + X_2 + \ldots + X_{20}$

We are trying to find E[X].

What we know:

- $E[X] = E[X_1 + X_2 + \ldots + X_{20}]$
- $E[X] = E[X_1] + E[X_2] + \ldots + E[X_{20}]$
- $E[X] = 20E[X_1]$

$$E[X_1] = E[E[X_1 \mid Y]]$$

$$= \sum_y E[X_1 \mid Y = y] \cdot Pr[Y = y]$$

$$= \tfrac{1}{2} \cdot 3/10 + \tfrac{1}{5} \cdot 5/10 + \tfrac{3}{4} \cdot 2/10$$

$$= 3/20 + 1/10 + 3/20$$

$$= 8/20$$

$$E[X] = 20E[X_i] = 8$$

Berkeley UNIVERSITY OF CALIFORNIA

# Example 4: INFINITE coins!!

Now let's say you pick a random number $N \geq 0$. Then, you perform this trial (20 flips) N times. What is the expected number of heads?

Is this just 8E[N]?

Let's define some random variables:

- $X_i$ = the number of heads in the $i^{th}$ trial
- X = the total number of heads
- $X = X_1 + X_2 + \ldots + X_N$

We are trying to find E[X].

Berkeley
UNIVERSITY OF CALIFORNIA

# Example 4: infinite coins

$E[X] = E[E[X \mid N]]$

$= \sum_n E[X \mid N = n] \cdot Pr[N = n]$

$= \sum_n n \cdot E[X_1] \cdot Pr[N = n]$

$= E[X_1] \cdot \sum_n n \cdot Pr[N = n]$

$= E[X_1] \cdot E[N]$

# Wald's identity

Let $X_1$, $X_2$, $X_3$, … be *independent and identically distributed* (iid).

Let N be a random variable with nonnegative integer values that is independent of each $X_i$.

Let $X = X_1 + X_2 + \ldots + X_N$.

**Wald's identity:**

$E[X] = E[X_1]E[N]$

Berkeley
UNIVERSITY OF CALIFORNIA

# More examples

# Example 5: rolling dice

Let's say you roll a fair 6-sided die until the outcome is not a 6.

What is the expected value of the outcome?

# Example 5: rolling dice

Let's define some random variables:

- X = the result of the die roll
- Y = the number of times we roll a 6

We are trying to find E[X].

What we know:

- $\Pr[Y = n] = \frac{1}{6}^n \cdot \frac{5}{6}$
- $\Pr[X = i \mid Y = n] = ?$
  - This is just Pr[X = i], which doesn't help
- $\Pr[X = i, Y = n] = \frac{1}{6}^{n+1}$
  - n sixes in a row, then one i

$$\Pr[X = i] = \sum_y \Pr[X = i, Y = y]$$

$$= \frac{1}{6} + \frac{1}{6}^2 + \frac{1}{6}^3 + \ldots$$

$$= \frac{1}{6} / (1 - \frac{1}{6})$$

$$= \frac{1}{6} / \frac{5}{6}$$

$$= \frac{1}{5}$$

$$E[X] = \sum_x x \cdot \Pr[X = x]$$

$$= \frac{1}{5} \cdot (1 + 2 + 3 + 4 + 5)$$

$$= 3$$

Berkeley
UNIVERSITY OF CALIFORNIA

# Example 5: rolling dice

Let's define some random variables:

- $X$ = the result of the die roll
- $Y$ = the number of flips

We are trying to find $E[X]$.

What we know:

- $\Pr[X = i, Y = n] = \frac{1}{6}^n$
  - roll a six $n - 1$ times, then $i$
- $\Pr[X = i] = \sum_y \Pr[X = i, Y = y]$

(this makes the bounds cleaner)

$$\Pr[X = i] = \sum_y \Pr[X = i, Y = y]$$

$$= \frac{1}{6} + \frac{1}{6}^2 + \frac{1}{6}^3 + \ldots$$

$$= \frac{1}{6} / (1 - \frac{1}{6})$$

$$= \frac{1}{6} / \frac{5}{6}$$

$$= \frac{1}{5}$$

$$E[X] = \sum_x x \cdot \Pr[X = x]$$

$$= \frac{1}{5} \cdot (1 + 2 + 3 + 4 + 5)$$

$$= 3$$

Berkeley UNIVERSITY OF CALIFORNIA

# Example 6: Geometric distribution revisited

Let's say you're like me and you forgot what Tail Sum is. Let's formalize the other way we computed it in the notes.

Let's define some random variables:

- $X \sim \text{Geometric}(p)$
- $Y$ = the result of the first flip
  - 1 if heads, with probability p
  - 0 if tails, with probability 1-p

$$E[X] = E[X \mid Y = 1] \cdot \Pr[Y = 1] + E[X \mid Y = 0] \cdot \Pr[Y = 0]$$

$$= 1 \cdot p + (1 - p)(1 + E[X])$$

$$= p + 1 - p + E[X] - (p \cdot E[X])$$

$$E[X] = 1 + E[X] - (p \cdot E[X])$$

$$p \cdot E[X] = 1$$

$$E[X] = 1/p$$

Berkeley
UNIVERSITY OF CALIFORNIA

# Example 7: we apparently only flip coins

What's the expected number of flips until we get two heads in a row?

Let's define some random variables:

- X = the number of flips until two heads in a row

X ~ Geometric($p^2$)?

This would only be if we flipped coins two at a time.

# Example 7: we apparently only flip coins

Let's define some (more) random variables:

- Y ~ Geometric(p)
    - in other words, the number of flips until one heads
- Z = the number of remaining flips

We are (still) trying to find E[X].

What we know:

- E[Y] = 1/p
- E[X] = E[Y] + E[Z]

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

$= 1/p + E[Z]$

$E[Z] = p + (1 - p)(1 + E[X])$

$E[X] = 1/p + p + 1 - p + E[X] - (p \cdot E[X])$

$p \cdot E[X] = 1/p + 1$

$$E[X] = \frac{p + 1}{p^2}$$

Berkeley
UNIVERSITY OF CALIFORNIA

# Example 7: we apparently only flip coins (Taylor's Version[1])

Let's consider the first two flips. Let Y = the first two flips[2]

- E[X | Y = HH] = 2
- E[X | Y = T*] = 1 + E[X]
    - E[X | Y = TH] = 3p + (1- p)(3 + E[X]) *
    - E[X | Y = TT] = 2 + E[X] *
- E[X | Y = HT] = 2 + E[X]

$E[X] = \sum_y E[X \mid Y = y] \cdot Pr[Y = y]$

$E[X] = p^2 \cdot 2 + (1 - p) \cdot (1 + E[X]) + p \cdot (1-p) \cdot (2 + E[X])$

(after a bunch of algebra, this will yield the same E[X])

[2] This is not formal RV notation (it's a shorthand for mapping numbers to outcomes). Be *very* careful about doing this!
* Using these values, with their associated probabilities ($p(1 - p)$ and $(1-p)^2$, respectively), also yields the same answer



[1] This is a reference to Taylor Swift. In this context, it means "alternate solution"

# Next class: variance

Quiz tomorrow. Best of luck!