Table of Contents

14: Conditional Expectation

Jerry Cain

February 9, 2024

Lecture Discussion on Ed

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:

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

Discrete probabilities of CS109

must be that $\sum_{y}\sum_{t} P(Y=y, T=t) = 1$

Each student responds with:

Year Y

- 1: Freshmen and Sophomores
- 2: Juniors and Seniors
- 3: Graduate Students and SCPD

Mood T:

- -1: 😕
- 0: 😐
- 1: 🥰

Joint PMF

	$Y = 1$	$Y = 2$	$Y = 3$
$T = -1$	.06	.01	.01
$T = 0$	.29	.14	.09
$T = 1$	.30	.08	$\boxed{.02}$

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

this column is concerned with the world where everyone is a frosh or sophomore

this row is concerned with just those feeling meh! (not good / not bad)

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?

Joint PMF

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

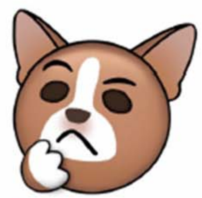
normalize

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

each column sums to 1... what does that mean?

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

these two rows sum to 1 why???



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
$T = 1$	$\boxed{.46}$	.35	.17

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

normalizes world to consist of juniors and seniors

(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

normalizes world so that everyone is happy

Conditional PMFs also sum to 1 conditioned on different events!

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. $P(X = x|Y = 5) = \frac{P(X=x, Y=5)}{P(Y=5)}$

   1-D function of x

3. $P(X = 2|Y = y) = \frac{P(X=2, Y=y)}{P(Y=y)}$

   1-D function of y

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

   2-D function of x, y

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. $\sum_x \left(\sum_y \underbrace{P(X = x|Y = y)P(Y = y)}_{\text{this is } P(X=x, Y=y)} \right) = 1$ true

Conditional Expectation

Conditional expectation

discrete, though continuous is defined similarly

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

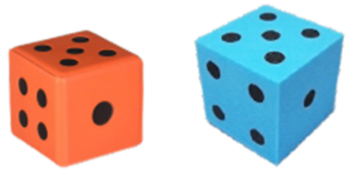
$$E[X|Y = y] = \sum_x x P(X = x|Y = y) = \sum_x x p_{X|Y}(x|y)$$

$Y=y$ is the new sample space

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 $D_1 + D_2$.



$S|D_2=6$ is a random variable!

1. What is $E[\underbrace{S|D_2 = 6}]$?

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

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

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

P.14

Intuitively: ✓ $6 + E[D_1] = 6 + 3.5 = 9.5$ → Let's prove this!

in a few slides

Properties of conditional expectation

1. LOTUS:

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

Y=y is the new sample space

2. Linearity of conditional expectation:

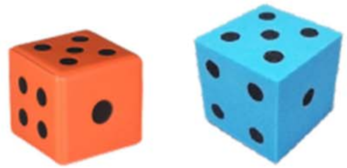
$$E\left[\sum_{i=1}^{n} X_i \mid Y=y\right] = \sum_{i=1}^{n} E[X_i|Y=y]$$

3. Law of total expectation (in, like, three slides)

CLIFFHANGER!!!

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 $E[S|D_2]$.

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

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

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

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

this sum is just 1

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

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

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:

$$E[X] = E\big[\underbrace{E[X|Y]}_{\text{this is a function on } Y}\big]$$

what?!

Proof of Law of Total Expectation

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

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

inner sum is just definition of conditional expectation

$$= \sum_y P(Y=y) \sum_x xP(X=x|Y=y)$$ (def of conditional expectation)

this is a constant with respect to x

$$= \sum_y \left(\sum_x xP(X=x|Y=y)P(Y=y) \right) = \sum_y \left(\sum_x xP(X=x, Y=y) \right)$$ (chain rule)

$$= \sum_x \sum_y xP(X=x, Y=y) \quad = \sum_x x \sum_y P(X=x, Y=y)$$ (switch order of summations)

this inner sum defines the marginal prob mass function of x!!

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

$$= E[X]$$ ...what?

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:

An assumption here is that it's difficult to compute marginal PDF of X.

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 of all conditional expectations of X, i.e., $E[X|Y=y]\ \forall y$ (where weights are $P(Y = y)$)

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

Useful for analyzing recursive code.

Analyzing recursive code

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

```python
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. $E[5+Y] = 5 + E[Y]$

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

Analyzing recursive code

If Y discrete

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

```python
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 $E[Y]$?

$$E[Y] = E[Y|X=1]P(X=1) + \boxed{E[Y|X=2]}P(X=2) + E[Y|X=3]P(X=3)$$

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

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

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

A. $E[5] + Y$

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

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

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

```python
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)?

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

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

Random number of random variables

Suppose you have a website: **`zerothworldproblems.com`**. Let:

- $X = \#$ of people per day who visit your site. $X \sim \text{Bin}(100, 0.5)$
- $Y_i = \#$ of minutes spent per day by visitor i. $Y_i \sim \text{Poi}(8)$
- X and all Y_i are independent.

The time spent by all visitors per day is $W = \sum_{i=1}^{X} Y_i$. What is $E[W]$?

$$E[W] = E\left[\sum_{i=1}^{X} Y_i\right] = E\left[\, E\left[\sum_{i=1}^{X} Y_i \,|\, X\right]\,\right]$$

Suppose $X = x$.

$$E\left[\sum_{i}^{x} Y_i \,|\, X = x\right] = \sum_{i=1}^{x} E[Y_i | X = x] \qquad \text{(linearity)}$$

$$= \sum_{i=1}^{x} E[Y_i] \qquad \text{(independence)}$$

$$= xE[Y_i]$$

$$= E\big[XE[Y_i]\big]$$

$$= E[Y_i]E[X] \qquad \text{(scalar } E[Y_i]\text{)}$$

$$= 8 \cdot 50$$