

Table of Contents

- 2 Discrete Conditional Distributions
- 9 Conditional Expectation
- 15 Law of Total Expectation

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

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

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

Different notation,
same idea:

Discrete probabilities of CS109

Each student responds with:

Year Y

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

Mood T :

- -1 : 😞
- 0 : 😐
- 1 : 😍

this middle row
is concerned with just
three who are feeling
meh 😐

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

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

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

this column is
concerned with the subset
of the world where
everyone is a fresh
or sophomore.

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?

	$Y = 1$	$Y = 2$	$Y = 3$
$T = -1$.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 up 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?

(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

restablishes
the world to
focus on just juniors and seniors
 $.30 / (.06 + .29 + .30)$

		<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

(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 distribution
to only include
those who are happy.

Conditional PMFs also sum to 1 conditioned on different events!

Quick check

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

1-D function in x

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

1-D function on y

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

2-D function on
 x and 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. $\sum_x \sum_y P(X = x|Y = y) = 1$ in general, false

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

Conditional 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

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

$Y = y$
is the
new
sample
space.

It's been so long, our dice friends

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

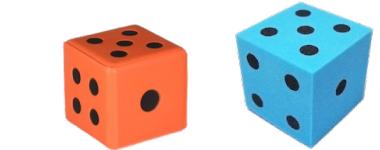
- Roll two 6-sided dice.
- Let roll 1 be D_1 , roll 2 be D_2 .
- Let $S = \text{value of } D_1 + D_2$.

1. What is $E[S|D_2 = 6]$? $E[S|D_2 = 6] = \sum_{x=7}^{12} x P(S = x|D_2 = 6)$

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

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

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



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. It's like the other value that Y can take in just don't exist any more.

2. Linearity of conditional expectation:

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

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

cliff hanger!

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 $S = \text{value of } D_1 + D_2$.

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 $S = \text{value of } D_1 + D_2$.

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 $E[S|D_2]$.

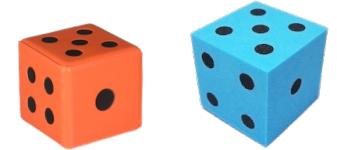
$$\begin{aligned} E[S|D_2 = d_2] &= E[D_1 + d_2|D_2 = d_2] \\ &= \sum_{d_1} (d_1 + d_2) P(D_1 = d_1 | D_2 = d_2) \\ &= \sum_{d_1} d_1 P(D_1 = d_1) + d_2 \sum_{d_1} P(D_1 = d_1) \end{aligned}$$

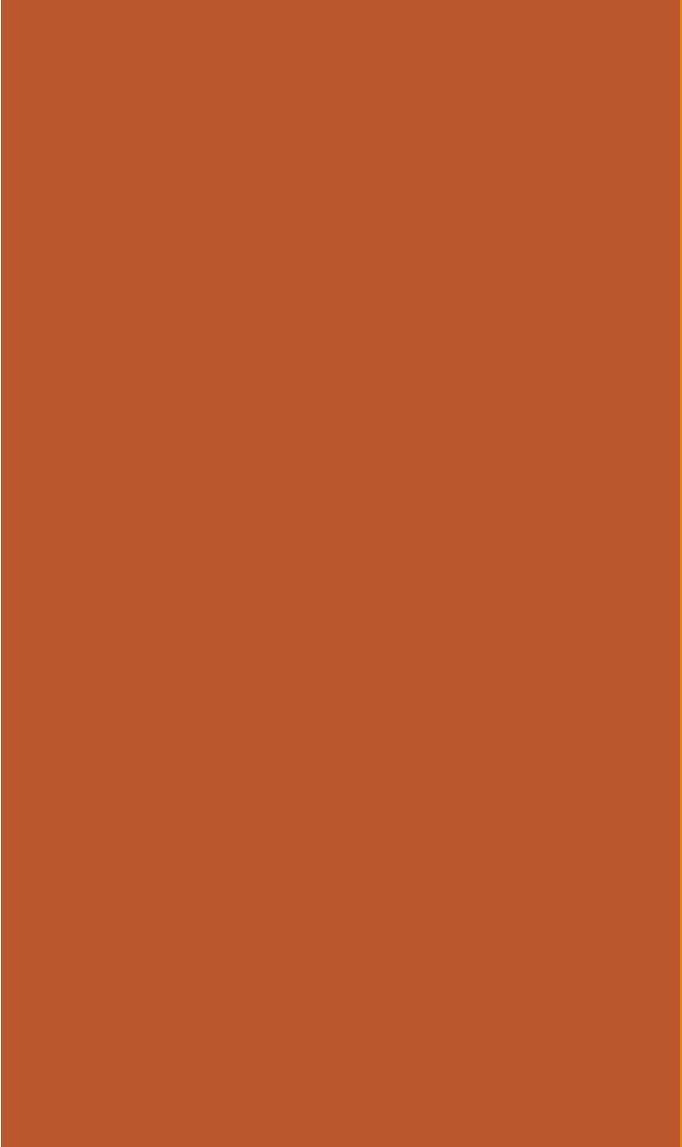
$(D_1 = d_1, D_2 = d_2$
independent
events)

d_1 definition of $E[D_2]$

d_1 this sums to 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:

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

3. Law of total expectation:

$$E[X] = E\left[E[X|Y]\right] \quad \text{what?}$$

this inner expectation
is a random variable in \mathcal{Y}

Proof of Law of Total Expectation

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

$$\begin{aligned}
 E[E[X|Y]] &= E[g(Y)] = \sum_y P(Y = y) E[X|Y = y] && (\text{LOTUS, } g(Y) = E[X|Y]) \\
 &= \sum_y P(Y = y) \sum_x x P(X = x | Y = y) && \text{(def of conditional expectation)} \\
 &= \sum_y \left(\sum_x x P(X = x | Y = y) P(Y = y) \right) = \sum_y \left(\sum_x x P(X = x, Y = y) \right) && (\text{chain rule}) \\
 &= \sum_x \sum_y x P(X = x, Y = y) && (\text{switch order of summations}) \\
 &= \sum_x x P(X = x, Y = y) && \text{this inner sum defines} \\
 &= \sum_x x P(X = x) && \text{the marginal probability mass function of } X \text{ (marginalization)} \\
 &= E[X]
 \end{aligned}$$

Handwritten annotations and clarifications:

- The first step shows $E[X|Y]$ as $E[g(Y)]$, where $g(Y) = E[X|Y]$.
- The second step shows the expansion of $E[g(Y)]$ using the Law of Total Probability.
- The third step shows the application of the definition of conditional expectation, where $P(Y=y)$ is highlighted in red and the inner sum is highlighted in green.
- The fourth step shows the application of the chain rule of probability.
- The fifth step shows the switching of the order of summations.
- The sixth step highlights the marginal probability mass function of X .
- The final step shows the result as $E[X]$.

Another way to compute $E[X]$

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

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

when might
this be the case?
when X behaves differently
depending on the value
of Y .

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.

Analyzing recursive code

```
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()
```

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

Let $Y = \text{return value of } \text{reurse}()$.
What is $E[Y]$?

Analyzing recursive code

```
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()
```

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

Let $Y = \text{return value of } \text{reurse}()$.
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

```
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()
```

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



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

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

```
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()
```

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

If Y discrete

Let $Y = \text{return value of } \text{reurse}()$.
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 $\text{reurse}()$.

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

```
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()
```

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

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

↑
↑

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

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

If Y discrete

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

Let $Y = \text{return value of } \text{recurse}()$.
What is $E[Y]$?

Analyzing recursive code

```
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()
```

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

If Y discrete

Let $Y = \text{return value of } \text{reurse}()$.
What is $E[Y]$?

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

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

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

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

$$E[Y] = 15$$

On your own: What is $\text{Var}(Y)$?

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

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

Random number of random variables

$$\begin{aligned} & \text{indep } X, Y \\ & E[X|Y = y] = E[X] \end{aligned}$$

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 $E[W]$?



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 $E[W]$?

$$\begin{aligned} E[W] &= E\left[\sum_{i=1}^X Y_i\right] = E\left[E\left[\sum_{i=1}^X Y_i | X\right]\right] \\ &\quad \text{this is really } E[E[W|x]] \\ &= E[XE[Y_i]] \\ &= E[Y_i]E[X] \quad (\text{scalar } E[Y_i]) \\ &= 8 \cdot 50 \end{aligned}$$

X *upper bound is a random variable !!*

Suppose $X = x$.

$$\begin{aligned} E\left[\sum_i^x Y_i | X = x\right] &= \sum_{i=1}^x E[Y_i | X = x] && (\text{linearity}) \\ &= \sum_{i=1}^x E[Y_i] && (\text{independence}) \\ &= xE[Y_i] \end{aligned}$$

Breaking News

Before You Go! CS109 Challenge



Do something cool and creative
with probability!

Grand Prize:

All exams replaced with 100%

All Serious Entries:

Extra credit (between %0.5 and
2% added to overall average)

Optional Proposal: Sun. 02/25, 11:59pm

Due: Sat. 03/09, 11:59pm

<https://cs109.stanford.edu/handouts/challenge.html>