14: Conditional Expectation

Jerry Cain
May 1st, 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: 😍

<table>
<thead>
<tr>
<th></th>
<th>$Y = 1$</th>
<th>$Y = 2$</th>
<th>$Y = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = -1$</td>
<td>.06</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>$T = 0$</td>
<td>.29</td>
<td>.14</td>
<td>.09</td>
</tr>
<tr>
<td>$T = 1$</td>
<td>.30</td>
<td><strong>.08</strong></td>
<td>.02</td>
</tr>
</tbody>
</table>

$P(Y = 2, T = 1)$

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?

<table>
<thead>
<tr>
<th></th>
<th>$Y = 1$</th>
<th>$Y = 2$</th>
<th>$Y = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = -1$</td>
<td>.09</td>
<td>.04</td>
<td>.08</td>
</tr>
<tr>
<td>$T = 0$</td>
<td>.45</td>
<td>.61</td>
<td>.75</td>
</tr>
<tr>
<td>$T = 1$</td>
<td>.46</td>
<td>.35</td>
<td>.17</td>
</tr>
</tbody>
</table>

Joint PMF

<table>
<thead>
<tr>
<th></th>
<th>$Y = 1$</th>
<th>$Y = 2$</th>
<th>$Y = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = -1$</td>
<td>.06</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>$T = 0$</td>
<td>.29</td>
<td>.14</td>
<td>.09</td>
</tr>
<tr>
<td>$T = 1$</td>
<td>.30</td>
<td>.08</td>
<td>.02</td>
</tr>
</tbody>
</table>

$T = -1$ | .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?

### Conditional PMFs

<table>
<thead>
<tr>
<th>$Y = 1$</th>
<th>$Y = 2$</th>
<th>$Y = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = -1$</td>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>$T = 0$</td>
<td>0.29</td>
<td>0.14</td>
</tr>
<tr>
<td>$T = 1$</td>
<td>0.30</td>
<td>0.08</td>
</tr>
</tbody>
</table>

| $T = -1$ | 0.75    | 0.125   | 0.125   |
| $T = 0$  | 0.56    | 0.27    | 0.17    |
| $T = 1$  | 0.75    | 0.2     | 0.05    |

Conditional PMFs also sum to 1 conditioned on different events!
### Quick check

<table>
<thead>
<tr>
<th>Number or function?</th>
<th>True or false?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. $P(X = 2</td>
<td>Y = 5)$</td>
</tr>
<tr>
<td>2. $P(X = x</td>
<td>Y = 5)$</td>
</tr>
<tr>
<td>3. $P(X = 2</td>
<td>Y = y)$</td>
</tr>
<tr>
<td>4. $P(X = x</td>
<td>Y = y)$</td>
</tr>
</tbody>
</table>

* $P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$
## Quick check

### 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
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

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

• Note that $E[X]$ is a well-defined statistic even when $X$ is one of many random variables in a multivariate distribution: $E[X] = \sum_x \sum_y xp_{X,Y}(x,y)$
• $E[X|Y = y]$ is the average value of $X$ when $Y$ is constrained to take on a specific value of $y$: $E[X|Y = y] = \sum_x xp_{X,Y}(x|y)$
It’s been so long, our dice friends

- 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 $E[S|D_2 = 6]$? 

\[
E[S|D_2 = 6] = \sum_x xP(S = x|D_2 = 6)
\]

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

\[
= \frac{57}{6} = 9.5
\]

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

We’ll prove in a moment
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 (in, like, three slides)
It’s been so long, our dice friends

- 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 $E[S|D_2 = 6]$?

2. What is $E[S|D_2]$?
   - A. A function of $S$
   - B. A function of $D_2$
   - C. A number


$$E[X|Y = y] = \sum_x x p_{x|y}(x|y)$$
It’s been so long, our dice friends

- 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 $E[S|D_2]$?
   \[
   E[S|D_2 = d_2] = E[D_1 + d_2|D_2 = d_2]
   = \sum_{d_1} (d_1 + d_2)P(D_1 = d_1|D_2 = d_2)
   = \sum_{d_1} d_1 P(D_1 = d_1) + d_2 \sum_{d_1} P(D_1 = d_1)
   = E[D_1] + d_2 = 3.5 + d_2
   \]
   $E[S|D_2] = 3.5 + D_2$


A. A function of $S$
B. A function of $D_2$
C. A number

Lisa Yan, Chris Piech, Mehran Sahami, and Jerry Cain, CS109, Spring 2024
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:

\[ E[X] = E[E[X|Y]] \quad \text{what?} \]
Proof of Law of Total Expectation

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

\[ = \sum_y P(Y = y) \sum_x xP(X = x|Y = y) \]

\[ = \sum_y \left( \sum_x xP(X = x|Y = y)P(Y = y) \right) = \sum_y \left( \sum_x xP(X = x, Y = y) \right) \]

\[ = \sum_x \sum_y xP(X = x, Y = y) = \sum_x \sum_y P(X = x, Y = y) \]

\[ = \sum_x xP(X = x) \]

\[ = E[X] \]

(LOTUS, \( g(Y) = E[X|Y] \))

(def of conditional expectation)

(chain rule)

(switch order of summations)

(marginalization)
Another way to compute $E[X]$

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

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 $P(Y = y)$)

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

Useful for analyzing recursive code.
Analyzing recursive code

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

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

\[ E[X] = E[E[X|Y]] = \sum_y E[X|Y = y]P(Y = y) \]
Analyzing recursive code

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

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

\[
\]

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

When $X = 1$, return 3.
Analyzing recursive code

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

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


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

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

B. $E[5 + Y] = 5 + E[Y]$
C. $5 + E[Y|X = 2]$
Analyzing recursive code

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

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

If $Y$ discrete

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


When $X = 2$, return $5 + \text{a future return value of} \ recurse()$. What is $E[Y|X = 2]$?

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
    if x == 2: return 5 + recurse()
    return 7 + recurse()
```

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


When $X = 3$, return $7 + \text{a future return value of } \text{recurse}()$.

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

If $Y$ discrete

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

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

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

\[
\]

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

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

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

\[
E[Y] = 15
\]

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

Suppose you have a website: dorisisthebeast.com. Let:

- $X = \#$ of people per day who visit your site. $X \sim \text{Poi}(50)$
- $Y_i = \#$ of minutes spent per day by visitor $i$. $Y_i \sim \text{Poi}(11)$
- $X$ 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: dorisisthebeast.com. Let:

- \( X = \# \) of people per day who visit your site. \( X \sim \text{Poi}(50) \)
- \( Y_i = \# \) of minutes spent per day by visitor \( i \). \( Y_i \sim \text{Poi}(11) \)
- \( X \) and all \( Y_i \) are independent.

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

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

\[
= E \left[ X E[Y_i] \right]
\]

\[
= E[Y_i] E[X]
\]

(\( E[Y_i] \) is a scalar)

\[
= 11 \cdot 50 = 550
\]

Suppose \( X = x \).

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

(linearity)

\[
= \sum_{i=1}^{x} E[Y_i]
\]

(independence)

\[
= x E[Y_i]
\]

= 550