14: Conditional Expectation

Jerry Cain
April 27, 2022
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

Each student responds with:

Year $Y$:
- 1: Frosh/Soph
- 2: Jr/Sr
- 3: Coterm/grad/SCPD

Timezone $T$ (12pm California time in my timezone is):
- $-1$: AM
- 0: noon
- 1: PM

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

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

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?

### 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>

### Conditional PMFs

<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>.75</td>
<td>.125</td>
<td>$?$</td>
</tr>
<tr>
<td>$T = 0$</td>
<td>.56</td>
<td>.27</td>
<td>.17</td>
</tr>
<tr>
<td>$T = 1$</td>
<td>.75</td>
<td>.2</td>
<td>.05</td>
</tr>
</tbody>
</table>
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>.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>

**Conditional PMFs also sum to 1 conditioned on different events!**

### (B) \( P(T = t|Y = y) \)

<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>

### (A) \( P(Y = y|T = t) \)

<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>.75</td>
<td>.125</td>
<td>.125</td>
</tr>
<tr>
<td>( T = 0 )</td>
<td>.56</td>
<td>.27</td>
<td>.17</td>
</tr>
<tr>
<td>( T = 1 )</td>
<td>.75</td>
<td>.2</td>
<td>.05</td>
</tr>
</tbody>
</table>

\[ .30/(.06+.29+.30) \]
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

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

\[ P(X = x | Y = y) = \frac{P(X = x,Y = y)}{P(Y = y)} \]
Conditional Expectation
Conditional expectation

Recall 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

$$E[X|Y = y] = \sum_x xP(X = x|Y = y) = \sum_x xp_{X|Y}(x|y)$$
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 $E[S|D_2 = 6]$?  

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

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

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

Intuitively:  

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

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. 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 (next time)
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 $E[S|D_2]$?
   
   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 $S =$ 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 function of $S$
   - A function of $D_2$
   - A number

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

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

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

what?!
Proof of Law of Total Expectation

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

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

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

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

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

\[ = E[X] \quad \text{...what?} \]
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
    else: 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
    elif (x == 2): return (5 + recurse())
    else: return (7 + recurse())
```

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

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

```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 $E[Y]$?


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

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

Let $Y =$ return value of `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

```
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 $E[Y]$?


When $X = 2$, return $5 + a$ 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

```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 $E[Y]$?

If $Y$ discrete

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


$E[Y|X = 1] = 3$ 
$E[Y|X = 2] = E[5 + Y]$ 
When $X = 3$, return $7 + a$ future return value of `recurse()`.

$E[Y|X = 3] = E[7 + Y]$
Analyzing recursive code

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


$$E[Y] = \frac{3}{3} + (5 + E[Y])(\frac{1}{3}) + (7 + E[Y])(\frac{1}{3})$$

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

$$E[Y] = 15$$

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

$$E[X|Y = y] = \sum_x x p_{X|Y}(x | y) = \sum_x x p_X(x) = E[X]$$
Random number of random variables

Say you have a website: BestJokesEver.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 $E[W]$?

[indep $X, Y$
\[
E[X|Y = y] = E[X]
\]
Random number of random variables

Say you have a website: BestJokesEver.com. Let:

- $X = \#$ of people per day who visit your site. $X \sim \text{Bin}(100, 0.5)$
- $Y_i = \#$ of minutes spent 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 $E[W]$?

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

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

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

\[
= 8 \cdot 50
\]
See you on Friday!

Have a super Wednesday!
Extra
Hiring software engineers

Your company has only one job opening for a software engineer.
• $n$ candidates interview, in order ($n!$ orderings equally likely)
• Must decide hire/no hire *immediately* after each interview

Strategy:
1. Interview $k$ (of $n$) candidates and reject all $k$
2. Accept the next candidate better than all of first $k$ candidates.

What is your target $k$ that maximizes $P($get best candidate$)$?

Fun fact:
• There is an $\alpha$-to-1 factor difference in productivity b/t the “best” and “average” software engineer.
• Steve jobs said $\alpha=25$, Mark Zuckerberg claims $\alpha=100$, some even claim $\alpha=300$
Hiring software engineers

Your company has only one job opening for a software engineer.

- *n* candidates interview, in order (*n!* orderings equally likely)
- Must decide hire/no hire *immediately* after each interview

**Strategy:**

1. Interview *k* (of *n*) candidates and reject all *k*
2. Accept the next candidate better than all of first *k* candidates.

What is your target *k* that maximizes P(get best candidate)?

**Define:**

- *X* = position of best engineer candidate (1, 2, ..., *n*)
- *B* = event that you hire the best engineer

Want to maximize for *k*:

\[ P_k(B) = \text{probability of } B \text{ when using strategy for a given } k \]

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

(law of total probability)
Hiring software engineers

Your company has only one job opening for a software engineer.

Strategy:
1. Interview \( k \) (of \( n \)) candidates and reject all \( k \)
2. Accept the next candidate better than all of first \( k \) candidates.

What is your target \( k \) that maximizes \( P(\text{get best candidate}) \)?

Define:
- \( X \) = position of best engineer candidate
- \( B \) = event that you hire the best engineer

If \( i \leq k \):
\[
P_k(B|X = i) = 0 \quad \text{(we fired best candidate already)}
\]

Else:
- We must not hire prior to the \( i \)-th candidate.
- We must have fired the best of the \( i-1 \) first candidates.
- The best of the \( i-1 \) needs to be our comparison point for positions \( k+1, \ldots, i-1 \).
- The best of the \( i-1 \) needs to be one of our first \( k \) comparison/auto-fire

\[
P_k(B) = \frac{1}{n} \sum_{i=1}^{n} P_k(B|X = i) = \frac{1}{n} \sum_{i=k+1}^{n} \frac{k}{i-1} \quad \Leftarrow \text{Want to maximize over } k
\]
Hiring software engineers

Your company has only one job opening for a software engineer.

Strategy: 1. Interview $k$ (of $n$) candidates and reject all $k$
2. Accept the next candidate better than all of first $k$ candidates.

What is your target $k$ that maximizes $P$(get best candidate)?

Want to maximize over $k$:

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

Maximize by differentiating w.r.t $k$, set to 0, solve for $k$:

$$\frac{d}{dk} \left( \frac{k}{n} \ln \frac{n}{k} \right) = \frac{1}{n} \ln \frac{n}{k} + \frac{k}{n} \cdot \frac{-n}{k^2} = 0$$

$$\ln \frac{n}{k} = 1$$

1. Interview $\frac{n}{e}$ candidates
2. Pick best based on strategy
3. $P_k(B) \approx 1/e \approx 0.368$