14: Conditional Expectation

Jerry Cain
April 27, 2022

Table of Contents

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

$$\sum_T \sum_Y p_{t,y} = 1$$

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

Joint PMF

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

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

Joint PMFs sum to 1.

Discrete probabilities of CS109

The below are **conditional probability tables** for conditional PMFs
(A) $\underbrace{P(Y = y|T = t)}$ and (B) $\overbrace{P(T = t|Y = y)}$.

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.

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

$.46 = \frac{.30}{.30 + .29 + .06}$

A

	$Y = 1$	$Y = 2$	$Y = 3$	
$T = -1$	.75	.125	? .125	$\to$ 1.0
$T = 0$	.56	.27	.17	$\to$ 1.0
$T = 1$	.75	.2	.05	$\to$ 1.0

$\frac{.01}{.01 + .01 + .06}$

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
$T = 1$	.46	.35	.17

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

(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

Conditional PMFs also sum to 1 conditioned on different events!

~~Quick~~ Long check

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

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. $P(X = x|Y = 5)$

   1-D function on X

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

   1-D function on y

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

   2-D function on x, y

   true

True or false?

5. $\sum_{x} P(X = x|Y = 5) = 1$ true

   true

6. $\sum_{y} P(X = 2|Y = y) = 1$ false

   false

7. $\sum_{x}\sum_{y} P(X = x|Y = y) = 1$ false

   $P_{X,Y}(x,y)$ ignore

8. $\sum_{x}\left(\sum_{y} P(X = x|Y = y)P(Y = y)\right) = 1$ true

Conditional Expectation

Conditional expectation discrete

Recall the the conditional PMF of X given $Y = y$:

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

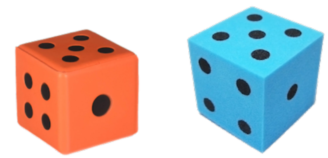
The conditional expectation of X given $Y = y$ is

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

$Y=y$ 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 $D_1 + D_2$.

$S|D_2=6$ random variable

1. What is $E[S|D_2 = 6]$?

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

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

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

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

Let's prove this!

Properties of conditional expectation

1. LOTUS:

is the
Y=y new sample
space

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

2. Linearity of conditional expectation:

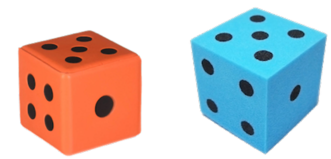
$$E\left[\sum_{i=1}^{n} X_i \mid Y = y\right] = \sum_{i=1}^{n} E[X_i|Y = y]$$

3. Law of total expectation (next time)

cliffhanger

It’s been so long, our dice friends

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

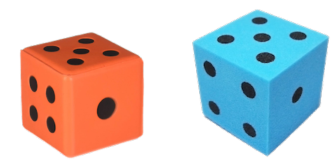


- Roll two 6-sided dice.
- Let roll 1 be D_1, roll 2 be D_2.
- Let S = value of $D_1 + D_2$.

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 $E[S|D_2]$.

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 $E[S|D_2]$.

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

$E[D_1]$

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

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

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

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

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:

$$\boxed{E[X] = E\left[\overbrace{E[X|Y]}\right]}$$

what?!

Proof of Law of Total Expectation

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

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

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

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

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

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

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

```
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 $E[Y]$?

Analyzing recursive code

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

```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 $E[Y]$?

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

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

When $X = 1$, return 3.

Analyzing recursive code

If Y discrete

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

```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 $E[Y]$?

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

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

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

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

(by yourself)

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 $E[Y]$?

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

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

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

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

```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 $E[Y]$?

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

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

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

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

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

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 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

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 $E[W]$?

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

Suppose $X = x$.

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

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

$$= xE[Y_i]$$

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

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

$$= 8 \cdot 50$$

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 α-to-1 factor difference in productivity b/t the "best" and "average" software engineer.
- Steve jobs said α=25, Mark Zuckerberg claims α=100, some even claim α=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: $P_k(B)$ = probability of B 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(get best candidate)?

Define: X = position of best engineer candidate
B = event that you hire the best engineer

If $i \le k$: $P_k(B|X=i) = 0$ (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, …, *i*−1.
→ The best of the *i*−1 needs to be one of our first *k* comparison/auto-fire

$$P_k(B|X=i) = \frac{k}{i-1}$$

$$P_k(B) = \frac{1}{n}\sum_{i=1}^{n} P_k(B|X=i) = \frac{1}{n}\sum_{i=k+1}^{n} \frac{k}{i-1}$$

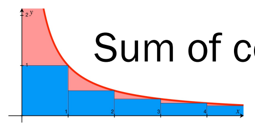
← Want to maximize over k

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)?



Sum of converging series

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:

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

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

$$k = \frac{n}{e}$$

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