Conditional Probability

Lecture 4

Please, sign in on iClicker

Outline

1. (Univariate) Conditional Probability
2. (Multivariate) Conditional Probability Distributions
3. Conditional Independence
4. Information about Quiz 1

Notation Review

- $A, B \subseteq S$ are events.
- $P(\cdot)$ gives the probability for an event.
- Upper case letters X, Y are random variables.
- Lower case letters x, y are possible values X, Y can take.
- $\{X = 1\}$ is an event, and $P(X = 1)$ gives the probability for that event.
- $P(X = x)$ for all possible values x is the PMF of X.
  - We often define a PMF as $p_X(x) = P(X = x)$ to represent the PMF compactly.

More Notation

- $\mathbb{E}(X)$ is the expected value for X.
- $P(X = x \cap Y = y)$ for all possible values x, y represents the joint PMF of X and Y
  - We often define a PMF as $p_{X,Y}(x, y) = P(X = x \cap Y = y)$
- For shorthand convention, you may use a comma to indicate the AND relationship for joint probability, that is, $P(X = x, Y = y)$. However, note that a comma is not technically a set operation.

Common Notation Issues

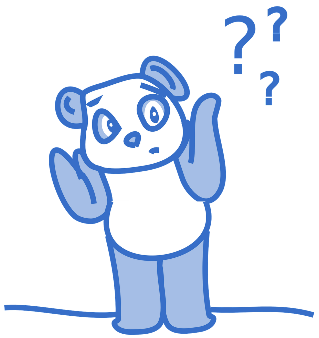
- Define X as a probability distribution
- $P(X)$ and $P(XY)$
- $E(X = 3)$

- If you want to use any notation different from what we have discussed during the lecture or labs, please define it clearly to avoid confusion. For example, you can write that you use a comma to indicate AND in joint probabilities.

1. Univariate Conditional Probability

- We have introduced joint probabilities, for example, probability of first 10 coin flips are all heads.
- We might want to update the probabilities after seeing these results.

> How can we update the probabilities if we have more information?



Conditional Probability for Events

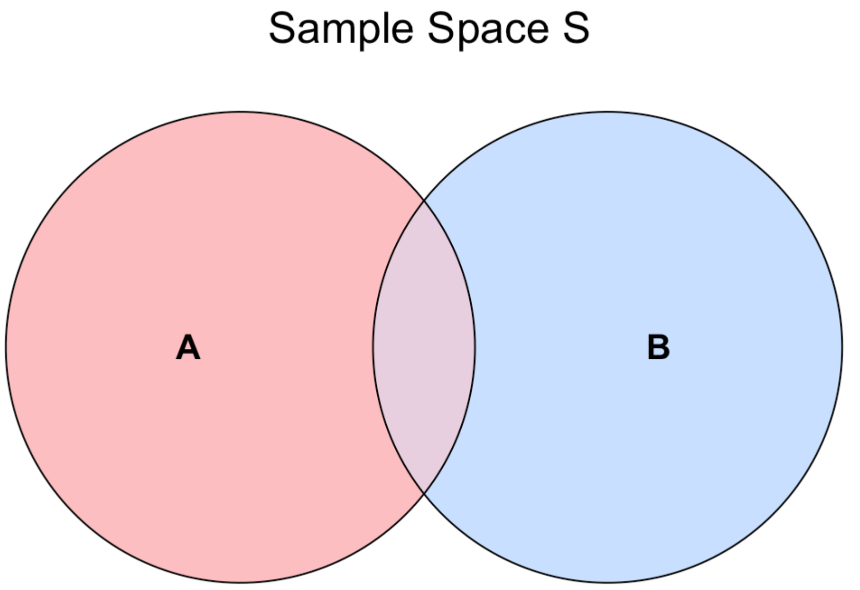
- Let A and B be two events of interest, and $P(B) > 0$, then the conditional probability of A given B is defined as:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

- Event B is becoming the new sample space.

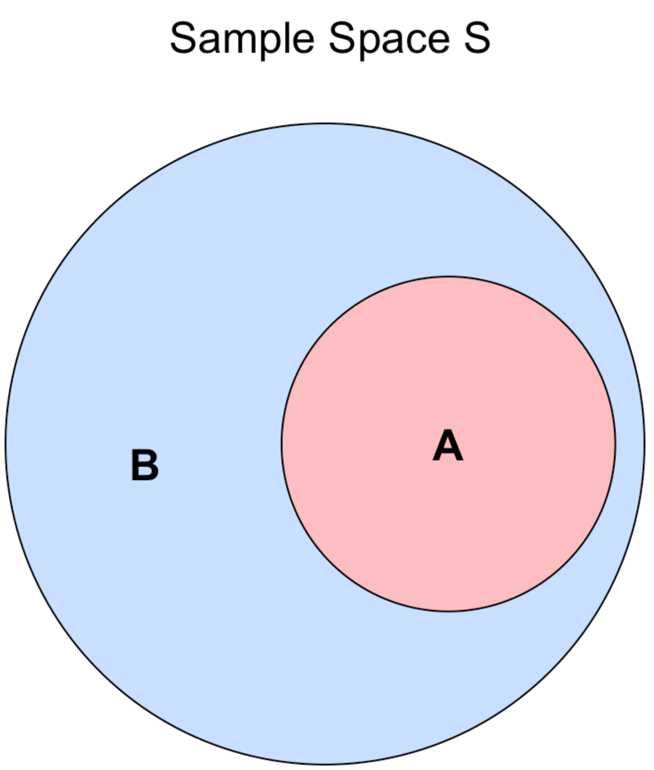
Graphically...

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$



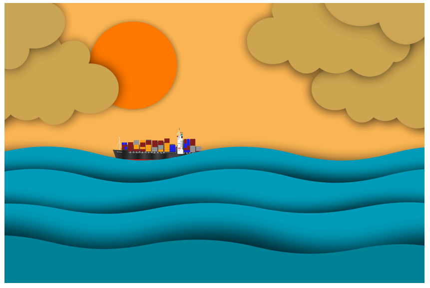
Another depiction…

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)}$$



Example: Length of Stay Versus Gang Demand

- Consider the example of ships arriving at the port of Vancouver again.
- Each ship will stay at port for a random number of days, which we will call the length of stay (L).

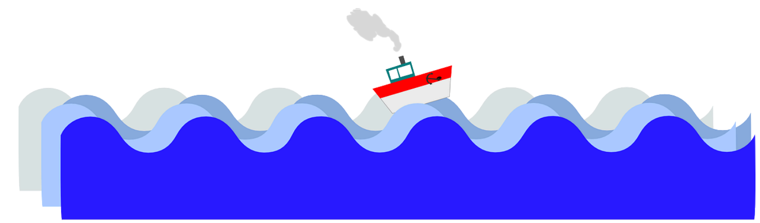


Probability Mass Function (PMF) of Length of Stay

- Let L denote the length of stay, which has the following distribution:

L (Days)	Probability
1	0.25
2	0.35
3	0.20
4	0.10
5	0.10

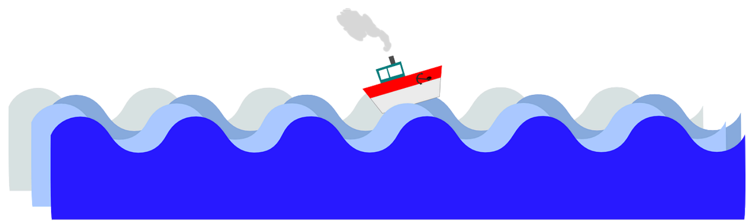
Univariate Conditional



- Suppose a ship has been at port for 2 days now, and it will be staying longer. This means that we know that L will be greater than 2.
- What is the distribution of L now?

$$P(L = l \mid L > 2) \text{ for } l = 3, 4, 5$$

Univariate Conditional Notation



UBC Master of Data Science

- $P(L = 3 \mid L > 2)$ is a conditional probability.
- $P(L = l \mid L > 2)$ gives the conditional probability of all possible outcome $L = l$. This is called a conditional probability distribution.
  - The distribution is under a new restricted sample space.
- The total probability within this restricted space should still sum to 1, i.e.,

$$\sum_{l=3}^{5} P(L = l \mid L > 2) = 1.$$

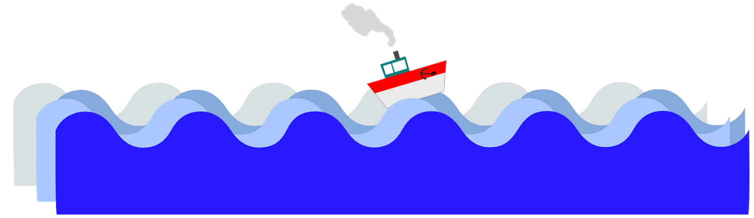
Approach 1: Formula Approach

- Let us use the conditional probability formula:

$$P(L = l \mid L > 2) = \frac{P(L = l \cap L > 2)}{P(L > 2)} = \frac{P(L = l)}{P(L > 2)}$$
$$\text{for} \quad l = 3, 4, 5.$$

- We can reduce the event $\{L = l \cap L > 2\}$ to a simple event $\{L = l\}$ for $l = 3, 4, 5$ (Why?)

Applying the Formula



- Let us recheck the formula:

$$P(L = l \mid L > 2) = \frac{P(L = l)}{P(L > 2)}$$
$$\text{for} \quad l = 3, 4, 5.$$

- The math is telling us to do the re-normalizing.
- For all cases satisfying the condition, we would divide by

$$\begin{aligned} P(L > 2) &= P(L = 3) + P(L = 4) + P(L = 5) \\ &= 0.20 + 0.10 + 0.10 = 0.40. \end{aligned}$$

UBC Master of Data Science

Therefore, we have

$$P(L=3 \mid L>2) = \frac{P(L=3)}{P(L>2)} = \frac{0.20}{0.40} = 0.50$$

$$P(L=4 \mid L>2) = \frac{P(L=4)}{P(L>2)} = \frac{0.10}{0.40} = 0.25$$

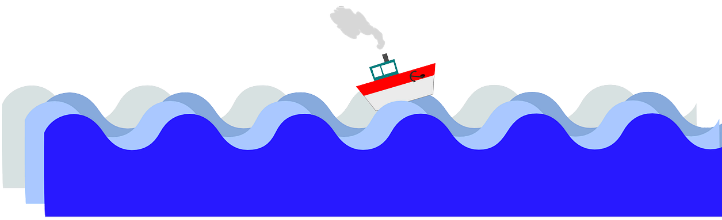
$$P(L=5 \mid L>2) = \frac{P(L=5)}{P(L>2)} = \frac{0.10}{0.40} = 0.25.$$

Approach 2: Table Approach

We will follow these steps:

1. Subset the PMF table to those outcomes that satisfy the condition $L > 2$ (we will have a "sub-table").
2. Re-normalize the remaining probabilities so that they add up to 1. We will end up with the conditional distribution.

Recall the original PMF in our ship example!



L (Days)	Probability
1	0.25
2	0.35
3	0.20
4	0.10
5	0.10

First Step of Re-normalization

- Now that we know $L > 2$, we have to "delete" some of these options:

L (Days)	Probability
1	IMPOSSIBLE
2	IMPOSSIBLE
3	Used to be 0.20
4	Used to be 0.10
5	Used to be 0.10

Second Step of Re-normalization

- We normalize their corresponding probabilities such that they all add up to 1 again. . . .
- We divide all the probabilities by $0.20 + 0.10 + 0.10 = 0.40$:

L (Days)	Probability
1	0
2	0
3	0.50
4	0.25
5	0.25

2. (Multivariate) Conditional Probability Distributions

- So far, we have considered conditioning in the one-variable (i.e., univariate) case.
- However, it is more useful to think about the distribution of one random variable when conditioned on a different random variable.

Definition: Conditional PMF

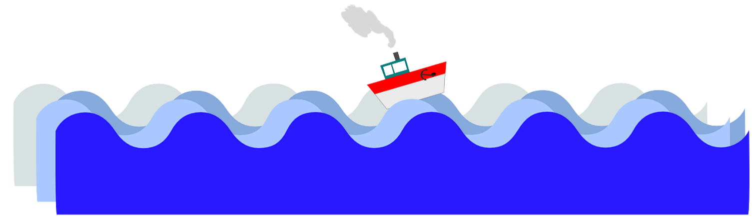
- Let X and Y be discrete random variables, then the conditional PMF of X given Y is defined as:

$$P(X = x \mid Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)}.$$

Conditional PMF

- A conditional PMF of X given $Y = y$ is a proper probability distribution.
- That is, $\sum_x P(X = x \mid Y = y) = 1$

Cargo ships come again!



- Let us revisit our 2-variable example where we looked at both the LOS (i.e., L) and the number of Gangs required (i.e., G):

	G = 1	G = 2	G = 3	G = 4
L = 1	0.0017	0.0425	0.1247	0.0811
L = 2	0.0266	0.1698	0.1360	0.0176
L = 3	0.0511	0.1156	0.0320	0.0013
L = 4	0.0465	0.0474	0.0059	0.0001
L = 5	0.0740	0.0246	0.0014	0.0000

Conditioning G on L

- Suppose a ship is arriving, and they have told you they will only be staying for 1 day.
- What is the distribution of G under this information for all possible g?

$$P(G = g \mid L = 1) \text{ for } g = 1, 2, 3, 4$$

Approach 1: Formula Approach

- By applying the formula for conditional probabilities, we get

$$P(G = g \mid L = 1) = \frac{P(G = g \cap L = 1)}{P(L = 1)} \quad \text{for} \quad g = 1, 2, 3$$

	G = 1	G = 2	G = 3	G = 4
L = 1	0.00170	0.04253	0.12471	0.08106
L = 2	0.02664	0.16981	0.13598	0.01757
L = 3	0.05109	0.11563	0.03203	0.00125
L = 4	0.04653	0.04744	0.00593	0.00010
L = 5	0.07404	0.02459	0.00135	0.00002

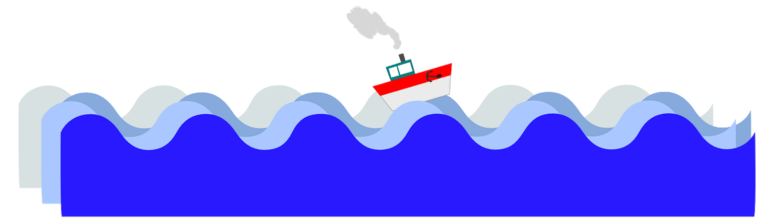
Obtaining Marginal $P(L = 1)$

	G = 1	**G = 2**	**G = 3**	**G = 4**	**Marginals**
L = 1	0.00170	0.04253	0.12471	0.08106	0.25

Then, each element of the conditional PMF…

$$P(G = 1 \mid L = 1) = \frac{0.0017}{0.25} = 0.0068$$
$$P(G = 2 \mid L = 1) = \frac{0.0425}{0.25} = 0.1701$$
$$P(G = 3 \mid L = 1) = \frac{0.1247}{0.25} = 0.4988$$
$$P(G = 4 \mid L = 1) = \frac{0.0811}{0.25} = 0.3242$$

Approach 2: Table Approach

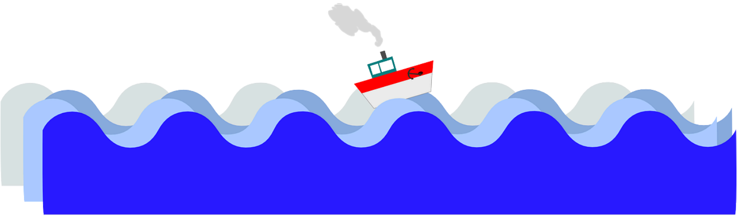


We will follow these steps:

1. We isolate the outcomes satisfying the condition ($L = 1$):

	G = 1	**G = 2**	**G = 3**	**G = 4**
L = 1	Used to be 0.0017	Used to be 0.0425	Used to be 0.1247	Used to be 0.0811
L = 2	IMPOSSIBLE	IMPOSSIBLE	IMPOSSIBLE	IMPOSSIBLE
L = 3	IMPOSSIBLE	IMPOSSIBLE	IMPOSSIBLE	IMPOSSIBLE
L = 4	IMPOSSIBLE	IMPOSSIBLE	IMPOSSIBLE	IMPOSSIBLE
L = 5	IMPOSSIBLE	IMPOSSIBLE	IMPOSSIBLE	IMPOSSIBLE

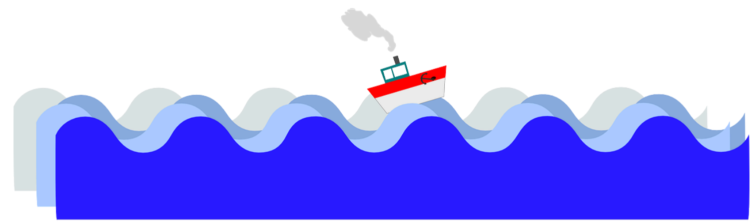
Then...



2. We re-normalize the probabilities so that they add up to 1, by dividing them by their sum, which is 0.25:

	G = 1	G = 2	G = 3	G = 4
L = 1	0.0068	0.1701	0.4988	0.3242

Sanity Check



- The previous four conditional probabilities are part of a proper conditional PMF:

$$\begin{aligned}\sum_{g=1}^{4} P(G = g \mid L = 1) &= 0.0068 + 0.1701 + 0.4988 + 0.3242 \\ &= 1.\end{aligned}$$

- Both approaches, table and formula, are equivalent!

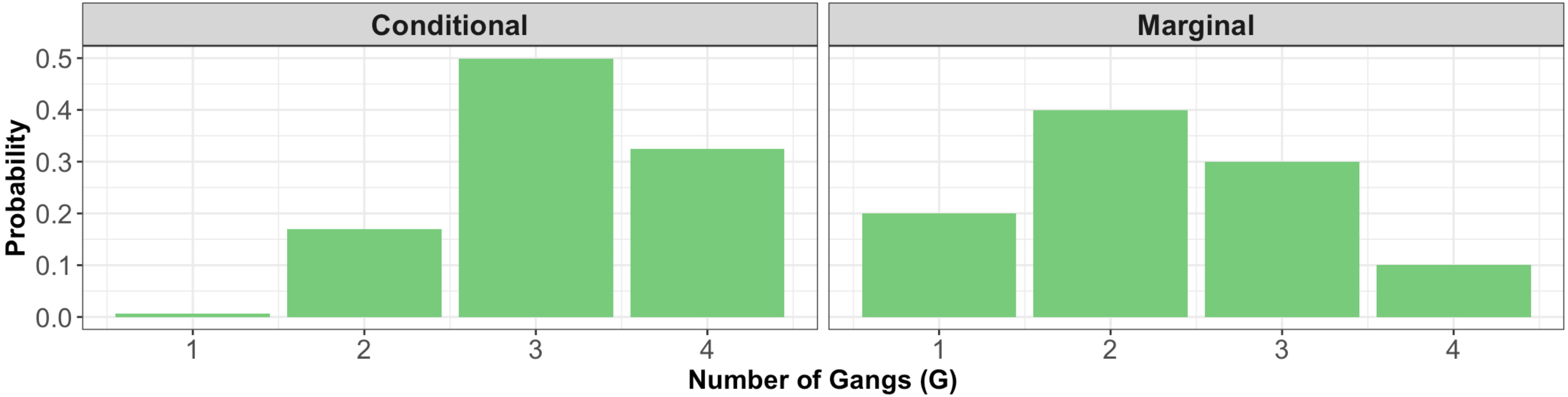
Comparing versus the marginal distribution of G

- Marginal distribution:

G = 1	G = 2	G = 3	G = 4
0.2	0.4	0.3	0.1

- Conditional distribution

	G = 1	G = 2	G = 3	G = 4
L = 1	0.0068	0.1701	0.4988	0.3242


Conditional
Marginal
Probability
0.5
0.4
0.3
0.2
0.1
0.0
1
2
3
4
Number of Gangs (G)

How about the expected values of G?

- Marginal Mean:
  $\mathbb{E}(G) = 1(0.2) + 2(0.4) + 3(0.3) + 4(0.1) = 2.3.$

G = 1	G = 2	G = 3	G = 4
0.2	0.4	0.3	0.1

- Conditional mean on $L = 1$:

$$\begin{aligned}\mathbb{E}(G \mid L = 1) &= 1(0.0068) + 2(0.1701) + 3(0.4988) + 4(0.3242) \\ &= 3.1406.\end{aligned}$$

	G = 1	G = 2	G = 3	G = 4
L = 1	0.0068	0.1701	0.4988	0.3242

Definition: Conditional Expectation

- For discrete random variables:

$$\mathbb{E}(X|Y=y) = \sum_x x P(X=x|Y=y)$$

- Machine learning & regression models are all about estimate conditional means of the target variable given the input features.

What We Have Learned So Far

- Joint probabilities as a table or a data frame
- Marginal probability: row sum or column sum
- Conditional probability: re-normalize probabilities

2.3. Conditional Probabilities for Independent Variables

- Random variables X and Y are independent if and only if

$$P(Y = y \cap X = x) = P(Y = y) \cdot P(X = x), \text{ for all } x \text{ and y.}$$

- With conditional probabilities introduced:

$$\begin{aligned} P(Y = y \mid X = x) &= \frac{P(Y = y \cap X = x)}{P(X = x)} \\ &= \frac{P(Y = y) \cdot P(X = x)}{P(X = x)} = P(Y = y). \end{aligned}$$

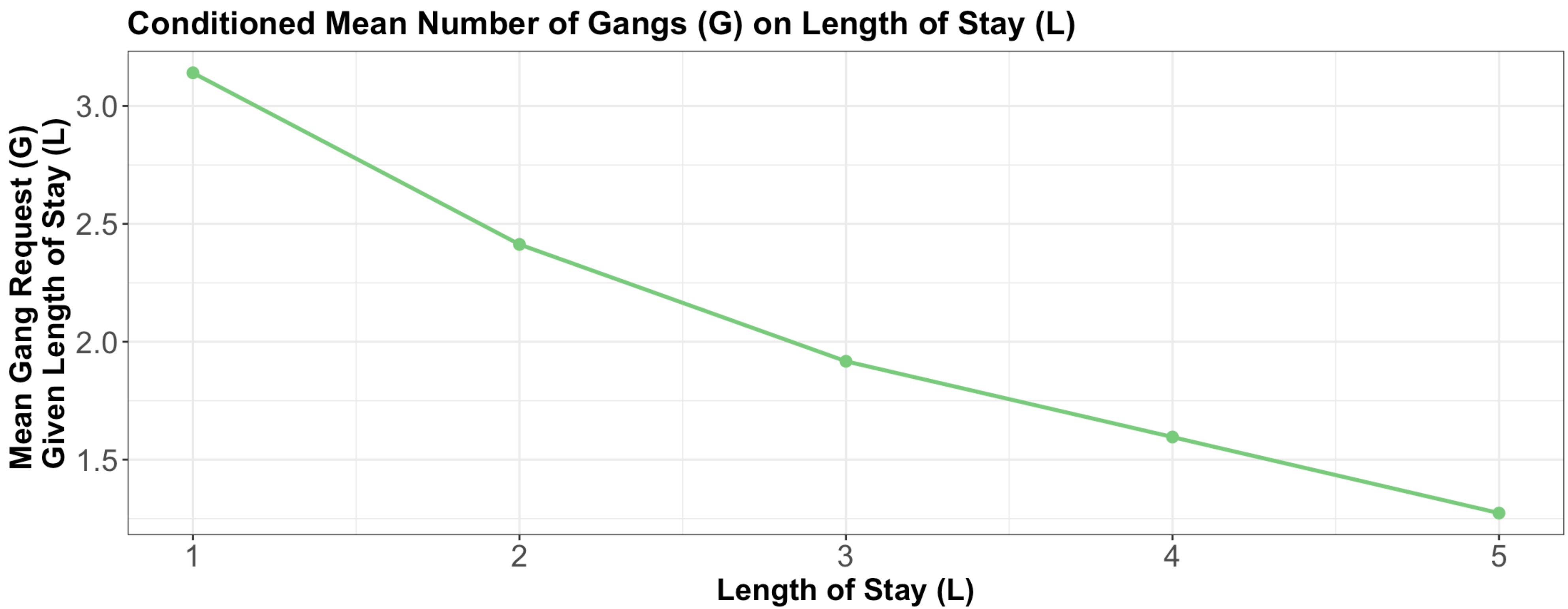
2.4. Law of Total Expectation

- A marginal mean can be computed from the conditional means and the marginal probabilities of the conditioning variable.
- This is a direct application of the Law of Total Expectation

$$\mathbb{E}_Y(Y) = \sum_x \mathbb{E}_Y(Y \mid X = x) \cdot P(X = x)$$
$$\mathbb{E}_Y(Y) = \mathbb{E}_X[\mathbb{E}_Y(Y \mid X)].$$

Proceeding with the Cargo Ships!

- Suppose we have the following conditional means of gang request G given the length of stay L of a ship as a model function:

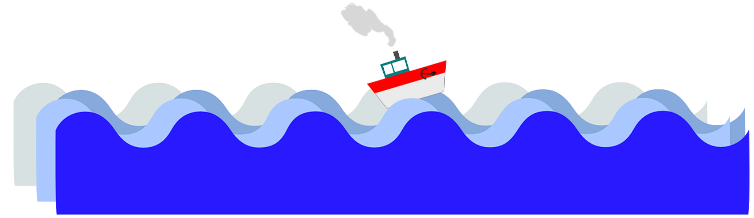


How can we compute $\mathbb{E}_G(G)$?

- The below table is based on the previous plot plus the marginal PMF of L:

l (Days)	E(G | L = l)	P(L = l)
1	3.1405	0.25
2	2.4128	0.35
3	1.9172	0.20
4	1.5960	0.10
5	1.2735	0.10

How can we compute $\mathbb{E}_G(G)$?



- Multiplying the last two columns together, and summing, gives us the marginal expectation:

$$
\begin{aligned}
\mathbb{E}_G(G) &= \sum_l \mathbb{E}_G(G \mid L = l) \cdot P(L = l) \\
&= \mathbf{2.3}.
\end{aligned}
$$

iClicker Question

In general for two random variables X and Y, $P(X = x \mid Y = y)$ is a normalized probability distribution in the sense that

A. $\sum_x P(X = x \mid Y = y) = 1.$

B. $\sum_y P(X = x \mid Y = y) = 1.$

Answer

- It is **A**.
- $Y = y$ is fixed.
- Following up with the Law of Total Probability, the sum applies to all the values of X.

iClicker Question

Answer **TRUE** or **FALSE**:

Let X be a random variable with non-zero entropy and Y be a random variable with zero entropy. Then, X and Y are independent.

A. TRUE

B. FALSE

Answer

- It is **TRUE**.
- Y is not random. Hence, knowing about X does not tell you anything about Y and also the other way around.

3. Conditional Independence

> Can the dependence/independence of random variables X and Y change if we condition on another random variable Z?

> If random variables X and Y are not independent, can they be independent given Z?

> If random variables X and Y are independent, can they be dependent given Z?

Yes!

- A box contains two coins. A regular coin and a two-headed coin. Choose a coin at random and toss it twice. Define

$$X = \text{First coin toss results}$$

$$Y = \text{Second coin toss results}$$

$$C = \text{Which coin has been selected}$$

- X and Y are NOT independent, but they are conditionally independent given $C =$ the regular coin

Definition of Conditional Independence

- Independence and conditional independence are different.
- Recall the independence definition:

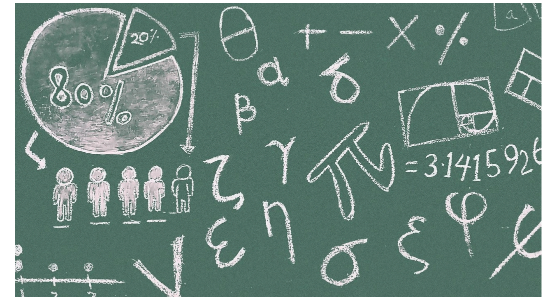
$$P(X = x \cap Y = y) = P(X = x) \cdot P(Y = y).$$

- For conditional independence, X and Y are conditionally independent given Z if and only if

$$P(X = x \cap Y = y \mid Z = z) = P(X = x \mid Z = z) \cdot P(Y = y \mid Z = z)$$

for all x, y, z.

Let us consider an example!



- Let L be a student's lab grade in DSCI 551 (low or high)
- Let Q be a student's quiz grade in DSCI 551 (low or high)
- Let S represents whether the student majored in Statistics in their undergraduate studies (yes or no)

Joint PMF

ℓ	q	s	$P(L=\ell \cap Q=q \cap S=s)$
low	low	yes	0.01
low	high	yes	0.03
high	low	yes	0.03
high	high	yes	0.09
low	low	no	0.21
low	high	no	0.21
high	low	no	0.21
high	high	no	0.21

Are L and Q independent?

- This question has nothing to do with S, so let us marginalize out S:

ℓ	q	$P(L = \ell \cap Q = q)$
low	low	0.22
low	high	0.24
high	low	0.24
high	high	0.30

So, are L and Q independent?

L/Q	low	high	Marginals for L
low	0.22	0.24	0.46
high	0.24	0.3	0.54
Marginals for Q	0.46	0.54	

- Apparently not
- Because

$$P(L = \text{low} \cap Q = \text{low}) = 0.22 \neq 0.46 \times 0.46.$$

Are L and Q conditionally independent given S?

- We need to check whether

  $$P(L = \ell \cap Q = q \mid S = s) = P(L = \ell \mid S = s) \cdot P(Q = q \mid S = s)$$

  for all l, q, s.
- We will check this for $S = \text{yes}$ and $S = \text{no}$.

For $S = \text{yes}$

- We re-normalize the table by $P(S = \text{yes}) = 0.16$:

ℓ	q	$P(L = \ell \cap Q = q \mid S = \text{yes})$
low	low	0.01 / 0.16 = 0.0625
low	high	0.03 / 0.16 = 0.1875
high	low	0.03 / 0.16 = 0.1875
high	high	0.09 / 0.16 = 0.5625

Checking Conditional Independence

- We can check that this distribution satisfies the definition of independence; e.g.:

$$
\begin{aligned}
&P(L = \text{low} \mid S = \text{yes}) \cdot P(Q = \text{low} \mid S = \text{yes}) \\
&= \left(\frac{0.04}{0.16}\right)\left(\frac{0.04}{0.16}\right) \\
&= 0.0625 \\
&= P(L = \text{low} \cap Q = \text{low} \mid S = \text{yes})
\end{aligned}
$$

- The condition also holds for other level of L and Q given $S = \text{yes}$.

Now for $S = \mathrm{no}$

- Again, we re-normalize the table by $P(S = \mathrm{no}) = 0.84$:

ℓ	q	$P(L = \ell \cap Q = q \mid S = \mathrm{no})$
low	low	0.21 / 0.84 = 0.25
low	high	0.21 / 0.84 = 0.25
high	low	0.21 / 0.84 = 0.25
high	high	0.21 / 0.84 = 0.25

Checking Conditional Independence Again

- We check the condition of conditional independence:

$$
\begin{aligned}
& P(L = \text{low} \mid S = \text{no}) \cdot P(Q = \text{low} \mid S = \text{no}) \\
& = \left(\frac{0.42}{0.84}\right)\left(\frac{0.42}{0.84}\right) \\
& = 0.25 \\
& = P(L = \text{low} \cap Q = \text{low} \mid S = \text{no})
\end{aligned}
$$

- The condition also holds for other level of L and Q given $S = \text{no}$.

Conclusion

- We can conclude the lab grade and quiz grade are not independent, but they are conditionally independent given information about whether the student was a Statistics major.
- If you already know that a person is (or is not) a Statistics major, then their lab and quiz grades are completely independent.

Finally…

- It is also possible to have the opposite case: two variables that are independent, but not conditionally independent given a third variable.

Today's Learning Goals

By the end of this lecture, we will be able to…

- Calculate **conditional distributions** from a full distribution.
- Obtain the marginal mean from conditional means and marginal probabilities, using the Law of Total Expectation.
- Compare and contrast independence versus **conditional independence**.

About Quiz 1

- General quiz policy for MDS: https://ubc-mds.github.io/resources_pages/quiz/
- For 551 Quiz 1
  - 10-12 questions: multiple choice, or multiple answer, or calculation
  - Review learning objectives for lecture 1 to 4
  - Try practice quiz on PrairieLearn
  - You will be able to open a RStudio workspace during the quiz