OI 3.2: Conditional Probability (Part 2)

Review

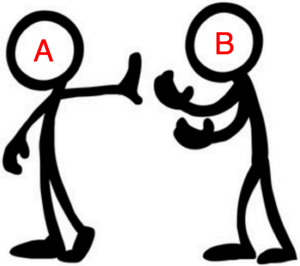
- What are some things we have already learned about for random processes?

Worksheet

- In your worksheet we calculated the probability of an event for a random process, *given* some information about a different random process.
- Did probabilities change when you learned information about the pet that was adopted?

Conditional Probability

- $P(A|B)$: The probability of event A happening *given* event B
- In other words: suppose we know B happens, what is the probability of A?



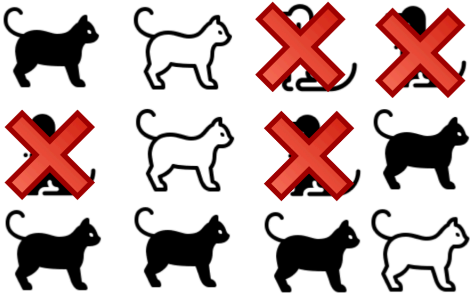
Conditional Probability

When we have outcomes that are all equally likely (count data):

$$P(A|B) = \frac{\text{\# of occurrences of variable } A \text{ and variable } B}{\text{\# of occurrences of variable } B}$$

Conditional Probability

Suppose you learn that the the animal adopted is a cat. Given this information, what is the probability that it has a black coat?



Conditional Probability

Suppose you learn that the the animal adopted is a cat. Given this information, what is the probability that it has a black coat?

Species\Coat	Black	White	Total
Dog	45	12	57
Cat	101	55	156
Total	146	67	213

Conditional Probability

When we have outcomes that *may not* by equally likely (probabilities/proportions):

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

The above formula is the **conditional probability rule**. This rule *always* works. The previous version of the rule only works with count data.

Conditional Probability

Suppose you learn that the the animal adopted is a cat. Given this information, what is the probability that it has a black coat?

Species\Coat	Black	White	Total
Dog	.45	.06	.51
Cat	.38	.11	.49
Total	.83	.17	1

Conditional Probability

Lets try Questions 1d and 2d on the handout again with the **conditional probability rule**.

Do the answers match?

Conditional Probability

- Conditional probabilities are essentially *rescaling* a probability.
- If the two variables A and B are dependent, then knowing information about B changes the probability of A.
- The notation $A|B$ reflects this newfound knowledge.
- For example:
  - What is the probability the pet adopted has a black coat?
  - What is the probability the pet adopted has black coat, given it is a cat?
- Knowing the pet was a cat changed the information!
- All conditional probability problems can be thought of in this way.

Multiplication Rule

- We can multiply and divide probabilities across the equal sign.
- For example, a simple rearrangement can give us the **multiplication rule**. If A and B represent two outcomes or events, then

$$P(A \text{ and } B) = P(A|B)P(B)$$

- The multiplication rule is the same thing as the conditional probability rule, this is just a rearrangement!

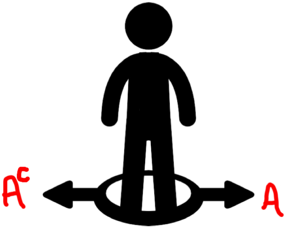
Multiplication Rule and Independence

- We have seen a little bit of the multiplication rule before when we talked about independent events.
- Recall, we said two events A and B are independent if $P(A \text{ and } B) = P(A)P(B)$. This is equivalent to saying that two events are independent if $P(A|B) = P(A)$.

Complement Rule with Conditional Probabilities

Recall our rule for complements:

$$1 = P(A^C) + P(A)$$

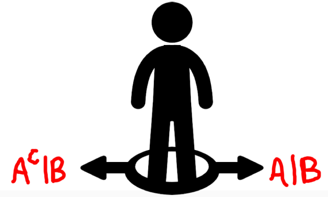


Since A^C is anything other than A, if we have $P(A)$ we can find $P(A^C)$.

Complement Rule with Conditional Probabilities

The same concept holds with conditional probabilities!

$$1 = P(A^C|B) + P(A|B)$$



There is only two options once we are given B:

$$A^C|B \qquad \text{and} \qquad A|B$$

Therefore, if we know $P(A|B)$ we can find $P(A^C|B)$ just as before!

Example

[OI 3.13] Joint and conditional probabilities.

Let $P(A) = 0.3$ and $P(B) = 0.7$.

a) Can you compute $P(A \text{ and } B)$ if you only know $P(A)$ and $P(B)$?

b) Assuming that events A and B arise from independent random processes,

   i. what is $P(A \text{ and } B)$?
   ii. what is $P(A \text{ or } B)$?
   iii. what is $P(A|B)$?

c) If we are given that $P(A \text{ and } B) = 0.1$, are the random variables giving rise to events A and B independent?

d) If we are given that $P(A \text{ and } B) = 0.1$, what is $P(A|B)$?

e) If $P(A|B) = .14$, what is $P(A^C|B)$?

Total Probability Rule

Let $A_1, ..., A_k$ represent all disjoint outcomes for a variable or process. Then if B is an event, possibly for another variable or process, we have

$$
\begin{aligned}
P(B) =& P(B \text{ and } A_1) + P(B \text{ and } A_2) + \cdots + P(B \text{ and } A_k) \\
=& P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)
\end{aligned}
$$

We get the second line from substituting the multiplication rule.

Total Probability Rule - Contingency Tables

We saw the total probability rule in with contingency tables!

MPAA Rating	**Box Office Gross** Low	High	Total
Not Rated	.001	.006	.007
G	.008	.014	.022
PG	.055	.101	.156
PG-13	.101	.267	.368
R	.140	.302	.442
NC-17	.003	.002	.005
Total	.308	.692	1

This specific example:

$$P(\text{PG-13}) = P(\text{PG-13 and Low}) + P(\text{PG-13 and High})$$

Generally:

$$P(B) = P(B \text{ and } A_1) + \cdots + P(B \text{ and } A_k)$$

Example (Part 1 - Together)

Let D indicate being infected with Covid19, and A represent a positive test result. Suppose scientists have concluded that rate of Covid19 infections for adults in the US is $P(D) = 0.01$. The probability of correctly detecting the Covid19 when it is present in the system is $P(A|D) = 0.95$, and the probability Covid19 is correctly not detected is $P(A^C|D^C) = 0.931$. Suppose we pick someone at random from the US. We want to find the probability that someone has Covid19 given that they test positive $P(D|A)$.

Testing statistics reported by John Hopkins in 2020.

Example (Part 1 - Together)

Did the answer surprise you?

Do you think Covid19 tests you take are accurate? Why or why not?

Example (Part 1 - Together)

- We found the probability that the test correctly detects if a person has Covid19 when testing someone in the **general population**.
- When actually preforming medical tests, we do not always test the general population, instead we test **those who are suspected to have the disease**.

Example (Part 2 - In groups)

Suppose that the Covid19 infection rate among those who suspect they have Covid19 (or have been exposed to it recently) is $P(D) = .40$. If $P(A|D) = 0.95$, and $P(A^C|D^C) = 0.931$. What is $P(D|A)$?

A Few Notes

- The medical field often reports *sensitivity* and *specificity* for a test.
- **Sensitivity**: a test's ability to designate an individual with disease as positive

$$P(A|D)$$

- **Specificity**: a test is its ability to designate an individual who does not have a disease as negative

$$P(A^C|D^C)$$

- Most people want to know if the test is accurate in general, or if the test is accurate for those at risk. Why report $P(A|D)$ or $P(A^C|D^C)$ instead?

Definitions and Rules

- Mutually Exclusive:

$$P(A \text{ and } B) = 0$$

- Independence:

$$P(A \text{ and } B) = P(A)P(B)$$

- Complement:

$$P(A) = 1 - P(A^C)$$

Definitions and Rules

- General Additivity:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

- Conditional:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

- General Multiplication:

$$P(A \text{ and } B) = P(A|B)P(B)$$

- Total Probability:

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)$$