Probability Worksheet Solutions

Worksheet 1

Introduction

These initial problems will help get us oriented in a context that is more familiar. In each, we will be referring to a standard "die" (singular of dice) that has 6 faces, each with an equal chance of being rolled

Part A What is the chance of getting a 1 when rolling a die once?

```
1 / 6
```

```
## [1] 0.1666667
```

Part B When rolling a die once, what is the chance of rolling a 1 *or* a 2?

```
2 / 6
```

```
## [1] 0.3333333
```

Part C When rolling a die once, what is the chance of rolling a 1 *and* a 2?

0. Can't happen. The events are disjoint.

Part D What is the chance of rolling a 1,2,3,4,5, or 6?

1. P(1,2,3,4,5, or 6) = P(1) + P(2) + P(3) + P(4) + P(5) + P(6) = 6*(1/6)

Part E What is the chance of *not* rolling a 2?

P(not 2) = 1 - P(2) = 1 - 1/6 = 5/6

Problem 1 (Addition Rule)

Question 1 Here, we concern ourselves with 10,000 individuals who either (1) rent their home (3,858), (2) have a mortgage on their home (4,789), or (3) own it outright (1,353).

- What proportion of individuals have either a mortgage or own it outright?
- If we select one person out of this 10,000 at random, what is the probability that this person either owns their own or has a mortgage?

```r
# prop. that own their home or have a mortgage
(4789 + 1353) / 10000
```

```
## [1] 0.6142
```

Probability is the same as this answer.

Question 2 Consider rolling a dice where we define three different events:

$$A = \{1, 2\}, \quad B = \{4, 6\}, \quad D = \{2, 3\}$$

- What is the probability of event A?
- Are events B and D disjoint? Confirm the addition rule by finding the probability that either B or D occurs.

P(A) = 2/6

B and D are disjoint since they share no outcomes.

B ∩ D = ∅

P(B ∩ D) = 0

P(B or D) = P(B) + P(D) - P(B ∩ D) = P(B) + P(D)

Problem 2 (General Addition Rule)

Question 1 If events A and B are disjoint, explain why this implies that $P(A \text{ and } B) = 0$. Verify that the General Addition Rule simplifies to the Addition Rule when A and B are disjoint.

They share no outcomes, so the probability of the intersection is zero (can't happen).

P(A or B) = P(A) + P(B) - P(A and B), but P(A and B) = 0, so P(A or B) = P(A) + P(B)

Question 2 In a sample of 10,000 homes, 1495 homes were painted blue, 4789 had a garage, and 950 homes had both of these properties. Create a Venn diagram illustrating this problem.

Problem 3 Using the information from Question 2, what is the probability that a home selected at random had a garage but was not painted blue?

P(G and 'not B') = P(G) - P(G and B)

```r
(4789 - 950) / 10000
```

```
## [1] 0.3839
```

Problem 3 (Compliments)

Question 1 For a single dice roll, let $D = \{2, 3\}$. What is D^C? Find $P(D)$ and $P(D^C)$

$D^C = \{1, 4, 5, 6\}$

P(D) = 2 / 6 = 1 / 3

P(D^C) = 4 / 6 = 2 / 3

P(D) + P(not D) = 1 / 3 + 2 / 3 = 1

	Heart Attack		
Treatment	Attack	No Attack	Total
Placebo	189	10,845	11,034
Aspirin	104	10,933	11,037
Total	293	21,778	22,071

Problem 4 (Empirical Probabilities)

This data is from a report published in 1988 that summarizes the results of a Harvard Medical School clinical trial determining effectiveness of aspirin in preventing heart attacks in middle-aged male physicians}.

For all questions, restate the probability in terms of probability notation you've seen.

Question 1 What is the probability that a randomly selected physician had a heart attack?

```
293 / 22071
```

```
## [1] 0.01327534
```

Question 2 What is the complement of the event in question 1? What is the probability of the complement?

complement is a randomly selected physician *not* having a heart attack

```
1 - (293 / 22071)
```

```
## [1] 0.9867247
```

Question 3 What is the probability that a randomly selected physician was taking aspirin?

```
11037 / 22071
```

```
## [1] 0.500068
```

Question 4 What is the probability that a randomly selected physician was taking aspirin *and* had a heart attack? Think about where this number comes from on the table and why it's called an intersection.

```
104 / 22071
```

```
## [1] 0.004712066
```

Question 5 Using your previous answers, what is the probability that a randomly selected physician was taking aspirin *or* had a heart attack?

```
(11037 + 293 - 104)/22071
```

```
## [1] 0.5086312
```

Worksheet 2

Rules

For each of these problem, please use notation that we have adopted in class, i.e., events A or B, probabilities $P(A)$, expressions $P(A|B)$ or $P(A \text{ or } B)$, etc., in addition to solving them numerically.

Independence

Question 1 It is estimated that 9% of people are left handed. We randomly sample five people from our population. Based on these five, answer the following:

- What is the probability that all five people are right-handed?
- What is the probability that all five people are left-handed?
- What is the probability that that not all five people are right-handed (i.e., probability that *at least one* is left-handed)?

P(all 5 right handed) = $P(righthanded)^5$ (they are independent)

```
(1 - .09)^5
```

```
## [1] 0.6240321
```

P(all 5 left handed)

```
.09 ^5
```

```
## [1] 5.9049e-06
```

P(not all 5 are right handed) = 1 - P(all 5 right handed)

```
1 - .6240321
```

```
## [1] 0.3759679
```

Question 2 Further assume that sex and handedness are independent, i.e.,

$$P(\text{right-handed and male}) = P(\text{right-handed}) \times P(\text{Male}).$$

In a population with equal proportions of men and women, answer the following:

- What is the probability that the first person is male and right-handed?
- What is the probability the the first two people are male and right-handed
- What is the probability that the third person is female and left-handed?
- What is the probability that the first two people are male and right-handed and the third person is female and left-handed?

P(first person M and R)

```
.91 * .5
```

```
## [1] 0.455
```

P(first 2 M and R)

```
.455^2
```

```
## [1] 0.207025
```

P(third person F and L) (independent of the other 2 people)

```
0.5 * .09
```

```
## [1] 0.045
```

P(first 2 are M and R and third person is F and L)

```
.455^2 * .045
```

```
## [1] 0.009316125
```

Question 3 In a standard 52 card deck, assume we draw one card at random. Is the event that a card is a Heart independent of the card being an Ace? Make your argument using the **multiplication rule**

H = {2H, 3H, 4H, 5H, 6H, 7H, 8H, 9H, 10H, JH, QH, KH, AH}

A = {AH, AD, AS, AC}

P(H | A) = 1/4 = P(H) = 13/52 = 1/4

P(H and A) = 1/52 = P(H)*P(A) = 1/4 x 1/13 = 1/52

They are independent.

Conditional Probability

Below is a table of 6,224 individuals from the year 1721 who were exposed to smallpox in Boston. Doctors at the time believed that exposing a person to the disease in a controlled form (inoculating them) could reduce the likelihood of death.

	inoculated		
	yes	no	Total
lived	238	5136	5374
died	6	844	850
Total	244	5980	6224

Here, the same table is reproduced, providing joint and marginal probabilities

	inoculated		
	yes	no	Total
lived	0.0382	0.8252	0.8634
died	0.0010	0.1356	0.1366
Total	0.0392	0.9608	1.0000

Question 1 Write in formal notation the probability that a randomly selected person who was not inoculated died from smallpox and compute the probability

P(died from smallpox | not vaccinated) = 0.1356

Question 2 Determine the probability that an inoculated person died from smallpox. How does this compare with what you found in Question 1?

P(died from smallpox | vaccinated) = 0.001

Question 3 Based on your results from Q1 and Q2, does it appear as if inoculation is effective at reducing the risk of smallpox?

Yes. The death rate is MUCH lower for the vaccination group.

Question 4 Let X and Y represent the outcomes of rolling two dice. Answer the following:

- What is the probability that the first die, X, is equal to 1?
- What is the probability that both X and Y are equal to 1?
- Use the formula for conditional probability to compute $P(Y = 1|X = 1)$
- What is $P(Y = 1)$? Is this different that what we found in the last part? Explain.

P(X = 1) = 1/6

P(X = Y = 1) = P(X = 1 and Y = 1) = P(X=1) x P(Y=1) = 1/6 x 1/6 = 1/36

P(Y=1|X=1) = P(X=1 and Y=1)/P(X=1) = 1/6

P(Y=1) = P(Y=1|X=1) = 1/6 (indepedent, X=1 doesn't change likelihood of Y=1)

Question 5 Suppose that 80% of people like peanut butter (obviously), 89% like jelly, and 78% like both. Given that a randomly sampled person likes peanut butter, what is the probability that they also like jelly?

P(likes jelly | likes PB) = P(likes both) / P(likes PB)

```
.78 / .8
```

```
## [1] 0.975
```

Question 6 Bob is watching a roulette table in a casino and notices that the last five outcomes landed on black. He figures that the probability of landing on black six times in a row is very small (1/64), so he places a bet on red. What is wrong with this logic? (This is known as the Gambler's Fallacy)

He is not accounting for the fact that all of the results are independent. The previous outcomes are not going to affect the next one. It is true that getting 6 in a row on black is extremely unlikely, but given that 5 already happened, the 6th isn't any more likely or unlikely than it would already be.

P(6th black | first 5 are black) = P(black) = 1/2

Worksheet 3

For each of these problem, (if applicable) please use notation that we have adopted in class, i.e., events A or B, probabilities $P(A)$, expressions $P(A|B)$ or $P(A \text{ or } B)$, etc., in addition to solving them numerically.

Question 1

This is the table of data for heart attacks (different than data in the slides).

Treatment	Heart_Attack	No_Heart_Attack
Aspirin	24	1533
Placebo	25	851
Total	49	2384

Part A: What are the odds for having a heart attack in the aspirin group? Simplify the result so it looks like 1:X.

```
# 24:1533 or roughly 1:64
```

Part B: What are the odds for having a heart attack in the placebo group? Simplify the result so it looks like 1:X.

```
# 25:851
# or roughly 1:34
```

Part C: What is the odds ratio for heart attacks for the aspirin group and placebo group?

```
(24 / 1533) / (25 / 851)
```

```
## [1] 0.5329159
```

Part D: According to the odds ratio, are aspirin use and heart attack occurence associated?

```
# Yes. The odds ratio is less than 1.
```

Part E: Explain to someone who has never taken a statistics class what the odds ratio value we got tells us about aspirin's effect on heart attack rates.

```
# Aspirin reduces the rate that people get heart attacks.
```

Question 2 – Titanic Data

```
##         Class
## Survived  1st  2nd  3rd Crew  Sum
##      No   122  167  528  673 1490
##      Yes  203  118  178  212  711
##      Sum  325  285  706  885 2201
```

Part A: For a randomly selected 1st-class passenger, what is the probability they survived?

```r
203/325
```

```
## [1] 0.6246154
```

Part B: For a randomly selected 1st-class passenger, what is their *risk* of death?

```r
122 / 325
```

```
## [1] 0.3753846
```

Part C: For a randomly selected 3rd-class passenger, what is their *risk* of death?

```r
528 / 706
```

```
## [1] 0.7478754
```

Part D: What is the value of *relative risk* of death for 3rd-class passengers compared to 1st-class passengers

```r
(528 / 706) / (122 / 325)
```

```
## [1] 1.992291
```

Part E: According to the *relative risk* value in Part D, is there an association between passenger class and survival?

```r
# Yes. The risk of death for 3rd class passengers is twice that of 1st-class
```

Part F: Explain what the *relative risk* value in Part D means to someone who has never taken a stats class.

```r
# 3rd-class passengers were twice as likely to die on the Titanic than 1st-class passengers
```

Part G: For a randomly selected 1st-class passenger, what are their odds of survival?

```r
203/122
```

```
## [1] 1.663934
```

1.66:1

Part H: For a randomly selected 3rd-class passenger, what are their odds of survival?

```r
178/673
```

```
## [1] 0.2644874
```

0.26:1

Part I: What is the odds ratio for survival between the 1st-class and 3rd-class passengers?

```r
(203/122) / (178/673)
```

```
## [1] 6.291168
```

Part J: According to the odds-ratio, is there an association between passenger class and survival?

Yes. The odds-ratio is not 1, so there is an association.

Part K: Do your answers to Parts E and J agree? Explain.

Yes. Both say the likelihood of dying for 3rd class is much higher whethere we look at odds or probabilities.

Part L: Explain what the odds-ratio means to someone who has never taken a stats class.

The odds of dying are 6.3 times higher for 3rd class passengers than for 1st class. Seems to be that 1st class passengers are more likely to survive than 3rd class (not surprising).

Question 3 – Intermittent Fasting and Heart Health

> Time-restricted eating, a type of intermittent fasting, involves limiting the hours for eating to a specific number of hours each day, which may range from a 4- to 12-hour time window in 24 hours. Many people who follow a time-restricted eating diet follow a 16:8 eating schedule, where they eat all their foods in an 8-hour window and fast for the remaining 16 hours each day, the researchers noted. Previous research has found that time-restricted eating improves several cardiometabolic health measures, such as blood pressure, blood glucose and cholesterol levels.
>
> In this study, researchers investigated the potential long-term health impact of following an 8-hour time-restricted eating plan. They reviewed information about dietary patterns for participants in the annual 2003-2018 National Health and Nutrition Examination Surveys (NHANES) in comparison to data about people who died in the U.S., from 2003 through December 2019.
>
> The study included approximately 20,000 adults in the U.S. with an average age of 49 years. They found that people who followed a pattern of eating all of their food across less than 8 hours per day had a 91% higher risk of death due to cardiovascular disease.

Part A: The study claims that those who follow the 8-hour time restricted diet had a 91% increased risk of dying from cardiovascular disease. *Explain what this means in terms of comparing probability of dying from cardiovascular disease for those who follow the 8-hour diet and those who do not.*

The probability of dying from cardiovascular disease for the group following the diet is 1.91 times the probability of dying from cardiovascular disease for the group who does not follow the diet.

Part B: Just knowing about this increased risk is not enough for us to tell how likely someone is to die from cardiovascular disease if they follow this 8-hour diet. What else do we need to know to figure this out?

We need to know the actual probability of dying for the group not following the diet.

Part C: An estimate I found of the *risk* of dying by cardiovascular disease before 50 in general is 0.000976. For the sake of comparison, let us use this to estimate the *risk* of dying from cardiovascular disease for someone who is not intermittent fasting. Using this information with info from Part A, what is the *risk* of dying from cardiovascular disease for someone intermittent fasting? Would you consider this risk low or high?

```
1.91 * .000976
```

```
## [1] 0.00186416
```

For an individual, it is relatively low.

Part D: The information in this study was not gathered from an experiment. What are some reasons the intermittent fasting group could have higher cardiovascular death rates even if intermittent fasting itself was not the cause?

Answers may vary.