Midterm

Name: Solutions

Question 1: R code interpretation

Consider this R code.

```
vec <- rbinom(n = 1000, size = 3, prob = .5)

## help file for rbinom says:
# "random generation for the binomial distribution with parameters `size` and `prob`.
# This is conventionally interpreted as the number of ‘successes’ in `size` trials."
## Explanation of function arguments in help file:
# n: number of observations.
# size: number of trials (zero or more).
# prob: probability of success on each trial.
```

(1a) What will the output of `mean(vec)` be (approximately)?

Answer: The PMF for this random variable X is

- $P(X = 0) = (1/2)^3 = 1/8$
- $P(X = 1) = 3(1/2)^3 = 3/8$
- $P(X = 2) = 3(1/2)^3 = 3/8$
- $P(X = 3) = (1/2)^3 = 1/8$

The 3s above are there because there are 3 ways to get 1 success and 2 failures or 2 successes and 1 failure.

So `mean(vec)` will be approximately 1.5.

Confirmation:

```
vec <- rbinom(n = 1000, size = 3, prob = .5)
mean(vec)
```

```
[1] 1.488
```

(1b) What will the output of `mean(vec <= 1)` be (approximately)?

Answer: Using the PMF above, $P(X \leq 1) = 1/2$.

```
mean(vec <= 1)
```

```
[1] 0.516
```

(1c) What will the output of `mean(vec[vec <= 1])` be (approximately)?

Answer: This is asking for the mean of X given that $X \leq 1$, i.e. $E[X \mid X \leq 1]$. You could reason about it this way: once you restrict to the observations with $X \leq 1$, you have only 0s and 1s; the proportion of these that are 1s is $(3/8)/(1/8 + 3/8) = 3/4$. So the mean will be .75.

More formally, the question is asking for $E[X \mid X \leq 1] = \sum_{x \leq 1} x f(x|X \leq 1)$. The latter probability is a conditional probability, and you can compute it as you compute other conditional probabilities.

```
mean(vec[vec <= 1])
```

```
[1] 0.7616279
```

Question 2: Proof interpretation

Theorem: If events A and B are independent and $P(B) > 0$, then $P(A \mid B) = P(A)$.

Proof:

$$P(A \cap B) = P(A)P(B) \quad \text{(Step 1)}$$
$$P(A \mid B)P(B) = P(A)P(B) \quad \text{(Step 2)}$$
$$P(A \mid B) = P(A) \quad \text{(Step 3)}$$

Explain what definition/property/mathematical operation is being used in each step of the proof.

(2a) Step 1:

Answer: Definition of independent events.

(2b) Step 2:

Answer: Definition of conditional probability.

(2c) Step 3:

Answer: Dividing through by $P(B)$, which uses the stated assumtion that $P(B) > 0$.

(2d) Explain in words what $P(A \mid B) = P(A)$ means.

Answer: In terms of subjective probabilities, the knowledge that B occurred does not change our assessment of the probability of A will occur or has occurred.

In terms of objective probabilities, the probability of A occurring is the same in the situation where B occurs as when it does not occur (or equivalently, averaging over situations when B either does or does not occur).

Question 3: Joint distribution of two random variables

Consider the joint PMF of two random variables, X and Y:

x	y	$f(x, y)$
0	0	1/6
1	0	1/3
2	0	1/8
2	1	3/8

(3a) What is the marginal distribution of X, i.e. $f_X(x)$?

Answer:

x	$f_X(x)$
0	**1/6**
1	**1/3**
2	**1/2**

(3b) What is the expectation of X, i.e. $E[X]$?

Answer: $E[X] = 1/6 \times 0 + 1/3 \times 1 + 1/2 \times 2 = 4/3$

(3c) What is the variance of X, i.e. $V[X]$?

Answer: We'll use $V[X] = E[X^2] - E[X]^2$.

$E[X^2] = 1/6 \times 0^2 + 1/3 \times 1^2 + 1/2 \times 2^2 = 7/3$

So $V[X] = E[X^2] - E[X]^2 = 7/3 - (4/3)^2 = 21/9 - 16/9 = 5/9$.

(3d) What is the conditional distribution of Y given X, i.e. $f_{Y|X}(y \mid x)$?

Answer:

x	y	$f_{Y\|X}(y \mid x)$
0	0	**1**
1	0	**1**
2	0	**1/4**
2	1	**3/4**