Midterm Solutions

Name: ________________________________

Question 1: `R` code interpretation

Consider this `R` code.

```
artist <- c("da vinci", "da vinci", "caravaggio", "delacroix", "da vinci")
area <- c(5, 6, 9, 12, 8)
rating <- c(7, 6, 10, 2, 5)
```

(1a) What is `mean(rating)`?

$$\frac{7+6+10+2+5}{5} = 6$$

```
#check
mean(rating)
```

```
[1] 6
```

(1b) What is `mean(rating[artist == "da vinci"])`?

$$\frac{7+6+5}{3} = 6$$

```
#check
mean(rating[artist == "da vinci"])
```

```
[1] 6
```

(1c) What is `mean(rating[area > 8])`?

$$\frac{10+2}{2} = 6$$

```r
#Check
mean(rating[area > 8])
```

```
[1] 6
```

```r
# From R help file:
# `sample` takes a sample of the specified `size` from the elements of `x`
# using either with or without replacement.
samp <- sample(x = area, size = 20000, replace = T, prob = rep(1/5, 5))
```

(1d) What is `mean(samp)`, approximately?

$$\frac{5+6+9+12+8}{5} = 8$$

Because the sampling is done with replacement, and because each element in the area vector is equally likely to be selected, `mean(samp)` should be approximately 8.

```r
#check
mean(samp)
```

```
[1] 8.02375
```

Question 2: Proof interpretation

The variance of a random variable X is defined as $\mathrm{V}[X] = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right]$. An alternative formula for variance is

$$\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2.$$

Here is a proof:

$$\begin{aligned}
\mathrm{V}[X] &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] && \text{(Step 1)} \\
&= \mathrm{E}\left[X^2 - 2\mathrm{E}[X]X + \mathrm{E}[X]^2\right] && \text{(Step 2)} \\
&= \mathrm{E}[X^2] - \mathrm{E}\left[2\mathrm{E}[X]X\right] + \mathrm{E}\left[\mathrm{E}[X]^2\right] && \text{(Step 3)} \\
&= \mathrm{E}[X^2] - 2\mathrm{E}[X]\mathrm{E}[X] + \mathrm{E}\left[\mathrm{E}[X]^2\right] && \text{(Step 4)} \\
&= \mathrm{E}[X^2] - 2\mathrm{E}[X]\mathrm{E}[X] + \mathrm{E}[X]^2 && \text{(Step 5)} \\
&= \mathrm{E}[X^2] - \mathrm{E}[X]^2 && \text{(Step 6)}
\end{aligned}$$

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

(2a) Step 1: Definition of Variance.

(2b) Step 2: Expand the squared term.

(2c) Step 3: Linearity of Expectations.

(2d) Step 4: Linearity of Expectations (moving $2\mathrm{E}[X]$ outside of the expectations operator, because it is a constant)

(2e) Step 5: Expectation of a constant is a constant. (Could also say linearity of expectations, or moving a constant outside of the expectations operator.)

(2f) Step 6: Combine like terms.

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. $f_X(x)$?

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

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

$$\mathrm{E}[X] = 0 * \frac{1}{6} + 1 * \frac{1}{3} + 2 * \frac{1}{2} = \frac{4}{3}$$

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

Given $\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$,

$$\begin{aligned} \mathrm{V}[X] &= \left(0^2 * \frac{1}{6} + 1^2 * \frac{1}{3} + 2^2 * \frac{1}{2}\right) - \left(\frac{4}{3}\right)^2 && (1) \\ &= \frac{7}{3} - \frac{16}{9} && (2) \\ &= \frac{5}{9} && (3) \end{aligned}$$

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

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