Midterm practice (2024),

Name: _Solutions___

Question 1: `R` code interpretation

Consider this `R` code.

```
gender <- c("male", "female", "female", "male", "female")
educ <- c(rep("HS", 3), "college", "college")
earnings <- c(1, 0, 2, 8, 4)
```

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

The mean of `earnings` is $(1 + 0 + 2 + 8 + 4)/5 = 3$.

(1b) What is `mean(earnings[gender == "female"])`?

`earnings[gender == "female"]` is the vector `c(0, 2, 4)`, so its mean is $(0+2+4)/3 = 2$.

(1c) What is `mean(earnings[gender == "female" & educ == "HS"])`?

`earnings[gender == "female" & educ == "HS"]` is the vector `c(0, 2)`, so its mean is 1.

(1d) What is `mean(earnings[earnings <= 4])`?

`earnings[earnings <= 4]` is the vector `c(0, 1, 2, 4)`, so its mean is $7/4$.

Question 2: Proof interpretation

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

Proof:

$$
\begin{aligned}
P(A \cap B) &= P(A)P(B) && \text{(Step 1)} \\
P(A \mid B)P(B) &= P(A)P(B) && \text{(Step 2)} \\
P(A \mid B) &= P(A) && \text{(Step 3)}
\end{aligned}
$$

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 (or product rule).

(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 (long-run frequency) 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	1/4
0	1	1/5
1	0	1/5
1	1	1/10
1	2	1/4

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

Answer:

x	$f_X(x)$
0	**9/20**
1	**11/20**

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

Answer: $E[X] = 9/20 \times 0 + 11/20 \times 1 = 11/20$

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

Answer: We'll use $V[X] = E[X^2] - E[X]^2$. (We could also note that this is a Bernoulli random variable, so the variance is $p(1-p)$.)

$E[X^2] = 9/20 \times 0^2 + 11/20 \times 1^2 = 11/20$

So $V[X] = E[X^2] - E[X]^2 = 11/20 - (11/20)^2 = 11/20 - 121/400 = 220/400 - 121/400 = 99/400$.

(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	5/9
0	1	4/9
1	0	4/11
1	1	2/11
1	2	5/11