Midterm exam

Name: _____

October 28, 2022

1. Probability and random variables

Suppose X and Y are random variables. The joint PMF for X and Y is the following:

$$f_{X,Y}(x,y) = \begin{cases} 1/4 & x=0, y=0 \\ 1/5 & x=0, y=1 \\ 1/5 & x=1, y=0 \\ 1/10 & x=1, y=1 \\ 1/4 & x=1, y=2 \\ 0 & \text{otherwise} \end{cases}$$

a. Write the marginal PMF of X.

Answer:

$$f_X(x) = \begin{cases} 9/20 & x=0, \\ 11/20 & x=1 \\ 0 & \text{otherwise} \end{cases}$$

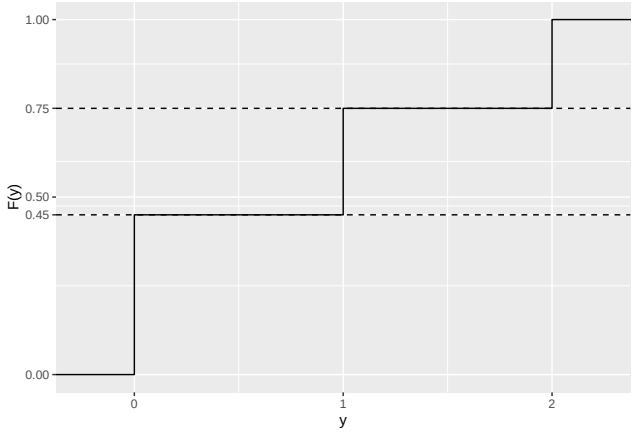
b. What is $E[X]$? Show your work.

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

c. What is $\Pr[X=1|Y=1]$? Show your work.

Answer: $\Pr[X=1|Y=1] = \frac{\Pr[X=1,Y=1]}{\Pr[Y=1]} = \frac{1/10}{1/5+1/10} = \frac{1}{3}$

d. Draw the CDF of Y.



Now suppose X and Y are Bernoulli random variables (i.e. random variables that take only the value 0 or 1), with $p = \Pr[X = 1]$ and $q = \Pr[Y = 1]$.

e. Write the joint PMF for X and Y, assuming that X and Y are independent.

Answer:

$$f_{X,Y}(x,y) = \begin{cases} (1-p)(1-q) & x = 0, y = 0 \\ (1-p)q & x = 0, y = 1 \\ p(1-q) & x = 1, y = 0 \\ pq & x = 1, y = 1 \\ 0 & \text{otherwise} \end{cases}$$

2. Causal quantities and coding exercises.

$Y_i(0)$	$Y_i(1)$	D_i
1	0	1
2	-3	0
0	1	1
3	2	1
1	1	1
5	-8	1
2	2	0
1	-1	1
7	4	1
0	0	0

a. Write out by hand the code you would use to create a tibble (data frame) representing the above table.

Answer:

```
df <- tibble(Y0 = c(1,2,0,3,1,5,2,1,7,0),
             Y1 = c(0, -3, 1,2,1,-8,2,-1,4,0),
             D = c(1,0,1,1,1,1,0,1,1,0))

df
```

```
## # A tibble: 10 x 3
##       Y0    Y1     D
##    <dbl> <dbl> <dbl>
##  1     1     0     1
##  2     2    -3     0
##  3     0     1     1
##  4     3     2     1
##  5     1     1     1
##  6     5    -8     1
##  7     2     2     0
##  8     1    -1     1
##  9     7     4     1
## 10     0     0     0
```

b. Write out by hand the code you would use to add a new column, `Y`, that represents the result of the "switching" function for every individual.

$$Y_i = D_i \times Y_i(1) + (1 - D_i) \times Y_i(0)$$

Answer:

```
# using case_when
df <- df |>
  mutate(Y = case_when(D == 1 ~ Y1,
                       TRUE ~ Y0))

# alternatively
df <- df |>
  mutate(Y = Y1*D + Y0*(1 - D))
```

c. What is the value of the average treatment effect for this population? Compute by hand.

$$-\frac{2}{10} - \frac{22}{10} = -\frac{12}{5}$$

d. What is the difference in means estimate for the average treatment effect using the observed data? Compute by hand.

Answer:

$$-\frac{1}{7} - \frac{4}{3} = -\frac{31}{21}$$