hw6 Beimnet Taye

2023-03-05

P1

1.

- That they are independent. $D \perp Y$.

2.

- No since those with higher blood pressure are more likely to be insured to cover potential adverse health outcomes. Healthier individuals are less likely to be motivated to buy health insurance.

3.

- $E[Y] = E[E[Y \mid X]]$ This is arrived at through the total expectation identity stating that $E[X] = E[E[X \mid Y]]$
- $= E[E[Y \mid D = 1, X]]$ This is arrived at through conditioning on an independent RV where $E[X \mid Y, Z] = E[X \mid Z]$ given X and Y are conditionally independent on Z, in this case Y and D are conditionally independent on X
- $= E[\mu(X)]$ This is arrived at with through the identity provided in the above problem where $E[Y \mid D = 1, X] = \mu(X)$

P2

1.

```
n = 1e6
full_data = tibble(
  X = runif(n),
  D = rbern(n, X),
  Y = X^2 + rnorm(n)
)

EY <- mean(full_data$Y)

observed_data <-  full_data %>%
  mutate(Y = ifelse(D, Y, NA))

yobv <- observed_data %>%
  filter(D == 1) %>%
  mutate(X2 = X^2) # used E[Y] = E[mu[X]] = E[X^2] since mu(x) = x^2
eyobv <- mean(yobv$X2)
eyobv
```

```
## [1] 0.4993496
```

2.

- 1st bullet: correct since Y given X,U is distributed normally with a mean of $X^2 + U$ and the expectation of a normal distribution is just its mean which is $X^2 + U$ in this case
- 2nd bullet: U cannot equal 1 given D is 1 since U and D are disjoint. If $U = 1$ then $D = (1-1)*Bern(X)$ which is equal to zero so D can only equal zero when U is 1. The same logic can be applied to the second part of this bullet; U can only equal zero when D is 1.
- 3rd bullet: This equality is arrived at through the law of total expectation. Since $P(U = 1 \mid D = 1, X) = 0$ and $P(U = 0 \mid D = 1, X) = 1$ we can substitute these values in resulting in: $E[Y \mid D = 1, X] = E[Y \mid D = 1, X, U = 1] * 0 + E[Y \mid D = 1, X, U = 0] * 1$ which simplifies to $E[Y \mid D = 1, X] = E[Y \mid D = 1, X, U = 0]$. The wizard knows from the DGP that $Y \mid X, U$ is normally distributed with mean $X^2 + U$ so with U being 0 and it being normally distributed its mean parameter is the expectation, $E[Y \mid D = 1, X, U = 0] = X^2 + 0 = X^2$. So $E[Y \mid D = 1, X] = \mu(X)$.

3.

```
mu = function(x) x^2
n = 1e6
full_datau = tibble(
  U = rbern(n, 1/2),
  X = runif(n),
  D = U*rbern(n, X),
  Y = X^2 + U + rnorm(n)
)
observed_datau = full_datau %>%
  mutate(Y = ifelse(D, Y, NA),
         X2 = mu(X)) %>%
  select(-U)

obsveyu <- observed_datau %>%
  filter(D==1)
Eyu <- mean(obsveyu$X2)

Eyu
```

```
## [1] 0.5004113
```

- Y is not conditionally independent with D on X in this DGP violating the assumption we used previously.

P3

1.

- Since the expectation of a conditioned RV is that function's output times the relative frequency of that RV's output given the condition summed over all possible RV values. Here function g(x,y)'s output is being multiplied by the relative frequency of said output given a particular z, this is summed over all possible values of x and y.

2.

- If $(X \perp Y) \mid Z$
- $f_{X,Y|Z}(x,y,z) = f_{X|Z}(x,z) * f_{Y|Z}(y,z)$: Property of conditional independence.
- $E[X \mid Y] = \sum_x x * f_{X|Y}(x,y)$: Definition of conditional expectation.
- $E[XY \mid Z] = \sum_x \sum_y xy * f_{X,Y|Z}(x,y,z)$: Applying definition of conditional expectation.
- $= \sum_x \sum_y xy * f_{X|Z}(x,z) * f_{Y|Z}(y,z)$: Applying property of conditional independence via substitution.
- $= \sum_x x * f_{X|Z}(x,z) * \sum_y y * f_{Y|Z}(y,z)$: Rearranging terms.
- $= E[X \mid Z] * E[Y \mid Z]$: Simplifying with definition of conditional expectation.

P4

- $C[X,Y] = E[XY] - E[X] * E[Y]$: Definition of Covariance
- $C[aX + bY, Z] = E[(aX + bY) * Z] - E[aX + bY] * E[Z]$: Applying definition
- $= E[aXZ + bZY] - E[aX + bY] * E[Z]$: Distributing Z
- $= E[aXZ + bZY] - ((E[aX] + E[bY]) * E[Z])$: linearity of expectation
- $= E[aXZ + bZY] - E[aX] * E[Z] - E[bY] * E[Z]$: simplifying
- $= E[aXZ] + E[bZY] - E[aX] * E[Z] - E[bY] * E[Z]$: linearity of expectation
- $= aE[XZ] + bE[ZY] - aE[X] * E[Z] - bE[Y] * E[Z]$: factoring constants outside of expectation
- $= aE[XZ] - aE[X] * E[Z] + bE[ZY] - bE[Y] * E[Z]$: rearranging terms
- $= a(E[XZ] - E[X] * E[Z]) + b(E[ZY] - E[Y] * E[Z])$: factoring constants
- $= aC[X,Z] + bC[Y,Z]$: simplifying using definition of covariance

P5

1

- Conditional variance is a function that returns the variance of a random variable given a particular value of another. In other words for Random Variables X and Y inputted, given a particular value for Y, it returns the variance of X. $Var(X \mid Y) = E[(X - E[X \mid Y])^2 \mid Y]$. This can be demonstrated visually with the below graphic:

```
graph <- tibble(Y = runif(10000),
                X = rnorm(10000,0,Y +.02)
                ) %>%
  ggplot()+
  geom_point(aes(x=Y, y=X)) +
  geom_vline(xintercept = .5, color = "blue") +
  geom_vline(xintercept = .75, color = "blue") +
  geom_vline(xintercept = 0, color = "red") +
  geom_vline(xintercept = .25, color = "red") +
  geom_hline(yintercept = 0, color ="green")
graph
```

Conditional Variance Figure
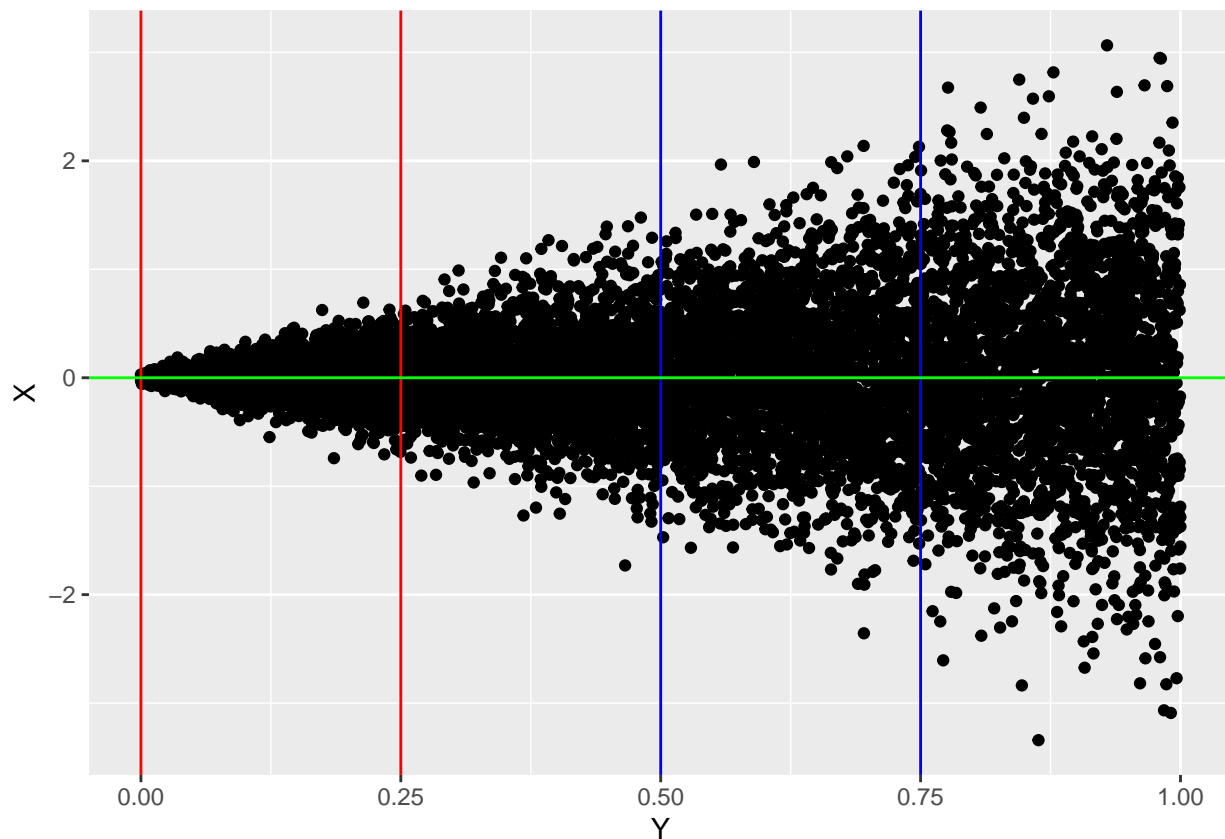
- As you can see in the figure above as Y changes the variance of X changes. Specifically as Y increases the variance around X's expectation, 0, also increases creating a "fan-out" like figure. The variance of X when 0<Y<.25 (the region between the red lines) does not equal the variance of X when 0.5<Y<0.75 (the region between the blue lines). This difference in conditional variances can also be quantitatively demonstrated using sampling.

```r
dgpv <- tibble(Y = runif(10000),
               X = rnorm(10000,0,Y +.02)
               )
rvv <- dgpv %>%
  filter(Y<0.25&Y>0) %>%
  pull(X) %>%
  var()

bvv <- dgpv %>%
  filter(Y<0.75&Y>0.5) %>%
  pull(X) %>%
  var()
conditional_variances <- c(rvv,bvv)
conditional_variances
```

```
## [1] 0.02438174 0.41889083
```

2

Given the DGP and the scatter plot below what can you tell me about the conditional variances of X given the two colored regions below. Use how the DGP is coded and the given graph to justify your answer.(Extra credit: Show this numerically using sampling.)

```
graphp <- tibble(Y = runif(10000),
                 X = rnorm(10000,0, 1)
                 ) %>%
  ggplot()+
  geom_point(aes(x=Y, y=X)) +
  geom_vline(xintercept = .5, color = "blue") +
  geom_vline(xintercept = .75, color = "blue") +
  geom_vline(xintercept = 0, color = "red") +
  geom_vline(xintercept = .25, color = "red") +
  geom_hline(yintercept = 0, color ="green")
graphp
```
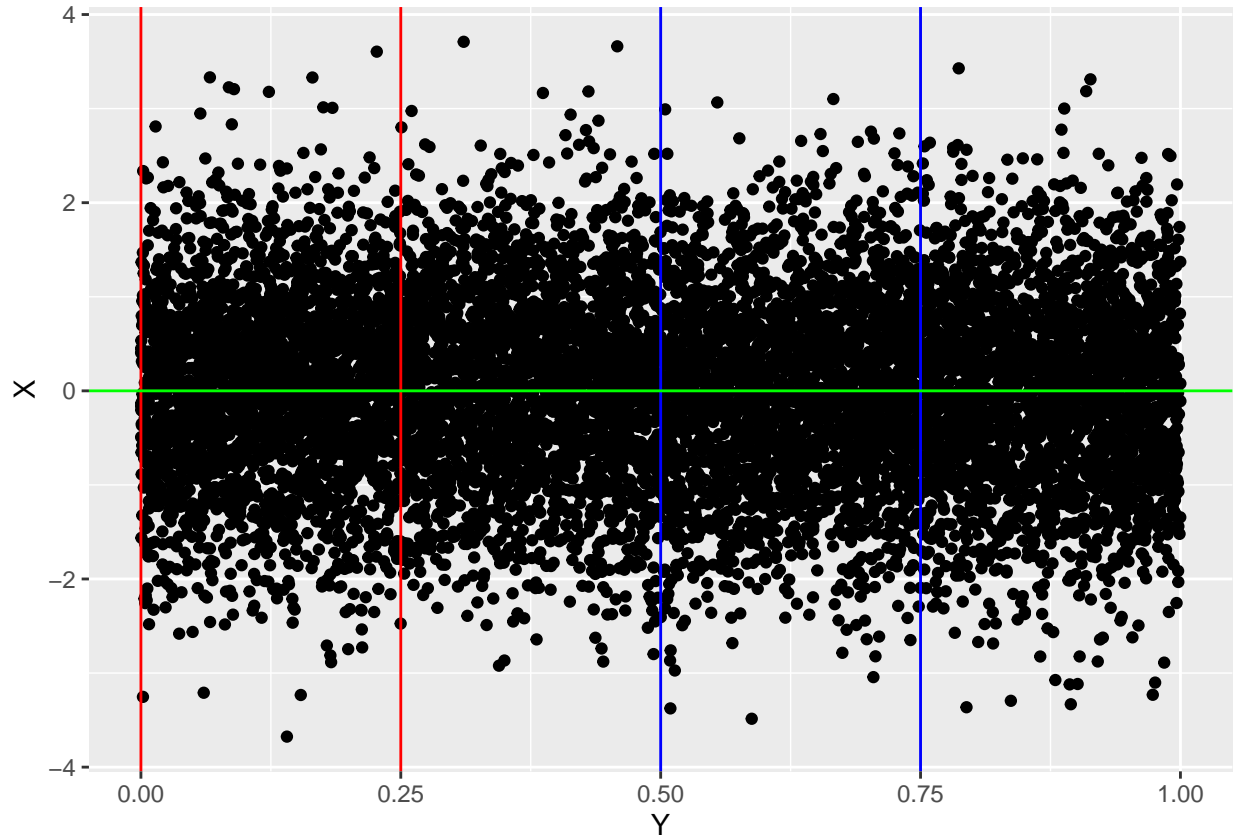

Solution:

- The conditional variance of the two regions should roughly be the same. If you visually inspect the scatter plot X varies from its expected value, 0, in roughly the same manner throughout the graph, meaning the variance does not change for any given interval of Y. Looking at the DGP, the standard deviation (and thus the variance) of X is not a function of Y meaning the variance of X should not

5

meaningfully depend on the value of Y. This can be proven mathematically using sampling with the code below:

```r
dgp <- tibble(Y = runif(10000),
              X = rnorm(10000,0, 1) )

rv <- dgp %>%
  filter(Y<0.25&Y>0) %>%
  pull(X) %>%
  var()

bv <- dgp %>%
  filter(Y<0.75&Y>0.5) %>%
  pull(X) %>%
  var()
sol_conditional_variances <- c(rv,bv)
sol_conditional_variances
```

```
## [1] 0.9457008 1.0329979
```

P6

1

```r
n = 1e4
X = rnorm(n)
var(X)
```

```
## [1] 0.9968908
```

- There is a difference since you are sampling from the DGP and thus only grabbing a subset of all possible values. This means that, through pure chance, you can grab a sample that tilts the variance slightly higher or lower than the true value. This can be rectified by increasing the sample size; as the sample size approaches infinity the variance will converge at 1.0.