Math 151 - Probability Theory - Homework 10

your name here

Due: Friday, October 30, 2020, midnight PDT

Important Note:

You should work to turn in assignments that are clear, communicative, and concise. Part of what you need to do is not print pages and pages of output. Additionally, you should remove these exact sentences and the information about HW scoring below.

Click on the *Knit to PDF* icon at the top of R Studio to run the R code and create a PDF document simultaneously. [PDF will only work if either (1) you are using R on the network, or (2) you have LaTeX installed on your computer. Lightweight LaTeX installation here: https://yihui.name/tinytex/]

> Either use the college's RStudio server (https://rstudio.pomona.edu/) or install R and R Studio on to your personal computer. See: https://research.pomona.edu/johardin/math151f20/ for resources.

Assignment

1: PodQ

Describe one thing you learned from someone in your pod this week (it could be: content, logistical help, background material, R information, etc.) 1-3 sentences.

2: 4.6.9

Suppose that X and Y are two random variables which may be dependent and $Var(X) = Var(Y)$. Assuming that $0 < Var(X+Y) < \infty$ and $0 < Var(X-Y) < \infty$, show that the random variables $X+Y$ and $X-Y$ are uncorrelated.

3: 4.6.12

Suppose that X and Y have a continuous joint distribution for which the pdf is

$$f(x,y) = \begin{cases} \frac{1}{3}(x+y) & 0 \le x \le 1 \text{ and } 0 \le y \le 2 \\ 0 & \text{else} \end{cases}$$

Determine the value of $Var(2X - 3Y + 8)$.

4: 4.6.14

Suppose that X, Y and Z are three random variables such that $Var(X) = 1$, $Var(Y) = 4$ and $Var(Z) = 8$. Suppose also that $Cov(X,Y) = 1, Cov(X,Z) = -1$ and $Cov(Y,Z) = 2$. Determine

a. $Var(X+Y+Z)$,
b. $Var(3X - Y - 2Z + 1)$.

5: 4.6.17

Let X and Y be random variables with finite variance. Prove that $|\rho(X,Y)| = 1$ implies that there exists constants a, b and c such that $aX + bY = c$ with probability 1.

6: R - zero covariance

Let's say that X and Y are both discrete. X can be any of the integers between -5 and +5. $Y = X^2$. Because Y is a function of X, they are not independent. That is, the distribution of Y given X is a point mass (i.e., the value of Y is fixed for a given X).

Consider the following distributions of X. **For which of the five distributions given below is the correlation (and covariance) equal to zero?** Give your response as (1) a simulation, and (2) an analytic or intuitive justification.

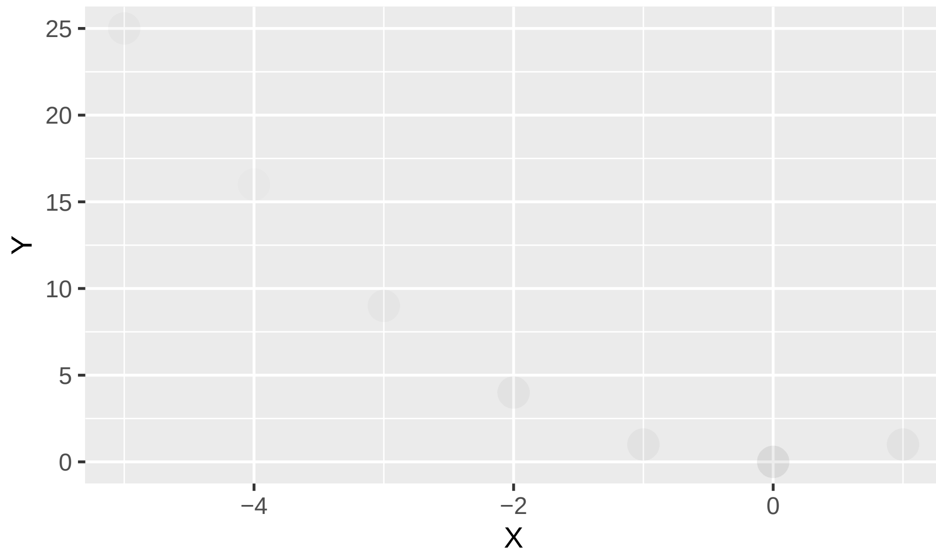
a. $f_X(x) = \frac{1}{11} \quad x = -5, -4, 3, -2, -1, 0, 1, 2, 3, 4, 5$
b. $f_X(x) = \frac{6-|x|}{36} \quad x = -5, -4, 3, -2, -1, 0, 1, 2, 3, 4, 5$
c. $f_X(x) = \frac{6-x}{66} \quad x = -5, -4, 3, -2, -1, 0, 1, 2, 3, 4, 5$
d. $f_X(x) = \frac{25-x^2}{165} \quad x = -5, -4, 3, -2, -1, 0, 1, 2, 3, 4, 5$
e. $f_X(x) = \frac{10+x}{110} \quad x = -5, -4, 3, -2, -1, 0, 1, 2, 3, 4, 5$

Some code that might be helpful for (2) is written below. There are three arguments for the `sample()` function. You should make sure that you understand all three arguments.

```r
# good idea to make your code reproducible.  how do you do that??

possible.ex = seq(-5,5,by=1)
nsamps <- 20  # you need this number to be bigger.  why?

# code the pdf, note you'll have to re-do this line for all 5 pdfs
sample.ex <- sample(possible.ex, nsamps, prob = (6 - abs(possible.ex))/36, replace=TRUE)
sample.why <- sample.ex^2

cor(sample.ex, sample.why)  # what are you measuring here?
```

```
## [1] -0.9060699
```

```r
# the code below will produce a scatterplot where the shading of the point
# is based on how many points are plotted at the exact coordinate.
# (as above, nsamps needs to be bigger to see the effect)
library(tidyverse)
data.frame(sample.ex, sample.why) %>%
  ggplot(aes(x=sample.ex, y=sample.why)) + geom_point(alpha=.01, cex=5) +
  xlab("X") + ylab("Y")
```


25
20
15
10
5
0
Y
−4
−2
0
X