14.310x: Data Analysis for Social Scientists Joint, Marginal, and Conditional distributions and Functions of Random Variables

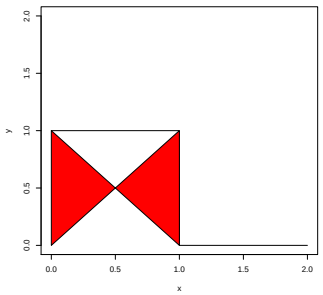
Welcome to your fourth homework assignment! You will have about one week to work through the assignment. We encourage you to get an early start, particularly if you still feel you need more experience using R. We have provided this PDF copy of the assignment so that you can print and work through the assignment offline. You can also go online directly to complete the assignment. If you choose to work on the assignment using this PDF, please go back to the online platform to submit your answers based on the output produced.

Good luck :)!

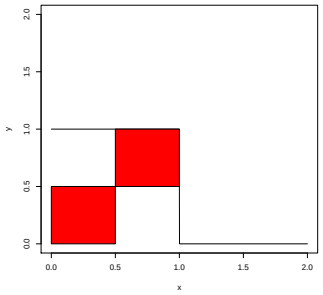
Suppose two sisters, Caroline and Anna, sleep in adjoining rooms. Each has a speaker over which she plays music, and each speaker has a volume dial going from 0 to 1. The joint distribution of the volumes of the two speakers is $f_{XY}(x, y) = c(x + y^2)$ over the unit square, 0 otherwise. (Caroline's volume is denoted by X, Anna's by Y.)

1) Which of the following figures represent the domain in which the density function is defined as $f_{XY}(x, y) = c(x + y^2)$?

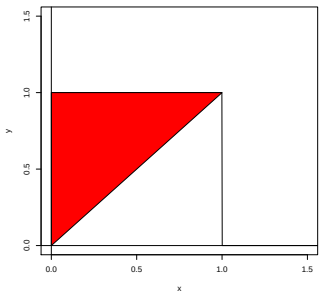
   (a) The domain is:



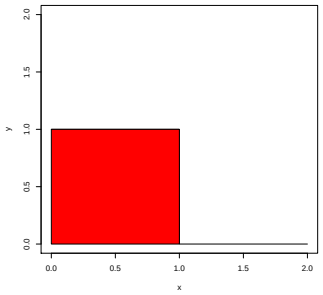
   (b) The domain is:



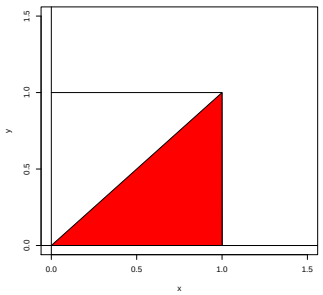
   (c) The domain is:



(d) The domain is:



(e) The domain is:



2) What does the constant c represent? (Select all that apply)

(a) The constant c is a parameter whose value assures that the joint PDF integrates to 1.

(b) The constant c represents a parameter that changes both the joint PDF and the joint CDF of the random variables X and Y.

(c) The constant c is an irrelevant parameter in the shape of the joint CDF of the random variables X and Y.

(d) The constant c is a parameter that helps to infer whether the random variables X and Y are independent.

3) What is the value of the constant c in this case?

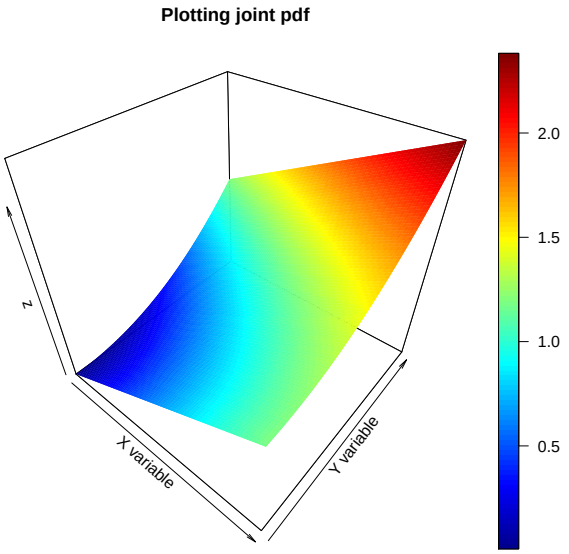
Now we are going to work in R to plot the bivariate pdf. Take a look at the following code in order to create a grid and a 3-dimensional plot of the pdf. Please note that you might need to install the package plot3D.

```
#Preliminaries
#-----------------------------------------------------------
rm(list = ls())
library("utils")
#install.packages('plot3D')
library(plot3D)
setwd("/Users/raz/Dropbox/14.31 edX Building the Course/Problem Sets/PSET 4")

#Creating the vector x and y
M <- mesh(seq(0, 1, length=100), seq(0, 1, length=100))
x <- M$x
y <- M$y
z <- 6/5*(M$x + M$y^2)

#Plotting this pdf
persp3D(x, y, z, xlab = 'X variable', ylab = 'Y variable', xlim = c(0,1), main = 'Plotting joint pdf')
```

4) The following plot was created by running the code. A student is claiming that this plot is wrong since there are certain regions in which the PDF shows values larger than 1. Is this student correct that there is a mistake and therefore the plot does not correspond to the information given in the problem?



(a) Yes

(b) No

5) Are the volumes of the two speakers independent random variables?

(a) Yes

(b) No

6) What is the formula for the marginal distribution of Anna's speaker volume?

a) $f_Y(y)$ is given by $\frac{5}{6}\left(\frac{1}{2}+y^2\right)$

b) $f_Y(y)$ is given by $\frac{6}{5}\left(\frac{1}{2}+y^2\right)$

c) $f_Y(y)$ is given by $\frac{6}{5}\left(\frac{1}{2}+\sqrt{y}\right)$

d) $f_Y(y)$ is given by $\frac{5}{6}\left(\frac{1}{2}+\sqrt{y}\right)$

7) What is the conditional distribution of Caroline's volume as a function of Anna's?

(a) This is given by $\frac{\left(x+y^2\right)}{\left(\frac{1}{2}+y^2\right)}$

(b) This is given by $\frac{\frac{5}{6}\left(x+y^2\right)}{\frac{6}{5}\left(\frac{1}{2}+y^2\right)}$

(c) This is given by $\frac{\left(x+\sqrt{y}\right)}{\left(\frac{1}{2}+y^2\right)}$

(d) This is given by $\frac{\frac{6}{5}\left(x+y^2\right)}{\left(\frac{1}{2}+y^2\right)}$

8) From this conditional distribution can you infer whether Caroline likes Anna's music or not? (think whether Caroline's stereo volume is lower when Anna's is higher)

(a) Caroline does like Anna's music.

(b) Caroline does not like Anna's music

9) What is the probability that Caroline's volume is less than $\frac{1}{2}$ if Anna's volume is $\frac{1}{2}$?

10) Now, what is the marginal distribution of Caroline's speaker volume?

(a) It is given by $\frac{5}{6}\left(x+\frac{2}{3}\right)$

(b) It is given by $\frac{5}{6}\left(x+\frac{1}{3}\right)$

(c) It is given by $\frac{6}{5}\left(x+\frac{2}{3}\right)$

(d) It is given by $\frac{6}{5}\left(x+\frac{1}{3}\right)$

11) Is there a First Order Stochastic Dominance relationship between the random variables X and Y? (We suggest you compute the cdf's of both variables and plot them in R.)

(a) The distribution of X FOSD the distribution of Y

(b) The distribution of Y FOSD the distribution of X

(c) There is no clear relationship

12) Can we say that Anna or Caroline prefer higher volumes?

a) Anna

b) Caroline

c) We can't say

13) Assume that the random variable X has a pdf given by $f_X(x) = 1$ for $0 < x < 1$. What is the pdf of the random variable $Y = X^2$?

(a) $f_Y(y) = \sqrt{y}$ for $0 < y < 1$

(b) $f_Y(y) = \frac{1}{2\sqrt{y}}$ for $0 < y < 1$

(c) $f_Y(y) = \frac{1}{2\sqrt{y}}$ for $-1 < y < 1$

(d) $f_Y(y) = \frac{1}{2}y^{-\frac{3}{2}}$ for $-1 < y < 1$

14) Suppose X has the geometric pmf $f_X(x) = \frac{1}{3}\left(\frac{2}{3}\right)^x$ for $x = 0, 1, 2, \cdots$. What is the probability distribution of $Y = \frac{X}{X+1}$ its pmf? *Note that both X and Y are discrete random variables.*

(a) $f_Y(y) = \frac{1}{3}\left(\frac{2}{3}\right)^{\frac{y}{1-y}}$ for $y = 0, 1, 2, \cdots$

(b) $f_Y(y) = \frac{1}{3}\left(\frac{2}{3}\right)^{\frac{1-y}{y}}$ for $y = 0, 1, 2, \cdots$

(c) $f_Y(y) = \frac{1}{3}\left(\frac{2}{3}\right)^{\frac{y}{1-y}}$ for $y = 0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \cdots, \frac{x}{x+1}, \cdots$

(d) $f_Y(y) = \frac{1}{3}\left(\frac{2}{3}\right)^{\frac{1-y}{y}}$ for $y = 0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \cdots, \frac{x}{x+1}, \cdots$

15) If the random variable X has a pdf given by:

$$f_X(x) = \begin{cases} \frac{x-1}{2} & 1 < x < 3 \\ 0 & otherwise \end{cases}$$

then it is possible to find a monotone function $u(x)$ such that the random variable $u(X)$ has a uniform distribution between 0 and 1?

(a) Yes

(b) No

16) If we have N i.i.d random variables from the uniform distribution between 0 and 1, and we know that $N = 1$, what is the probability that the n^{th} order statistic is less or equal to the value x.

The following code can be run in R to create a draw of 1000 numbers from the uniform distribution.

```
#Creating a random draw of 1000 numbers
u <- runif(1000)
```

17) Is it possible to create from this vector a random draw of a uniform distribution between 2 and 5?

a) Yes

b) No

18) What is the PDF of the minimum of the draw created in R?

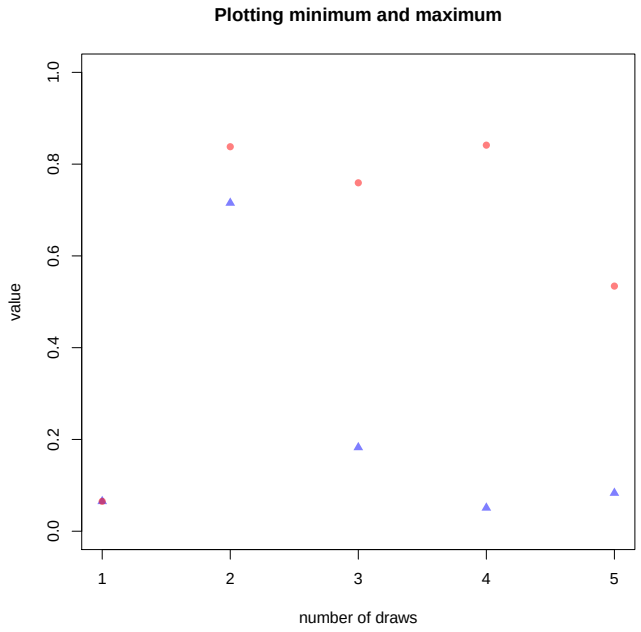
a) It is given by $f_{y^{(1)}}(y) = 999(1-y)^{998}$

b) It is given by $f_{y^{(1)}}(y) = 1000(1-y)^{999}$

c) It is given by $f_{y^{(1)}}(y) = 999(1-y)^{1000}$

d) It is given by $f_{y^{(1)}}(y) = 999y^{998}$

19) The following plot shows the maximum and the minimum of a uniform distribution by changing the number of draws.



A student is claiming that this plot is wrong since both, the maximum and the minimum should show a monotonous relationship with the number of draws. Is this student's statement True or False?

a) True

b) False

c) We can't tell

20) What is the command in R that allows you to transform this draw of random numbers into the one of a Standardized Normal distribution? (Please just enter the name of the command, without any parentheses or arguments)