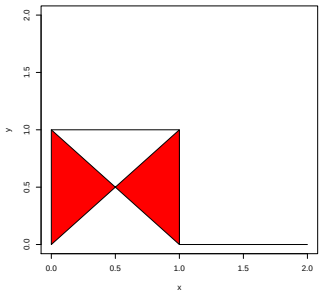
14.310x: Data Analysis for Social Scientists Describing Data, Joint and Conditional distributions - Part 2

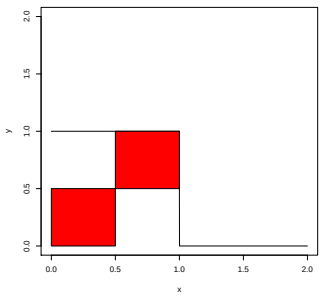
Suppose two sisters, Caroline and Anna, sleep in adjoining rooms. Each has a speaker over which she plays music, and each speaker has a volume dial going from 0 to 1. The joint distribution of the volumes of the two speakers is $f_{XY}(x,y) = c(x+y^2)$ over the unit square, 0 otherwise. (Caroline's volume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

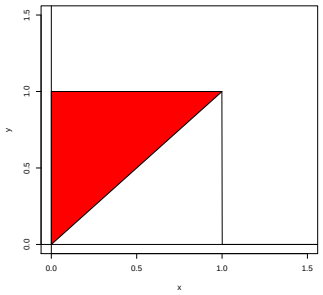
   (a) The domain is:



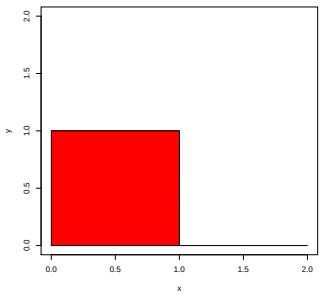
   (b) The domain is:



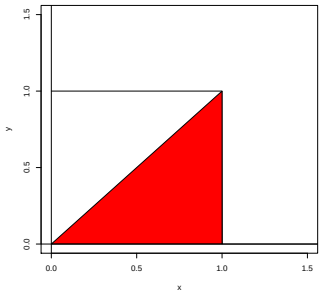
   (c) The domain is:



   (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?

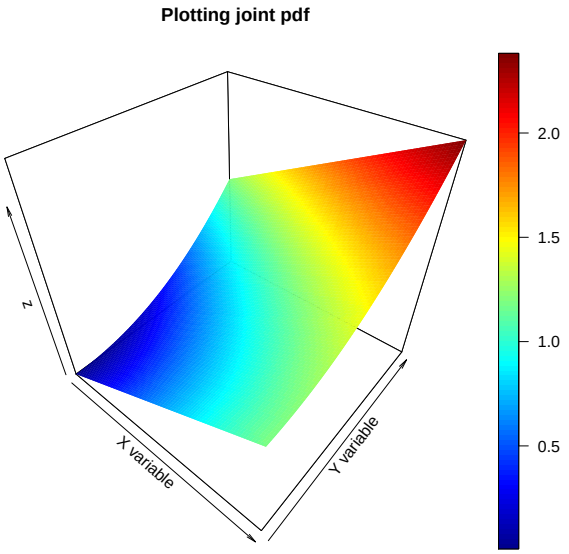
Now we are going to work in R to plot the bivariate PDF. Take a look at the following code in order to create a grid and a 3-dimensional plot of the PDF. Please note that you might need to install the package plot3D.

```
#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)

Hint: 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