1 Joint random variables

some review questions, Dec 27, 2009

1. If X and Y are two random variables with N observations. If $\overline{X}$ and $\overline{Y}$ are the sample means, can you show that $\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y}) = \sum_{i=1}^{N}(x_i - \overline{X})y_i + \sum_{i=1}^{N}((y_i - \overline{Y})x_i$, and that $\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y}) = \sum_{i=1}^{N} x_i y_i - N\overline{XY}$.
answer: .....

2. Two random variables X and Y, have the following joint density function $f(x, y) = 2 - x - y$ if $\leq x \leq 1$ and $0 \leq y \leq 1$, and zero otherwise. Show that $f(x, y)$ is a proper density function. Calculate the marginal density function of x, $f(x)$ and the conditional density function of y given x, $f(y\,|x)$. Determine if the two random variables are statistically independent. Calculate $E[Y\,|X = 1]$ and $Var[Y\,|X = 1]$.

3. Make up a simple joint density function with two random variables (X and Y), where X and Y are not independent. They can be either continuous or discrete random variables, but not a mixture. Given your $f(x, y)$. Find, $f_X(x)$, $f_Y(y)$, $E[X]$, $E[Y]$, $var(X)$, $var(Y)$, $cov(X, Y)$. Prove that X and Y are not independent.

4. Make up a simple joint density function with three random variables (X, Y and Z), where X, Y and Z are not independent. They can be either continuous or discrete random variables, but not a mixture. Given your $f(x, y, z)$. Find, $f_X(x)$, $f_Y(y)$, $f_Z(z)$, $E[X]$, $E[Y]$, $E[Z]$, $var(X)$, $var(Y)$, $var(Z)$, $cov(X, Y)$, and $cov(X, Z)$. Prove that X, Y and Z are not independent.

5. Assume that X and Y are both continuous random variables. Derive $var(X + Y)$ as a function of $var(X)$, $var(Y)$, and $cov(X, Y)$. How does this simplify if X and Y are independent?

6. Consider conditional expectations. Assume some joint density function $f(x, y)$, and some function of x and y, $g(x, y)$. Assume both x and y are continuous random variables. What is?

$$E[g(x, y) : X = x]$$

7. Let the random variables X and Y have the following joint pdf

X, Y	$1,1$	$1,2$	$1,3$	$2,1$	$2,2$	$2,3$
$f(X, Y)$	$2/15$	$4/15$	$3/15$	$1/15$	$1/15$	$4/15$

Determine and report its CDF, $F(a,b)$. Calculate the marginal distributions $f_X(x)$ and $f_Y(y)$. Calculate the conditional mean $E[Y\,|X=1]$. Calculate the conditional variance $Var[Y\,|X=1]$. Calculate the correlation coefficient ρ between X and Y.

8. Consider two random variables, X and Y. Provide an example density function where the two variables are uncorrelated, but dependent (not independent). Not an example from the book. Make sure to explain why the are uncorrelated and why they not independent.

9. Assume two jointly distributed random variables, X and Y. Specifiy a joint density function. Derive three different sorts of conditional density functions for X.

10. Consider two random variables, X and Y. Provide an example where the two variables are independent, but correlated. Not an example from the book.

    answer: there are no examples. If two variables are independent, they are not correlated. Can you prove that there are no examples.

11. Consider two random variables, X and Y. Provide an example of a density function where the two variables are independent, and uncorrelated. Not an example from the book.

    answer: Any example where X and Y are independent will suffice. Independent implies uncorrelated.

12. Assume the random variables, X, Y and Z are jointly distributed. Specify a joint density function (not the Normal or the Uniform - try to be creative) for these three variables and convince the reader that what you have specified is in fact a joint density function. For this density function derive $f(x\,|a<Y<b,\ c<Z)$.

13. Correlation comes in degress (something can be more or less correlated). Can two variables be more independent than two other variables? That is, does dependence come in degress?

14. Consider the experiment of sequentially tossing two fair "four-sided" die - each side has a one in four chance of being up. Let d_1 be the out outcome of the first die, and d_2 the outcome of the second die. How many outcomes are there and what is the probability of each of them. Now define a bunch of events in terms of two random variables: the value of the first die and the sum of the values of the two die. Denote each of the events as the vector (d_1, s), where, for example $(2,4)$ is an event, and $(1,2)$ is a different event. How many events are there? What is the probability of each event. Specify the density function for the vector of random variables, d_1 and s. Specify the CDF for this vector of two random variables. Specify the two marginal density functions. Specify the two conditional densities: $f(d_1\,|s=4)$ and

answer: Consider the sample space

$d_1 \backslash d_2$	1	2	3	4
4	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
2	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

There are 16 outcomes and each has probability of $\frac{1}{16}$. But we are not directly concerned with the probability of each outcome, we are concerned with the probabilities of the events of interest. How many events are there. Is $(1,1)$ an event? Yes, but it can't happen so has probability zero. There are 32 events.

$d_1 \backslash s$	1	2	3	4	5	6	7	8
4	0	0	0	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
3	0	0	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	0
2	0	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	0	0
1	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	0	0	0

Now let's figure out the probabilities. Events where $s \leq d_1$ cannot happen, so these all have zero probability. There are 10 of these. If $s - d_1 > 4$, the event is also impossible. so, $(1,6),(1,7),(1,8),(2,7)(2,8),(3,8)$ have probability zero; 6 of these. So 16 events are impossible. If I am thinking about this correctly, all the rest of the 16 other outcomes are equally likely, so the probability of each is $\frac{1}{16}$ So, the density function, in table form, is

$f(d_1, s) =$

$d_1 \backslash s$	1	2	3	4	5	6	7	8
4	0	0	0	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
3	0	0	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	0
2	0	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	0	0
1	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	0	0	0

Now, what is the corresponding CDF, $F(a,b) = \Pr[d_1 \leq a \text{ and } s \leq b \leq]$? Let's figure a few of them out. If $a = 1$ and $b = 2$, $F(1,2) = \frac{1}{16}$ - there is only one event involved. If $a = 1$ and $b = 3$, $F(1,3) = \frac{2}{16}$, $F(1,4) = \frac{3}{16}$,...,$F(1,7) = F(1,8) = \frac{4}{16}$. At the other extreme, $F(4,8) = 1$. If $d_1 = 2$ and $b = 3$, $F(2,3) = \frac{3}{16}$- Notice this is the sum of all of the probabilities in the lower-left rectangle defined by $d_1 \leq 2$ and $s \leq 3$. Just keep expanding out the lower rectangle. And,

$F(a,b) =$

$d_1 \backslash s$	1	2	3	4	5	6	7	8
4	0	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{6}{16}$	$\frac{10}{16}$	$\frac{13}{16}$	$\frac{15}{16}$	1
3	0	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{6}{16}$	$\frac{9}{16}$	$\frac{11}{16}$	$\frac{12}{16}$	$\frac{12}{16}$
2	0	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$	$\frac{8}{16}$	$\frac{8}{16}$	$\frac{8}{16}$
1	0	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{4}{16}$

Now specify two marginal density functions. $f_{D_1}(d_1)$. When $d_1 = 1$ this can happen with six different positive sums, so $f_{D_1}(1) = \frac{4}{16} = \sum_{s=1}^{6} f(1,s)$. $f_{D_1}(2) = \sum_{s=1}^{8} f(2,s) = \frac{4}{16}$. $f_{D_1}(3) = \sum_{s=1}^{8} f_{D_1}(3,s) = \frac{4}{16}$... So, $f_{D_1}(d_1) =$

d_1	1	2	3	4
	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

each value of d is equally likely,

just what one would expect. How about the marginal density for S, $f_S(s)$? $f_S(s) = \sum_{d_1=1}^{4} f_{D_1}(d_1, s)$. For example, $f_S(4) = \frac{3}{16}$. So $f_S(s) =$

s	1	2	3	4	5	6	7	8
	0	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Now specify two conditional density functions. Remember that

$$f(d_1 \,|s = c\,) = \frac{f(d_1, c)}{f_S(c)}$$

The conditional probability $f(d_1 \,|s = 4\,) = \frac{f(d_1,4)}{f_S(4)} = \frac{f(d_1,4)}{\frac{3}{16}} =$

d_1	1	2	3	4
	$\frac{(\frac{1}{16})}{(\frac{3}{16})}$	$\frac{(\frac{1}{16})}{(\frac{3}{16})}$	$\frac{(\frac{1}{16})}{(\frac{3}{16})}$	$\frac{0}{(\frac{3}{16})}$

$=$

d_1	1	2	3	4
	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0

Since

$$f(s \,|d_1 = c\,) = \frac{f(c, s)}{f_{D_1}(c)}$$

The conditional probability $f(s \,|d_1 = 1\,) = \frac{f(1,s)}{f_{D_1}(1)} = \frac{f(1,s)}{(\frac{4}{16})} =$

s	1	2	3	4	5	6
	$\frac{0}{(\frac{4}{16})}$	$\frac{(\frac{1}{16})}{(\frac{4}{16})}$	$\frac{(\frac{1}{16})}{(\frac{4}{16})}$	$\frac{(\frac{1}{16})}{(\frac{4}{16})}$	$\frac{(\frac{1}{16})}{(\frac{4}{16})}$	$\frac{0}{(\frac{4}{16})}$

s	1	2	3	4	5	6	7	8
	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	0	0	0

15. So, those weirdos M,G and B assert the following: if $x_1 < x_2$ and $y_1 < y_2$, then $\Pr[x_1 < X \leq x_2; y_1 < Y \leq y_2] = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geq 0$. God only know where they got this from. I don't want you to prove that this assertion is correct. Rather, assume that X and Y are discretely-distributed integers with positive density for $1 \leq X \leq 4$ and $1 \leq Y \leq 4$. Specifically assume,

$f_{XY}(x, y) =$

$X \backslash Y$	1	2	3	4
4	$\frac{5}{32}$	$\frac{3}{32}$	$\frac{1}{32}$	$\frac{2}{32}$
3	$\frac{2}{32}$	$\frac{0}{32}$	$\frac{1}{32}$	$\frac{1}{32}$
2	$\frac{1}{32}$	$\frac{3}{32}$	$\frac{2}{32}$	$\frac{0}{32}$
1	$\frac{1}{32}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{4}{32}$

Show me that MGB's assertion correctly determines $\Pr[-\infty < X \leq x_2; -\infty < Y \leq y_2]$

For my example, show me that MGB's assertion correctly determines $\Pr[1 < X \leq 3; 0 < Y \leq 4]$.

For my example, show me that the MGB's assertion correctly determines $\Pr[1 < X \leq 3; 2 < Y \leq 4]$

answer: First determine that

$F_{XY}(x, y) =$

$X \backslash Y$	1	2	3	4
4	$\frac{9}{32}$	$\frac{18}{32}$	$\frac{25}{32}$	1
3	$\frac{4}{32}$	$\frac{10}{32}$	$\frac{16}{32}$	$\frac{21}{32}$
2	$\frac{2}{32}$	$\frac{8}{32}$	$\frac{13}{32}$	$\frac{17}{32}$
1	$\frac{1}{32}$	$\frac{4}{32}$	$\frac{7}{32}$	$\frac{11}{32}$

Show me that MGB's assertion correctly determines $\Pr[-\infty < X \le x_2; -\infty < Y \le y_2]$. Answer $\Pr[-\infty < X \le x_2; -\infty < Y \le y_2] = F(x_2, y_2) - F(x_2, -\infty) - F(-\infty, y_2) + F(-\infty, -\infty) = F(x_2, y_2)$ by definition of the CDF. So, for my example. $\Pr[-\infty < X \le 3; -\infty < Y \le 2] = F(3, 2) = \frac{10}{32}$

For my example, show me that MGB's assertion correctly determines $\Pr[1 < X \le 3; 0 < Y \le 4]$. Answer $\Pr[1 < X \le 3; 0 < Y \le 4] = F(3, 4) - F(3, 0) - F(1, 4) + F(1, 0) = F(3, 4) - F(1, 4) = \frac{21}{32} - \frac{11}{32} = \frac{10}{32}$ which by inspection is the correct answer. Wow, Geometrically, notice what happened in terms of the density function. We took the area about the rectangle bounded on the bottom left by $(0, 0)$ and on the upper right by $(3, 4)$ and then subtracted from it the bottom row.

For my example show me that the MGB's assertion correctly determines $\Pr[1 < X \le 3; 2 < Y \le 4] = F(3, 4) - F(3, 2) - F(1, 4) + F(1, 2) = \frac{21}{32} - \frac{10}{32} - \frac{11}{32} + \frac{4}{32} = \frac{4}{32}$ which is correct by inspection. In terms of the density function we want the area above $(2, 3) + (2, 4) + (3, 3) + (3, 4)$ Who know why the MGB rule works? Do you?

16. Consider professors that are residents of Cambridge Mass. As you know, all professors want to be cool. Cambridge professors judge coolness in terms of how much one's beer consumption per week deviates from the average, and how much one's watching NASCAR (stock-car) racing on television per week deviates from the average. Too much deviation is not cool. Let's assume that beer drinking minus the average, B, and NASCAR watching minus the average, N, have the following bivariate uniform distribution.

$$f(b, n) = \begin{cases} 1 & if \quad -.5 \le b \le .5 \text{ and } -.5 \le n \le .5 \\ 0 & if \quad \text{otherwise} \end{cases}$$

Consider a circle with radius α and centered on no deviations. Here is the deal: if the amount of beer one drinks and the amount of NASCAR one watches put one in the circle, one is cool - otherwise, you are not a cool professor. Note that this coolness criteria applies to professors of both genders.

answer: we have to do is find the area under this density function that lies above the circle. The equation for our circle is $b^2 + n^2 = \alpha^2$.

So, the probability of being cool is

$$\begin{aligned} \Pr[cool] &= \\ & \int_{-\alpha}^{+\alpha} \int_{-\sqrt{\alpha^2 - m^2}}^{\sqrt{\alpha^2 - m^2}} 1 dk dm \\ &= 2 \int_{-\alpha}^{\alpha} \sqrt{\alpha^2 - m^2}\, dm \end{aligned}$$

Hopefully, I got that right. So, for example, if $\alpha = .25$, $2 \int_{-.25}^{.25} \sqrt{(.25)^2 - m^2}\, dm = 0.196\,35$: 19% are cool. If alternatively $\alpha = .50$, $2 \int_{-.5}^{.5} \sqrt{(.5)^2 - m^2}\, dm$:

0.785 40: 78% are cool. Why is this percentage less than one? Because we are not integrating over the rectangle bounded by $-.5$ and $+.5$, but rather the circle of radius .5 lying in this rectangle.

17. Consider professors that are residents of Cambridge Mass. As you know, all professors want to be cool. Cambridge professors judge coolness in terms of how much beer one drinks per week and how many hours of NASCAR (stock-car) racing one watches on television per week. Too little of either is not cool, neither is too much of either. Let's assume that beer drinking, B, and NASCAR watching, N, are bivarate normally distributed with means of $\mu_B > 0$ and $\mu_N > 0$, variances σ_B^2 and σ_N^2 and correlation coefficient $\rho = \frac{cov(B,N)}{\sigma_B \sigma_N}$. Consider a circle with center μ_B and μ_M and radius α. Here is the deal: if the amount of beer one drinks and the amount of NASCAR one watches are both in the circle, one is cool - otherwise, you are not a cool professor. Note that this coolness criteria applies to professors of both genders. To make this problem a bit simpler (at least for me), I am going to create and work with the random variable $\widetilde{b} = b - \mu_B$ and $\widetilde{n} = n - \mu_N$ - you should as well. Don't change notation on me.

18. Assume the X and Y are discretely distributed random variables with the joint density function.

$$f_{X,Y}(x,y) = \begin{cases} Y \backslash X & -1 & 0 & 1 \\ -1 & \frac{1}{4} & 0 & \frac{1}{4} \\ 1 & 0 & \frac{1}{2} & 0 \end{cases}$$

For this density function, show and discuss the relationships between X and Y.

answer: Note that the marginal density functions are $f_Y(y) = \begin{cases} Y & \\ -1 & \frac{1}{2} \\ 1 & \frac{1}{2} \end{cases}$ and $f_X(x) = \begin{cases} X & -1 & 0 & 1 \\ & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{cases}$. $E[Y] = E[X] = 0$. By examination we see that X and Y are not independent; for example, the probability that that $x = 0$ depends on the value of Y, which it would not if the two were independent. Let's check the covariance to see if they are correlated. $cov(X,Y) = (-1)(-1)\frac{1}{4} + (0)(-1)0 + (1)(-1)\frac{1}{4} + (-1)(1)0 + (0)(1)\frac{1}{2} + (1)(1)0 = 0$. So the relationship between X and Y is unusual: they are uncorrelated but dependent; most bivariate density functions will not have this property.

19. X and Y are discretely distributed random variables such that

$$f_{X,Y}(x,y) = \begin{cases} Y \backslash X & -1 & 0 & 1 \\ -1 & \frac{1}{4} & 0 & \frac{1}{4} \\ 1 & 0 & \frac{1}{2} & 0 \end{cases}$$

Derive the marginal density functions $f_Y(y)$ and $f_X(x)$, the expected values $E[X]$ and $E[Y]$, the variances $var(X)$ and $var(Y)$, the conditional density functions $f_Y(y\,|x\,)$ and $f_X(x\,|y\,)$, and $cov(X,Y)$.

answer: The marginal density functions are $f_Y(y) = \begin{cases} Y & \\ -1 & \frac{1}{2} \\ 1 & \frac{1}{2} \end{cases}$ and $f_X(x) = \begin{cases} X & -1 & 0 & 1 \\ & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{cases}$, both X and Y have uniform distributions. $E[Y] = E[X] = 0$.

By examination we see that X and Y are not independent; for example, the probability that that $x = 0$ depends on the value of Y (its 0 if $y = -1$ and 1 if $y = 1$), which it would not if the two were independent. Examine the conditional density functions

$$f_Y(y\,|x\,) = \begin{cases} Y & -1 & 1 & & \\ & 1 & 0 & if & x = -1 \\ & 0 & 1 & if & x = 0 \\ & 1 & 0 & if & x = 1 \end{cases}$$

and

$$f_X(x\,|y\,) = \begin{cases} X & -1 & 0 & 1 & & \\ & \frac{1}{2} & 0 & \frac{1}{2} & if & y = -1 \\ & 0 & 1 & 0 & if & y = 1 \end{cases}$$

These conditional density functions also indicate dependence: not all of the rows are the same.

Now consider the variances: $var(X) = E[(X - E(X)^2)] = \frac{1}{4}(-1-0)^2 + \frac{1}{2}(0-0)^2 + \frac{1}{4}(1-0)^2 = 0.5$

$var(Y) = E[(Y - E(Y)^2)] = \frac{1}{2}(-1-0)^2 + \frac{1}{2}(1-0)^2 = 1.0$

Now calculate $cov(X,Y)$. I will use the fact that/ $cov(X,Y) = E(XY) - E(X)E(Y)$. Since in this example $E[Y] = E[X] = 0$, $cov(X,Y) = E(XY)$. Determine that $E(XY) = (-1)(-1)\frac{1}{4} +$ $(0)(-1)0 +$ $(1)(-1)\frac{1}{4} +$ $(1)(1)0 +$ $(0)(1)\frac{1}{2} +$ $(1)(1)0 = 0.0$.

So, $cov(X,Y) = 0$. So the correlation coefficient is also zero.

20. Carolyn's wicked boss, David, and his friend Rich are interested in how residents of Colorado value the presence of mountain lions, so have done a valuation survey. They did not have the time nor money to get a random sample of Colorado residents, so collected instead a "convenience" sample - they sat in a bar and bought everyone who came through the door a drink in exchange for filling out their survey. Ten residents came through the door, and they all filled out the survey. At the end of the survey David and Rich asked each respondent whether they were a binge drinker; four said yes, five said no, and one was not sure.

In Colorado, thirty-two percent of the population identifies themselves as binge drinkers, fifty percent say they are not binge drinkers, and eighteen percent answer "don't know."

David is going to fire Carolyn unless she figures out the probability of randomly drawing their sample of 10 from the population of Colorado residents. Help her figure out the probability, and tell her how to explain to David and Rich that the number makes sense.

answer? The answer to the binge question is a disrete rv (it can take one of only three values) where $p_b = .32$, $p_{nb} = .50$ and $p_{ns} = (1 - p_b - p_{nb}) = .18$. The thing we need to determine is what is the probability that a sample of 10 Colorado residents contains 4 binge drinkers, 1 don't know and 5 non-bingers. I think it is the multinomial

$$\Pr[n_b, n_{nb}, n_{ns} \,|n] = \frac{n!}{n_b!n_{nb}!n_{ns}}(p_b)^{n_b}(p_{nb})^{n_{nb}}(1 - p_b - p_{nb})^{n_{ns}}$$

where n is the number of observations in the sample, n_b is the number of binge drinkers in the sample, etc. So

$$\Pr[4, 5, 1\,|10] = \frac{10!}{(4!)(5!)(1!)}(.32)^4(.50)^5(.18)^{(1)} = 7.431\,8 \times 10^{-2}$$

21. Consider a random variable X. Assume for simplicity that $f_X(x) > 0$ for all values of X. Denote the CDF $F_X(x)$. Assuming $x_2 > x_1$, what is

$F_X(x_2) - F_X(x_1)$?

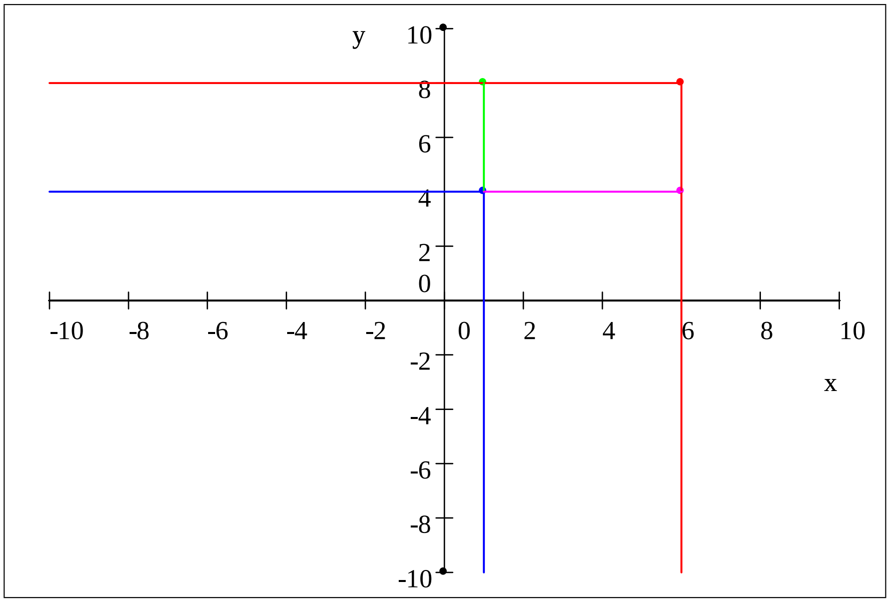
Now consider the joint CDF $F_{XY}(x, y)$ where $f_{XY}(x, y) > 0$ for all values of X and Y. Assume $x_1 < x_2$and $y_1 < y_2$. Denote the CDF $F_{XY}(x, y)$. $F_{XY}(x_2, y_2) - F_{XY}(x_1, y_1)$ is the probability of what? Explain/demonstrate.

answer: $F_X(x_2) - F_X(x_1)$ is the probability that x is between x_1 and x_2, $\Pr[x_1 < x < x_2]$

By assertion, $F_{XY}(x_2, y_2) - F_{XY}(x_1, y_1) \neq \Pr[x_1 < X < x_2, y_1 < Y < y_2,]$. MGB say (page 132)

$$\begin{aligned}
&\Pr[x_1 < X < x_2, y_1 < Y < y_2,] \\
&\quad = F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1) \\
&\quad = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{XY}(x, y)dydx
\end{aligned}$$

Convincement: Let's draw a grid, and specify some points: $(x_2, y_2) = (6, 8)$, $(x_1, y_1) = (1, 4)$, $(x_1, y_2) = (1, 8)$, and $(x_2, y_1) = (6, 4)$



$F_{XY}(6, 8)$ is the volume above that area south of the horizontal red line and west (North to the top) of the vertical red line. $F_{XY}(1, 4)$ is the volume above the area south of the horizontal blue line west of the vertical blue line. $\Pr[x_1 < X < x_2, y_1 < Y < y_2,]$ is the volume above the rectangle traced by red to the right and top, purple on the bottom and green on the left. $F_{XY}(x_2, y_2) - F_{XY}(x_1, y_1)$ is the volume above the cornered hallway between the red and blue lines, much larger than $\Pr[x_1 < X < x_2, y_1 < Y < y_2,]$.

Not part of the answer, but explaining the MGB assertion that $\Pr[x_1 < X < x_2, y_1 < Y < y_2,] = F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1)$. Start with the LHS of the MGB equality, and consider the area that we want the volume over. It is the rectangle defined by the area south and west of the red lines, minus the area west of the vertical blue-green line and minus the area south of the horizontal blue-purple line. The volume above this rectangle is $F_{XY}(x_1, y_2) + F_{XY}(x_2, y_1) - F_{XY}(x_1, y_1)$: $F_{XY}(x_1, y_2) + F_{XY}(x_2, y_1)$ is this area, but with the area to the left of the blue lines, $F_{XY}(x_1, y_1)$, counted twice, explaining the subtraction of $F_{XY}(x_1, y_1)$.