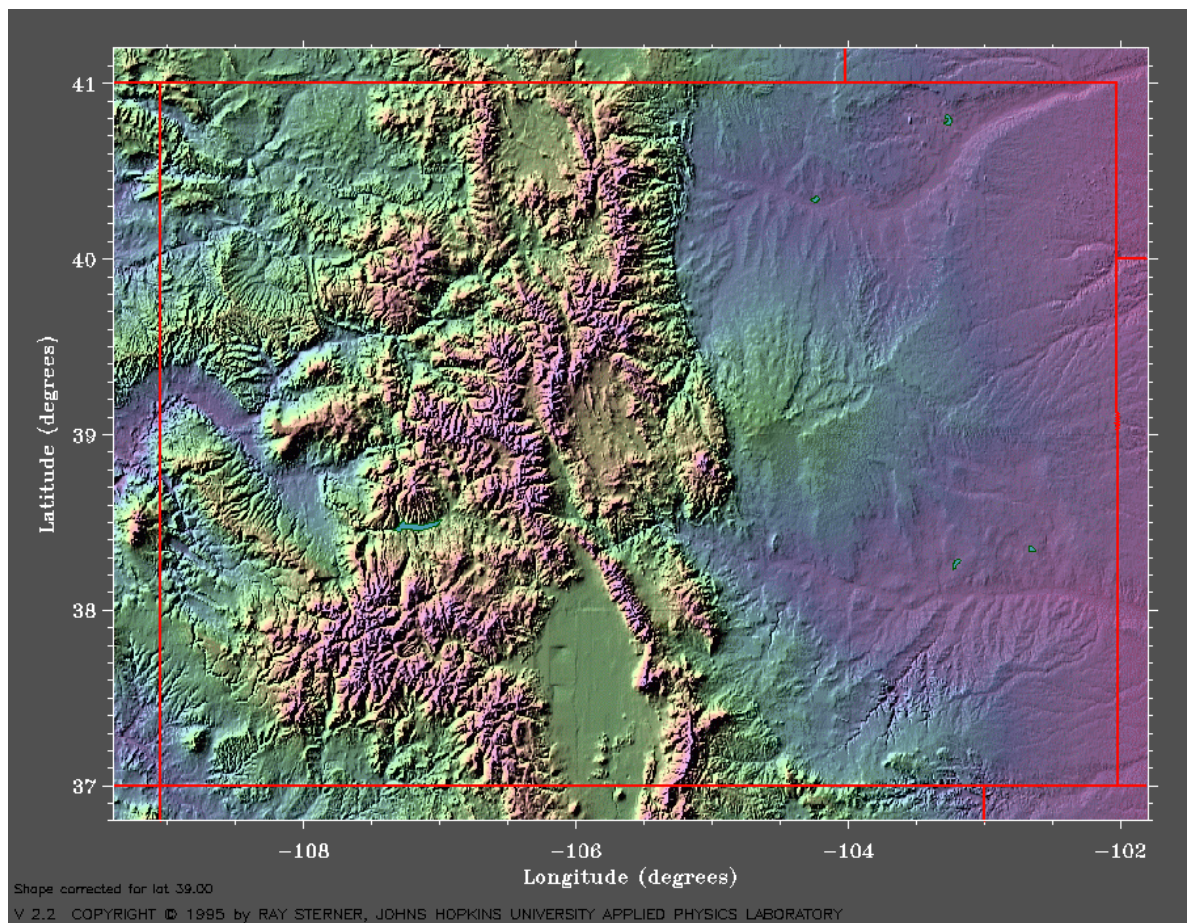


1 Joint Density Functions, Marginal Density Functions, Conditional Density Functions, Expectations and Independence

\ Oct. 19, 2010
erm

A joint density function???



1.1 Some motivation

Up to now, we have considered one random variable.

The presumption has been that there is some data generation process/random process whose outcome is a scalar random variable. For example, weight, and weight alone, is generated by height, consumption of Big Macs, and a random component.

But, a data generation process that generates one and only one output is unlikely: experiments typically have output in multiple dimensions, and we typically care about more than one of those dimensions.

You, for example, are the outcome of a random process involving your parents meeting and greeting, genetics, and environment, and you have more than one dimension.

Each of the outputs of an experiment can often be expressed as a variable realizing some specific value. Since the process generating the realization of those variables has one or more random components, they are all random variables.

So, put simply, we typically deal with a vector of random variables rather than single random variable.

One must therefore represent the distribution of vectors of random variables.

For example, the outcome of a random process might be that you ski Vail 5 times, Aspen once, and Winter Park 16 times.

1.2 Define the joint density

Assume two random variables, X and Y , and consider their joint density function

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

To be a joint density function requires $f_{XY}(x, y) \geq 0 \forall x$ and y , $-\infty < x, y < +\infty$, and

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = 1.$$

I like to describe this condition as the area under the joint density function is 1, but Anil, a former student, correctly insisted that the double integral represents a volume not an area.

Any function that fulfills these properties is a joint density function.

1.2.1 Make up a two variable joint density function and demonstrate that it is a density function.

First example: Is the following function a joint density function?

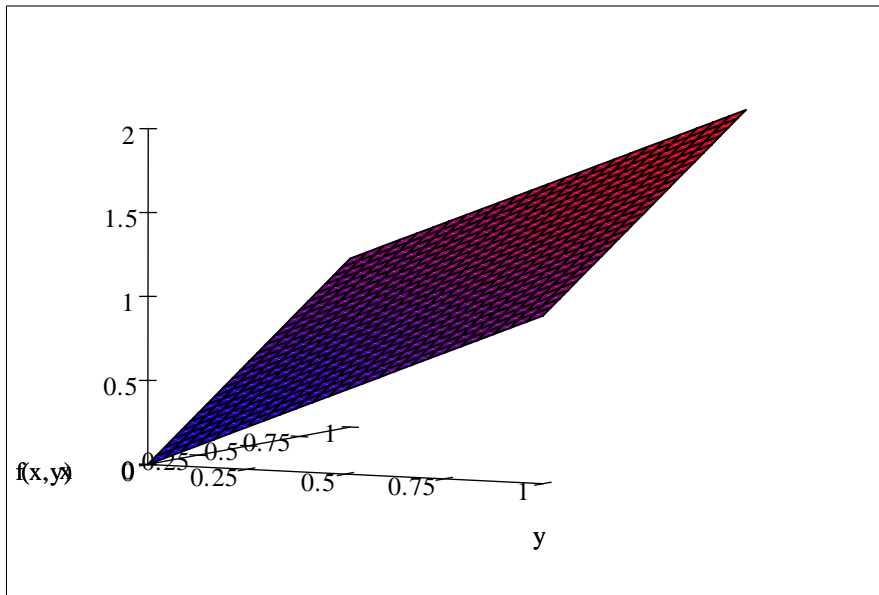
$$f_{XY}(x,y) = \begin{cases} x+y & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

That is, does the function have the following properties: $f_{XY}(x,y) \geq 0 \forall x,y$ and $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x,y) dx dy = 1$?

Yes to the first. Check the second.

$$\begin{aligned} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x,y) dx dy &= \int_0^1 \int_0^1 (x+y) dx dy \\ &= \int_0^1 [.5x^2 + xy]_0^1 dy \\ &= \int_0^1 (.5 + y) dy \\ &= .5y + .5y^2 \Big|_0^1 \\ &= .5 + .5 = 1 \end{aligned}$$

So yes, it is a joint density function. And it looks like



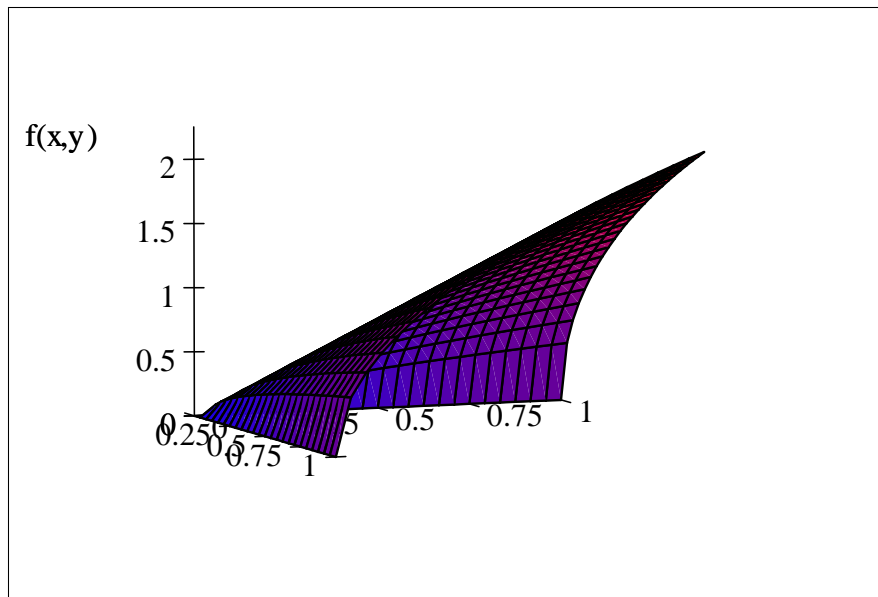
Example 2: I started by assuming $h(x, y) = x^a y^{1-a}$ if $0 \leq x \leq 1$ and $0 \leq y \leq 1$, and 0 otherwise, where $0 < a < 1$ —sort of a Cobb-Douglas density function. This function never takes negative values, so fulfills the first property of a density function. I then determined that

$$\begin{aligned}
 & \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x, y) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^a y^{1-a} dx dy \\
 &= \int_0^1 \int_0^1 x^a y^{1-a} dx dy = \int_0^1 y^{1-a} \left[\int_0^1 x^a dx \right] dy \\
 &= \frac{1}{(1+a)} \int_0^1 y^{1-a} dy = \frac{1}{(1+a)} \frac{1}{(2-a)} \\
 &= 1/(2+a-a^2) \neq 1
 \end{aligned}$$

so $h(x, y)$ is not a density function.¹

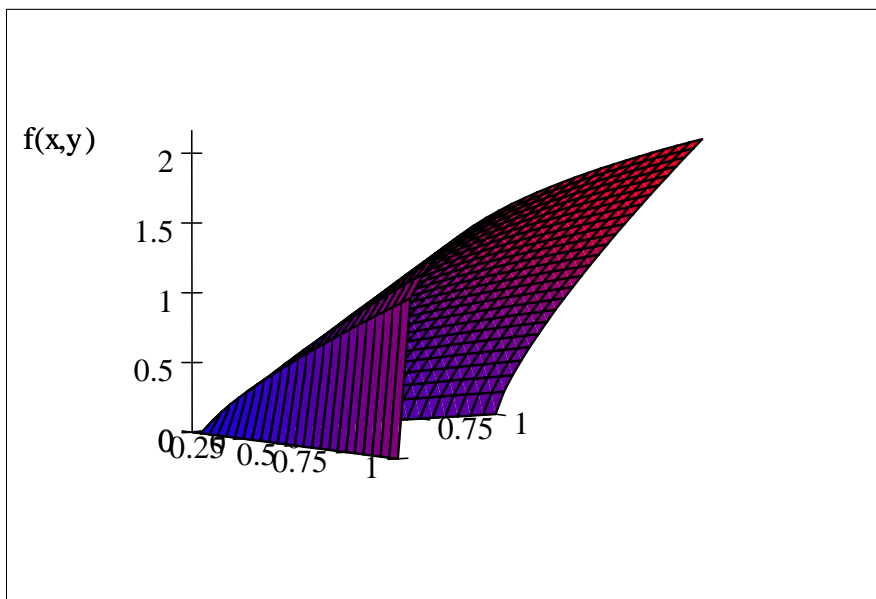
However one can easily turn it into a density function by multiplying $h(x, y)$ by $(2 + a - a^2)$ to obtain the density function $f_{XY}(x, y) = (2 + a - a^2)x^a y^{1-a}$ if $0 \leq x \leq 1$ and $0 \leq y \leq 1$, and 0 otherwise, where $0 < a < 1$.

For example, if $a = .5$, $f_{XY}(x, y) = 2.25x^{.5}y^{.5}$



And if $a = .8$, $f_{XY}(x, y) = 2.16x^{.8}y^{.2}$

¹ $1/(2+a-a^2) < 1 \forall 0 < a < 1$.



Example 3 Is the following function a joint density function?

$$f_{XY}(x, y) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

That is, does the function have the following properties: $f_{XY}(x, y) \geq 0 \forall x, y$ and $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = 1$?

Yes to the first. Check the second.

$$\begin{aligned} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy &= \int_0^1 \int_0^1 1 dx dy \\ &= \int_0^1 [x]_0^1 dy \\ &= \int_0^1 1 dy \\ &= y \Big|_0^1 \\ &= 1 \end{aligned}$$

So yes, it is a joint density function - it is the uniform density function on the unit rectangle.

1.3 Cumulative density functions

Using the basic definition of a density function

$$\begin{aligned}\Pr(x \leq b \text{ and } y \leq d) &= \Pr(-\infty \leq x \leq b \text{ and } -\infty \leq y \leq d) \\ &= \int_{-\infty}^d \int_{-\infty}^b f_{XY}(x, y) dx dy \\ &= F(b, d)\end{aligned}$$

where $F(b, d)$ is the joint cumulative density function evaluated at b and d . Note that $F(b, d)$ is defined for all b and d , $-\infty < b, d < \infty$. So, $F(x, y) = \Pr(X \leq x \text{ and } Y \leq y)$

1.3.1 An example:

Consider our first example density

$$f_{XY}(x, y) = \begin{cases} x + y & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

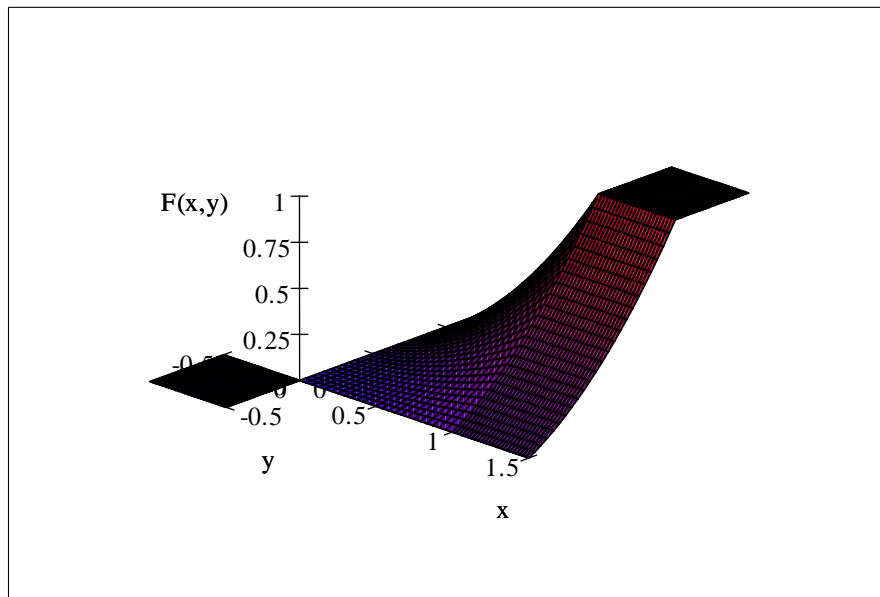
If $0 \leq b \leq 1$ and $0 \leq d \leq 1$

$$\begin{aligned} F(b, d) &= \int_0^d \int_0^b (x + y) dx dy = \int_0^d \left[.5x^2 + xy \Big|_0^b \right] dy \\ &= \int_0^d (.5b^2 + by) dy \\ &= .5b^2y + .5by^2 \Big|_0^d \\ &= .5b^2d + .5bd^2 \\ &= .5bd(b + d) \end{aligned}$$

Therefore, (this is a bit tricky)

$$F(x, y) = \begin{cases} 0 & \text{if } x < 0 \text{ or } y < 0 \\ .5xy(x + y) & \text{if } 0 \leq x < 1 \text{ and } 0 \leq y \leq 1 \\ 1 & \text{if } x > 1 \text{ and } y > 1 \\ .5x(x + 1) & \text{if } 0 \leq x \leq 1 \text{ and } y > 1 \\ .5y(1 + y) & \text{if } x > 1 \text{ and } 0 \leq y \leq 1 \end{cases}$$

Graphing this CDF, which is cool looking:



So, for example,

$$F(.5, .25) = .5(.5)(.25)(.5 + .25) = 4.6875 \times 10^{-2} = \frac{3}{64}$$

which is the probability that $x \leq .5$ and $y \leq .25$. It is also the volume under the surface $x + y$ over the region $\{(x, y) : 0 \leq x \leq .5, 0 \leq y \leq .25\}$. I think you will find this example of a joint density function on around page 139 of MGB

1.3.2 Example 2:

If

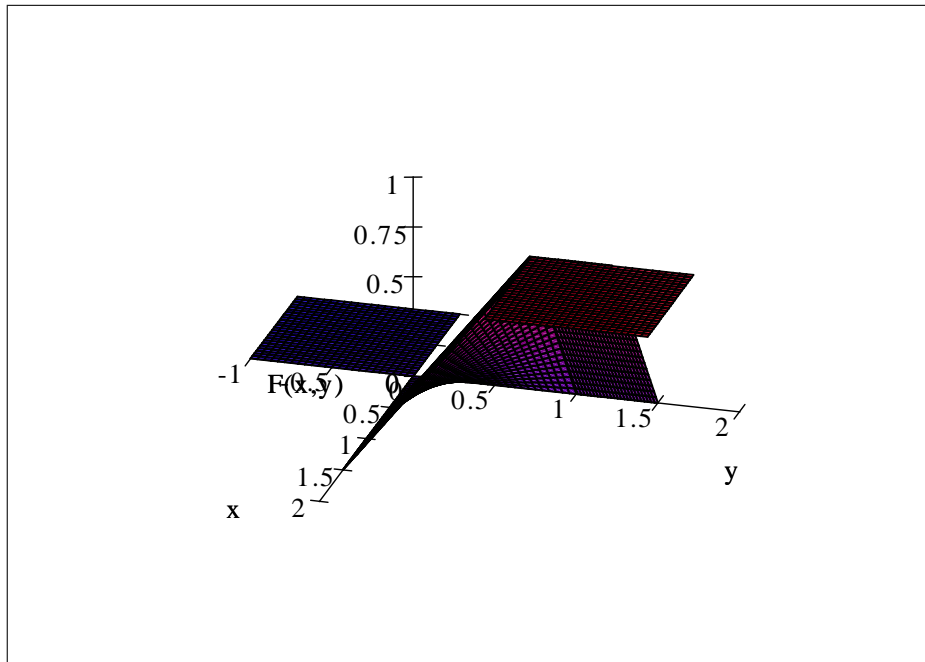
$$f_{XY}(x, y) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

In this example, if $0 \leq b \leq 1$ and $0 \leq d \leq 1$

$$\begin{aligned} F(b, d) &= \int_0^d \int_0^b 1 dx dy = \int_0^d [x]_0^b dy \\ &= \int_0^d b dy \\ &= by \Big|_0^d \\ &= bd \end{aligned}$$

Therefore

$$F(x, y) = \begin{cases} 0 & \text{if } x < 0 \text{ or } y < 0 \\ xy & \text{if } 0 \leq x < 1 \text{ and } 0 \leq y \leq 1 \\ 1 & \text{if } x > 1 \text{ and } y > 1 \\ x & \text{if } 0 \leq x \leq 1 \text{ and } y > 1 \\ y & \text{if } x > 1 \text{ and } 0 \leq y \leq 1 \end{cases}$$



So, for example,

$$F(.5, .5) = .5(.5) = .25$$

which is the probability that $x \leq .5$ and $y \leq .5$. It is also the volume under the surface $f_{XY}(x, y) = 1$ over the region $\{(x, y) : -\infty \leq x \leq .5, -\infty \leq y \leq .5\}$

Note that I did not have an example where I derived the normal CDF from the normal density function. **Why?**

Note the following

$$f_{X,Y}(x,y) = \frac{\partial F_{X,Y}(x,y)}{\partial x \partial y}$$

Can you convince that this assertion is correct?

Insert an example starting with the joint CDF

$$F(x, y) =$$

1.4 Marginal density functions

Consider the joint density function $f_{XY}(x, y)$. If

$$\Pr(a \leq x \leq b \text{ and } c \leq y \leq d) = \int_a^b \int_c^d f_{XY}(x, y) dy dx$$

then what is the probability that $a < x < b$? It is

$$\Pr(a \leq x \leq b \text{ and } -\infty \leq y \leq +\infty) = \int_a^b \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx$$

If we define $f_X(x) \equiv \int_{-\infty}^{+\infty} f_{XY}(x, y) dy$, then $\Pr(a \leq x \leq b \text{ and } -\infty \leq y \leq +\infty)$ can be rewritten as

$$\begin{aligned} \Pr(a \leq x \leq b \text{ and } -\infty \leq y \leq +\infty) \\ = \Pr(a \leq x \leq b) = \int_a^b f_X(x) dx \end{aligned}$$

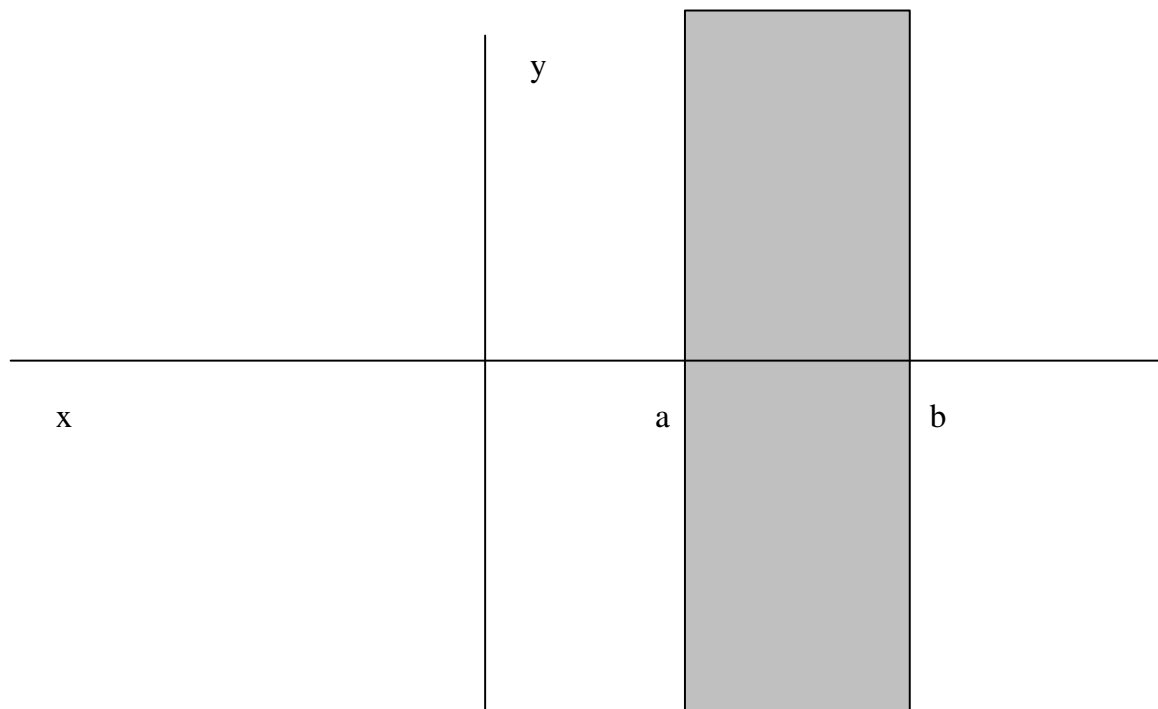
where $f_X(x)$ is called the marginal density function of the random variable X .² Note that $f_X(x)$ is not a function of y .

Are you sure $f_X(x)$ is a density function? For $f_X(x)$ to be a density function it must be the case that $f_X(x) \geq 0 \forall x$ and $\int_{-\infty}^{+\infty} f_X(x) dx = 1$. Check that it is

$$\begin{aligned} & \int_{-\infty}^{+\infty} f_X(x) dx \\ = & \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx \text{ by definition of } f_X(x) \\ = & 1 \text{ because } f_{XY}(x, y) \text{ was defined as the joint density function} \end{aligned}$$

Calling $f_X(x)$ the marginal density function emphasizes that X is jointly distributed with other variables (in our simple case, Y) Note that $f_X(x)$ is not a function of y .

²Note that one could interpret $f_X(x)$ as simply the density function for X .



1.4.1 In parallel ,

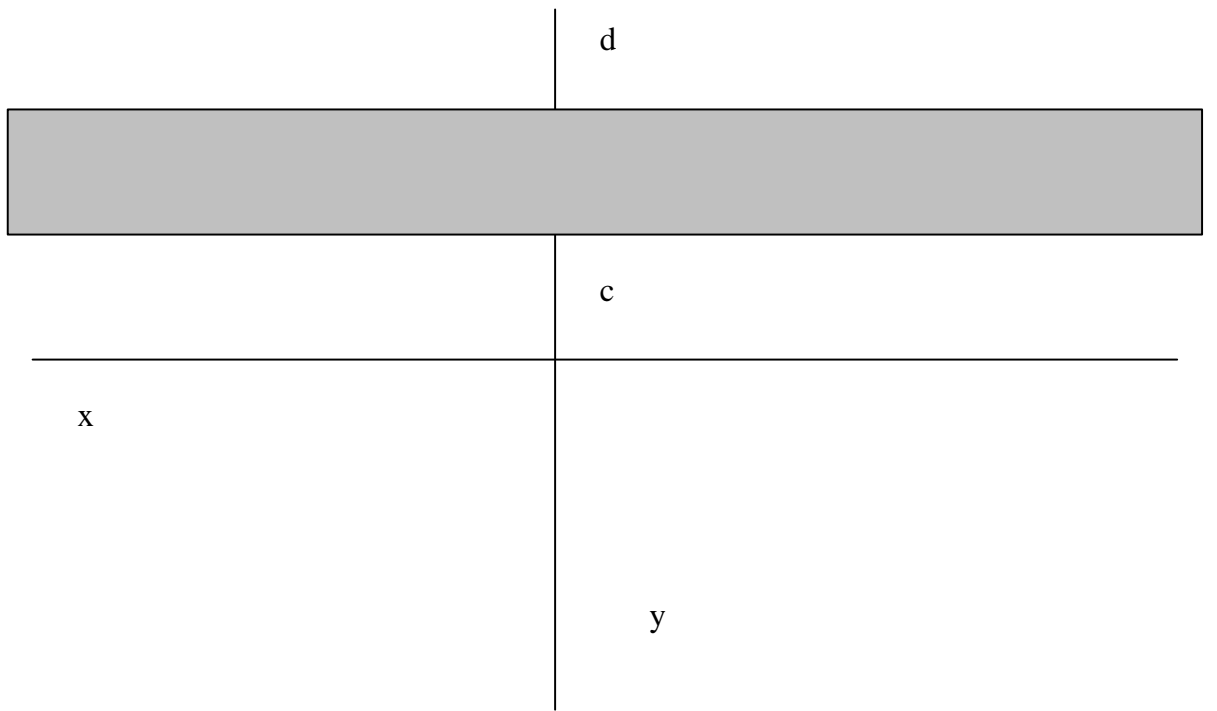
$$\begin{aligned} \Pr(c \leq y \leq d) &= \Pr(-\infty \leq x \leq +\infty \text{ and } c \leq y \leq d) \\ &= \int_c^d \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = \int_c^d f_Y(y) dy \end{aligned}$$

where $f_Y(y)$ is the marginal density function of Y . That is

$$f_Y(y) \equiv \int_{-\infty}^{+\infty} f_{XY}(x, y) dx$$

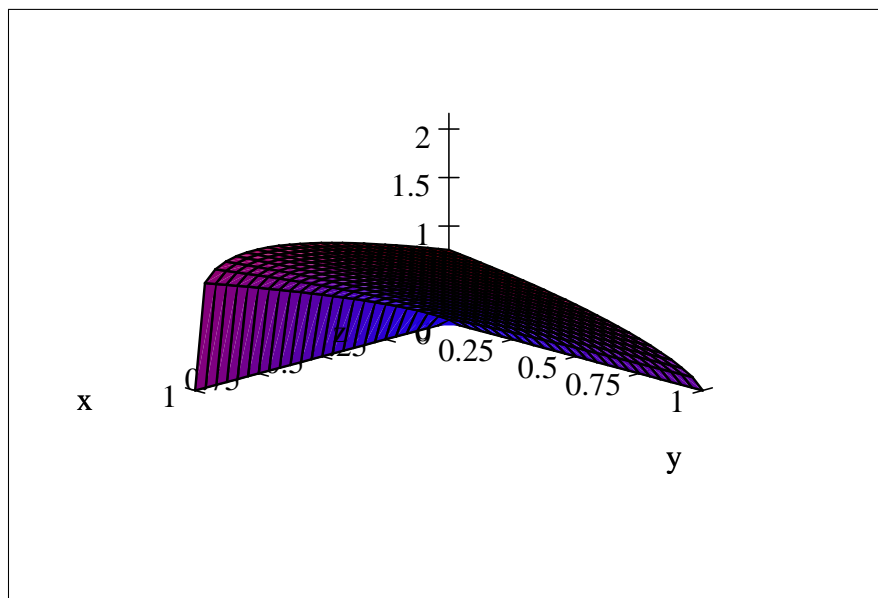
Graphically, $\Pr(a < x < b)$ is the volume under $f_{XY}(x, y)$ above the shaded area

and, $\Pr(c < y < d)$ is the volume under $f_{XY}(x, y)$ above the shaded area



1.4.2 Consider the following example: Assume $f_{XY}(x, y) = 2.16x^8y^2$,

$$0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1.$$



Imagine that X is the percent of the population richer than you (e.g. $x = .9$ means almost everyone is richer than you), and Y is percent of the population richer than your parents. (Some might have trouble with this example because the random variables are percentages.)

In which case, the marginal density function for Y is $f_Y(y) = \int_0^1 2.16x^8y^2 dx = 1.2y^{0.2}$; this is the density function for the percentage of the population that is richer than your parents.

And the marginal density function for X is $f_X(x) = \int_0^1 2.16x^8y^2 dy = 8x^{0.8}$; this is the density function for the percentage of the population that is richer than you.

Note that

$F_X(x) = F_{X,Y}(x, \infty)$ and $F_Y(y) = F_{X,Y}(\infty, y)$. Can you convince that these

two assertions are correct? How about a graphical convincing.

1.5 Now consider some conditional density functions

If X and Y are jointly distributed with density function $f_{XY}(x, y)$, what is the probability that $(a < x < b)$ given that $Y = c$? Define

$$\Pr(a < x < b | Y = c) \equiv \int_a^b f(x | Y = c) dx$$

where $f(x | Y = c)$ is the density of x given that $Y = c$. $f(x | Y = c)$ is called a conditional density function. Let's try and identify $f(x | Y = c)$ and see how it relates to the functions $f_{XY}(x, y)$ and $f(x, c)$, where $f(x, c)$ is simply the function $f(x, y)$ with y fixed at the level c .

Note the difference, in notation, between $f(x, c)$ and $f(x | Y = c) \equiv f(x | c)$

One could visualize $f(x, c)$ as a slice of $f(x, y)$ with the slice in the X direction taken at $y = c$

One might first guess that $f(x | Y = c) = f(x, c)$, but this is not the case. They are not equal because $f(x, c)$ is not a density function: the area under it is not equal to one.³ To make $f(x, c)$ a density function, one needs to scale it (adjust it) so that the area under it is equal to 1. Consider the following adjustment

$$f(x | Y = c) = \frac{f(x, c)}{f_Y(c)}$$

where $f_Y(c)$ is the marginal density of Y , evaluated at c . By definition of $f_Y(y)$,

$$f_Y(c) = \int_{-\infty}^{+\infty} f(x, c) dx$$

Then integrate $\frac{f(x, c)}{\int_{-\infty}^{+\infty} f(x, c) dx}$ with respect to x to see if the area under it is equal to one.

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x | Y = c) dx &= \int_{-\infty}^{+\infty} \frac{f(x, c)}{\int_{-\infty}^{+\infty} f(x, c) dx} dx = \frac{\int_{-\infty}^{+\infty} f(x, c) dx}{\int_{-\infty}^{+\infty} f(x, c) dx} \\ &\text{because } \int_{-\infty}^{+\infty} f(x, c) dx \text{ in the denominator is not a function of } x; \text{ it is a constant} \\ &= 1 \end{aligned}$$

³ $\int_{-\infty}^{+\infty} f(x, c) dx \neq 1$ because $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$.

so $f(x|Y=c) \equiv f(x|c) = \frac{f(x,c)}{f_y(c)}$ is a legitimate density function.

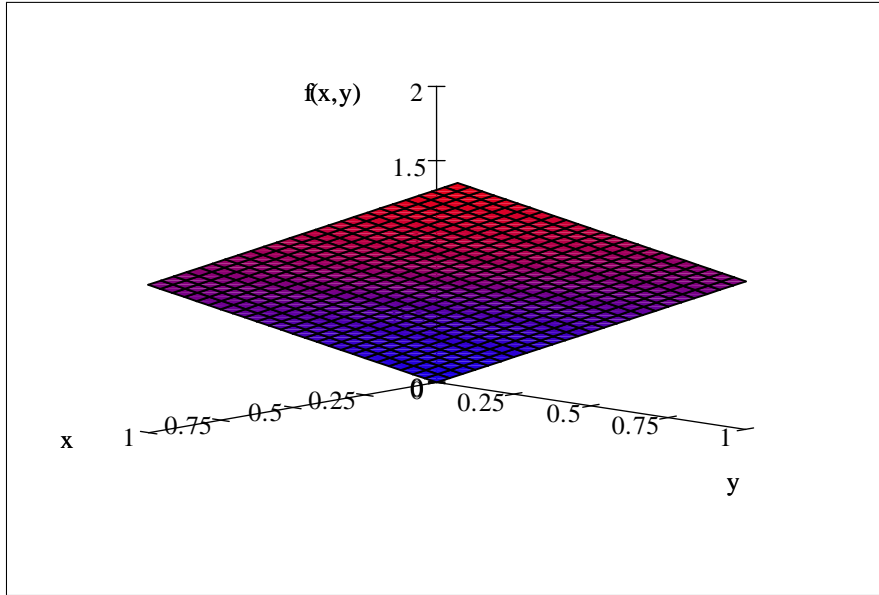
Convince yourself that it can be used to correctly estimate the $\Pr(a < x < b|Y=c)$.

$f(x|Y=c)$ is often written, more loosely, as $f(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$.

Consider the following earlier examined joint density function $f_{XY}(x, y) =$

$x + y$, $0 \leq x \leq 1$ and $0 \leq y \leq 1$. As we saw earlier, graphically, this density is rectangle with one corner resting at $(0, 0, 0)$ and the opposite corner topping out at $(1, 1, 2)$

$$x + y$$



The corresponding marginals are

$$f_Y(y) = \int_0^1 (x + y) dx = y + \frac{1}{2}$$

and

$$f_X(x) = \int_0^1 (x + y) dy = x + \frac{1}{2}$$

$$f_{XY}(x, c) = x + c$$

$$f_{XY}(c, y) = y + c$$

and the simple conditionals are

$$f(x | Y = c) = \frac{f_{XY}(x, c)}{f_Y(c)} = \frac{x + c}{c + \frac{1}{2}} = \text{and}$$

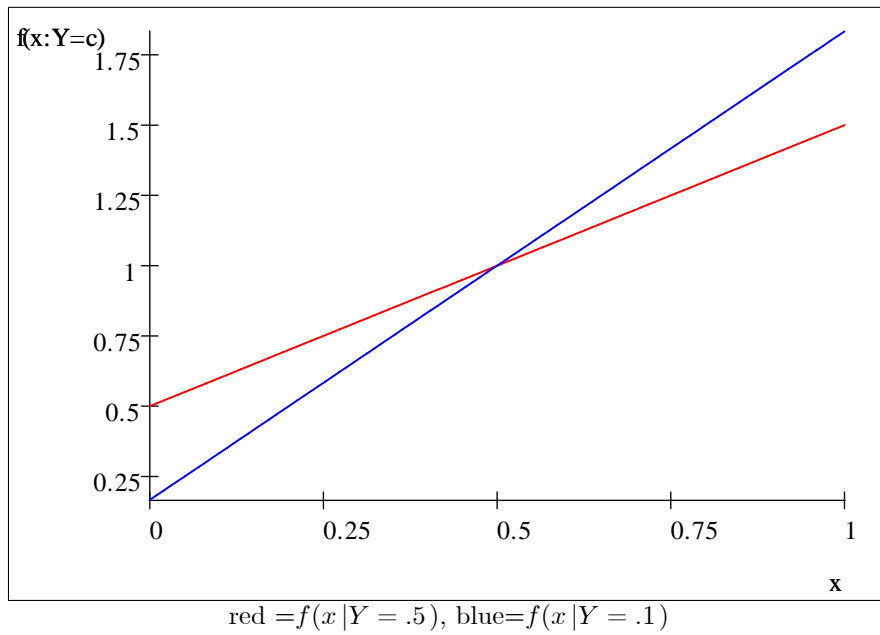
$$f(y | X = c) = \frac{f_{XY}(x, y)}{f_X(c)} = \frac{c + y}{c + \frac{1}{2}}$$

For example if $y = .5$

$$f(x | Y = .5) = \frac{x + .5}{.5 + \frac{1}{2}} = 1.0x + 0.5 \text{ and if } Y = .1, f(x | Y = .1) = \frac{x + .1}{.1 + \frac{1}{2}} = 1.$$

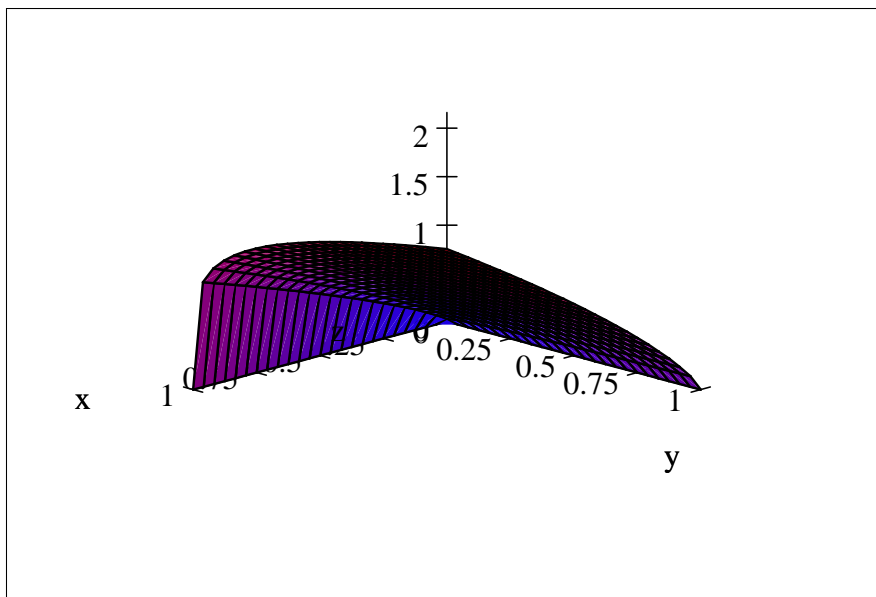
$$.6667x + 0.16667.^4$$

⁴ $\int_0^1 (1.0x + 0.5) dx = 1.0$ and
 $\int_0^1 (1.6667x + 0.16667) dx = 1.0$



Note how $f(x|Y = c) = \frac{x+c}{c+\frac{1}{2}}$ depends on c , something you likely anticipated.

Consider the another earlier example $f_{XY}(x, y) = 2.16x^8y^2$, $0 \leq x \leq 1$ and $0 \leq y \leq 1$.



where X is the percent of the population richer than you (e.g. $x = .9$ means almost everyone is richer than you), and Y is percent of the population richer than your parents. (Some might have trouble with this example because the random variables are percentages.)

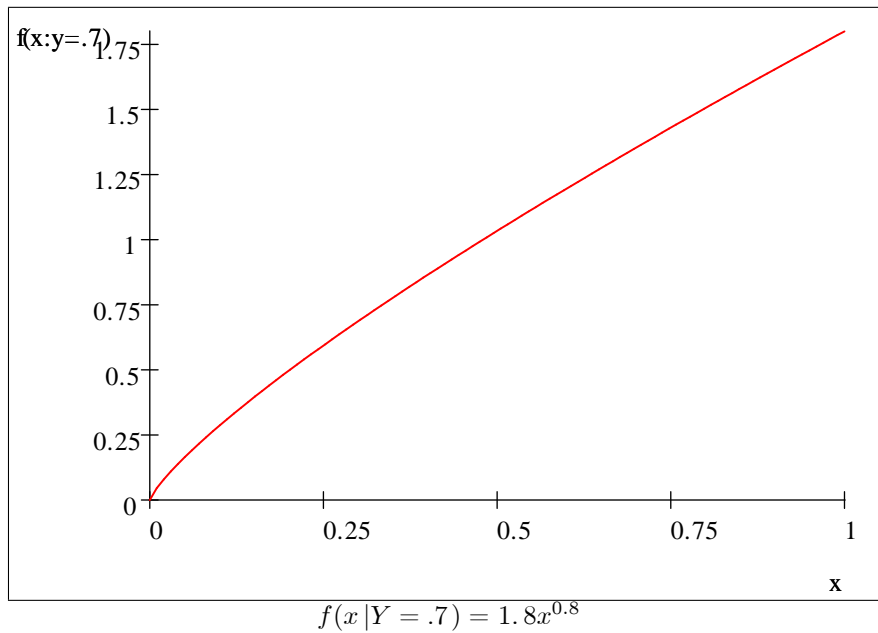
This was originally my only example of a simple conditional density function, but it was not a good example, by itself. It is informative now that I have combined it with the previous example.

So, the density function for X , the percent of the population richer than you, given that 70% of the population is richer than your parents, is

$$\begin{aligned} f(x|Y = .7) &= \frac{f(x, .7)}{f_Y(.7)} \\ &= \frac{2.16x^8(.7)^2}{1.2(.7)^{0.2}} \\ &= 1.8x^{0.8}, \quad 0 \leq x \leq 1 \end{aligned}$$

As required, $\int_0^1 1.8x^{0.8}dx = 1.5$ Graphing $f(x|Y = .7) = 1.8x^{0.8}$

⁵Note that this conditional density function takes values greater than one.



So, for this example,

$$\begin{aligned} \Pr(.25 \leq x \leq .50 | Y = .7) \\ &= \int_{.25}^{.5} 1.8x^{0.8} dx = 0.20471 \end{aligned}$$

Which says that if 70% of the population is richer than your parents, there is a 20.5% change that between 25% and 50% of the population will be richer than you.

$$\begin{aligned} \Pr(x \leq .50 | Y = .5) \\ &= \int_0^{.5} 1.8x^{0.8} dx \\ &= 0.28717 \end{aligned}$$

In words, if your parents income was in the lower half, you have a 28% change of being in the lower half.

As noted above, this second example turns out not to be a highly restrictive

example. Note that

$$\begin{aligned} f(x|Y=c) &= \frac{f(x,c)}{f_Y(c)} \\ &= \frac{2.16x^8(c)^2}{1.2(c)^{0.2}} \\ &= 1.8x^{0.8} \quad 0 \leq x \leq 1 \end{aligned}$$

does not depend on c —it cancels out. What does this mean? The conditional density $f(x|Y=c)$ does not depend on c . I accidentally chose a density function where X and Y are independent (the density of X does not depend on the level of Y), but we have yet to define independence in terms of two random variables. We will soon.

1.5.1 Let's get a bit more complicated with conditional densities.

Consider the more complicated conditional density function ; that is, the density of X conditional on $Y \leq b$. It can! be shown that⁶

$$\begin{aligned} f(x|Y \leq b) &= \frac{\int_{-\infty}^b f_{XY}(x,y)dy}{\int_{-\infty}^{+\infty} \int_{-\infty}^b f_{XY}(x,y)dydx} \\ &= \frac{\int_{-\infty}^b f_{XY}(x,y)dy}{\Pr(y \leq b)} \end{aligned}$$

The denominator, $\Pr(y \leq b)$ is the volume, which is less than one, above the shaded area.⁷

Given, $Y \leq b$, any pair-wise draw, (x, y) has to to be an element in the shaded set.

The numerator, $\int_{-\infty}^b f_{XY}(x,y)dy$, depends on x and is the area above the vertical line (bottom to top) from $(x, -\infty)$ to (x, b) . So $f(x|Y \leq b)$ is the proportion of the volume above the shaded area that is above the line from $(x, -\infty)$ to (x, b) .

⁶ Compare this to $f(x|Y=c) = \frac{f(x,c)}{\int_{-\infty}^{+\infty} f(x,c)dx}$.

⁷ Check that $\int_{-\infty}^{+\infty} f(x|y \leq b) = 1$.

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x|y \leq b) &= \int_{-\infty}^{+\infty} \frac{\int_{-\infty}^b f(x,y)dy}{\int_{-\infty}^{+\infty} \int_{-\infty}^b f(x,y)dydx} \\ &= \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^b f(x,y)dy}{\int_{-\infty}^{+\infty} \int_{-\infty}^b f(x,y)dydx} = 1 \end{aligned}$$

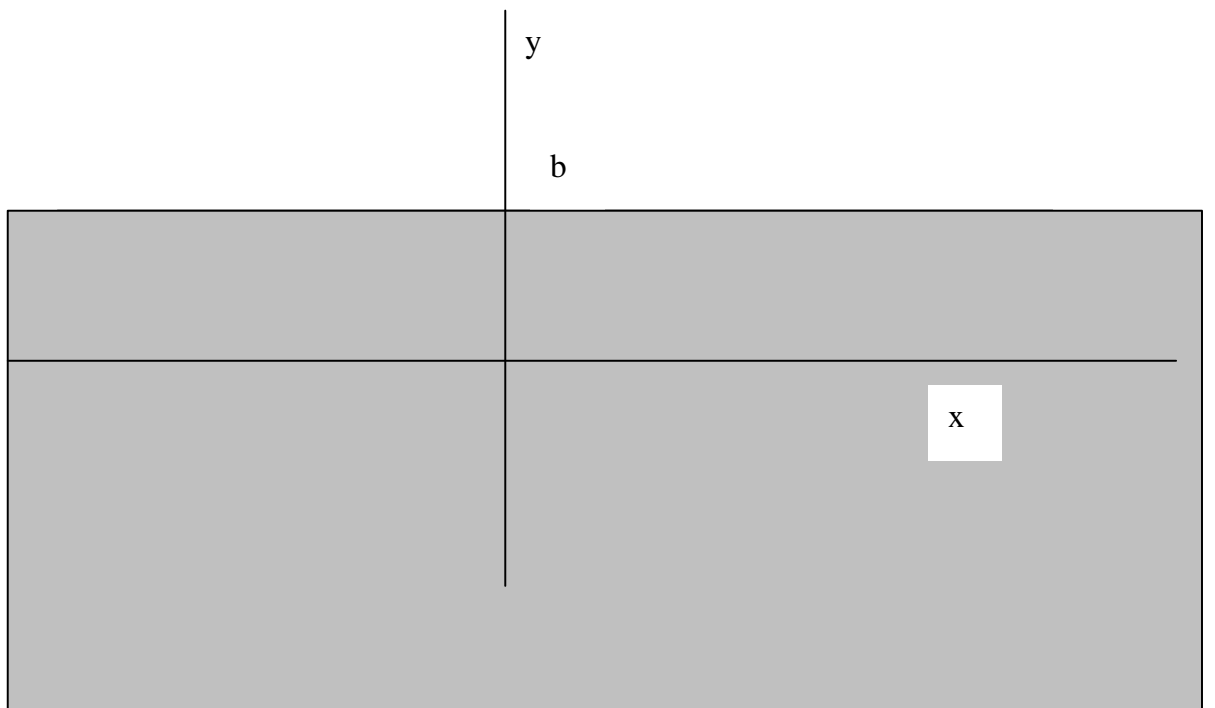


Figure 1: X is on the horizontal axis, Y is on the vertical axis

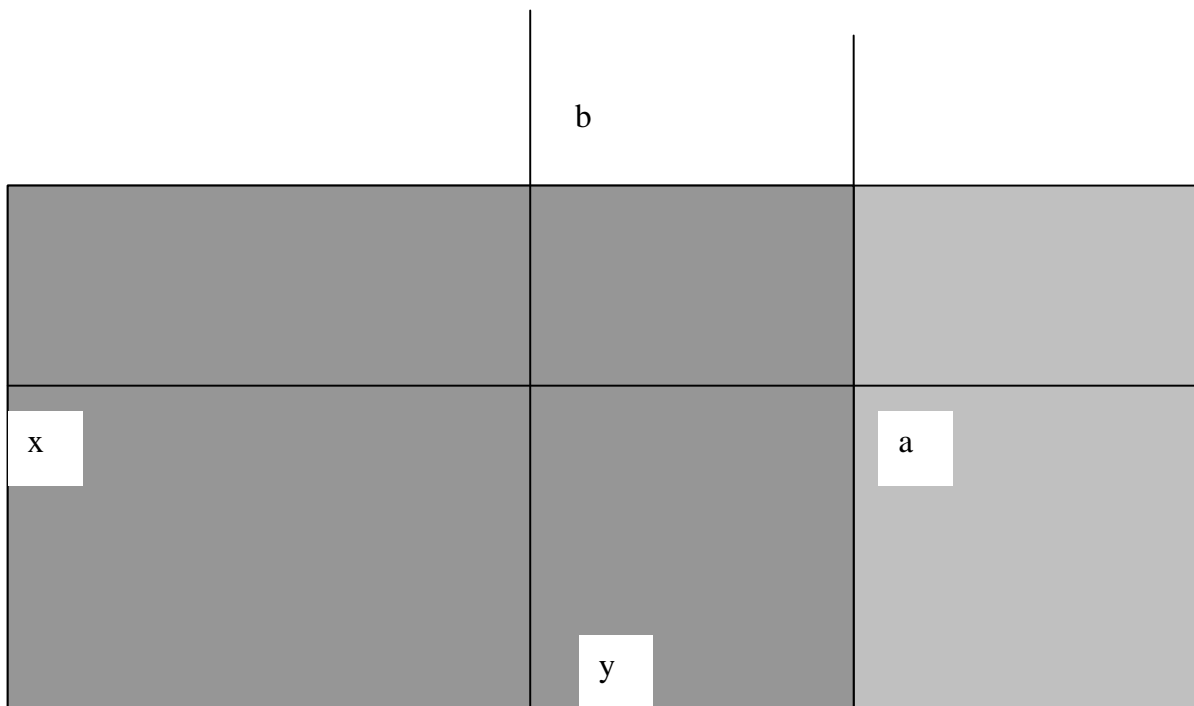


Figure 2: X is on the horizontal axis, Y is on the vertical axis

Now consider the probability that $(x \leq a | Y \leq b)$.

$$\begin{aligned} \Pr(x \leq a | Y \leq b) &= \int_{-\infty}^{+a} f(x|y \leq b) dx \\ &= \frac{\int_{-\infty}^a \int_{-\infty}^b f_{XY}(x, y) dy dx}{\int_{-\infty}^{+\infty} \int_{-\infty}^b f_{XY}(x, y) dy dx} \end{aligned}$$

The denominator is the volume above the two shaded areas and the numerator is the volume above the darker shaded region.

The dark shaded set is the set of all (x, y) pairs such that $x \leq a$ and $Y \leq b$. So, the probability is the volume above the dark shaded region as a proportion of the volume above the whole shaded region.

Consider the following example: $f(b, t) = \frac{1}{8}(6 - b - t)$, if $0 \leq b \leq 2$ and $2 \leq t \leq 4$, and zero otherwise, where b is the number of Big Macs consumed and t is the number of tequila bottles emptied.

$$\begin{aligned}
 \Pr(b \leq 1 | t \leq 3.5) &= \int_0^1 f(b | t \leq 3.5) db \\
 &= \int_0^1 \left(\frac{\int_2^{3.5} f(b, t) dt}{\int_0^2 \int_2^{3.5} f(b, t) dt db} \right) db = \frac{\int_0^1 \int_2^{3.5} f(b, t) dt db}{\int_0^2 \int_2^{3.5} f(b, t) dt db} \\
 &= \frac{\int_0^1 \int_2^{3.5} \frac{1}{8}(6 - b - t) dt db}{\int_0^2 \int_2^{3.5} \frac{1}{8}(6 - b - t) dt db} = \frac{\int_0^1 (0.60938 - 0.1875b) db}{\int_0^2 (0.60938 - 0.1875b) db} \\
 &= \frac{0.51563}{0.84376} = 0.61111
 \end{aligned}$$

If I managed to do this correctly, the probability that you will eat one Big Mac or less, given that you have emptied 3.5 bottles or less is 61%

For a possible review question, I would like you to come up with one or two informative and amusing examples of conditional probability(s) in terms of a cumulative conditional density function. Relate what you find in terms of the conditional CDF to the conditional df. **Does a group want this right now? It would be due a week from Wednesday.**

1.6 Expected values

If x and y are jointly distributed with density function $f_{XY}(x, y)$,⁸

$$E[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{XY}(x, y) dx dy$$

Begin by considering the special case $g(x, y) = x$. In which case,

$$\begin{aligned} E[X] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f_{XY}(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} x \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx \\ &= \int_{-\infty}^{+\infty} x f_X(x) dx \end{aligned}$$

and by analogy $E[Y] = \int_{-\infty}^{+\infty} y f_Y(y) dy$.

Now consider

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

So

$$\begin{aligned} &E[(x - E[X])(y - E[Y])] \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [(x - E[X])(y - E[Y])] f_{XY}(x, y) dx dy \end{aligned}$$

This number is called the covariance between the random variables x and y . $cov(X, Y)$ is a measure of the **linear** relationship between x and y . Note that

$$cov(X, Y) \leq 0$$

if $cov(x, y) = 0$, x and y are uncorrelated. If $cov(x, y) > 0$, x and y are positively correlated. If $cov(x, y) < 0$, x and y are negatively correlated.

Looking ahead, zero covariance is **not** the same thing as independence.

Can you show me that

$$cov(X, Y) = E[XY] - E[X]E[Y]$$

proof

$$\begin{aligned} cov(X, Y) &= E[(x - E[X])(y - E[Y])] \\ &= E[xy - xE[Y] - E[X]y + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

⁸ $E[g(x)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$, which we saw before, is a special case.

So, if $cov(X, Y) = 0$, $E[XY] = E[X]E[Y]$, but remember this does not imply independence of X and Y ; it implies not **linear** relationship.

$cov(X, Y)$ can take almost any value and it is not possible to tell whether the correlation between X and Y is high or low by observing the covariance; it will, for example, depend on the units in which X and Y are denoted. To "fix" this

we sometimes re-scale it so it falls in the -1 to $+1$ interval. This is accomplished by dividing it by $\sigma_x\sigma_y$ where $\sigma_x^2 \equiv E[(x - E[X])^2]$ and $\sigma_y^2 \equiv E[(y - E[Y])^2]$. That is,

$$-1 \leq \rho \equiv \frac{cov(X, Y)}{\sigma_x\sigma_y} \leq 1$$

ρ is called the correlation coefficient.

Note that $|\rho| = 1$ if and only if one of the random variables is linear function of the other. For fun, generate 100 pairs of random variables, X and

Y , and calculate the covariance and correlation between X and Y . Make sure you can explain.

1.7 Independence

If $f_{XY}(x, y) = f_x(x)f_y(y)$, then x and y are defined as being statistically independent.

That is, if the joint density function can be written as the product of the marginal density functions of its random variables, those random variables are statistically independent.

Note that the independence condition $f_{XY}(x, y) = f_x(x)f_y(y)$ can be rearranged to $f_x(x) = \frac{f_{XY}(x, y)}{f_y(y)}$ or $f_y(y) = \frac{f_{XY}(x, y)}{f_x(x)}$.

And, since $f(x|y) = \frac{f_{XY}(x, y)}{f_y(y)}$ is a conditional density function,

$$\text{independence of } x \text{ and } y \Leftrightarrow f_x(x) = f(x|y) \Leftrightarrow f_y(y) = f(y|x)$$

In words, if the random variables x and y are independent, the conditional distribution for each variable is the same as that variable's marginal distribution. If x and y are independent, the distribution of x does not depend on y and vice versa. Independence is a strong condition.⁹

It can be shown (MGB around p150) that

$$f_{XY}(x, y) = f_x(x)f_y(y) \Leftrightarrow F(x, y) = F_x(x)F_y(y)$$

So, independence is equivalent to the joint CDF being the product of the CDFs for each of its random variables.

If X and Y are independent, one can show that (MGB 160 with proof) that¹⁰

$$E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]$$

⁹Looking ahead, random samples have the property that each observation is an independent draw from the same or different distributions. In which case, the joint density for the sample, by assumption, has the independence property; it is the product of the separate density functions for each draw.

¹⁰A special case of this is $E[XY] = E[X]E[Y]$ if X and Y are independent. Earlier we showed that if $cov(X, Y) = 0$, $E[XY] = E[X]E[Y]$, so independence of X and Y is sufficient but not necessary for $cov(X, Y) = 0$. If X and Y are independent $f_{XY}(x, y) = f_X(x)f_Y(y)$ and

$$\begin{aligned} E[XY] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y)dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf_X(x)f_Y(y)dx dy \\ &= \left(\int_{-\infty}^{+\infty} xf_X(x)dx \right) \left(\int_{-\infty}^{+\infty} yf_Y(y)dy \right) \\ &= E[X]E[Y] \end{aligned}$$

Recollect that uncorrelated is not the same thing as independent. Simply put

$$X \text{ and } Y \text{ independent} \Rightarrow \text{cov}(X, Y) = 0 \iff E[XY] = E[X]E[Y]$$

but

$$\text{cov}(X, Y) = 0 \not\Rightarrow X \text{ and } Y \text{ independent}$$

Independence tells us that X and Y do not vary together in any way. $\text{cov}(X, Y) = 0$ tells us that there is no linear relationship between X and Y . MGB have an example on page 161 where X and Y are uncorrelated, but not independent. There is also an example in the review questions.

In the terminology of necessary and sufficient, X and Y independent is sufficient for $\text{cov}(X, Y) = 0$ and $\text{cov}(X, Y) = 0$ is necessary, but not sufficient for X and Y independent

Provide me with another example of two variables that are uncorrelated, but not independent.

2 Some asides

a conditional expectation:

$$E[g(X, Y) | X = x] = \int_{-\infty}^{+\infty} g(x, y) f(y|x) dy$$

The integral wrt to y of $g(x, y)$ weighted by the conditional density function $f(y|x)$.

Note the following, which are obvious when you think about them

$$E[g(X)] = E[E[g(X|Y)]]$$

$$E[g(Y)] = E[E[g(Y|X)]]$$

Repeating, an earlier footnote: Looking ahead, random samples have the property that each observation is an independent draw from the same or different distributions. In which case, the joint density for the sample is, by assumption, the product of the separate density functions for each draw.

Note that linear regression is much about conditional expectations. Consider $y_i = \alpha + \beta x_i + \varepsilon_i$ and ask what is the expected value of Y given $X = x_i$. In fact, $E[Y | X = x]$ is called the *regression curve of Y on x* .