Functions of several random variables

April 27, 2020

In this section we are given several random variables, we'll focus on 2 random variables, but the general ideas apply to more random variables. We'll also just consider the continuous case.

1 The distribution of a function of two random variables

Given X and Y continuous random variables and a function $g:\mathbb{R}^2\to\mathbb{R}$ let $Z=g(X,Y)$. Then we can compute the pdf of Z by the same general philosophy as in the last section. We begin with the CDF of Z and once we've computed it, we differentiate it to get the pdf.

$$F_Z(z)=\mathbb{P}(Z\leq z)=\mathbb{P}(g(X,Y)\leq z))$$

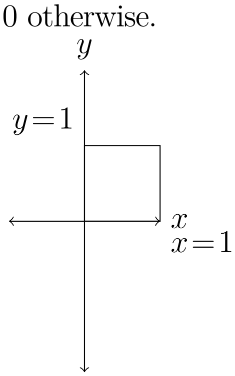
Then the final expression must be manipulated a bit to write it as an integral of some function times the joint pdf of X and Y. The step is somewhat problem dependent. With these types of problems (and really in general), if you're working with just one random variable drawing the pdf of it is usually a good idea (see the last notes). If you're working with multiple random variables, sketching the jpdf can be a bit hard, but most the useful information is in the support of the jpdf (meaning the set where it is positive). Your first step on all these types of problems is to sketch the support of the jpdf.

Let's look at an example to see how this works.

EXAMPLE 1.1. *Let X and Y each be independent random variables uniformly distributed on the interval $[0,1]$.*

Compute the pdf of $Z=Y-X$.

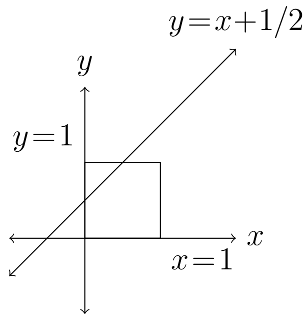
Solution: We begin by plotting the support of the jpdf of X and Y. Since the rv's are independent, the jpdf is just the product of the individual rv's. So $f_{X,Y}(x,y)=1$ if $0\leq x\leq 1$ and $0\leq y\leq 1$, and 0 otherwise.



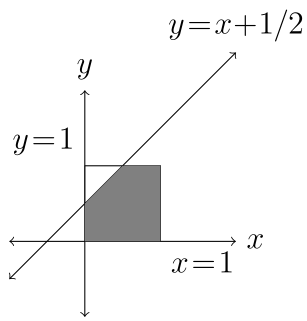
Then it is useful to determine what possible values Z can take, before we proceed. The smallest Z can be is when X is at its biggest and Y at its smallest. So at $X=1$ and $Y=0$ we'll have $Z=-1$. The biggest Z can be is when X is at its smallest and Y at its biggest. So at $X=0$ and $Y=1$ we'll have $Z=1$. So we need to compute the pdf of Z between -1 and 1. We'll actually compute the CDF and then take it's derivative.

$$F_Z(z)=\mathbb{P}(Z<z)=\mathbb{P}(Y-X<z)=\iint_{y-x<z} f_{X,Y}(x,y)\ dxdy=\iint_{y-x<z} 1\ dxdy$$

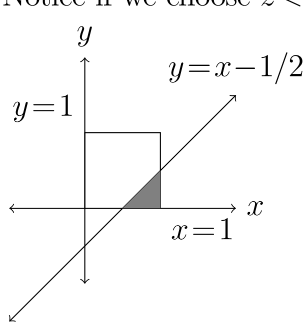
So we'll need to integrate over the region $y-x<z$, so we plot the line $y-x=z$:



where I have chosen $z=1/2$, and then we shade the region where $y-x<z$:



Notice if we choose $z<0$ (say $z=-1/2$) we get a different picture:



So the computation of the integral $\mathbb{P}(Y-X<z)=\iint_{y-x<z} 1\ dxdy$ will be different for $-1\leq z\leq 0$ and for $0\leq z\leq 1$. Since we're just integrating the constant function 1, we just need to compute the area of the shaded regions in the last picture above, for each value of z.

When $-1\leq z\leq 0$, we compute the area of the triangle in the picture above to get:

$$F_Z(z)=\frac{1}{2}(1+z)^2$$

or by computing the integral (because we can only use the area trick when the pdf is constant):

$$\begin{aligned} F_Z(z) &= \iint_{y-x<z} 1\, dxdy \\ &= \int_{-z}^{1}\int_{0}^{x+z} dydx \\ &= \int_{-z}^{1} x+z\, dx \\ &= \frac{x^2}{2}+zx\Big|_{-z}^{1} \\ &= \frac{1}{2}+z-\frac{(-z)^2}{2}-z(-z) \\ &= \frac{1}{2}(1+z)^2. \end{aligned}$$

In the other case, when $0 \leq z \leq 1$ we should compute 1 minus the unshaded region in the second picure, because the unshaded region is just a triangle instead of the pentagon. Once again we can use an area argument, because the pdf is constant to get:

$$F_Z(z) = 1-\frac{1}{2}(1-z)^2$$

or by computing the integral (we're integrating over the unshaded region)

$$\begin{aligned} F_Z(z) &= 1-\iint_{y-x\geq z} 1\, dxdy \\ &= 1-\int_{0}^{1-z}\int_{x+z}^{1} dydx \\ &= 1-\int_{0}^{z} 1-x-z\, dx \\ &= 1-\int_{0}^{z} (1-z)-x\, dx \\ &= 1-\left((1-z)x-\frac{x^2}{2}\right)_{0}^{1-z} \\ &= 1-\left((1-z)^2-\frac{(1-z)^2}{2}\right) \\ &= 1-\frac{1}{2}(1-z)^2 \end{aligned}$$

Now we differentiate the two expressions we got above to get the final answer:

$$f_Z(z) = \begin{cases} (1+z) & \text{if } -1 \leq z \leq 0 \\ (1-z) & \text{if } 0 \leq z \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Here is another example with a bit more involved function.

EXAMPLE 1.2. *Let X and Y be independent normal random variables with mean 0 and variance 1.*

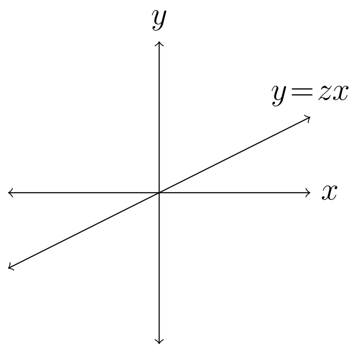
Let $Z = Y/X$, compute the pdf of Z.

Solution:

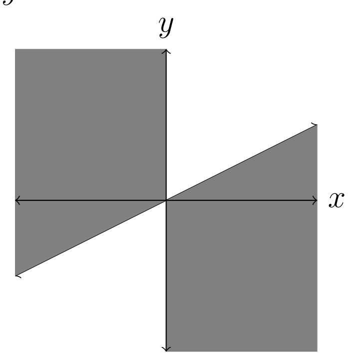
We begin with the CDF of Z:

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(Y/X \leq z)$$

To determine the region for this inequality, we first consider the the equality $Y/X = z$ or $Y = zX$, and then consider which side of this line we need to consider. We must be a little careful here, when we multiplied both sides by X, we can change the direction of the inequality so we consider the positive and negative values of X separately.



In our picture we have chosen z to be positive. Then we shade in the regions where $y/x \leq z$. When $x > 0$, we need y to be small or negative. When x is negative we need, the absolute value of y to be small or for it to be positive.



So we see

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(Y/X \leq z) = \int_{-\infty}^{0} \int_{zx}^{\infty} f_{X,Y} \; dydx + \int_{0}^{\infty} \int_{-\infty}^{zx} f_{X,Y} \; dydx$$

Note that up to this point we haven't used anything about the jpdf of X and Y. Now we can substitute it in. Because X and Y are independent their jpdf is the product of their marginal pdfs.

$$
\begin{aligned}
F_Z(z) &= \int_{-\infty}^{0}\int_{zx}^{\infty} f_{X,Y}\ dydx + \int_{0}^{\infty}\int_{-\infty}^{zx} f_{X,Y}\ dydx \\
&= \int_{-\infty}^{0}\int_{zx}^{\infty} \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}\ dydx + \int_{0}^{\infty}\int_{-\infty}^{zx} \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}\ dydx \\
&= \int_{-\tan^{-1}(z)}^{\pi/2}\int_{0}^{\infty} \frac{1}{2\pi}e^{-\frac{r^2}{2}}\ rdrd\theta + \int_{-\pi/2}^{\tan^{-1}(z)}\int_{0}^{\infty} \frac{1}{2\pi}e^{-\frac{r^2}{2}}\ rdrd\theta \\
&= \frac{1}{\pi}\tan^{-1}(z)
\end{aligned}
$$

Where we switched to polar coordinates.

so

$$f_Z(z) = \frac{d}{dz}F_Z(z) = \frac{1}{\pi}\frac{1}{z^2+1}$$

As you can see as the function g gets more complicated determining the region of integration gets more complicated.

Remark: in the two examples above I assume X and Y are independent. The situation is not much different if they are not independent. You plot the support of the joint pdf, and then integrate it over the appropriate region as above.

2 The distribution of the sum of two random variables

An important special case is when $g(x,y) = x+y$. We'll look at this case in more detail.

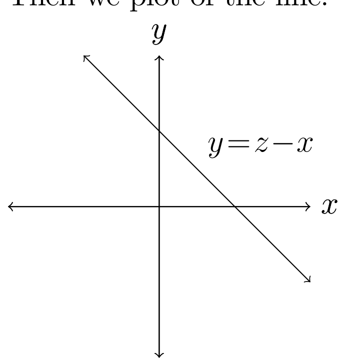
Example 2.1. *Let X and Y be independent continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$, respectively. Give an expression for the pdf of $Z = X+Y$ in terms of $f_X(x)$ and $f_Y(y)$.*

Solution: As usual we begin with the CDF of Z

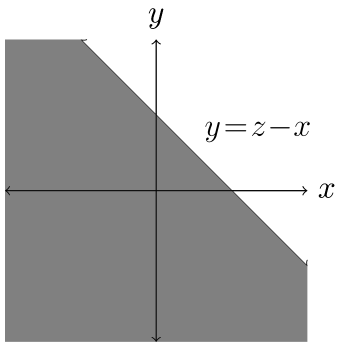
$$F_Z(z) = \mathbb{P}(Z \le z) = \mathbb{P}(X+Y \le z) = \iint_{x+y\le z} f_X(x)f_Y(y)dxdxy$$

Note on the last equality, we used the assumption that X and Y are independent, to write the jpdf as the product of the marginal pdfs.

Then we plot of the line: $x+y=z$



Then we want to integrate over the region $x+y \leq z$:



$$
\begin{aligned}
F_Z(z) &= \iint_{x+y \leq z} f_X(x) f_Y(y) dx dxy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_X(x) f_Y(y) dy dx \\
&= \int_{-\infty}^{\infty} f_X(x) \int_{-\infty}^{z-x} f_Y(y) dy dx \\
&= \int_{-\infty}^{\infty} f_X(x) F_Y(z-x) dx
\end{aligned}
$$

The last line just uses the definition of the CDF. We can now get the pdf by differentiating both sides

$$
\begin{aligned}
f_Z(Z) &= \frac{d}{dz} F_Z(z) \\
&= \frac{d}{dz} \int_{-\infty}^{\infty} f_X(x) F_Y(z-x) dx \\
&= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx
\end{aligned}
$$

So we can write the pdf of $Z=X+Y$ as a single integral of the pdfs of X and Y. This integral called the convolution of $f_X(x)$ and $f_Y(y)$ and shows up in a wide range of mathematical applications.

This expression looks relatively simple, but one always must take care with the bounds of integration. This amounts to figuring out which values of x and $z-x$ the pdfs of X and Y are positive.

The formula

$$\begin{aligned} f_Z(Z) &= \frac{d}{dz} F_Z(z) \\ &= \frac{d}{dz} \int_{-\infty}^{\infty} f_X(x) F_Y(z-x)\,dx \\ &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx \end{aligned}$$

in the last example is sort of like in the last section, where we had a shortcut formula for computing the pdf of $g(X)$ with g is a monotone function.

EXAMPLE 2.2. *Let X and Y be independent exponential random variables. Compute the pdf of $Z = X + Y$.*

Solution:

For the last example we have

$$f_Z(Z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx$$

The pdf of X and Y are both

$$f(t) = e^{-t}$$

for $t > 0$ and 0 otherwise. So the above integral will have a positive argument when $x > 0$ and $z - x > 0$. Putting these two inequalities together we see the bounds of integration will be $0 \le x \le z$.

For $z > 0$ we then have:

$$\begin{aligned} f_Z(Z) &= \int_{-0}^{z} e^{-x} e^{-(z-x)}\,dx \\ &= \int_{-0}^{z} e^{-x+x-z}\,dx \\ &= e^{-z} \int_{-0}^{z} dx \\ &= z e^{-z} \end{aligned}$$

So

$$f_Z(z) = z e^{-z}$$

for $z > 0$ and 0 otherwise. Note that this is a gamma random variable with $\alpha = 2$ and $\beta = 1$.