

# Random Variables Pt. 3

## Functions of a Random Variable

Elliot Pickens

Jul 10, 2021

### 1 Intro

Once we have the distribution of a random variable why not manipulate it? Maybe for some reason we would like five times the distribution, or it's square, or whatever else we might be able to think up for the purpose of the problem at hand. With some random variable  $X$  already in hand it seems reasonable to believe we could also find the distribution of  $5X$  or  $X^2$ . And it is, but before we can determine such distributions we have to take a dip into the world of functions of random variables.

### 2 PF/PDFs and CDFs of Functions of a Random Variable

#### 2.1 For Discrete Distributions

Starting with a discrete distribution  $X$  (with pf  $f$ ) let's define a function on all possible values of  $X$  that we'll call  $r$ . Now let's use  $r$  to define a new random variable  $Y = r(X)$  with a pf  $g$ . Since  $g$  outputs probability values for  $Y$ , which are related to the probability values of  $X$  then we should be able to define  $g$  using  $f$ . The precise way we do this is by summing the probabilities of all  $X$  value that could produce  $Y = y$ . We can write out this definition of  $g$  as

$$g(y) = P(Y = y) = P(r(X) = y) = \sum_{\{x|r(x)=y\}} f(x) \quad (1)$$

#### 2.2 For Continuous Distributions

We define the pdfs and cdfs of the distributions of functions of a random variable slightly differently in the continuous case. Instead of the standard format of the definition of a pdf of a continuous random variable that invokes the existence of some integral that outputs the probability on some interval, we instead derive the cdf of  $Y$  directly from  $f$  with

$$G(y) = P(Y \leq y) = P(r(X) \leq y) = \int_{\{x|r(x) \leq y\}} f(x) dx \quad (2)$$

from which we can derive the pdf (where  $G$  is differentiable) using

$$g(y) = \frac{dG(y)}{dy} \quad (3)$$

##### 2.2.1 Linear Functions

A linear function is a function of the generic  $y = mx + b$  sort, and happens to be quite often used in statistical models and elsewhere. In the world of random variables we can create  $Y$  as the result of a linear function  $Y = mX + b$  (where  $m \neq 0$ ). We can then find the pdf of  $Y$  to be

$$g(y) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right) \text{ for } -\infty < y < \infty \quad (4)$$

where  $f$  is the pdf of  $X$ .

### 3 Probability Integral Transforms

The probability integral transform is an operation where we use a random variable  $X$  with a continuous cdf  $F$  to create a new random variable  $Y = F(x)$ . Due to the nature of  $F$  we now have a random variable in  $Y$  that is equivalent to the uniform distribution on  $[0, 1]$ .

The transformation works, because we know that  $0 \leq F(x) \leq 1$  then  $P(Y < 0) = P(Y > 1) = 0$ . And we know that  $F(x) = y$  for some set of  $x$  values that exist on a bounded interval  $[x_0, x_1]$ . Therefore, we can say that  $Y \leq y$  if and only if  $X \leq x_1$  (where  $x_1$  is the upper bound of the interval we just mentioned). Thus, the cdf of  $Y$  is

$$G(y) = P(Y \leq y) = P(X \leq x_1) = F(x_1) = y \tag{5}$$

and the distribution of  $Y$  is the uniform distribution on  $[0, 1]$ . To quickly restate, since the cdf  $G(y) = y$  for  $0 < y < 1$  is also the cdf of the uniform distribution we can draw the equivalence.

Transforming a distribution to a uniform one is not all that interesting on its own, but with just a another short step we can then convert this new uniform distribution into any sort of our choosing. To do this we just need another continuous cdf, but first we need to introduce a corollary. Assume  $Y$  represents a uniform distribution on  $[0, 1]$  and  $F$  is a continuous cdf with the quantile function  $F^{-1}$ . Then  $F$  is the cdf of  $X = F^{-1}(Y)$ .

We can put this corollary into action in order to transform  $Y$  into a third variable  $Z$  by noticing that for some continuous cdf  $G$  we can define  $Z$  as  $G^{-1}(Y)$ . Then by our corollary we get that  $G$  is the cdf of  $Z = G^{-1}(F(X))$ .

#### 3.1 Pseudo-Random Numbers

Although I won't go into much detail about it on this post, it should be noted that the probability integral transform can be quite useful when we need to generate pseudo-random numbers. Due to our newfound ability to convert a uniform distribution into whatever distribution we can think of (with a continuous cdf), so long as we can generate uniform pseudo-random numbers we can also generate pseudo random numbers from your desired distribution. To formalize this a bit we can say that if we have a random variable  $X$  with a uniform distribution on  $[0, 1]$  and another random variable  $Y = G^{-1}(X)$  has a continuous cdf  $G$  with quantile  $G^{-1}$ . Then we can use  $X$  to produce a sequence of independent values  $x_1, \dots, x_n$ , and then transform them to  $y_1, \dots, y_n$  using  $G^{-1}$  while maintaining its random sample-ness.

### 4 Functions of Multiple Random Variables

When we have more than a single random variable things proceed much the same way they do when we work with multivariate distributions. That is to say that the definitions are altered to fit the new conditions, but the intuition remains the same. Here in this section we'll retool the ideas presented in section 2 to make things gel with as many variables as we might need.

#### 4.1 For Discrete *Joint* Distributions

Let's say that we have a group of random variables  $X_1, \dots, X_n$  that hold a discrete joint distribution with pf  $f$ . Then we can define a function of this set of random variables as  $Y = r(X_1, \dots, X_n)$ . Now using that form let's define  $m$  functions of our set of random variables as

$$Y_1 = r_1(X_1, \dots, X_n) \tag{6}$$

$$Y_2 = r_2(X_1, \dots, X_n) \tag{7}$$

$$\vdots \tag{8}$$

$$Y_m = r_m(X_1, \dots, X_n) \tag{9}$$

To define a joint pdf of these functions we take some number of values  $y_1, \dots, y_m$  produced by the random variables  $Y_1, \dots, Y_m$ , as well as the set of all points  $(x_1, \dots, x_n) \in A$  that relate to the  $y$  values via the relationship

$$r_1(x_1, \dots, x_n) = y_1 \tag{10}$$

$$r_2(x_1, \dots, x_n) = y_2 \tag{11}$$

$$\vdots \tag{12}$$

$$r_m(x_1, \dots, x_n) = y_m \tag{13}$$

Thus, by following the process we used in 1 we can solve for the probability  $g(y_1, \dots, y_m)$  by summing over the probabilities of all points  $(x_1, \dots, x_n)$  that map to our  $y$  values through  $r$ . This sum will then take the form

$$g(y_1, \dots, y_m) = \sum_{(x_1, \dots, x_n) \in A} f(x_1, \dots, x_n) \tag{14}$$

#### 4.1.1 Relation to Binomial and Bernoulli Distributions

We often think of binomial distributions as being composed of a sequence of Bernoulli trials. Therefore we should be able to relate a sequence of i.i.d. random variables  $X_1, \dots, X_n$  following a Bernoulli distribution (with parameter  $p$ ) to a random variable  $Y$  that follows a binomial distribution (that has parameters  $p$  and  $n$ ) with  $Y = X_1 + \dots + X_n$ .

Beyond simple intuition, we can show that this relationship holds weight starting with the assertion that  $Y = y$  if and only if we have exactly  $y$  of our Bernoulli random variables equaling 1 and the remaining  $n - y$  equaling 0. We also know that the vector  $(X_1, \dots, X_n)$  has  $\binom{n}{y}$  possible values that have our desired number of zeroes and ones. All of these vector configurations have a probability of occurring that is tied to the parameter  $p$ , since each element of the each vector has a probability  $p$  of being a one and probability  $1 - p$  of being a zero. Therefore, the probability of each individual vector is  $p^y(1 - p)^{n-y}$ , and the probability that  $Y = y$  is the sum of all these vector probabilities, which equates to  $\binom{n}{y}p^y(1 - p)^{n-y}$ . This result shows our proposed relationship is correct, because it is the pdf of the binomial distribution.

## 4.2 For Continuous *Joint* Distributions

Functions of continuous joint distributions follow the same behavioral path, but with the standard "we're now working with the reals" twist. Put directly: it's once again time to integrate. To set things up let's say that we have a set of random variables  $\mathbf{X} = (X_1, \dots, X_n)$  that have a joint pdf  $f$ , and that  $\mathbf{Y} = r(\mathbf{X})$ . We will also define a set  $A_y = \{\mathbf{x} | r(\mathbf{x}) \leq y\}$  for each value  $y$  (which is very similar to the  $A$  in 14). Then we define the cdf of  $Y$  to be

$$G(y) = \int_{A_y} \dots \int f(\mathbf{x}) d\mathbf{x} \tag{15}$$

We get this definition from the equality

$$G(y) = P(Y \leq y) = P(r(\mathbf{X}) \leq y) = P(\mathbf{X} \in A_y) \tag{16}$$

whose final term  $P(\mathbf{X} \in A_y)$  is equivalent to 15.

We can also find the pdf  $Y$  by taking the derivative of  $G$ , so long as  $Y$  is continuous.

#### 4.2.1 Bivariate Linear Functions

In 2.2.1 we showed how we can find the pdf of a linear function of a single variable. Here we'll expand upon such univariate functions to develop cdfs and pdfs for linear functions of two variables. Although this won't quite reach the point of having a general form for linear functions of an arbitrary number of random variables, we'll still make good progress towards that here.

Suppose we have the joint pdf  $f$  of two random variables  $X_1$  &  $X_2$ , and let  $Y = m_1X_1 + m_2X_2 + b$  (where  $m_1 \neq 0$ ). Then the pdf  $g$  of  $Y$  (which will have a continuous distribution) will be

$$g(y) = \int_{-\infty}^{\infty} f\left(\frac{y-b-m_2x_2}{m_1}, x_2\right) \frac{1}{|m_1|} dx_2 \quad (17)$$

To prove 17 we will use essentially the same process used to show 15. We begin by creating our  $A_y = \{(x_1, x_2) | m_1x_1 + m_2x_2 + b \leq y\}$ , and then setting up our integral for  $G$  (while assuming  $m_1 > 0$ )

$$G(y) = \int_{A_y} \int f(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{(y-b-m_2x_2)/m_1} f(x_1, x_2) dx_1 dx_2 \quad (18)$$

To proceed we need to modify the inner integral by carrying out a change of variables. The change we'll use is  $z = m_1x_1 + m_2x_2 + b$ , which becomes  $x_1 = \frac{z - m_2x_2 - b}{m_1}$  with  $dx_1 = \frac{dz}{m_1}$ . Plugging this in causes the inner integral to become

$$\int_{-\infty}^y f\left(\frac{z-b-m_2x_2}{m_1}, x_2\right) \frac{1}{m_1} dz \quad (19)$$

that we can insert into 18 to create

$$G(y) = \int_{-\infty}^{\infty} \int_{-\infty}^y f\left(\frac{z-b-m_2x_2}{m_1}, x_2\right) \frac{1}{m_1} dz dx_2 \quad (20)$$

$$= \int_{-\infty}^y \int_{-\infty}^{\infty} f\left(\frac{z-b-m_2x_2}{m_1}, x_2\right) \frac{1}{m_1} dx_2 dz \quad (21)$$

Then by substituting the  $g(z)$  for the inner integral we get  $G(y) = \int_{-\infty}^y g(z) dz$ . Therefore, since the derivative of  $G(y)$  is  $g(y)$  we have our proof, because  $g(y)$  is equal to 17.

## 4.2.2 Convolutions

A convolution is a special case of the theorem we expressed in 17 where  $X_1$  and  $X_2$  are independent,  $m_1 = m_2 = 1$ , and  $b = 0$ . These conditions leave us with the arrangement  $Y = X_1 + X_2$ , where the distribution of  $Y$  is called the convolution of  $X_1$  and  $X_2$ . Similarly, we can call the pdf of  $Y$  the convolution of the pdfs of  $X_1$  and  $X_2$ .

Then if we let the pdfs of  $X_1$  and  $X_2$  be  $f_1$  and  $f_2$  respectively, we can use 17 to find that the pdf of  $Y = X_1 + X_2$  is

$$g(y) = \int_{-\infty}^{\infty} f(y - x_2, x_2) dx_2 \quad (22)$$

$$= \int_{-\infty}^{\infty} f_1(y - x_2) f_2(x_2) dx_2 \quad (23)$$

or if we flip the variables we could get<sup>1</sup>

$$g(y) = \int_{-\infty}^{\infty} f_1(x_1) f_2(y - x_1) dx_1 \quad (24)$$

While I'm going to end this brief aside on convolutions here, it should be noted that the convolution operation is a surprising powerful one with applications in probability and beyond.

---

<sup>1</sup>I typed up these last few convolution equations fairly late at night and have a feeling I made a mistake, but I'm not currently seeing it. If anyone happens to read this and notice something's off please let me know :)

## 5 Summary

In this post we covered a number of important topics related to functions of random variables. At this point we should have a solid enough understanding of the basics, but there are still a fair few details I left out and didn't touch on at all. The most glaring of these are the direct transformation/derivations of the pdfs of functions of a random variable. I might dive deeper into this at some point in the future, but if I do I think it'll be part of a slightly different series of posts centered on the details I've been skirting past. There are also a few minor things I could've spent a bit of time describing, but didn't find warranted a place in this post (mostly special transformations like linear transformations). Those holes aside, I hope this post was informative enough for what it is.

My next post will be the final installment in this little random variables sub-series. It'll focus on Markov chains and should work as an ultra abridged introduction to all of the great topics that branch off from them including stochastic processes and interesting statistical models like HMMs (hidden Markov models).

## 6 Acknowledgments

These notes were based on *Probability and Statistics (Fourth Edition)* by DeGroot & Schervish.