

All of our examples so far have involved functions, g of x , that are monotonic in X , at least over the range of x 's of interest. Let us now look at an example that involves a non-monotonic function. We're going to consider the square function, which has the shape shown in this diagram.

And we'll assume that X has a general distribution so that it can take both positive and negative values. And so over the range of values of X , the function that we're dealing with is decreasing and then increasing. So it is not monotonic. How can we find the distribution of Y ?

As a warm-up, let's look at the discrete case. And as an example of the calculation, let us find the formula for the probability that the random variable Y takes a value of 9. This event can happen in two ways.

It can happen if X is equal to 3. But it can also happen if x is equal to negative 3. And these are the two and only two ways that y can take a value of 9. We can generalize this calculation. The probability that the random variable y takes on a specific value little y , this probability can be found by adding the probabilities of all of the different x 's that lead to this particular value.

Now, for X squared to be equal to little y , we need to have X to be equal either to the positive square root of y or to be equal to the negative square root of y . And this is the general formula for the PMF of the random variable Y . And it involves two terms, because any given value of little y can happen in two ways, either by having X equal to the negative square root of y or by having X be equal to the positive square root of y .

We have here a situation where the function that we're dealing with is not invertible. For a given value of y , we cannot find a single value of x that will lead to that y . But instead, we typically have two values of x that lead to that particular y , namely the positive and the negative square roots of y . So we cannot use the tools that we used before in the monotonic case, where we dealt with the inverse function.

What we can do instead is to proceed from first principles and calculate the CDF of the random variable Y . The CDF of Y is the probability that the random variable is less than or equal to a certain number. And we're going to focus only on the case where that certain number is non-negative.

If little y is negative, then we know that this probability is going to be 0, because the random variable Y

cannot take negative values. Now, this is the probability that the random variable X squared is less than or equal to y . So what we did here is to express this event in terms of the original random variable, capital X , whose PDF is presumably available.

Now, this event is the same as requiring the absolute value of X to be less than or equal to the square root of y . And this event, again, is the same as having the random variable X be between the negative and the positive square root of y . In terms of a picture, the random variable capital Y takes a value less than or equal to this particular little y if and only if the random variable x falls inside this range.

Now, we want to express this probability in terms of the CDF of X . The probability that we're looking at, the probability of this interval, is equal to the probability that x is less than or equal to the square root of y . This is the probability from minus infinity up to the square root of y .

But from this, we need to subtract the probability of this interval. And that would be the CDF of the random variable x up to the point [negative] square root of y . So we now have an expression for the CDF of Y in terms of the CDF of X . At this point, now we can take derivatives and use the chain rule.

The PDF of Y is going to be equal to the derivative of this expression. The derivative of the first term, by the chain rule, is the PDF of X , evaluated at the square root of y times the derivative of this argument with respect to y , which is 1 over 2 square root of y . And then we need the derivative of the second term.

We have a minus sign, then the derivative of the CDF which is the PDF. Evaluate it at minus square root of y . And then the derivative of this term with respect to y , which is minus 1 over two square root of y . Now, we have a minus sign here and a minus sign there. So the two cancel out.

We can get rid of this minus 1 term and change this minus into a plus. And this is the final form of the answer. So we see that the PDF of Y evaluated at a particular point, which tells us something about the probability that the random variable takes values around this point, has to do with the probabilities that the random variable X takes values around here or around there. There are two contributions, and this is because there are two different ways that a value of y may occur, either X falling here or X falling there.