

The Transformation Technique



By Matt Van Wyhe, Tim Larsen, and Yongho Choi

If one is trying to find the distribution of a function (statistic) of a given random variable X with known distribution, it is often helpful to use the $Y=g(X)$ transformation. This transformation can be used to find the distributions of many statistics of interest, such as finding the distribution of a sample variance or the distribution of a sample mean. The following notes will conceptually explain how this is done for both discrete random variables and continuous random variables, both in the univariate and multivariate cases. Because these processes can be almost impossible to implement, the third section will show a technique for estimating the distributions of statistics.

1. Univariate Random Variables

Given X , we can define Y as the random variable generated by transforming X by some function $g(\cdot)$ —i.e. $Y=g(X)$. The distribution of $Y=g(X)$ will then depend both on the known distribution of X as well as the transformation itself.

1.1 Discrete Univariate Random Variables

In the discrete case, for a random variable X that can take on values $x_1, x_2, x_3, \dots, x_n$, with probabilities $\Pr(x_1), \Pr(x_2), \Pr(x_3), \dots, \Pr(x_n)$, then the possible values for Y are found by simply plugging $x_1, x_2, x_3, \dots, x_n$ into the transformation $Y=g(X)$. These values need not be unique for each x_i , several values of X may yield identical values of Y . Think for example of $g(X) = X^2$, $g(X) = |X|$, $g(X) = \max\{X, 10\}$, etc., each of which have multiple values of x that could generate the same value for y .

The density function of Y is then the sum of x 's that yield the same values for y :

$$\Pr(y_j) = \sum_{i:g(x_i)=y_j} \Pr(x_i)$$

Or in other terms:

$$f_Y(y_j) = \sum_{i:g(x_i)=y_j} f_X(x_i)$$

And the cumulative distribution function:

$$F_Y(y) = Pr(Y \leq y) = \sum_{i:g(x) \leq y_j} Pr(x_i)$$

Example 1-1

Let X have possible values of -2, 1, 2, 3, 6 each with a probability of .20. Define the function:

$$Y = g(X) = (X - \bar{X})^2,$$

the squared deviation from the mean for each x_j .

To find the distribution of $Y=g(X)$, first calculate $\bar{X} = 2$ for these x_j . With a mean of 2, $(X - \bar{X})^2$ can take on values of 16, 1, and 0 with $Pr(16) = .4$, $Pr(1) = .4$, and $Pr(0) = .2$, the sums of the probabilities of the corresponding x_j 's. This is the density function for $Y=g(X)$. Formally:

$$f_Y(y) = \begin{cases} 0.2 & \text{for } y = 2 \\ 0.4 & \text{for } y = 1 \text{ or } y = 16 \\ 0 & \text{otherwise} \end{cases}$$

1.2 Continuous Univariate Random Variables

To derive the density function for a continuous random variable, use the following theorem from MGB:

Theorem 11 (p. 200 in MGB)

Suppose X is a continuous random variable with probability density function $f_x(\cdot)$, and $\mathcal{X} = \{x : f_x(x) > 0\}$ the set of possible values of $f_x(x)$. Assume that:

1. $y = g(x)$ defines a one-to-one transformation of \mathcal{X} onto its domain \mathcal{D} .
2. The derivative of $x = g^{-1}(y)$ is continuous and nonzero for $y \in \mathcal{D}$, where $g^{-1}(y)$ is the inverse function of $g(x)$; that is, $g^{-1}(y)$ is that x for which $g(x) = y$.

Then $Y = g(X)$ is a continuous random variable with density:

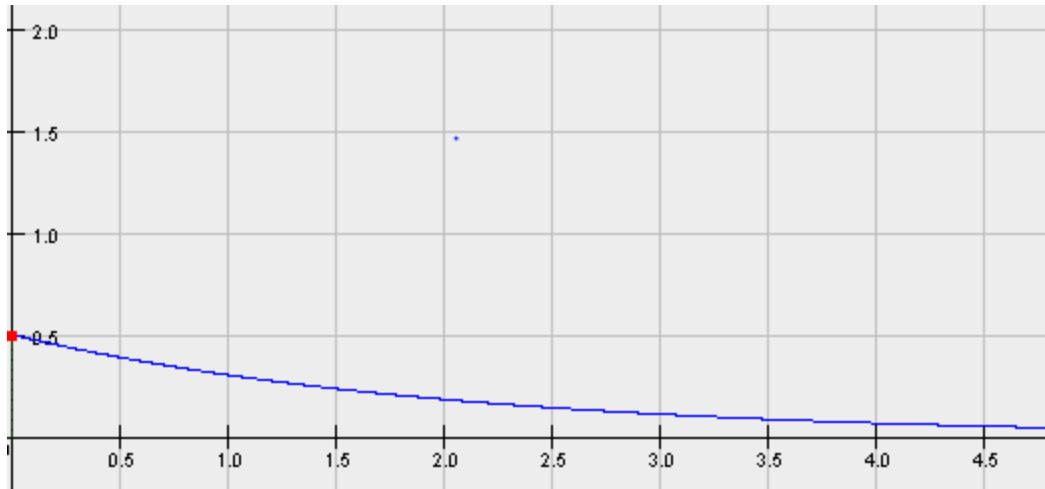
$$f_y(y) = \begin{cases} \left| \frac{d}{dy} g^{-1}(y) \right| f_x(g^{-1}(y)) & \text{if } y \in \mathcal{D}, \\ 0 & \text{otherwise} \end{cases}$$

Proof—see MGB p. 200.

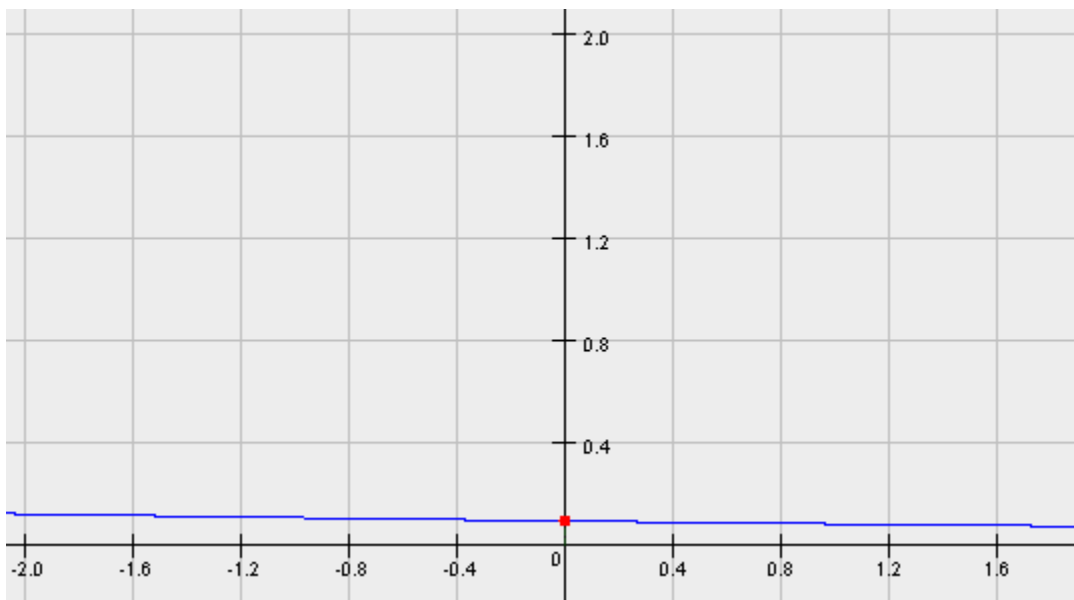
Example 1-2

Let's now define a function $Y = g(X) = 4X - 2$ and assume X has an exponential distribution. Note that $f_x(x) = \lambda e^{-\lambda x}$ where $\lambda > 0$ and $X = g^{-1}(Y) = \frac{y+2}{4}$. Since the mean for the exponential is $1/\lambda$, a simplifying assumption of a mean of 2 will give us the parameter $\lambda = 1/2$. Applying Theorem 11, we get the following distribution for $Y = g(X)$:

$$f_y(y) = \begin{cases} \left| \frac{d}{dy} \frac{y+2}{4} \right| \frac{1}{2} e^{-\frac{1}{2} \frac{(y+2)}{4}} = \frac{1}{8} e^{-\frac{y+2}{8}} & \text{if } \frac{y+2}{4} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$



Density function for X (exponential distribution with $\lambda=1/2$)



Density function of the transformation $Y = g(X) = 4X - 2$
with X distributed exponentially with $\lambda=1/2$

Clearly from this example, we could also solve the transformation more generally without a restriction on λ as well.

If $y=g(x)$ does not define a one-to-one transformation of X onto D , (and hence the inverse $x=g^{-1}(y)$ does not exist), we can still use the transformation technique, we just need to split the domain up into sections (disjoint sets) on which $y=g(x)$ is monotonic and its inverse will exist.

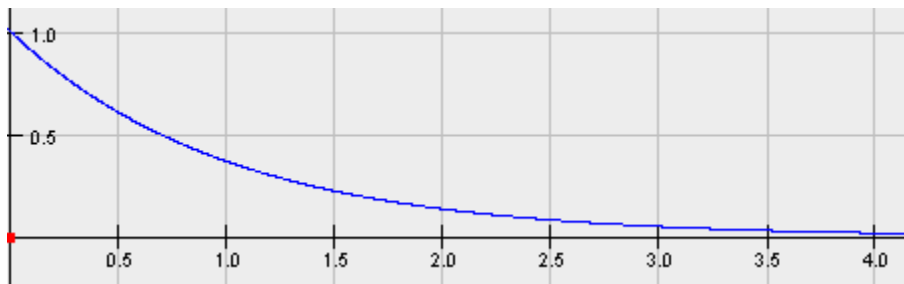
Example 1-3

Let $Y = g(x) = \sin x$ on the interval $[0, \pi]$. To ensure that inverse functions exist, split the domain into $[0, \pi/2)$ and $[\pi/2, \pi]$. Also, let x again have an exponential distribution, this time with mean 1 and $\lambda = 1$. Solving for x yields $x = \sin^{-1} y$, and applying Theorem 11, we have:

$$f_y(y) = \begin{cases} \left| \frac{d}{dy} \sin^{-1} y \right| e^{-\sin^{-1} y} & \text{if } \sin^{-1} y \in [0, 1], & x \in [0, \pi/2) \\ \left| \frac{d}{dy} \sin^{-1} y \right| e^{-\sin^{-1} y} & \text{if } \sin^{-1} y \in [0, 1], & x \in [\pi/2, \pi] \\ 0 & \text{otherwise} \end{cases}$$

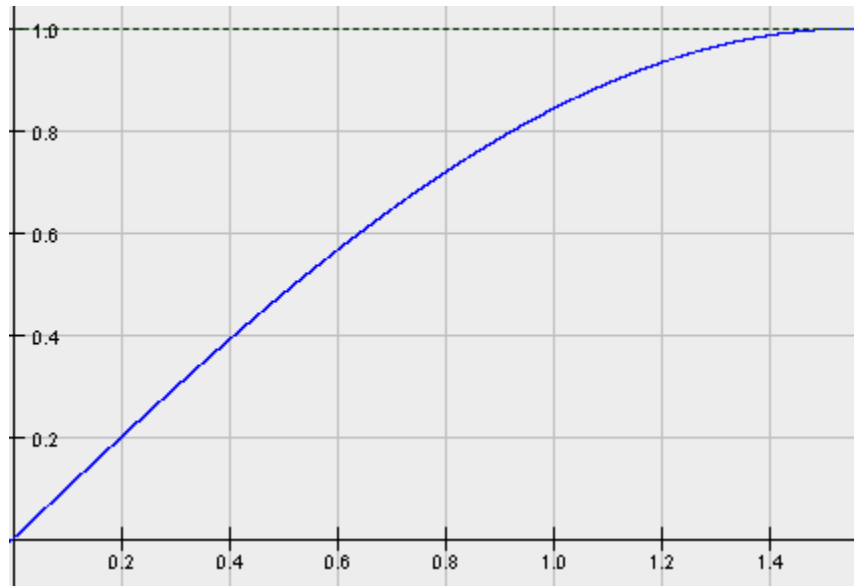
This becomes:

$$f_y(y) = \begin{cases} \frac{1}{\sqrt{1-y^2}} \left(e^{-\sin^{-1} y} \right) & \text{if } \sin^{-1} y \in [0, 1], & x \in [0, \pi] \\ 0 & \text{otherwise} \end{cases}$$

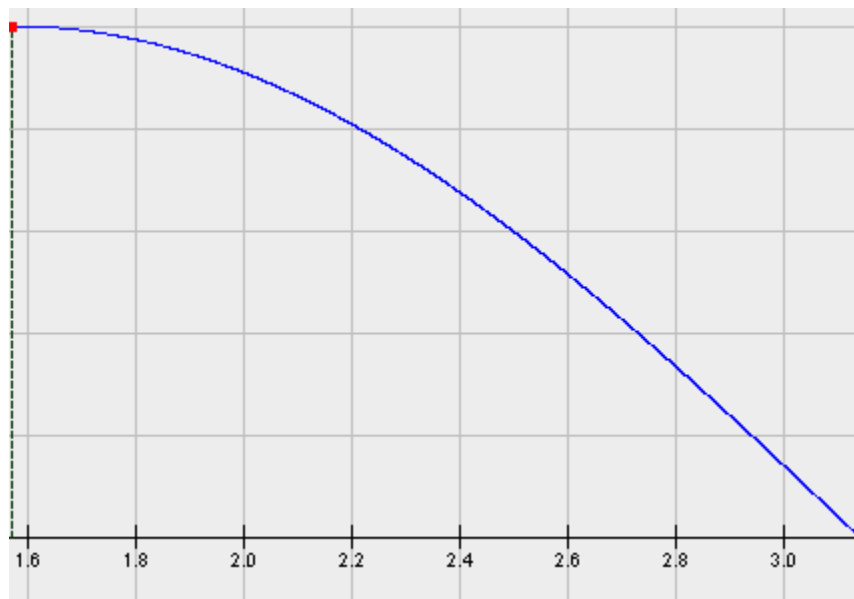


Density function for X (exponential distribution with $\lambda=1$)

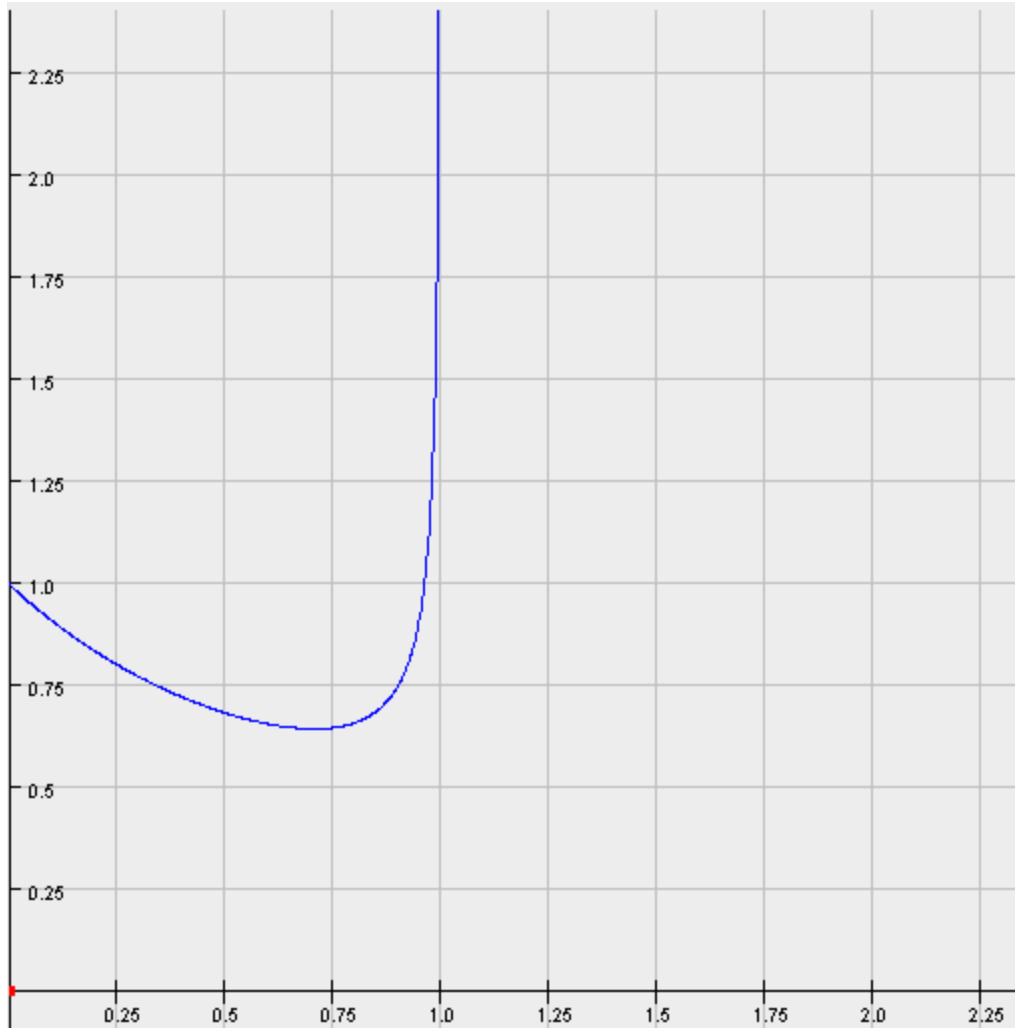
Splitting the function Y into invertible halves (both halves monotonic)



$$Y = \sin(X) \text{ for } X \in [0, \pi/2]$$



$$Y = \sin(X) \text{ for } X \in [\pi/2, \pi]$$



Density Function for $Y = g(X) = \sin(X)$ with X distributed exponentially with $\lambda=1$

1.3 Probability Integral Transform

As noted previously in class, a common application of the $Y = g(X)$ transformation is creating random samples for various continuous distributions. This involves creating a function $Y = g(X)$ = the unit uniform distribution, and then finding the distribution of this given some underlying distribution of X . In other words, we set $Y = 1$, plug in the known distribution of X , and then solve for X and apply Theorem 11 as above to get the transformed density. Then, starting with a random sample from the unit uniform distribution, we can plug in values for Y and generate a random sample for the distribution of X .

2. Multivariate Transformations

The procedure that is given in the book for finding the distribution of transformations of discrete random variables is notation heavy. It is also hard to understand how to apply the procedure as given to an actual example. Here we will instead explain the procedure with an example.

2.1 Discrete Multivariate Transformations

Example 2-1

Mort has three coins, a penny, a dime, and a quarter. These coins are not necessarily fair. The penny only has a $1/3$ chance of landing on heads, the dime has a $1/2$ chance of landing on heads, and the quarter has a $4/5$ chance of landing on heads. Mort flips each of the coins once in a round.

Let 1 mean “heads” and let 0 mean “tails”. Also, we will refer to a possible outcome of the experiment as follows: {penny, dime, quarter} so that {1,1,0} means the penny and the dime land on heads while the quarter lands on tails.

Below is the joint distribution of the variables “P”, “D”, and “Q”:

| Cases | Outcomes | Probability |
|-------|----------------------|-------------|
| 1 | $f_{PDQ}(0, 0, 0) =$ | $1/15$ |
| 2 | $f_{PDQ}(0, 0, 1) =$ | $4/15$ |
| 3 | $f_{PDQ}(0, 1, 0) =$ | $1/15$ |
| 4 | $f_{PDQ}(1, 0, 0) =$ | $1/30$ |
| 5 | $f_{PDQ}(1, 1, 0) =$ | $1/30$ |
| 6 | $f_{PDQ}(1, 0, 1) =$ | $2/15$ |
| 7 | $f_{PDQ}(0, 1, 1) =$ | $4/15$ |
| 8 | $f_{PDQ}(1, 1, 1) =$ | $2/15$ |

Let’s say we care about two functions of the above random variables: How many heads there are in a round (Y_1), and how many heads there are between pennies and quarters, without caring about the result of the dimes (Y_2).

We can either find the distributions individually, or the joint distribution of Y_1 and Y_2 . Lets start with finding them individually.

To do this, we find all of the cases of the pre-transformed joint distribution that apply to all the possible outcomes of Y_1 and Y_2 and add them up.

Distribution of Y_1 , $f_{Y_1}(y_1) =$

| # of Heads | Applicable Cases | Probabilities |
|------------|------------------|---------------|
| 0 | 1 | 1/15 |
| 1 | 2, 3, 4 | 11/30 |
| 2 | 5, 6, 7 | 13/30 |
| 3 | 8 | 2/15 |

Distribution of Y_2 , $f_{Y_2}(y_2) =$

| P or Q heads | Applicable Cases | Probabilities |
|--------------|------------------|---------------|
| 0 | 1, 3 | 2/15 |
| 1 | 2, 4, 5, 7 | 18/30 |
| 2 | 6, 8 | 4/15 |

We will do this same process for finding the joint distribution of Y_1 and Y_2 .

$f_{Y_1, Y_2}(y_1, y_2) =$

| Outcome | Applicable Cases | Probabilities |
|----------------------|------------------|---------------|
| $f_{Y_1, Y_2}(0, 0)$ | 1 | 1/15 |
| $f_{Y_1, Y_2}(0, 1)$ | N/A | 0 |
| $f_{Y_1, Y_2}(0, 2)$ | N/A | 0 |
| $f_{Y_1, Y_2}(1, 0)$ | 3 | 1/15 |
| $f_{Y_1, Y_2}(1, 1)$ | 2, 4 | 9/30 |
| $f_{Y_1, Y_2}(1, 2)$ | N/A | 0 |
| $f_{Y_1, Y_2}(2, 0)$ | N/A | 0 |
| $f_{Y_1, Y_2}(2, 1)$ | 5, 7 | 9/30 |
| $f_{Y_1, Y_2}(2, 2)$ | 6 | 2/15 |
| $f_{Y_1, Y_2}(3, 0)$ | N/A | 0 |
| $f_{Y_1, Y_2}(3, 1)$ | N/A | 0 |

| | | |
|----------------------|---|------|
| $f_{Y_1, Y_2}(3, 2)$ | 8 | 2/15 |
|----------------------|---|------|

To Simplify: $f_{Y_1, Y_2}(y_1, y_2) =$

| Outcome | Probabilities |
|----------------------|---------------|
| $f_{Y_1, Y_2}(0, 0)$ | 1/15 |
| $f_{Y_1, Y_2}(1, 0)$ | 1/15 |
| $f_{Y_1, Y_2}(1, 1)$ | 9/30 |
| $f_{Y_1, Y_2}(2, 1)$ | 9/30 |
| $f_{Y_1, Y_2}(2, 2)$ | 2/15 |
| $f_{Y_1, Y_2}(3, 2)$ | 2/15 |

To see something interesting, let Y_3 be the random variable of the number of heads of the dime in a round. The distribution of Y_3 , $f_{Y_3}(y_3) =$

| Outcome | Probabilities |
|--------------|---------------|
| $f_{Y_3}(0)$ | $\frac{1}{2}$ |
| $f_{Y_3}(1)$ | $\frac{1}{2}$ |

The joint distribution of Y_2 and Y_3 is then $f_{Y_2, Y_3}(y_2, y_3) =$

| Outcome | Applicable Cases | Probabilities |
|----------------------|------------------|---------------|
| $f_{Y_2, Y_3}(0, 0)$ | 1 | 1/15 |
| $f_{Y_2, Y_3}(1, 0)$ | 2, 4 | 9/30 |
| $f_{Y_2, Y_3}(2, 0)$ | 6 | 2/15 |
| $f_{Y_2, Y_3}(0, 1)$ | 3 | 1/15 |
| $f_{Y_2, Y_3}(1, 1)$ | 5, 7 | 9/30 |
| $f_{Y_2, Y_3}(2, 1)$ | 8 | 2/15 |

Because Y_2 and Y_3 are independent, $f_{Y_2, Y_3}(y_2, y_3) = f_{Y_2}(y_2) * f_{Y_3}(y_3)$, which is really easy to see here.

To apply the process used above to different problems can be straightforward or very hard, but most of the problems we have seen in the book or outside the book have been quite solvable. The key to doing this type of transformation is to make sure you know exactly which sample point from the untransformed distribution corresponds to which sample point from the transformed distribution. Once you have figured this out, solving for the transformed distribution is straightforward.

We think the process outlined in this example is applicable to all discrete multivariate transformations, except perhaps in the case where the number of sample points with positive probability is infinite.

2.2 Continuous Multivariate Transformations

Extending the transformation process from a single continuous random variable to multiple random variables is conceptually intuitive, but often very difficult to implement.

The general method for solving continuous multivariate transformations is given below. In this method, we will transform multiple random variables into multiple functions of these random variables. Below is the theorem we will use in the method.

This is **Theorem 15** directly from Mood, Graybill, and Boes;

Let X_1, X_2, \dots, X_n be jointly continuous random variables with density function $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$. We want to transform these variables into Y_1, Y_2, \dots, Y_n where $Y_1, Y_2, \dots, Y_n \in \mathfrak{R}$. Let $\mathfrak{X} = \{(x_1, x_2, \dots, x_n): f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) > 0\}$. Assume that \mathfrak{X} can be decomposed into sets $\mathfrak{X}_1, \dots, \mathfrak{X}_m$ such that $y_1 = g_1(x_1, \dots, x_n), \dots, y_n = g_n(x_1, \dots, x_n)$ is a one to one transformation of \mathfrak{X}_i onto \mathfrak{R} , $i = 1, \dots, m$. Let $x_1 = g_{1i}^{-1}(y_1, y_2, \dots, y_n), \dots, x_n = g_{ni}^{-1}(y_1, y_2, \dots, y_n)$ denote the inverse transformation of \mathfrak{R} onto \mathfrak{X}_i , $i = 1, \dots, m$. Define J_i as the determinant of the matrix of the derivatives of inverse transformations with respect to the transformed variables, y_1, y_2, \dots, y_n , for all $i = 1, \dots, m$.

Assuming that all the partial derivatives in J_i are continuous over \mathfrak{R} and the determinant J_i is non-zero for all the i 's. Then

$$f_{Y_1, Y_2, \dots, Y_n}(y_1, y_2, \dots, y_n) = \sum_{i=1}^m |J_i| f_{X_1, X_2, \dots, X_n}(g_{1i}^{-1}(y_1, y_2, \dots, y_n), \dots, g_{ni}^{-1}(y_1, y_2, \dots, y_n))$$

for $(y_1, y_2, \dots, y_n) \in \mathfrak{R}$.

$|J|$ in this theorem is the absolute value of the determinate of the Jacobian matrix. Using the notation of the theorem, the Jacobian matrix is the matrix of the

first-order partial derivatives of all of the X 's with respect to all of the Y 's. For a transformation from two variables, X and Y , to two functions of the two variables, Z and U , the Jacobian matrix is:

$$\text{Jacobian Matrix} = \begin{bmatrix} \frac{dx}{dz} & \frac{dx}{du} \\ \frac{dy}{dz} & \frac{dy}{du} \end{bmatrix}$$

The determinate of the Jacobian Matrix, often called simply a "Jacobian", is then

$$J = \begin{vmatrix} \frac{dx}{dz} & \frac{dx}{du} \\ \frac{dy}{dz} & \frac{dy}{du} \end{vmatrix} = \frac{dx}{dz} \frac{dy}{du} - \frac{dy}{dz} \frac{dx}{du}$$

The reason we need the Jacobian in Theorem 15 is that we are doing what mathematicians refer to as a coordinate transformation. It has been proven that an absolute value of a Jacobian must be used in this case (refer to a graduate level calculus book for a proof).

Notice from Theorem 15 that we take n number of X 's and transform them into n number of Y 's. This must be the case for the theorem to work, even if we only care about k number of Y 's where $k \leq n$.

Let's say that you have a set of random variables, X_1, X_2, \dots, X_n and you would like to know the distributions of functions of these random variables. Let's say the functions you would like to know are called $Y_i(X_1, X_2, \dots, X_n)$ for $i \in (0, k)$.

In order to make the dimensions the same (for the theorem to work), we will need to add transformations Y_{i+1} to Y_n , even though we do not care about them. Note the choice of these extra Y 's can greatly affect the complexity of the problem, although we have no intuition about what choices will make the calculations the easiest. After the joint density is found with this method, you can integrate out the variables you do not care about to make a joint density of only the 0 through k Y_i 's you do care about.

Another important point in this theorem is an extension of a problem we saw above in the univariate case. If our transformations are not 1 to 1, then we must divide up the space into areas where we do have 1 to 1 correspondence, do our calculations, and then sum them up, which is exactly what the theorem does.

Example 2-2

Let (X, Y) have a joint density function of $f_{X,Y}(x, y) = \begin{cases} xe^{-x}e^{-y} & \text{for } x, y > 0 \\ 0 & \text{if else} \end{cases}$

Note that $f_X(x) = xe^{-x}$ for $x > 0$ and $f_Y(y) = e^{-y}$ for $y > 0$.

{Also note: $\int_0^\infty \int_0^\infty xe^{-x}e^{-y} dx dy = 1$, $\int_0^\infty xe^{-x} dx = 1$, and $\int_0^\infty e^{-y} dy = 1$ }

We want to find the distribution for the variables Z and U , where $Z = X + Y$ and $U = Y/X$.

We first find the Jacobian. Note: $Z = X + Y$ and $U = Y/X$ and therefore:

$$X = g_X^{-1}(z, u) = \frac{Z}{U+1} \text{ and } Y = g_Y^{-1}(z, u) = \frac{UZ}{U+1}$$

So:

$$\frac{dx}{dz} = \frac{1}{U+1} \text{ and } \frac{dx}{du} = -\frac{Z}{(U+1)^2}$$

And

$$\frac{dy}{dz} = \frac{U}{U+1} \text{ and } \frac{dy}{du} = \frac{Z}{(U+1)^2}$$

Plugging these into the definition:

$$J = \begin{vmatrix} \frac{dx}{dz} & \frac{dx}{du} \\ \frac{dy}{dz} & \frac{dy}{du} \end{vmatrix} = \frac{dx}{dz} \frac{dy}{du} - \frac{dy}{dz} \frac{dx}{du}$$

So the determinate of the Jacobian Matrix is $J = \frac{z}{(u+1)^2}$

Z is the sum of only positive numbers and U is a positive number divided by a positive number, so the absolute value of J is just J :

$$|J| = \left| \frac{z}{(u+1)^2} \right| = \frac{z}{(u+1)^2}$$

Notice that in this example, the transformations are always one-to-one, so we do not need to worry about segmenting up the density function.

So using the theorem:

$$f_{Z,U}(z, u) = |J| f_{X,Y}(g_X^{-1}(z, u), g_Y^{-1}(z, u))$$

Substituting in $|J| = \frac{z}{(u+1)^2}$, $g_X^{-1}(z, u) = X = \frac{z}{u+1}$, and $g_Y^{-1}(z, u) = Y = \frac{UZ}{u+1}$, we get:

$$f_{Z,U}(z, u) = \frac{z}{(u+1)^2} f_{X,Y}\left(\frac{z}{u+1}, \frac{UZ}{u+1}\right)$$

Remember from above that $f_{X,Y}(x, y) = \begin{cases} xe^{-x}e^{-y} & \text{for } x, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$

So our joint density becomes:

$$f_{Z,U}(z, u) = \frac{z}{(u+1)^2} * \frac{z}{u+1} e^{-z/(u+1)} e^{-uz/(u+1)} \text{ for } z, u \geq 0$$

$$f_{Z,U}(z, u) = \frac{z^2}{(u+1)^3} e^{-z/(u+1)} e^{-uz/(u+1)} \text{ for } z, u \geq 0$$

If we wanted to know the density functions of Z and U separately, just integrate out the other variable in the normal way:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z,U}(z, u) du = \int_{-\infty}^0 f_{Z,U}(z, u) du + \int_0^{\infty} f_{Z,U}(z, u) du$$

But U cannot be less than zero so:

$$\int_{-\infty}^0 f_{Z,U}(z, u) du = 0$$

Simplifying:

$$f_Z(z) = \int_0^{\infty} f_{Z,U}(z, u) du$$

$$= \frac{z^2}{2} e^{-z} \text{ for } z > 0.$$

And

$$f_U(u) = \int_0^{\infty} f_{Z,U}(z, u) dz$$

$$= \frac{2}{(1+u)^3} \text{ for } u > 0.$$

3. Simulating

Often obtaining an analytical solution to these sorts of transformations is nearly impossible. This may be because the integral has no closed form solution or that the transformed bounds of the integral are very hard to translate. In these situations, a different approach to estimating the distribution of a new statistic is to run simulations. We have used Matlab for the following examples because there are some cool functions you can download that will randomly draw from many different kinds of distributions. (If you don't have access to this function, you can do the trick where you randomly sample over the uniform distribution between 0 and 1 and then back out the random value through the CDF. This technique was covered earlier in the class notes).

A boring example: (We started with boring to illustrate the technique, below this are more interesting examples. Also, we have tried to solve this simple example using the method above and found it very difficult).

Let's say we have two uniformly distributed variables, X_1 and X_2 such that $X_1 \sim U(0,1)$ and $X_2 \sim U(0,1)$. We want to find the distribution of Y_1 where $Y_1 = X_1 + X_2$.

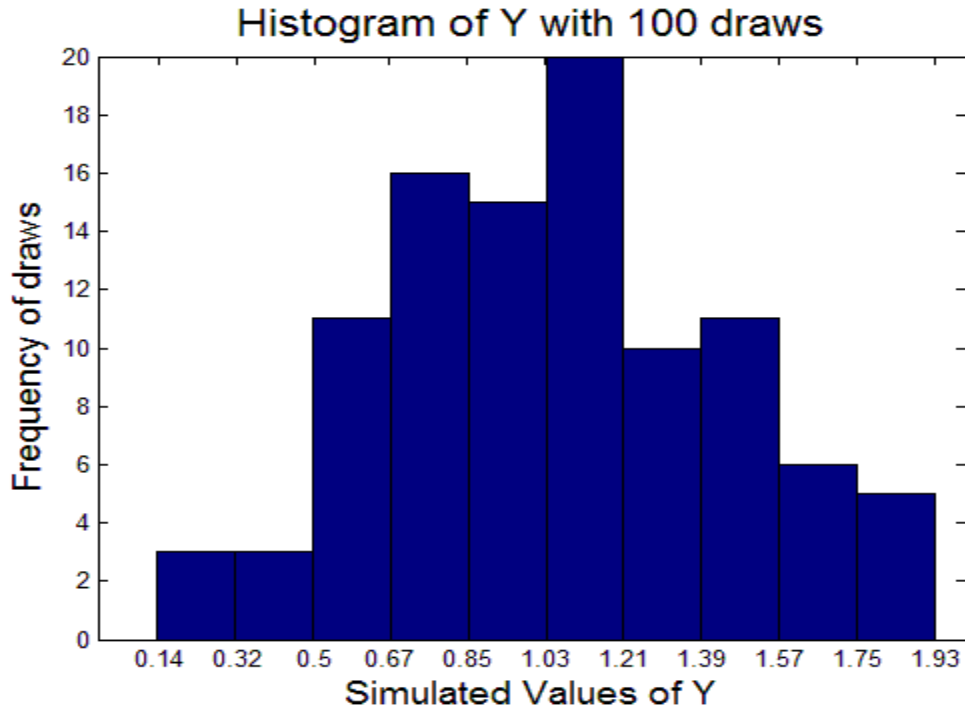
Choose how many simulation points you would like to have and then take that many draws from the uniform distribution for X_1 . When we say take a draw from a particular distribution, that means asking Matlab to return a realization of a random variable according to its distribution, so Matlab will give us a number. The number could potentially be any number where its distribution function is positive. Each draw that we take is a single simulated observation from that distribution.

Let's say we have chosen to take a sample with 100 simulated observations, so we now have a 100 element vector of numbers that are the realized values for X_1 that we observed. We will call these x_1 . Do the same thing for X_2 to get a 100 element vector of simulated observations from X_2 that we will call x_2 .

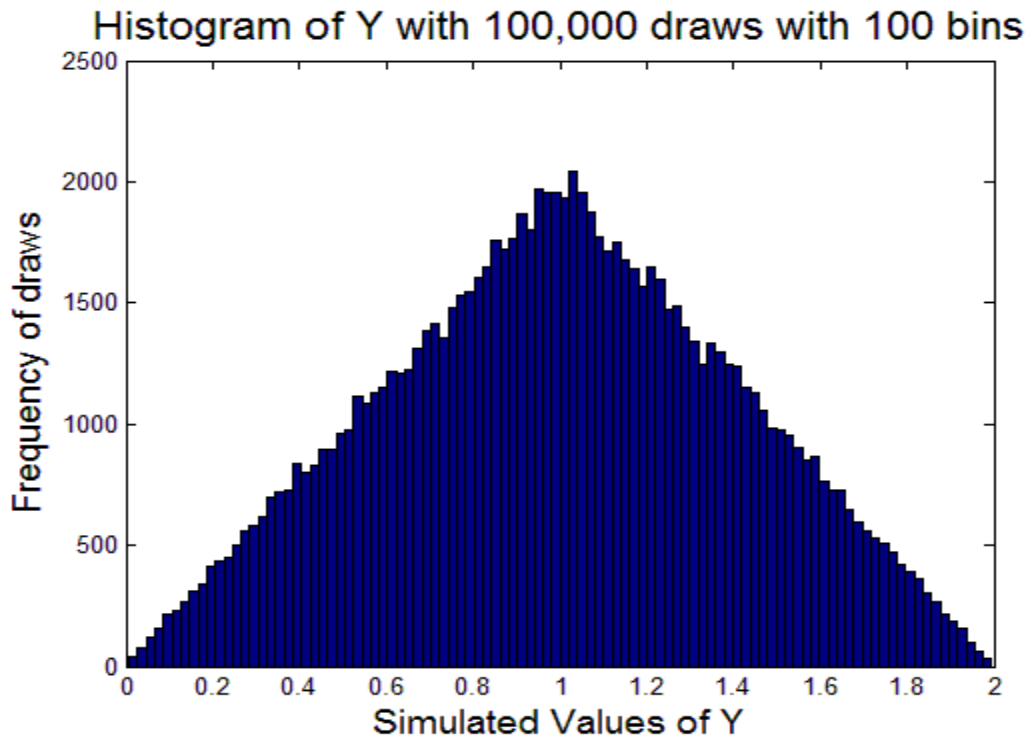
Now we create the vector for realized values for Y_1 . The first element of y_1 is the first element of x_1 plus the first element of x_2 . We now have a realized y_1 vector that is 100 elements long. We treat each element in this vector as an observation of the random variable Y_1 .

We now graph these observations of Y_1 by making a histogram. The realized values of y_1 will be on the horizontal axis while the frequency of the different values will be along the vertical axis. You can alter how good the graph looks by adjusting the number of different bins you have in your histogram.

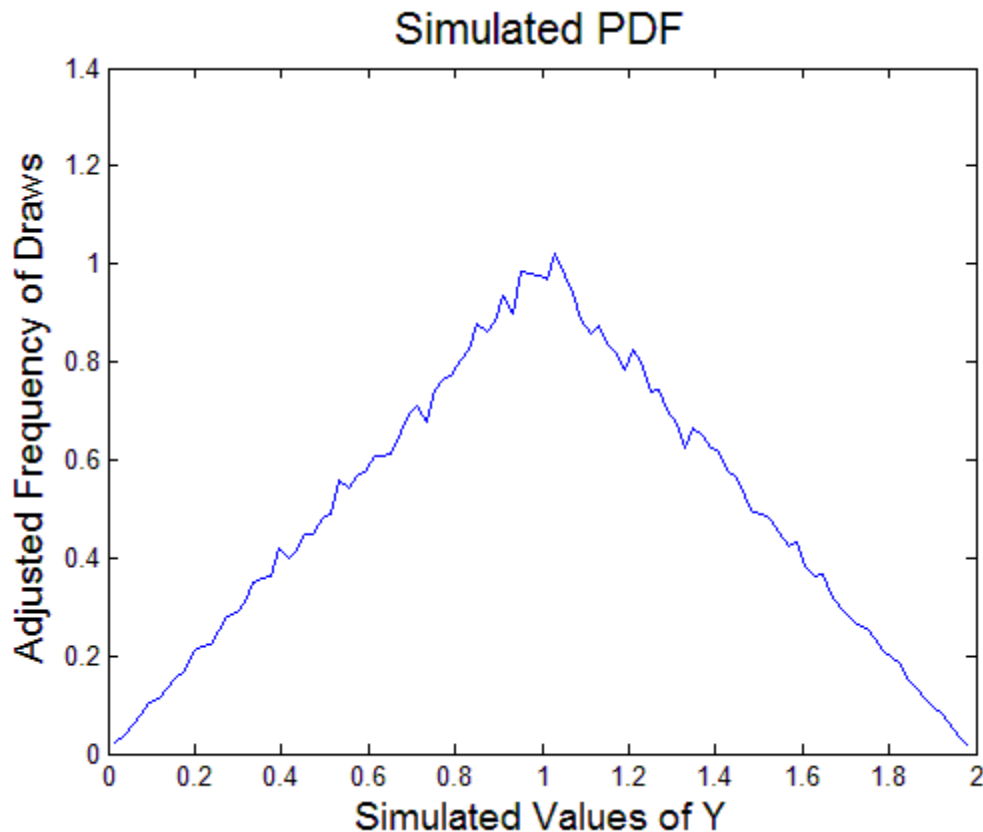
Below is a graph of 100 draws and 10 bins:



We will try to make the graph look better by trying 100,000 simulated observations with 100 bins:



This graph looks a lot better, but it is not yet an estimation of the distribution of Y_1 because the area underneath the curve is not even close to one. To fix this problem, add up the area underneath and divide each frequency bar in the histogram by this number. Here, our area is 2000, so dividing the frequency levels of each bin by 200 gives us the following graph with an area of 1:



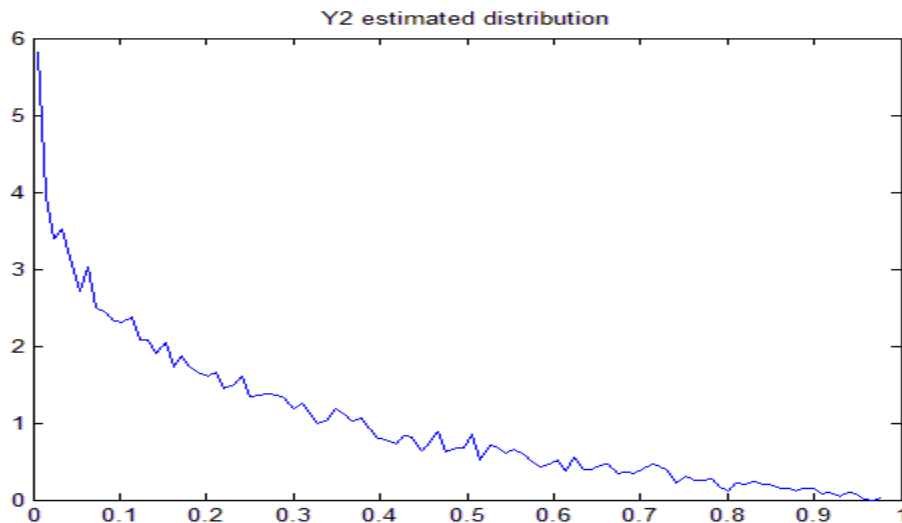
It looks as if the distribution here is $Y_1 = \begin{cases} y_1 & \text{if } 0 \leq y_1 \leq 1 \\ 2 - y_1 & \text{if } 1 \leq y_1 \leq 2 \end{cases}$

This seems to make intuitive sense and is close to what we would expect.

We will not go further into estimating the functional form of this distribution because that is a nuanced art that seems beyond the scope of these notes. It is good to know however that there are many different ways the computer can assist here, whether it is fitting a polynomial or doing something called

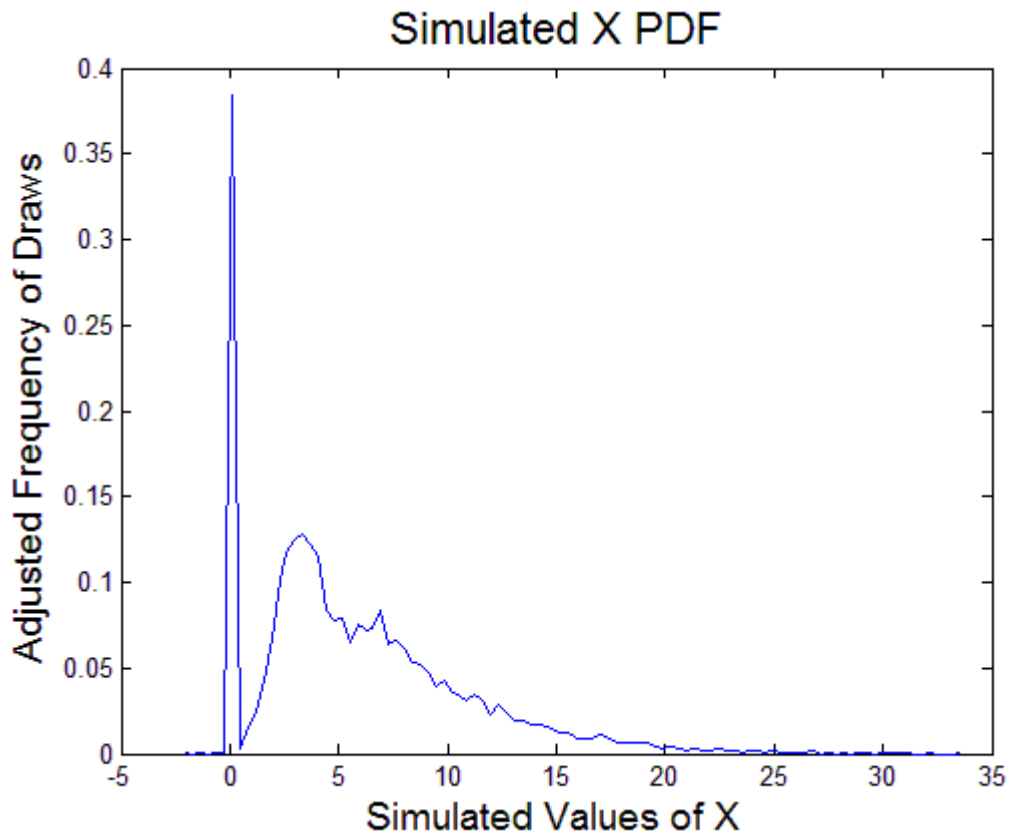
interpolating which doesn't give an explicit functional form, but will return an estimate of the dependent variable given different inputs.

Using the above frame work, the graph of the estimated distribution of $Y_2 = X_1/X_2$ is:



We can extend this approach further to estimate the distributions of statistics that would be nearly impossible to calculate any other way. For example, lets say we have a variable $Y \sim N(3, 1)$, so Y is normally distributed with a mean around 3 and a standard deviation of 1. We also have a variable Z that has a Poisson distribution with the parameter $\lambda = 2$.

We want to know the distribution of $X = Y * Z$, so we are looking for a statistic that is a function of a continuous and a discrete random variable. Using the process from above, we simulated 10,000 observations of this statistic and so we estimate the distribution to look something like this:



Let's see a very crazy example:

Let's say the variable Z has an extreme-value distribution with mean 0 and scale 1 and Y is a draw from the uniform distribution between 0 and 1.

We want to find the distribution of the statistic X where

$$X = (\cot^{-1} Z)^{1/2} * Y$$

Using 10,000 simulated observations and 100 bins, an estimate of the distribution of X is:

Simulated X PDF

