# TRANSFORMATIONS OF RANDOM VARIABLES

## 1. Introduction

1.1. **Definition.** We are often interested in the probability distributions or densities of functions of one or more random variables. Suppose we have a set of random variables, $X_1$, $X_2$, $X_3$, . . . $X_n$, with a known joint probability and/or density function. We may want to know the distribution of some function of these random variables $Y = \phi(X_1, X_2, X_3, ...X_n)$. Realized values of $y$ will be related to realized values of the $X$'s as follows

$$y = \Phi(x_1, x_2, x_3, \ldots, x_n)$$

A simple example might be a single random variable $X$ with transformation

$$y = \Phi(x) = log(x)$$

1.2. **Techniques for finding the distribution of a transformation of random variables.**

1.2.1. *Distribution function technique.* We find the region in $x_1, x_2, x_3, ...x_n$ space such that $\Phi(x_1, x_2, ...x_n) \leq \phi$. We can then find the probability that $\Phi(x_1, x_2, ...x_n) \leq \phi$, i.e., $P[\, \Phi(x_1, x_2, ...x_n) \leq \phi]$ by integrating the density function $f(x_1, x_2, ...x_n)$ over this region. Of course, $F_{\Phi}(\phi)$ is just $P[\Phi \leq \phi]$. Once we have $F_{\Phi}(\phi)$, we can find the density by differentiation.

1.2.2. *Method of transformations (inverse mappings).* Suppose we know the density function of X. Also suppose that the function $y = \Phi(x)$ is differentiable and monotonic for values within its range for which the density $f(x) \neq 0$. This means that we can solve the equation $y = \Phi(x)$ for $x$ as a function of y. We can then use this inverse mapping to find the density function of $y$. We can do a similar thing when there is more than one variable $X$ and then there is more than one mapping $\Phi$.

## 2. Distribution Function Technique

2.1. **Using the Distribution Function Technique.** Below are examples where we use the method described above.

2.2. **Example 1.** Let the probability density function of $X$ be given by

$$f(x) = \begin{cases} 6\,x\,(1 - x) & 0 < x < 1 \\ 0 & otherwise \end{cases}$$

*Date*: September 21, 2006.

Now find the probability density of $Y = X^3$. Let $G(y)$ denote the value of the distribution function of $Y$ at $y$ and write

$$
\begin{aligned}
G(y) &= P(Y \leq y) = P(X^3 \leq y) \\
&= P\left(X \leq y^{1/3}\right) = \int_0^{y^{1/3}} 6x(1-x)\,dx \\
&= \int_0^{y^{1/3}} \left(6x - 6x^2\right) dx = \left(3x^2 - 2x^3\right)\Big|_0^{y^{1/3}} \\
&= 3y^{2/3} - 2y
\end{aligned}
$$

Now differentiate $G(y)$ to obtain the density function $g(y)$

$$
\begin{aligned}
g(y) &= \frac{dG(y)}{dy} = \frac{d}{dy}\left(3y^{2/3} - 2y\right) \\
&= 2y^{-1/3} - 2 = 2(y^{-1/3} - 1), \quad 0 < y < 1
\end{aligned}
$$

2.3. **Example 2.** Let the probability density function of $X_1$ and of $X_2$ be given by

$$
f(x_1, x_2) = \begin{cases} 2e^{-x_1 - 2x_2} & x_1 > 0,\ x_2 > 0 \\ 0 & otherwise \end{cases}
$$

Now find the probability density of $Y = X_1 + X_2$ or $X_1 = Y - X_2$. Given that $Y$ is a linear function of $X_1$ and $X_2$, we can easily find $F(y)$ as follows: Let $F_Y(y)$ denote the value of the distribution function of $Y$ at $y$ and write

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) = \int_0^y \int_0^{y-x_2} 2e^{-x_1 - 2x_2}\,dx_1\,dx_2 \\
&= \int_0^y -2e^{-x_1 - 2x_2}\Big|_0^{y-x_2}\,dx_2 \\
&= \int_0^y \left[\left(-2e^{-y + x_2 - 2x_2}\right) - \left(-2e^{-2x2}\right)\right] dx_2 \\
&= \int_0^y -2e^{-y - x_2} + 2e^{-2x2}\,dx_2 \\
&= \int_0^y 2e^{-2x2} - 2e^{-y - x_2}\,dx_2
\end{aligned}
$$

Now integrate with respect to $x_2$ as follows

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) = \int_0^y 2e^{-2x2} - 2e^{-y - x_2}\,dx_2 \\
&= -e^{-2x2} + 2e^{-y - x_2}\Big|_0^y \\
&= -e^{-2y} + 2e^{-y - y} - \left[-e^0 + 2e^{-y}\right] \\
&= e^{-2y} - 2e^{-y} + 1
\end{aligned}
$$

Now differentiate $F_Y(y)$ to obtain the density function $f(y)$

$$\begin{aligned} f_Y(y) &= \frac{dF(y)}{dy} = \frac{d}{dy}\left(e^{-2y} - 2e^{-y} + 1\right) \\ &= -2e^{-2y} + 2e^{-y} = 2e^{-2y}\,(-1 + e^{y}) \end{aligned}$$

2.4. **Example 3.** Let the probability density function of $X$ be given by

$$\begin{aligned} f_X(x) &= \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < \infty \end{aligned}$$

Now let $Y = \Phi(X) = e^{X}$. We can then find the distribution of $Y$ by integrating the density function of $X$ over the appropriate area that is defined as a function of $y$. Let $F_Y(y)$ denote the value of the distribution function of $Y$ at $y$ and write

$$\begin{aligned} F_Y(y) &= P(Y \le y) \qquad (1) \\ &= P\left(e^X \le y\right) = P(X \le \ln y), \quad y > 0 \\ &= \int_{-\infty}^{\ln y} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx, \quad y > 0 \end{aligned}$$

Now differentiate $F_Y(y)$ to obtain the density function $f(y)$. In this case we will need the rules for differentiating under the integral sign. They are given by theorem 1 which we state below without proof.

**Theorem 1.** *Suppose that $f$ and $\frac{\partial f}{\partial x}$ are continuous in the rectangle*

$$R = \{(x, t) : a \le x \le b, c \le t \le d\}$$

*and suppose that $u_0(x)$ and $u_1(x)$ are continuously differentiable for $a \le x \le b$ with the range of $u_0(x)$ and $u_1(x)$ in $(c, d)$. If $\psi$ is given by*

$$\psi(x) = \int_{u_0(x)}^{u_1(x)} f(x, t)\, dt$$

*then*

$$\begin{aligned} \frac{d\psi}{dx} &= \frac{\partial}{\partial x}\int_{u_0(x)}^{u_1(x)} f(x, t)\, dt \\ &= f(x, u_1(x))\,\frac{du_1(x)}{dx} - f(x, u_0(x))\,\frac{du_0(x)}{dx} + \int_{u_0(x)}^{u_1(x)} \frac{\partial f(x,t)}{\partial x} dt \end{aligned}$$

*If one of the bounds of integration does not depend on $x$, then the term involving its derivative will be zero.*

For a proof of theorem 1 see (Protter [3, p. 425] ). Applying this to equation 1 where $y$ takes the role of $x$, $\ln y$ takes the role of $u_1(x)$, and $x$ takes the role of $t$ in the theorem we obtain

$$\begin{aligned} F_Y(y) &= \int_{-\infty}^{\ln y} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx, \quad y > 0 \\ F_Y'(y) &= f_Y(y) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(\ln y-\mu)^2}{2\sigma^2}\right]\left(\frac{1}{y}\right)\right) \\ &\quad + \int_{-\infty}^{\ln y} \frac{d}{dy}\left(\frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\right) dx \\ &= \left(\frac{1}{y\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(\ln y-\mu)^2}{2\sigma^2}\right]\right) \end{aligned} \tag{2}$$

The last line of equation 2 follows because

$$\frac{d}{dy}\left(\frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\right) = 0$$

## 3. Method of transformations (single variable)

### 3.1. **Discrete examples of the method of transformations.**

3.1.1. *One-to-one function.* Find a formula for the probability distribution of the total number of heads obtained in four tosses of a coin where the probability of a head is $0.60$.

The sample space, probabilities and the value of the random variable are given in Table 1.
From the table we can determine the probabilities as

$$P(X=0) = \frac{16}{625},\ P(X=1) = \frac{96}{625},\ P(X=2) = \frac{216}{625},\ P(X=3) = \frac{216}{625},\ P(X=4) = \frac{81}{625}$$

We can also compute these probabilities using counting rules. The probability of one head and then three tails is

$$\left(\frac{3}{5}\right)\left(\frac{2}{5}\right)\left(\frac{2}{5}\right)\left(\frac{2}{5}\right)$$

or

$$\left(\frac{3}{5}\right)^1\left(\frac{2}{5}\right)^3 = \frac{24}{625}$$

The probability of 3 heads and then one tail is

$$\left(\frac{3}{5}\right)\left(\frac{3}{5}\right)\left(\frac{3}{5}\right)\left(\frac{2}{5}\right)$$

or

$$\left(\frac{3}{5}\right)^3\left(\frac{2}{5}\right)^1 = \frac{54}{625}.$$

Of course there are other ways to obtain $1$ head and three tails besides one head and then three tails. In particular there are $\binom{4}{1} = 4$ ways to obtain one head. And there are $\binom{4}{0} = 1$ way to obtain

| Element of sample space | Probability | Value of random variable X (x) |
|---|---|---|
| HHHH | 81/625 | 4 |
| HHHT | 54/625 | 3 |
| HHTH | 54/625 | 3 |
| HTHH | 54/625 | 3 |
| THHH | 54/625 | 3 |
| HHTT | 36/625 | 2 |
| HTHT | 36/625 | 2 |
| HTTH | 36/625 | 2 |
| THHT | 36/625 | 2 |
| THTH | 36/625 | 2 |
| TTHH | 36/625 | 2 |
| HTTT | 24/625 | 1 |
| THTT | 24/625 | 1 |
| TTHT | 24/625 | 1 |
| TTTH | 24/625 | 1 |
| TTTT | 16/625 | 0 |

TABLE 1. **Outcomes, Probabilities and Number of Heads from Tossing a Coin Four Times.**

zero heads. Similarly, there are six ways to obtain two heads, four ways to obtain three heads and one way to obtain four heads. We can then write the probability mass function as

$$f(x) \;=\; \binom{4}{x} \left(\frac{3}{5}\right)^{x} \left(\frac{2}{5}\right)^{4-x} \; for\; x \;=\; 0,\, 1,\, 2,\, 3,\, 4$$

This, of course, is the binomial distribution. The probabilities of the various possible random variables are contained in table 2.

| Number of Heads x | Probability f(x) |
|---|---|
| 0 | 16/625 |
| 1 | 96/625 |
| 2 | 216/625 |
| 3 | 216/625 |
| 4 | 81/625 |

TABLE 2. **Probability of Number of Heads from Tossing a Coin Four Times**

Now consider a transformation of $X$ in the form $Y = 2X^{\,2} + X$. There are five possible outcomes for $Y$, i.e., $0, 3, 10, 21, 36$. Given that the function is one-to-one, we can make up a table describing the probability distribution for $Y$ (see Table 3).

3.1.2. *Case where the transformation is not one-to-one.* Now let the transformation of $X$ be given by $Z = (6 - 2X)^{\,2}$. The possible values for $Z$ are $0, 4, 16, 36$. When $X = 2$ and when $X = 4$, $Y = 4$. We can find the probability of $Z$ by adding the probabilities for cases when $X$ gives more than one value as shown in table 4.

| Y = 2 * (# heads) $^2$ + # of heads | | | |
|---|---|---|---|
| Number of Heads x | Probability f(x) | **y** | g(y) |
| 0 | 16/625 | 0 | 16/625 |
| 1 | 96/625 | 3 | 96/625 |
| 2 | 216/625 | 10 | 216/625 |
| 3 | 216/625 | 21 | 216/625 |
| 4 | 81/625 | 36 | 81/625 |

TABLE 3. **Probability of a Function of the Number of Heads from Tossing a Coin Four Times.**

| Y = (6 - (# heads))$^2$ | | |
|---|---|---|
| y | Number of Heads x | g(y) |
| 0 | 3 | $\frac{216}{625}$ |
| 4 | 2, 4 | $\frac{216}{625} + \frac{81}{625} = \frac{297}{625}$ |
| 16 | 1 | $\frac{96}{625}$ |
| 36 | 0 | $\frac{16}{625}$ |

TABLE 4. **Probability of a Function of the Number of Heads from Tossing a Coin Four Times (not one-to-one).**

3.2. **Intuitive Idea of the Method of Transformations.** The idea of a transformation is to consider the function that maps the random variable $X$ into the random variable $Y$. The idea is that if we can determine the values of $X$ that lead to any particular value of $Y$, we can obtain the probability of $Y$ by summing the probabilities of those values of $X$ that mapped into $Y$. In the continuous case, to find the distribution function, we want to integrate the density of $X$ over the portion of its space that is mapped into the portion of $Y$ in which we are interested. Suppose for example that both $X$ and $Y$ are defined on the real line with $0 \leq X \leq 1$ and $0 \leq Y \leq 10$. If we want to know $G(5)$, we need to integrate the density of $X$ over all values of $x$ leading to a value of $y$ less than five, where $G(5)$ is the probability that $Y$ is less than five.

3.3. **General formula when the random variable is discrete.** Consider a transformation defined by $y = \Phi(x)$. The function $\Phi$ defines a mapping from the sample space of the variable $X$, to a sample space for the random variable $Y$. If $X$ is discrete with frequency function $p_{\,X}$, then $\Phi(X)$ is discrete and has frequency function

$$p_{\Phi(X)}(t) \;=\; \sum_{x:\,\Phi(x)=t} p_X(x) = \sum_{x \epsilon \Phi^{-1}(t)} p_X(x)$$

The process is simple in this case. One identifies $g^{\,-1}(t)$ for each $t$ in the sample space of the random variable $Y$, and then sums the probabilities which is what we did in section 3.1.

3.4. **General change of variable or transformation formula.**

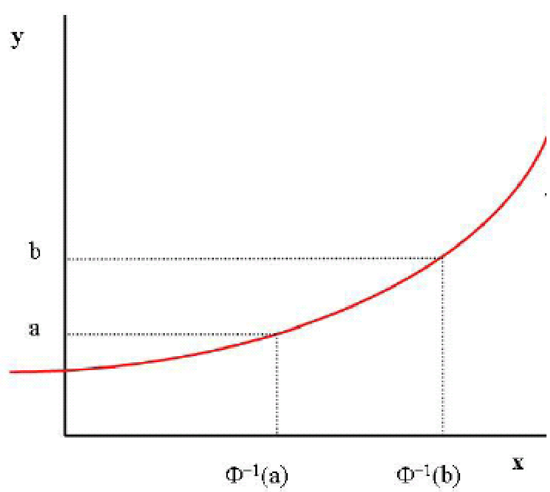
**Theorem 2.** *Let $f_X(x)$ be the value of the probability density of the continuous random variable $X$ at $x$. If the function $y = \Phi(x)$ is differentiable and either increasing or decreasing (monotonic) for all values within the range of $X$ for which $f_X(x) \neq 0$, then for these values of $x$, the equation $y = \Phi(x)$ can be uniquely solved for $x$ to give $x = \Phi^{-1}(y) = w(y)$ where $w(\cdot) = \Phi^{-1}(\cdot)$. Then for the corresponding values of $y$, the probability density of $Y = \Phi(X)$ is given by*

$$g(y) = f_Y(y) = \begin{cases} f_X\left[\Phi^{-1}(y)\right] \cdot \left|\frac{d\Phi^{-1}(y)}{dy}\right| & when \ \frac{d\Phi(x)}{dx} \neq 0 \\ \\ 0 & otherwise \end{cases}$$

Note that

$$f_X\left[\Phi^{-1}(y)\right] \cdot \left|\frac{d\Phi^{-1}(y)}{dy}\right| = f_X\left[\, w\,(y)\right] \cdot \left|\frac{d\,w\,(y)}{dy}\right| = f_X\ [\, w\ (y)\,] \cdot |\, w\,'(y)\,|$$

*Proof.* Consider the digram in the figure below:.




y = $\Phi(x)$ is an increasing function

As can be seen from the figure, each point on the $y$ axis maps into a point on the $x$ axis, that is, $X$ must take on a value between $\Phi^{-1}(a)$ and $\Phi^{-1}(b)$ when $Y$ takes on a value between $a$ and $b$. Therefore

$$\begin{aligned} P(a < Y < b) &= P\left(\Phi^{-1}(a) < X < \Phi^{-1}(b)\right) \\ &= \int_{\Phi^{-1}(a)}^{\Phi^{-1}(b)} f_X(x)\,dx \end{aligned} \tag{3}$$

What we would like to do is replace $x$ in the second line with $y$, and $\Phi^{-1}(a)$ and $\Phi^{-1}(b)$ with $a$ and $b$. To do so we need to make a change of variable. Consider how we make a $u$ substitution when we perform integration or use the chain rule for differentiation. For example if $u = h(x)$ then

$du = h'(x)dx$. So if $x = \Phi^{-1}(y)$, then

$$dx = \frac{d\,\Phi^{-1}(y)}{dy}\,dy.$$

Then we can write

$$\int f_X(x)\,dx = \int f_X\left(\Phi^{-1}(y)\right)\frac{d\,\Phi^{-1}(y)}{dy}\,dy \tag{4}$$

For the case of a definite integral the following lemma applies.

**Lemma 1.** *If the function $u = h(x)$ has a continuous derivative on the closed interval $[a, b]$ and $f$ is continuous on the range of $h$, then*

$$\int_a^b f(h(x))\,h'(x)\,dx = \int_{h(a)}^{h(b)} f(u)\,du$$

Using this lemma or the intuition from equation 4 we can then rewrite equation 3 as follows

$$\begin{aligned} P(a < Y < b) &= P\left(\Phi^{-1}(a) < X < \Phi^{-1}(b)\right) \\ &= \int_{\Phi^{-1}(a)}^{\Phi^{-1}(b)} f_X(x)\,dx \\ &= \int_a^b f_X\left(\Phi^{-1}(y)\right)\frac{d\,\Phi^{-1}(y)}{dy}\,dy \end{aligned} \tag{5}$$

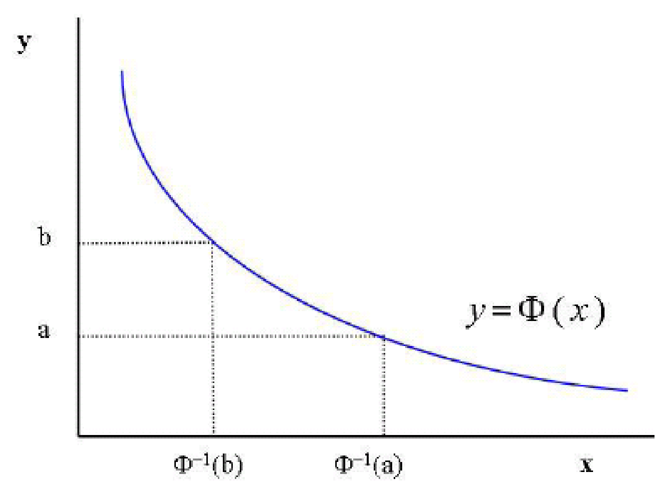
Recall that the probability density function, $f_Y(y)$, of a continuous random variable $Y$ is the function $f(\cdot)$ that satisfies

$$P(a < Y < b) = F(b) - F(a) = \int_a^b f_Y(t)\,dt$$

This then implies that the integrand in equation 5 is the density of $Y$, i.e., $g(y)$, so we obtain

$$g(y) = f_Y(y) = f_X\left[\Phi^{-1}(y)\right]\cdot\frac{d\,\Phi^{-1}(y)}{dy}$$

as long as $\frac{d\,\Phi^{-1}(y)}{dy}$ exists. This proves the lemma if $\Phi$ is an increasing function. Now consider the case where $\Phi$ is a decreasing function as the figure below:



y = $\Phi(x)$ is a decreasing function.

As can be seen from the figure, each point on the $y$ axis maps into a point on the $x$ axis, that is, $X$ must take on a value between $\Phi^{-1}(a)$ and $\Phi^{-1}(b)$ when $Y$ takes on a value between $a$ and $b$. Therefore

$$\begin{aligned} P(a < Y < b) &= P\left(\Phi^{-1}(b) < X < \Phi^{-1}(a)\right) \\ &= \int_{\Phi^{-1}(b)}^{\Phi^{-1}(a)} f_X(x)\,dx \end{aligned}$$

Making a change of variable for $x = \Phi^{-1}(y)$ as before we can write

$$\begin{aligned} P(a < Y < b) &= P\left(\Phi^{-1}(b) < X < \Phi^{-1}(a)\right) \\ &= \int_{\Phi^{-1}(b)}^{\Phi^{-1}(a)} f_X(x)\,dx \\ &= \int_b^a f_X\left(\Phi^{-1}(y)\right) \frac{d\,\Phi^{-1}(y)}{dy}\,dy \\ &= -\int_a^b f_X\left(\Phi^{-1}(y)\right) \frac{d\,\Phi^{-1}(y)}{dy}\,dy \\ &= \int_a^b f_X\left(\Phi^{-1}(y)\right) \left(-\frac{d\,\Phi^{-1}(y)}{dy}\right) dy \end{aligned}$$

Because

$$\frac{d\,\Phi^{-1}(y)}{dy} = \frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} > 0$$

when the function $y = \Phi(x)$ is increasing and

$$-\frac{d\,\Phi^{-1}(y)}{dy}$$

is positive when $y = \Phi(x)$ is decreasing, we can combine the two cases by writing

$$g(y) = f_Y(y) = f_X\left[\Phi^{-1}(y)\right] \cdot \left|\frac{d\,\Phi^{-1}(y)}{dy}\right|$$

□

3.5. **Examples.**

3.5.1. *Example 1.* Let $X$ have the probability density function given by

$$f_X(x) = \begin{cases} \frac{1}{2}\,x, & 0 \le x \le 2 \\ 0, & elsewhere \end{cases}$$

Find the density function of $Y = \Phi(X) = 6X - 3$.
Notice that $f_X(x)$ is positive for all $x$ such that $0 \le x \le 2$. The function $\Phi$ is increasing for all $X$. We can then find the inverse function $\Phi^{-1}$ as follows

$$\begin{aligned} y &= 6x - 3 \\ \Rightarrow 6x &= y + 3 \\ \Rightarrow x &= \frac{y+3}{6} = \Phi^{-1}(y) \end{aligned}$$

We can then find the derivative of $\Phi^{-1}$ with respect to $y$ as

$$\frac{d\Phi^{-1}}{dy} = \frac{d}{dy}\left(\frac{y + 3}{6}\right) = \frac{1}{6}$$

The density of $Y$ is then

$$\begin{aligned} g(y) = \quad & f_Y(y) = f_X\left[\Phi^{-1}(y)\right] \cdot \left|\frac{d\Phi^{-1}(y)}{dy}\right| \\ = \quad & \left(\frac{1}{2}\right)\left(\frac{3 + y}{6}\right)\left|\frac{1}{6}\right|, \quad 0 \leq \frac{3 + y}{6} \leq 2 \end{aligned}$$

For all other values of $y$, $g(y) = 0$. Simplifying the density and the bounds we obtain

$$g(y) = f_Y y) = \begin{cases} \frac{3 + y}{72}, & -3 \leq y \leq 9 \\ 0 & elsewhere \end{cases}$$

3.5.2. *Example 2.* Let $X$ have the probability density function given by $f_X(x)$. Then consider the transformation $Y = \Phi(X) = \sigma X + \mu$, $\sigma \neq 0$. The function $\Phi$ is increasing for all $X$. We can then find the inverse function $\Phi^{-1}$ as follows

$$\begin{aligned} y = \quad & \sigma x + \mu \\ \Rightarrow \sigma x = \quad & y - \mu \\ \Rightarrow x = \quad & \frac{y - \mu}{\sigma} = \Phi^{-1}(y) \end{aligned}$$

We can then find the derivative of $\Phi^{-1}$ with respect to $y$ as

$$\frac{d\Phi^{-1}}{dy} = \frac{d}{dy}\left(\frac{y - \mu}{\sigma}\right) = \frac{1}{\sigma}$$

The density of $Y$ is then

$$f_Y(y) = f_X\left[\Phi^{-1}(y)\right] \cdot \left|\frac{d\,\Phi^{-1}(y)}{dy}\right| = f_X\left[\frac{y - \mu}{\sigma}\right] \cdot \left|\frac{1}{\sigma}\right|$$

3.5.3. *Example 3.* Let $X$ have the probability density function given by

$$f_X(x) = \begin{cases} e^{-x}, & 0 \leq x \leq \infty \\ 0, & elsewhere \end{cases}$$

Find the density function of $Y = X^{1/2}$.
Notice that $f_X(x)$ is positive for all $x$ such that $0 \leq x \leq \infty$. The function $\Phi$ is increasing for all $X$. We can then find the inverse function $\Phi^{-1}$ as follows

$$\begin{aligned} y = \quad & x^{\frac{1}{2}} \\ \Rightarrow y^2 = \quad & x \\ \Rightarrow x = \quad & \Phi^{-1}(y) = y^2 \end{aligned}$$

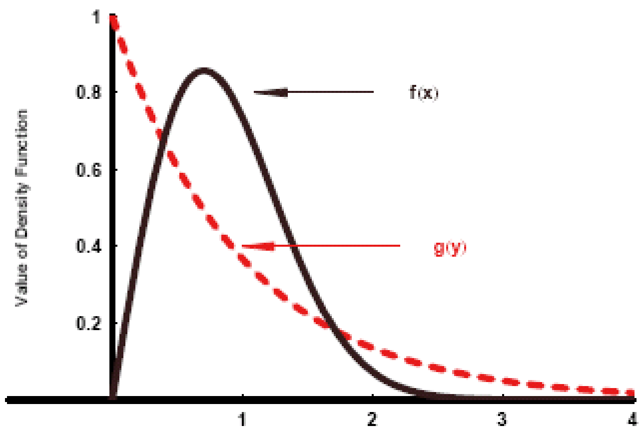
We can then find the derivative of $\Phi^{-1}$ with respect to $y$ as

$$\frac{d\,\Phi^{-1}}{dy} = \frac{d}{dy}\,y^2 = 2y$$

The density of $Y$ is then

$$f_Y(y) \; = \; f_X \left[ \, \Phi^{-1}(y) \, \right] \; \cdot \; \left| \frac{d\Phi^{-1}(y)}{dy} \right| = \; e^{-y^2} \; |2y|$$

A graph of the two density functions is shown in the figure below:




The Two Density Functions.

## 4. METHOD OF TRANSFORMATIONS (MULTIPLE VARIABLES)

4.1. **General definition of a transformation.** Let $\Phi$ be any function from $R^k$ to $R^m$, $k, m \geq 1$, such that $\Phi^{-1}(A) = \{x \epsilon R^k : \Phi(x) \epsilon A\}$. If we write $y = \Phi(x)$, the function $\Phi$ defines a mapping from the sample space of the variable $X (\Xi)$ to a sample space of the random variable $Y \; (\Psi)$. Specifically

$$\Phi \, ( \, x \, ) \; : \; \Xi \; \rightarrow \; \Psi$$

and

$$\Phi^{-1} \, (A) \; = \; \{ \, x \, \epsilon \, \Psi \; : \; \Phi \, (x) \, \epsilon \, A \, \}$$

4.2. **Transformations involving multiple functions of multiple random variables.**

**Theorem 3.** *Let $f_{X_1 \, X_2} \, (x_1 \;\, x_2)$ be the value of the joint probability density of the continuous random variables $X_1$ and $X_2$ at $(x_1, x_2)$. If the functions given by $y_1 = u_1(x_1, x_2)$ and $y_2 = u_2(x_1, x_2)$ are partially differentiable with respect to $x_1$ and $x_2$ and represent a one-to-one transformation for all values within the range of $X_1$ and $X_2$ for which $f_{X_1 \, X_2} \, (x_1 \;\, x_2) \;\; \neq \;\; 0$, then, for these values of $x_1$ and $x_2$, the equations $y_1 \; = \; u_1(x_1, x_2)$ and $y_2 \; = \; u_2(x_1, x_2)$ can be uniquely solved for $x_1$ and $x_2$ to give $x_1 \; = \; w_1(y_1, \; y_2)$ and $x_2 = w_2(y_1, \; y_2)$ and for corresponding values of $y_1$ and $y_2$, the joint probability density of $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$ is given by*

$$f_{Y_1 \, Y_2} \, ( \, y_1 \, y_2 ) \;\; = \;\; f_{X_1 \;\, X_2} \;\; [ \, w_1 \; ( \, y_1 \, , \;\, y_2 ) \, , \;\, w_2 \, ( \, y_1 \, y_2 ) \; ] \; \cdot \; | \; J \; |$$

*where $J$ is the Jacobian of the transformation and is defined as the determinant*

$$J \;\; = \;\; \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

*At all other points $f_{Y_1 Y_2}(y_1 \; y_2) \;\; = \;\; 0$.*

4.3. **Example.** Let the probability density function of $X_1$ and $X_2$ be given by

$$f_{X_1 X_2}(x_1, x_2) = \begin{cases} e^{-(x_1 + x_2)} & x_1 \geq 0,\ x_2 \geq 0 \\ 0 & elsewhere \end{cases}$$

Consider two random variables $Y_1$ and $Y_2$, defined in the following manner.

$$Y_1 = X_1 + X_2 \tag{6}$$

$$Y_2 = \frac{X_1}{X_1 + X_2} \tag{7}$$

To find the joint density of $Y_1$ and $Y_2$ we first need to solve the system of equations in 6 and 7 for $X_1$ and $X_2$. Equation 6 provides

$$X_1 = Y_1 - X_2 \tag{8}$$

Plugging this into 7, we get

$$\begin{aligned} Y_2 &= \frac{Y_1 - X_2}{Y_1 - X_2 + X_2} = \frac{Y_1 - X_2}{Y_1} \Leftrightarrow \\ Y_1\, Y_2 &= Y_1 - X_2 \Leftrightarrow \\ X_2 &= Y_1 - Y_1\, Y_2 = Y_1\,(1 - Y_2) \end{aligned}$$

Plugging this into 8, we obtain

$$X_1 = Y_1 - X_2 = Y_1 - (Y_1 - Y_1\, Y_2) = Y_1\, Y_2$$

The Jacobian is given by

$$\begin{aligned} J &= \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} y_2 & y_1 \\ 1 - y_2 & -y_1 \end{vmatrix} \\ &= -y_2\, y_1 - y_1\,(1 - y_2) \\ &= -y_2\, y_1 - y_1 + y_1\, y_2 = -y_1 \end{aligned}$$

This transformation is one-to-one and maps the domain of $X(\Xi)$ given by $x_1 > 0$ and $x_2 > 0$ in the $x_1\, x_2$-plane into the domain of $Y(\Psi)$ in the $y_1\, y_2$-plane given by $y_1 > 0$ and $0 < y_2 < 1$. If we apply theorem 3 we obtain

$$\begin{aligned} f_{Y_1 Y_2}(y_1\, y_2) &= f_{X_1\, X_2}\,[\,w_1\,(y_1\,,\, y_2),\, w_2\,(\,y_1\, y_2\,)\,] \cdot |\ J\ | \\ &= e^{-\,(\,y_1\, y_2\, +\, y_1\, -\, y_1\, y_2\,)}\ |-y_1\,| \\ &= e^{-\,y_1}\ |-y_1\,| = y_1\, e^{-\,y_1} \end{aligned}$$

Considering all possible values of values of $y_1$ and $y_2$ we obtain

$$f_{Y_1\, Y_2}(y_1, y_2) = \begin{cases} y_1 e^{-y_1} & for\ y_1 \geq 0\,,\ 0 < y_2 < 1 \\ 0 & elsewhere \end{cases}$$

We can then find the marginal density of $Y_2$ by integrating over $y_1$ as follows

$$f_{Y_2}(y_1,\, y_2) = \int_0^\infty f_{Y_1\, Y_2}\,(y_1,\, y_2)\, dy_1 = \int_0^\infty y_1\ e^{-\,y_1}\ dy_1$$

We make a $uv$ substitution to integrate. $u$, $v$, $du$, and $dv$ are defined as

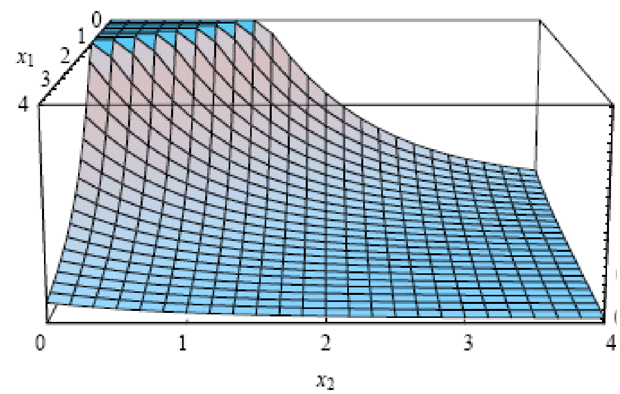
$$\begin{aligned} u &= y_1 & v &= -e^{-y_1} \\ du &= dy_1 & dv &= e^{-y_1} dy_1 \end{aligned}$$

This then implies

$$\begin{aligned} f_{Y_2}(y_1,\, y_2) &= \int_0^\infty y_1\, e^{-y_1} dy_1 = -y_1\, e^{-y_1}\, |_0^\infty - \int_0^\infty -e^{-y_1}\, d\,y_1 \\ &= (0 - 0) - \left(e^{-y_1}\right)\Big|_0^\infty = 0 - \left(e^{-\infty} - e^0\right) \\ &= 0 - 0 + 1 = 1 \end{aligned}$$

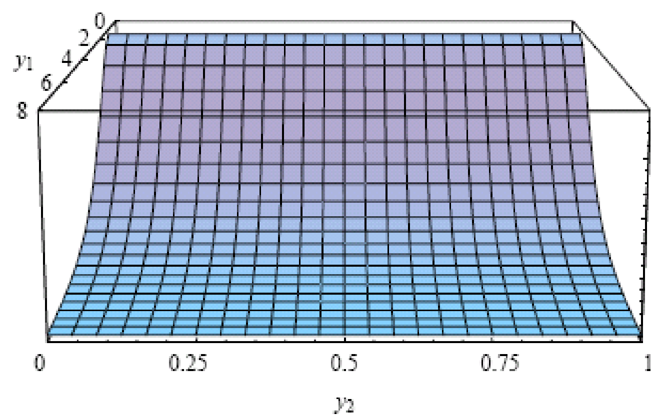
for all $y_2$ such that $0 < y_2 < 1$.

A graph of the joint densities and the marginal density follows. The joint density of $(X_1, X_2)$ is shown in the first figure



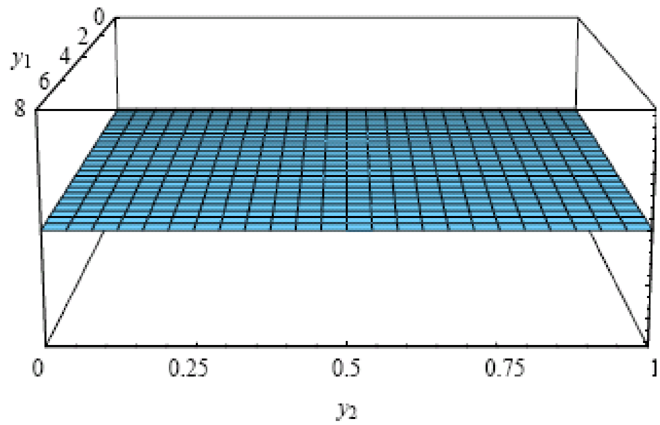

Joint Density of $X_1$ and $X_2$.

The joint density of $(Y_1, Y_2)$ is depicted in the following figure:




Joint Density of $Y_1$ and $Y_2$.

The marginal density of $Y_2$ is shown graphically in this figure:




Marginal Density of $Y_2$.

## REFERENCES

[1] Billingsley, P. *Probability and Measure.* 3rd edition. New York: Wiley, 1995
[2] Casella, G. and R.L. Berger. *Statistical Inference*. Pacific Grove, CA: Duxbury, 2002
[3] Protter, Murray H. and Charles B. Morrey, Jr. *Intermediate Calculus*. New York: Springer-Verlag, 1985