A random variable X has a *probability density function* if there is a function $f : \mathbb{R} \to \mathbb{R}$ so that

$$P(a < X \leq b) = \int_a^b f(x)dx$$

for any $a < b$. A random variable with a density is called *continuous*. We remind the reader that the distribution of a continuous random variable is determined by its density function.

The goal of this discussion is to discuss the determination of the distribution of the random variable $Y = h(X)$, where h is a differentiable function and X is a continuous random variable with density function f.

We first consider the case where the following hold:

1. $h : D \to R$ is differentiable on its domain D, which is an open interval of $\mathbb{R}$.
2. h is one-to-one, which means that $h(x) = h(y)$ implies that $x = y$.
3. The support of X, defined as

$$\operatorname{supp}(f) \stackrel{\text{def}}{=} \{x : f(x) > 0\} \tag{1}$$

   is contained in D.

For any set $S \subset \mathbb{R}$, define $h(S) = \{y : y = f(x) \text{ for some } x \in D\}$.

Theorem 1. *Suppose that the conditions above hold. Then $Y = h(X)$ is a continuous random variable with density function*

$$f_Y(y) = f_X(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right| \mathbf{1}\{y \in h(D)\}, \tag{2}$$

for all points y so that f_X is continuous at $h^{-1}(y)$.

Proof. Since h is one-to-one, it is either always increasing or always decreasing on D. Assume that h is increasing, the other case is similar. We begin by computing the distribution function of Y:

$$\begin{aligned}
P(Y \leq y) &= P(h(X) \leq y) \\
&= P\big(X \leq h^{-1}(y)\big) \\
&= \int_{-\infty}^{h^{-1}(y)} f_X(x)dx \\
&= \int_{-\infty}^{y} f_X(h^{-1}(z)) \frac{d}{dy} h^{-1}(z)dz && \text{by change of variables} \\
&= \int_{-\infty}^{y} f_X(h^{-1}(z)) \left| \frac{d}{dy} h^{-1}(z) \right| dz && \text{since } h \text{ is increasing}.
\end{aligned}$$

Now suppose that y is such that $h^{-1}(y)$ is a point of continuity for f_X. Then by the Fundamental Theorem of Calculus, F_Y is differentiable at y with derivative

$$f_Y(y) = \frac{d}{dy}F_Y(y) = f_X(h^{-1}(y))\left|\frac{d}{dy}h^{-1}(y)\right|.$$

$\square$

Now we consider the case where h is not one-to-one. We will assume the following: For each $y \in h(D)$, the set $h^{-1}(\{y\}) = \{x : h(x) = y\}$ is a finite set.

We recall the following theorem from calculus:

Theorem (Inverse Function Theorem)**.** *Let $h : D \to R$ be a differentiable function. Let $y = f(x)$ for some $x \in D$. Suppose that $f'(x) \neq 0$. Then there is an open interval I containing x and an open interval J containing y, so that h restricted to I is one-to-one, and there is a differentiable inverse $h_x^{-1} : J \to I$.*

Thus, for each $x \in h^{-1}(\{y\})$, there is a function g_x defined in a neighborhood of x so that $h \circ g_i(y') = y'$ for all y' in a neighborhood of y, and $g_i \circ h(x') = x'$ for all x' in a neighborhood of x.

Assume that for each $x \in h^{-1}(\{y\})$, we have $h'(x) \neq 0$. Now, since $h^{-1}(\{y\})$ is finite, say equal to $\{x_1, \ldots, x_r\}$, letting $g_i = g_{x_i}$ there is an interval J containing y on which each of the g_i is defined. We can take J small enough so that $\{g_i(J)\}$ are disjoint intervals.

We have for $a \leq y \leq b$ with $a < b$ and $a, b \in J$,

$$\begin{aligned}
P(a \leq Y \leq b) &= P\left(\bigcup_{x \in h^{-1}(\{y\})} \{X \in g_x([a,b])\}\right) \\
&= \sum_{x \in h^{-1}(\{y\})} P(X \in g_x([a,b])) \\
&= \sum_{x \in h^{-1}(\{y\})} \int_{g_x([a,b])} f(u)du \\
&= \sum_{x \in h^{-1}(\{y\})} \int_a^b f(g_x(s))\left|g_x'(s)\right| ds \\
&= \int_a^b \sum_{x \in h^{-1}(\{y\})} f(g_x(s))\left|g_x'(s)\right| ds
\end{aligned}$$

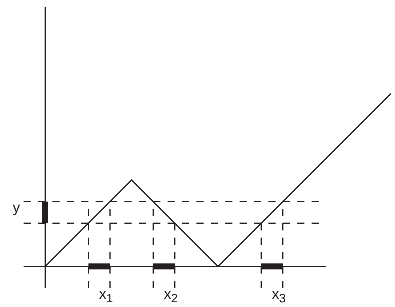

FIGURE 1. A many-to-one function

Then taking $a = y$ and $b = y + \Delta y$, differentiating with respect to Δy, and evaluating at $\Delta y = 0$ yields

$$f_Y(y) = \sum_{x \in h^{-1}(\{y\})} f(g_x(s)) \left| g_x'(s) \right| . \tag{3}$$

Figure 1 shows a many-to-one function. Note how a little neighborhood around y maps to neighborhoods surrounding the three points in $h^{-1}(\{y\})$. For this y, the sum in (3) will have three terms.

Let us consider an example. Suppose that X has an exponential(1) distribution, and let

$$h(x) = \begin{cases} 3x & \text{if } 0 < x \leq \frac{1}{3} \\ 1 - 5\left(x - \frac{1}{3}\right) & \text{if } \frac{1}{3} < x < \frac{8}{15} \\ 2\left(x - \frac{8}{15}\right) & \text{if } x \geq \frac{8}{15} . \end{cases}$$

The reader should graph this function. Let $0 < y < 1$. The there are three x so that $h(x) = y$. Namely,

$$\begin{aligned} x &= \frac{1}{3} y \\ x &= -\frac{1}{5} y + \frac{4}{15} \\ x &= \frac{1}{2} y + \frac{8}{15} . \end{aligned}$$

The three functions on the right of the above equation are then $g_1(x), g_2(x)$ and $g_3(x)$. Thus we have

$$f_Y(y) = \frac{e^{-y/3}}{3} + \frac{e^{y/5 - 4/15}}{5} + \frac{e^{-y/2 - 8/15}}{2} .$$

Here is another example. Suppose that X has the density

$$f(x) = \frac{x}{2\pi^2} \mathbf{1}\{0 < x < 2\pi\} ,$$

and consider the random variable $Y = \sin X$. We will now find the density of Y.

First take $y > 0$. Then

$$\sin^{-1}(\{y\}) \cap (0, 2\pi) = \{\arcsin(y), \arcsin(y) + \pi/2\}.$$

This follows since, by convention, $\arcsin(y)$ is defined to take values in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ for $y \in [-1, 1]$.

Thus using the notation as above, we have

$$g_1(y) = \arcsin(y),$$
$$g_2(y) = \arcsin(y) + \frac{\pi}{2}.$$

Thus we have

$$f_Y(y) = \frac{\arcsin(y)}{2\pi^2} \frac{1}{\sqrt{1-y^2}} + \frac{\arcsin(y) + \frac{\pi}{2}}{2\pi^2} \frac{1}{\sqrt{1-y^2}}$$
$$= \frac{\arcsin(y) + \frac{\pi}{4}}{\pi^2 \sqrt{1-y^2}}.$$

Let us review some facts from multivariate calculus. Let $h : D \to R$ be a one-to-one function, where $D \subset \mathbb{R}^n$ and $R \subset \mathbb{R}^n$. We write

$$h(x_1, \ldots, x_n) = (h_1(x_1, \ldots, x_n), \ldots, h_n(x_1, \ldots, x_n)).$$

The *total derivative* of h at $x = (x_1, \ldots, x_n)$ is defined as the matrix

$$Dh(x) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1}(x) & \frac{\partial h_1}{\partial x_2}(x) & \cdots & \frac{\partial h_1}{\partial x_n}(x) \\ \frac{\partial h_2}{\partial x_1}(x) & \frac{\partial h_2}{\partial x_2}(x) & \cdots & \frac{\partial h_2}{\partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial x_1}(x) & \frac{\partial h_n}{\partial x_2}(x) & \cdots & \frac{\partial h_n}{\partial x_n}(x) \end{bmatrix}$$

The *Jacobian* of h at x, which we will denote by $J_h(x)$ is defined as

$$J_h(x) \stackrel{\text{def}}{=} \det Dh(x). \tag{4}$$

Now the change of variables formula says the following: Let $h : D \to R$ be a function which has continuous first partial derivatives. Then

$$\int_E f(y) dy = \int_{h^{-1}(E)} f(h(x)) |J_h(x)| dx.$$

Now we can state the formula for finding the density of $Y = h(X)$, where X is a random vector in $\mathbb{R}^n$, and h is a one-to-one function defined on $D \subset \mathbb{R}^n$, where the support of f_X is contained in D.

Theorem 2. *Let h be as above, and let X be a continuous random vector in $\mathbb{R}^n$ with density f_X. Then the density of Y is given by*

$$f_Y(y) = f_X(h^{-1}(y))|J_{h^{-1}}(y)|.$$

A fact which is often very useful is that

$$|J_{h^{-1}}(y)| = \frac{1}{|J_h(h^{-1}(y))|}. \tag{5}$$

Let X, Y be independent standard Normal random variables. Let

$$(D, \Theta) = h(X, Y) = (X^2 + Y^2, \arctan(Y/X)).$$

Then

$$h^{-1}(d, \theta) = (\sqrt{d}\cos\theta, \sqrt{d}\sin\theta).$$

We have

$$Dh^{-1}(d, \theta) = \begin{bmatrix} \frac{1}{2\sqrt{d}}\cos\theta & -\sqrt{d}\sin\theta \\ \frac{1}{2\sqrt{d}}\sin\theta & \sqrt{d}\cos\theta \end{bmatrix}$$

Thus

$$|J_{h^{-1}}(d, \theta)| = \frac{1}{2}\cos^2\theta + \frac{1}{2}\sin^2\theta = \frac{1}{2}.$$

Then we have for $d > 0$ and $\theta \in [0, 2\pi)$:

$$f_{D,\Theta}(d, \theta) = \frac{1}{2\pi} e^{-\frac{1}{2}\left(d\cos^2\theta + d\sin^2\theta\right)} \frac{1}{2} = \frac{1}{2} e^{-\frac{d}{2}} \frac{1}{2\pi}$$

This shows that D and Θ are independent, and D is exponential$(1/2)$, and Θ is Uniform$[0, 2\pi)$. [Why?]

Note we could run this in reverse: Suppose we start with D an exponential$(1/2)$ random variable, and Θ an independent Uniform$[0, 2\pi)$ random variable. Then let

$$g(d, \theta) = (\sqrt{d}\cos\theta, \sqrt{d}\sin\theta).$$

Then $g^{-1}(x, y) = h(x, y)$, where h is defined as above. Now finding the Jacobian of g^{-1} itself can be done, but it is perhaps easier to use (5):

$$J_{g^{-1}}(x, y) = J_h(x, y) = \frac{1}{J_{h^{-1}}(h(x, y))} = \frac{1}{1/2} = 2.$$

Thus,

$$f_{X,Y}(x, y) = \frac{1}{2} e^{-(x^2+y^2)/2} \frac{1}{2\pi} 2 = \frac{1}{2\pi} e^{-(x^2+y^2)/2}.$$

Thus $g(D, \Theta)$ gives a pair of independent standard normal random variables. [Why?]

This gives a method of simulating a pair of Normal random variable. It is relatively easy to simulate a uniform and an exponential random variable. Then applying the function g to them gives a pair of Normal random variables.