Variable Transformation to Generate New Distributions

Rahul Bhadani

May 24, 2022

1 Introduction

Often we are required to generate new distribution or density function with closed forms from a given distribution. Programmatically, it might be easy, but sometimes looking beyond just numbers is required to obtain specific parameters for new distributions such as mean, standard deviation, moment generating functions, etc. Hence, knowing the methodology of generating new distribution by the transformation of a random variable is significant. In this article, we will look at the transformation of a random variable to create a new distribution from a continuous distribution given the transformation function.

2 Functions of Random Variable

Consider X to be a random variable with a continuous probability density function (pdf) f_X with sample space $\mathcal{X}$. We can have a function g applied to X so that we can obtain a new random variable Y, i.e. $Y = g(X)$. In such a case, a natural case to ask is how the distribution of Y related to the distribution of X.

Consider the following case:

$$\mathcal{X} = \{x : f_X(x) \geq 0\}, \quad \mathcal{Y} = \{y : y = g(x) \text{ for some } x \in \mathcal{X}\} \tag{1}$$

Here, $f_X(x) \geq 0$ is the support set of X. From the first principle, we can write the cumulative distribution function (CDF) of Y as:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(g(X) \leq y) = P(x \in \mathcal{X} : g(x) \leq y) \\ &= \int_{\{x \in \mathcal{X} : g(x) \leq y\}} f_X(x) dx \\ &= F_X(g^{-1}(y)) \end{aligned} \tag{2}$$

The density function of Y can be derived using the chain rule of differentiation as (considering g to be monotonically increasing):

$$f_Y(y) = F_Y'(y) = \frac{d}{dy} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) \tag{3}$$

For monotonically decreasing g, we have (derivation omitted):

$$F_Y(y) = 1 - F_X(g^{-1}(y)) \tag{4}$$

and hence,

$$f_Y(y) = -f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y) \tag{5}$$

Thus, the overall rule of obtaining density function for Y can be written as:

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|, & y \in \mathcal{Y} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

2.1 Example

Consider a random variable X with a uniform distribution with a transformation on Y:

$$\begin{aligned} X &\sim \mathrm{U}[0, 2] \\ Y &= X^3 \end{aligned} \tag{7}$$

Uniform distribution $U[a, b]$ is defined as:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

In this case, uniform distribution's density function looks like:

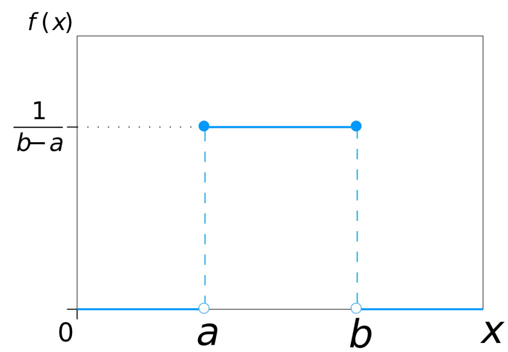

Figure 1: The density function of uniform distribution.

Let's look at the density function plot as obtained from simulation in Python (see code at `https://gist.github.com/rahulbhadani/17e25e80195a5c1f2016d7b64e5cd280`):

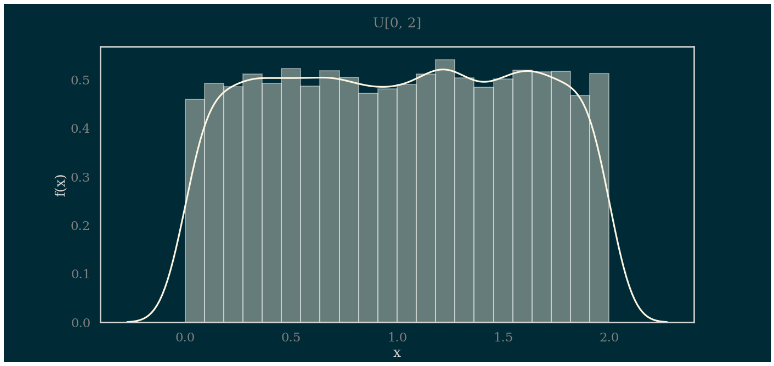

Figure 2: Simulation of Uniform Distribution U[0, 2].

From Figure 2, we see that randomly generated numbers sampled from a uniform distribution resemble the density function shown in Figure 1.

Now, a naive understand thinking for density function of $Y = X^3$ would be that amplitude in Figure 1 should be (1/8) instead of (1/2). However, this is not true. This can be easily validated by plotting density function using `displot(A**3)` (see the Code Snippet from the link provided) which gives Figure 3.

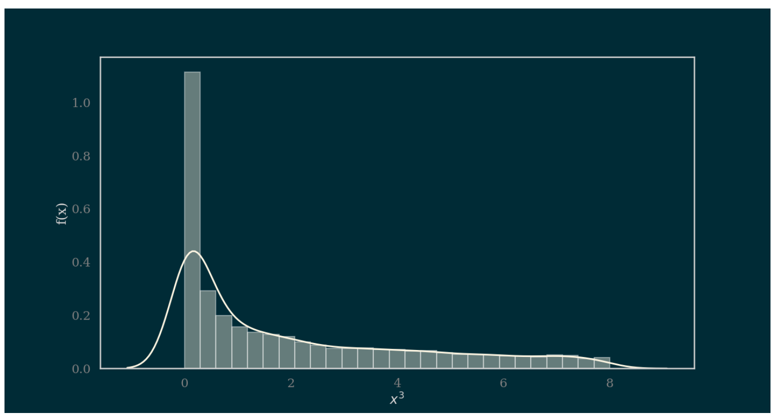

Figure 3: Density plot with $Y = X^3$.

As you can see, the density plot of the transformed random variable is not at all similar to the original density. One can obtain the density function of transformed random variable can be obtained using Equation (6) as detailed earlier.

2.1.1 Obtaining Density Function for Example

As $Y = X^3$, we can write

$$
\begin{aligned}
y &= x^3 \\
x &= y^{1/3} \\
g^{-1}(y) &= y^{1/3} \\
\frac{d}{dy} g^{-1}(y) &= \frac{1}{3} y^{-2/3}
\end{aligned} \tag{9}
$$

and we don't need to consider $|\cdot|$ as in the range of $[0, 2]$, x^3 is monotonically increasing. Thus,

$$
\begin{aligned}
f_Y(y) &= \begin{cases} 0 \cdot \frac{1}{3} y^{-2/3}, & \sqrt[3]{y} < a \\ \frac{1}{2} \cdot \frac{1}{3} y^{-2/3}, & a \leq \sqrt[3]{y} \leq b \\ 0 \cdot \frac{1}{3} y^{-2/3}, & b < \sqrt[3]{y} \end{cases} \\
&= \begin{cases} 0, & \sqrt[3]{y} < a \\ \frac{1}{6} y^{-2/3}, & a \leq \sqrt[3]{y} \leq b \\ 0, & b < \sqrt[3]{y} \end{cases}
\end{aligned} \tag{10}
$$

We can simulate the density function from Equation (10) using scipy package's `rv_continuous` module. The code-snippet is provided in `https://gist.github.com/rahulbhadani/3f62b0ae49cc55cbb31c5f0160b1a6a6`. The code generates a density plot similar to the one shown in Figure 3 as you can see in Figure 4.

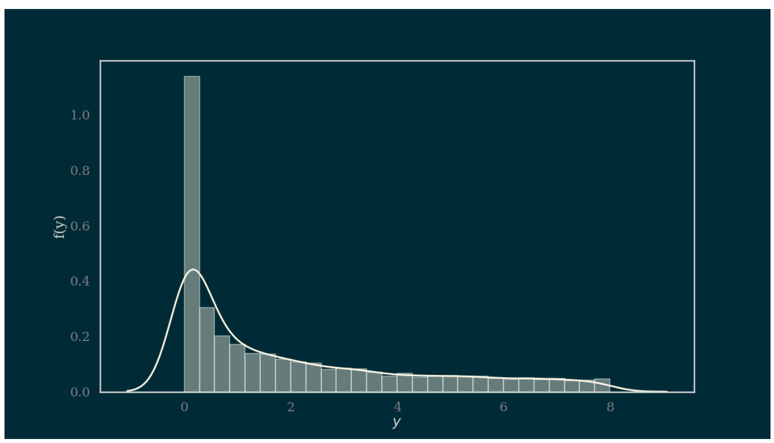

Figure 4: Density plot with y from Equation (10).

As you can see after applying transformation and deriving density from Equation 10, we arrive at the same conclusion. However, with a closed-form density function, we can also obtain mean, variance, moment generating function, survival functions, etc. A similar technique can be used to generate data from a more complicated distribution given a simpler distribution.