
Transformations of Random Variables: In this lecture we are going to focus on finding the probability distribution of functions of random variables. That is, if X is a random variable defined on $\mathbb{S}_X$, then the probability distribution of X can be used to obtain the probability distribution of $Y = g(X)$, where $g(\cdot)$ is some known function. For example, if we know the probability distribution of the random variable X, we would like know the distribution of $Y = \ln(X)$. Similarly, if $X_1, X_2, ..., X_k$ are k random variables defined on some k-dimensional space, then the distribution of $g(X_1, X_2, ..., X_k)$ can be identified using the joint distribution of $X_1, X_2, ..., X_k$. There are several methods for finding the probability density function of transformed random variables. We will be focusing on three transformation methods in this course.

1- The CDF method: This method is used to obtain the distribution of a function of a single random variable (*univariate*). Where the CDF of $Y = g(X)$ is derived using the CDF of X. Then the density of Y is determined by differentiating the CDF of Y. That is, if $F_X(x)$ is the CDF of X, and $Y = g(X)$, then the CDF of Y is defined as

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) \\ &= F_X(g^{-1}(y)) \end{aligned}$$

then the PDF of Y is $f_Y(y) = \frac{dF_Y(y)}{dy}$. And the support of Y is calculated based on the support of X and the transformation function, $g(X)$.

Example 7.1: Let $X \sim Uniform(a, b)$. Find the PDF of $Y = \frac{X}{a}$ using the CDF method.

Solution: Here we have

$$F_X(x) = \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a}, \qquad \text{for } x \in [a, b]$$

Hence

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P\left(\frac{X}{a} \leq y\right) \\ &= P\left(X \leq ay\right) \\ &= F_X(ay) \\ &= \frac{ay-a}{b-a} = \frac{y-1}{\frac{b}{a}-1} \end{aligned}$$

and

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy} \\ &= \frac{d}{dy}\left(\frac{y-1}{\frac{b}{a}-1}\right) = \frac{1}{\frac{b}{a}-1} \end{aligned}$$

if $X = a \implies Y = \frac{a}{a} = 1$, and if $X = b \implies Y = \frac{b}{a}$. Since $b > a$, therefore $\frac{b}{a} > 1$. Thus, the PDF of Y is

$$f_Y(y) = \begin{cases} \frac{1}{\frac{b}{a}-1} & \text{for } y \in \left[1, \frac{b}{a}\right] \\ 0 & \text{otherwise} \end{cases}$$

What is the distribution of Y?

Example 7.2: Let X be a random variable with the following PDF:

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{for } -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

what is the probability density function $W = X^2$?

Solution: Notice that $X \sim Uniform(-1, 1)$, i.e. $a = -1$ and $b = 1$. Therefore the CDF of X is

$$F_X(x) = \int_{-1}^{x} \frac{1}{2} dx = \frac{x+1}{2}, \qquad \text{for } x \in [-1, 1]$$

Now, to obtain the density of W, we first need to find the CDF of W. That is

$$
\begin{aligned}
F_W(w) &= P(W \leq w) \\
&= P(X^2 \leq w) \\
&= P(-\sqrt{w} \leq X \leq \sqrt{w}) \\
&= F_X(\sqrt{w}) - F_X(-\sqrt{w}) \\
&= \frac{\sqrt{w}+1}{2} - \frac{-\sqrt{w}+1}{2} \\
&= \sqrt{w}
\end{aligned}
$$

notice that if $X = \pm 1 \implies W = (\pm 1)^2 = 1$, and if $X = 0 \implies W = 0$.

Hence

$$
F_W(w) = \begin{cases} \sqrt{w} & \text{for } 0 \leq w \leq 1 \\ 1 & \text{for } w \geq 1 \\ 0 & \text{otherwise} \end{cases}
$$

Now the PDF of W is

$$
\begin{aligned}
f_W(w) &= \frac{dF_W(w)}{dw} \\
&= \frac{d}{dw}\sqrt{w} = \frac{1}{2\sqrt{w}}
\end{aligned}
$$

Thus

$$f_W(w) = \begin{cases} \frac{1}{2\sqrt{w}} & \text{for } 0 \leq w \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Can you name the distribution W?

Example 7.3: Let $X \sim Uniform(0,1)$. If $Y = -\frac{\ln X}{\theta}$, for $\theta > 0$. Find the PDF of Y using the CDF method.

Solution: Here we have

$$F_X(x) = \int_0^x 1dx = x, \qquad \text{for } x \in [0,1]$$

Hence

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P\left(-\frac{\ln X}{\theta} \leq y\right) \\ &= P\left(-\ln X \leq \theta y\right) \\ &= P\left(\ln X > -\theta y\right), \qquad \text{multiply both sides by -1} \\ &= P\left(X > e^{-\theta y}\right) \\ &= 1 - P\left(X \leq e^{-\theta y}\right) \\ &= 1 - F_X(e^{-\theta y}) \\ &= 1 - e^{-\theta y} \end{aligned}$$

and

$$
\begin{aligned}
f_Y(y) &= \frac{dF_Y(y)}{dy} \\
&= \frac{d}{dy}\left(1 - e^{-\theta y}\right) = \theta e^{-\theta y}
\end{aligned}
$$

if $X = 0 \implies Y = -\frac{\ln(0)}{\theta} = \infty$, and if $X = 1 \implies Y = -\frac{\ln(1)}{\theta} = 0$.

Therefore, the PDF of Y is

$$
f_Y(y) = \begin{cases} \theta e^{-\theta y} & \text{for } y \geq 0 \\ 0 & \text{otherwise} \end{cases}
$$

Notice that the distribution of Y is exponential with rate parameter θ.

Example 7.4 (Homework): Let $X \sim Uniform(0, 1)$. Find the PDF of $Y = X - c$ using the CDF method.

Example 7.5 (Homework): Let $X \sim Uniform(a, b)$. Find the PDF of $Y = \frac{X}{b}$ using the CDF method.

Drawbacks of the CDF method:

1. Limited to Univariate Transformations: The CDF method is primarily applicable to univariate transformations, meaning it can only be used to find the distribution of a single transformed random variable.

2. Existence of CDF: The CDF method requires that the Cumulative Distribution Function of the transformed random variable exists. In cases where the CDF is not well-defined or difficult to obtain, this

method may not be applicable.

Overall, while the CDF method is useful for obtaining the distribution of functions of random variables under certain conditions, its limitations make it important to consider alternative methods depending on the complexity and nature of the transformation being analyzed.

2- Change-of-variable method: This method relies on the change-of-variable technique from calculus. Let's take the following theorem to understand the mechanism of this approach.

Theorem: Let X be a continuous random variable with probability density function $f_X(x)$. And let $Y = g(X)$ be an *increasing* (or *decreasing*) function, i.e. *monotone function*. Then the PDF of the random variable $Y = g(X)$ is given by

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|$$

provided that $g(X)$ is invertible.

Example 7.6: Redo example 7.1 using the change-of-variable approach.

Solution: We have $X \sim Uniform(a, b)$, so $f_X(x) = \frac{1}{b-a}$, for $a \leq x \leq b$). We also have $y = g(x) = \frac{x}{a}$, and hence $x = g^{-1}(y) = ay$, and $\left|\frac{dg^{-1}(y)}{dy}\right| =$

$\left|\frac{dx}{dy}\right| = \left|\frac{d}{dy}(ay)\right| = a$. So the PDF of Y is

$$
\begin{aligned}
f_Y(y) =& f_X(g^{-1}(y)) \left|\frac{dg^{-1}(y)}{dy}\right| \\
&= f_X(ay) \times a \\
&= \frac{1}{b-a} \times a = \frac{1}{\frac{b}{a}-1}, \qquad \text{for } 1 \le y \le \frac{b}{a}
\end{aligned}
$$

meaning that $Y \sim Uniform(1, \frac{b}{a})$.

Example 7.7 (Homework): Apply the change-of-variable method to find the PDF of W in example 7.2, and then check if your result matches the answer given in example 7.2. If they are not the same, why do you think they differ, and which one do you think is correct? Justify your answer.

Example 7.8: Let $X \sim Uniform(0, 1)$. If $Y = -\frac{\ln X}{\theta}$, for $\theta > 0$. Find the distribution of Y using the change-of-variable approach.

Solution: Here we have $f_X(x) = 1$, for $0 \le x \le 1$. We also have $y = -\frac{\ln x}{\theta}$, implies $x = e^{-\theta y}$, so $\left|\frac{dx}{dy}\right| = \left|-\theta e^{-\theta y}\right| = \theta e^{-\theta y}$. Hence the PDF of Y is

$$
\begin{aligned}
f_Y(y) &= f_X(g^{-1}(y)) \left|\frac{dx}{dy}\right| \\
&= 1 \times \theta e^{-\theta y} = \theta e^{-\theta y}, \qquad \text{for } y \ge 0
\end{aligned}
$$

therefore $Y \sim Expo(\theta)$.

Example 7.9: Let $X \sim N(\mu, \sigma^2)$. Use the change-of variable technique to find the distribution of $Z = \frac{X-\mu}{\sigma}$, where $\mu \in \mathbb{R}$ and $\sigma > 0$.

Solution: Here we have $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, for $x \in \mathbb{R}$, $\mu \in \mathbb{R}$,and $\sigma > 0$. We also have $z = \frac{x-\mu}{\sigma}$, implies $x = \sigma z + \mu$, so $\left|\frac{dx}{dz}\right| = \sigma$. Also when $X = \pm\infty \implies Z = \pm\infty$, i.e. $Z \in \mathbb{R}$. Hence the PDF of Z is

$$\begin{aligned} f_Z(z) &= f_X(g^{-1}(z)) \left|\frac{dx}{dz}\right| \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2}} \times \sigma = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \qquad \text{for } z \in \mathbb{R} \end{aligned}$$

therefore $Z \sim N(0, 1)$.