STAT 512: Statistical Inference **Autumn 2020**

Lecture 2: Transforming continuous random variables

Instructor: Yen-Chi Chen

Reference: Casella and Berger Chapter 2.1.

In the previous lecture, we have seen a couple of distributions that have nice properties. When working with data, we may perform some transformation of random variables. Suppose we know the distribution of a random variable before the transformation, does this give us any hint on the distribution of the transformed variable?

2.1 One function of one random variable

Let X be a continuous random variable whose PDF $p_X(x)$ is known. Consider a given function f and another random variable $Y = f(X)$. Since the input X is random, the output Y is often random as well. What will the distribution of Y be?

When f is differentiable, we have the following useful theorem.

Theorem 2.1 *In the above setting and assume that $X \in [a,b]$ and $f'(x) > 0$ (strictly increasing) over $[a,b]$, then the PDF of Y*

$$p_Y(y) = \begin{cases} \frac{p_X(f^{-1}(y))}{f'(f^{-1}(y))}, & f(a) \le y \le f(b) \\ 0, & \text{otherwise.} \end{cases}$$

Proof:

To start with, we consider the CDF of Y:

$$\begin{aligned} P(Y \le y) &= P(f(X) \le y) \\ &= P(X \le f^{-1}(y)). \end{aligned}$$

The PDF will be the derivative of the CDF, leading to

$$\begin{aligned} p_Y(y) &= \frac{d}{dy} P(Y \le y) \\ &= \frac{d}{dy} P(X \le f^{-1}(y)) \\ &= p_X(f^{-1}(y)) \frac{d}{dy} f^{-1}(y) \\ &= \frac{p_X(f^{-1}(y))}{f'(f^{-1}(y))}, \end{aligned}$$

which completes the proof.

■

Example. Suppose $f(x) = x^2$ and $X \sim \mathsf{Uniform}[0,1]$. And we are interested in the PDF of $Y = f(X) = X^2$. Because $f'(x) = 2x$ and $X \geq 0$ so $f^{-1}(y) = \sqrt{y}$, we have

$$p_Y(y) = \frac{1}{2\sqrt{y}} I(0 \leq y \leq 1).$$

Example. Assume $X \sim \mathsf{Uniform}[0,1]$ and consider $f(x) = -2\log X$ and let $Y = -2\log X$. In this case, $f'(x) = -\frac{2}{X}$ and $f^{-1}(y) = e^{-\frac{1}{2}y}$. However, $f'(x)$ is negative so we cannot directly apply Theorem 2.1. A simple modification shows that the same formula holds as long as we replace $f'(f^{-1}(y))$ by $|f'(f^{-1}(y))|$ (think about why).

Then the PDF of Y will be

$$p_Y(y) = \frac{1}{2} e^{-\frac{1}{2}y} I(0 \leq y)$$

which is the Exponential distribution with parameter $\lambda = \frac{1}{2}$.

Example. Suppose that Y is a continuous random variable with CDF F_Y and X is a uniform random variable within $[0,1]$. Then you can show that $Z = F_Y^{-1}(X)$ has a CDF $F_Z(z) = F_Y(z)$.

Example. Consider $X \sim N(0,1)$ and $Y = X^2$. What is the distribution of Y? Note that the underlying transformation $f(x) = x^2$ is not always increasing or decreasing since $x \in \mathbb{R}$. In this case, a general strategy is to work out the CDF:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

Thus,

$$\begin{aligned} p_Y(y) &= \frac{d}{dy}[F_X(\sqrt{y}) - F_X(-\sqrt{y})] \\ &= \frac{1}{2\sqrt{y}}(p_X(\sqrt{y}) + p_X(-\sqrt{y})). \end{aligned}$$

In this case, because $X \sim N(0,1)$, it is symmetric so we further have

$$p_Y(y) = \frac{1}{\sqrt{y}} p_X(\sqrt{y}).$$

Putting $p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ into the above equation, we obtain

$$p_Y(y) = \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-y/2} = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y},$$

which is Gamma $(\frac{1}{2}, \frac{1}{2})$. Note: Gamma $(\frac{1}{2}, \frac{1}{2})$ is the same as χ^2_1, the chi-squared distribution with degree of freedom 1.

2.2 One function of two or more random variables

In practice, we may encounter problems involving a function of two or more random variables. Namely, we have X, Y two random variables whose joint distribution $p(x,y)$ is known and we are interested in the distribution of another random variable $U = f(X, Y)$ for some given function f. In this case, a general strategy is to investigate the underlying CDF and take the derivative to obtain the corresponding PDF. Here we will illustrate the idea via a few examples.

Example. Consider (X, Y) to be a uniform distribution over $[0, 1] \times [0, 1]$. Note that in this case, they are from two independent uniform distributions.

- **Case 1:** $U = X + Y$**.** Note that the event $\{U = X + Y \leq u\}$ will be the region of $[0, 1] \times [0, 1]$ intersecting with $x + y \leq u$. So it will be 0 when $u \leq 0$ and 1 when $u \geq 2$. When $u \in [0, 2]$, we can easily work it out using the area of a triangle, which leads to

$$\begin{aligned} F_U(u) &= P(U \leq u) \\ &= \begin{cases} 0, & u < 0 \\ u^2/2, & 0 \leq u \leq 1 \\ 1 - (2-u)^2/2, & 1 \leq u \leq 2 \\ 1, & u > 2 \end{cases}. \end{aligned}$$

  The PDF $p_U(u)$ will be

$$p_U(u) = \begin{cases} 0, & u < 0 \\ u, & 0 \leq u \leq 1 \\ 2 - u, & 1 \leq u \leq 2 \\ 0, & u > 2 \end{cases}.$$

- **Case 2:** $U = \max\{X, Y\}$**.** A common trick to compute the distribution of a maximum of two or more independent random variables is based on the following insight:

$$\{\max\{X, Y\} \leq u\} \equiv \{X \leq u, Y \leq u\}.$$

  Therefore,

$$F_U(u) = P(U \leq u) = P(\max\{X, Y\} \leq u) = P(X \leq u, Y \leq u) = P(X \leq u)P(U \leq u),$$

  which implies $F_U(u) = u^2$ and $p_U(u) = 2u$ when $u \in [0, 1]$.

- **Case 3:** $U = \min\{X, Y\}$**.** The case of minimum is similar to the case of maximal but we will consider a reverse event:

$$\{\min\{X, Y\} > u\} \equiv \{X > u, Y > u\}.$$

  Therefore.

$$1 - F_U(u) = P(U > u) = P(\min\{X, Y\} > u) = P(X > u, Y > u) = P(X > u)P(U > u) = (1 - u)^2,$$

  Thus, $F_U(u) = 1 - (1 - u)^2$ so $p_U(u) = 2 - 2u$ for $u \in [0, 1]$.

Example (minimum of many uniforms). Now consider $X_1, \cdots, X_n$ that are IID from a uniform distribution over $[0, 1]$. Define $U = n \min\{X_1, \cdots, X_n\}$. What will the distribution of U be when n is large? Using the trick that we have discussed,

$$\{\min\{X_1, \cdots, X_n\} > u\} \equiv \{X_1 > u, \cdots, X_n > u\},$$

so

$$1 - F_U(u) = P\left(\min\{X_1, \cdots, X_n\} > \frac{u}{n}\right) = \prod_{i=1}^{n} P\left(X_i > \frac{u}{n}\right) = \left(1 - \frac{u}{n}\right)^n \to e^{-u}.$$

As a result, $F_U(u) \to 1 - e^{-u}$ and $p_U(u) \to e^{-u}$ so when n is large, U behaves like from an Exponential distribution.

Example (exponential distributions). Consider X, Y are IID from exponential distribution with parameter 1.

- **Sum of two exponentials.** What is the distribution of $U = X + Y$? A simple trick is to fixed one variable at a time and make good use of integration. Specifically, for a given $u > 0$,

$$\begin{aligned}
F_U(u) &= P(U \leq u) \\
&= P(X + Y \leq u) \\
&= \int_{x+y \leq u} e^{-x-y} dx dy \\
&= \int_{x=0}^{u} \int_{y=0}^{u-x} e^{-x-y} dy dx \\
&= \int_{x=0}^{u} e^{-x}(1 - e^{x-u}) dx \\
&= 1 - e^{-u} - ue^{-u}.
\end{aligned}$$

  Thus, $p_U(u) = ue^{-u}$.

- **Minimum of two exponentials.** Now we consider $V = \min\{X, Y\}$. Using the same trick as the minimum of many uniforms, i.e.,

$$\{\min\{X, Y\} > v\} \equiv \{X > v, Y > v\}$$

  so

$$1 - F_V(v) = P(X > v)P(Y > v) = e^{-2v},$$

  which implies that $V \sim \mathsf{Exp}(2)$. In fact, you can easily generalize it to showing that if $X_1, \cdots, X_n \sim \mathsf{Exp}(\lambda)$, then $\min\{X_1, \cdots, X_n\} \sim \mathsf{Exp}(n\lambda)$.

- **Difference.** Consider

$$Z = \max\{X, Y\} - \min\{X, Y\} = |X - Y|.$$

  What will the distribution of Z be?

  Using a direct computation, we see that

$$\begin{aligned}
F_Z(z) &= P(Z \leq z) \\
&= P(|X - Y| \leq z) \\
&= P(-z \leq X - Y \leq z) \\
&= P(X - Y \leq z) - P(X - Y < -z) \\
&= P(X \leq Y + z) - 1 + P(X - Y \geq -z) \\
&= -1 + P(X \leq Y + z) + P(Y \leq X + z) \\
&= -1 + 2P(X \leq Y + z) \qquad X, Y \text{ are symmetriic.}
\end{aligned}$$

Moreover,

$$\begin{aligned} P(X \leq Y + z) &= \int_{y=0}^{\infty} \int_{x=0}^{y+z} e^{-x} dx e^{-y} dy \\ &= \int_{y=0}^{\infty} (1 - e^{-y-z}) e^{-y} dy \\ &= 1 - e^{-z} \int_{0}^{\infty} e^{-2y} dy \\ &= 1 - \frac{1}{2} e^{-z}. \end{aligned}$$

As a result,

$$F_Z(z) = -1 + 2P(X \leq Y + z) = 1 - e^{-z},$$

which is the CDF of $\mathsf{Exp}(1)$! This is another *memoryless property.*

- **Ratio.** Finally, we consider $W = \frac{X}{X+Y}$ and studies its distribution. Clearly, $0 \leq w \leq 1$ so we will focus on the range $[0, 1]$.

$$\begin{aligned} F_W(w) &= P\left(\frac{X}{X+Y} \leq w\right) \\ &= P(X \leq w(X+Y)) \\ &= P((1-w)X \leq wY) \\ &= P\left(X \leq \frac{w}{1-w} Y\right) \\ &= \int_{y=0}^{\infty} \int_{x=0}^{\frac{w}{1-w} y} e^{-x} dx e^{-y} dy \\ &= \int_{y=0}^{\infty} (1 - e^{\frac{-w}{1-w} y}) e^{-y} dy \\ &= 1 - \int_{0}^{\infty} e^{-\frac{1}{1-w} y} dy \\ &= 1 - 1 + w = w. \end{aligned}$$

Thus $W \sim \mathsf{Unif}[0, 1]$.

Useful properties about normal (please verify them).

- Let $X \sim N(\mu_1, \sigma^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ be independent. Then

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Also, for any real number a,

$$aX \sim N(a\mu_1, a^2 \sigma_1^2).$$

- Let $X_1, \cdots, X_n$ be IID normal random variables from $N(\mu, \sigma^2)$. Then the sample mean

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n).$$

- Let $X_1, \cdots, X_n$ be IID normal random variables from $N(0, 1)$. Then $Z_1 = X_1^2$ follows the χ^2 distribution with a degree of freedom 1. And $Z_n = \sum_{i=1}^{n} X_i^2$ follows the χ^2 distribution with a degree of freedom n.