Transformation

PROBLEM: X is a rv. Find the distribution of $Y = g(X)$.

- The very 1st step: specify the supports $\mathcal{S}_X$ and $\mathcal{S}_Y$.
- X is discrete – straightforward.
- X is continuous – pick one from the toolbox 
  - CDF approach: a general method applicable to all situations
  - A simple formula: g must be one-to-one and differentiable
  - MGF approach: not a general method but could be very powerful in some situations

The CDF Approach

Find the distribution of $Y = g(X)$.

1. Derive the CDF for Y:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \int_A f(x)dx$$

   where A is a set containing values $\{x : g(x) \leq y\}$.

2. Then find the pdf

$$f_Y(y) = \frac{dF_Y(y)}{dy}.$$

A Simple Formula

X is a continuous rv with pdf $f_X(x)$. Find the distribution of $Y = g(X)$. Suppose g is one-to-one and differentiable.

1. Solve $x = h(y)$ from $y = g(x)$ (i.e., $h(\cdot) = g^{-1}(\cdot)$)

$$y = g(x) \Longrightarrow x = h(y);$$

2. Calculate dx/dy, the derivative of $h(y)$ with respect to y

$$\frac{dx}{dy} = \frac{dh(y)}{dy} = h'(y);$$

3. Plug all the expressions into the following formula (Theorem 1.7.1, p. 47),

$$f_Y(y) = f_X(h(y))\Big|h'(y)\Big|, \quad y \in \mathcal{S}_Y. \tag{1}$$

How to remember this formula? In some sense (I'll explain this in class), the following holds true

$$f_X(x)|dx| = f_Y(y)|dy|.$$

So the density function for Y is given by

$$f_Y(y) = f_X(x)\Big|\frac{dx}{dy}\Big|, \tag{2}$$

where the absolute value ensures $f_Y(y) \geq 0$. Since the right side of (2) should be a function of y, we then replace every x in (2) by $h(y)$, which gives rises to that simple formula (1).

The MGF Approch

Find the distribution of $Y = g(X)$.

- Note that the mgf could uniquely determine a distribution when it exists.
- Compute the mgf of Y

$$M_Y(t) = \mathbb{E}_{Y}\left[e^{Y \cdot t}\right] = \mathbb{E}_{X}\left[e^{g(X) \cdot t}\right].$$

- Identify the distribution of Y based on its mgf.

The mgf of Y doesn't give us pdf/CDF, unless it is from some known parametric family. So the mgf approach may not be applicable in many situations. However, it could be very powerful in some special situations, e.g., it can be used to show CLT (Theorem 5.3.1, p.307).

Examples[a]

- Ex 1, 2, 4 from `Review2.pdf`
- Ex 4: mgf approach
- Ex 5: This is the trick I mentioned before — when computing an integral $\int g(x)dx$, if multiplying $g(x)$ with some constant A makes $Ag(x)$ a valid pdf, then $\int g(x)dx = 1/A$.
- Ex 6: Note that $X \in (-3, -1) \to Y \in (1, 9)$ and $X \in (0, 2) \to Y \in (0, 4)$. so $\mathcal{S}_Y = (0, 9)$ and the transformation is NOT one-to-one.

[a] Examples are from `Functions_of_Random_Variable.pdf` unless otherwise specified.

- $X \sim \mathsf{Unif}(0,1), \quad Y = \frac{1}{\lambda} \ln \frac{1}{1-X}, \quad \lambda > 0$. Show that $Y \sim \mathsf{Ex}(\lambda)$.

$$1 > X > 0 \qquad \Longrightarrow \qquad Y > 0$$

$$\begin{aligned} Y &= \frac{1}{\lambda} \ln \frac{1}{1-X} \qquad \Longrightarrow \qquad X = 1 - e^{-\lambda Y} \\ \left| \frac{dx}{dy} \right| &= \lambda e^{-\lambda y} \\ f_Y(y) &= 1 \cdot \left| \frac{dx}{dy} \right| = \lambda e^{-\lambda y}, \quad y > 0 \end{aligned}$$

- X is a continuous r.v. with cdf F, then

$$Y = F(X) \sim \text{Unif}(0, 1).$$

- Suppose F is strictly increasing. So $F^{-1}(u)$ is well-defined for $u \in (0, 1)$. Let $U \sim \text{Unif}(0, 1)$, then

$$Y = F^{-1}(U)$$

has the same distribution as X.

This is called the inverse cdf method that can be used to generates random samples from F by using samples from Unif$(0, 1)$ – it's easy to generating uniform random numbers. The previous example is a special case of this method.