

# **Lecture 4**

## **Transformation of Random Variables**

**Chao Song**

College of Ecology  
Lanzhou University

September 25, 2023

## Transformation of discrete random variables

A problem often encountered in statistics is the following. We have a random variable  $X$  and we know its distribution. We are interested in a random variable  $Y$  which is some **transformation** of  $X$ , say  $Y = g(X)$ . We want to determine the distribution of  $Y$ .

Let  $X$  be the number of trials until we get the first success. Let  $p$  be the probability of success. The probability mass function of  $X$  is thus  $P(X = x) = p(1 - p)^{x-1}$ . Let  $Y = X - 1$ , i.e.,  $Y$  is the number of failures before first success. What is the PMF of  $Y$ ?

$$\begin{aligned}P(Y = y) &= P(X - 1 = y) = P(X = y + 1) \\&= p(1 - p)^{y-1+1} = p(1 - p)^y\end{aligned}$$

In general, for discrete random variable, we can directly use the probability mass function of the original random variable to derive the probability mass function of the transformed random variable.

## Transformation of continuous random variables

Recall the theorem about standard normal distribution. If  $X \sim N(\mu, \sigma^2)$ , then  $Z = \frac{x-\mu}{\sigma}$  is  $N(0, 1)$ . Why is this the case?

**Proof:** The cumulative distribution function of  $Z$  is

$$\begin{aligned} P(Z \leq z) &= P\left(\frac{X - \mu}{\sigma} \leq z\right) = P(X \leq z\sigma + \mu) \\ &= \int_{-\infty}^{z\sigma + \mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \end{aligned}$$

We now use the change of variable integration given by  $w = (x - \mu)/\sigma$  (i.e.,  $x = w\sigma + \mu$ ) to obtain

$$P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw$$

# Transformation of continuous random variables

**Theorem:** Let  $X$  be a continuous random variable with PDF  $f_X(x)$  and support  $S_X$ . Let  $Y = g(x)$ , where  $g(x)$  is a one-to-one differentiable function, on the support of  $X$ . Denote the inverse of  $g$  by  $x = g^{-1}(y)$  and let  $dx/dy = d[g^{-1}(y)]$ . Then the PDF of  $Y$  is given by

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

**Proof:** Since  $g(x)$  is one-to-one and continuous, it is either monotonically increasing or decreasing. When it is strictly monotonically increasing, the CDF for  $Y$  is

$$F_Y(y) = P(Y \leq y) = P(g(x) \leq y) = P(x \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

Hence the PDF of  $Y$  is

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

Similarly, when  $g(x)$  is monotonically decreasing,

$$F_Y(y) = P(Y \leq y) = P(g(x) \leq y) = P(x \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y))$$

Hence the PDF of  $Y$  is

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

## log-normal distributions

Let  $X \sim N(0, \sigma^2)$ , then  $Y = e^X$  has a log-normal distribution.

**Proof:** When  $Y = e^X$ , we have  $X = \ln(Y)$ . Using the general conclusions about transformation of continuous random variable, the PDF of  $Y$  is

$$\begin{aligned} f_Y(y) &= f_X(\ln(y)) \left| \frac{dx}{dy} \right| \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln(y)-\mu)^2}{2\sigma^2}} \left| \frac{d\ln(y)}{dy} \right| \\ &= \frac{1}{\sigma y\sqrt{2\pi}} e^{-\frac{(\ln(y)-\mu)^2}{2\sigma^2}} \end{aligned}$$

## log-normal distributions

Let  $X \sim N(0, \sigma^2)$ , then  $Y = e^X$  has a log-normal distribution. What is the mean and variance of  $Y$ ?

$$\begin{aligned} E(Y) &= \int_0^{\infty} y f_Y(y) dy \\ &= \int_0^{\infty} y \frac{1}{\sigma y \sqrt{2\pi}} e^{-\frac{(\ln(y) - \mu)^2}{2\sigma^2}} dy \end{aligned}$$

For convenience of integration, use change of variable  $t = (\ln(y) - \mu)/\sigma$  so that  $y = e^{\sigma t + \mu}$  and  $dy = \sigma e^{\sigma t + \mu} dt$ , we have

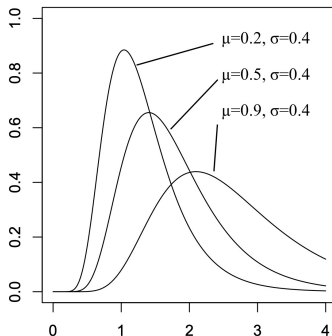
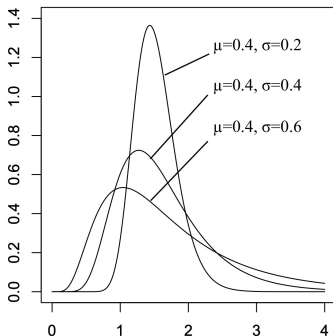
$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}t^2} \sigma e^{\sigma t + \mu} dt \\ &= e^{\mu + \frac{1}{2}\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-\sigma)^2}{2}} dt \\ &= e^{\mu + \frac{1}{2}\sigma^2} \end{aligned}$$

Similarly, we could calculate the variance of  $Y$  to be  $\text{Var}(Y) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$ .

# log-normal distribution

If  $X \sim N(\mu, \sigma^2)$ , then  $Y = e^X$  has a log-normal distribution with mean  $e^{\mu + \frac{1}{2}\sigma^2}$  and variance  $(e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$ .

Note that if  $X \sim N(\mu, \sigma^2)$ , then the mean of  $Y = e^X$  **is not**  $e^\mu$  because  $e^X$  is a non-linear transformation.



## $\chi^2$ -distribution

Let  $X$  follows a standard normal distribution. Find the PDF of  $Y = X^2$

$$\begin{aligned}F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\&= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\&= \int_{-\infty}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} dy - \int_{-\infty}^{-\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} dy\end{aligned}$$

Thus, the PDF of  $Y$  is

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}$$

This is the PDF of a  $\chi^2$ -distribution with 1 degree of freedom.



## Universality of the uniform

Let  $X$  be a continuous random variable and  $F_X(x)$  be its cumulative distribution function. What is the PDF of  $Y = F_X(x)$ ?

Using the method of distribution function, we have

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

Thus, the PDF of  $Y$  is

$$f_Y(y) = \frac{d}{dy} F_Y(y) = 1$$

**Theorem:** For a continuous random variable  $X$ , its cumulative distribution function  $F_X(x)$  follows a uniform distribution between 0 and 1,  $U(0, 1)$

**Corollary:** The fact that cumulative distribution function is  $U(0, 1)$  provides a universal way to simulate continuous random variable. Specifically, one can draw random numbers from  $U(0, 1)$  and then compute any random variable by the inverse of its cumulative distribution function.

## Order statistics

**Definition:** Let  $X_1, X_2, \dots, X_n$  be a random sample from a distribution. Let  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  be the random variables sorted from the smallest to the largest. We call  $X_{(j)}$  the  $j$ th order statistics of the random sample. We use  $f_{(j)}$  and  $F_{(j)}$  to denote its PDF and CDF respectively.

Let  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  be the order statistics of a random sample from a distribution. What is the probability density function of the maximum  $X_{(n)}$ ?

$$\begin{aligned} F_{(n)}(x) &= P(X_{(n)} \leq x) = P(X_1 \leq x, \dots, X_n \leq x) \\ &= \prod_{i=1}^n P(X_i \leq x) = F_X(x)^n \\ f_n(x) &= \frac{d}{dx} F_{(n)}(x) = n F_X(x)^{n-1} f_X(x) \end{aligned}$$

## Order statistics

Let  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  be the order statistics of a random sample from a distribution. What is the probability density function of the minimum  $X_{(1)}$ ?

$$\begin{aligned}F_{(1)}(x) &= P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x) \\&= 1 - P(X_1 > x, \dots, X_n > x) \\&= 1 - \prod_{i=1}^n P(X_i > x) \\&= 1 - \prod_{i=1}^n (1 - P(X_i \leq x)) \\&= 1 - (1 - F_X(x))^n \\f_{(1)}(x) &= \frac{d}{dx} F_{(1)}(x) = n(1 - F_X(x))^{n-1} f_X(x)\end{aligned}$$

# Moment generating function

**Definition:** Let  $X$  be a random variable, continuous or discrete. We define the moment generating function of  $X$  to be

$$m_X(t) = E(e^{tX})$$

Moment generating function, as its name suggests, can be used to find moments of a random variable.

$$\frac{d}{dt}m_X(t) = E(Xe^{tX}),$$

which when we evaluate at  $t = 0$  becomes  $E(X)$ . More generally, the  $n$ th derivative of  $m_X(t)$  evaluated at zero is the expected value of  $X^n$ , i.e.,  $m^{(n)}(0) = E(X^n)$

**The moment generating function determines the distribution of  $X$ .** From this point of view, knowing a random variables moment generating function gives as much information as its PDF or CDF.

## Moment generating function

If  $X \sim N(0, 1)$ , what is the moment generating function of  $X$ ?

$$\begin{aligned}m_X(t) &= E(e^{tX}) = \int_{-\infty}^{\infty} e^{xt} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\&= \int_{-\infty}^{\infty} e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} dx \\&= e^{\frac{t^2}{2}}\end{aligned}$$

Using similar technique with a change of variable in the integration, we can get the moment generating function for  $X \sim N(\mu, \sigma^2)$  to be

$$m_X(t) = e^{\mu t} e^{\frac{1}{2} \sigma^2 t^2}$$

## Moment generating function

**Theorem:** Let  $X$  and  $Y$  be random variables with moment generating functions  $m_X(t)$  and  $m_Y(t)$ . if  $X$  and  $Y$  are independent, the moment generating function of  $aX + bY$  is

$$m_{aX+bY}(t) = m_X(at)m_Y(bt)$$

**Proof:** According to the definition of moment generating function:

$$m_{aX+bY}(t) = E(e^{(aX+bY)t}) = E(e^{aXt+bYt}) = E(e^{Xat}e^{Ybt})$$

Because  $X$  and  $Y$  are independent,  $E(e^{Xt}e^{Yt}) = E(e^{Xat})E(e^{Ybt})$ . Thus

$$m_{aX+bY}(t) = m_X(at)m_Y(bt)$$

## Methods of moment generating function

Because moment generating functions uniquely identifies a distribution. We can use the moment generating function to find the distribution of a transformed random variable.

**Example:** Recall that the moment generating function of  $X \sim N(\mu, \sigma^2)$  is  $m_X(t) = e^{\mu t} e^{\frac{1}{2}\sigma^2 t^2}$ . If  $X_1 \sim N(\mu_1, \sigma_1^2)$  and  $X_2 \sim N(\mu_2, \sigma_2^2)$  are independent, what is the distribution of  $X_1 + X_2$ ?

$$\begin{aligned} m_{X_1+X_2}(t) &= m_{X_1}(t)m_{X_2}(t) \\ &= e^{\mu_1 t} e^{\frac{1}{2}\sigma_1^2 t^2} e^{\mu_2 t} e^{\frac{1}{2}\sigma_2^2 t^2} \\ &= e^{(\mu_1+\mu_2)t} e^{\frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2} \end{aligned}$$

This is the moment generating function of a normal distribution with mean  $\mu_1 + \mu_2$  and variance  $\sigma_1^2 + \sigma_2^2$ .

## Methods of moment generating function

Using the method of moment generating function, we get the following theorem about transformation of normally distributed random variables.

**Theorem:** if  $X_1, \dots, X_n$  are mutually independent normal variables with mean  $\mu_i$  and variance  $\sigma_i^2$ , then the linear function

$$Y = \sum_{i=1}^n c_i X_i$$

has the normal distribution

$$N\left(\sum_{i=1}^n c_i \mu_i, \sum_{i=1}^n c_i^2 \sigma_i^2\right)$$

**Theorem:** if  $X_1, X_2, \dots, X_n$  are observations of a random sample of size  $n$  from the normal distribution  $N(\mu, \sigma^2)$ , then the sample mean

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



## Method of moment generating function

If  $X$  has a Poisson distribution with parameter  $\lambda$ , its moment generating function is

$$\begin{aligned} E(e^{tX}) &= \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} \\ &= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)} \end{aligned}$$

If  $X_1$  and  $X_2$  are independent Poisson distributed random variables with parameters  $\lambda_1$  and  $\lambda_2$ , what is the distribution of  $X_1 + X_2$ ?

$$\begin{aligned} m_{X_1+X_2}(t) &= m_{X_1}(t)m_{X_2}(t) \\ &= e^{\lambda_1(e^t - 1)} e^{\lambda_2(e^t - 1)} \\ &= e^{(\lambda_1 + \lambda_2)(e^t - 1)} \end{aligned}$$

**Theorem:** If  $X_1$  and  $X_2$  are independent Poisson distributed random variables with parameters  $\lambda_1$  and  $\lambda_2$ , then  $X_1 + X_2$  follows a Poisson distribution with parameter  $\lambda_1 + \lambda_2$ .