Lecture 5
Transformation of Random Variables

Chao Song

College of Ecology

Lanzhou University

October 13, 2025

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

$$P(Z \leqslant z) = P(\frac{X-\mu}{\sigma} \leqslant z) = P(X \leqslant z\sigma + \mu)$$
$$= \int_{-\infty}^{z\sigma+\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

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

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

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

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

Hence the PDF of Y is

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

Transformation of continuous random variables

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

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

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 distribution

Let $X \sim N(\mu, \sigma^2)$, 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

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

Log-normal distribution

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

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

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

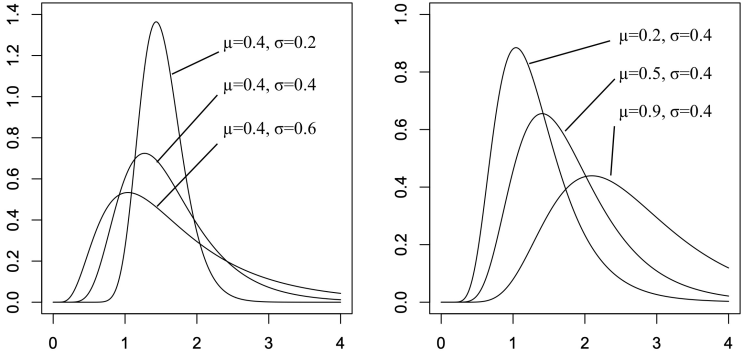
$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \sigma e^{\sigma t + \mu} dt \\ &= e^{\mu + \frac{1}{2}\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-\sigma)^2}{2}} dt \\ &= e^{\mu + \frac{1}{2}\sigma^2} \end{aligned}$$

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

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^μ because e^X is a non-linear transformation.



Chi-square distribution

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

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

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-square 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

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

Thus, the PDF of Y is

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

Universality of the uniform

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

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

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

Order statistics

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

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

Method of 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

Theorem: if $X_1, \ldots, X_n$ are mutually independent normal variables with mean μ_i and variance σ_i^2, then the linear function

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

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

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

Method of moment generating function

If X_1 and X_2 are independent Poisson distributed random variables with parameters λ_1 and λ_2, what is the distribution of $X_1 + X_2$?

The MGF of a Poisson random variable is $m(t) = e^{\lambda(e^t - 1)}$. Thus,

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

Theorem: If X_1 and X_2 are independent Poisson distributed random variables with parameters λ_1 and λ_2, then $X_1 + X_2$ follows a Poisson distribution with parameter $\lambda_1 + \lambda_2$.

Transformation of multiple random variables

We can extend the techniques of transforming a single random variable to the cases of multiple random variables. If we want to transform multiple random variables into a single random variable, we can address this problem from the cumulative distribution function.

Example: Let X and Y be two random variables with joint PDF

$$f(x, y) = e^{-x-y}, \quad x > 0, y > 0$$

Let $V = X + Y$. What is the PDF of V?

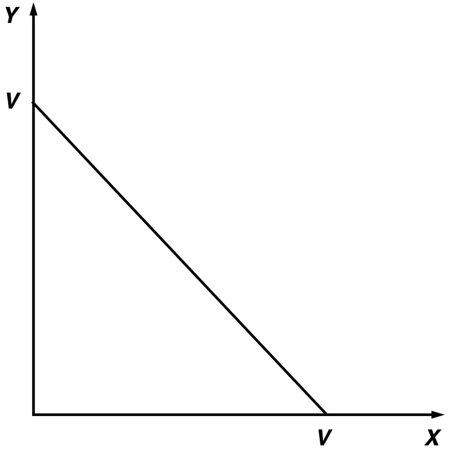
Answer: To find the PDF of V, we start from its CDF

$$P(V \leqslant v) = P(X + Y \leqslant v)$$

Since we know the joint PDF of X and Y, all we need to do is to integrate the joint PDF function over the region defined by $X + Y < V$.

Transformation of multiple random variables

Given that $X > 0$ and $Y > 0$, the region defined by $X + Y < V$ is a triangle shown in the figure below.



Transformation of multiple random variables

Integrating the joint PDF of X and Y over the support of V, we have

$$
\begin{aligned}
P(V \leqslant v) = P(X + Y \leqslant v) &= \int_0^v \int_0^{v-y} e^{-x-y} dxdy \\
&= \int_0^v (-e^{-x})\Big|_0^{v-y} e^{-y} dy = \int_0^v (1 - e^{y-v}) e^{-y} dy \\
&= \int_0^v (e^{-y} - e^{-v}) dy = (-e^{-y} - e^{-v} y)\Big|_0^v \\
&= 1 - e^{-v} - ve^{-v}
\end{aligned}
$$

Differentiate the CDF with respect to v

$$
f_V(v) = \frac{dF_V(v)}{dv} = e^{-v} - e^{-v} + ve^{-v} = ve^{-v}, \quad v > 0.
$$

Transformation of multiple random variables

Using similar techniques with a bit help from matrix algebra, we can derive a general formula for joint PDF of transformation of multiple random variables.

If X_1 and X_2 are two continuous random variables with joint PDF $f(x_1, x_2)$, and if $Y_1 = u_1(X_1, X_2)$, $Y_2 = u_2(X_1, X_2)$ has the single-valued inverse $X_1 = v_1(Y_1, Y_2)$, $X_2 = v_2(Y_1, Y_2)$, then the joint PDF of Y_1 and Y_2 is

$$g(y_1, y_2) = f(v_1(y_1, y_2), v_2(y_1, y_2))\left|\boldsymbol{J}\right|$$

where the Jacobian $\boldsymbol{J}$ is the determinant

$$\left|\boldsymbol{J}\right| = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{vmatrix}$$

Transformation of multiple random variables

Example: Let X_1 and X_2 have the joint PDF

$$f(x_1, x_2) = 2, \quad 0 < x_1 < x_2 < 1$$

Consider the transformation

$$Y_1 = \frac{X_1}{X_2}, \quad Y_2 = X_2$$

what is the joint PDF of Y_1 and Y_2?

Transformation of multiple random variables

Answer: Based on the transformation, we have

$$X_1 = Y_1 Y_2, \quad X_2 = Y_2$$

and the Jacobian of the transformation is

$$|\boldsymbol{J}| = \begin{vmatrix} \frac{\partial X_1}{\partial Y_1} & \frac{\partial X_1}{\partial Y_2} \\ \frac{\partial X_2}{\partial Y_1} & \frac{\partial X_2}{\partial Y_2} \end{vmatrix} = \begin{vmatrix} y_2 & y_1 \\ 0 & 1 \end{vmatrix} = y_2$$

Thus, the joint PDF of Y_1 and Y_2 is

$$g(y_1, y_2) = f(y_1 y_2, y_2)|\boldsymbol{J}| = 2y_2$$

Here, the support of Y_1 and Y_2 is

$$0 < y_1 < 1, \quad 0 < y_2 < 1.$$