Functions of random variables

Dr. Qiuzhuang Sun

STAT3023

Transformation of a random variable

Distribution of functions of a random variable

Let X be a random variable with cdf $F_X(x) = P(X \leq x)$, then for any function g, we have $Y = g(X)$ is also a random variable. The distribution of Y can be determined from

$$F_Y(y) = P(Y \leq y) = P\left(g(X) \leq y\right) = P(X \in \mathcal{A})$$

where $\mathcal{A} = \{x : g(x) \leq y\}$.

Example: What is the distribution of $Y = e^X$ with $X \sim N(0,1)$?

Example

What is the distribution of $Y = X^2$ with $X \sim N(0, 1)$?

χ^2 and gamma distributions

A random variable Y follows a gamma distribution with shape k and scale θ, denoted by Gamma(k, θ), if its pdf is given by

$$f_Y(y) = \frac{1}{\Gamma(k)\theta^k} y^{k-1} e^{-\frac{y}{\theta}}.$$

Compare Gamma$(\frac{1}{2}, 2)$ and χ_1^2.

χ^2 and gamma distributions

Show Gamma$(\frac{n}{2}, 2)$ is equivalent to χ^2_n.

Distribution of functions of a random variable

If g is **monotone increasing**, the calculation can be simplified. In this case, the function g has also a monotone increasing inverse function g^{-1}. Then we have

$$F_Y(y) = P(Y \leq y) =$$

$$f_Y(y) = F'_Y(y) =$$

If g is **monotone decreasing**, then we have

$$F_Y(y) = P(Y \leq y) =$$

$$f_Y(y) = F'_Y(y) =$$

Density of $g(X)$

From the last slide, we have for monotone g,

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

Example: What is the density of $Y = -\log(X)$ with $X \sim$ Uniform$(0, 1)$?

Bivariate transformation

Joint distribution

Let X and Y be two random variables with known joint distribution. Consider $U = g_1(X, Y)$ and $V = g_2(X, Y)$ for some functions g_1 and g_2. Using the cdf argument, we can also find the joint distribution of U and V:

$$F_{U,V}(u, v) = P(U \leq u, V \leq v) = P((X, Y) \in \mathcal{A}_{u,v})$$

for some set $\mathcal{A}_{u,v} = \{(x, y) \mid g_1(x, y) \leq u,\ g_2(x, y) \leq v\}$, and

$$P((X, Y) \in \mathcal{A}_{u,v}) = \iint_{\mathcal{A}_{u,v}} f_{X,Y}(x, y) dx dy.$$

Typically, computing this is not easy!

Jacobian technique

Consider a case when both X and Y are continuous with joint pdf $f_{X,Y}(x,y)$, and there is a one-to-one transformation between (X,Y) and (U,V). In this case, we can write $X = h_1(U,V)$ and $Y = h_2(U,V)$, and the joint density of U and V is given by

$$f_{U,V}(u,v) = f_{X,Y}(h_1(u,v), h_2(u,v))|\det(J)|, \quad J = \begin{bmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{bmatrix}.$$

Are the following two transformations one-to-one?

(a) $u = x + y$, $v = x - y$;

(b) $u = x^2 + y^2$, $v = x^2 - y^2$.

Examples

Assume X, Y has a joint pdf $f_{X,Y}(x, y) = e^{-(x+y)}$, $x, y > 0$. What is the joint pdf of U, V with $U = X + Y$ and $V = \frac{X}{X+Y}$?

Examples

Assume $X \sim N(0,1)$, $Y \sim N(0,1)$, and X and Y are independent. What is the distribution of $Z = \frac{X}{Y}$?

Examples

(Example continued)

Examples

Assume $\mathbf{U} = (X, Y)^T$ follows a bivariate normal distribution. What is the distribution of $\mathbf{V} = \mathbf{A}\mathbf{U}$ with a non-singular $\mathbf{A} \in \mathbb{R}^{2\times 2}$?

Examples

(Example continued)

Examples

(Example continued)

More on bivariate normal

If (X, Y) follows a bivariate normal distribution, then the marginal distributions of X and Y are both normal.

However, marginal normality does not imply bivariate normality.

Example: Let $X \sim N(0, 1)$ and $Y = -X \sim N(0, 1)$. However, $X + Y = 0$ is not bivariate normal.

Extension to multivariate distributions

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector. Consider the transformation:

$$\begin{cases} U_1 = g_1(X_1, \ldots, X_n), \\ \qquad \vdots \\ U_n = g_n(X_1, \ldots, X_n). \end{cases}$$

If the transformation is one-to-one, i.e., we can get $\mathbf{X}$ from $\mathbf{U} = (U_1, \ldots, U_n)$ using $\mathbf{X} = \mathbf{h}(\mathbf{U}) = (h_1(\mathbf{U}), \ldots, h_n(\mathbf{U}))$, then

$$g_{\mathbf{U}}(\mathbf{u}) = f_{\mathbf{X}}(\mathbf{h}(\mathbf{u}))|\det(J)|, \quad J = \begin{bmatrix} \frac{\partial h_1}{\partial u_1} & \cdots & \frac{\partial h_1}{\partial u_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial u_1} & \cdots & \frac{\partial h_n}{\partial u_n} \end{bmatrix}.$$

Sample from normal distributions

Sample mean and variances

Let $X_1, \ldots, X_n$ denote iid samples from $N(\mu, \sigma^2)$ distribution. Let $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ and $s_n^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$.

Then:

1. $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$;
2. $\bar{X}_n$ and s_n^2 are independent;
3. $\frac{(n-1)s_n^2}{\sigma^2} \sim \chi^2_{n-1}$.

Proof:

Proofs

Proofs

Proofs

Proofs

The t-distribution

Let $Z \sim N(0,1)$ and $V \sim \chi^2_d$ independent of Z. Then the random variable

$$T = \frac{Z}{\sqrt{V/d}} \sim t_d,$$

the t distribution with d degrees of freedom.

Example: What is the density of T?

Density of T

(Example continued)

Density of T

(Example continued)

The t-distribution

Let $X_1, \dots, X_n$ denote iid samples from $N(\mu, \sigma^2)$ distribution, with $\bar{X}_n$ and s_n^2 the sample mean and sample variance. Then

$$T = \frac{\bar{X}_n - \mu}{s_n / \sqrt{n}} \sim t_{n-1}.$$

Compare T with

$$Z = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \sim N(0, 1).$$

Proof: