

**Wee 13**  
Dec. 3 – Dec. 7

**Lecture 35. Central Limit Theorem**

**Theorem:** Let  $X_1, X_2, \dots, X_n$  be i.i.d. with  $EX_i = \mu$  and  $Var(X_i) = \sigma^2$ . Let  $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ . Then the distribution of  $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$  is approximately  $N(0, 1)$  in the sense that

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq b\right) = \Phi(b) - \Phi(a) \text{ for all } a, b.$$

**Fact:** If two random variables have close moment generating functions, then their cumulative distribution functions are close!

Sorry. We are not proving this fact, but using it.

Without loss of generality we assume that  $\mu = 0$  and  $\sigma = 1$ . Let  $g(t)$  be the moment generating function of  $X_i$ . The moment generating function of  $X_1 + X_2 + \dots + X_n$  is

$$(g(t))^n$$

then the moment generating function of  $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sqrt{n}\sigma} = \frac{X_1 + X_2 + \dots + X_n}{\sqrt{n}}$  (for  $\mu = 0$  and  $\sigma = 1$ ) is

$$g(t) = \left(g\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

(**Recall:** Let  $Y = \frac{X - \mu}{\sigma}$ , then  $g_Y(t) = e^{-\mu t/\sigma} g_X\left(\frac{t}{\sigma}\right)$ .)

**Remark:**

$$g(0) = 1, g'(0) = 0, g''(0) = 1.$$

This remark implies

$$g\left(\frac{t}{\sqrt{n}}\right) \approx 1 + \frac{t^2}{2n}$$

then

$$g(t) = \left(g\left(\frac{t}{\sqrt{n}}\right)\right)^n \approx \left(1 + \frac{t^2}{2n}\right)^n \rightarrow e^{t^2/2} \text{ as } n \rightarrow \infty$$

and  $e^{t^2/2}$  is exactly the moment generating function of *Normal*  $(0, 1)$ .

**Remark:** If the moment generating function of  $X_i$  doesn't exist, our calculation here doesn't make sense.

**Example:** For  $X \sim \text{Cauchy}(0, 1)$ , i.e.,

$$f_X(x) = \frac{1}{\pi} \frac{1}{1 + x^2}$$

it doesn't have a moment generating function, since

$$\int_{-\infty}^{\infty} e^{tx} \frac{1}{\pi} \frac{1}{1 + x^2} dx$$

doesn't exist!

**A fun fact:** Let  $X_1, X_2, \dots, X_n$  be i.i.d. *Cauchy*  $(0, 1)$ , then

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \sim \text{Cauchy}(0, 1).$$

then

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{\bar{X}}{\sqrt{n}} \leq b\right) = \lim_{n \rightarrow \infty} P(\sqrt{n}a \leq \bar{X} \leq b\sqrt{n}) = 1.$$

### Lecture 36. Multivariate Cases

#### Joint densities for linear combinations

Suppose  $X$  and  $Y$  have a jointly continuous distribution with joint density  $f(x, y)$ . For constants  $a, b, c, d$ , define

$$U = aX + bY \text{ and } V = cX + dY$$

Find the joint density function  $\psi(u, v)$  for  $(U, V)$ , under the assumption that the quantity  $\kappa = ad - bc$  is nonzero.

**Example:** Suppose  $X$  and  $Y$  are independent random variables, each distributed  $N(0, 1)$ . For constants  $a, b, c, d$ , define

$$U = aX + bY \text{ and } V = cX + dY$$

Find the joint density function  $\psi(u, v)$  for  $(U, V)$ , under the assumption that the quantity  $\kappa = ad - bc$  is nonzero.

Think of the pair  $(U, V)$  as defining a new random point in  $\mathbb{R}^2$ . That is  $(U, V) = T(X, Y)$ , where  $T$  maps the point  $(x, y) \in \mathbb{R}^2$  to the point  $(u, v) \in \mathbb{R}^2$  with

$$u = ax + by \text{ and } v = cx + dy,$$

or in matrix notation,

$$(u, v) = (x, y)A, \text{ where } A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

Notice that  $\det(A) = ad - bc = \kappa$ . The assumption that  $\kappa \neq 0$  ensures that the transformation is invertible:

$$(u, v)A^{-1} = (x, y), \text{ where } A^{-1} = \frac{1}{\kappa} \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$$

That is,

$$x = \frac{1}{\kappa}(du - bv) \text{ and } y = \frac{1}{\kappa}(-cu + av)$$

Notice that  $\det(A^{-1}) = 1/\kappa = 1/\det(A)$ . It helps to distinguish between the two roles for  $\mathbb{R}^2$ , referring to the domain of  $T$  as the  $(X, Y)$ -plane and the range as the  $(U, V)$ -plane. The joint density function  $\psi(u, v)$  for  $(U, V)$  is characterized by the property that

$$P\{u_0 \leq U \leq u_0 + \delta_1, v_0 \leq V \leq v_0 + \delta_2\} \approx \psi(u_0, v_0)\delta_1\delta_2$$

for each  $(u_0, v_0)$  in the  $(U, V)$ -plane, and small  $(\delta_1, \delta_2)$ . To calculate the probability on the left-hand side we need to find the region  $R$  in the  $(X, Y)$ -plane corresponding to the small rectangle, with corners at  $(u_0, v_0)$  and  $(u_0 + \delta_1, v_0 + \delta_2)$ , in the  $(U, V)$ -plane. The linear transformation  $A^{-1}$  maps parallel straight lines

in the  $(U, V)$ -plane into parallel straight lines in the  $(X, Y)$ -plane. The region  $R$  must be a parallelogram, with vertices

$$(u_0, v_0)A^{-1}, (u_0 + \delta_1, v_0)A^{-1}, (u_0, v_0 + \delta_2)A^{-1}, (u_0 + \delta_1, v_0 + \delta_2)A^{-1}.$$

Let

$$(x_0, y_0) = (u_0, v_0)A^{-1}$$

then four vertices can be also written as

$$(x_0, y_0), (x_0, y_0) + (\delta_1, 0)A^{-1}, (x_0, y_0) + (0, \delta_2)A^{-1}, (x_0, y_0) + (\delta_1, \delta_2)A^{-1}$$

**Fact:**

$$\frac{\text{area}(R)}{\delta_1 \delta_2} = \frac{1}{|\det(A)|}.$$

We also know

$$P\{(X, Y) \in R\} \approx f(x_0, y_0)\text{area}(R) = f(x_0, y_0)\delta_1\delta_2 \frac{1}{|\det(A)|}$$

and

$$P\{(X, Y) \in R\} = P\{u_0 \leq U \leq u_0 + \delta_1, v_0 \leq V \leq v_0 + \delta_2\}$$

then

$$\psi(u_0, v_0)\delta_1\delta_2 \approx f(x_0, y_0)\delta_1\delta_2 \frac{1}{|\det(A)|}$$

thus

$$\psi(u_0, v_0) = \frac{1}{|\det(A)|} f((u_0, v_0)A^{-1})$$

**Conclusion:** Suppose  $X$  and  $Y$  have a jointly continuous distribution with joint density  $f(x, y)$ . For constants  $a, b, c, d$ , define

$$(U, V) = (X, Y)A, \text{ where } A = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \text{ with } \det(A) \neq 0.$$

The joint density function  $\psi(u, v)$  for  $(U, V)$  is

$$\psi(u, v) = \frac{1}{|\det(A)|} f((u, v)A^{-1}).$$

**Conclusion (Generalization):** Let  $f(x_1, x_2, \dots, x_n)$  be a jointly continuous density of  $X_1, X_2, \dots, X_n$ . Define

$$(U_1, U_2, \dots, U_n) = (X_1, X_2, \dots, X_n)A, \text{ where } A \text{ is an } n \times n \text{ matrix with } \det(A) \neq 0.$$

The joint density function  $\psi(u_1, u_2, \dots, u_n)$  for  $(U_1, U_2, \dots, U_n)$  is

$$\psi(u_1, u_2, \dots, u_n) = \frac{1}{|\det(A)|} f((u_1, u_2, \dots, u_n)A^{-1}).$$

Remark: if you don't know linear algebra, it is fine to assume that  $n = 2$ .

### Lecture 37. Multivariate Normal Distribution

#### A special case of the generalized conclusion

Let  $X_1, X_2, \dots, X_n$  be i.i.d.  $N(0, 1)$ . Define

$$(U_1, U_2, \dots, U_n) = (X_1, X_2, \dots, X_n) A,$$

where  $A$  is an  $n \times n$  matrix with  $\det(A) \neq 0$ . The joint density function  $\psi(u_1, u_2, \dots, u_n)$  for  $(U_1, U_2, \dots, U_n)$  is

$$\psi(u_1, u_2, \dots, u_n) = \frac{1}{|\det(A)|} f((u_1, u_2, \dots, u_n)A^{-1}).$$

**Recall that**

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{x_1^2 + x_2^2 + \dots + x_n^2}{2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{(x_1, x_2, \dots, x_n)(x_1, x_2, \dots, x_n)^T}{2}\right) \end{aligned}$$

Then

$$\begin{aligned} &\psi(u_1, u_2, \dots, u_n) \\ &= \frac{1}{|\det(A)|} f((u_1, u_2, \dots, u_n)A^{-1}) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{|\det(A)|} \exp\left(-\frac{(u_1, u_2, \dots, u_n)A^{-1}((u_1, u_2, \dots, u_n)A^{-1})^T}{2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{|\det(A)|} \exp\left(-\frac{(u_1, u_2, \dots, u_n)(A^T A)^{-1}(u_1, u_2, \dots, u_n)^T}{2}\right) \end{aligned}$$

Let  $B = A^T A$ . We write

$$\begin{aligned} &\psi(u_1, u_2, \dots, u_n) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{|\det(B)|^{1/2}} \exp\left(-\frac{(u_1, u_2, \dots, u_n)B^{-1}(u_1, u_2, \dots, u_n)^T}{2}\right) \end{aligned}$$

**Definition:** We say  $(U_1, U_2, \dots, U_n) \sim N(0, \Sigma)$ , if the joint density function of  $(U_1, U_2, \dots, U_n)$  is

$$\begin{aligned} &\psi(u_1, u_2, \dots, u_n) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{|\det(\Sigma)|^{1/2}} \exp\left(-\frac{(u_1, u_2, \dots, u_n)\Sigma^{-1}(u_1, u_2, \dots, u_n)^T}{2}\right) \end{aligned}$$

**Remark:** This general definition coincides with the definition for  $n = 1$ . From this general definition  $X \sim N(0, \sigma^2)$  means

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

since  $\Sigma = \sigma^2$ .

**Example:** Let  $X_1$  and  $X_2$  be i.i.d.  $N(0, 1)$ ,

$$(U_1, U_2) = (X_1, X_2) A, \text{ where } A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Then

$$B = A^T A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\begin{aligned} \psi(u_1, u_2, \dots, u_n) &= \left( \frac{1}{\sqrt{2\pi}} \right)^2 \frac{1}{|\det(B)|^{1/2}} \exp \left( -\frac{(u_1, u_2) B^{-1} (u_1, u_2)^T}{2} \right) \\ &= \left( \frac{1}{\sqrt{2\pi}} \right)^2 \frac{1}{2} \exp \left( -\frac{(u_1, u_2) \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} (u_1, u_2)^T}{2} \right) \\ &= \frac{1}{4\pi} \exp \left( -\frac{u_1^2 + u_2^2}{4} \right) \end{aligned}$$

**Remark:** It is very interesting in this example that  $U_1$  and  $U_2$  are independent!

**Remark (Generalization):** Let  $X_1, X_2, \dots, X_n$  be i.i.d.  $N(0, 1)$ . Define

$$(U_1, U_2, \dots, U_n) = (X_1, X_2, \dots, X_n) A, \text{ where } A A^T = a I_{n \times n}.$$

Then  $(U_1, U_2, \dots, U_n)$  are i.i.d. again! When  $a = 1$ ,  $(U_1, U_2, \dots, U_n)$  are i.i.d.  $N(0, 1)$ .

### Variance–Covariance Matrix

**Definition:** Covariance of  $X$  and  $Y$

$$\text{cov}(X, Y) = E((X - EX)(Y - EY))$$

**Remark:** If  $X = Y$ , then

$$\text{cov}(X, Y) = \text{Var}(X)$$

**Remark:** If  $X$  and  $Y$  are independent,

$$\text{cov}(X, Y) = E((X - EX)(Y - EY)) = 0$$

**Definition:** The variance–covariance matrix of  $(X_1, X_2, \dots, X_n)$  is

$$\text{Var}(X_1, X_2, \dots, X_n) = (v_{ij})_{n \times n}$$

where

$$v_{ij} = E((X_i - EX_i)(X_j - EX_j)).$$

We often write

$$\begin{aligned} & \text{Var}(X_1, X_2, \dots, X_n) \\ &= E \left( (X_1 - EX_1, X_2 - EX_2, \dots, X_n - EX_n)^T (X_1 - EX_1, X_2 - EX_2, \dots, X_n - EX_n) \right) \end{aligned}$$

**Question:** For  $X \sim N(0, \sigma^2)$ , i.e.,

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

we have  $\text{var}(X) = \sigma^2$ .

For  $(X_1, X_2, \dots, X_n) \sim N(0, \Sigma)$ , i.e.,

$$\begin{aligned} & f(x_1, x_2, \dots, x_n) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{|\det(\Sigma)|^{1/2}} \exp\left(-\frac{(x_1, x_2, \dots, x_n)\Sigma^{-1}(x_1, x_2, \dots, x_n)^T}{2}\right) \end{aligned}$$

do we have  $\text{Var}(X_1, X_2, \dots, X_n) = \Sigma$ ? Here we assume that  $\Sigma$  is positive definite.

**Remark:** Let  $C$  be a matrix such that  $C^2 = \Sigma$  with  $C = C^T$  (this is always possible for a positive definite matrix  $\Sigma$ ). Define

$$(U_1, U_2, \dots, U_n) = (X_1, X_2, \dots, X_n) C^{-1}.$$

The joint density function  $\psi(u_1, u_2, \dots, u_n)$  for  $(U_1, U_2, \dots, U_n)$  is

$$\begin{aligned} & \psi(u_1, u_2, \dots, u_n) \\ &= \frac{1}{|\det(B^{-1})|} f((u_1, u_2, \dots, u_n)C) \\ &= \frac{1}{|\det(C^{-1})|} \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{|\det(\Sigma)|^{1/2}} \exp\left(-\frac{(u_1, u_2, \dots, u_n)C\Sigma^{-1}C(u_1, u_2, \dots, u_n)^T}{2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{u_1^2 + u_2^2 + \dots + u_n^2}{2}\right) \end{aligned}$$

Then  $U_1, U_2, \dots, U_n$  are i.i.d.  $N(0, 1)$ .

For  $(X_1, X_2, \dots, X_n) \sim N(0, \Sigma)$ , find  $C = (c_{ij})$  be a matrix such that  $C^2 = \Sigma$  and define

$$(U_1, U_2, \dots, U_n) = (X_1, X_2, \dots, X_n) C^{-1}.$$

Then we can write

$$(X_1, X_2, \dots, X_n) = (U_1, U_2, \dots, U_n) C.$$

It is easy to see

$$EX_i = 0 \text{ for all } i$$

and

$$\begin{aligned} & E((X_i - EX_i)(X_j - EX_j)) \\ &= E(X_i X_j) = \sum_{k=1}^n c_{ki} c_{kj} \end{aligned}$$

By definition of  $C$  we have  $\sum_{k=1}^n c_{ki} c_{kj}$  is exactly the  $(i, j)$ -th element of  $\Sigma = C^T C$ .