556: Mathematical Statistics I

THE CENTRAL LIMIT THEOREM

THEOREM (THE LINDEBERG-LÉVY CENTRAL LIMIT THEOREM)
Suppose $X_1, ..., X_n$ are i.i.d. random variables with mgf M_X, with

$$\mathrm{E}_{f_X}[X_i] = \mu \qquad \mathrm{Var}_{f_X}[X_i] = \sigma^2$$

both finite. Let the random variable Z_n be defined by

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$$

where

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

and denote by M_{Z_n} the mgf of Z_n. Then, as $n \longrightarrow \infty$,

$$M_{Z_n}(t) \longrightarrow \exp\{t^2/2\}$$

irrespective of the form of M_X. Thus, as $n \longrightarrow \infty$, $Z_n \xrightarrow{d} Z \sim N(0,1)$.

Proof. First, let $Y_i = (X_i - \mu)/\sigma$ for $i = 1, ..., n$. Then $Y_1, ..., Y_n$ are i.i.d. with mgf M_Y say, and by the elementary properties of expectation, $\mathrm{E}_{f_Y}[Y_i] = 0$,$\mathrm{Var}_{f_Y}[Y_i] = 1$ for each i. Using the power series expansion result for mgfs, we have that

$$M_Y(t) = 1 + tE_{f_Y}[Y] + \frac{t^2}{2!}E_{f_Y}[Y^2] + \frac{t^3}{3!}E_{f_Y}[Y^3] + \ldots = 1 + \frac{t^2}{2!} + O(t^3)$$

using the $O(t^3)$ notation to capture all terms involving t^3 and higher power. Now, the random variable Z_n can be rewritten

$$Z_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i$$

and thus, again by a standard mgf result, as $Y_1, ..., Y_n$ are independent, we have that

$$M_{Z_n}(t) = \prod_{i=1}^{n}\left\{M_Y(t/\sqrt{n})\right\} = \left\{1 + \frac{t^2}{2n} + O(n^{-3/2})\right\}^n = \left\{1 + \frac{t^2}{2n} + o(n^{-1})\right\}^n.$$

As $n \longrightarrow \infty$, by the definition of the exponential function

$$M_{Z_n}(t) \longrightarrow \exp\{t^2/2\} \qquad \therefore \qquad Z_n \xrightarrow{d} Z \sim N(0,1)$$

where no further assumptions on M_X are required. ∎

Alternative statement: The theorem can also be stated in terms of

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}} = \sqrt{n}(\overline{X}_n - \mu)$$

so that

$$Z_n \xrightarrow{d} Z \sim N(0, \sigma^2).$$

and σ^2 is termed the **asymptotic variance** of Z_n.

Notes :

(i) The theorem requires the **existence of the mgf** M_X.

(ii) The theorem holds for the i.i.d. case, but there are similar theorems for **non identically distributed**, and **dependent** random variables.

(iii) The theorem allows the construction of **asymptotic normal approximations**. For example, for **large but finite** n, by using the properties of the Normal distribution,

$$\begin{aligned} \overline{X}_n &\sim AN(\mu, \sigma^2/n) \\ S_n = \sum_{i=1}^{n} X_i &\sim AN(n\mu, n\sigma^2). \end{aligned}$$

where $AN(\mu, \sigma^2)$ denotes an asymptotic normal distribution. The notation

$$\overline{X}_n \mathrel{\dot{\sim}} N(\mu, \sigma^2/n)$$

is sometimes used.

(iv) The **multivariate version** of this theorem can be stated as follows: Suppose $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ are i.i.d. k-dimensional random variables with mgf $M_{\underset{\sim}{X}}$, with

$$\mathrm{E}_{f_{\underset{\sim}{X}}}[\underset{\sim}{X}_i] = \underset{\sim}{\mu} \qquad \mathrm{Var}_{f_{\underset{\sim}{X}}}[\underset{\sim}{X}_i] = \Sigma$$

where Σ is a positive definite, symmetric $k \times k$ matrix defining the variance-covariance matrix of the $\underset{\sim}{X}_i$. Let the random variable $\underset{\sim}{Z}_n$ be defined by

$$\underset{\sim}{Z}_n = \sqrt{n}(\overline{\underset{\sim}{X}}_n - \underset{\sim}{\mu})$$

where

$$\overline{\underset{\sim}{X}}_n = \frac{1}{n} \sum_{i=1}^{n} \underset{\sim}{X}_i.$$

Then

$$\underset{\sim}{Z}_n \xrightarrow{d} \underset{\sim}{Z} \sim N(0, \Sigma)$$

as $n \longrightarrow \infty$.