1. Week 13 Supplementary material

Remark 1.1. Let $X_1, X_2, \cdots$ be i.i.d. RVs such that $\mathbb{E}X_1^2$ exists. Consider the sample variance $S_n^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{n}{n-1}(\bar{X}_n)^2$. By the assumption, the RVs X_i^2 are i.i.d. with finite expectation $\mathbb{E}X_1^2$, and hence by the WLLN

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow[n\to\infty]{P} \mathbb{E}X_1^2.$$

Again by WLLN $\bar{X}_n \xrightarrow[n\to\infty]{P} \mathbb{E}X_1$ and by the continuous mapping theorem applied to the function $h(x) = x^2, \forall x \in \mathbb{R}$, we have

$$(\bar{X}_n)^2 \xrightarrow[n\to\infty]{P} (\mathbb{E}X_1)^2.$$

Since $\frac{n}{n-1} \xrightarrow{n\to\infty} 1$, we have $S_n^2 \xrightarrow[n\to\infty]{P} Var(X_1)$. By the continuous mapping theorem, we have $S_n \xrightarrow[n\to\infty]{P} \sqrt{Var(X_1)}$.

Example 1.2. Consider the discrete RVs X_n with the p.m.f.s and DFs given by

$$f_{X_n}(x) = \begin{cases} \frac{1}{2}, & \text{if } x \in \{\frac{1}{2n}, \frac{1}{n}\}, \\ 0, & \text{otherwise} \end{cases}, \quad F_{X_n}(x) = \begin{cases} 0, & \text{if } x < \frac{1}{2n}, \\ \frac{1}{2}, & \text{if } \frac{1}{2n} \le x < \frac{1}{n}, \\ 1, & \text{if } x \ge \frac{1}{n}. \end{cases}$$

Since

$$\lim_n F_{X_n}(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1, & \text{if } x > 0 \end{cases}$$

equals the DF F of the degenerate RV at 0, except at the point of discontinuity 0 of F, we have $X_n \xrightarrow[n\to\infty]{d} 0$. However, $\lim_n f_{X_n}(0) = 0 \neq 1 = f_X(0)$. Here, the pointwise convergence of the p.m.f.s do not hold.

Example 1.3. Let $X, X_1, X_2, \cdots$ be independent RVs with $X \sim N(0,1)$ and $X_n \sim N(0, 1 + \frac{1}{n})$. Looking at the MGFs we have

$$\lim_n M_{X_n}(t) = \lim_n \exp\left(\frac{1}{2}\left(1 + \frac{1}{n}\right)t^2\right) = \exp\left(\frac{1}{2}t^2\right) = M_X(t), \forall t \in \mathbb{R}.$$

Therefore, $X_n \xrightarrow[n\to\infty]{d} X$. However, using the independence of $X, X_1, X_2, \cdots$, we have $X_n - X \sim N(0, 2 + \frac{1}{n})$ and an argument similar to above shows that $X_n - X \xrightarrow[n\to\infty]{d} Z$, where $Z \sim N(0, 2)$. Here, $X_n - X$ does not converge to the degenerate RV at 0.

Example 1.4 (Application of Poisson approximation to Binomial distribution)**.** If

$$X \sim Binomial(1000, 0.003),$$

then the exact value of

$$\mathbb{P}(X = 5) = \binom{1000}{5}(0.003)^5(0.997)^{995}$$

is hard to compute. Instead, we can approximate the value by $\mathbb{P}(Y = 5)$ where $Y \sim Poisson(1000 \times 0.003) = Poisson(3)$. Here, $\mathbb{P}(Y = 5) = e^{-3}\frac{3^5}{5!}$ is comparatively easier to compute.

Theorem 1.5 (Lindeberg-Lévy Central Limit Theorem (CLT))**.** *Let $X_1, X_2, \cdots$ be i.i.d. RVs such that $\mathbb{E}X_1^2$ exists and $Var(X_1) = \sigma^2 > 0$. Then,*

$$\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} \xrightarrow[n\to\infty]{d} Z,$$

where $Z \sim N(0, 1)$.

Remark 1.6 (Restatements of the CLT)*.* Under the assumptions of the CLT above, we can restate the conclusion in various useful ways. Note that the DF Φ of $Z \sim N(0, 1)$ is continuous everywhere on $\mathbb{R}$.

(a) $\lim_{n\to\infty} \mathbb{P}(\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} \leq x) = \Phi(x), \forall x \in \mathbb{R}$.

(b) For all $a < b$, we have

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(a < \sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} \leq b) \\
&= \lim_{n\to\infty} \mathbb{P}(\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} \leq b) - \lim_{n\to\infty} \mathbb{P}(\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} \leq a) \\
&= \Phi(b) - \Phi(a) \\
&= \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt
\end{aligned}$$

(c) Writing $Y_n = X_1 + X_2 + \cdots + X_n$, for all $a < b$ we have

$$\lim_{n\to\infty} \mathbb{P}(a < \frac{Y_n - n\mathbb{E}X_1}{\sigma\sqrt{n}} \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)\, dt$$

Theorem 1.7 (Lindeberg-Feller CLT)**.** *Let $\{X_n\}_n$ be independent RVs with $\mathbb{E}X_n = m_n$ and $Var(X_n) = \sigma_n^2$. We have $C_n^2 := Var(S_n) = \sum_{j=1}^n \sigma_j^2$, where $S_n := X_1 + \cdots + X_n, n \geq 1$. If for all $\epsilon > 0$,*

$$\lim_{n\to\infty} \frac{1}{C_n^2} \sum_{j=1}^n \mathbb{E}\left[(X_j - m_j)^2 1_{\{|X_j - m_j| \geq \epsilon C_n\}}\right] = 0 \tag{1.1}$$

then

$$\frac{S_n - \mathbb{E}S_n}{C_n} \xrightarrow[n\to\infty]{d} Z,$$

where $Z \sim N(0,1)$.

Theorem 1.8 (Lyapunov CLT)**.** *Continue with the notations of the Lindeberg-Feller CLT. If*

$$\lim_{n\to\infty} \frac{1}{C_n^{2+\delta}} \sum_{j=1}^n \mathbb{E}|X_j - m_j|^{2+\delta} = 0$$

for some $\delta > 0$, then the condition (1.1) *stated in the Lindeberg-Feller CLT holds for all $\epsilon > 0$ and*

$$\frac{S_n - \mathbb{E}S_n}{C_n} \xrightarrow[n\to\infty]{d} Z,$$

where $Z \sim N(0,1)$.

Example 1.9 (Application of CLT)**.** It is known that the lifetime of a batch of electric bulbs follow the *Exponential*(10) distribution. Given a box containing 100 electric bulbs, we are interested in the probability (approximately) that 20 bulbs or more have a lifetime at least $10\ln 2$. Using CLT, we can get an approximate value in terms of Φ, the DF of $N(0,1)$ distribution.
For $X \sim$ *Exponential*(10),

$$\mathbb{P}(X \geq 10\ln 2) = \int_{10\ln 2}^{\infty} \frac{1}{10} \exp\left(-\frac{t}{10}\right)\, dt \overset{t=10x}{=} \int_{\ln 2}^{\infty} \exp(-x)\, dx = \exp(-\ln 2) = \frac{1}{2}.$$

Let $X_1, \cdots, X_{100}$ denote the lifetime of the electric bulbs. Define

$$Y_i = \begin{cases} 1, & \text{if } X_i \geq 10 \ln 2, \\ 0, & \text{otherwise} \end{cases}, \forall i = 1, \cdots, 100.$$

Then, Y_i's are i.i.d. with $Y_1 \sim Bernoulli(\mathbb{P}(X > 10 \ln 2)) = Bernoulli(\frac{1}{2})$. Here, $\mathbb{E} Y_1 = \frac{1}{2}$ and $Var(Y_1) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.
By the CLT, for large n the distribution of

$$\sqrt{n} \frac{\frac{1}{n} \sum_{i=1}^n Y_i - \frac{1}{2}}{\sqrt{\frac{1}{4}}}$$

is close to $N(0,1)$ in the sense of convergence in distribution. Putting $n = 100$, the distribution of $20 \left(\frac{1}{100} \sum_{i=1}^{100} Y_i - \frac{1}{2} \right)$ is close to $N(0,1)$ in the sense of convergence in distribution.
The required probability is

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{100} Y_i \geq 20 \right) &= \mathbb{P}\left(\frac{1}{100} \sum_{i=1}^{100} Y_i - \frac{1}{2} \geq \frac{20}{100} - \frac{1}{2} \right) \\ &= \mathbb{P}\left(\frac{1}{100} \sum_{i=1}^{100} Y_i - \frac{1}{2} \geq -\frac{3}{10} \right) \\ &= \mathbb{P}\left(20 \left\{ \frac{1}{100} \sum_{i=1}^{100} Y_i - \frac{1}{2} \right\} \geq -6 \right) \\ &= 1 - \mathbb{P}\left(20 \left\{ \frac{1}{100} \sum_{i=1}^{100} Y_i - \frac{1}{2} \right\} < -6 \right) \end{aligned}$$

Using the above convergence, an approximate value of the required probability is $1 - \Phi(-6)$, where Φ is the DF of $N(0,1)$ distribution.

Remark 1.10 (From CLT to WLLN). Our motivation to study CLT type results was to find a 'rate of convergence' for the WLLN. In principle, a convergence result with a 'rate of convergence' is stronger than another convergence result without any clear 'rate of convergence'. We illustrate this idea by deriving the WLLN from the CLT. Under the assumptions of CLT (i.i.d. RVs with finite second moment), we have

$$\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} \xrightarrow[n\to\infty]{d} Z,$$

where $Z \sim N(0,1)$. Since $\frac{\sigma}{\sqrt{n}} \xrightarrow{n\to\infty} 0$, by Slutky's theorem,

$$\bar{X}_n - \mathbb{E}X_1 \xrightarrow[n\to\infty]{d} 0 \times Z = Y,$$

where Y denotes an RV degenerate at 0. This is equivalent to $\bar{X}_n - \mathbb{E}X_1 \xrightarrow[n\to\infty]{P} 0$ and hence, we conclude $\bar{X}_n \xrightarrow[n\to\infty]{P} \mathbb{E}X_1$, which is the WLLN. Note that, however, to show this we needed the additional assumption on the second moments, which is not required if we are only interested in the WLLN.

Remark 1.11. Let $X_1, X_2, \cdots$ be i.i.d. RVs such that $\mathbb{E}X_1^2$ exists and $Var(X_1) = \sigma^2 > 0$. By the CLT,

$$\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{\sigma} \xrightarrow[n\to\infty]{d} Z,$$

where $Z \sim N(0,1)$. From Remark 1.1 we have

$$\frac{\sigma}{S_n} \xrightarrow[n\to\infty]{P} 1,$$

where $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ is the sample variance. By Slutky's theorem,

$$\sqrt{n}\frac{\bar{X}_n - \mathbb{E}X_1}{S_n} \xrightarrow[n\to\infty]{d} Z.$$

Remark 1.12. Let $X_1, X_2, \cdots$ be i.i.d. $Uniform(0,\theta)$ RVs, for some $\theta > 0$. Recall that $X_{(n)} = \max\{X_1, X_2, \cdots, X_n\} \xrightarrow[n\to\infty]{P} \theta$. We can now ask about the limiting distribution of $(\theta - X_{(n)})$ to understand the rate of convergence. Recall that the p.d.f. of $X_{(n)}$ is given by

$$g_{X_{(n)}}(x) = \begin{cases} \frac{n}{\theta^n}x^{n-1}, \text{ if } x \in (0,\theta), \\ 0, \text{ otherwise.} \end{cases}$$

Look at $Y_n := n(\theta - X_{(n)})$. Then for all $y \in \mathbb{R}$,

$$F_{Y_n}(y) = \mathbb{P}(Y_n \leq y)$$

$$
\begin{aligned}
&= \mathbb{P}\left(X_{(n)} \geq \theta - \frac{y}{n}\right) \\
&= \int_{\theta - \frac{y}{n}}^{\infty} g_{X_{(n)}}(x)\, dx \\
&= \begin{cases} 0, \text{ if } y \leq 0 \\ 1 - \left(1 - \frac{y}{n\theta}\right)^n, \text{ if } 0 < y < n\theta, \\ 1, \text{ if } y > n\theta \end{cases} \\
&\xrightarrow{n\to\infty} \begin{cases} 0, \text{ if } y \leq 0 \\ 1 - \exp\left(-\frac{y}{\theta}\right), \text{ if } y > 0 \end{cases} \\
&= F_Y(y)
\end{aligned}
$$

where $Y \sim Exponential(\theta)$. Since the DF F_Y of Y is continuous everywhere, from the above computation we conclude that $Y_n = n(\theta - X_{(n)}) \xrightarrow[n\to\infty]{d} Y \sim Exponential(\theta)$. The sequence $\{X_{(n)}\}_n$ another example where Delta method can be applied.

Remark 1.13. If $Z \sim N(0,1)$, it can be checked that $\mathbb{P}(|Z| \leq 3) \approx 0.997$ and $\mathbb{P}(|Z| \leq 6) \approx 0.9997$. More generally, for $X \sim N(\mu, \sigma^2)$, we have $\mathbb{P}(|X - \mu| \leq 3\sigma) \approx 0.997$ and $\mathbb{P}(|X - \mu| \leq 6\sigma) \approx 0.9997$. This shows that the values of a normal RV is quite concentrated near its mean.

2. Examples of Characteristic Functions

For a given random variable/vector X, the corresponding Characteristic function is denoted by Φ_X.

(a) If X is degenerate at c, then $\Phi_X(u) = \exp(icu), u \in \mathbb{R}$.
(b) $X \sim Bernoulli(p)$, then $\Phi_X(u) = 1 - p + pe^{iu}, u \in \mathbb{R}$.
(c) $X \sim Binomial(n, p)$, then $\Phi_X(u) = (1 - p + pe^{iu})^n, u \in \mathbb{R}$.
(d) $X \sim Poisson(\lambda)$, then $\Phi_X(u) = \exp(\lambda(\exp(iu) - 1)), u \in \mathbb{R}$.
(e) $X \sim Negative\,Binomial(n, p)$, then $\Phi_X(u) = \left(\frac{p}{1 - e^{iu} + pe^{iu}}\right)^n, u \in \mathbb{R}$.
(f) $X \sim Uniform(a, b)$, then $\Phi_X(u) = \frac{e^{iub} - e^{iua}}{iu(b-a)}, u \in \mathbb{R}$.
(g) $X \sim N(\mu, \sigma^2)$, then $\Phi_X(u) = \exp(i\mu u - \frac{1}{2}\sigma^2 u^2), u \in \mathbb{R}$.

(h) $X \sim Gamma(\alpha, \beta)$, then $\Phi_X(u) = (1 - iu\beta)^{-\alpha}, u \in \mathbb{R}$.
(i) $X \sim Exponential(\lambda)$, then $\Phi_X(u) = (1 - iu\lambda)^{-1}, u \in \mathbb{R}$.
(j) $X \sim \chi^2_k$, then $\Phi_X(u) = (1 - 2iu)^{-\frac{k}{2}}, u \in \mathbb{R}$.
(k) $X \sim N_p(\mu, \Sigma)$, then $\Phi_X(u) = \exp(iu^t\mu - \frac{1}{2}u^t\Sigma u), u \in \mathbb{R}^p$.

3. Descriptive

Remark 3.1 (Descriptive Measures of Probability Distributions). The distribution of an RV provides numerical values through which we can quantify/understand the manner in which the RV takes values in various subsets of the real line. However, at times, it is difficult to grasp the features of the RV from the distribution. As an alternative, we typically use four types of numerical quantities associated with the distribution to summarize the information. We refer to them as descriptive measures of the probability distribution.

(a) Measures of Central Tendency or location: here, we try to find a 'central' value around which the possible values of the RV are distributed.
(b) Measures of Dispersion: once we have an idea of the 'central' value of the RV (equivalently, the probability distribution), we check the scattering/dispersion of the all the possible values of the RV around this 'central' value.
(c) Measures of Skewness: here, we try to quantify the asymmetry of the probability distribution.
(d) Measures of Kurtosis: here, we try to measure the thickness of the tails of the RV (equivalently, the probability distribution) while comparing with the Normal distribution.

We describe these measures along with examples.

Example 3.2 (Measures of Central Tendency).

(a) The Mean of an RV is a good example of a measure of central tendency. It also has the useful property of linearity. However, it may be affected by few extreme values, referred to as the outliers. The mean may not exist for all distributions.
(b) Median, i.e. a quantile of order $\frac{1}{2}$ of an RV is always defined and is usually not affected by a few outliers. However, the median lacks the linearity property, i.e. a median of $X + Y$

has no general relationship with the medians of X and Y. Further, a median focuses on the probabilities with which the values of the RV occur rather than the exact numerical values. A median need not be unique.

(c) The mode m_0 of a probability distribution is the value that occurs with 'highest probability', and is defined by $f_X(m_0) = \sup\{f_X(x) : x \in S_X\}$, where f_X denotes the p.m.f./p.d.f. of X, as appropriate and S_X denotes the support of X. Mode need not be unique. Distributions with one, two or multiple modes are called unimodal, bimodal or multimodal distributions, respectively. Usually, it is easy calculate. However, it may so happen that a distribution has more than multiple modes situated far apart, in which case it may not be suitable for a measure of central tendency.

Example 3.3 (Measures of Dispersion). (a) If the support S_X of an RV X is contained in the interval $[a, b]$ and this is the smallest such interval, then we define $b - a$ to be the range of X. This measure of dispersion does not take into account the probabilities with which the values of X are distributed.

(b) Mean Deviation about a point $c \in \mathbb{R}$: If $\mathbb{E}|X - c|$ exists, we define it to be the mean deviation of X about the point c. Usually, we take c to be the mean (if it exists) or the median and obtain mean deviation about the mean or median, respectively. However, it may be difficult to compute and even may not exist. The mean deviations are also affected by a few outliers.

(c) Standard Deviation: As defined earlier, the standard deviation of an RV X is $\sqrt{Var(X)}$, if it exists. Compared to the mean deviation, the standard deviation is usually easier to compute. The standard deviation is affected by a few outliers.

(d) Quartile Deviation: Recall that $\mathfrak{z}_{0.25}$ and $\mathfrak{z}_{0.75}$ denotes the lower and upper quartiles. We define $\mathfrak{z}_{0.75} - \mathfrak{z}_{0.25}$ to be the inter-quartile range and refer to $\frac{1}{2}[\mathfrak{z}_{0.75} - \mathfrak{z}_{0.25}]$ as the semi-inter-quartile range or the quartile deviation. This measures the spread in the middle half of the distribution and is therefore not influenced by extreme values. However, it does not take into account the numerical values of the RV.

(e) Coefficient of Variation: The coefficient of variation of X is defined as $\frac{\sqrt{Var(X)}}{\mathbb{E}X}$, provided $\mathbb{E}X \neq 0$. This aims to measure the variation per unit of mean. It, by definition, does not depend on the unit of measurement. However, it may be sensitive to small changes in the mean, if it is close to zero.

Note 3.4 (A Measure of Skewness)**.** If the distribution of an RV X is symmetric about the mean μ, then $f_X(\mu + x) = f_X(\mu - x), \forall x \in \mathbb{R}$, where f_X denotes the p.m.f./p.d.f. of X. If this is not the case, then two cases may occur.

(a) (Positively skewed) the distribution may have more probability mass towards the right hand side of the graph of f_X. In this case, the tails on the right hand side are longer.

(b) (Negatively skewed) the distribution may have more probability mass towards the left hand side of the graph of f_X. In this case, the tails on the left hand side are longer.

To measure this asymmetry, we usually look at $\mathbb{E}Z^3$, where $Z = \frac{X - \mathbb{E}X}{\sqrt{Var(X)}}$, provided the moments exist. Note that Z is independent of the units of measurement and

$$\mathbb{E}Z^3 = \frac{\mathbb{E}(X - \mathbb{E}X)^3}{(Var(X))^{\frac{3}{2}}} = \frac{\mu_3(X)}{(\mu_2(X))^{3/2}}.$$

We may refer to a distribution being positively or negatively skewed according as the above quantity being positive or negative. If $X \sim Exponential(\lambda)$, then $\mathbb{E}Z^3 = 2$ and hence the distribution of X is positively skewed.

Note 3.5. There are many other measures of skewness used in practice. However, we do not discuss them in this course.

Note 3.6 (A measure of Kurtosis)**.** The probability distribution of X is said to have higher (respectively, lower) kurtosis than the Normal distribution, if its p.m.f./p.d.f., in comparison with the p.d.f. of a Normal distribution, has a sharper (respectively, rounded) peak and longer/fatter (respectively, shorter/thinner) tails. To measure the kurtosis of X, we look at $\mathbb{E}Z^4$, where $Z = \frac{X - \mathbb{E}X}{\sqrt{Var(X)}}$, provided the moments exist. Note that Z is independent of the units of measurement and

$$\mathbb{E}Z^4 = \frac{\mathbb{E}(X - \mathbb{E}X)^4}{(Var(X))^2} = \frac{\mu_4(X)}{(\mu_2(X))^2}.$$

If $X \sim N(\mu, \sigma^2)$, then $Z \sim N(0,1)$ and hence $\mathbb{E}Z^4 = 3$. For a general RV X, the quantity $\frac{\mu_4(X)}{(\mu_2(X))^2} - 3$ is referred to as the excess kurtosis of X. If the excess kurtosis is zero, positive or negative, then we refer to the corresponding probability distribution as mesokurtic, leptokurtic or platykurtic, respectively. If $X \sim Exponential(\lambda)$, then $\mathbb{E}Z^4 = 9$ and hence the distribution of X is leptokurtic.