12. Week 12

Remark 12.1 (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 12.2 (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 12.3 (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 12.4 (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 12.5. There are many other measures of skewness used in practice. However, we do not discuss them in this course.

Note 12.6 (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 $\mathbb{E}Z^4 = 3$ (see Remark 8.1). 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.

Definition 12.7 (Quantile function of an RV)**.** Let X be an RV with the DF F_X. The function $Q_X : (0,1) \to \mathbb{R}$ defined by

$$Q_X(p) := \inf\{x \in \mathbb{R} : F_X(x) \geq p\}, \forall p \in (0,1)$$

is called the quantile function of X.

Proposition 12.8 (Probability integral transform)**.** *Let X be a continuous RV with the DF F_X, p.d.f. f_X and quantile function Q_X.*

(a) We have $F_X(X) \sim Uniform(0,1)$.

(b) For any $U \sim Uniform(0,1)$, we have $Q_X(U) \stackrel{d}{=} X$.

Proof. We prove only the first statement. The proof of the second statement is similar. Take $Y = F_X(X)$. Then,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(F_X(X) \leq y) = \begin{cases} 0, & \text{if } y < 0, \\ 1, & \text{if } y \geq 1. \end{cases}$$

For $y \in [0,1)$, we have

$$\mathbb{P}(F_X(X) = y) = \mathbb{P}(x_1 \leq X \leq x_2) = 0$$

for some $x_1, x_2 \in \mathbb{R}$ with $F_X(x_1) = F_X(x_2)$. Here, we have used the fact that X is a continuous RV. Now, for $y \in [0,1)$,

$$\begin{aligned} \mathbb{P}(F_X(X) \leq y) &= \mathbb{P}(F_X(X) < y) \\ &= 1 - \mathbb{P}(F_X(X) \geq y) \\ &= 1 - \mathbb{P}(X \geq Q_X(y)) \\ &= 1 - \mathbb{P}(X > Q_X(y)) \\ &= \mathbb{P}(X \leq Q_X(y)) \end{aligned}$$

$$= F_X(Q_X(y))$$
$$= y.$$

Hence, $Y = F_X(X) \sim Uniform(0,1)$. This completes the proof. □

Note 12.9. Let X be an RV with the quantile function Q_X. If we can generate random samples $U_1, U_2, \cdots, U_n$ from $U \sim Uniform(0,1)$, then $Q_X(U_1), Q_X(U_2), \cdots, Q_X(U_n)$ are random samples from the distribution of X. This observation may be used in practice to generate random samples for known distributions from the $Uniform(0,1)$ distribution.

Note 12.10 (Moments do not determine the distribution of an RV)**.** Let $X \sim N(0,1)$ and consider $Y = e^X$. The distribution of Y is usually called the lognormal distribution, since $\ln Y = X \sim N(0,1)$. Using standard techniques, we can compute the p.d.f. of Y:

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-1} \exp\left[-\frac{(\ln y)^2}{2}\right], & \text{if } y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

It can be shown that the continuous RVs $X_\alpha, \alpha \in [-1,1]$ with the p.d.fs

$$f_{X_\alpha}(y) = f_Y(y)\left[1 + \alpha \sin(2\pi \ln y)\right], \forall y \in \mathbb{R}$$

has the same moments as Y. However, the distributions are different. This shows that the moments of an RV do not determine the distribution. (see the article 'On a property of the lognormal distribution' by C.C. Heyde, published in Journal of the Royal Statistical Society: Series B, volume 29 (1963).)

Note 12.11 (Operations on DFs)**.** Recall that a DF $F : \mathbb{R} \to [0,1]$ is characterized by the properties that it is right continuous, non-decreasing and $\lim_{x\to\infty} F(x) = 1, \lim_{x\to-\infty} F(x) = 0$. Given two DFs $F, G : \mathbb{R} \to [0,1]$ and $\alpha \in [0,1]$, we make the following observations.

(a) (Convex combination of DFs) The function $H : \mathbb{R} \to [0,1]$ defined by $H(x) := \alpha F(x) + (1-\alpha)G(x), \forall x \in \mathbb{R}$ has the relevant properties and hence is a DF.

(b) (Product of DFs) The function $H : \mathbb{R} \to [0,1]$ defined by $H(x) := F(x)G(x), \forall x \in \mathbb{R}$ has the relevant properties and hence is a DF. In particular, F^2 is a DF, if F is so.

In fact, a general DF can be written as a convex combination of discrete DFs and some special continuous DFs. We do not discuss such results in this course.

Remark 12.12. In practice, given a known RV X, many times we need to find out the distribution of $h(X)$ for some function $h : \mathbb{R} \to \mathbb{R}$ or even, simply, compute the expectations of the form $\mathbb{E}h(X)$. As already discussed earlier in the course, we can theoretically (i.e., in principle) compute $\mathbb{E}h(X)$ as $\int_{-\infty}^{\infty} h(x) f_X(x)\, dx$, when X is a continuous RV with p.d.f. f_X, for example. However, in practice, it may happen that this integral does not have a closed form expression – which makes it challenging to evaluate. The problem becomes more intractable when we look at similar problems where X is a random vector and the joint/marginal distributions need to be considered. In such situations, as an alternative, we try to find 'good' approximations for the quantities of interest, where the approximation terms are easier to compute than the original expression. This motivation leads to the various notions for convergence of RVs. If some quantity of interest involving an RV X, say $\mathbb{E}X$, is difficult to compute, then we find an appropriate 'approximating' sequence of RVs $\{X_n\}_n$ for X and use the values $\mathbb{E}X_n$ as an approximation for $\mathbb{E}X$.

Remark 12.13. Given a random sample $X_1, X_2, \cdots, X_n$ from $N(\mu, \sigma^2)$ distribution, consider the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Here, we have written $\bar{X}_n$, instead of just $\bar{X}$, to highlight the dependence of the sample mean on the sample size n. Recall that $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. The behaviour of $\bar{X}_n$ for large n is of interest. This is also another motivation for us to study the convergence of sequences of RVs.

We now discuss concepts for convergence of sequences of RVs.

Definition 12.14 (Convergence in r-th mean)**.** Let $X, X_1, X_2, \cdots$ be RVs defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $r \geq 1$. If $\mathbb{E}|X|^r < \infty, \mathbb{E}|X_n|^r < \infty, \forall n$ and if

$$\lim_{n\to\infty} \mathbb{E}|X_n - X|^r = 0,$$

then we say that the sequence $\{X_n\}_n$ converges to X in r-th mean.

Note 12.15. (a) If a sequence $\{X_n\}_n$ converges to X in r-th mean for some $r \geq 1$, then we have

$$\lim_{n\to\infty} \mathbb{E}|X_n|^r = \mathbb{E}|X|^r,$$

and

$$\lim_{n\to\infty} \mathbb{E}X_n^r = \mathbb{E}X^r,$$

i.e., we have the convergence of the r-th moments.

(b) The sequence $\{X_n\}_n$ converges to X in r-th mean if and only if the sequence $\{X_n - X\}_n$ converges to 0 in r-th mean.

Remark 12.16. Even though we have defined the r-th order moments for $0 < r < 1$, for technical reasons we do not consider the convergence in r-th mean in this case. The details are beyond the scope of this course. In what follows, whenever we consider the convergence in r-th mean, we assume $r \geq 1$.

Definition 12.17 (Convergence in Probability). Let $X, X_1, X_2, \cdots$ be RVs defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If for all $\epsilon > 0$, we have

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0,$$

then we say that the sequence $\{X_n\}_n$ converges to X in probability and write $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$.

Note 12.18. (a) Suppose that a sequence $\{X_n\}_n$ converges to X in probability. Now, for all $\epsilon > 0$, note that

$$\mathbb{P}(|X_n - X| \geq 2\epsilon) \leq \mathbb{P}(|X_n - X| > \epsilon) \leq \mathbb{P}(|X_n - X| \geq \epsilon).$$

Convergence in probability is equivalent to the fact that

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0.$$

(b) The sequence $\{X_n\}_n$ converges to X in probability if and only if the sequence $\{X_n - X\}_n$ converges to 0 in probability.

Proposition 12.19. *Let $X, X_1, X_2, \cdots$ be RVs defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If the sequence $\{X_n\}_n$ converges to X in r-th mean for some $r \geq 1$, then $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$.*

Proof. By Markov's inequality (Corollary 8.9), we have

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \epsilon^{-r}\ \mathbb{E}|X_n - X|^r.$$

Since $\lim_{n\to\infty} \mathbb{E}|X_n - X|^r = 0$, we have the result. $\square$

Corollary 12.20. *Let $\{X_n\}_n$ be a sequence of RVs with finite second moments. If $\lim_n \mathbb{E}X_n = \mu$ and $\lim_n Var(X_n) = 0$, then $\{X_n\}_n$ converges to μ in 2nd mean and in particular, in probability.*

Proof. We have $\mathbb{E}|X_n - \mu|^2 = \mathbb{E}\left[(X_n - \mu_n) + (\mu_n - \mu)\right]^2 = \mathbb{E}(X_n - \mu_n)^2 + (\mu_n - \mu)^2 = Var(X_n) + (\mu_n - \mu)^2$. By our hypothesis, $\lim_n \mathbb{E}|X_n - \mu|^2 = 0$. Hence, $\{X_n\}_n$ converges to μ in 2nd mean. By Proposition 12.19, the sequence also converges in probability. $\square$

Example 12.21. Let $X_1, X_2, \cdots$ be i.i.d. $Uniform(0, \theta)$ RVs, for some $\theta > 0$. The sequence $\{X_n\}_n$ being i.i.d. means that the collection $\{X_n : n \geq 1\}$ is mutually independent and that all the RVs have the same law/distribution. Here, the common p.d.f. and the common DF are given by

$$f(x) = \begin{cases} \frac{1}{\theta}, & \text{if } x \in (0, \theta), \\ 0, & \text{otherwise} \end{cases}, \quad F(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{x}{\theta}, & \text{if } 0 \leq x < \theta, \\ 1, & \text{if } x \geq \theta. \end{cases}$$

Consider $X_{(n)} = \max\{X_1, X_2, \cdots, X_n\}$. Using Proposition 10.15, we have the marginal p.d.f. of $X_{(n)}$ is given by

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

Then,

$$\mathbb{E}X_{(n)} = \int_0^\theta x \,\frac{n}{\theta^n} x^{n-1}\, dx = \frac{n}{n+1}\theta, \quad \mathbb{E}X_{(n)}^2 = \int_0^\theta x^2\, \frac{n}{\theta^n} x^{n-1}\, dx = \frac{n}{n+2}\theta^2$$

and

$$Var(X_{(n)}) = \theta^2 \left[\frac{n}{n+2} - \left(\frac{n}{n+1} \right)^2 \right] = \theta^2 \frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2} = \theta^2 \frac{n}{(n+2)(n+1)^2}.$$

Now, $\lim_n \mathbb{E}X_{(n)} = \theta$ and $\lim_n Var(X_{(n)}) = 0$. Hence, by Corollary 12.20, $\{X_{(n)}\}_n$ converges in 2nd mean to θ and also in probability.

Remark 12.22 (Convergence in probability does not imply convergence in r-th mean). Consider a sequence of discrete RVs $\{X_n\}_n$ with $X_n \sim Bernoulli(\frac{1}{n}), \forall n$. Consider $Y_n := nX_n, \forall n$. Then Y_n's are also discrete with the p.m.f.s given by

$$f_{Y_n}(y) = \begin{cases} 1 - \frac{1}{n}, & \text{if } y = 0, \\ \frac{1}{n}, & \text{if } y = n, \\ 0, & \text{otherwise.} \end{cases}$$

For all $\epsilon > 0$, we have $\mathbb{P}(|Y_n| \geq \epsilon) = \frac{1}{n} \xrightarrow{n\to\infty} 0$ and hence $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$. But, for any $r > 1$, $\mathbb{E}|Y_n|^r = n^{r-1}, \forall n$. Here, $\{Y_n\}_n$ does not converge to 0 in r-th mean.

Example 12.23. $X_1, X_2, \cdots$ be i.i.d. RVs following $N(\mu, \sigma^2)$ distribution. Recall that $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. Then $\lim_n \mathbb{E}\bar{X}_n = \lim_n \mu = \mu$ and $\lim_n Var(\bar{X}_n) = \lim_n \frac{\sigma^2}{n} = 0$. By Corollary 12.20, $\{\bar{X}_n\}_n$ converges in 2nd mean to μ and also in probability.

The above example leads to the following result.

Theorem 12.24 (Weak Law of Large Numbers (WLLN))**.** *Let $X_1, X_2, \cdots$ be i.i.d. RVs such that $\mathbb{E}X_1$ exists. Then, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}X_1$.*

Remark 12.25. We only discuss the proof of Theorem 12.24, when $\mathbb{E}X_1^2$ exists. The proof of the theorem when $\mathbb{E}X_1^2$ does not exist is beyond the scope of this course. However, we shall use this theorem in its full generality.

Proof of WLLN (Theorem 12.24) (assuming $\mathbb{E}X_1^2 < \infty$). Observe that $\mathbb{E}\bar{X}_n = \frac{1}{n}\sum_{i=1}^n \mathbb{E}X_i = \frac{1}{n}n\mathbb{E}X_1 = \mathbb{E}X_1$ and, using independence of X_i's we have

$$Var(\bar{X}_n) = \frac{1}{n^2}Var\left(\sum_{i=1}^n X_i\right) = \frac{1}{n}Var(X_1) \xrightarrow{n\to\infty} 0.$$

By Corollary 12.20, the result follows. $\square$