8 SAMPLING DISTRIBUTIONS AND LIMIT THEOREMS

Let $n \geq 1$, $X_1, X_2, \ldots, X_n$ be an i.i.d. random sample from a population. Recall the sample mean

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

and sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

We have seen in the previous chapter the significance of the above two sample statistics. In this chapter we shall discuss their distributional properties and limiting behaviour. In the next chapter we shall discuss how these results can be effectively used to verify specific hypotheses about the population. The corresponding field of study is called Hypothesis Testing or Test of Significance.

We will find the distribution of the sample mean and sample variance given the distribution of X_1. One immediately observes that these are somewhat complicated functions of independent random variables. However in Section 3.3 and Section 5.5 we have seen examples of functions for which we were able to explicitly compute their distribution. To understand sampling statistics we must also understand the notion of joint distribution of more than two continuous random variables (See Section 3.3 for discrete random variables).

8.1 MULTI-DIMENSIONAL CONTINUOUS RANDOM VARIABLES

In Chapter 3, while discussing discrete random variables we had considered a finite collection of random variables $(X_1, X_2, \ldots, X_n)$. In Definition 3.2.7, we had described how to define their joint distribution and we used this to understand the multinomial distribution in Example 3.2.12.

In the continuous setting as well there are many instances where it is relevant to study the joint distribution of a finite collection of random variables. Suppose X is a point chosen randomly in the unit sphere in the 3 dimensions. Then X has three coordinates and say $X = (X_1, X_2, X_3)$ where each X_i is a random variable in $(0, 1)$. Also they are dependent because we know that, $\sqrt{X_1^2 + X_2^2 + X_3^2} \leq 1$. It is useful and needed to understand their "joint distribution". We have already seen the usefulness of sample mean and sample variance which are a function of $X_1, X_2, \ldots, X_n$. To understand the distribution of sample mean and sample variance the joint distribution of $X_1, X_2, \ldots, X_n$ will be needed to be understood first. We define the joint distribution function first.

DEFINITION 8.1.1. *Let $n \geq 1$ and $X_1, X_2, \ldots, X_n$ be random variables defined on the same probability space. The joint distribution function $F : \mathbb{R}^n \to [0, 1]$ is given by*

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n),$$

for $x_1, x_2, \ldots, x_n \in \mathbb{R}$.

As in single variable and two variable situations, the joint distribution function determines the entire joint distribution of $X_1, X_2, \ldots, X_n$,. More precisely, if all the random variables were discrete with $X_i : S \to T_i$ with T_i being countable subsets of $\subset \mathbb{R}$ for $1 \leq i \leq n$ from the joint distribution function one can determine

$$P(X_1 = t_1, X_2 = t_2, \ldots, X_n = t_n),$$

for all $t_i \in T_i, 1 \le i \le n$. To understand the random variables in the continuous setting we need to set up some notation.

Let $n \ge 1$ and $f : \mathbb{R}^n \to \mathbb{R}$ be a non-negative function, piecewise-continuous in each variable for which

$$\int_{\mathbb{R}^n} f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n = 1.$$

For a Borel set $A \subset \mathbb{R}^n$ if

$$P(A) = \int_A f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n.$$

Then one can show as in Theorem 5.1.5 that P is a probability on $\mathbb{R}^n$. f is called the density function for P. A sequence of random variables $(X_1, X_2, X_3, \ldots, X_n)$ is said to have a joint density $f : \mathbb{R}^n \to \mathbb{R}$ if for every event $A \subset \mathbb{R}^n$

$$P((X_1, X_2, X_3, \ldots, X_n) \in A) = \int_A f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n.$$

In this setting, the joint distribution of $(X_1, X_2, \ldots, X_n)$ is determined by joint density f. . Using multivariable calculus we can can state and prove a similar type of result as Theorem 5.2.5 for random variables $(X_1, X_2, \ldots, X_n)$ that have a joint density. In particular, we can conclude that since the joint densities are assumed to be piecewise continuous, the corresponding distribution functions are piecewise differentiable. Further, the joint distribution of the continuous random variables $(X_1, X_2, \ldots, X_n)$ are completely determined by their joint distribution function F. That is, if we know $F(x_1, x_2, \ldots, x_n)$ for all $x_1, x_2, \ldots, x_n \in \mathbb{R}$ we could use multivariable calculus to differentiate F to find f. Then integrating this joint density over the event A we can calculate $P((X_1, X_2, \ldots, X_n) \in A)$.

As in the $n = 2$ case one can recover the marginal density of each X_i for i between 1 and n by integrating over the other indices. So, the marginal density of X_i at a is given by

$$f_{X_i}(a) = \int_{\mathbb{R}^{n-1}} f(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n.$$

Further for $n \ge 3$, we can deduce the joint density for any sub-collection $m \le n$ random variables by integrating over the other variables. For instance, if we were interested in the joint density of (X_1, X_3, X_7) we would obtain

$$f_{X_1, X_3, X_7}(a_1, a_3, a_7) = \int_{\mathbb{R}^{n-3}} f(a_1, x_2, a_3, x_4, x_5, x_6, a_7, x_8 \ldots, x_n) dx_2 dx_4 dx_5 dx_6 \ldots dx_n.$$

Suppose X_1, X_2, $\ldots$, X_n are random variables defined on a single sample space S with joint density $f : \mathbb{R}^n \to \mathbb{R}$. Let $g : \mathbb{R}^n \to \mathbb{R}$ be a function of n variables for which $g(X_1, X_2, \ldots, X_n)$ is defined in the range of the X_j variables. Let B be an event in the range of g. Then, following the proof of Theorem 3.3.5, we can show that

$$P(g(X_1, X_2, \ldots, X_n) \in B) = P((X_1, X_2, \ldots, X_n) \in g^{-1}(B)).$$

The above provides an abstract method of finding the distribution of the random variable $Y = g(X_1, X_2, \ldots, X_n)$ but it might be difficult to calculate it explicitly. For $n = 1$, we discussed this question in detail in Section 5.3, for $n = 2$ we did explore how to find the distributions of sums and ratios of independent random variables (see Section 5.5). In a few cases by induction n, this method could be extended but in general it is not possible. In Appendix B, Section B.2 we discuss the Jacobian method of finding the joint density of the transformed random variable.

The notion of independence also extends to the multi-dimensional continuous random variable as in discrete setting. As discussed in Definition 3.2.3, a finite collection of continuous random variables $X_1, X_2, \ldots, X_n$ is mutually independent if the sets $(X_j \in A_j)$ are mutually independent for all events A_j in the ranges of the corresponding X_j. As proved for the $n = 2$ case in Theorem 5.4.7, we can similarly deduce that if $(X_1, X_2, X_3, \ldots, X_n)$ are mutually independent continuous random variables with marginal densities f_{X_i} then their joint density is given by

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i), \tag{8.1.1}$$

for $x_i \in \mathbb{R}$ and $1 \le i \le n$. Further for any finite sub-collection $\{X_{i_1}, X_{i_2}, \ldots, X_{i_m}\}$ of the above independent random variables, their joint density is given by

$$f(a_1, a_2, \ldots, a_m) = \prod_{i=j}^{m} f_{X_{i_j}}(a_j). \tag{8.1.2}$$

We conclude this section with a result that we will repeatedly use.

THEOREM 8.1.2. *For each $j \in \{1, 2, \ldots, n\}$ define a positive integer m_j and suppose $X_{i,j}$ is an array of mutually independent continuous random variables for $j \in \{1, 2, \ldots, n\}$ and $i \in \{1, 2, \ldots, m_j\}$. Let $g_j(\cdot)$ be functions such that the quantity*

$$Y_j = g_j(X_{1,j}, X_{2,j}, \ldots X_{m_j,j})$$

is defined for the outputs of the $X_{i,j}$ variables. Then the resulting variables $Y_1, Y_2, \ldots, Y_n$ are mutually independent.

Proof- Follows by the same proof presented in Theorem 3.3.6. ■

8.1.1 *Order Statistics and their Distributions*

Let $n \ge 1$ and let $X_1, X_2, \ldots, X_n$ be a i.i.d random sample from a population. Let F be the common distribution function. Let the X's be arranged in increasing order of magnitude denoted by

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}.$$

These ordered values are called the order statistics of the sample $X_1, X_2, \ldots, X_n$. For, $1 \le r \le n$, $X_{(r)}$ is called the r-th order statistic. One can computer $F_{(r)}$, the distribution function of $X_{(r)}$, for $1 \le r \le n$ in terms of n and F. We have,

$$\begin{aligned} F_{(1)}(x) &= P(X_{(1)} \le x) = 1 - P(X_{(1)} > x) = 1 - P(\cap_{1=1}^{n}(X_i > x)) \\ &= 1 - \prod_{i=1}^{n} P(X_i > x) = 1 - \prod_{i=1}^{n}(1 - P(X_i \le x)) \\ &= 1 - (1 - F(x))^n, \end{aligned}$$

$$F_{(n)}(x) = P(X_{(n)} \le x) = P(\cap_{i=1}^{n}(X_i \le x)) = \prod_{i=1}^{n} P(X_i \le x) = (F(x))^n,$$

and for $1 < r < n$,

$$\begin{aligned} F_{(r)}(x) &= P(X_{(r)} \le x) = P(\text{at least } r \text{ elements from the sample are} \le x) \\ &= \sum_{j=r}^{n} P(\text{exactly } j \text{ elements from the sample are} \le x) \\ &= \sum_{j=r}^{n} \binom{n}{j} P(\text{chosen } j \text{ elements from the sample are} \le x) \times \\ &\qquad \times P((\ n-j \text{ elements not chosen from the sample are} > x) \\ &= \sum_{j=r}^{n} \binom{n}{j} F(x)^j (1 - F(x))^{n-j} \end{aligned}$$

If the distribution function F had a probability density function f then each $X_{(r)}$ has a probability density function $f_{(r)}$. This can be obtained by differentiating $F_{(r)}$ and is given by the below expression.

$$f_{(r)}(x) = \begin{cases} n(1-F(x))^{n-1}f(x) & r=1 \\ nf(x)(F(x))^{n-1} & r=n \\ \frac{n!}{(r-1)!(n-r)!}f(x)(F(x))^{r-1}(1-F(x))^{n-r} & 1<r<n \end{cases} \tag{8.1.3}$$

Example 8.1.3. Let $n \geq 1$ and let $X_1, X_2, \ldots, X_n$ be a i.i.d random sample from a population whose common distribution F is an Exponential (λ) random variable. Then we know that

$$F(x) = \begin{cases} 0 & x<0 \\ 1-e^{-\lambda x} & x \geq 0 \end{cases}$$

Therefore using (8.1.3) and substituting for F as above we have that the densities of the order statistics are given by

$$f_{(r)}(x) = \begin{cases} n(e^{-\lambda x}))^{n-1}\lambda e^{-\lambda x} & r=1 \\ n\lambda e^{-\lambda x}(1-e^{-\lambda x})^{n-1} & r=n \\ \lambda e^{-\lambda x}\frac{n!}{(r-1)!(n-r)!}(1-e^{-\lambda x})^{r-1}(e^{-\lambda x})^{n-r} & 1<r<n, \end{cases}$$

for $x>0$. Simplifying the algebra we obtain,

$$f_{(r)}(x) = \begin{cases} n\lambda e^{-n\lambda x} & r=1 \\ n\lambda e^{-\lambda x}(1-e^{-\lambda x})^{n-1} & r=n \\ \frac{\lambda n!}{(r-1)!(n-r)!}(1-e^{-\lambda x})^{r-1}(e^{-\lambda x})^{n-r+1} & 1<r<n, \end{cases}$$

for $x>0$. We note from the above that $X_{(1)}$, i.e minimum of exponentials, is Exponential ($n\lambda$) random variable. However the other order statistics are not exponentially distributed.

■

In many applications one is interested in the range of values a random variable X assumes. A method to understand this to sample $X_1, X_2, \ldots, X_n$ i.i.d. X and examine $R = X_{(n)} - X_{(1)}$. Suppose X has a probability density function $f : \mathbb{R} \to \mathbb{R}$ and distribution function $F : \mathbb{R} \to [0,1]$. As before we can can calculate the joint density of $X_{(1)}, X_{(n)}$ by first computing the joint distribution function. This is done by using the i.i.d. nature of the sample and the definition of the order statistics.

$$\begin{aligned} P(X_{(1)} \leq x, X_{(n)} \leq y) &= P(X_{(n)} \leq y) - P(x < X_{(1)}, X_{(n)} \leq y) \\ &= P(\cap_{i=1}^n \{X_i \leq y\}) - P(\cap_{i=1}^n \{x < X_i \leq y\}) \\ &= [P(X \leq y)]^n - [P(x < X \leq y)]^n \\ &= \begin{cases} [F(x)]^n - [F(y)-F(x)]^n & x<y \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

From the above, differentiating partially in x and y we see that the joint density of $(X_{(1)}, X_{(n)})$ is given by

$$f_{X_{(1)},X_{(n)}}(x,y) \begin{cases} n(f(x)-f(y))[F(y)-F(x)]^{n-1} & x<y \\ 0 & \text{otherwise.} \end{cases} \tag{8.1.4}$$

To calculate the distribution of R, we compute its distribution function. For $r \leq 0$, $P(R \leq r) = 0$ and for $r > 0$, using the above joint density of $(X_{(1)}, X_{(n)})$ we have

$$\begin{aligned} P(R \leq r) &= P(X_{(n)} \leq X(1) + r) \\ &= \int_{-\infty}^{\infty} \left[\int_0^r f_{X_{(1)},X_{(n)}}(x, z+x) dz \right] dx \\ &= \int_0^r \left[\int_{-\infty}^{\infty} f_{X_{(1)},X_{(n)}}(x, z+x) dx \right] dz, \end{aligned}$$

where we have done a change of variable $y = z + x$ in the second last line and a change in the order of integration in the last line. Differentiating the above we conclude that R has a joint density given by

$$f_R(r) = \begin{cases} \int_{-\infty}^{\infty} f_{X_{(1)},X_{(n)}}(x, r+x) dx & \text{if } r > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{8.1.5}$$

EXAMPLE 8.1.4. Let $X_1, X_2, \ldots, X_n$ be i.i.d Uniform $(0,1)$. The probability density function and distribution function of a uniform $(0,1)$ random variable are given by

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

Using (8.1.3), we have the probability density function of

$$X_{(1)} \text{ is given by} \quad f_{X_{(1)}}(x) = \begin{cases} n(1-x)^{n-1} & \text{if } x \in (0,1) \\ 0 & \text{otherwise,} \end{cases}$$

$$X_{(n)} \text{ is given by} \quad f_{X_{(n)}}(x) = \begin{cases} nx^{n-1} & \text{if } x \in (0,1) \\ 0 & \text{otherwise.} \end{cases}$$

$$X_{(r)} \text{ is given by} \quad f_{X_{(r)}}(x) = \begin{cases} \frac{n!}{(r-1)!(n-r)!} x^{r-1}(1-x)^{n-r} & \text{if } x \in (0,1) \\ 0 & \text{otherwise.} \end{cases}$$

for $1 < r < n$.

Using (8.1.4), we have the joint density of

$$(X_{(1)}, X_{(n)}) \text{ is given by} \quad f_{X_{(1)},X_{(n)}}(x,y) = \begin{cases} n(n-1)(y-x)^{n-1} & \text{if } 0 \leq x \leq y \leq 1) \\ 0 & \text{otherwise,} \end{cases}$$

Using (8.1.5), we have the probability density function of the range

$$\begin{aligned} R = X_{(n)} - X_{(1)} \text{ is given by } f_R(r) &= \begin{cases} \int_0^{1-r} n(n-1)(x+r-x)^{n-1} dx & \text{if } 0 < r < 1 \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} n(n-1)r^{n-1}(1-r) & \text{if } 0 < r < 1 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

We see that $X_{(r)} \sim$ Beta $(r, n-r+1)$ for $1 \leq r \leq n$ and the range $R \sim$ Beta$(n, 2)$ ■

In general we can also understand the joint-distribution of the order statistics. Suppose we have an i.i.d sample $X_1, X_2, \ldots, X_n$ having distribution X. If X has a probability density function $f : \mathbb{R} \to \mathbb{R}$ then one can show that the order statistic $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ has a joint density $h : \mathbb{R}^n \to \mathbb{R}$ by

$$h(u_1, u_2, \ldots, u_n) \begin{cases} n! f(u_1) f(u_2) \ldots f(u_n) & u_1 < u_2 < \ldots < u_n, \\ 0 & \text{otherwise.} \end{cases}$$

The above fact intuitively is clear. Any ordering $u_1 < u_2 < \ldots < u_n$ "has a probability" $f(u_1)f(u_2)\ldots f(u_n)$. Each of the X_i can assume any of the u_k's. The total number of possible orderings is $n!$. A formal proof involves using the Jacobian method and will be discussed in Appendix B.

8.1.2 χ^2, *F and t*

χ^2, F and t distributions arise naturally when considering functions of i.i.d. normal random variables $(X_1, X_2, X_3, \ldots, X_n)$ for $n \geq 1$. They also are useful in Hypothesis testing as well. We discuss these via three examples.

Example 8.1.5. (Chi-Square) Let $n \geq 1$ and $(X_1, X_2, X_3, \ldots, X_n)$ be a collection of independent Normal random variables with mean 0 and variance 1. Then the joint density is given by

$$f(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) = \frac{1}{\sqrt[n]{2\pi}} e^{-\sum_{i=1}^{n} \frac{x_i^2}{2}},$$

for $x_i \in \mathbb{R}$ and $1 \leq i \leq n$.

Let $Z = \sum_{i=1}^{n} X_i^2$. We shall find the distribution of Z in two steps. First, clearly the range of X_1^2 is non-negative. The distribution function for X_1^2 at $z \geq 0$, is given by

$$\begin{aligned} F_1(z) &= P(X_1^2 \leq z) \\ &= P(X_1 \leq \sqrt{z}) \\ &= \int_0^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \int_0^{z} \frac{1}{2\sqrt{2\pi}} e^{-\frac{u}{2}} u^{-\frac{1}{2}} du \end{aligned}$$

Comparing it with the Gamma (α, λ) random variable defined in Definition 5.5.5 and using Exercise 5.5.10, we see that X_1^2 is distributed as a Gamma $(\frac{1}{2}, \frac{1}{2})$ random variable. Using the calculation done in Example 5.5.6 for $n = 2$ and by induction we have that $Z = \sum_{i=1}^{n} X_i^2$ will be Gamma $(\frac{n}{2}, \frac{1}{2})$. This distribution is referred to as Chi-Square with n- degrees of freedom. We define it precisely next. ■

Definition 8.1.6. *(Chi-Square with n degrees of freedom) A random variable X whose distribution is Gamma $(\frac{n}{2}, \frac{1}{2})$ is said to have Chi-square distribution with n-degrees of freedom (i.e number of parameters). Gamma $(\frac{n}{2}, \frac{1}{2})$ is denoted by χ_n^2 and as discussed earlier it has density given by*

$$\begin{aligned} f(x) &= \frac{2^{-\frac{n}{2}}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \\ &= \begin{cases} \frac{2^{-\frac{n}{2}}}{(\frac{n}{2}-1)!} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{when } n \text{ is even.} \\ \\ \frac{2^{n-\frac{n}{2}-1}(\frac{n-1}{2})!}{(n-1)!\sqrt{\pi}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{when } n \text{ is odd.} \end{cases} \end{aligned}$$

when $x > 0$.

We shall show in the next subsection that sample variance from a Normal population is a Chi-square random variable. In the next chapter we shall construct a test to make inferences about the variances of the two population. In that context we shall compare sample variances and this is where the F distribution arises naturally.

EXAMPLE 8.1.7. (F-distribution) Suppose $X_1, X_2, \ldots, X_{n_1}$ be an i.i.d. random sample from a Normal mean 0 and variance σ_1^2 population and $Y_1, Y_2, \ldots, Y_{n_2}$ be an i.i.d. random sample from a Normal mean 0 and variance σ_2^2 population. We have already seen in Example 8.1.5 that $U = \sum_{i=1}^{n_1} \left(\frac{X_i}{\sigma_1}\right)^2$ is a $\chi^2_{n_1}$ random variable and $V = \sum_{i=1}^{n_2} \left(\frac{Y_i}{\sigma_2}\right)^2$ is a $\chi^2_{n_2}$ random variable. Further U and V are independent. Let $Z = \frac{U}{n_1} / \frac{V}{n_2}$. Let $Y = \frac{n_1}{n_2} Z = \frac{U}{V}$. As done in Example 5.5.10 the density of Y for $y > 0$ is given by

$$f_Y(y) = \frac{y^{\frac{n_1}{2}-1}}{(1+y)^{\frac{n_1+n_2}{2}}} \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}$$

Therefore, for $z > 0$

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P(Y \leq \frac{n_2}{n_1} z) \\ &= \int_{-\infty}^{\frac{n_2}{n_1}z} \frac{y^{\frac{n_1}{2}-1}}{(1+y)^{\frac{n_1+n_2}{2}}} \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})} dy \end{aligned}$$

making a u-substitution with $\frac{n_1}{n_2} y = u$

$$= \int_{-\infty}^{z} \left(\frac{n_2}{n_1}\right)^{\frac{n_1}{2}} \frac{u^{\frac{n_1}{2}-1}}{(1+\frac{n_1}{n_2}u)^{\frac{n_1+n_2}{2}}} \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})} du$$

Therefore the density of Z, for $z > 0$ is given by

$$f(z) = \left(\frac{n_2}{n_1}\right)^{\frac{n_1}{2}} \frac{z^{\frac{n_1}{2}-1}}{(1+\frac{n_1}{n_2}z)^{\frac{n_1+n_2}{2}}} \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}.$$

Z is said to have $\mathrm{F}(n_1, n_2)$ distribution. Z is close to a widely used distribution in statistics called F-distribution. ■

The distribution of the ratio of sample mean and sample variance plays an important role in Hypothesis testing. This forms the motivation for the next example where the t distribution arises naturally.

EXAMPLE 8.1.8. (t-distribution) Let X_1 be a Normal random variable with mean 0 and variance 1. Let X_2 be an independent χ^2_n random variable. Let

$$Z = \frac{X_1}{\sqrt{\frac{X_2}{n}}}.$$

We wish to find the density of Z. Observe that $U = Z^2$ is given by $\frac{X_1^2}{\frac{X_2}{n}}$. Now, X_1^2 has χ^2_1 distribution (See Example 8.1.5). So applying Example 8.1.7 with $n_1 = 1$ and $n_2 = n$, we find that U has $F(1, n)$ distribution. The density of U is given by

$$\begin{aligned} f_U(u) &= \left(\frac{1}{n}\right)^{\frac{1}{2}} \frac{u^{\frac{1}{2}-1}}{(1+\frac{1}{n}u)^{\frac{n+1}{2}}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})} \\ &= \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \frac{u^{-\frac{1}{2}}}{(1+\frac{u}{n})^{\frac{n+1}{2}}}. \end{aligned}$$

Since X_1 is a symmetric random variable and $\sqrt{\frac{X_2}{n}}$ is positive valued we conclude that Z is a symmetric random variable (Exercise 8.1.10). So, for $u > 0$

$$\begin{aligned}
P(U \leq u) &= P(Z^2 \leq u) \\
&= P(-\sqrt{u} \leq Z \leq \sqrt{u}) \\
&= P(Z \leq \sqrt{u}) - P(Z \leq -\sqrt{u}) \\
&= P(Z \leq \sqrt{u}) - P(Z \geq \sqrt{u}) \\
&= 2P(Z \leq \sqrt{u}) - 1
\end{aligned}$$

Therefore if $f_Z(\cdot)$ is the density of Z then

$$f_U(u) = \frac{1}{\sqrt{u}}(f_Z(\sqrt{u})).$$

Hence for any $z \in \mathbb{R}$ the density of Z is given by

$$\begin{aligned}
f_Z(z) &= \mid z \mid f_U(z^2) \\
&= \mid z \mid \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \frac{z^{2-\frac{1}{2}}}{\left(1 + \frac{u}{n}\right)^{\frac{n+1}{2}}} \\
&= \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{z^2}{n}\right)^{-\frac{n+1}{2}}
\end{aligned}$$

Z is said to have t-distribution with n-degrees of freedom. We will denote this by the notation $Z \sim t_n$. ■

8.1.3 *Distribution of Sampling Statistics from a Normal population*

Let $n \geq 1$, $X_1, X_2, \ldots, X_n$ be an i.i.d. random sample from a population having mean μ and variance σ^2. Consider the sample mean

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

and sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

We have already seen in Theorem 7.2.2 that $E[\overline{X}] = \mu$ and in Theorem 7.2.4 that $E[S^2] = \sigma^2$. It turns out that it is not easy to understand the precise distribution of $\overline{X}$ or S^2 in general. However, this can be done when the population is normally distributed. The main result of this section is the following.

Theorem 8.1.9. *Let $n \geq 1$, $X_1, X_2, \ldots X_n$, be an i.i.d random sample with distribution $X \sim$ Normal(μ, σ^2). Let $\overline{X}$ and S^2 be as above. Then,*

(a) $\overline{X}$ is a Normal random variable with mean μ and variance $\frac{\sigma^2}{n}$.

(b) $\frac{(n-1)}{\sigma^2}S^2$ has the χ^2_{n-1} distribution.

(c) $\overline{X}$ and S^2 are independent.

Proof - (a) follows from Theorem 6.3.13. The typical proof for (b) and (c) is via Helmert's transformation (see [Rao73]) and requires some knowledge of Linear Algebra. Here we will follow Kruskal's proof as illustrated in [Stig84]. The proof is by the method of induction. For implementing the inductive step on the sample size n, we shall replace $\overline{X}$ and S^2 with $\overline{X}_n$ and S^2_n for the rest of the proof.

Step 1: (Proof for $n = 2$) Here

$$\overline{X}_2 = \frac{X_1 + X_2}{2} \quad \text{and} \quad S_2^2 = \left(X_1 - \frac{X_1 + X_2}{2}\right)^2 + \left(X_2 - \frac{X_1 + X_2}{2}\right)^2 = \frac{(X_1 - X_2)^2}{2}. \tag{8.1.6}$$

(a) Follows from Theorem 6.3.13.

(b) As X_1 and X_2 are independent Normal random variables with mean μ and variance σ^2, by Theorem 6.3.13, $\frac{(X_1 - X_2)}{\sigma\sqrt{2}}$ is a Normal random variable with mean 0 and variance 1. Using Example 8.1.5, we know that $\frac{S_2^2}{\sigma^2}$ has χ_1^2 distribution and this proves (b).

(c) From (8.1.6), $\overline{X}_2$ is a function of $X_1 + X_2$ and S_2^2 is a function of $X_1 - X_2$. Theorem 8.1.2 will imply that $\overline{X}_2$ and S_2^2 are independent if we show $X_1 + X_2$ and $X_1 - X_2$ are independent. Let $\alpha, \beta \in \mathbb{R}$. Then using Theorem 6.3.13 again we have that $\alpha(X_1 + X_2) + \beta(X_1 - X_2) = (\alpha + \beta)X_1 + (\alpha - \beta)X_2$ is normally distributed. As this is true for any $\alpha, \beta \in \mathbb{R}$, by Definition 6.4.1 $(X_1 + X_2, X_1 - X_2)$ is a bivariate normal random variable. Using Theorem 6.2.2 (f) and (g), along with the fact that X_1 and X_2 are independent Normal random variables with mean μ and variance σ^2,we have

$$Cov[X_1 + X_2, X_1 - X_2] = Var[X_1] + Cov[X_2, X_1] - Cov[X_1, X_2] - Var[X_2] = 0.$$

Theorem 6.4.3 then implies that $X_1 + X_2$ and $X_1 - X_2$ are independent.

Step 2: (inductive hypothesis) Let us inductively assume that (a),(b), and (c) are true when $n = k$ for some $k \in \mathbb{N}$.

Step 3: (Proof for $n = k + 1$) We shall rewrite $\overline{X}_{k+1}$ and S_{k+1}^2 using some elementary algebra.

$$\overline{X}_k - \overline{X}_{k+1} = \overline{X}_k - \frac{1}{k+1}\sum_{i=1}^{k+1} X_i = \left(1 - \frac{k}{k+1}\right)\overline{X}_k - \frac{1}{k+1}X_{k+1} = \frac{1}{k+1}(\overline{X}_k - X_{k+1}). \tag{8.1.7}$$

Adding and subtracting $\overline{X}_k$ inside the summand of S_{k+1}^2, we have

$$\begin{aligned}
S_{k+1}^2 &= \frac{1}{k}\sum_{i=1}^{k+1}(X_i - \overline{X}_{k+1})^2 = \frac{1}{k}\sum_{i=1}^{k+1}(X_i - \overline{X}_k + \overline{X}_k - \overline{X}_{k+1})^2 \\
&= \frac{1}{k}\sum_{i=1}^{k+1}(X_i - \overline{X}_k)^2 + 2(X_i - \overline{X}_k)(\overline{X}_k - \overline{X}_{k+1}) + (\overline{X}_k - \overline{X}_{k+1})^2 \\
&= \frac{k-1}{k}S_k^2 + \frac{1}{k}(X_{k+1} - \overline{X}_k)^2 + \frac{1}{k}\left(2(X_{k+1} - \overline{X}_k)(\overline{X}_k - \overline{X}_{k+1}) + (k+1)(\overline{X}_k - \overline{X}_{k+1})^2\right) \\
&= \frac{k-1}{k}S_k^2 + \frac{1}{k}(X_{k+1} - \overline{X}_k)^2 - \frac{1}{k}\left(2(X_{k+1} - \overline{X}_k)\frac{(X_{k+1} - \overline{X}_k)}{k+1} + \frac{(X_{k+1} - \overline{X}_k)^2}{k+1}\right) \\
&= \frac{k-1}{k}S_k^2 + \frac{1}{k+1}(X_{k+1} - \overline{X}_k)^2,
\end{aligned}$$

where we have used (8.1.7) in the second last inequality. Dividing across by σ^2 and multiplying by k we have

$$\frac{k}{\sigma^2}S_{k+1}^2 = \frac{k-1}{\sigma^2}S_k^2 + \frac{k}{\sigma^2(k+1)}(X_{k+1} - \overline{X}_k)^2. \tag{8.1.8}$$

(a) Follows from Theorem 6.3.13.

(b) To prove (b), it is enough to show that:

$$\left(\sqrt{\frac{k}{(k+1)\sigma^2}}\right)(X_{k+1} - \overline{X}_k) \text{ is a standard normal random variable and is independent of } \frac{(k-1)}{\sigma^2}S_k^2.$$

The reason being: $\frac{k}{\sigma^2(k+1)}(X_{k+1} - \overline{X}_k)^2$ then has χ_1^2 distribution by Example 8.1.5 and is independent of $\frac{(k-1)}{\sigma^2}S_k^2$ by Theorem 8.1.2; by the induction hypothesis $\frac{(k-1)}{\sigma^2}S_k^2$ has the χ_{k-1}^2 distribution; and finally using (8.1.8) along with Example 5.5.6, will imply that $\frac{k}{\sigma^2}S_{k+1}^2$ has χ_k^2 distribution.

As

$$\left(\sqrt{\frac{k}{(k+1)\sigma^2}}\right)(X_{k+1}-\overline{X}_k) = \left(\sqrt{\frac{(k+1)\sigma^2}{k}}\right)X_{k+1} - \sum_{i=1}^{k}\frac{1}{k}\left(\sqrt{\frac{k}{(k+1)\sigma^2}}\right)X_i$$

It is routine calculation using Theorem 6.3.13 to see that is a standard normal random variable.

By induction hypothesis $\overline{X}_k$ and $\frac{k-1}{\sigma^2}S_k^2$ are independent. Since $X_1,\dots,X_k,X_{k+1}$ are mutually independent, Theorem 8.1.2 implies that X_{k+1} is independent of $\overline{X}_k$ and $\frac{k-1}{\sigma^2}S_k^2$. Therefore,

$$\overline{X}_k, \quad \frac{k-1}{\sigma^2}S_k^2, \quad X_{k+1} \qquad \text{are mutually independent random variables.} \tag{8.1.9}$$

Consequently, another application of Theorem 8.1.2 will then imply that $\frac{k}{\sigma^2(k+1)}(X_{k+1}-\overline{X}_k)^2$ and $\frac{(k-1)}{\sigma^2}S_k^2$ are independent random variables.

(c) To prove (c), it is enough to show that $\overline{X}_{k+1}$ and $X_{k+1}-\overline{X}_k$ are independent. The reason is the following:

(i) Theorem 8.1.2 then implies $\overline{X}_{k+1}$ is independent of $\frac{k}{\sigma^2(k+1)}(X_{k+1}-\overline{X}_k)^2$;

(ii) $\overline{X}_{k+1}$ is a function of X_{k+1} and $\overline{X}_k$. So (8.1.9) and Theorem 8.1.2 will then imply $\overline{X}_{k+1}$ is independent of $\frac{(k-1)}{\sigma^2}S_k^2$ and also $\frac{k}{\sigma^2(k+1)}(X_{k+1}-\overline{X}_k)^2$ is independent of $\frac{(k-1)}{\sigma^2}S_k^2$;

(iii) Using (i) and (ii) we can conclude that $\overline{X}_{k+1}$, $\frac{(k-1)}{\sigma^2}S_k^2$, and $\frac{k}{\sigma^2(k+1)}(X_{k+1}-\overline{X}_k)^2$ are mutually independent; and

(iv) finally S_{k+1}^2 is a function $\frac{(k-1)}{\sigma^2}S_k^2$, and $\frac{k}{\sigma^2(k+1)}(X_{k+1}-\overline{X}_k)^2$ by (8.1.8). Then (iii) and Theorem 8.1.2 will imply that S_{k+1}^2 and $\overline{X}_{k+1}$ are independent.

Let $\alpha,\beta\in\mathbb{R}$. We have

$$\alpha(\overline{X}_{k+1}) + \beta(X_{k+1}-\overline{X}_k) = \sum_{i=1}^{k}\left(\frac{\alpha}{k+1}-\frac{\beta}{k}\right)X_i + \left(\frac{\alpha}{k+1}-\beta\right)X_{k+1}.$$

Theorem 6.3.13 will imply that $\alpha(\overline{X}_{k+1}) + \beta(X_{k+1}-\overline{X}_k)$ is is normally distributed random variable for any $\alpha,\beta\in\mathbb{R}$. So by Definition 6.4.1 $(\overline{X}_{k+1}, X_{k+1}-\overline{X}_k)$ is a bivariate normal random variable. Further, from Theorem 6.2.2 (f) and (g), we have

$$\begin{aligned}
Cov[\overline{X}_{k+1}, X_{k+1}-\overline{X}_k] &= Cov[\frac{k\overline{X}_k + X_{k+1}}{k+1}, X_{k+1}-\overline{X}_k] \\
&= \frac{1}{k+1}Var[X_{k+1}] - Cov[\overline{X}_k, X_{k+1}] - \frac{k}{k+1}Var[\overline{X}_k] \\
&= \frac{1}{k+1}\sigma^2 + 0 + -\frac{k}{k+1}\frac{\sigma^2}{k} = 0,
\end{aligned}$$

where we have used (8.1.9) in the last line. From Theorem 6.4.3 we conclude that $\overline{X}_{k+1}, X_{k+1}-\overline{X}_k$ are independent. ■

The following Corollary will be used in Chapter 9

Corollary 8.1.10. *Let $n\geq 1$, $X_1, X_2, \dots X_n$, be an i.i.d random sample with distribution $X \sim Normal(\mu,\sigma^2)$. Let $\overline{X}$ and S^2 be as above. Then*

$$\frac{\sqrt{n}(\overline{X}-\mu)}{S}$$

has the t_{n-1} distribution.

Proof - From Theorem 8.1.9 it is clear that

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

is a Normal random variable with mean 0 and variance 1, and

$$\frac{(n-1)}{\sigma^2} S^2$$

is a χ^2_{n-1} random variable. Note

$$\frac{\sqrt{n}(\overline{X} - \mu)}{S} = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{1}{n-1}\frac{(n-1)S^2}{\sigma^2}}}$$

So by Example 8.1.8 we have the result. ■

EXERCISES

Ex. 8.1.1. Verify that each of $f : \mathbb{R}^3 \to \mathbb{R}$ are density functions on $\mathbb{R}^3$.

(a) $f(x_1, x_2, x_3) = \begin{cases} \frac{2}{3}(x_1 + x_2 + x_3) & \text{if } 0 < x_i < 1, i = 1, 2, 3. \\ 0 & \text{otherwise} \end{cases}$

(b) $f(x_1, x_2, x_3) = \begin{cases} \frac{1}{8}(x_1^2 + x_2^2 + x_3^2) & \text{if } 0 < x_i < 2, i = 1, 2, 3. \\ 0 & \text{otherwise} \end{cases}$

(c) $f(x_1, x_2, x_3) = \begin{cases} \frac{2}{81}x_1x_2x_3 & \text{if } 0 < x_i < 3, i = 1, 2, 3. \\ 0 & \text{otherwise} \end{cases}$

(d) $f(x_1, x_2, x_3) = \begin{cases} \frac{3}{4}(x_1x_2 + x_1x_3 + x_2x_3) & \text{if } 0 < x_i < 1, i = 1, 2, 3. \\ 0 & \text{otherwise} \end{cases}$

Ex. 8.1.2. Suppose (X_1, X_2, X_3) have a joint density $f : \mathbb{R}^3 \to \mathbb{R}$ given by

$$f(x_1, x_2, x_3) = \begin{cases} \frac{4}{3}(x_1^3 + x_2^3 + x_3^3) & \text{if } 0 < x_i < 1, i = 1, 2, 3. \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $P(X_1 < \frac{1}{2}, X_3 > \frac{1}{2})$.

(b) Find the joint density of (X_1, X_2),(X_1, X_3), (X_2, X_3).

(c) Find the marginal densities of X_1, X_2, and X_3.

Ex. 8.1.3. Let D be a set in $\mathbb{R}^3$ with a well defined volume. (X_1, X_2, X_3) are said be uniform on a set D if they have a joint density given by

$$f(x_1, x_2, x_3) = \begin{cases} \frac{1}{\text{Volume}(D)} & \text{if } x \in D \\ 0 & \text{otherwise.} \end{cases}$$

Suppose D is a cube of dimension R.

(a) Find the joint density (X_1, X_2, X_3) which is uniform on D.

(b) Find the marginal density of X_1, X_2, X_3.

(c) Find the joint density of (X_1, X_2),(X_1, X_3),(X_3, X_2).

Ex. 8.1.4. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables having a common distribution function $F : \mathbb{R} \to [0,1]$ and probability density function $f : \mathbb{R} \to \mathbb{R}$. Let $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ be the corresponding order statistic. Show that for $1 \le i < j \le n$, $(X_{(i)}, X_{(j)})$ has a joint density function given by

$$f_{X_{(i)},X_{(j)}}(x,y) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f(x)f(y)[F(x)]^{i-1}[F(y)-F(x)]^{j-1-i}[1-F(y)]^{n-j},$$

for $-\infty < x < y < \infty$.

Ex. 8.1.5. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables having a common distribution $X \sim$ Uniform $(0,1)$. Let $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ be the corresponding order statistic. Show that $\frac{X_{(1)}}{X_{(n)}}$ and $X_{(n)}$ are independent random variables.

Ex. 8.1.6. Let $\{U_i : i \ge 1\}$ be a sequence of i.i.d. uniform $(0,1)$ random variables and Let $N \sim$ Poisson (λ). Find the distribution of $V = \min\{U_1, U_2, \ldots, U_{N+1}\}$.

Ex. 8.1.7. Let $-\infty < a < b < \infty$. Let $X_1, X_2, \ldots, X_n$ i.i.d $X \sim$ Uniform (a,b). Find the probability density function of $M = \frac{X_{(1)}+X_{(n)}}{2}$.

Ex. 8.1.8. Let X_1, X_2 be two independent standard normal random variables. Find the distribution of $Z = X_{(1)}^2$.

Ex. 8.1.9. Let $X_1, X_2, \ldots, X_n$ be i.i.d. Uniform $(0,1)$ random variables.

(a) Find the conditional distribution of $X_{(n)} \mid X_{(1)} = x$ for some $0 < x < 1$.

(b) Find $E[X_{(n)} \mid X_{(1)} = x]$ and $Var[X_{(n)} \mid X_{(1)} = x]$.

Ex. 8.1.10. Suppose X is a symmetric continuous random variable. Let Y be a continuous random variable such that $P(Y > 0) = 1$. Show that $\frac{X}{Y}$ is symmetric.

Ex. 8.1.11. Verify (8.1.3).

Ex. 8.1.12. Suppose $X_1, X_2, \ldots$ are i.i.d. Cauchy $(0,1)$ random variables.

(a) Fix $z \in \mathbb{R}$. Find a, b, c, d such that

$$\frac{1}{1+x^2}\frac{1}{1+(z-x)^2} = \frac{ax+b}{1+x^2} + \frac{cx+d}{1+(z-x)^2},$$

for all $x \in \mathbb{R}$.

(b) Show that $X_1 + X_2 \sim$ Cauchy $(0,2)$.

(c) Use induction to show that $X_1 + X_2 + \ldots + X_n \sim$ Cauchy $(0,n)$.

(d) Use Lemma 5.3.2 to show that $\overline{X}_n \sim$ Cauchy $(0,1)$.

8.2 WEAK LAW OF LARGE NUMBERS

Let $n \ge 1$, $X_1, X_2, \ldots, X_n$ be an i.i.d. random sample from a population whose distribution is given by a random variable X which has mean μ. In Chapter 7 we considered the sample mean

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

and showed in Theorem 7.2.2 that $E[\overline{X}] = \mu$. We also discussed that $\overline{X}$ could be considered as an estimate for μ. The below result makes this precise and is referred to as the weak law of large numbers.

In the statement and proof of the below Theorem we shall denote $\overline{X}$ by $\overline{X}_n$ to emphasise the dependence on n.

Theorem 8.2.1. **(Weak Law of Large Numbers)** *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables. Assume that X_1 has finite mean μ and finite variance σ^2. Then for any $\epsilon > 0$*

$$\lim_{n\to\infty} P(\mid \overline{X}_n - \mu \mid > \epsilon) = 0, \tag{8.2.1}$$

Proof- Let $\epsilon > 0$ be given. We note that

$$E(\overline{X}_n) = E(\frac{\sum_{i=1}^n X_i}{n}) = \sum_{i=1}^n \frac{1}{n} E(X_i) = \frac{n\mu}{n} = \mu.$$

Using Theorem 4.2.4, Theorem 4.2.6 and Exercise 6.2.17 we have

$$\begin{aligned} \text{Var}[\overline{X}_n] &= \text{Var}[\frac{\sum_{i=1}^n X_i}{n}] \\ &= \frac{1}{n^2}\text{Var}[\sum_{i=1}^n X_i] \\ &= \frac{1}{n^2}\sum_{i=1}^n \text{Var}[X_i] \\ &= \frac{\sigma^2}{n} \end{aligned}$$

So we have shown that the random variable $\overline{X}_n$ has finite expectation variance. By Chebychev's inequality,(apply Theorem 6.1.13 (a) with $k = \frac{\epsilon}{\sigma}$), we have

$$P(|\overline{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Therefore as $0 \leq P(|\overline{X}_n - \mu| > \epsilon)$ for all $n \geq 1$ and $\frac{\sigma^2}{n\epsilon^2} \to 0$ as $n \to \infty$, by standard results in Real Analysis we conclude that

$$\lim_{n\to\infty} P(| \overline{X}_n - \mu |> \epsilon) = 0.$$

■

REMARK 8.2.2. The convergence of sample mean to μ actually happens with Probability one. That is, suppose we denote the event $A = \{\lim_{n\to\infty} \overline{X}_n = \mu\}$, then $P(A) = 1$. The result is referred to as the Strong Law of large numbers. We prove it in Appendix C (see Theorem C.0.1).

Theorem 8.2.1 states that, for any $\epsilon > 0$, the $P(| \overline{X}_n - \mu |> \epsilon)$, goes to zero as $n \to \infty$. This mode of convergence of the sample to the true mean is called "convergence in probability" . We define it precisely below.

DEFINITION 8.2.3. *A sequence $X_1, X_2, \ldots$ is said to converge in probability to a random variable X if for any $\epsilon > 0$*

$$\lim_{n\to\infty} P(| X_n - X |> \epsilon) = 0, \tag{8.2.2}$$

The following notation

$$X_n \xrightarrow{p} X$$

is typically used to convey that the sequence $X_1, X_2, \ldots$ converges in probability to X.

EXAMPLE 8.2.4. Let X_1, X_2, $\cdots$, X_n be i.i.d random variables that are uniformly distributed over the interval $(0, 1)$. We already know by the law of large numbers that $\overline{X}$ converges to $E(X_1) = \frac{1}{2}$ in probability. Often we are interested in other functionals of the sample and their convergence properties. We illustrate one such example below.

Consider the n-th order statistic $X_{(n)} = \max\{X_1, \cdots, X_n\}$. For any $0 < \epsilon < 1$,

$$\begin{aligned} P\big(| X_{(n)} - 1 |\geq \epsilon\big) &= P\big(X_{(n)} \leq 1-\epsilon\big) + P\big(X_{(n)} \geq 1+\epsilon\big) \\ &= P\big(X_{(n)} \leq 1-\epsilon\big) + 0 \\ &= P\big(\cap_{i=1}^n (X_i \leq 1-\epsilon)\big) \\ &= \big(1-\epsilon\big)^n. \end{aligned}$$

and for $\epsilon > 1$,

$$P\big(\mid X_{(n)} - 1 \mid \geq \epsilon\big) = P\big(X_{(n)} \leq 1-\epsilon\big) + P\big(X_{(n)} \geq 1+\epsilon\big) = 0.$$

For $0 < \epsilon < 1$,

$$\lim_{n\to\infty} \big(1-\epsilon\big)^n = 0.$$

So we have shown that $X_{(n)}$ converges in probability to 1 as $n \to \infty$. ■

Another application of the weak law of large numbers is to sample proportion discussed in Section 7.2.3.

Example 8.2.5. Suppose we are interested in an event A and want to estimate $p = P(X \in A)$. We consider a sample $X_1, X_2, \ldots, X_n$ which is i.i.d. X. We define a sequence of random variables $\{Y_n\}_{n\geq 1}$ by

$$Y_n = \begin{cases} 1 & \text{if } X_n \in A \\ 0 & \text{if } X_n \notin A \end{cases}$$

Clearly Y_n are independent (as the X_n are) and further they are identically distributed as $P(Y_n = 1) = P(X_n \in A) = p$. In particular $\{Y_n\}$ are an i.i.d. Bernoulli (p) sequence of random variables. We readily observe (as done in Chapter 7) that

$$\overline{Y}_n = \frac{\sum_{i=1}^n Y_i}{n} = \frac{\#\{X_i \in A\}}{n} = \hat{p}.$$

Hence the Weak law of large numbers (applied to the sequence Y_n) will imply that sample proportion converges to the true proportion p in probability. Consequently, as discussed earlier, this provides legitimacy to the relationship between Probability and relative frequency. ■

EXERCISES

Ex. 8.2.1. Let $X, X_1, X_2, \cdots, X_n$ be i.i.d random variables that are uniformly distributed over the interval $(0, 1)$. Consider the first order statistic $X_{(1)} = \max\{X_1, \cdots, X_n\}$. Show that $X_{(1)}$ converges to 0 in probability.

Ex. 8.2.2. Let $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. random variables with finite mean and variance. Define

$$Y_n = \frac{2}{n(n+1)} \sum_{i=1}^n iX_i.$$

Show that $Y_n \xrightarrow{p} E(X_1)$ as $n \to \infty$.

8.3 CONVERGENCE IN DISTRIBUTION

When discussing a collection of random variables it makes sense to think of them as a sequence of objects, and as with any sequence in calculus we may ask whether the sequence converges in any way. We have already seen "convergence in probability" in the previous section. Here we be interested in what is known as "convergence in distribution". . This type of convergence plays a major role in the understand the limiting distribution of the sample mean (See Central Limit Theorem, Theorem 8.4.1).

Definition 8.3.1. *A sequence $X_1, X_2, \ldots$ is said to converge in distribution to a random variable X if $F_{X_n}(x)$ converges to $F_X(x)$ at every point x for which F_X is continuous. The following notation*

$$X_n \xrightarrow{d} X$$

is typically used to convey that the sequence $X_1, X_2, \ldots$ converges in distribution to X.

Example 8.3.2. Let $X_n \sim \text{Uniform}(0, \frac{1}{n})$ so that the distribution function is

$$F_{X_n}(x) = \begin{cases} 0 & \text{if } 0 \leq x \\ nx & \text{if } 0 < x < \frac{1}{n} \\ 1 & \text{if } x \geq \frac{1}{n} \end{cases}$$

and it is then easy to see that $F_{X_n}(x)$ converges to

$$F(x) = \begin{cases} 0 & \text{if } 0 \leq x \\ 1 & \text{if } x > 0 \end{cases}$$

If X is the constant random variable for which $P(X = 0) = 1$, then X has distribution function

$$F_X(x) = \begin{cases} 0 & \text{if } 0 < x \\ 1 & \text{if } x \geq 0 \end{cases}$$

It is not true that $F_X(x) = F(x)$, but the two are equal are points where they are continuous. Therefore the sequence $X_1, X_2, \ldots$ converges in distribution to the constant random variable 0. ■

Note that this form of convergence does not generally guarantee that probabilities associated with X can be derived as limits of probabilities associated with X_n. For instance, in the example above $P(X_n = 0) = 0$ for all n while $P(X = 0) = 1$. However, with a few additional assumptions a stronger claim may be made.

Theorem 8.3.3. *Let $f_{X_1}, f_{X_2}, \ldots$ be the respective densities of continuous random variables $X_1, X_2, \ldots$. Suppose they converge in distribution to a continuous random variable X with density f_X. Then for every interval A we have $P(X_n \in A) \to P(X \in A)$.*

Proof - Since X is a continuous random variable $F_X(x)$ is the integral of a density, and thus a continuous function. Therefore convergence in distribution guarantees that $F_{X_n}(x)$ converges to $F_X(x)$ everywhere. Let $A = (a, b)$ (and note that whether or not endpoints are included does not matter since all random variables are taken to be continuous). Then

$$\begin{aligned} P(X_n \in A) &= \int_a^b f_{X_n}(x)\, dx \\ &= F_{X_n}(b) - F_{X_n}(a) \\ &\to F_X(b) - F_X(a) \\ &= \int_a^b f_X(x)\, dx = P(X \in A). \end{aligned}$$

■

The second theorem about moment generating functions that we will state, but leave unproven, is the following:

Theorem 8.3.4. *(M.G.F. Convergence Theorem) If $X_1, X_2, \ldots$ are a sequence of random variables whose moment generating functions $M_n(t)$ exist in an interval containing zero, and if $M_n(t) \to M(t)$ on that interval where $M(t)$ is the moment generating function of a random variable X, then X_n converges to X in distribution.*

To illustrate the use of this fact, consider an alternate proof of the limiting relationship between binomial and Poisson random variables (See Theorem 2.2.2).

Example 8.3.5. Let $X \sim \text{Poisson}(\lambda)$ and let $X_n \sim \text{Binomial}(n, \frac{\lambda}{n})$. Then X_n converges in distribution to X.

The moment generating function of a binomial variable was already computed in Example 6.3.7. Therefore,

$$\begin{aligned} M_{X_n}(t) &= (\frac{\lambda}{n}e^t + (1-\frac{\lambda}{n}))^n \\ &= (1+\frac{\lambda(e^t-1)}{n})^n \end{aligned}$$

Using Exercise 8.4.4, we see that

$$M_{X_n}(t) \to e^{\lambda(e^t-1)}.$$

On the other hand, the moment generating function of X is

$$\begin{aligned} M_X(t) &= E[e^{tX}] \\ &= \sum_{j=0}^{\infty} e^{tj} P(X=j) \\ &= \sum_{j=0}^{\infty} e^{tj} \frac{\lambda^j e^{-\lambda}}{j!} \\ &= e^{\lambda e^t} \cdot e^{-\lambda} \cdot \sum_{j=0}^{\infty} \frac{(\lambda e^t)^j e^{-\lambda e^t}}{j!} \\ &= e^{\lambda(e^t-1)} \end{aligned}$$

where the series equals 1 since it is simply the sum of the probabilities of a Poisson(λe^t) random variable.

Since $M_{X_n}(t) \to M_X(t)$, by the m.g.f. convergence theorem (Theorem 8.3.4), X_n converges in distribution to X. That is, Binomial(n,p) random variables converge in distribution to a Poisson(λ) distribution when $p = \frac{\lambda}{n}$ and $n \to \infty$. ■

EXERCISES

Ex. 8.3.1. Suppose a sequence X_n, $n \geq 1$ of random variables converges to a random variable X in probability then show that X_n converges in distribution to X. That is show that

$$F_{X_n}(x) \to F_X(x) \text{as } n \to \infty,$$

for all continuity points of $F_X : \mathbb{R} \to [0,1]$ with F_{X_n}, F_X being the distribution functions of X_n and X respectively.

Ex. 8.3.2. Let X_n have the t-distribution with n degrees of freedom. Show that $X_n \xrightarrow{d} X$ where X is standard Normal distribution.

Ex. 8.3.3. Let $X_n \xrightarrow{d} X$. Show that $X_n^2 \xrightarrow{d} X^2$.

8.4 CENTRAL LIMIT THEOREM

Let $n \geq 1$, $X_1, X_2, \ldots, X_n$ be an i.i.d. random sample from a population with mean μ and variance σ^2. Consider the sample mean

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

As observed in Theorem 7.2.2, $E(\overline{X}) = \mu$ and SD$(\overline{X}) = \frac{\sigma}{\sqrt{n}}$. As discussed before, we might view this information as $\overline{X}$ being typically close to μ up to an error of $\frac{\sigma}{\sqrt{n}}$ with high probability. As $n \to \infty$, $\frac{\sigma}{\sqrt{n}} \to 0$ and this indicates that $\overline{X}$ approaches μ. We have already verified that $\overline{X}$ converges in probability to μ

courtesy of the weak law of large numbers (in fact it converges with probability 1 by the strong law of large numbers).

To get a better understanding of the limiting distribution of $\overline{X}$ we standardise it and consider,

$$Y_n = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma}.$$

Finding the probabilities of events connected with Y_n for each n exactly may not be possible in all cases but one can find good approximate values. It turns out that for a large class of random variables the distribution of Y_n is close to that of the standard Normal random variable particularly for large n. This remarkable fact is referred to as the Central Limit Theorem and we prove it next.

As done earlier, in the statement and proof of the below Theorem we shall denote $\overline{X}$ by $\overline{X}_n$ to emphasise the dependence on n.

Theorem 8.4.1. **(Central Limit Theorem)** *Let $X_1, X_2, \ldots$ be i.i.d. random variables with finite mean μ, finite variance σ^2, and possessing common moment generating function $M_X()$. Then*

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} Z, \tag{8.4.1}$$

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

Proof- Let $Y_n = \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}$. We will verify that

$$\lim_{n\to\infty} M_{Y_n}(t) = e^{\frac{t^2}{2}}.$$

Now, using the definition of the moment generating function and some elementary algebra we have

$$\begin{aligned} M_{Y_n}(t) &= E[\exp(tY_n))] = E[\exp(t\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma})] \\ &= E[\exp(\frac{t}{\sigma}\sqrt{n}(\frac{\sum_{i=1}^n X_i}{n} - \mu))] = E[\exp(\sum_{i=1}^n \frac{t}{\sigma\sqrt{n}}(X_i - \mu))] \\ &= E[\prod_{i=1}^n \exp(\frac{t}{\sigma\sqrt{n}}(X_i-\mu))]. \end{aligned} \tag{8.4.2}$$

As $X_1, X_2, \ldots, X_n$ are independent, from Theorem 8.1.2 we can conclude that

$$\exp(\frac{t}{\sigma\sqrt{n}}(X_1-\mu)), \exp(\frac{t}{\sigma\sqrt{n}}(X_2-\mu)), \ldots, \exp(\frac{t}{\sigma\sqrt{n}}(X_n-\mu))$$

are also independent. From Exercise 7.2.2 and 7.2.3, they also have the same distribution. So from the calculation in (8.4.2) and using Exercise 6.3.4 inductively we have

$$\begin{aligned} M_{Y_n}(t) = E[\prod_{i=1}^n \exp(\frac{t}{\sigma\sqrt{n}}(X_i-\mu))] &= \prod_{i=1}^n E[\exp(\frac{t}{\sigma\sqrt{n}}(X_i-\mu))] \\ \text{(Using Theorem 6.3.9}(a)) & \\ &= \left(E[\exp(\frac{t}{\sigma\sqrt{n}}(X_1-\mu))]\right)^n. \end{aligned} \tag{8.4.3}$$

Let $U = \frac{X_1-\mu}{\sigma}$. As $E(U) = 0, E(U^2) = 1$ we have that $M_U'(0) = 0$ and $M_U''(0) = 1$. From Exercise 8.4.5, we have that for $t \in \mathbb{R}$

$$M_U(t) = 1 + \frac{t^2}{2} + g(t) \tag{8.4.4}$$

where $\lim_{s\to 0} \frac{g(s)}{s^2} = 0$. Therefore from (8.4.3) and (8.4.4) we have

$$M_{Y_n}(t)) = [M_U(t)]^n = \left[1 + \frac{t^2}{2n} + g(\frac{t}{\sqrt{n}})\right]^n = \left[1 + \frac{1}{n}(\frac{t^2}{2} + ng(\frac{t}{\sqrt{n}}))\right]^n$$

Using the fact $\frac{t^2}{2} + ng(\frac{t}{\sqrt{n}}) \to \frac{t^2}{2}$ and Exercise 8.4.4 it follows that ,

$$\lim_{n\to\infty} M_{Y_n}(t) = e^{\frac{t^2}{2}}.$$

Theorem 8.3.4 will then imply the result. ■

Remark 8.4.2. The existence of moment generating function is not essential for the Central Limit Theorem. (8.4.1) holds when $X, X_1, X_2, \ldots$ are i.i.d. random variables with finite mean μ and finite variance σ^2. The proof is more complicated in this case.

Further we shall often use an equivalent formulation of (8.4.1). By definition of $\overline{X}$ and elementary algebra we see that $Y_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$, where $S_n = \sum_{i=1}^{n} X_i$.

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} Z, \tag{8.4.5}$$

where $S_n = \sum_{i=1}^{n} X_i$.

8.4.1 *Normal Approximation and Continuity Correction*

A typical application of the central limit theorem is to find approximate value of the probability of events related to S_n or $\overline{X}$. For instance, suppose we were interested in calculating for any $a, b \in \mathbb{R}$, $P(a < S_n \le b)$ for large n. We would proceed in the following way. We know from (8.4.5) that

$$P(\frac{S_n - n\mu}{\sqrt{n}\sigma} \le x) \to P(Z \le x) \tag{8.4.6}$$

as $n \to \infty$ for all $x \in \mathbb{R}$.

$$\begin{aligned} P(a < S_n \le b) &= P(\frac{a - n\mu}{\sqrt{n}\sigma} < \frac{S_n - n\mu}{\sqrt{n}\sigma} \le \frac{b - n\mu}{\sqrt{n}\sigma}) \\ &= P(\frac{S_n - n\mu}{\sqrt{n}\sigma} \le \frac{b - n\mu}{\sqrt{n}\sigma}) - P(\frac{S_n - n\mu}{\sqrt{n}\sigma} \le \frac{a - n\mu}{\sqrt{n}\sigma}) \\ & \quad \text{from (8.4.6)for large enough } n \\ &\approx P(Z \le \frac{b - n\mu}{\sqrt{n}\sigma}) - P(Z \le \frac{a - n\mu}{\sqrt{n}\sigma}) \\ &= P(\frac{a - n\mu}{\sqrt{n}\sigma} < Z \le \frac{b - n\mu}{\sqrt{n}\sigma}), \end{aligned}$$

where in the second last line we have used the notation $\approx$ to indicate that the right hand side is an approximation. Therefore we would conclude that for large n,

$$P(a < S_n \le b) \approx P(\frac{a - n\mu}{\sqrt{n}\sigma} < Z \le \frac{b - n\mu}{\sqrt{n}\sigma}). \tag{8.4.7}$$

We would then use the R function `pnorm()` or Normal Tables (See Table D.2) to compute the right hand side.

A similar computation would also yield

$$P(a < \overline{X} \le b) \approx P(\frac{\sqrt{n}(a - \mu)}{\sigma} < Z \le \frac{\sqrt{n}(b - \mu)}{\sigma}). \tag{8.4.8}$$

EXAMPLE 8.4.3. Let Y be a random variable distributed as Gamma$(100, 4)$. Suppose we were interested in finding

$$P(20 < Y \leq 30).$$

Suppose $X_1, X_2, \ldots, X_{100}$ are independent Exponential (4) random variables then Y and $S_{100} = \sum_{i=1}^{100} X_i$ have the same distribution. Therefore, applying the Central Limit Theorem with $\mu = E(X_1) = \frac{1}{4}, \sigma = \text{SD}(X_1) = \frac{1}{4}$, we have

$$\begin{aligned}
P(20 < Y \leq 30) &= P(20 < S_{100} \leq 30) \\
&\quad \text{by (8.4.7)} \\
&\approx P(\frac{20 - 100(0.25)}{\sqrt{100}(0.25)} < Z \leq \frac{30 - 100(0.25)}{\sqrt{100}(0.25)}) \\
&= P(\frac{-5}{2.5} < Z \leq \frac{5}{2.5}) \\
&= P(-2 < Z \leq 2) \\
&= P(Z \leq 2) - P(Z \leq -2) \\
&\quad \text{using symmetry of Normal distribution} \\
&= P(Z \leq 2) - (1 - P(Z \leq 2)) \\
&= 2P(Z \leq 2) - 1
\end{aligned}$$

Looking up Table D.2, we see that this value comes out to be approximately $2 \times 0.9772 - 1 = 0.9544$. A more precise answer is given by R as

```
> 2 * pnorm(2) - 1
[1] 0.9544997
```

Using R, we can also compare this with the exact probability that we are approximating.

```
> pgamma(30, 100, 4) - pgamma(20, 100, 4)
[1] 0.9550279
```

■

Continuity Correction: Suppose $X_1, X_2, X_3, \ldots$ are all integer valued random variables. Then $S_n = \sum_{i=1}^{n} X_i$ is also a integer random variable. Now,

$$P(S_n = k) = P(k - h < S_n \leq k + h)$$

for all natural numbers k and $0 < h < 1$. However it is easy to see that two distinct values of h will lead to two different answers if we use the Normal approximation provided by the Central Limit Theorem. One can also observe that this will increase with h. So it is customary to use $h = \frac{1}{2}$ while computing such probabilities using the Normal approximation. So when $X_1, X_2, X_3, \ldots$ are all integer valued random variables we use,

$$\begin{aligned}
P(a < S_n \leq b) &= P(a - 0.5 < S_n \leq b + 0.5) \\
&\approx P(\frac{a + 0.5 - n\mu}{\sqrt{n}\sigma} < Z \leq \frac{b + 0.5 - n\mu}{\sqrt{n}\sigma})
\end{aligned} \tag{8.4.9}$$

whenever a, b are in the range of S_n. This convention is referred to as the "continuity correction".

EXAMPLE 8.4.4. Two types of coin are produced at a factory: a fair coin and a biased one that comes up heads 55% of the time. Priya is the quality control scientist at the factory. She wants to design an experiment that will test whether a coin is fair or biased. In order to ascertain which type of coin she has, she prescribes the following experiment as a test:- *Toss the given coin 1000 times, if the coin comes up heads 525 or more times conclude that it is a biased coin. Otherwise conclude that it is fair.* Factory manager Ayesha is interested in the following question: What is the probability that Priya's test shall reach a false conclusion for a fair coin ?

Let S_{1000} be the number of heads in 1000 tosses of a coin. As discussed in earlier chapters, we know that $S_{1000} = \sum_{i=1}^{1000} X_i$ where each X_i are i.i.d. Bernoulli random variables with parameter p.

If the coin is fair, then $p = 0.5$ and $E[X_1] = 0.5, Var[X_1] = 0.25$, and therefore $E[S_{1000}] = 500$ and $SD[S_{1000}] = \sqrt{250} = 15.8114$. We want to approximate

$$P(S_{1000} \geq 525) = 1 - P(S_{1000} \leq 524) = 1 - P(S_{1000} \leq 524.5)$$

Without the continuity correction, we would approximate this probability as

$$1 - P\left(Z \leq \frac{24}{15.8114}\right) = 1 - P\left(Z \leq 1.52\right)$$

which can be computed using Table D.2 as $1 - 0.9357 = 0.0643$, or using R as

```
> 1 - pnorm(24 / sqrt(250))
[1] 0.06452065
```

With the continuity correction, the approximate value would instead use $z = 24.5/15.8114 = 1.55$, giving $1 - 0.9394 = 0.0606$ using Table D.2 or

```
> 1 - pnorm(24.5 / sqrt(250))
[1] 0.06062886
```

in R. We can also compute the exact probability that we are trying to approximate, namely $P(S_{1000} \geq 525)$, in R as

```
> 1 - pbinom(524, 1000, 0.5)
[1] 0.06060713
```

As we can see, the continuity correction gives us a slightly better approximation. These calculations tell us that the probability of Priya's test reaching a false conclusion if the coin is fair is approximately 0.061. We shall examine the topic of Hypothesis testing, that Priya was trying to do, more in Chapter 9. ■

Example 8.4.5. We return to the Birthday problem. Suppose a small town has 1095 students. What is the probability that five or more students were born on independence day ? Assume that birthrates are constant throughout the year and that each year has 365 days.

The probability that any given student was born on independence day is $\frac{1}{365}$. So the exact probability that five or more students were born on independence day is

$$1 - \sum_{k=0}^{4} \binom{1095}{k} \left(\frac{1}{365}\right)^k \left(\frac{364}{365}\right)^{1095-k}.$$

In Example 2.2.1 we have used the Poisson approximation with $\lambda = 4$ to estimate the above as

$$\begin{aligned} & 1 - \sum_{k=0}^{4} \binom{1460}{k} \left(\frac{1}{365}\right)^k \left(\frac{364}{365}\right)^{1460-k} \\ \approx & \quad 1 - \left[e^{-4} + 4e^{-4} + \frac{4^2}{2}e^{-4} + \frac{1}{6}4^3e^{-4} + \frac{1}{24}4^4e^{-4}\right] \\ = & \quad 0.3711631 \end{aligned}$$

We can do another approximation using central limit theorem, which is typically called the normal approximation. For $1 \leq i \leq 1460$, define

$$X_i = \begin{cases} 1 & \text{if } i\text{th person's birthday is on independence day} \\ 0 & \text{otherwise} \end{cases}$$

Given the assumptions above on birthrates we know X_i are i.i.d random variables distributed as Bernoulli($\frac{1}{365}$). Note that $S_{1460} = \sum_{i=1}^{1460} X_i$ is the number of people born on independence day and we are interested in calculating

$$P(S_{1460} \geq 5).$$

Observe that $E(X_1) = \frac{1}{365}$, $\text{Var}(X_1) = \frac{1}{365}(1 - \frac{1}{365}) = \frac{364}{365^2}$. By the central limit theorem, we know that

$$\begin{aligned} P(S_{1460} \geq 5) &= 1 - P(S_{1460} \leq 4) = 1 - P(S_{1460} \leq 4.5) \\ &\approx 1 - P(Z \leq \frac{4.5 - (1460)(\frac{1}{365})}{\sqrt{(1460)(\frac{364}{365^2})}}) \\ &= 1 - P(Z \leq \frac{0.5}{1.9973}) \\ &= 0.401 \end{aligned}$$

Recall from the calculations done in Example 2.2.1 that the exact answer for this problem is 0.3711629. So in this example, the Poisson approximation seems to work better then the Normal approximation. This is due to the fact that more asymmetry in the underlying Bernoulli distribution worsens the normal approximation, just as it improves the Poisson approximation as we saw in Figure 2.2. ■

EXERCISES

Ex. 8.4.1. Suppose S_n is binomially distributed with parameters $n = 200$ and $p = 0.3$ Use the central limit theorem to find an approximation for $P(99 \leq S_n \leq 101)$.

Ex. 8.4.2. Toss a fair coin 400 times. Use the central limit theorem to

(a) find an approximation for the probability of at most 190 heads.

(b) find an approximation for the probability of at least 70 heads.

(c) find an approximation for the probability of at least 120 heads.

(d) find an approximation for the probability that the number of heads is between 140 and least 160.

Ex. 8.4.3. Suppose that the weight of open packets of daal in a home is uniformly distributed from 200 to 600 gms. In random survey of 64 homes, find the (approximate) probability that the total weight of open boxes is less than 25 kgs.

Ex. 8.4.4. Let $\{a_n\}_{n\geq 1}$ be a sequence of real numbers such that $a_n \to a$ as $n \to \infty$. Then

$$\lim_{n\to\infty} \left(1 + \frac{a_n}{n}\right)^n = e^a.$$

Ex. 8.4.5. Suppose U is a random variable (discrete or continuous) and $M_U(t) = E(e^{tU})$ exists for all t. Then show that

$$M_U(t) = 1 + tM_U'(0) + \frac{t^2}{2}M_U''(0) + g(t)$$

where $\lim_{t\to 0} \frac{g(t)}{t^2} = 0$.

Ex. 8.4.6. Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with $X_1 \sim$, Exponential (1). Find

$$\lim_{n\to\infty} P\left(\frac{n}{2} - \frac{\sqrt{n}}{2\sqrt{3}} \leq \sum_{i=1}^{n}[1 - \exp(-X_i)] \leq \frac{n}{2} + \frac{\sqrt{n}}{2\sqrt{3}}\right).$$

Ex. 8.4.7. Let $a_n = \sum_{k=0}^{n} \frac{n^k}{k!}e^{-n}$, $n \geq 1$. Using the Central Limit Theorem evaluate $\lim_{n\to\infty} a_n$.

Ex. 8.4.8. How often should you toss a coin:

(a) to be at least 90 % sure that your estimate of the P(head) is within 0.1 of its true value ?

(b) to be at least 90 % sure that your estimate of the P(head) is within 0.01 of its true value ?

Ex. 8.4.9. To forecast the outcome of the election in which two parties are contesting, an internet poll via Facebook is conducted. How many people should be surveyed to be at least 95% sure that the estimated proportion is within 0.05 of the true value ?

Ex. 8.4.10. A medical study is conducted to estimate the proportion of people suffering from April allergies in Bangalore. How many people should be surveyed to be at least 99% sure that the estimate is within 0.02 of the true value ?