# Chapter 13

# SEQUENCES OF RANDOM VARIABLES AND ORDER STASTISTICS

In this chapter, we generalize some of the results we have studied in the previous chapters. We do these generalizations because the generalizations are needed in the subsequent chapters relating mathematical statistics. In this chapter, we also examine the weak law of large numbers, Bernoulli's law of large numbers, the strong law of large numbers, and the central limit theorem. Further, in this chapter, we treat the order statistics and percentiles.

## 13.1. Distribution of sample mean and variance

Consider a random experiment. Let $X$ be the random variable associated with this experiment. Let $f(x)$ be the probability density function of $X$. Let us repeat this experiment $n$ times. Let $X_k$ be the random variable associated with the $k^{\text{th}}$ repetition. Then the collection of the random variables $\{ X_1, X_2, ..., X_n \}$ is a random sample of size $n$. From here after, we simply denote $X_1, X_2, ..., X_n$ as a random sample of size $n$. The random variables $X_1, X_2, ..., X_n$ are independent and identically distributed with the common probability density function $f(x)$.

For a random sample, functions such as the sample mean $\overline{X}$, the sample variance $S^2$ are called *statistics*. In a particular sample, say $x_1, x_2, ..., x_n$, we observed $\overline{x}$ and $s^2$. We may consider

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

and

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

as random variables and $\overline{x}$ and $s^2$ are the realizations from a particular sample.

In this section, we are mainly interested in finding the probability distributions of the sample mean $\overline{X}$ and sample variance $S^2$, that is the distribution of the statistics of samples.

**Example 13.1.** Let $X_1$ and $X_2$ be a random sample of size 2 from a distribution with probability density function

$$f(x) = \begin{cases} 6x(1-x) & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

What are the mean and variance of sample sum $Y = X_1 + X_2$?

**Answer:** The population mean

$$\begin{aligned} \mu_X &= E\,(X) \\ &= \int_0^1 x\, 6x(1-x)\, dx \\ &= 6 \int_0^1 x^2\,(1-x)\, dx \\ &= 6\, B(3,2) \qquad \text{(here } B \text{ denotes the beta function)} \\ &= 6\, \frac{\Gamma(3)\,\Gamma(2)}{\Gamma(5)} \\ &= 6 \left(\frac{1}{12}\right) \\ &= \frac{1}{2}. \end{aligned}$$

Since $X_1$ and $X_2$ have the same distribution, we obtain $\mu_{X_1} = \frac{1}{2} = \mu_{X_2}$. Hence the mean of $Y$ is given by

$$\begin{aligned} E(Y) &= E(X_1 + X_2) \\ &= E(X_1) + E(X_2) \\ &= \frac{1}{2} + \frac{1}{2} \\ &= 1. \end{aligned}$$

Next, we compute the variance of the population $X$. The variance of $X$ is given by

$$
\begin{aligned}
Var(X) &= E\left(X^2\right) - E(X)^2 \\
&= \int_0^1 6x^3(1-x)\,dx - \left(\frac{1}{2}\right)^2 \\
&= 6\int_0^1 x^3\,(1-x)\,dx - \left(\frac{1}{4}\right) \\
&= 6\,B(4,2) - \left(\frac{1}{4}\right) \\
&= 6\,\frac{\Gamma(4)\,\Gamma(2)}{\Gamma(6)} - \left(\frac{1}{4}\right) \\
&= 6\,\left(\frac{1}{20}\right) - \left(\frac{1}{4}\right) \\
&= \frac{6}{20} - \frac{5}{20} \\
&= \frac{1}{20}.
\end{aligned}
$$

Since $X_1$ and $X_2$ have the same distribution as the population $X$, we get

$$Var(X_1) = \frac{1}{20} = Var(X_2).$$

Hence, the variance of the sample sum $Y$ is given by

$$
\begin{aligned}
Var(Y) &= Var\,(X_1 + X_2) \\
&= Var\,(X_1) + Var\,(X_2) + 2\,Cov\,(X_1, X_2) \\
&= Var\,(X_1) + Var\,(X_2) \\
&= \frac{1}{20} + \frac{1}{20} \\
&= \frac{1}{10}.
\end{aligned}
$$

**Example 13.2.** Let $X_1$ and $X_2$ be a random sample of size 2 from a distribution with density

$$f(x) = \begin{cases} \frac{1}{4} & \text{for } x = 1, 2, 3, 4 \\ \\ 0 & \text{otherwise.} \end{cases}$$

What is the distribution of the sample sum $Y = X_1 + X_2$ ?

**Answer:** Since the range space of $X_1$ as well as $X_2$ is $\{1, 2, 3, 4\}$, the range space of $Y = X_1 + X_2$ is

$$R_Y = \{2, 3, 4, 5, 6, 7, 8\}.$$

Let $g(y)$ be the density function of $Y$. We want to find this density function. First, we find $g(2)$, $g(3)$ and so on.

$$\begin{aligned} g(2) &= P(Y = 2) \\ &= P(X_1 + X_2 = 2) \\ &= P(X_1 = 1 \text{ and } X_2 = 1) \\ &= P(X_1 = 1)\, P(X_2 = 1) \qquad \text{(by independence of } X_1 \text{ and } X_2) \\ &= f(1)\, f(1) \\ &= \left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = \frac{1}{16}. \end{aligned}$$

$$\begin{aligned} g(3) &= P(Y = 3) \\ &= P(X_1 + X_2 = 3) \\ &= P(X_1 = 1 \text{ and } X_2 = 2) + P(X_1 = 2 \text{ and } X_2 = 1) \\ &= P(X_1 = 1)\, P(X_2 = 2) \\ &\quad + P(X_1 = 2)\, P(X_2 = 1) \qquad \text{(by independence of } X_1 \text{ and } X_2) \\ &= f(1)\, f(2) + f(2)\, f(1) \\ &= \left(\frac{1}{4}\right)\left(\frac{1}{4}\right) + \left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = \frac{2}{16}. \end{aligned}$$

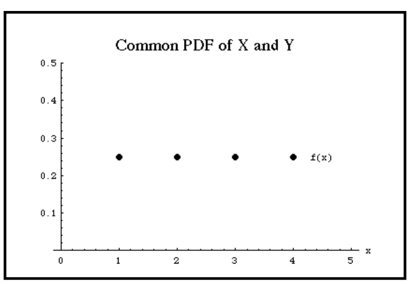
$$
\begin{aligned}
g(4) &= P\left(Y=4\right) \\
&= P\left(X_1+X_2=4\right) \\
&= P\left(X_1=1 \text{ and } X_2=3\right)+P\left(X_1=3 \text{ and } X_2=1\right) \\
&\quad + P\left(X_1=2 \text{ and } X_2=2\right) \\
&= P\left(X_1=3\right)\, P\left(X_2=1\right)+P\left(X_1=1\right)\, P\left(X_2=3\right) \\
&\quad + P\left(X_1=2\right)\, P\left(X_2=2\right) \qquad \text{(by independence of } X_1 \text{ and } X_2\text{)} \\
&= f(1)\, f(3)+f(3)\, f(1)+f(2)\, f(2) \\
&= \left(\frac{1}{4}\right)\, \left(\frac{1}{4}\right)+\left(\frac{1}{4}\right)\, \left(\frac{1}{4}\right)+\left(\frac{1}{4}\right)\, \left(\frac{1}{4}\right) \\
&= \frac{3}{16}.
\end{aligned}
$$

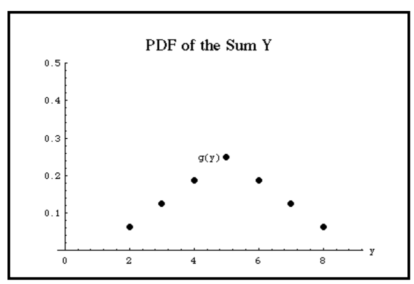
Similarly, we get

$$
g(5)=\frac{4}{16}, \qquad g(6)=\frac{3}{16}, \qquad g(7)=\frac{2}{16}, \qquad g(8)=\frac{1}{16}.
$$

Thus, putting these into one expression, we get

$$
\begin{aligned}
g(y) &= P(Y=y) \\
&= \sum_{k=1}^{y-1} f(k)\, f(y-k) \\
&= \frac{4-|y-5|}{16}, \qquad y=2,3,4,...,8.
\end{aligned}
$$




**Remark 13.1.** Note that $g(y)=\sum_{k=1}^{y-1} f(k)\, f(y-k)$ is the discrete convolution of $f$ with itself. The concept of convolution was introduced in chapter 10.

The above example can also be done using the moment generating function method as follows:

$$\begin{aligned}
M_Y(t) &= M_{X_1+X_2}(t) \\
&= M_{X_1}(t)\, M_{X_2}(t) \\
&= \left(\frac{e^t + e^{2t} + e^{3t} + e^{4t}}{4}\right) \left(\frac{e^t + e^{2t} + e^{3t} + e^{4t}}{4}\right) \\
&= \left(\frac{e^t + e^{2t} + e^{3t} + e^{4t}}{4}\right)^2 \\
&= \frac{e^{2t} + 2e^{3t} + 3e^{4t} + 4e^{5t} + 3e^{6t} + 2e^{7t} + e^{8t}}{16}.
\end{aligned}$$

Hence, the density of $Y$ is given by

$$g(y) = \frac{4 - |y-5|}{16}, \qquad y = 2, 3, 4, ..., 8.$$

**Theorem 13.1.** If $X_1, X_2, ..., X_n$ are mutually independent random variables with densities $f_1(x_1)$, $f_2(x_2), ..., f_n(x_n)$ and $E[u_i(X_i)]$, $i = 1, 2, ..., n$ exist, then

$$E\left[\prod_{i=1}^n u_i(X_i)\right] = \prod_{i=1}^n E[u_i(X_i)],$$

where $u_i$ $(i = 1, 2, ..., n)$ are arbitrary functions.

**Proof:** We prove the theorem assuming that the random variables $X_1, X_2, ..., X_n$ are continuous. If the random variables are not continuous, then the proof follows exactly in the same manner if one replaces the integrals by summations. Since

$$\begin{aligned}
&E\left(\prod_{i=1}^n u_i(X_i)\right) \\
&= E(u_1(X_1) \cdots u_n(X_n)) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u_1(x_1) \cdots u_n(x_n) f(x_1, ..., x_n) dx_1 \cdots dx_n \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u_1(x_1) \cdots u_n(x_n) f_1(x_1) \cdots f_n(x_n) dx_1 \cdots dx_n \\
&= \int_{-\infty}^{\infty} u_1(x_1) f_1(x_1) dx_1 \cdots \int_{-\infty}^{\infty} u_n(x_n) f_n(x_n) dx_n \\
&= E\left(u_1(X_1)\right) \cdots E\left(u_n(X_n)\right) \\
&= \prod_{i=1}^n E\left(u_i(X_i)\right),
\end{aligned}$$

the proof of the theorem is now complete.

**Example 13.3.** Let $X$ and $Y$ be two random variables with the joint density

$$f(x,y) = \begin{cases} e^{-(x+y)} & \text{for } 0 < x,\, y < \infty \\ \\ 0 & \text{otherwise.} \end{cases}$$

What is the expected value of the continuous random variable $Z = X^2Y^2 + XY^2 + X^2 + X$?

**Answer:** Since

$$\begin{aligned} f(x,y) &= e^{-(x+y)} \\ &= e^{-x}\, e^{-y} \\ &= f_1(x)\, f_2(y), \end{aligned}$$

the random variables $X$ and $Y$ are mutually independent. Hence, the expected value of $X$ is

$$\begin{aligned} E(X) &= \int_0^\infty x\, f_1(x)\, dx \\ &= \int_0^\infty x e^{-x}\, dx \\ &= \Gamma(2) \\ &= 1. \end{aligned}$$

Similarly, the expected value of $X^2$ is given by

$$\begin{aligned} E\left(X^2\right) &= \int_0^\infty x^2\, f_1(x)\, dx \\ &= \int_0^\infty x^2 e^{-x}\, dx \\ &= \Gamma(3) \\ &= 2. \end{aligned}$$

Since the marginals of $X$ and $Y$ are same, we also get $E(Y) = 1$ and $E(Y^2) = 2$. Further, by Theorem 13.1, we get

$$\begin{aligned} E\left[Z\right] &= E\left[X^2Y^2 + XY^2 + X^2 + X\right] \\ &= E\left[\left(X^2 + X\right)\, \left(Y^2 + 1\right)\right] \\ &= E\left[X^2 + X\right]\, E\left[Y^2 + 1\right] \qquad \text{(by Theorem 13.1)} \\ &= \left(E\left[X^2\right] + E\left[X\right]\right)\, \left(E\left[Y^2\right] + 1\right) \\ &= (2+1)\, (2+1) \\ &= 9. \end{aligned}$$

**Theorem 13.2.** If $X_1, X_2, ..., X_n$ are mutually independent random variables with respective means $\mu_1,\, \mu_2,\, ...,\mu_n$ and variances $\sigma_1^2,\, \sigma_2^2,\, ...,\sigma_n^2$, then the mean and variance of $Y = \sum_{i=1}^n a_i\, X_i$, where $a_1,\, a_2,\, ..., a_n$ are real constants, are given by

$$\mu_Y = \sum_{i=1}^n a_i\, \mu_i \qquad \text{and} \qquad \sigma_Y^2 = \sum_{i=1}^n a_i^2\, \sigma_i^2.$$

**Proof:** First we show that $\mu_Y = \sum_{i=1}^n a_i\, \mu_i$. Since

$$\begin{aligned} \mu_Y &= E(Y) \\ &= E\left(\sum_{i=1}^n a_i\, X_i\right) \\ &= \sum_{i=1}^n a_i E(X_i) \\ &= \sum_{i=1}^n a_i \mu_i \end{aligned}$$

we have asserted result. Next we show $\sigma_Y^2 = \sum_{i=1}^n a_i^2\, \sigma_i^2$. Since $Cov(X_i,\, X_j) = 0$ for $i \neq j$, we have

$$\begin{aligned} \sigma_Y^2 &= Var(Y) \\ &= Var\,(a_i\, X_i) \\ &= \sum_{i=1}^n a_i^2 Var\,(X_i) \\ &= \sum_{i=1}^n a_i^2\, \sigma_i^2. \end{aligned}$$

This completes the proof of the theorem.

**Example 13.4.** Let the independent random variables $X_1$ and $X_2$ have means $\mu_1 = -4$ and $\mu_2 = 3$, respectively and variances $\sigma_1^2 = 4$ and $\sigma_2^2 = 9$. What are the mean and variance of $Y = 3X_1 - 2X_2$?

**Answer:** The mean of $Y$ is

$$\begin{aligned} \mu_Y &= 3\mu_1 - 2\mu_2 \\ &= 3(-4) - 2(3) \\ &= -18. \end{aligned}$$

Similarly, the variance of $Y$ is

$$\begin{aligned}
\sigma_Y^2 &= (3)^2\, \sigma_1^2 + (-2)^2\, \sigma_2^2 \\
&= 9\, \sigma_1^2 + 4\, \sigma_2^2 \\
&= 9(4) + 4(9) \\
&= 72.
\end{aligned}$$

**Example 13.5.** Let $X_1, X_2, ..., X_{50}$ be a random sample of size 50 from a distribution with density

$$f(x) = \begin{cases} \frac{1}{\theta}\, e^{-\frac{x}{\theta}} & \text{for } 0 \leq x < \infty \\ 0 & \text{otherwise.} \end{cases}$$

What are the mean and variance of the sample mean $\overline{X}$?

**Answer:** Since the distribution of the population $X$ is exponential, the mean and variance of $X$ are given by

$$\mu_X = \theta, \qquad \text{and} \qquad \sigma_X^2 = \theta^2.$$

Thus, the mean of the sample mean is

$$\begin{aligned}
E\left(\overline{X}\right) &= E\left(\frac{X_1 + X_2 + \cdots + X_{50}}{50}\right) \\
&= \frac{1}{50} \sum_{i=1}^{50} E\left(X_i\right) \\
&= \frac{1}{50} \sum_{i=1}^{50} \theta \\
&= \frac{1}{50}\; 50\, \theta = \theta.
\end{aligned}$$

The variance of the sample mean is given by

$$\begin{aligned}
Var\left(\overline{X}\right) &= Var\left(\sum_{i=1}^{50} \frac{1}{50}\, X_i\right) \\
&= \sum_{i=1}^{50} \left(\frac{1}{50}\right)^2\; \sigma_{X_i}^2 \\
&= \sum_{i=1}^{50} \left(\frac{1}{50}\right)^2\; \theta^2 \\
&= 50\; \left(\frac{1}{50}\right)^2\; \theta^2 \\
&= \frac{\theta^2}{50}.
\end{aligned}$$

**Theorem 13.3.** If $X_1, X_2, ..., X_n$ are independent random variables with respective moment generating functions $M_{X_i}(t)$, $i = 1, 2, ..., n$, then the moment generating function of $Y = \sum_{i=1}^{n} a_i X_i$ is given by

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}\left(a_i t\right).$$

**Proof:** Since

$$\begin{aligned} M_Y(t) &= M_{\sum_{i=1}^{n} a_i X_i}(t) \\ &= \prod_{i=1}^{n} M_{a_i X_i}(t) \\ &= \prod_{i=1}^{n} M_{X_i}(a_i t) \end{aligned}$$

we have the asserted result and the proof of the theorem is now complete.

**Example 13.6.** Let $X_1, X_2, ..., X_{10}$ be the observations from a random sample of size 10 from a distribution with density

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}, \qquad -\infty < x < \infty.$$

What is the moment generating function of the sample mean?

**Answer:** The density of the population $X$ is a standard normal. Hence, the moment generating function of each $X_i$ is

$$M_{X_i}(t) = e^{\frac{1}{2}t^2}, \qquad i = 1, 2, ..., 10.$$

The moment generating function of the sample mean is

$$\begin{aligned} M_{\overline{X}}(t) &= M_{\sum_{i=1}^{10} \frac{1}{10} X_i}(t) \\ &= \prod_{i=1}^{10} M_{X_i}\left(\frac{1}{10}\, t\right) \\ &= \prod_{i=1}^{10} e^{\frac{t^2}{200}} \\ &= \left[\, e^{\frac{t^2}{200}} \,\right]^{10} = e^{\left(\frac{1}{10}\, \frac{t^2}{2}\right)}. \end{aligned}$$

Hence $\overline{X} \sim N\left(0,\, \frac{1}{10}\right)$.

The last example tells us that if we take a sample of any size from a standard normal population, then the sample mean also has a normal distribution.

The following theorem says that a linear combination of random variables with normal distributions is again normal.

**Theorem 13.4.** If $X_1, X_2, ..., X_n$ are mutually independent random variables such that

$$X_i \sim N\left(\mu_i,\, \sigma_i^2\right), \qquad i = 1, 2, ..., n.$$

Then the random variable $Y = \displaystyle\sum_{i=1}^{n} a_i X_i$ is a normal random variable with mean

$$\mu_Y = \sum_{i=1}^{n} a_i\, \mu_i \qquad \text{and} \qquad \sigma_Y^2 = \sum_{i=1}^{n} a_i^2\, \sigma_i^2,$$

that is $Y \sim N\left(\sum_{i=1}^{n} a_i \mu_i,\, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$.

**Proof:** Since each $X_i \sim N\left(\mu_i, \sigma_i^2\right)$, the moment generating function of each $X_i$ is given by

$$M_{X_i}(t) = e^{\mu_i t + \frac{1}{2}\sigma_i^2 t^2}.$$

Hence using Theorem 13.3, we have

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_{X_i}(a_i t) \\ &= \prod_{i=1}^{n} e^{a_i \mu_i t + \frac{1}{2} a_i^2 \sigma_i^2 t^2} \\ &= e^{\sum_{i=1}^{n} a_i \mu_i t + \frac{1}{2} \sum_{i=1}^{n} a_i^2 \sigma_i^2 t^2}. \end{aligned}$$

Thus the random variable $Y \sim N\left(\displaystyle\sum_{i=1}^{n} a_i \mu_i,\, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$. The proof of the theorem is now complete.

**Example 13.7.** Let $X_1, X_2, ..., X_n$ be the observations from a random sample of size $n$ from a normal distribution with mean $\mu$ and variance $\sigma^2 > 0$. What are the mean and variance of the sample mean $\overline{X}$?

**Answer:** The expected value (or mean) of the sample mean is given by

$$
\begin{aligned}
E\left(\overline{X}\right) &= \frac{1}{n} \sum_{i=1}^{n} E\left(X_i\right) \\
&= \frac{1}{n} \sum_{i=1}^{n} \mu \\
&= \mu.
\end{aligned}
$$

Similarly, the variance of the sample mean is

$$
Var\left(\overline{X}\right) = \sum_{i=1}^{n} Var\left(\frac{X_i}{n}\right) = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 \sigma^2 = \frac{\sigma^2}{n}.
$$

This example along with the previous theorem says that if we take a random sample of size $n$ from a normal population with mean $\mu$ and variance $\sigma^2$, then the sample mean is also normal with mean $\mu$ and variance $\frac{\sigma^2}{n}$, that is $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

**Example 13.8.** Let $X_1, X_2, ..., X_{64}$ be a random sample of size 64 from a normal distribution with $\mu = 50$ and $\sigma^2 = 16$. What are $P\left(49 < X_8 < 51\right)$ and $P\left(49 < \overline{X} < 51\right)$?

**Answer:** Since $X_8 \sim N(50, 16)$, we get

$$
\begin{aligned}
P\left(49 < X_8 < 51\right) &= P\left(49 - 50 < X_8 - 50 < 51 - 50\right) \\
&= P\left(\frac{49 - 50}{4} < \frac{X_8 - 50}{4} < \frac{51 - 50}{4}\right) \\
&= P\left(-\frac{1}{4} < \frac{X_8 - 50}{4} < \frac{1}{4}\right) \\
&= P\left(-\frac{1}{4} < Z < \frac{1}{4}\right) \\
&= 2P\left(Z < \frac{1}{4}\right) - 1 \\
&= 0.1974 \qquad \text{(from normal table).}
\end{aligned}
$$

By the previous theorem, we see that $\overline{X} \sim N\left(50, \frac{16}{64}\right)$. Hence

$$\begin{aligned}
P\left(49 < \overline{X} < 51\right) &= P\left(49 - 50 < \overline{X} - 50 < 51 - 50\right) \\
&= P\left(\frac{49-50}{\sqrt{\frac{16}{64}}} < \frac{\overline{X}-50}{\sqrt{\frac{16}{64}}} < \frac{51-50}{\sqrt{\frac{16}{64}}}\right) \\
&= P\left(-2 < \frac{\overline{X}-50}{\sqrt{\frac{16}{64}}} < 2\right) \\
&= P\left(-2 < Z < 2\right) \\
&= 2P\left(Z < 2\right) - 1 \\
&= 0.9544 \qquad \text{(from normal table).}
\end{aligned}$$

This example tells us that $\overline{X}$ has a greater probability of falling in an interval containing $\mu$, than a single observation, say $X_8$ (or in general any $X_i$).

**Theorem 13.5.** Let the distributions of the random variables $X_1, X_2, ..., X_n$ be $\chi^2(r_1)$, $\chi^2(r_2)$, ..., $\chi^2(r_n)$, respectively. If $X_1, X_2, ..., X_n$ are mutually independent, then $Y = X_1 + X_2 + \cdots + X_n \sim \chi^2\left(\sum_{i=1}^n r_i\right)$.

**Proof:** Since each $X_i \sim \chi^2(r_i)$, the moment generating function of each $X_i$ is given by

$$M_{X_i}(t) = (1-2t)^{-\frac{r_i}{2}}.$$

By Theorem 13.3, we have

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n (1-2t)^{-\frac{r_i}{2}} = (1-2t)^{-\frac{1}{2}\sum_{i=1}^n r_i}.$$

Hence $Y \sim \chi^2\left(\sum_{i=1}^n r_i\right)$ and the proof of the theorem is now complete.

The proof of the following theorem is an easy consequence of Theorem 13.5 and we leave the proof to the reader.

**Theorem 13.6.** If $Z_1, Z_2, ..., Z_n$ are mutually independent and each one is standard normal, then $Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi^2(n)$, that is the sum is chi-square with $n$ degrees of freedom.

The following theorem is very useful in mathematical statistics and its proof is beyond the scope of this introductory book.

**Theorem 13.7.** If $X_1, X_2, ..., X_n$ are observations of a random sample of size $n$ from the normal distribution $N\left(\mu, \sigma^2\right)$, then the sample mean $\overline{X} =$

$\frac{1}{n}\sum_{i=1}^{n} X_i$ and the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ have the following properties:

(A) $\overline{X}$ and $S^2$ are independent, and

(B) $(n-1)\,\frac{S^2}{\sigma^2} \sim \chi^2(n-1)$.

**Remark 13.2.** At first sight the statement (A) might seem odd since the sample mean $\overline{X}$ occurs explicitly in the definition of the sample variance $S^2$. This remarkable independence of $\overline{X}$ and $S^2$ is a unique property that distinguishes normal distribution from all other probability distributions.

**Example 13.9.** Let $X_1, X_2, ..., X_n$ denote a random sample from a normal distribution with variance $\sigma^2 > 0$. If the first percentile of the statistics $W = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2}$ is 1.24, where $\overline{X}$ denotes the sample mean, what is the sample size $n$?

**Answer:**

$$\begin{aligned}
\frac{1}{100} &= P\left(W \le 1.24\right) \\
&= P\left(\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \le 1.24\right) \\
&= P\left((n-1)\,\frac{S^2}{\sigma^2} \le 1.24\right) \\
&= P\left(\chi^2(n-1) \le 1.24\right).
\end{aligned}$$

Thus from $\chi^2$-table, we get

$$n - 1 = 7$$

and hence the sample size $n$ is 8.

**Example 13.10.** Let $X_1, X_2, ..., X_4$ be a random sample from a normal distribution with unknown mean and variance equal to 9. Let $S^2 = \frac{1}{3}\sum_{i=1}^{4}\left(X_i - \overline{X}\right)$. If $P\left(S^2 \le k\right) = 0.05$, then what is $k$?

**Answer:**

$$\begin{aligned}
0.05 &= P\left(S^2 \le k\right) \\
&= P\left(\frac{3S^2}{9} \le \frac{3}{9}\,k\right) \\
&= P\left(\chi^2(3) \le \frac{3}{9}\,k\right).
\end{aligned}$$

From $\chi^2$-table with 3 degrees of freedom, we get

$$\frac{3}{9}\,k = 0.35$$

and thus the constant $k$ is given by

$$k = 3(0.35) = 1.05.$$

## 13.2. Laws of Large Numbers

In this section, we mainly examine the weak law of large numbers. The weak law of large numbers states that if $X_1, X_2, ..., X_n$ is a random sample of size $n$ from a population $X$ with mean $\mu$, then the sample mean $\overline{X}$ rarely deviates from the population mean $\mu$ when the sample size $n$ is very large. In other words, the sample mean $\overline{X}$ converges in probability to the population mean $\mu$. We begin this section with a result known as Markov inequality which is needed to establish the weak law of large numbers.

**Theorem 13.8 (Markov Inequality).** Suppose $X$ is a nonnegative random variable with mean $E(X)$. Then

$$P(X \geq t) \leq \frac{E(X)}{t}$$

for all $t > 0$.

**Proof:** We assume the random variable $X$ is continuous. If $X$ is not continuous, then a proof can be obtained for this case by replacing the integrals with summations in the following proof. Since

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_{-\infty}^{t} x f(x) dx + \int_{t}^{\infty} x f(x) dx \\
&\geq \int_{t}^{\infty} x f(x) dx \\
&\geq \int_{t}^{\infty} t f(x) dx \qquad \text{because } x \in [t, \infty) \\
&= t \int_{t}^{\infty} f(x) dx \\
&= t\, P(X \geq t),
\end{aligned}$$

we see that

$$P(X \geq t) \leq \frac{E(X)}{t}.$$

This completes the proof of the theorem.

In Theorem 4.4 of the chapter 4, Chebychev inequality was treated. Let $X$ be a random variable with mean $\mu$ and standard deviation $\sigma$. Then Chebychev inequality says that

$$P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

for any nonzero positive constant $k$. This result can be obtained easily using Theorem 13.8 as follows. By Markov inequality, we have

$$P((X-\mu)^2 \geq t^2) \leq \frac{E((X-\mu)^2)}{t^2}$$

for all $t > 0$. Since the events $(X-\mu)^2 \geq t^2$ and $|X-\mu| \geq t$ are same, we get

$$P((X-\mu)^2 \geq t^2) = P(|X-\mu| \geq t) \leq \frac{E((X-\mu)^2)}{t^2}$$

for all $t > 0$. Hence

$$P(|X-\mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

Letting $t = k\sigma$ in the above equality, we see that

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Hence

$$1 - P(|X-\mu| < k\sigma) \leq \frac{1}{k^2}.$$

The last inequality yields the Chebychev inequality

$$P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

Now we are ready to treat the weak law of large numbers.

**Theorem 13.9.** Let $X_1, X_2, ...$ be a sequence of independent and identically distributed random variables with $\mu = E(X_i)$ and $\sigma^2 = Var(X_i) < \infty$ for $i = 1, 2, ..., \infty$. Then

$$\lim_{n\to\infty} P(|\overline{S}_n - \mu| \geq \varepsilon) = 0$$

for every $\varepsilon$. Here $\overline{S}_n$ denotes $\frac{X_1+X_2+\cdots+X_n}{n}$.

**Proof:** By Theorem 13.2 (or Example 13.7) we have

$$E(\overline{S}_n) = \mu \qquad \text{and} \qquad Var(\overline{S}_n) = \frac{\sigma^2}{n}.$$

By Chebychev's inequality

$$P(|\overline{S}_n - E(\overline{S}_n)| \geq \varepsilon) \leq \frac{Var(\overline{S}_n)}{\varepsilon^2}$$

for $\varepsilon > 0$. Hence

$$P(|\overline{S}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\,\varepsilon^2}.$$

Taking the limit as $n$ tends to infinity, we get

$$\lim_{n\to\infty} P(|\overline{S}_n - \mu| \geq \varepsilon) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\,\varepsilon^2}$$

which yields

$$\lim_{n\to\infty} P(|\overline{S}_n - \mu| \geq \varepsilon) = 0$$

and the proof of the theorem is now complete.

It is possible to prove the weak law of large numbers assuming only $E(X)$ to exist and finite but the proof is more involved.

The weak law of large numbers says that the sequence of sample means $\left\{\overline{S}_n\right\}_{n=1}^{\infty}$ from a population $X$ stays close to the population mean $E(X)$ most of the time. Let us consider an experiment that consists of tossing a coin infinitely many times. Let $X_i$ be 1 if the $i^{\text{th}}$ toss results in a Head, and 0 otherwise. The weak law of large numbers says that

$$\overline{S}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \to \frac{1}{2} \qquad \text{as} \quad n \to \infty \tag{13.0}$$

but it is easy to come up with sequences of tosses for which (13.0) is false:

$$\begin{array}{cccccccccccccc} H & H & H & H & H & H & H & H & H & H & H & H & \cdots & \cdots \\ H & H & T & H & H & T & H & H & T & H & H & T & \cdots & \cdots \end{array}$$

The strong law of large numbers (Theorem 13.11) states that the set of "bad sequences" like the ones given above has probability zero.

Note that the assertion of Theorem 13.9 for any $\varepsilon > 0$ can also be written as

$$\lim_{n\to\infty} P(|\overline{S}_n - \mu| < \varepsilon) = 1.$$

The type of convergence we saw in the weak law of large numbers is not the type of convergence discussed in calculus. This type of convergence is called convergence in probability and defined as follows.

**Definition 13.1.** Suppose $X_1, X_2, ...$ is a sequence of random variables defined on a sample space $S$. The sequence *converges in probability* to the random variable $X$ if, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1.$$

In view of the above definition, the weak law of large numbers states that the sample mean $\overline{X}$ converges in probability to the population mean $\mu$.

The following theorem is known as the Bernoulli law of large numbers and is a special case of the weak law of large numbers.

**Theorem 13.10.** Let $X_1, X_2, ...$ be a sequence of independent and identically distributed Bernoulli random variables with probability of success $p$. Then, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|\overline{S}_n - p| < \varepsilon) = 1$$

where $\overline{S}_n$ denotes $\frac{X_1+X_2+\cdots+X_n}{n}$.

The fact that the relative frequency of occurrence of an event $E$ is very likely to be close to its probability $P(E)$ for large $n$ can be derived from the weak law of large numbers. Consider a repeatable random experiment repeated large number of time independently. Let $X_i = 1$ if $E$ occurs on the $i^{\text{th}}$ repetition and $X_i = 0$ if $E$ does not occur on $i^{\text{th}}$ repetition. Then

$$\mu = E(X_i) = 1 \cdot P(E) + 0 \cdot P(E) = P(E) \qquad \text{for } i = 1, 2, 3, ...$$

and

$$X_1 + X_2 + \cdots + X_n = N(E)$$

where $N(E)$ denotes the number of times $E$ occurs. Hence by the weak law of large numbers, we have

$$\begin{aligned}\lim_{n\to\infty} P\left(\left|\frac{N(E)}{n} - P(E)\right| \geq \varepsilon\right) &= \lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right)\\ &= \lim_{n\to\infty} P\left(\left|\overline{S}_n - \mu\right| \geq \varepsilon\right)\\ &= 0.\end{aligned}$$

Hence, for large $n$, the relative frequency of occurrence of the event $E$ is very likely to be close to its probability $P(E)$.

Now we present the strong law of large numbers without a proof.

**Theorem 13.11.** Let $X_1, X_2, ...$ be a sequence of independent and identically distributed random variables with $\mu = E(X_i)$ and $\sigma^2 = Var(X_i) < \infty$ for $i = 1, 2, ..., \infty$. Then

$$P\left(\lim_{n\to\infty} \overline{S}_n = \mu\right) = 1$$

for every $\varepsilon > 0$. Here $\overline{S}_n$ denotes $\frac{X_1+X_2+\cdots+X_n}{n}$.

The type convergence in Theorem 13.11 is called almost sure convergence. The notion of almost sure convergence is defined as follows.

**Definition 13.2** Suppose the random variable $X$ and the sequence $X_1, X_2, ...$, of random variables are defined on a sample space $S$. The sequence $X_n(w)$ *converges almost surely* to $X(w)$ if

$$P\left(\left\{w \in S \;\Big|\; \lim_{n\to\infty} X_n(w) = X(w)\right\}\right) = 1.$$

It can be shown that the convergence in probability implies the almost sure convergence but not the converse.

## 13.3. The Central Limit Theorem

Consider a random sample of measurement $\{X_i\}_{i=1}^n$. The $X_i$'s are identically distributed and their common distribution is the distribution of the population. We have seen that if the population distribution is normal, then the sample mean $\overline{X}$ is also normal. More precisely, if $X_1, X_2, ..., X_n$ is a random sample from a normal distribution with density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\; e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

then

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

The central limit theorem (also known as Lindeberg-Levy Theorem) states that even though the population distribution may be far from being normal, still for large sample size $n$, the distribution of the standardized sample mean is approximately standard normal with better approximations obtained with the larger sample size. Mathematically this can be stated as follows.

**Theorem 13.12 (Central Limit Theorem).** Let $X_1, X_2, ..., X_n$ be a random sample of size $n$ from a distribution with mean $\mu$ and variance $\sigma^2 < \infty$, then the limiting distribution of

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

is standard normal, that is $Z_n$ converges in distribution to $Z$ where $Z$ denotes a standard normal random variable.

The type of convergence used in the central limit theorem is called the convergence in distribution and is defined as follows.

**Definition 13.3.** Suppose $X$ is a random variable with cumulative density function $F(x)$ and the sequence $X_1, X_2, ...$ of random variables with cumulative density functions $F_1(x), F_2(x), ...$, respectively. The sequence $X_n$ *converges in distribution* to $X$ if

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for all values $x$ at which $F(x)$ is continuous. The distribution of $X$ is called the *limiting distribution* of $X_n$.

Whenever a sequence of random variables $X_1, X_2, ...$ converges in distribution to the random variable $X$, it will be denoted by $X_n \xrightarrow{d} X$.

**Example 13.11.** Let $Y = X_1 + X_2 + \cdots + X_{15}$ be the sum of a random sample of size 15 from the distribution whose density function is

$$f(x) = \begin{cases} \frac{3}{2}\,x^2 & \text{if } -1 < x < 1 \\ \\ 0 & \text{otherwise.} \end{cases}$$

What is the approximate value of $P(-0.3 \leq Y \leq 1.5)$ when one uses the central limit theorem?

**Answer:** First, we find the mean $\mu$ and variance $\sigma^2$ for the density function $f(x)$. The mean for this distribution is given by

$$\begin{aligned} \mu &= \int_{-1}^{1} \frac{3}{2} x^3 \, dx \\ &= \frac{3}{2} \left[ \frac{x^4}{4} \right]_{-1}^{1} \\ &= 0. \end{aligned}$$

Hence the variance of this distribution is given by

$$\begin{aligned} Var(X) &= E(X^2) - [\,E(X)\,]^2 \\ &= \int_{-1}^{1} \frac{3}{2} x^4 \, dx \\ &= \frac{3}{2} \left[ \frac{x^5}{5} \right]_{-1}^{1} \\ &= \frac{3}{5} \\ &= 0.6. \end{aligned}$$

$$\begin{aligned} P(-0.3 \leq Y \leq 1.5) &= P(-0.3 - 0 \leq Y - 0 \leq 1.5 - 0) \\ &= P\left( \frac{-0.3}{\sqrt{15(0.6)}} \leq \frac{Y-0}{\sqrt{15(0.6)}} \leq \frac{1.5}{\sqrt{15(0.6)}} \right) \\ &= P(-0.10 \leq Z \leq 0.50) \\ &= P(Z \leq 0.50) + P(Z \leq 0.10) - 1 \\ &= 0.6915 + 0.5398 - 1 \\ &= 0.2313. \end{aligned}$$

**Example 13.12.** Let $X_1, X_2, ..., X_n$ be a random sample of size $n = 25$ from a population that has a mean $\mu = 71.43$ and variance $\sigma^2 = 56.25$. Let $\overline{X}$ be the sample mean. What is the probability that the sample mean is between 68.91 and 71.97?

**Answer:** The mean of $\overline{X}$ is given by $E\left(\overline{X}\right) = 71.43$. The variance of $\overline{X}$ is given by

$$Var\left(\overline{X}\right) = \frac{\sigma^2}{n} = \frac{56.25}{25} = 2.25.$$

In order to find the probability that the sample mean is between 68.91 and 71.97, we need the distribution of the population. However, the population distribution is unknown. Therefore, we use the central limit theorem. The central limit theorem says that $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,\,1)$ as $n$ approaches infinity. Therefore

$$\begin{aligned} & P\left(68.91 \leq \overline{X} \leq 71.97\right) \\ &= \left( \frac{68.91 - 71.43}{\sqrt{2.25}} \leq \frac{\overline{X} - 71.43}{\sqrt{2.25}} \leq \frac{71.97 - 71.43}{\sqrt{2.25}} \right) \\ &= P\,(-0.68 \leq W \leq 0.36) \\ &= P\,(W \leq 0.36) + P\,(W \leq 0.68) - 1 \\ &= 0.5941. \end{aligned}$$

**Example 13.13.** Light bulbs are installed successively into a socket. If we assume that each light bulb has a mean life of 2 months with a standard deviation of 0.25 months, what is the probability that 40 bulbs last at least 7 years?

**Answer:** Let $X_i$ denote the life time of the $i^{\text{th}}$ bulb installed. The 40 light bulbs last a total time of

$$S_{40} = X_1 + X_2 + \cdots + X_{40}.$$

By the central limit theorem

$$\frac{\sum_{i=1}^{40} X_i - n\mu}{\sqrt{n\sigma^2}} \sim N(0,1) \qquad \text{as} \qquad n \to \infty.$$

Thus

$$\frac{S_{40} - (40)(2)}{\sqrt{(40)(0.25)^2}} \sim N(0,1).$$

That is

$$\frac{S_{40} - 80}{1.581} \sim N(0,1).$$

Therefore

$$\begin{aligned} & P\left(S_{40} \geq 7(12)\right) \\ & \quad = P\left(\frac{S_{40} - 80}{1.581} \geq \frac{84 - 80}{1.581}\right) \\ & \quad = P\left(Z \geq 2.530\right) \\ & \quad = 0.0057. \end{aligned}$$

**Example 13.14.** Light bulbs are installed into a socket. Assume that each has a mean life of 2 months with standard deviation of 0.25 month. How many bulbs $n$ should be bought so that one can be 95% sure that the supply of $n$ bulbs will last 5 years?

**Answer:** Let $X_i$ denote the life time of the $i^{\text{th}}$ bulb installed. The $n$ light bulbs last a total time of

$$S_n = X_1 + X_2 + \cdots + X_n.$$

The total average life span $S_n$ has

$$E\left(S_n\right) = 2n \qquad \text{and} \qquad Var(S_n) = \frac{n}{16}.$$

By the central limit theorem, we get

$$\frac{S_n - E\left(S_n\right)}{\frac{\sqrt{n}}{4}} \sim N(0,1).$$

Thus, we seek $n$ such that

$$\begin{aligned}
0.95 &= P\left(S_n \geq 60\right) \\
&= P\left(\frac{S_n - 2n}{\frac{\sqrt{n}}{4}} \geq \frac{60 - 2n}{\frac{\sqrt{n}}{4}}\right) \\
&= P\left(Z \geq \frac{240 - 8n}{\sqrt{n}}\right) \\
&= 1 - P\left(Z \leq \frac{240 - 8n}{\sqrt{n}}\right).
\end{aligned}$$

From the standard normal table, we get

$$\frac{240 - 8n}{\sqrt{n}} = -1.645$$

which implies

$$1.645\sqrt{n} + 8n - 240 = 0.$$

Solving this quadratic equation for $\sqrt{n}$, we get

$$\sqrt{n} = -5.375 \quad \text{or} \quad 5.581.$$

Thus $n = 31.15$. So we should buy 32 bulbs.

**Example 13.15.** American Airlines claims that the average number of people who pay for in-flight movies, when the plane is fully loaded, is 42 with a standard deviation of 8. A sample of 36 fully loaded planes is taken. What is the probability that fewer than 38 people paid for the in-flight movies?

**Answer:** Here, we like to find $P(\overline{X} < 38)$. Since, we do not know the distribution of $\overline{X}$, we will use the central limit theorem. We are given that the population mean is $\mu = 42$ and population standard deviation is $\sigma = 8$. Moreover, we are dealing with sample of size $n = 36$. Thus

$$\begin{aligned}
P(\overline{X} < 38) &= P\left(\frac{\overline{X} - 42}{\frac{8}{6}} < \frac{38 - 42}{\frac{8}{6}}\right) \\
&= P\left(Z < -3\right) \\
&= 1 - P(Z < 3) \\
&= 1 - 0.9987 \\
&= 0.0013.
\end{aligned}$$

Since we have not yet seen the proof of the central limit theorem, first let us go through some examples to see the main idea behind the proof of the central limit theorem. Later, at the end of this section a proof of the central limit theorem will be given. We know from the central limit theorem that if $X_1, X_2, ..., X_n$ is a random sample of size $n$ from a distribution with mean $\mu$ and variance $\sigma^2$, then

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} Z \sim N(0,\, 1) \qquad \text{as} \qquad n \to \infty.$$

However, the above expression is *not equivalent* to

$$\overline{X} \xrightarrow{d} Z \sim N\left(\mu,\, \frac{\sigma^2}{n}\right) \qquad \text{as} \qquad n \to \infty$$

as the following example shows.

**Example 13.16.** Let $X_1, X_2, ..., X_n$ be a random sample of size $n$ from a gamma distribution with parameters $\theta = 1$ and $\alpha = 1$. What is the distribution of the sample mean $\overline{X}$? Also, what is the limiting distribution of $\overline{X}$ as $n \to \infty$?

**Answer:** Since, each $X_i \sim GAM(1, 1)$, the probability density function of each $X_i$ is given by

$$f(x) = \begin{cases} e^{-x} & \text{if } \; x \geq 0 \\ \\ 0 & \text{otherwise} \end{cases}$$

and hence the moment generating function of each $X_i$ is

$$M_{X_i}(t) = \frac{1}{1-t}.$$

First we determine the moment generating function of the sample mean $\overline{X}$, and then examine this moment generating function to find the probability distribution of $\overline{X}$. Since

$$\begin{aligned} M_{\overline{X}}(t) &= M_{\frac{1}{n}\sum_{i=1}^{n} X_i}(t) \\ &= \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{n}\right) \\ &= \prod_{i=1}^{n} \frac{1}{\left(1 - \frac{t}{n}\right)} \\ &= \frac{1}{\left(1 - \frac{t}{n}\right)^n}, \end{aligned}$$

therefore $\overline{X} \sim GAM\left(\frac{1}{n},\, n\right)$.

Next, we find the limiting distribution of $\overline{X}$ as $n \to \infty$. This can be done again by finding the limiting moment generating function of $\overline{X}$ and identifying the distribution of $\overline{X}$. Consider

$$\begin{aligned}
\lim_{n\to\infty} M_{\overline{X}}(t) &= \lim_{n\to\infty} \frac{1}{\left(1-\frac{t}{n}\right)^n} \\
&= \frac{1}{\lim_{n\to\infty}\left(1-\frac{t}{n}\right)^n} \\
&= \frac{1}{e^{-t}} \\
&= e^t.
\end{aligned}$$

Thus, the sample mean $\overline{X}$ has a degenerate distribution, that is all the probability mass is concentrated at one point of the space of $\overline{X}$.

**Example 13.17.** Let $X_1, X_2, ..., X_n$ be a random sample of size $n$ from a gamma distribution with parameters $\theta = 1$ and $\alpha = 1$. What is the distribution of

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \qquad \text{as} \qquad n \to \infty$$

where $\mu$ and $\sigma$ are the population mean and variance, respectively?

**Answer:** From Example 13.7, we know that

$$M_{\overline{X}}(t) = \frac{1}{\left(1-\frac{t}{n}\right)^n}.$$

Since the population distribution is gamma with $\theta = 1$ and $\alpha = 1$, the population mean $\mu$ is 1 and population variance $\sigma^2$ is also 1. Therefore

$$\begin{aligned}
M_{\frac{\overline{X}-1}{\frac{1}{\sqrt{n}}}}(t) &= M_{\sqrt{n}\,\overline{X}-\sqrt{n}}(t) \\
&= e^{-\sqrt{n}t}\, M_{\overline{X}}\left(\sqrt{n}\,t\right) \\
&= e^{-\sqrt{n}t}\, \frac{1}{\left(1-\frac{\sqrt{n}t}{n}\right)^n} \\
&= \frac{1}{e^{\sqrt{n}t}\left(1-\frac{t}{\sqrt{n}}\right)^n}.
\end{aligned}$$

The limiting moment generating function can be obtained by taking the limit of the above expression as $n$ tends to infinity. That is,

$$\begin{aligned}
\lim_{n\to\infty} M_{\frac{\overline{X}-1}{\frac{1}{\sqrt{n}}}}(t) &= \lim_{n\to\infty} \frac{1}{e^{\sqrt{n}t}\left(1-\frac{t}{\sqrt{n}}\right)^n} \\
&= e^{\frac{1}{2}t^2} \qquad \text{(using MAPLE)} \\
&= \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1).
\end{aligned}$$

The following theorem is used to prove the central limit theorem.

**Theorem 13.13 (Lévy Continuity Theorem).** Let $X_1, X_2, ...$ be a sequence of random variables with distribution functions $F_1(x), F_2(x), ...$ and moment generating functions $M_{X_1}(t), M_{X_2}(t), ...$, respectively. Let $X$ be a random variable with distribution function $F(x)$ and moment generating function $M_X(t)$. If for all $t$ in the open interval $(-h, h)$ for some $h > 0$

$$\lim_{n\to\infty} M_{X_n}(t) = M_X(t),$$

then at the points of continuity of $F(x)$

$$\lim_{n\to\infty} F_n(x) = F(x).$$

The proof of this theorem is beyond the scope of this book.

The following limit

$$\lim_{n\to\infty} \left[1 + \frac{t}{n} + \frac{d(n)}{n}\right]^n = e^t, \qquad \text{if } \lim_{n\to\infty} d(n) = 0, \tag{13.1}$$

whose proof we leave it to the reader, can be established using advanced calculus. Here $t$ is independent of $n$.

Now we proceed to prove the central limit theorem assuming that the moment generating function of the population $X$ exists. Let $M_{X-\mu}(t)$ be the moment generating function of the random variable $X - \mu$. We denote $M_{X-\mu}(t)$ as $M(t)$ when there is no danger of confusion. Then

$$\left.\begin{aligned}
M(0) &= 1, \\
M'(0) &= E(X-\mu) = E(X) - \mu = \mu - \mu = 0, \\
M''(0) &= E\left((X-\mu)^2\right) = \sigma^2.
\end{aligned}\right\} \tag{13.2}$$

By Taylor series expansion of $M(t)$ about 0, we get

$$M(t) = M(0) + M'(0)\, t + \frac{1}{2}\, M''(\eta)\, t^2$$

where $\eta \in (0, t)$. Hence using (13.2), we have

$$\begin{aligned}
M(t) &= 1 + \frac{1}{2}\, M''(\eta)\, t^2 \\
&= 1 + \frac{1}{2}\, \sigma^2\, t^2 + \frac{1}{2}\, M''(\eta)\, t^2 - \frac{1}{2}\, \sigma^2\, t^2 \\
&= 1 + \frac{1}{2}\, \sigma^2\, t^2 + \frac{1}{2}\, \left[ M''(\eta) - \sigma^2 \right]\, t^2.
\end{aligned}$$

Now using $M(t)$ we compute the moment generating function of $Z_n$. Note that

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

Hence

$$\begin{aligned}
M_{Z_n}(t) &= \prod_{i=1}^{n} M_{X_i - \mu} \left( \frac{t}{\sigma\,\sqrt{n}} \right) \\
&= \prod_{i=1}^{n} M_{X - \mu} \left( \frac{t}{\sigma\,\sqrt{n}} \right) \\
&= \left[ M \left( \frac{t}{\sigma\,\sqrt{n}} \right) \right]^n \\
&= \left[ 1 + \frac{t^2}{2\,n} + \frac{(M''(\eta) - \sigma^2)\, t^2}{2\,n\,\sigma^2} \right]^n
\end{aligned}$$

for $0 < |\eta| < \frac{1}{\sigma\,\sqrt{n}}\, |t|$. Note that since $0 < |\eta| < \frac{1}{\sigma\,\sqrt{n}}\, |t|$, we have

$$\lim_{n\to\infty} \frac{t}{\sigma\,\sqrt{n}} = 0, \quad \lim_{n\to\infty} \eta = 0, \quad \text{and} \quad \lim_{n\to\infty} M''(\eta) - \sigma^2 = 0. \tag{13.3}$$

Letting

$$d(n) = \frac{(M''(\eta) - \sigma^2)\, t^2}{2\,\sigma^2}$$

and using (13.3), we see that $\lim_{n\to\infty} d(n) = 0$, and

$$M_{Z_n}(t) = \left[ 1 + \frac{t^2}{2\,n} + \frac{d(n)}{n} \right]^n. \tag{13.4}$$

Using (13.1) we have

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty} \left[ 1 + \frac{t^2}{2\,n} + \frac{d(n)}{n} \right]^n = e^{\frac{1}{2}\, t^2}.$$

Hence by the Lévy continuity theorem, we obtain

$$\lim_{n\to\infty} F_n(x) = \Phi(x)$$

where $\Phi(x)$ is the cumulative density function of the standard normal distribution. Thus $Z_n \xrightarrow{d} Z$ and the proof of the theorem is now complete.

Now we give another proof of the central limit theorem using L'Hospital rule. This proof is essentially due to Tardiff (1980).

As before, let $Z_n = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$. Then $M_{Z_n}(t) = \left[ M\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$ where $M(t)$ is the moment generating function of the random variable $X-\mu$. Hence from (13.2), we have $M(0) = 1,\ M'(0) = 0$, and $M''(0) = \sigma^2$. Now applying the L'Hospital rule twice we compute

$$\begin{aligned}
&\lim_{n\to\infty} M_{Z_n}(t) \\
&= \lim_{n\to\infty} \left[ M\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n \\
&= \lim_{n\to\infty} exp\left( n \ln\left( M\left(\frac{t}{\sigma\sqrt{n}}\right)\right)\right) \\
&= \lim_{n\to\infty} exp\left( \frac{\ln\left(M\left(\frac{t}{\sigma\sqrt{n}}\right)\right)}{\frac{1}{n}}\right) \qquad \left(\frac{0}{0} \text{ form since M(0) = 1}\right) \\
&= \lim_{n\to\infty} exp\left( \frac{t}{2\sigma}\, \frac{M\left(\frac{t}{\sigma\sqrt{n}}\right)^{-1} M'\left(\frac{t}{\sigma\sqrt{n}}\right)\left(-\frac{1}{n\sqrt{n}}\right)}{-\frac{1}{n^2}}\right) \text{(L'Hospital rule)} \\
&= \lim_{n\to\infty} exp\left( \frac{t}{2\sigma}\, \frac{M\left(\frac{t}{\sigma\sqrt{n}}\right)^{-1} M'\left(\frac{t}{\sigma\sqrt{n}}\right)}{\frac{1}{\sqrt{n}}}\right) \left(\frac{0}{0} \text{ form since M}'(0) = 0\right) \\
&= \lim_{n\to\infty} exp\left( \frac{t^2}{2\sigma^2}\, \frac{M\left(\frac{t}{\sigma\sqrt{n}}\right) M''\left(\frac{t}{\sigma\sqrt{n}}\right) - \left\{M'\left(\frac{t}{\sigma\sqrt{n}}\right)\right\}^2}{M\left(\frac{t}{\sigma\sqrt{n}}\right)^2}\right) \\
&= exp\left( \frac{t^2}{2\sigma^2}\, \frac{M(0)\, M''(0) - \{M'(0)\}^2}{M(0)^2}\right) \\
&= exp\left( \frac{t^2}{2\sigma^2}\, \left[1\cdot\sigma^2 - 0^2\right]\right) \\
&= exp\left(\frac{1}{2}\, t^2\right).
\end{aligned}$$

Hence by the Lévy continuity theorem, we obtain

$$\lim_{n\to\infty} F_n(x) = \Phi(x)$$

where $\Phi(x)$ is the cumulative density function of the standard normal distribution. Thus as $n \to \infty$, the random variable $Z_n \stackrel{d}{\to} Z$, where $Z \sim N(0,1)$.

**Remark 13.3.** In contrast to the moment generating function, since the characteristic function of a random variable always exists, the original proof of the central limit theorem involved the characteristic function (see for example *An Introduction to Probability Theory and Its Applications, Volume II* by Feller). In 1988, Brown gave an elementary proof using very clever Taylor series expansions, where the use of the characteristic function has been avoided.

## 13.4. Order Statistics

Often, sample values such as the smallest, largest, or middle observation from a random sample provide important information. For example, the highest flood water or lowest winter temperature recorded during the last 50 years might be useful when planning for future emergencies. The median price of houses sold during the previous month might be useful for estimating the cost of living. The statistics highest, lowest or median are examples of order statistics.

**Definition 13.4.** Let $X_1, X_2, ..., X_n$ be observations from a random sample of size $n$ from a distribution $f(x)$. Let $X_{(1)}$ denote the smallest of $\{X_1, X_2, ..., X_n\}$, $X_{(2)}$ denote the second smallest of $\{X_1, X_2, ..., X_n\}$, and similarly $X_{(r)}$ denote the $r^{\text{th}}$ smallest of $\{X_1, X_2, ..., X_n\}$. Then the random variables $X_{(1)}, X_{(2)}, ..., X_{(n)}$ are called the order statistics of the sample $X_1, X_2, ..., X_n$. In particular, $X_{(r)}$ is called the $r^{\text{th}}$-order statistic of $X_1, X_2, ..., X_n$.

The sample range, $R$, is the distance between the smallest and the largest observation. That is,

$$R = X_{(n)} - X_{(1)}.$$

This is an important statistic which is defined using order statistics.

The distribution of the order statistics are very important when one uses these in any statistical investigation. The next theorem gives the distribution of an order statistic.

**Theorem 13.14.** Let $X_1, X_2, ..., X_n$ be a random sample of size $n$ from a distribution with density function $f(x)$. Then the probability density function of the $r^{\text{th}}$ order statistic, $X_{(r)}$, is

$$g(x) = \frac{n!}{(r-1)!\,(n-r)!}\,[F(x)]^{r-1}\,f(x)\,[1-F(x)]^{n-r},$$

where $F(x)$ denotes the cdf of $f(x)$.

**Proof:** We prove the theorem assuming $f(x)$ continuous. In the case $f(x)$ is discrete the proof has to be modified appropriately. Let $h$ be a positive real number and $x$ be an arbitrary point in the domain of $f$. Let us divide the real line into three segments, namely

$$\mathbb{R} = (-\infty,\, x) \bigcup [x,\, x+h) \bigcup [x+h,\, \infty).$$

The probability, say $p_1$, of a sample value falls into the first interval $(-\infty,\, x]$ and is given by

$$p_1 = \int_{-\infty}^{x} f(t)\, dt = F(x).$$

Similarly, the probability $p_2$ of a sample value falls into the second interval $[x,\, x+h)$ is

$$p_2 = \int_{x}^{x+h} f(t)\, dt = F(x+h) - F(x).$$

In the same token, we can compute the probability $p_3$ of a sample value which falls into the third interval

$$p_3 = \int_{x+h}^{\infty} f(t)\, dt = 1 - F(x+h).$$

Then the probability, $P_h(x)$, that $(r-1)$ sample values fall in the first interval, one falls in the second interval, and $(n-r)$ fall in the third interval is

$$P_h(x) = \binom{n}{r-1,\, 1,\, n-r} p_1^{r-1}\, p_2^{1}\, p_3^{n-r} = \frac{n!}{(r-1)!\,(n-r)!}\, p_1^{r-1}\, p_2\, p_3^{n-r}.$$

Hence the probability density function $g(x)$ of the $r^{\text{th}}$ statistics is given by

$$\begin{aligned}
&g(x)\\
&= \lim_{h\to 0} \frac{P_h(x)}{h}\\
&= \lim_{h\to 0} \left[ \frac{n!}{(r-1)!\,(n-r)!}\, p_1^{r-1}\, \frac{p_2}{h}\, p_3^{n-r} \right]\\
&= \frac{n!}{(r-1)!\,(n-r)!}\, [F(x)]^{r-1} \lim_{h\to 0} \frac{F(x+h)-F(x)}{h} \lim_{h\to 0} [1-F(x+h)]^{n-r}\\
&= \frac{n!}{(r-1)!\,(n-r)!}\, [F(x)]^{r-1}\, F'(x)\, [1-F(x)]^{n-r}\\
&= \frac{n!}{(r-1)!\,(n-r)!}\, [F(x)]^{r-1}\, f(x)\, [1-F(x)]^{n-r}.
\end{aligned}$$

**Example 13.18.** Let $X_1, X_2$ be a random sample from a distribution with density function

$$f(x) = \begin{cases} e^{-x} & \text{for } 0 \leq x < \infty \\ \\ 0 & \text{otherwise.} \end{cases}$$

What is the density function of $Y = \min\{X_1, X_2\}$ where nonzero?

**Answer:** The cumulative distribution function of $f(x)$ is

$$\begin{aligned}
F(x) &= \int_0^x e^{-t}\, dt\\
&= 1 - e^{-x}
\end{aligned}$$

In this example, $n = 2$ and $r = 1$. Hence, the density of $Y$ is

$$\begin{aligned}
g(y) &= \frac{2!}{0!\,1!}\, [F(y)]^0\, f(y)\, [1-F(y)]\\
&= 2f(y)\, [1-F(y)]\\
&= 2\, e^{-y} \left(1 - 1 + e^{-y}\right)\\
&= 2\, e^{-2y}.
\end{aligned}$$

**Example 13.19.** Let $Y_1 < Y_2 < \cdots < Y_6$ be the order statistics from a random sample of size 6 from a distribution with density function

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1 \\ \\ 0 & \text{otherwise.} \end{cases}$$

What is the expected value of $Y_6$?

**Answer:**

$$\begin{aligned} f(x) &= 2x \\ F(x) &= \int_0^x 2t\,dt \\ &= x^2. \end{aligned}$$

The density function of $Y_6$ is given by

$$\begin{aligned} g(y) &= \frac{6!}{5!\,0!}\;[F(y)]^5\; f(y) \\ &= 6\left(y^2\right)^5\; 2y \\ &= 12y^{11}. \end{aligned}$$

Hence, the expected value of $Y_6$ is

$$\begin{aligned} E\left(Y_6\right) &= \int_0^1 y\,g(y)\,dy \\ &= \int_0^1 y\,12y^{11}\,dy \\ &= \frac{12}{13}\;\left[y^{13}\right]_0^1 \\ &= \frac{12}{13}. \end{aligned}$$

**Example 13.20.** Let $X, Y$ and $Z$ be independent uniform random variables on the interval $(0, a)$. Let $W = \min\{X, Y, Z\}$. What is the expected value of $\left(1 - \frac{W}{a}\right)^2$?

**Answer:** The probability distribution of $X$ (or $Y$ or $Z$) is

$$f(x) = \begin{cases} \frac{1}{a} & \text{if } 0 < x < a \\ \\ 0 & \text{otherwise.} \end{cases}$$

Thus the cumulative distribution of function of $f(x)$ is given by

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \\ \frac{x}{a} & \text{if } 0 < x < a \\ \\ 1 & \text{if } x \geq a. \end{cases}$$

Since $W = \min\{X, Y, Z\}$, $W$ is the first order statistic of the random sample $X, Y, Z$. Thus, the density function of $W$ is given by

$$\begin{aligned}
g(w) &= \frac{3!}{0!\,1!\,2!}\;[F(w)]^0\;f(w)\;[1-F(w)]^2 \\
&= 3f(w)\;[1-F(w)]^2 \\
&= 3\left(1-\frac{w}{a}\right)^2\;\left(\frac{1}{a}\right) \\
&= \frac{3}{a}\;\left(1-\frac{w}{a}\right)^2.
\end{aligned}$$

Thus, the pdf of $W$ is given by

$$g(w) = \begin{cases} \frac{3}{a}\;\left(1-\frac{w}{a}\right)^2 & \text{if } 0 < w < a \\ \\ 0 & \text{otherwise.} \end{cases}$$

The expected value of $W$ is

$$\begin{aligned}
&E\left[\left(1-\frac{W}{a}\right)^2\right] \\
&\quad = \int_0^a \left(1-\frac{w}{a}\right)^2\; g(w)\,dw \\
&\quad = \int_0^a \left(1-\frac{w}{a}\right)^2\;\frac{3}{a}\;\left(1-\frac{w}{a}\right)^2\;dw \\
&\quad = \int_0^a \frac{3}{a}\;\left(1-\frac{w}{a}\right)^4\;dw \\
&\quad = -\frac{3}{5}\;\left[\left(1-\frac{w}{a}\right)^5\right]_0^a \\
&\quad = \frac{3}{5}.
\end{aligned}$$

**Example 13.21.** Let $X_1, X_2, ..., X_n$ be a random sample from a population $X$ with uniform distribution on the interval $[0, 1]$. What is the probability distribution of the sample range $W := X_{(n)} - X_{(1)}$?

**Answer:** To find the distribution of $W$, we need the joint distribution of the random variable $\big(X_{(n)}, X_{(1)}\big)$. The joint distribution of $\big(X_{(n)}, X_{(1)}\big)$ is given by

$$h(x_1, x_n) = n(n-1)f(x_1)f(x_n)\,[F(x_n)-F(x_1)]^{n-2},$$

where $x_n \geq x_1$ and $f(x)$ is the probability density function of $X$. To determine the probability distribution of the sample range $W$, we consider the transformation

$$\left.\begin{aligned} U &= X_{(1)} \\ W &= X_{(n)} - X_{(1)} \end{aligned}\right\}$$

which has an inverse

$$\left.\begin{aligned} X_{(1)} &= U \\ X_{(n)} &= U + W. \end{aligned}\right\}$$

The Jacobian of this transformation is

$$J = det \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = 1.$$

Hence the joint density of $(U, W)$ is given by

$$\begin{aligned} g(u, w) &= |J| \, h(x_1, x_n) \\ &= n(n-1) f(u) f(u+w) [F(u+w) - F(u)]^{n-2} \end{aligned}$$

where $w \geq 0$. Since $f(u)$ and $f(u+w)$ are simultaneously nonzero if $0 \leq u \leq 1$ and $0 \leq u + w \leq 1$. Hence $f(u)$ and $f(u+w)$ are simultaneously nonzero if $0 \leq u \leq 1 - w$. Thus, the probability of $W$ is given by

$$\begin{aligned} j(w) &= \int_{-\infty}^{\infty} g(u, w) \, du \\ &= \int_{-\infty}^{\infty} n(n-1) f(u) f(u+w) [F(u+w) - F(u)]^{n-2} \, du \\ &= n(n-1) \, w^{n-2} \int_0^{1-w} du \\ &= n(n-1) \, (1-w) \, w^{n-2} \end{aligned}$$

where $0 \leq w \leq 1$.

## 13.5. Sample Percentiles

The sample median, $M$, is a number such that approximately one-half of the observations are less than $M$ and one-half are greater than $M$.

**Definition 13.5.** Let $X_1, X_2, ..., X_n$ be a random sample. The sample median $M$ is defined as

$$M = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{if } n \text{ is odd} \\ \frac{1}{2} \left[ X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n+2}{2}\right)} \right] & \text{if } n \text{ is even.} \end{cases}$$

The median is a measure of location like sample mean.

Recall that for continuous distribution, $100p^{\text{th}}$ percentile, $\pi_p$, is a number such that

$$p = \int_{-\infty}^{\pi_p} f(x)\, dx.$$

**Definition 13.6.** The $100p^{\text{th}}$ sample percentile is defined as

$$\pi_p = \begin{cases} X_{([np])} & \text{if } p < 0.5 \\ M & \text{if } p = 0.5 \\ X_{(n+1-[n(1-p)])} & \text{if } p > 0.5. \end{cases}$$

where $[b]$ denote the number $b$ rounded to the nearest integer.

**Example 13.22.** Let $X_1, X_2, ..., X_{12}$ be a random sample of size 12. What is the $65^{\text{th}}$ percentile of this sample?

**Answer:**

$$\begin{aligned} 100p &= 65 \\ p &= 0.65 \\ n(1-p) &= (12)(1-0.65) = 4.2 \\ [n(1-p)] &= [4.2] = 4 \end{aligned}$$

Hence by definition of $65^{\text{th}}$ percentile is

$$\begin{aligned} \pi_{0.65} &= X_{(n+1-[n(1-p)])} \\ &= X_{(13-4)} \\ &= X_{(9)}. \end{aligned}$$

Thus, the $65^{\text{th}}$ percentile of the random sample $X_1, X_2, ..., X_{12}$ is the $9^{\text{th}}$-order statistic.

For any number $p$ between 0 and 1, the $100p^{\text{th}}$ sample percentile is an observation such that approximately $np$ observations are less than this observation and $n(1-p)$ observations are greater than this.

**Definition 13.7.** The $25^{\text{th}}$ percentile is called the lower quartile while the $75^{\text{th}}$ percentile is called the upper quartile. The distance between these two quartiles is called the interquartile range.

**Example 13.23.** If a sample of size 3 from a uniform distribution over $[0,1]$ is observed, what is the probability that the sample median is between $\frac{1}{4}$ and $\frac{3}{4}$?

**Answer:** When a sample of $(2n+1)$ random variables are observed, the $(n+1)^{\text{th}}$ smallest random variable is called the sample median. For our problem, the sample median is given by

$$X_{(2)} = 2^{\text{nd}} \text{ smallest } \{X_1, X_2, X_3\}.$$

Let $Y = X_{(2)}$. The density function of each $X_i$ is given by

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ \\ 0 & \text{otherwise.} \end{cases}$$

Hence, the cumulative density function of $f(x)$ is

$$F(x) = x.$$

Thus the density function of $Y$ is given by

$$\begin{aligned} g(y) &= \frac{3!}{1!\,1!}\; [F(y)]^{2-1}\; f(y)\; [1 - F(y)]^{3-2} \\ &= 6\, F(y)\, f(y)\; [1 - F(y)] \\ &= 6y\,(1-y). \end{aligned}$$

Therefore

$$\begin{aligned} P\left(\frac{1}{4} < Y < \frac{3}{4}\right) &= \int_{\frac{1}{4}}^{\frac{3}{4}} g(y)\, dy \\ &= \int_{\frac{1}{4}}^{\frac{3}{4}} 6\, y\,(1-y)\, dy \\ &= 6\, \left[\frac{y^2}{2} - \frac{y^3}{3}\right]_{\frac{1}{4}}^{\frac{3}{4}} \\ &= \frac{11}{16}. \end{aligned}$$

## 13.6. Review Exercises

**1.** Suppose we roll a die 1000 times. What is the probability that the sum of the numbers obtained lies between 3000 and 4000?

**2.** Suppose Kathy flip a coin 1000 times. What is the probability she will get at least 600 heads?

**3.** At a certain large university the weight of the male students and female students are approximately normally distributed with means and standard deviations of 180, and 20, and 130 and 15, respectively. If a male and female are selected at random, what is the probability that the sum of their weights is less than 280?

**4.** Seven observations are drawn from a population with an unknown continuous distribution. What is the probability that the least and the greatest observations bracket the median?

**5.** If the random variable $X$ has the density function

$$f(x) = \begin{cases} 2\,(1-x) & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise,} \end{cases}$$

what is the probability that the larger of 2 independent observations of $X$ will exceed $\frac{1}{2}$?

**6.** Let $X_1, X_2, X_3$ be a random sample from the uniform distribution on the interval (0, 1). What is the probability that the sample median is less than 0.4?

**7.** Let $X_1, X_2, X_3, X_4, X_5$ be a random sample from the uniform distribution on the interval (0, $\theta$), where $\theta$ is unknown, and let $X_{max}$ denote the largest observation. For what value of the constant $k$, the expected value of the random variable $kX_{max}$ is equal to $\theta$?

**8.** A random sample of size 16 is to be taken from a normal population having mean 100 and variance 4. What is the $90^{\text{th}}$ percentile of the distribution of the sample mean?

**9.** If the density function of a random variable $X$ is given by

$$f(x) = \begin{cases} \frac{1}{2x} & \text{for } \frac{1}{e} < x < e \\ 0 & \text{otherwise,} \end{cases}$$

what is the probability that one of the two independent observations of $X$ is less than 2 and the other is greater than 1?

**10.** Five observations have been drawn independently and at random from a continuous distribution. What is the probability that the next observation will be less than all of the first 5?

**11.** Let the random variable $X$ denote the length of time it takes to complete a mathematics assignment. Suppose the density function of $X$ is given by

$$f(x) = \begin{cases} e^{-(x-\theta)} & \text{for } \theta < x < \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta$ is a positive constant that represents the minimum time to complete a mathematics assignment. If $X_1, X_2, ..., X_5$ is a random sample from this distribution. What is the expected value of $X_{(1)}$?

**12.** Let $X$ and $Y$ be two independent random variables with identical probability density function given by

$$f(x) = \begin{cases} e^{-x} & \text{for } x > 0 \\ \\ 0 & \text{elsewhere.} \end{cases}$$

What is the probability density function of $W = \max\{X, Y\}$ ?

**13.** Let $X$ and $Y$ be two independent random variables with identical probability density function given by

$$f(x) = \begin{cases} \frac{3\,x^2}{\theta^3} & \text{for } 0 \leq x \leq \theta \\ \\ 0 & \text{elsewhere,} \end{cases}$$

for some $\theta > 0$. What is the probability density function of $W = \min\{X, Y\}$?

**14.** Let $X_1, X_2, ..., X_n$ be a random sample from a uniform distribution on the interval from 0 to 5. What is the limiting moment generating function of $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ as $n \to \infty$?

**15.** Let $X_1, X_2, ..., X_n$ be a random sample of size $n$ from a normal distribution with mean $\mu$ and variance 1. If the $75^{\text{th}}$ percentile of the statistic $W = \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$ is 28.24, what is the sample size $n$ ?

**16.** Let $X_1, X_2, ..., X_n$ be a random sample of size $n$ from a Bernoulli distribution with probability of success $p = \frac{1}{2}$. What is the limiting distribution the sample mean $\overline{X}$ ?

**17.** Let $X_1, X_2, ..., X_{1995}$ be a random sample of size 1995 from a distribution with probability density function

$$f(x) = \frac{e^{-\lambda}\,\lambda^x}{x!} \qquad\qquad x = 0, 1, 2, 3, ..., \infty.$$

What is the distribution of $1995\overline{X}$ ?

**18.** Suppose $X_1, X_2, ..., X_n$ is a random sample from the uniform distribution on $(0, 1)$ and $Z$ be the sample range. What is the probability that $Z$ is less than or equal to 0.5?

**19.** Let $X_1, X_2, ..., X_9$ be a random sample from a uniform distribution on the interval $[1, 12]$. Find the probability that the next to smallest is greater than or equal to 4?

**20.** A machine needs 4 out of its 6 independent components to operate. Let $X_1, X_2, ..., X_6$ be the lifetime of the respective components. Suppose each is exponentially distributed with parameter $\theta$. What is the probability density function of the machine lifetime?

**21.** Suppose $X_1, X_2, ..., X_{2n+1}$ is a random sample from the uniform distribution on $(0, 1)$. What is the probability density function of the sample median $X_{(n+1)}$?

**22.** Let $X$ and $Y$ be two random variables with joint density

$$f(x, y) = \begin{cases} 12x & \text{if } 0 < y < 2x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

What is the expected value of the random variable $Z = X^2Y^3 + X^2 - X - Y^3$?

**23.** Let $X_1, X_2, ..., X_{50}$ be a random sample of size 50 from a distribution with density

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\,\theta^\alpha}\, x^{\alpha-1} e^{-\frac{x}{\theta}} & \text{for } 0 < x < \infty \\ \\ 0 & \text{otherwise.} \end{cases}$$

What are the mean and variance of the sample mean $\overline{X}$?

**24.** Let $X_1, X_2, ..., X_{100}$ be a random sample of size 100 from a distribution with density

$$f(x) = \begin{cases} \frac{e^{-\lambda}\,\lambda^x}{x!} & \text{for } x = 0, 1, 2, ..., \infty \\ 0 & \text{otherwise.} \end{cases}$$

What is the probability that $\overline{X}$ greater than or equal to 1?