# Chapter 11

# Expectation and Variance of a random variable

The aim of this lecture is to define and introduce mathematical Expectation and variance of a function of discrete & continuous random variables and the distribution of the sampling mean.

## 11. Introduction

Based on the type of experiment, the outcomes could belong to limited categories, for example, Heads and Tails from tossing a coin; or 1,2,3,4,5,6 from throwing a die. Such a random variable is known as a *discrete random* variable. On the other hand, a random variable which is neither categorical nor discrete is referred to as *continuous random* variable. It is possible to study the *probability distribution* or simply the *distribution* of a random variable *X* along with the associated probabilities of its outcomes.

**Notation:** A random variable is normally denoted by capital letters, *X*, *Y*, etc and particular values of a random variable by small letters, *x*, *y*, etc.

## 11.1 Expectation and Variance

**Definition:** The mathematical expectation of a random variable *X*, written *E*(*X*), is the mean value of *X* and is defined as

$$\mu = E(X) = \sum xP(X = x) \text{ when } X \text{ is discrete}$$

or

$$\mu = E(X) = \int xf(x)dx \quad \text{when } X \text{ is continuous}$$

Population mean (or Expected value) of a function $g(X)$ of a random variable $X$ is defined as

$$E[g(X)] = \sum g(x)p(X = x) \quad \text{when } X \text{ is discrete}$$

or

$$E[g(X)] = \int g(x)f(x)dx \quad \text{when } X \text{ is continuous}$$

For example $E(x^2) = \sum x^2 p(X = x)$ or $E(x^2) = \int x^2 f(x)dx$ depending on the nature of $X$.

### 11.1.1 Some Properties of *E*(*X*)

1. $E(c) = c$, i.e. the Expected value of a constant $c$ is $c$.
2. $E(cX) = cE(X)$, i.e. the Expected value of $cX$ is $c$ times the Expected value of $X$. This can be proved as follows.

   **Proof**

   $$E(cX) = \sum c\, x\, p(X = x)$$

$$= c \sum x \; p(X = x)$$
$$= c \; E(X)$$

3. $E(X \pm Y) = E(X) \pm E(Y)$
4. $E(aX \pm bY) = a\,E(X) \pm b\,E(Y)$
5. $E(XY) = E(X)E(Y)$ (provided variables $X$ and $Y$ are independent of each other).

**Definition:** The variance of a random variable *X*, written *V*(*X*), is defined as
$$\sigma^2 = V(X) = E\left[(X - \mu)^2\right] = E\left[X^2\right] - \mu^2$$

since $E\left[(X-\mu)^2\right] = E\left[X^2 - 2\mu X + \mu^2\right] = E\left[X^2\right] - \mu^2$. Hence the variance of $X$ is the average squared distance of $X$ to its mean $\mu$.

Population Variance of a function $g(X)$ of a random variable $X$ is

$$V\left[g(X)\right] = E\left\{\left[g(X) - \mu_g\right]^2\right\} = E\left\{\left[g(X)\right]^2\right\} - \mu_g^2 \text{ where } \mu_g = E\left[g(X)\right]$$

Hence the variance of $g(X)$ is the average squared distance of $g(X)$ to its mean $\mu_g$.

### 11.1.2 Some Properties of *V(X)*

1. $V(c) = 0$ since $V(c) = E\left[c^2\right] - \{E[c]\}^2 = c^2 - c^2 = 0$

2. $V(cX) = c^2\,V(X)$ since
$$V(cX) = E\left\{[cX]^2\right\} - \{E[cX]\}^2 = c^2\left\{E\left[X^2\right] - \left[E(X)\right]^2\right\} = c^2\,V(X)$$
3. $V(X \pm Y) = V(X) + V(Y)$ , provided $X$ and $Y$ are independent.
4. $V(aX \pm bY) = a^2\,V(X) + b^2\,V(Y)$, provided $X$ and $Y$ are independent.

**Note:** (i) $V(-Y) = V(Y)$ since $V(-Y) = V(-1.Y) = (-1)^2\,V(Y) = V(Y)$

(ii) $V(X - Y) = V(X) + V(Y)$ provided $X$ and $Y$ are independent.

since $V(X - Y) = V\left[(X + (-Y))\right] = V(X) + V(-Y) = V(X) + V(Y)$

**Example 1:** ***X* and *Y* are two independent random variables with mean values of 5 and 7 and variances of 1.5 and 2 respectively. Find**

(i) *E*(*X*+2*Y*)
(ii) *E*(5-3*X*)
(iii) *V*(5*X*-4*Y*)

**Solution 1:** (i) *E*(*X* +2*Y*) = *E*(*X*) + 2*E*(*Y*) = 5 + 2*7 = 19
(ii) *E*(5 – 3*X*) = 5 – 3*E*(*X*) = 5 – 3*5 = - 10

(iii) $V(5X - 4Y) = V(5X) + V(4Y) = 5^2V(X) + 4^2V(Y)$
$= 25*1.5 + 16*2 = 69.5$

## 11.2 The distribution of linear combinations of normal random variables.

**Theorem 1.** Let $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ where $X$ and $Y$ are independent. Let $T = aX + bY$ where $a$ and $b$ are constants, then $T \sim N(\mu_T, \sigma_T^2)$ where $\mu_T = a\mu_X + b\mu_Y$ and $\sigma_T^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$

**Proof**

$$\mu_T = E(T) = E(aX + bY) = E(aX) + E(bY) = a\,E(X) + b\,E(Y) = a\,\mu_X + b\,\mu_Y$$

$$\sigma_T^2 = V(T) = V(aX + bY) = V(aX) + V(bY) = a^2V(X) + b^2V(Y) = a^2\sigma_X^2 + b^2\sigma_Y^2$$

since $X$ and $Y$ are independent.

- The **Proof** that $T$ is Normally distributed is difficult.

**Example 2:** The weight $X$ of an apple has a $N(80, 25)$ distribution and weight $Y$ of an orange has a $N(150, 39)$ distribution, both weights are measured in grams.

(a) Suppose an apple and an orange at random are selected, what is the distribution of the total weight, $T$?

**Solution (a):** Let $T = X + Y$. Then $E(T) = E(X + Y)$

$= E(X) + E(Y)$
$= 80 + 150$
$= 230$ gm

and

$V(T) = V(X + Y)$
$= V(X) + V(Y)$ since $X$ and $Y$ are independent
$= 25 + 39$
$= 64$ gm$^2$

$\therefore \quad T \sim N(230, 64)$

(b) Suppose two apples at random are selected, what is the distribution of the total weight, $T$?

**Solution (b):**

Let $X_1$ measure the weight of the first apple and $X_2$ measure the weight of the second apple. And let $T = X_1 + X_2$. Then

$$E(T) = E(X_1 + X_2)$$

$$= E(X_1) + E(X_2)$$
$$= 80 + 80$$
$$= 160 \text{ gm}$$

and

$$V(T) = V(X_1 + X_2)$$
$$= V(X_1) + V(X_2)$$
$$= 25 + 25$$
$$= 50 \text{ gm}^2 \quad \text{since } X_1 \text{ and } X_2 \text{ are independent.}$$

$$\therefore\ T \sim N(160, 50)$$

(c) Suppose an apple at random was selected, and then found another apple with exactly the same weight. What is the distribution of the total weight, *T*?

**Solution (c):**

Let $T = 2X$ Then $E(T) = E(2X) = 2E(X)$
$$= 2 * 80$$
$$= 160 \text{ gm}$$

and

$$V(T) = V(2X) = 2^2 V(X)$$
$$= 4 * 25$$
$$= 100 \text{ gm}^2$$

$$\therefore T \sim N(160, 100)$$

**Example 3.** Male heights $M \sim N(68,5)$ and Female heights $F \sim N(63,4)$ where heights are measured in inches. If a male and a female at random are selected, what is the probability that the female is taller than the male?

**Solution 3:** We are given that $E(M) = 68$, $V(M) = 5$, $E(F) = 63$ and $V(F) = 4$

**Hence** $p(F > M) = p(M - F) < 0 = p(D) < 0$ where $D = M - F$ *the* difference between male and female heights. Thus

$$E(D) = E(M - F) = E(M) - E(F) = 68 - 63 = 5 \text{ inches}$$
$$V(D) = V(M - F) = V(M) + V(F) = 5 + 4 = 9 \text{ inches}^2$$
$$\therefore D \sim N(5, 9)$$

**Hence** $p(F > M) = p(D < 0)$ where $D \sim N(5,9)$

$$= p\left(\frac{D-5}{3} < \frac{0-5}{3}\right)$$
$$= p\left(Z < -\frac{5}{3}\right) \text{ where } Z \sim N(0,1)$$
$$= p(Z < -1.67)$$
$$= 0.0475$$

**Conclusion:** The probability that a female will be taller than a male (if both are selected at random) is 0.0475 (i.e. a 4.75% chance).

## 11.3 The distribution of $\overline{X}$, the sample mean.

If you compute the **mean** of a sample of 10 numbers chosen at random from a population of size *N* (*N* being large), the value you obtain will not equal the **population** mean exactly; by chance it will be a little bit higher or a little bit lower. If you sampled sets of 10 numbers over and over again (computing the mean for each set), you would find that some sample means come much closer to the population mean than others. Some would be higher than the population mean and some would be lower.
Imagine sampling 10 numbers and computing the mean over and over again, say about 1,000 times, and then constructing a relative **frequency distribution** of those 1,000 means. Interest then would be on how the means are distributed. This distribution of means is a very good approximation to the *sampling distribution* of the mean.

### 11.3.1 Sampling Distribution of the Mean

**<u>Definition</u>:** The *sampling distribution* of the mean is a theoretical distribution that is approached as the number of samples in the relative frequency distribution increases. A *sampling distribution* can also be defined as the relative frequency distribution that would be obtained if all possible samples of a particular sample size were taken.

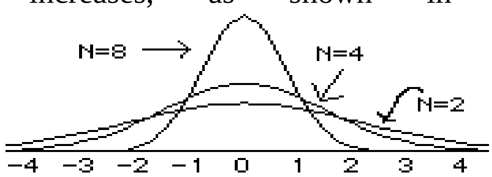
With 1,000 samples, the relative frequency distribution is quite close; with 10,000 it is even closer. As the number of samples approaches infinity, the relative frequency distribution approaches the sampling distribution of the mean.
The sampling distribution of the mean for a sample size of 10 was just an example; there is a different sampling distribution for other sample sizes. Also, keep in mind that the relative frequency distribution approaches a sampling distribution as **the number of samples** increases, not as the sample size increases since there is a different sampling distribution for each sample size.

The ***sampling distribution*** of the mean is a very important distribution. In later lectures you will see that it is used to construct **confidence intervals** for the mean and for **significance testing.**

Given a **population** with a mean of μ and a standard deviation of σ, the sampling distribution of the mean has a mean of $\mu$ and a standard deviation of $\frac{\sigma}{\sqrt{n}}$ where *n* is the sample size.

**<u>Definition</u>:** The standard deviation of the sampling distribution of the mean is called the ***standard error of the mean*** and is given by the formula**,** $\frac{\sigma}{\sqrt{n}}$**.**

Note that the **spread** of the sampling distribution of the mean decreases as the sample size increases, as shown in the following diagram.



In general, the larger the sample size the smaller the standard error.

### 11.3.2 Definition of a random sample of size $n$

The random variables $X_1$ , $X_2$ …. , $X_n$ are called random sample if

(i) they are **independent** of each other and

(ii) they have the **same** distribution.

**Example:** Suppose we select $n$ apples at random from an infinite population and weigh them. Let $X_i$ be the random variable measuring the weight of the $i$th apple for $i = 1, 2, 3, \ldots, n$. Then $X_1$ , $X_2$ , $X_3$ … , $X_n$ is a random sample of size $n$.

**Definition:** The random variable $\bar{X}$ is the sample mean of the random samples $X_1$ , $X_2$ , $X_3$ , … , $X_n$ defined by $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ .

### 11.3.3 The Expectation and Variance of $\bar{X}$

**Theorem 2.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with mean $\mu$ and variance $\sigma^2$, then

$$E(\bar{X}) = E(X) = \mu$$

and

$$V(\bar{X}) = \frac{V(X)}{n} = \frac{\sigma^2}{n}$$

**Proof 2**

$$E\left[\bar{X}\right] = E\left[\frac{1}{n}\sum_{i=1} X_i\right] = \frac{1}{n} E\left[\sum X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{n\mu}{n} = \mu$$

$$V(\bar{X}) = V(\frac{1}{n}\sum_{i=1}^{n} X_i$$

$$= \frac{1}{n^2} V(\sum_{i=1}^{n} X_i)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} V(X_i)$$ since $X_i$’s are independent.

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$\therefore V\left(\bar{X}\right) = \frac{\sigma^2}{n} \qquad \text{i.e.} \quad \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

**Summary of the results:**

| Variable | Population Mean or Expected Value | Population Variance |
|---|---|---|
| ***X*** | $\mu$ | $\sigma^2$ |
| $\bar{X}$ (**sample mean**) | $\mu$ | $\dfrac{\sigma^2}{n}$ |

**Theorem 3. The distribution of $\bar{X}$**

(a) Let $X_1, X_2, X_3 \ldots X_n$ be a random sample from a $N(\mu, \sigma^2)$ population then

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

**(b)** Let $X_1$ , $X_2$ , ... , $X_n$ be a random sample from **ANY** population (with a finite mean $\mu$ and finite variance $\sigma^2$), then

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

(The Central Limit Theorem, see below.)

**Example:** Suppose the weight of an apple, $X \sim N(80,25)$. Let $X_1\ X_2\ X_3 \ldots X_{100}$ be the weights of a random sample of 100 apples. Find (i) $P(X > 85)$ and compare this with (ii) $P(\bar{X} > 85)$

**Solution:**

(i) $P(X > 85)$ where $X \sim N(80,25)$

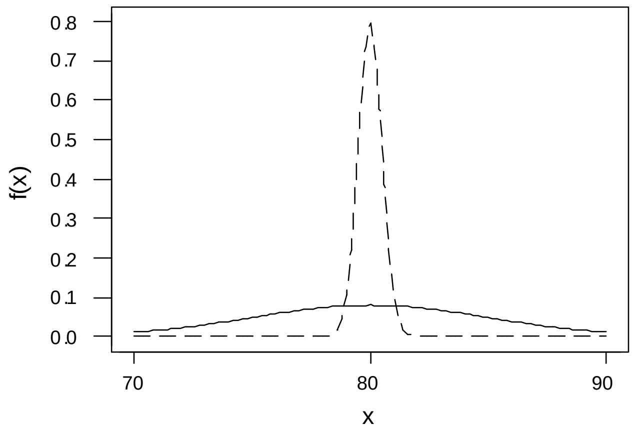
$$= P\left(Z > \frac{85-80}{\sqrt{25}}\right) = P(Z > 1) \quad \text{where } Z \sim N(0,1)$$

$$= 0.16 \quad \text{i.e. a } 16\% \text{ chance}$$

(ii) From Theorem 3, we have $\bar{X} \sim N\left(80\ ,\ \dfrac{25}{100}\right)$

$$P(\bar{X} > 85) = P\left(Z > \frac{85-80}{\sqrt{0.25}}\right) \quad \text{where } Z \sim N(0,1)$$

$$= P(Z > 10) = 0.0000\ldots.$$




### 11.3.4 Calculating intervals for $\bar{X}$

**Theorem 4.** Let $X_1$ , $X_2$ ….. $X_n$ be a random sample from a normal population of $N(\mu, \sigma^2)$. A $100(1-\alpha)\%$ interval for $\bar{X} = \left(\mu \pm z_{\alpha/2}\sigma_{\bar{X}}\right) = \left(\mu \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right)$

**Proof 4:** $\bar{X} \sim N\left(\mu\, , \,\sigma_{\bar{X}}^2\right) = N\left(\mu\, , \,\dfrac{\sigma^2}{n}\right)$ where $\sigma_{\bar{X}}^2 = \dfrac{\sigma^2}{n}$

$$\therefore Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \sim N(0\, , 1)$$

**Steps** 1. Find an interval for $Z$
2. Substitute for $Z$
3. Rearrange to give an interval for $\bar{X}$.

**Step** 1. $P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1 - \alpha$

**Step** 2 Substitute for Z

$$P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} < z_{\alpha/2}\right) = 1 - \alpha$$

**Step** 3. $P\left(-z_{\alpha/2}\,\sigma_{\bar{X}} < \bar{X} - \mu < z_{\alpha/2}\,\sigma_{\bar{X}}\right) = 1 - \alpha$

$$p\left(\boxed{\mu - z_{\alpha/2}\ \sigma_{\bar{X}}} < \bar{X} < \boxed{\mu - z_{\alpha/2}\ \sigma_{\bar{X}}}\right) = 1 - \alpha$$

A $100(1-\alpha)\%$ interval for $\bar{X}$ is given by

$$\left(\mu - z_{\alpha/2}\ \sigma_{\bar{X}}\ ,\ \mu + z_{\alpha/2}\ \sigma_{\bar{X}}\right)$$

$= \left(\mu \pm z_{\alpha/2}\ \sigma_{\bar{X}}\right)$ **where** $\sigma_{\bar{X}} = \sqrt{\sigma_{\bar{X}}^2} = \sqrt{\dfrac{\sigma^2}{n}} = \dfrac{\sigma}{\sqrt{n}}$

**Example:** A random sample of 100 apples was taken from a $X \sim N(80, 25)$ population distribution. Find a 95% interval for $\bar{X}$, the sample mean of 100 apples.

**Solution:**

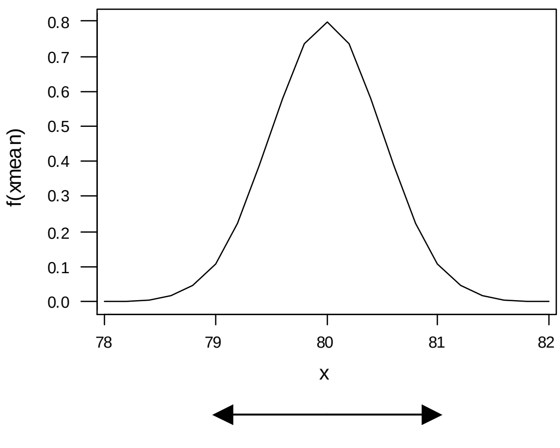
First find the distribution of $\bar{X}$

$$\bar{X} \sim N\left(\mu\ ,\ \frac{\sigma^2}{n}\right) = N\left(80\ ,\ \frac{25}{100}\right) = N\left(80\ ,\ 0.25\right)$$

Then find 95% CI for $\bar{X}$

$$\begin{aligned} A \text{ 95\% CI for } \bar{X} &= \left(\mu \pm z_{0.025}\ \sigma_{\bar{X}}\right) \\ &= \left(80 \pm z_{0.025}\ \sqrt{0.25}\right) \\ &= \left(80 \pm 1.96 * 0.5\right) \\ &= \left(80 \pm 0.98\right) \\ &= \left(79.02, 80.98\right) \\ &\approx \left(79.0, 81.0\right) \end{aligned}$$

**Conclusion:** There is a probability 0.95 (i.e. a 95% chance) that $\bar{X}$ (the sample mean weight of 100 apples) lies between 79.0 and 81.0 gm.




95% interval for $\bar{X}$

## 11.3.5 Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be a random sample from ANY distribution with finite mean $\mu$ and finite variance $\sigma^2$, then when n is large $\bar{X} \approx N\left(\mu\ ,\ \frac{\sigma^2}{n}\right)$ and $\sum_{i=1}^{n} X_i = n\bar{X} \approx N(n\mu,\ n\sigma^2)$

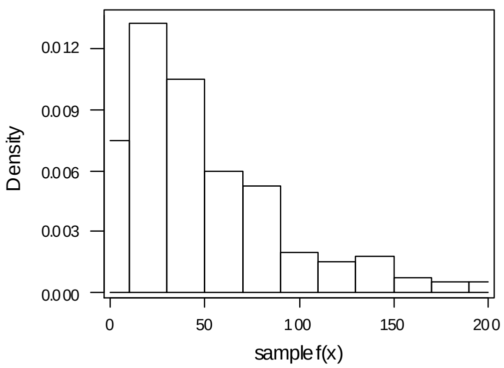
where $\approx$ means 'has the approximate distribution'.

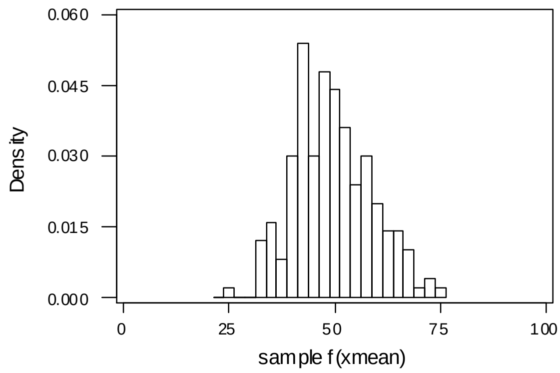
**Example:** Let $X_1, X_2, X_3, \ldots X_n$ be a random sample from an Exponential population where $E(X) = \mu$ and $V(X) = \mu^2$. By the Central Limit Theorem $\bar{X} \approx N\left(\mu\,,\, \frac{\mu^2}{n}\right)$

Consider $r$ samples of size $n$ from the above distribution

| | | $X_1$ $X_2$ $X_3$ $----$ $X_n \to \bar{X}$ |
|---|---|---|
| Sample | 1 | $X_{11}$ $X_{12}$ $X_{13}$ $----$ $X_{1n} \to \bar{X}_1$ |
| | 2 | $X_{21}$ $X_{22}$ $X_{23}$ $----$ $X_{2n} \to \bar{X}_2$ |
| | 3 | $X_{31}$ $X_{32}$ $X_{33}$ $----$ $X_{3n} \to \bar{X}_3$ |
| | . | |
| | . | |
| | $r$ | $X_{r1}$ $X_{r2}$ $X_{r3}$ $----$ $X_{rn} \to \bar{X}_r$ |
| | | $\uparrow$ $X_1$ $\qquad$ $\uparrow$ $\bar{X}$ |

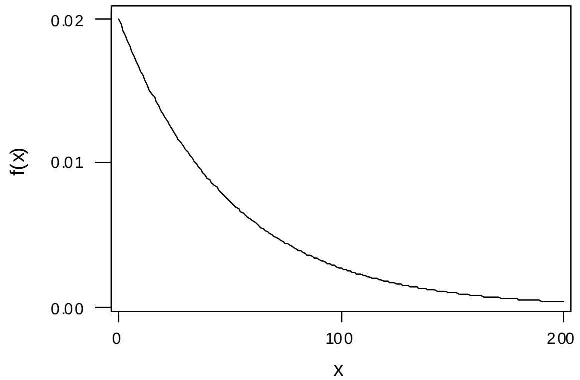
$X_1$ $\qquad\qquad$ $\bar{X}$



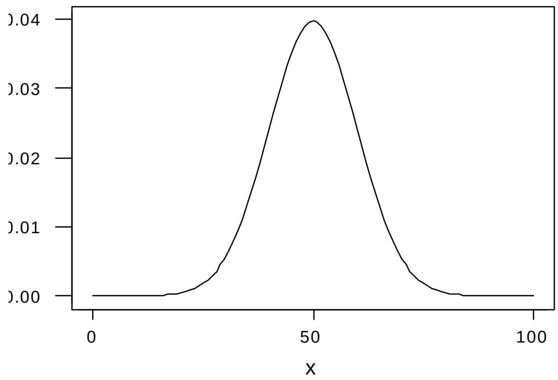




Sample pdf of $X$, $\hat{f}_X(x)$ $\qquad$ $n \longrightarrow \infty$ $\qquad$ Sample pdf of $\hat{f}_X(\bar{x})$

Population pdf of $X$, $f_X(x)$




Population pdf of $\bar{X}$, $f_{\bar{X}}(\bar{x})$




**Example:** Let $X_1$ , $X_2$ ... , $X_n$ be a random sample from a uniform $U$(a, b) distribution $\left[\text{where } E(X) = \mu = \frac{(a+b)}{2} \text{ and } V(X) = \sigma^2 = \frac{(b-a)^2}{12}\right]$.

For example let $X_1$ , $X_2$ ... , $X_n$ be a random sample from a uniform $U$(0, 100) distribution.

By the Central Limit Theorem, $\bar{X} \approx N\left(50 , \frac{100^2}{12n}\right)$.

# Practical 7

1. A fair die is rolled 40 times. Find the probability that the mean of the 40 scores is more than 4.
2. If 30 observations are taken from a population with distribution given by the pdf

$$f(x) = \begin{cases} \frac{2}{9}x, & 0 \le x \le 3 \\ 0 & otherwise \end{cases}$$

   a) Find the mean and variance of this population.
   b) Find the probability that the mean of the 30 observations is more than 1.85.

3. a) A population is known to have a standard deviation of 2.34 but unknown mean $\mu$. A random sample size of 81 provided a mean of 9.72. For the population mean construct:
   (i) a 90% confidence interval;
   (ii) a 98% confidence interval;
   (iii) comment on the precision of the results.

   b) A random sample of size 50 was drawn from a normal population of variance 0.20. If the sample produced a mean of 13.45, construct a 99% confidence interval for the mean of the population.

4. In a class of 25 pupils if each pupil rolls a fair die 30 times and records the number of sixes they obtain, find the probability that the mean number of sixes recorded for the class is less than 4.3

5. A sample of 100 tins filled by a machine has an average weight equal to 510 gm and a standard deviation of weight equal to 12 gm.
   a) Construct a 95% confidence interval for the mean
   b) Calculate the probability that the mean is less than 512 gm.

6. On the basis of a large sample a 95% confidence interval for the population mean is (55.8, 60.4). Using this information compute:
   a) estimates of $\mu$ and the standard error;
   b) a 90% confidence interval for $\mu$.

7. The heights of female employees of a certain company have a mean of 170 cm and a standard deviation of 2.5 cm while the heights of the male employees have a mean of 178 cm with a standard deviation of 2.1 cm. If random samples of 35 female and 45 male employees are taken and their heights recorded, find the probability that the mean of the male employees height is greater than the female's by more than 9 cm.