CHAPTER 6

RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

Definition: A random variable is a numerical description of the outcomes of the experiment or a numerical valued function defined on sample space, usually denoted by capital letters. **Example:** If X is a random variable, then it is a function from the elements of the sample space to the set of real numbers. i.e. X is a function X: S $\subseteq$ R

➔A random variable takes a possible outcome and assigns a number to it.

Example: Flip a coin three times, let X be the number of heads in three tosses.

$$\Rightarrow S = \{(HHH),(HHT),(HTH),(HTT),(THH),(THT),(TTH),(TTT)\}$$

$$\Rightarrow \quad X(HHH) = 3,$$

$$X(HHT) = X(HTH) = X(THH) = 2,$$

$$X(HTT) = X(THT) = X(TTH) = 1$$

$$X(TTT) = 0$$

X = {0, 1, 2, 3, 4, 5}

➔X assumes a specific number of values with some probabilities.

Random variables are of two types:

1. **Discrete random variable**: are variables which can assume only a specific number of values. They have values that can be counted

Examples:

- Toss coin n times and count the number of heads.
- Number of children in a family.
- Number of car accidents per week.
- Number of defective items in a given company.
- Number of bacteria per two cubic centimeter of water.

2. **Continuous random variable**: are variables that can assume all values between any two give values.

Examples:

- Height of students at certain college.
- Mark of a student.
- Life time of light bulbs.
- Length of time required to complete a given training.

Probability Distribution

Definition: a probability distribution consists of value that a random variable can assume and the corresponding probabilities of the values.

Example: Consider the experiment of tossing a coin three times. Let X is the number of heads. Construct the probability distribution of X.

Solution:

- First identify the possible value that X can assume.
- Calculate the probability of each possible distinct value of X and express X in the form of frequency distribution.

$X = x$	0	1	2	3
$P(X = x)$	1/8	3/8	3/8	1/8

- Probability distribution is denoted by **P** for discrete and by **f** for continuous random variable.

Properties of Probability Distribution:

1.

$$P(x) \geq 0, \quad if \quad X \quad is \ discrete.$$

$$f(x) \geq 0, \ if \ X \ is \ continuous.$$

2.

$$\sum_x P(X = x) = 1, \ if \ X \ is \ discrete.$$

$$\int_x f(x)dx = 1, \ if \ is \ continuous.$$

Note:

1. If X is a continuous random variable then

$$P(a < X < b) = \int_a^b f(x)dx$$

2. Probability of a fixed value of a continuous random variable is zero.

$$\Rightarrow P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$$

3. If X is discrete random variable then

$$P(a < X < b) = \sum_{x=a+1}^{b-1} P(x)$$

$$P(a \le X < b) = \sum_{x=a}^{b-1} p(x)$$

$$P(a < X \le b) = \sum_{x=a+1}^{b} P(x)$$

$$P(a \le X \le b) = \sum_{x=a}^{b} P(x)$$

4. Probability means area for continuous random variable.

Introduction to expectation

Definition:

1. Let a discrete random variable X assume the values X_1, X_2, ….,X_n with the probabilities $P(X_1)$, $P(X_2)$, ….,$P(X_n)$ respectively. Then the expected value of X, denoted as E(X) is defined as:

$$E(X) = X_1P(X_1) + X_2P(X_2) + ... + X_nP(X_n)$$
$$= \sum_{i=1}^{n} X_i P(X_i)$$

2. Let X be a continuous random variable assuming the values in the interval (a, b) such that $\int_a^b f(x)dx = 1$,then $E(X) = \int_a^b x\, f(x)dx$

Examples:

1. What is the expected value of a random variable X obtained by tossing a coin three times where X is the number of heads?

Solution:

First construct the probability distribution of X

$X = x$	0	1	2	3
$P(X = x)$	$1/8$	$3/8$	$3/8$	$1/8$

$$\Rightarrow E(X) = X_1P(X_1) + X_2P(X_2) + ... + X_nP(X_n)$$
$$= 0*1/8 + 1*3/8 + ... + 2*1/8$$
$$= 1.5$$

2. Suppose a charity organization is

mailing printed return-address stickers to over one million homes in Ethiopia. Each recipient is asked to donate either \$1, \$2, \$5, \$10, \$15, or \$20. Based on past experience, the amount a person donates is believed to follow the following probability distribution:

$X = x$	\$1	\$2	\$5	\$10	\$15	\$20
$P(X = x)$	0.1	0.2	0.3	0.2	0.15	0.05

What is expected that an average donor to contribute?

Solution:

$X = x$	\$1	\$2	\$5	\$10	\$15	\$20	Total
$P(X = x)$	0.1	0.2	0.3	0.2	0.15	0.05	1
$xP(X = x)$	0.1	0.4	1.5	2	2.25	1	7.25

$$\Rightarrow E(X) = \sum_{i=1}^{6} x_i P(X = x_i) = \$7.25$$

Mean and Variance of a random variable

Let X is given random variable.

1. The expected value of X is its mean

$$\Rightarrow Mean\,of\;\; X = E(X)$$

2. The variance of X is given by:

$$Variance\,of\;\; X = \mathrm{var}(X) = E(X^2) - [E(X)]^2$$

Where:

$$E(X^2) = \sum_{i=1}^{n} x_i^{\,2} P(X = x_i) \;\;,\quad if\;\; X \;\; is\;\; discrete$$

$$= \int_x x^2 f(x)dx \;,\qquad if\;\; X\; is\; continuous.$$

Examples:

1. Find the mean and the variance of a random variable X in example 2 above.

Solution:

$X = x$	\$1	\$2	\$5	\$10	\$15	\$20	Total
$P(X = x)$	0.1	0.2	0.3	0.2	0.15	0.05	1
$xP(X = x)$	0.1	0.4	1.5	2	2.25	1	7.25
$x^2P(X = x)$	0.1	0.8	7.5	20	33.75	20	82.15

$$\Rightarrow E(X) = 7.25$$

$$Var(X) = E(X^2) - [E(X)]^2 = 82.15 - 7.25^2 = 29.59$$

Exercise: Two dice are rolled. Let X is a random variable denoting the sum of the numbers on the two dice.

i) Give the probability distribution of X

ii) Compute the expected value of X and its variance

$\rightarrow$ There are some general rules for mathematical expectation.

Let X and Y are random variables and k is a constant.

RULE 1: $E(k) = k$

RULE 2: $Var(k) = 0$

RULE 3: $E(kX) = kE(X)$

RULE 4: $Var(kX) = k^2 Var(X)$

RULE 5: $E(X + Y) = E(X) + E(Y)$

Common Discrete Probability Distributions

1. **<u>Binomial Distribution</u>**

A binomial experiment is a probability experiment that satisfies the following four requirements called assumptions of a binomial distribution.

1. The experiment consists of n identical trials.
2. Each trial has only one of the two possible mutually exclusive outcomes, success or a failure.
3. The probability of each outcome does not change from trial to trial, and
4. The trials are independent, thus we must sample with replacement.

Examples of binomial experiments

- Tossing a coin 20 times to see how many tails occur.
- Asking 200 people if they watch BBC news.
- Registering a newly produced product as defective or non defective.

- Asking 100 people if they favor the ruling party.
- Rolling a die to see if a 5 appears.

Definition: The outcomes of the binomial experiment and the corresponding probabilities of these outcomes are called **Binomial Distribution**.

$$Let\ P = the\ probability\ of\ success$$

$$q = 1 - p = the\ probability\ of\ \ failure\ on\ any\ given\ trial$$

Then the probability of getting x successes in n trials becomes:

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0,1,2,....,n$$

And this is some times written as: $X \sim Bin(n, p)$

When using the binomial formula to solve problems, we have to identify three things:

- The number of trials (n)
- The probability of a success on any one trial (p) and
- The number of successes desired (X).

Examples:

1. What is the probability of getting three heads by tossing a fair con four times?

Solution: Let X be the number of heads in tossing a fair coin four times

$X \sim Bin(n = 4, p = 0.50)$

$$\Rightarrow P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0,1,2,3,4$$

$$= \binom{4}{x} 0.5^x 0.5^{4-x}$$

$$= \binom{4}{x} 0.5^4$$

$$\Rightarrow P(X = 3) = \binom{4}{3} 0.5^4 = 0.25$$

2. Suppose that an examination consists of six true and false questions, and assume that a student has no knowledge of the subject matter. The probability that the student will guess the correct answer to the first question is 30%. Likewise, the probability of guessing each of the remaining questions correctly is also 30%.

a) What is the probability of getting more than three correct answers?
b) What is the probability of getting at least two correct answers?
c) What is the probability of getting at most three correct answers?
d) What is the probability of getting less than five correct answers?

Solution: Let X = the number of correct answers that the student gets.

$$X \sim Bin(n = 6, p = 0.30)$$

a) $P(X > 3) = ?$

$$\Rightarrow P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0,1,2,..6$$

$$= \binom{6}{x} 0.3^x 0.7^{6-x}$$

$$\Rightarrow P(X > 3) = P(X = 4) + P(X = 5) + P(X = 6)$$
$$= 0.060 + 0.010 + 0.001$$
$$= 0.071$$

Thus, we may conclude that if 30% of the exam questions are answered by guessing, the probability is 0.071 (or 7.1%) that more than four of the questions are answered correctly by the student.

b) $P(X \geq 2) = ?$

$$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6)$$
$$= 0.324 + 0.185 + 0.060 + 0.010 + 0.001$$
$$= 0.58$$

c) $P(X \leq 3) = ?$

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$
$$= 0.118 + 0.303 + 0.324 + 0.185$$
$$= 0.93$$

d) $P(X < 5) = ?$

$$P(X < 5) = 1 - P(X \geq 5)$$
$$= 1 - \{P(X = 5) + P(X = 6)\}$$
$$= 1 - (0.010 + 0.001)$$
$$= 0.989$$

Exercises:

a. Suppose that 4% of all TVs made by A&B Company in 2000 are defective. If eight of these TVs are randomly selected from across the country and tested, what is the probability that *exactly* three of them are defective? Assume that each TV is made independently of the others.

b. An allergist claims that 45% of the patients she tests are allergic to some type of weed. What is the probability that

I. Exactly 3 of her next 4 patients are allergic to weeds?

II. None of her next 4 patients are allergic to weeds?

c. Explain why the following experiments are not Binomial

I. Rolling a die until a 6 appears.

II. Asking 20 people how old they are.

III. Drawing 5 cards from a deck for a poker hand.

Remark: If X is a binomial random variable with parameters n and p then

$$\boxed{E(X) = np \quad , \qquad Var(X) = npq}$$

2. Poisson Distribution

A random variable X is said to have a Poisson distribution if its probability distribution is given by:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0,1,2,......$$

$$Where \;\; \lambda = the \;\; average \, number.$$

The Poisson distribution depends *only* on the average number of occurrences per unit time of space.

The Poisson distribution is used as a distribution of rare events, such as: Arrivals, Accidents, Number of misprints, Hereditary, Natural disasters like earth quake, etc.

The process that gives rise to such events is called Poisson process.

Example: If 1.6 accidents can be expected an intersection on any given day, what is the probability that there will be 3 accidents on any given day?

Solution: Let X =the number of accidents, $\lambda = 1.6$

$$X = poisson(1.6) \Rightarrow p(X = x) = \frac{1.6^x e^{-1.6}}{x!}$$

$$p(X = 3) = \frac{1.6^3 e^{-1.6}}{3!} = 0.1380$$

Exercise: On the average, five smokers pass a certain street corners every ten minutes, what is the probability that during a given 10 minutes the number of smokers passing will be

a. 6 or fewer
b. 7 or more
c. Exactly 8…….

If X is a Poisson random variable with parameter λ then

$$\boxed{E(X)=\lambda \quad , \qquad Var(X)=\lambda}$$

Note: The Poisson probability distribution provides a close approximation to the binomial probability distribution when n is large and p is quite small or quite large with $\lambda = np$.

$$P(X=x)=\frac{(np)^x e^{-(np)}}{x!}, \quad x=0,1,2,......$$

$$Where \ \ \lambda = np = the \ \ average\, number.$$

Usually we use this approximation if $np \le 5$. In other words, if $n > 20$ and $np \le 5$ [or $n(1-p) \le 5$], then we may use Poisson distribution as an approximation to binomial distribution.

Example: Find the binomial probability P(X=3) by using the Poisson distribution if $p = 0.01$ and $n = 200$. **Solution:**

$$Using\ Poisson \ \ ,\lambda = np = 0.01*200 = 2$$

$$\Rightarrow P(X=3)=\frac{2^3 e^{-2}}{3!}=0.1804$$

$$Using \ \ Binomial,\ n=200, \ \ p=0.01$$

$$\Rightarrow P(X=3)=\binom{200}{3}(0.01)^3(0.99)^{99}=0.1814$$

Common Continuous Probability Distributions

1. **<u>Normal Distribution</u>**

A random variable X is said to have a normal distribution if its probability density function is

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad -\infty < x < \infty, \qquad -\infty < \mu < \infty, \quad \sigma > 0$$

$$Where \ \ \mu = E(X), \quad \sigma^2 = Variance(X)$$

$$\mu \ \ and \quad \sigma^2 \ \ are\, the \ \ Parameters\, of\ the\, Normal \ \ Distribution.$$

Properties of Normal Distribution:

1. It is bell shaped and is symmetrical about its mean and it is mesokurtic. The maximum ordinate is at $x = \mu$ and is given by $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}}$
2. It is asymptotic to the axis, i.e., it extends indefinitely in either direction from the mean.
3. It is a continuous distribution.
4. It is a family of curves, i.e., every unique pair of mean and standard deviation defines a different normal distribution. Thus, the normal distribution is completely described by two parameters: mean and standard deviation.
5. Total area under the curve sums to 1, i.e., the area of the distribution on each side of the mean is 0.5. $\Rightarrow \int_{-\infty}^{\infty} f(x)dx = 1$
6. It is unimodal, i.e., values mound up only in the center of the curve.
7. $Mean = Median = \text{mod}\, e = \mu$
8. The probability that a random variable will have a value between any two points is equal to the area under the curve between those points.

Note: To facilitate the use of normal distribution, the following distribution known as the standard normal distribution was derived by using the transformation

$$Z = \frac{X - \mu}{\sigma} \qquad \Rightarrow f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Properties of the Standard Normal Distribution:

- Same as a normal distribution, but also mean is zero, variance is one, standard Deviation is one
- Areas under the standard normal distribution curve have been tabulated in various ways. The most common ones are the areas between $Z = 0$ *and a positive value of* Z.
- Given normal distributed random variable X with mean μ *and standard deviation* σ

$$P(a < X < b) = P(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma})$$

$$\Rightarrow P(a < X < b) = P(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma})$$

Note:

$$P(a < X < b) = P(a \le X < b)$$
$$= P(a < X \le b)$$
$$= P(a \le X \le b)$$

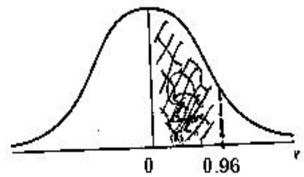
Examples:

1. Find the area under the standard normal distribution which lies

a) Between $Z = 0 \;\; and \;\; Z = 0.96$

Solution:

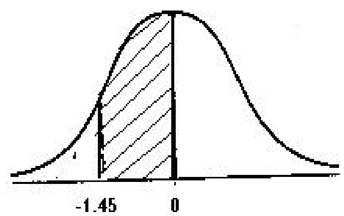
$$Area = P(0 < Z < 0.96) = 0.3315$$



b) Between $Z = -1.45 \;\; and \;\; Z = 0$

Solution:

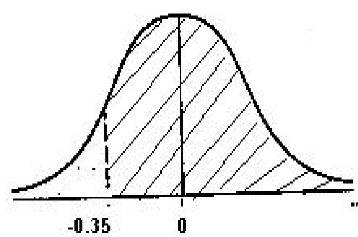
$$Area = P(-1.45 < Z < 0)$$
$$= P(0 < Z < 1.45)$$
$$= 0.4265$$



c) To the right of $Z = -0.35$

Solution:

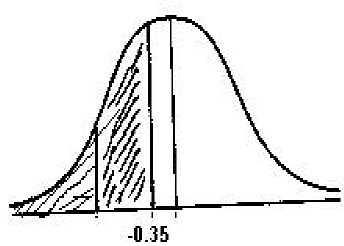
$$Area = P(Z > -0.35)$$
$$= P(-0.35 < Z < 0) + P(Z > 0)$$
$$= P(0 < Z < 0.35) + P(Z > 0)$$
$$= 0.1368 + 0.50 = 0.6368$$



d) To the left of $Z = -0.35$

Solution:

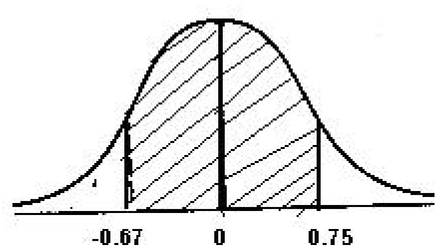
$$Area = P(Z < -0.35)$$
$$= 1 - P(Z > -0.35)$$
$$= 1 - 0.6368 = 0.3632$$



e) Between

$Z = -0.67 \;\; and \; Z = 0.75$

Solution:

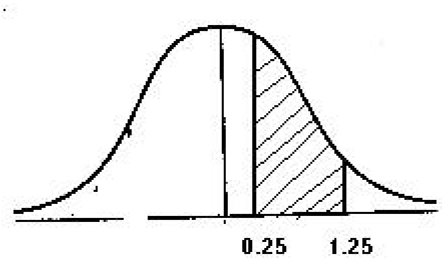


$$
\begin{aligned}
Area &= P(-0.67 < Z < 0.75) \\
&= P(-0.67 < Z < 0) + P(0 < Z < 0.75) \\
&= P(0 < Z < 0.67) + P(0 < Z < 0.75) \\
&= 0.2486 + 0.2734 = 0.5220
\end{aligned}
$$

f) Between $Z = 0.25 \;\; and \; Z = 1.25$

Solution:

$$
\begin{aligned}
Area &= P(0.25 < Z < 1.25) \\
&= P(0 < Z < 1.25) - P(0 < Z < 0.25) \\
&= 0.3934 - 0.0987 = 0.2957
\end{aligned}
$$

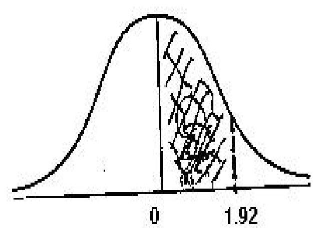


2. Find the value of Z if

a) The normal curve area between 0 and z(positive) is 0.4726

Solution

$$
\begin{aligned}
&P(0 < Z < z) = 0.4726 \;\; and \;\; from \;\; table \\
&P(0 < Z < 1.92) = 0.4726 \\
&\Leftrightarrow z = 1.92.....uniqueness \; of \;\; Areea.
\end{aligned}
$$



b) The area to the left of z is 0.9868

Solution

$$
\begin{aligned}
P(Z < z) &= 0.9868 \\
&= P(Z < 0) + P(0 < Z < z) \\
&= 0.50 + P(0 < Z < z) \\
\Rightarrow P(0 < Z < z) &= 0.9868 - 0.50 = 0.4868 \\
&and \;\; from \quad table \\
P(0 < Z < 2.2) &= 0.4868 \\
&\Leftrightarrow z = 2.2
\end{aligned}
$$

3. A random variable X has a normal distribution with mean 80 and standard deviation 4.8. What is the probability that it will take a value

a) Less than 87.2

b) Greater than 76.4

c) Between 81.2 and 86.0

Solution

$$X \text{ is normal with mean}, \mu = 80, \text{ standard deviation}, \sigma = 4.8$$

a)

$$P(X < 87.2) = P(\frac{X-\mu}{\sigma} < \frac{87.2-\mu}{\sigma})$$
$$= P(Z < \frac{87.2-80}{4.8})$$
$$= P(Z < 1.5)$$
$$= P(Z < 0) + P(0 < Z < 1.5)$$
$$= 0.50 + 0.4332 = \underline{\underline{0.9332}}$$

b)

$$P(X > 76.4) = P(\frac{X-\mu}{\sigma} > \frac{76.4-\mu}{\sigma})$$
$$= P(Z > \frac{76.4-80}{4.8})$$
$$= P(Z > -0.75)$$
$$= P(Z > 0) + P(0 < Z < 0.75)$$
$$= 0.50 + 0.2734 = \underline{\underline{0.7734}}$$

c)

$$P(81.2 < X < 86.0) = P(\frac{81.2-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{86.0-\mu}{\sigma})$$
$$= P(\frac{81.2-80}{4.8} < Z < \frac{86.0-80}{4.8})$$
$$= P(0.25 < Z < 1.25)$$
$$= P(0 < Z < 1.25) - P(0 < Z < 1.25)$$
$$= 0.3934 - 0.0987 = \underline{\underline{0.2957}}$$

4. A normal distribution has mean 62.4.Find its standard deviation if 20.0% of the area under the normal curve lies to the right of 72.9

Solution

$$P(X > 72.9) = 0.20 \Rightarrow P(\frac{X-\mu}{\sigma} > \frac{72.9-\mu}{\sigma}) = 0.20$$

$$\Rightarrow P(Z > \frac{72.9-62.4}{\sigma}) = 0.20$$

$$\Rightarrow P(Z > \frac{10.5}{\sigma}) = 0.20$$

$$\Rightarrow P(0 < Z < \frac{10.5}{\sigma}) = 0.50 - 0.20 = 0.30$$

$$And \ \ from \ table \ \ P(0 < Z < 0.84) = 0.30$$

$$\Leftrightarrow \frac{10.5}{\sigma} = 0.84$$

$$\Rightarrow \sigma = 12.5$$

5. A random variable has a normal distribution with $\sigma = 5$ .Find its mean if the probability that the random variable will assume a value less than 52.5 is 0.6915.

Solution

$$P(Z < z) = P(Z < \frac{52.5-\mu}{5}) = 0.6915$$

$$\Rightarrow P(0 < Z < z) = 0.6915 - 0.50 = 0.1915.$$

$$But \ from \ the \, table$$

$$\Rightarrow P(0 < Z < 0.5) = 0.1915$$

$$\Leftrightarrow z = \frac{52.5-\mu}{5} = 0.5$$

$$\Rightarrow \mu = \underline{\underline{50}}$$

Exercise: Of a large group of men, 5% are less than 60 inches in height and 40% are between 60 & 65 inches. Assuming a normal distribution, find the mean and standard deviation of heights.