Lecture 4

Probability Distributions

Many decisions in business, insurance, and other real-life situations are made by assigning probabilities to all possible outcomes pertaining to the situation and then evaluating the results.

For example, a saleswoman can compute the probability that she will make

0, 1, 2, or 3 or more sales in a single day. An insurance company might be able to assign probabilities to the number of vehicles a family owns.

A self-employed speaker might be able to compute the probabilities for giving 0, 1, 2, 3, or 4 or more speeches each week.

Once these probabilities are assigned, statistics such as the mean, variance, and standard deviation can be computed for these events. With these statistics, various decisions can be made. The saleswoman will be able to compute the average number of sales she makes per week, and if she is working on commission, she will be able to approximate her weekly income over a period of time, say, monthly.

The public speaker will be able to plan ahead and approximate his average income and expenses. The insurance company can use its information to design special computer forms and programs to accommodate its customers' future needs.

We will explain the concepts and applications of what is called a *probability distribution.*

Before probability distribution is defined formally, the definition of a variable must be reviewed.

a *variable* was defined as a characteristic or attribute that can assume different values. Various letters of the alphabet, such as *X, Y,* or *Z,* are used to represent variables.

Since the variables in this lesson are associated with probability, they are called *random variables.*

For example, if a die is rolled, a letter such as *X* can be used to represent the outcomes. Then the value that *X* can assume is 1, 2, 3, 4, 5, or 6, corresponding to the outcomes of rolling a single die.

If two coins are tossed, a letter, say *Y*, can be used to represent the number of heads, in this case 0, 1, or 2.

As another example, if the temperature at 8:00 A.M. is 43 and at noon it is 53, then the values *T* that the temperature assumes are said to be random, since they are due to various atmospheric conditions at the time the temperature was taken.

A random variable is a variable whose values are determined by chance.

Also we can classify variables as discrete or continuous

by observing the values the variable can assume. If a variable can assume only a specific number of values, such as the outcomes for the roll of a die or the outcomes for the toss of a coin, then the variable is called a ***discrete variable***.

Discrete variables **have a finite number of possible values or an infinite number of values that can be counted**

The word *counted* means that they can be enumerated using

the numbers 1, 2, 3, etc. For example, the number of boys in a class and the number of phone calls received after a TV commercial airs are examples of discrete variables, since they can be counted.

Variables that can assume all values in the interval between any two given values are called ***continuous variables.*** For example, if the temperature goes from 62 to 78 in a 24-hour period, it has passed through every possible number from 62 to 78. ***Continuous random variables are obtained from data that can be measured rather than counted.***

Continuous random variables can assume an infinite number of values and can be decimal and fractional values. On a continuous scale, a person's weight might be exactly 183.426 pounds if a scale could measure weight to the thousandths place; however, on a digital scale that measures only to tenths of pounds, the weight would be 183.4 pounds. Examples of continuous variables are heights, weights, temperatures, and time. In this lesson only discrete random variables are used.

The procedure shown here for constructing a probability distribution for a discrete random variable uses the probability experiment of tossing three coins.

Recall that when three coins are tossed, the sample space is represented as TTT, TTH, THT, HTT, HHT, HTH, THH, HHH; and if X is the random variable for the number of heads,

then X

assumes the value 0, 1, 2, or 3.

Probabilities for the values of X can be determined as follows:

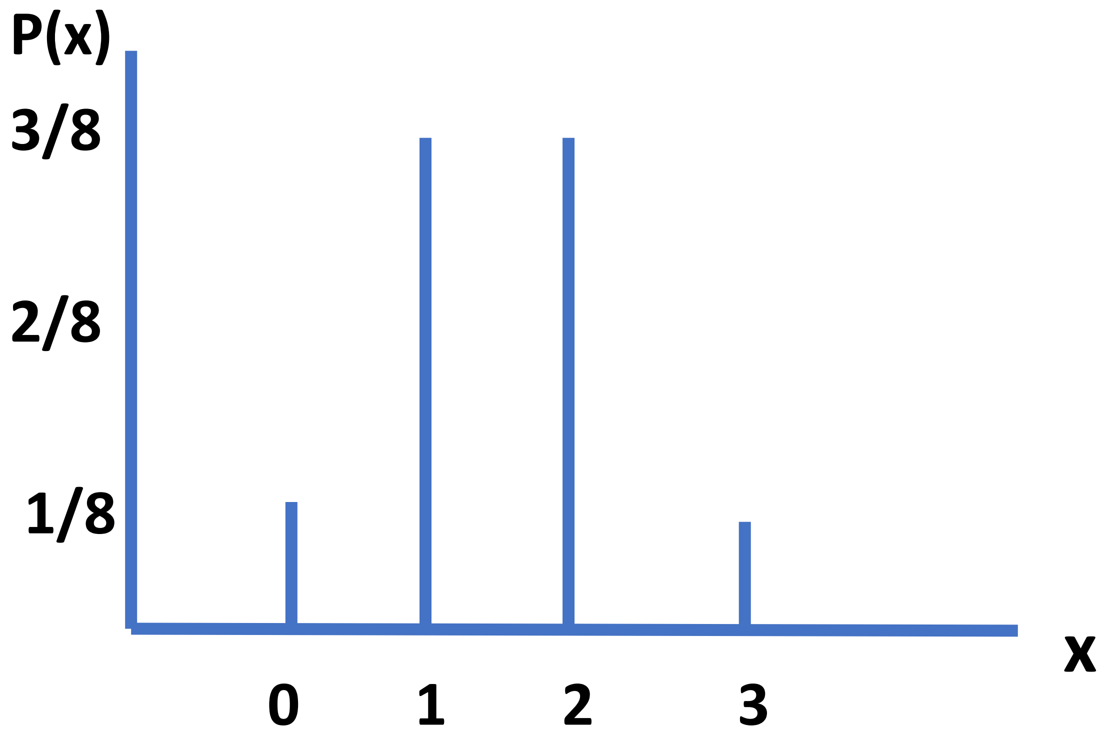
No heads	**One head**	**Two heads**	**Three heads**
TTT	**TTH THT HTT**	**HHT HTH THH**	**HHH**
1/8	**1/8**	**1/8**	**1/8**
P=(1/8)	**p=(3/8)**	**p= (3/8)**	**p= (1/8)**

Hence, the probability of getting no heads is=1/8 , one head is=3/8 , two heads is=3/8 , and three heads is=1/8 .

From these values, a probability distribution can be constructed by listing the outcomes and assigning the probability of each outcome, as shown here.

Number of heads X	Probability $P(X)$
0	1/8
1	3/8
2	3/8
3	1/8

The graph



A discrete probability distribution consists of the values a random variable can assume and the corresponding probabilities of the values. The probabilities are determined theoretically or by observation.

Example 1: Rolling a Die

Construct a probability distribution for rolling a single die.

Solution

Since the sample space is 1, 2, 3, 4, 5, 6 and each outcome has a probability of , the distribution is as shown.

Outcome X	Probability $P(X)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Example 2:

The baseball World Series is played by the winner of the National League and the American League. The first team to win four games wins the World Series.

In other words, the series will consist of four to seven games, depending on the individual victories. The data shown consist of 40 World Series events.

The number of games played in each series is represented by the variable X.

Find the probability $P(X)$ for each X, construct a probability distribution, and draw a graph for the data

Solution

The probability *P*(*X*) can be computed for each *X* by dividing the number of games *X* by the total

For 4 games,= 8/40 = 0.200

For 5 games = 7/40 = 0.175

For 6 games =9/40 = 0.225

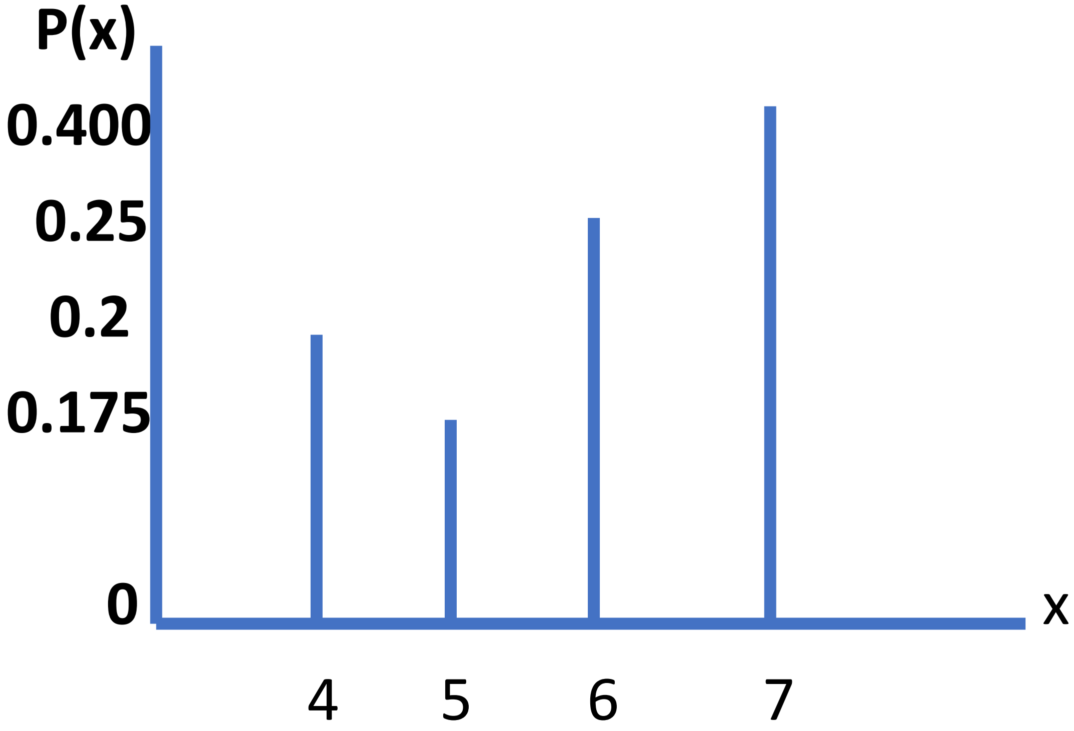
For 5 games = 16/40= 0.400

X	Number of games played
4	8
5	7
6	9
7	16
total	40

The probability distribution is:

Number of games *X*	4	5	6	7
Probability *P*(*X*)	0.200	0.175	0.225	0.400

The graph



Mean, Variance, Standard Deviation, and Expectation

The mean, variance, and standard deviation for a probability distribution are computed differently from the mean, variance, and standard deviation for samples.

This section explains how these measures—as well as a new measure called the *expectation*—are calculated for probability distributions.

Mean:

the mean for a sample or population was computed by adding the values and dividing by the total number of values, as shown in these formulas:

$$\bar{x} = \frac{\sum x}{n} \qquad \mu = \frac{\sum x}{N}$$

But how would you compute the mean of the number of spots that show on top when a die is rolled? You could try rolling the die, say, 10 times, recording the number of spots, and finding the mean; however, this answer would only approximate the true mean. What about 50 rolls or 100 rolls? Actually, the more times the die is rolled, the better the approximation.

You might ask, then, How many times must the die be rolled to get the exact answer? It must be rolled an infinite number of times. Since this task is impossible, the previous formulas cannot be used because the denominators would be infinity.

Hence, a new method of computing the mean is necessary. This method gives the exact theoretical value of the mean as if it were possible to roll the die an infinite number of times.

Before the formula is stated, an example will be used to explain the concept. Suppose two coins are tossed repeatedly, and the number of heads that occurred is recorded.

What will be the mean of the number of heads? The sample space is

HH, HT, TH, TT

and each outcome has a probability of 1/4 . Now, in the long run, you would *expect* two heads (HH) to occur approximately 1/4of the time, one head to occur approximately 1/2 of the time (HT or TH), and no heads (TT) to occur approximately 1/4 of the time. Hence, on average, you would expect the number of heads to be

$$1/4.2+ 1/2.1+1/4.0=1$$

That is, if it were possible to toss the coins many times or an infinite number of times,

the *average* of the number of heads would be 1.

Hence, to find the mean for a probability distribution, you must multiply each possible outcome by its corresponding probability and find the sum of the products.

Formula for the Mean of a Probability Distribution

The mean of a random variable with a discrete probability distribution is

$$\mu = X1 \,.\, P(X1) + X2 \,.\, P(X2) + X3 \,.\, P(X3) \ldots\ldots + \quad Xn \,.\, P(Xn)$$

$$\underline{\mu = \Sigma X \,.\, P(X)}$$

where $X1, X2, X3, \ldots, Xn$ are the outcomes and $P(X1), P(X2), P(X3), \ldots, P(Xn)$ are the

corresponding probabilities.

Note that $\Sigma X \,.\, P(X)$ means to sum the products.

Example3

Find the mean of the number of spots that appear when a die is tossed.

Solution

In the toss of a die, the mean can be computed thus.

Outcome X	1	2	3	4	5	6
Probability $P(X)$	1/6	1/6	1/6	1/6	1/6	1/6

$\mu = \Sigma X \,.\, P(X)=$

$1.1/6+2.1/6+3.1/6+4.1/6+5.1/6+6.1/6= 21/6 = 3.5$

Example 4:

In a family with two children, find the mean of the number of children who will be girls.

Solution

The probability distribution is as follows:

Number of girls *X*	0	1	2
Probability p(x)	1/4	1/2	1/4

Hence, the mean is

$\mu = \Sigma X \,.\, P(X) =$

0.1/4 +1.1/2+2.1/4= 1

Variance and Standard Deviation

For a probability distribution, the mean of the random variable describes the measure of the so-called long-run or theoretical average, but it does not tell anything about the spread of the distribution. to measure this spread or variability, statisticians use the variance and standard deviation.

Formula for the Variance of a Probability Distribution

Find the variance of a probability distribution by multiplying the square of each outcome by its corresponding probability, summing those products, and subtracting the square of the mean. The formula for the variance of a probability distribution is

$$\sigma^2 = \left(\Sigma x^2 \cdot p(x)\right) - \mu^2$$

The standard deviation of a probability distribution is

$$\sigma = \sqrt{\Sigma x^2 \cdot p_{(x)} - \mu^2}$$

Example 5 Rolling a Die

Compute the variance and standard deviation of spots that appear when a die is tossed.

.

Outcome X	1	2	3	4	5	6
Probability $P(X)$	1/6	1/6	1/6	1/6	1/6	1/6

Solution

Recall that the mean μ= 3.5, as computed in Example 3. Square each outcome and multiply by the corresponding probability, sum those products, and then subtract the square of the mean.

$$\sigma^2 = \left(\Sigma x^2 \cdot p(x)\right) - \mu^2$$

σ^2 = (1^2.1/6+ 2^2.1/6 +3^2.1/6+ 4^2.1/6 +5^2.1/6 +6^2.1/ 6 – $(3.5)^2$ = 2.9

To get the standard deviation, find the square root of the variance.

$$\sigma = \sqrt{2.9} = 1.7$$