

**Applications of Probabilistic
Models in day
to day life
by
Dr. Javaid Ahmad Reshi**

Probability Distributions

A **random variable** is a **numerical** measure of the outcome from a probability experiment, so its value is determined by chance. Random variables are denoted using letters such as X .

A **discrete random variable** is a random variable that has values that has either a finite number of possible values or a countable number of possible values.

A **continuous random variable** is a random variable that has an infinite number of possible values that is not countable.

EXAMPLE *Distinguishing Between Discrete and Continuous Random Variables*

Determine whether the following random variables are discrete or continuous. State possible values for the random variable.

- (a) The number of light bulbs that burn out in a room of 10 light bulbs in the next year.
- (b) The number of leaves on a randomly selected Oak tree.
- (c) The length of time between calls to 911.
- (d) A single die is cast. The number of pips showing on the die.

- We use capital letter , like X , to denote the random variable and use small letter to list the possible values of the random variable.
- Example. A single die is cast, X represent the number of pips showing on the die and the possible values of X are $x=1,2,3,4,5,6$.

A **probability distribution** provides the possible values of the random variable and **their corresponding probabilities**. A probability distribution can be in the form of a **table**, **graph** or mathematical **formula**.

The table below shows the probability distribution for the random variable X , where X represents the number of DVDs a person rents from a video store during a single visit.

x	$P(X = x)$
0	0.06
1	0.58
2	0.22
3	0.10
4	0.03
5	0.01

Requirements for a Discrete Probability Distribution

Let $P(X = x)$ denote the probability the random variable X equals x , then

$$\sum P(X = x) = 1 \quad \text{and}$$

$$0 \leq P(X = x) \leq 1.$$

EXAMPLE *Identifying Probability Distributions*

Is the following a probability distribution?

x	$P(X = x)$
0	0.16
1	0.18
2	0.22
3	0.10
4	0.3
5	0.01

EXAMPLE *Identifying Probability Distributions*

Is the following a probability distribution?

x	$P(X = x)$
0	0.16
1	0.18
2	0.22
3	0.10
4	0.3
5	0.01

Answer:

$$**0.16 + 0.18 + 0.22 + 0.10 + 0.3 + 0.01 = 0.97**$$

<1 , Not a probability distribution.

EXAMPLE *Identifying Probability Distributions*

Is the following a probability distribution?

x	$P(X = x)$
0	0.16
1	0.18
2	0.22
3	0.10
4	0.3
5	0.04

Answer:

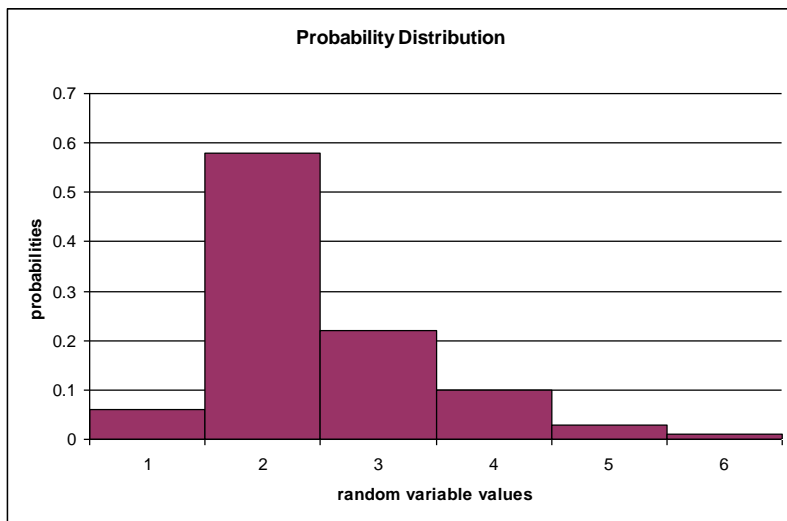
- **$0.16 + 0.18 + 0.22 + 0.10 + 0.3 + 0.04 = 1$**
- **It is a probability distribution**

A **probability histogram** is a histogram in which the horizontal axis corresponds to the value of the random variable and the vertical axis represents the probability of that value of the random variable.

EXAMPLE*Drawing a Probability Histogram*

Draw a probability histogram of the following probability distribution which represents the number of DVDs a person rents from a video store during a single visit.

x	$P(X = x)$
0	0.06
1	0.58
2	0.22
3	0.10
4	0.03
5	0.01



x	prob
0	0.06
1	0.58
2	0.22
3	0.1
4	0.03
5	0.01

The Mean of a Discrete Random Variable

The mean of a discrete random variable is given by the formula

$$\mu_x = \sum [x \cdot P(X = x)]$$

where x is the value of the random variable and $P(X = x)$ is the probability of observing the random variable x .

Expected Value

The **expected value** of a discrete random variable X , denoted $E(X)$, is obtained using the formula

$$E(X) = \sum [x \cdot P(X = x)]$$

EXAMPLE *The Mean of a Discrete Random Variable*

Compute the mean of the following probability distribution which represents the number of DVDs a person rents from a video store during a single visit.

x	$P(X = x)$
0	0.06
1	0.58
2	0.22
3	0.10
4	0.03
5	0.01

$$\begin{aligned}\text{Mean} &= 0 * 0.06 + 1 * 0.58 + 2 \\ &\quad * 0.22 + 3 * 0.1 + 4 \\ &\quad * 0.03 + 5 * 0.01 \\ &= \mathbf{1.49}\end{aligned}$$

Interpretation of the Mean of a Discrete Random Variable

Suppose an experiment is repeated n independent times and the value of the random variable X is recorded. As the number of repetitions of the experiment, n , increases the arithmetic mean value of the n trials will approach μ_X , the mean of the random variable X . In other words, let x_1 be the value of the random variable X after the first experiment, x_2 will be the value of the random variable X after the second experiment, and so on. Then,

$$\bar{X} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

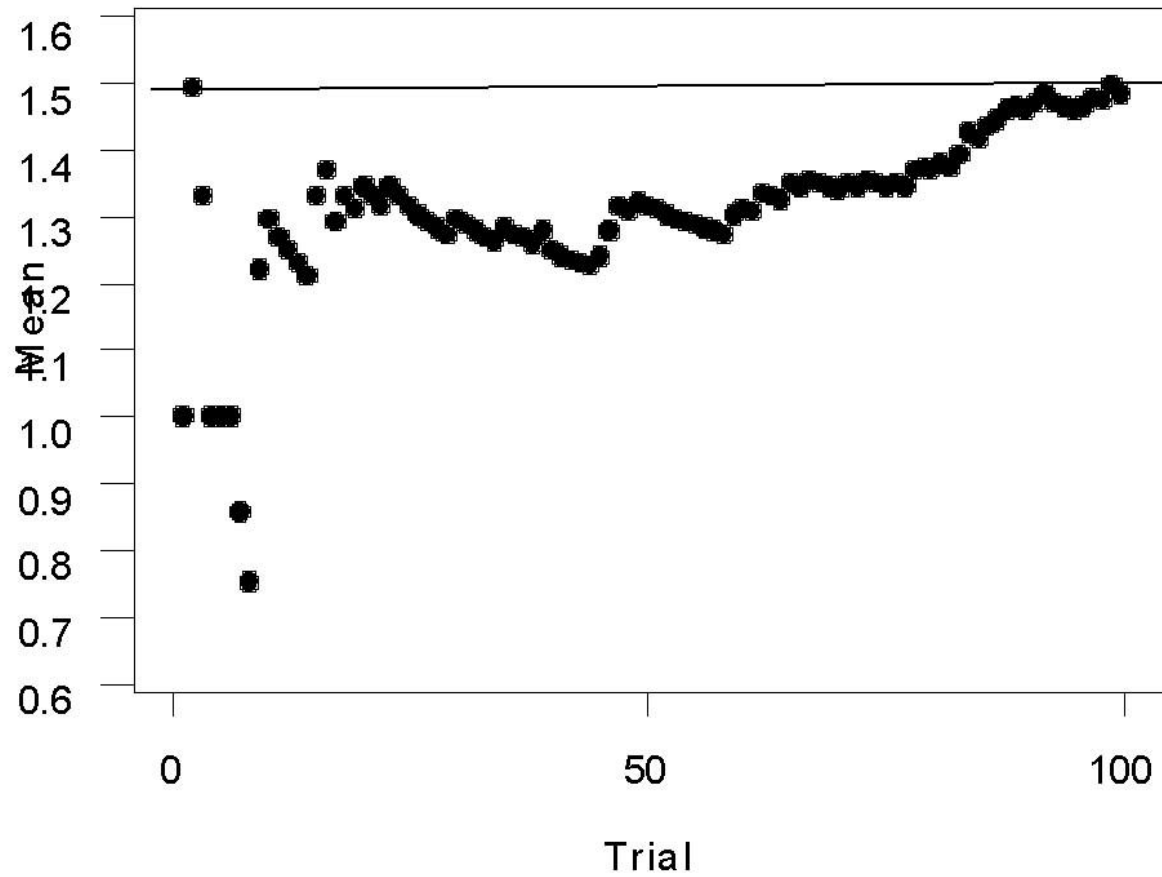
The difference between \bar{X} and μ_X will get closer to 0 as n increases.

The following data represent the number of DVDs rented by 100 randomly selected customers in a single visit. Compute the mean number of DVDs rented.

1	1	1	1	1	1	1	2	2	2
2	1	1	1	1	1	3	1	1	3
1	1	2	1	1	1	1	2	3	0
0	1	1	1	1	1	1	1	4	1
1	3	1	2	2	1	3	1	1	1
1	2	1	1	3	1	1	2	3	2
0	0	1	1	3	1	2	1	2	3
0	2	1	1	1	1	1	3	3	1
5	1	1	2	2	3	1	2	2	4
2	2	2	0	1	2	1	1	1	0

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_{100}}{100} = 1.49$$

Demonstrating the Law of Large Numbers



Variance and Standard Deviation of a Discrete Random Variable

The variance of a discrete random variable is given by

$$\sigma_X^2 = \sum \left[(x - \mu_X)^2 \cdot P(X = x) \right]$$

To find the standard deviation of the discrete random variable, take the square root of the variance. That is,

$$\sigma_X = \sqrt{\sigma_X^2}$$

EXAMPLE*Variance and Std*

Compute the variance and standard deviation of the following probability distribution which represents the number of DVDs a person rents from a video store during a single visit.

x	$P(X = x)$
0	0.06
1	0.58
2	0.22
3	0.10
4	0.03
5	0.01

- The variance= $(0-1.49)^2*0.06+(1-1.49)^2*0.58$
 $+(2-1.49)^2*0.22+(3-1.49)^2*0.1$
 $+(4-1.49)^2*0.03+(5-1.49)^2*0.01$
 $=0.8699$
- Standard Deviation= 0.932684

EXAMPLE

Expected Value

A term life insurance policy will pay a beneficiary a certain sum of money upon the death of the policy holder. These policies have premiums that must be paid annually. Suppose a life insurance company sells a \$250,000 one year term life insurance policy to a 49-year-old female for \$520. According to the National Vital Statistics Report, Vol. 47, No. 28, the probability the female will survive the year is 0.99791. Compute the expected value of this policy to the insurance company.

- Correct:

$$\begin{aligned} E(X) &= (520-250,000)*(1- \\ &0.99791)+520*0.99791 \\ &= \mathbf{-2.5} \end{aligned}$$

- Wrong:

$$\begin{aligned} E(X) &= -250,000*(1-0.99791)+520*0.99791 \\ &= \mathbf{-3.5868} \end{aligned}$$

The Binomial Probability Distribution

Criteria for a Binomial Probability Experiment

An experiment is said to be a **binomial experiment** provided

1. The experiment is performed a fixed number of times. Each repetition of the experiment is called a **trial**.
2. The trials are independent. This means the outcome of one trial will not affect the outcome of the other trials.
3. For each trial, there are two mutually exclusive outcomes, success or failure.
4. The probability of success is fixed for each trial of the experiment.

Notation Used in the Binomial Probability Distribution

- There are n independent trials of the experiment
- Let p denote the probability of success so that $1 - p$ is the probability of failure.
- Let x denote the number of successes in n independent trials of the experiment. So, $0 \leq x \leq n$.

EXAMPLE *Identifying Binomial Experiments*

Which of the following are binomial experiments?

(a) A player rolls a pair of fair die 10 times. The number X of 7's rolled is recorded.

(b) The 11 largest airlines had an on-time percentage of 84.7% in November, 2001 according to the Air Travel Consumer Report. In order to assess reasons for delays, an official with the FAA randomly selects flights until she finds 10 that were not on time. The number of flights X that need to be selected is recorded.

(c) In a class of 30 students, 55% are female. The instructor randomly selects 4 students. The number X of females selected is recorded.

EXAMPLE *Constructing a Binomial Probability Distribution*

According to the Air Travel Consumer Report, the 11 largest air carriers had an on-time percentage of 84.7% in November, 2001.

Suppose that 4 flights are randomly selected from November, 2001 and the number of on-time flights X is recorded. Construct a probability distribution for the random variable X using a tree diagram.

- $X=(x_1,x_2,x_3,x_4)$
- X =the number of on-time

Binomial Probability Distribution Function

The probability of obtaining x successes in n independent trials of a binomial experiment where the probability of success is p is given by

$$P(X = x) = {}_n C_x p^x (1 - p)^{n-x} \quad x = 0, 1, \dots, n$$

where p = probability of success

Phrase	<i>Math Symbol</i>
“at least”	\geq
“more than” or “greater than”	$>$
“fewer than” or “less than”	$<$
“no more than”	\leq
“exactly”	$=$

EXAMPLE *Using the Binomial Probability Distribution Function*

According to the United States Census Bureau, 18.3% of all households have 3 or more cars.

- (a) In a random sample of 20 households, what is the probability that exactly 5 have 3 or more cars?
- (b) In a random sample of 20 households, what is the probability that less than 4 have 3 or more cars?
- (c) In a random sample of 20 households, what is the probability that at least 4 have 3 or more cars?

Mean and Standard Deviation of a Binomial Random Variable

A binomial experiment with n independent trials and probability of success p will have a mean and standard deviation given by the formulas

$$\mu = np \quad \text{and} \quad \sigma = \sqrt{np(1-p)}$$

EXAMPLE *Finding the Mean and Standard Deviation of a Binomial Random Variable*

According to the United States Census Bureau, 18.3% of all households have 3 or more cars. In a simple random sample of 400 households, determine the mean and standard deviation number of households that will have 3 or more cars.

EXAMPLE *Constructing Binomial Probability Histograms*

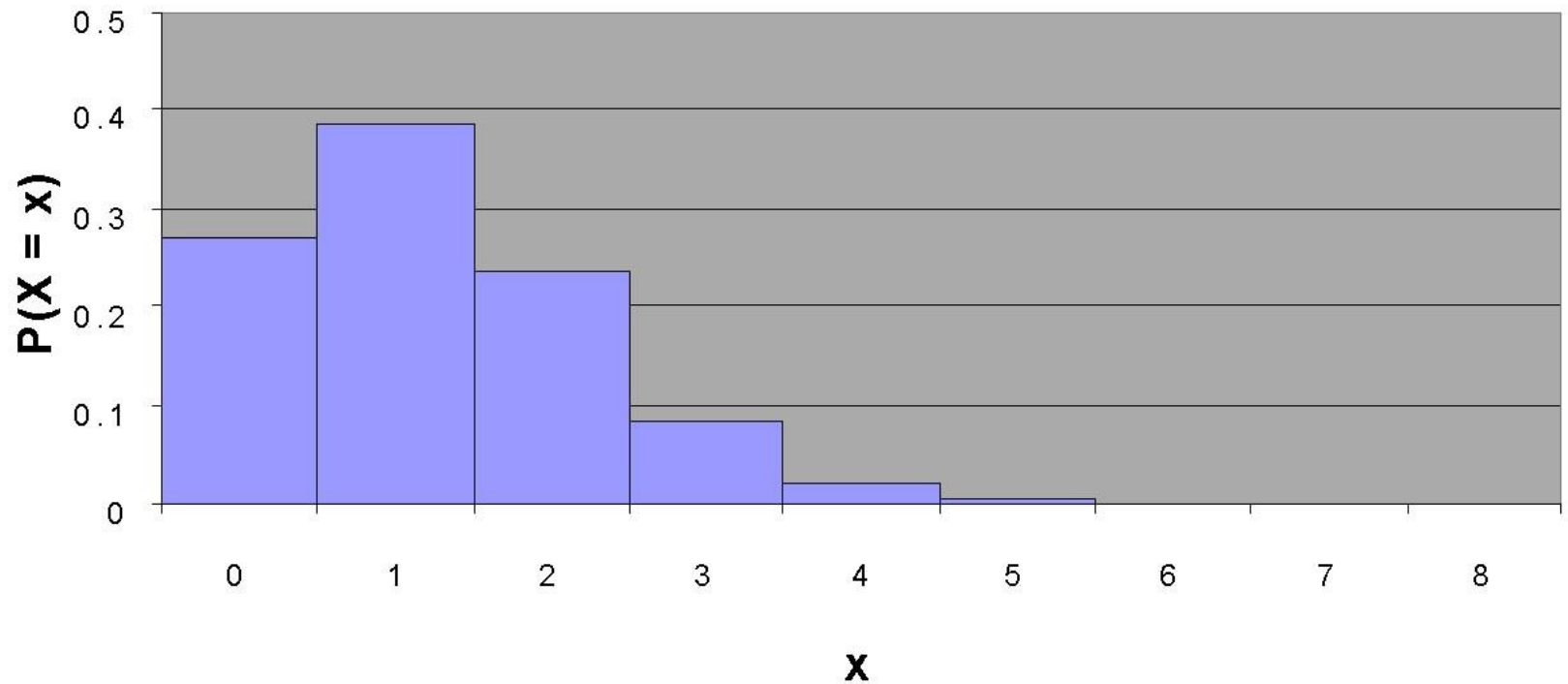
(a) Construct a binomial probability histogram with $n = 8$ and $p = 0.15$.

(b) Construct a binomial probability histogram with $n = 8$ and $p = 0.5$.

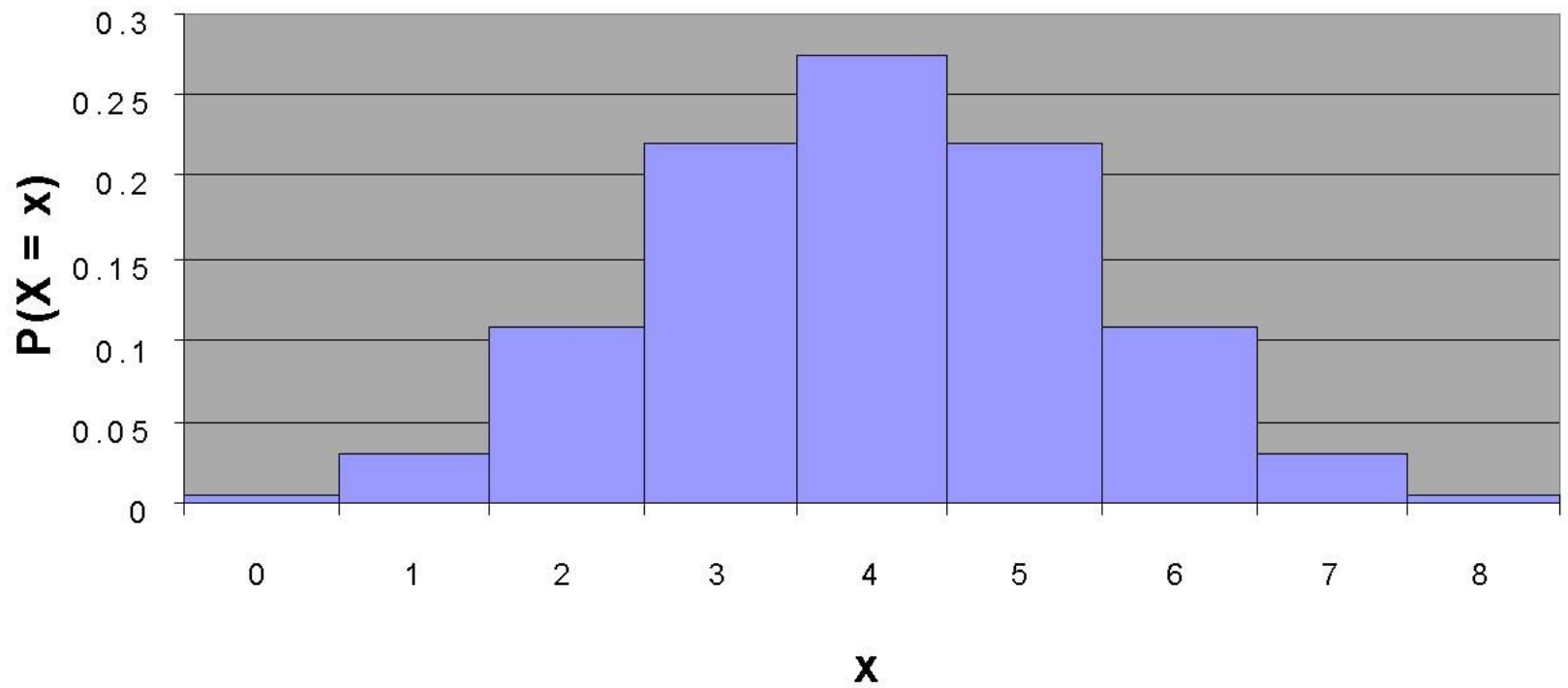
(c) Construct a binomial probability histogram with $n = 8$ and $p = 0.85$.

For each histogram, comment on the shape of the distribution.

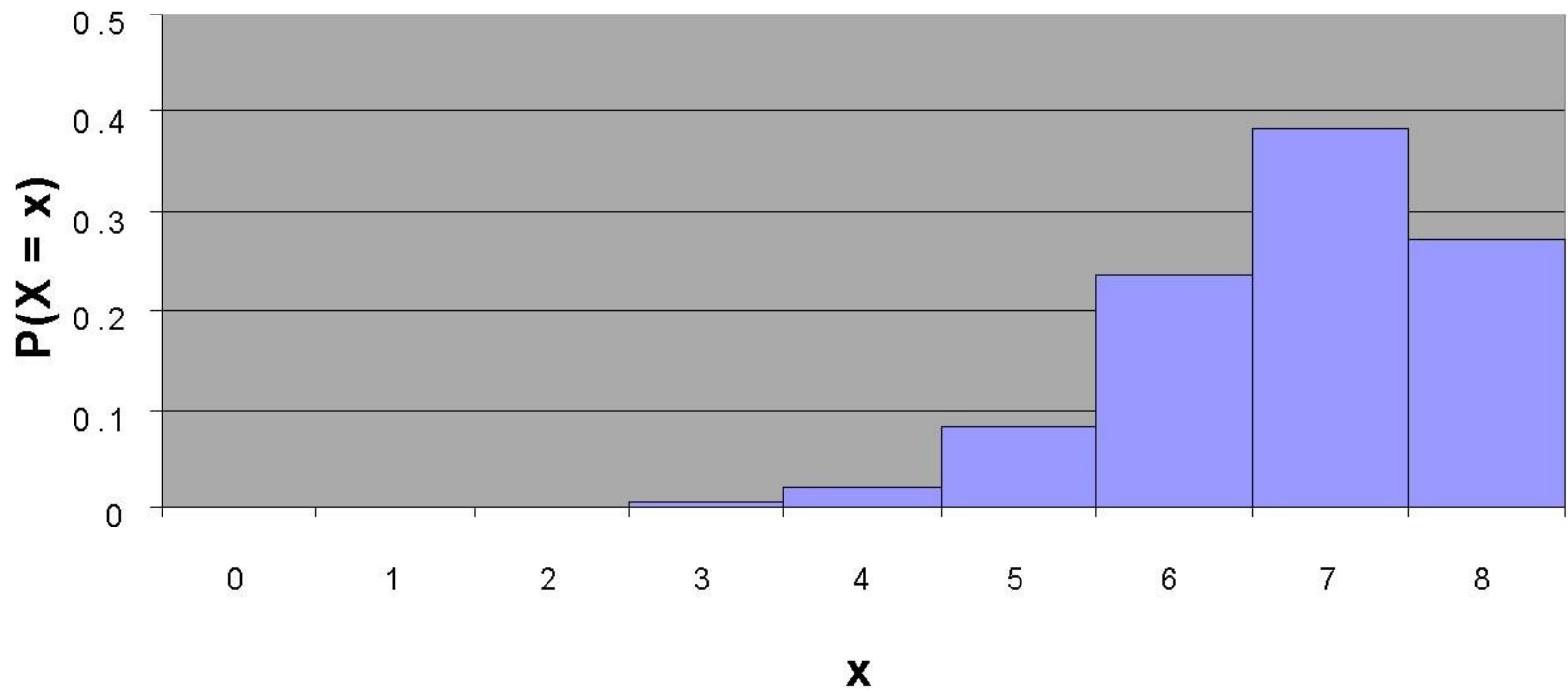
Binomial Histogram, $n = 8$, $p = 0.15$



Binomial Histogram, $n = 8, p = 0.5$

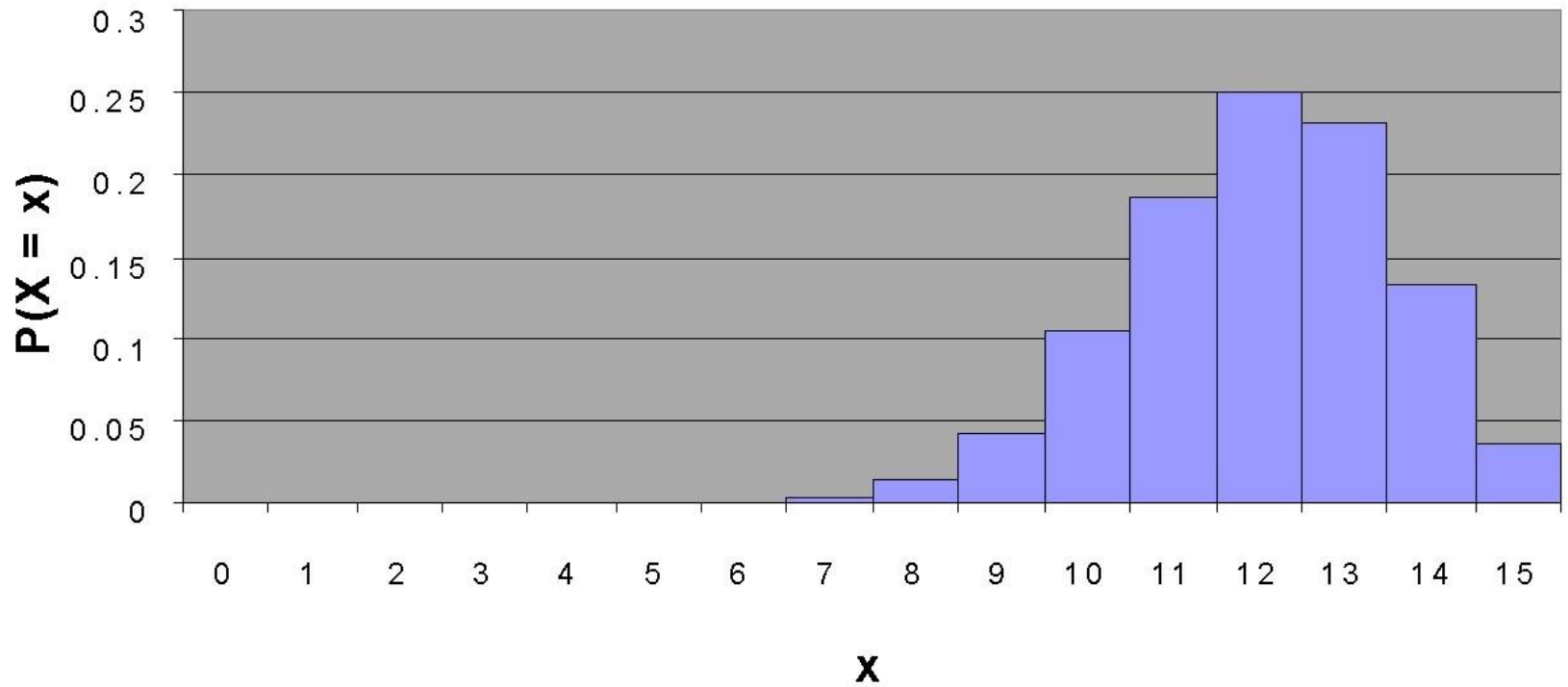


Binomial Histogram, $n = 8$, $p = 0.85$



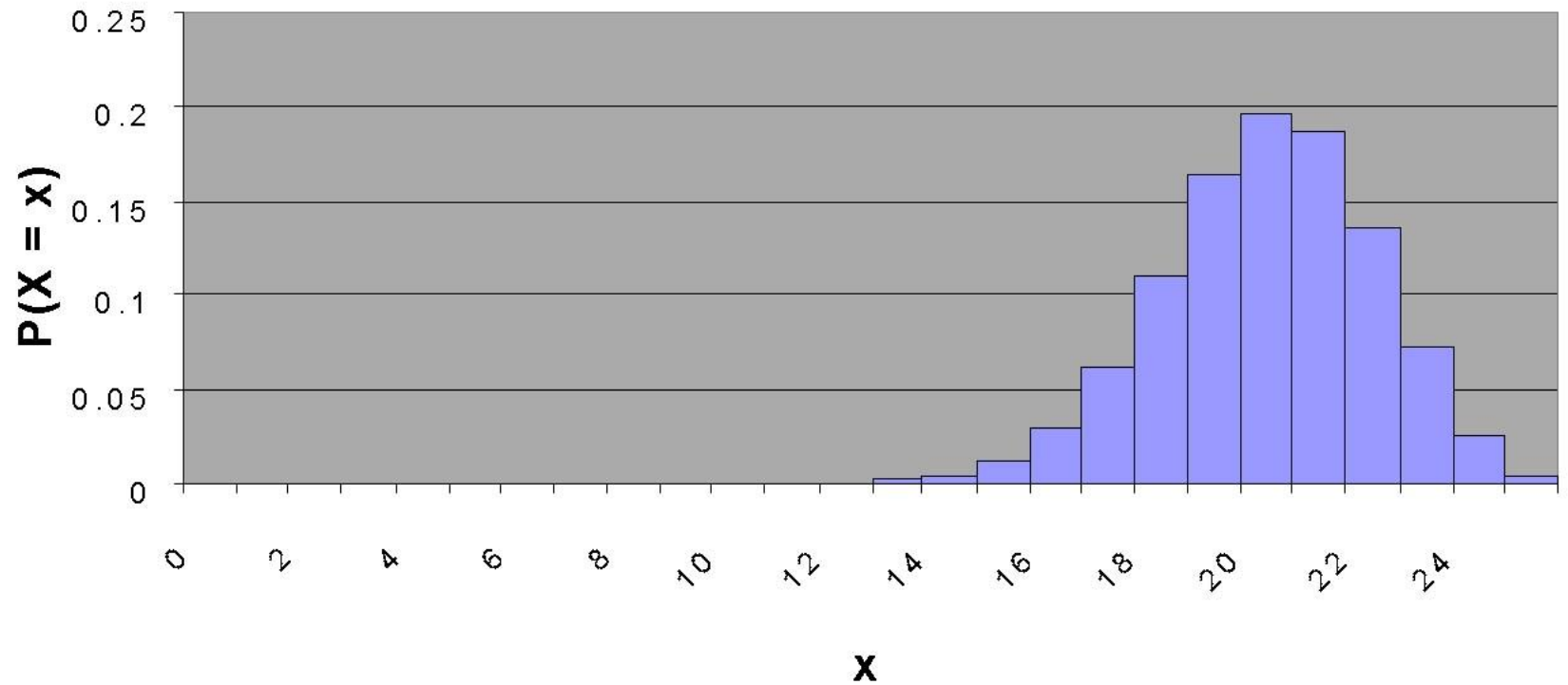
Construct a binomial probability histogram with $n = 15$ and $p = 0.8$. Comment on the shape of the distribution.

Binomial Histogram, $n = 15$, $p = 0.8$



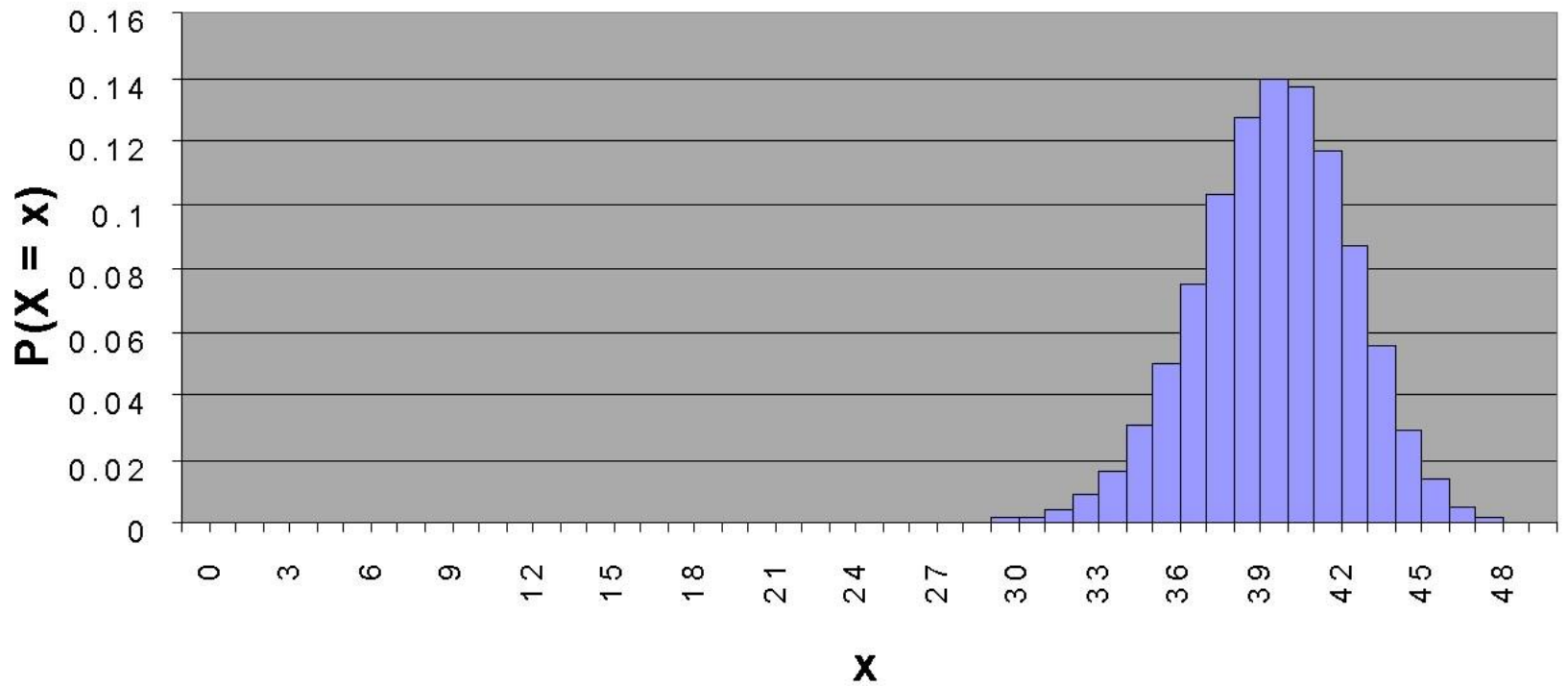
Construct a binomial probability histogram with $n = 25$ and $p = 0.8$. Comment on the shape of the distribution.

Binomial Histogram, $n = 25$, $p = 0.8$



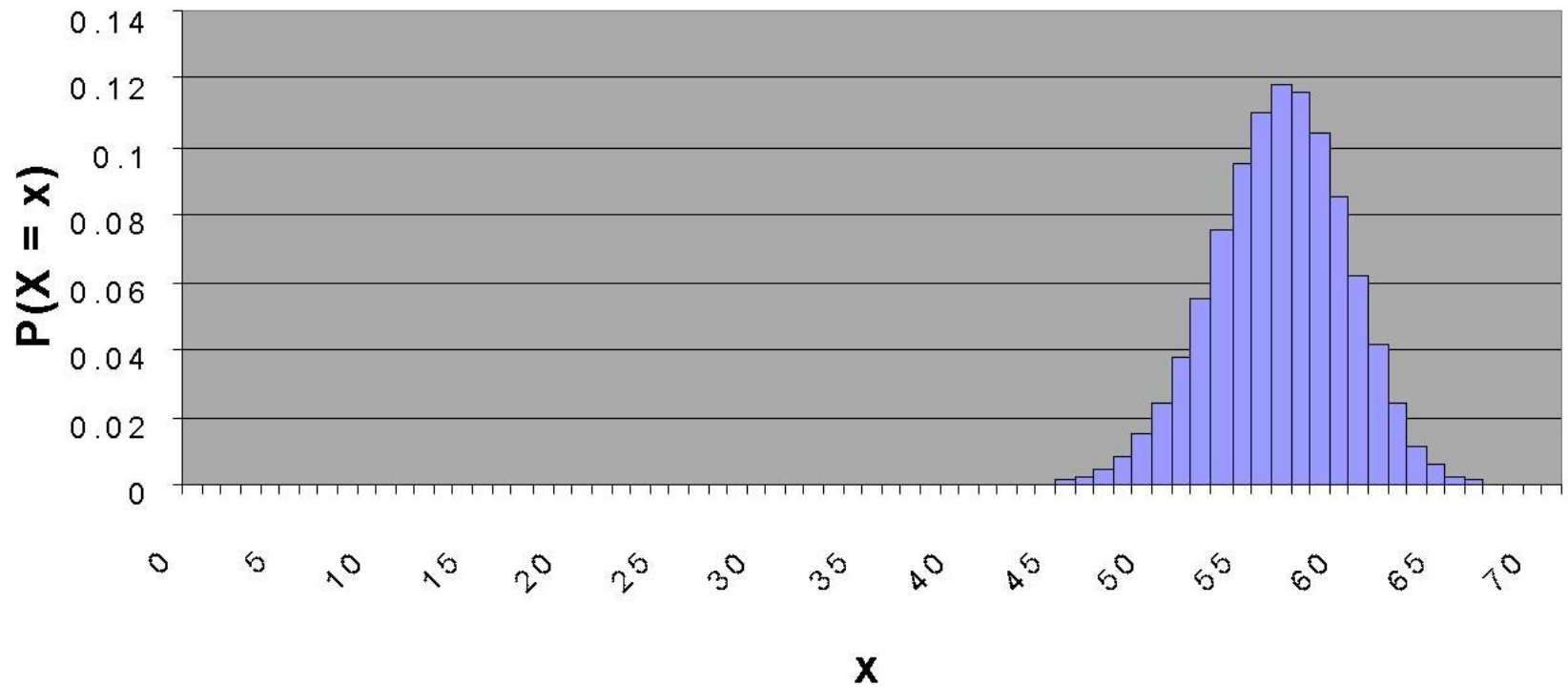
Construct a binomial probability histogram with $n = 50$ and $p = 0.8$. Comment on the shape of the distribution.

Binomial Histogram, $n = 50$, $p = 0.8$



Construct a binomial probability histogram with $n = 70$ and $p = 0.8$. Comment on the shape of the distribution.

Binomial Histogram, $n = 70$, $p = 0.8$



As the number of trials n in a binomial experiment increase, the probability distribution of the random variable X becomes bell-shaped. As a general rule of thumb, if $np(1 - p) \geq 10$, then the probability distribution will be approximately bell-shaped.

This result allows us to use the Empirical Rule to identify unusual observations in a binomial experiment. Recall, the Empirical Rule states that in a bell-shaped distribution, about 95% of all observations lie within 2 standard deviations of the mean. That is, 95% of the observations lie between $\mu - 2\sigma$ and $\mu + 2\sigma$. Any observation that lies outside these intervals may be considered unusual because the observation occurs only 5% of the time.

EXAMPLE *Using the Mean, Standard Deviation and Empirical Rule to Check for Unusual Results in a Binomial Experiment*

According to the United States Census Bureau, in 2000, 18.3% of all households have 3 or more cars. A researcher believes this percentage has increased since then. He conducts a simple random sample of 400 households and found that 82 households had 3 or more cars. Is this result unusual if the percentage of households with 3 or more cars is still 18.3%?

EXAMPLE *Using the Binomial Probability Distribution Function to Perform Inference*

According to the United States Census Bureau, in 2000, 18.3% of all households have 3 or more cars. A researcher believes this percentage has increased since then. He conducts a simple random sample of 20 households and found that 5 households had 3 or more cars.

Is this result unusual if the percentage of households with 3 or more cars is still 18.3%?

EXAMPLE *Using the Binomial Probability Distribution Function to Perform Inference*

According to the United States Census Bureau, in 2000, 18.3% of all households have 3 or more cars. One year later, the same researcher conducts a simple random sample of 20 households and found that 8 households had 3 or more cars.

Is this result unusual if the percentage of households with 3 or more cars is still 18.3%?

The Poisson Probability Distribution

A random variable X , the number of successes in a fixed interval, follows a **Poisson process** provided the following conditions are met

1. The probability of two or more successes in any sufficiently small subinterval is 0.
2. The probability of success is the same for any two intervals of equal length.
3. The number of successes in any interval is independent of the number of successes in any other interval provided the intervals are not overlapping.

EXAMPLE

A Poisson Process

The Food and Drug Administration sets a Food Defect Action Level (FDAL) for various foreign substances that inevitably end up in the food we eat and liquids we drink. For example, the FDAL level for insect filth in chocolate is 0.6 insect fragments (larvae, eggs, body parts, and so on) per 1 gram.

- For a sufficiently small interval, the probability of two successes is 0.
- The probability of insect filth in one region of a candy bar is equal to the probability of insect filth in some other region of the candy bar.
- The number of successes in any random sample is independent of the number of successes in any other random sample.

Mean and Standard Deviation of a Binomial Random Variable

A binomial experiment with n independent trials and probability of success p will have a mean and standard deviation given by the formulas

$$\mu = np \quad \text{and} \quad \sigma = \sqrt{np(1-p)}$$

EXAMPLE

Computing Poisson Probabilities

The Food and Drug Administration sets a Food Defect Action Level (FDAL) for various foreign substances that inevitably end up in the food we eat and liquids we drink. For example, the FDAL level for insect filth in chocolate is 0.6 insect fragments (larvae, eggs, body parts, and so on) per 1 gram.

- (a) Determine the mean number of insect fragments in a 5 gram sample of chocolate.
- (b) What is the standard deviation?

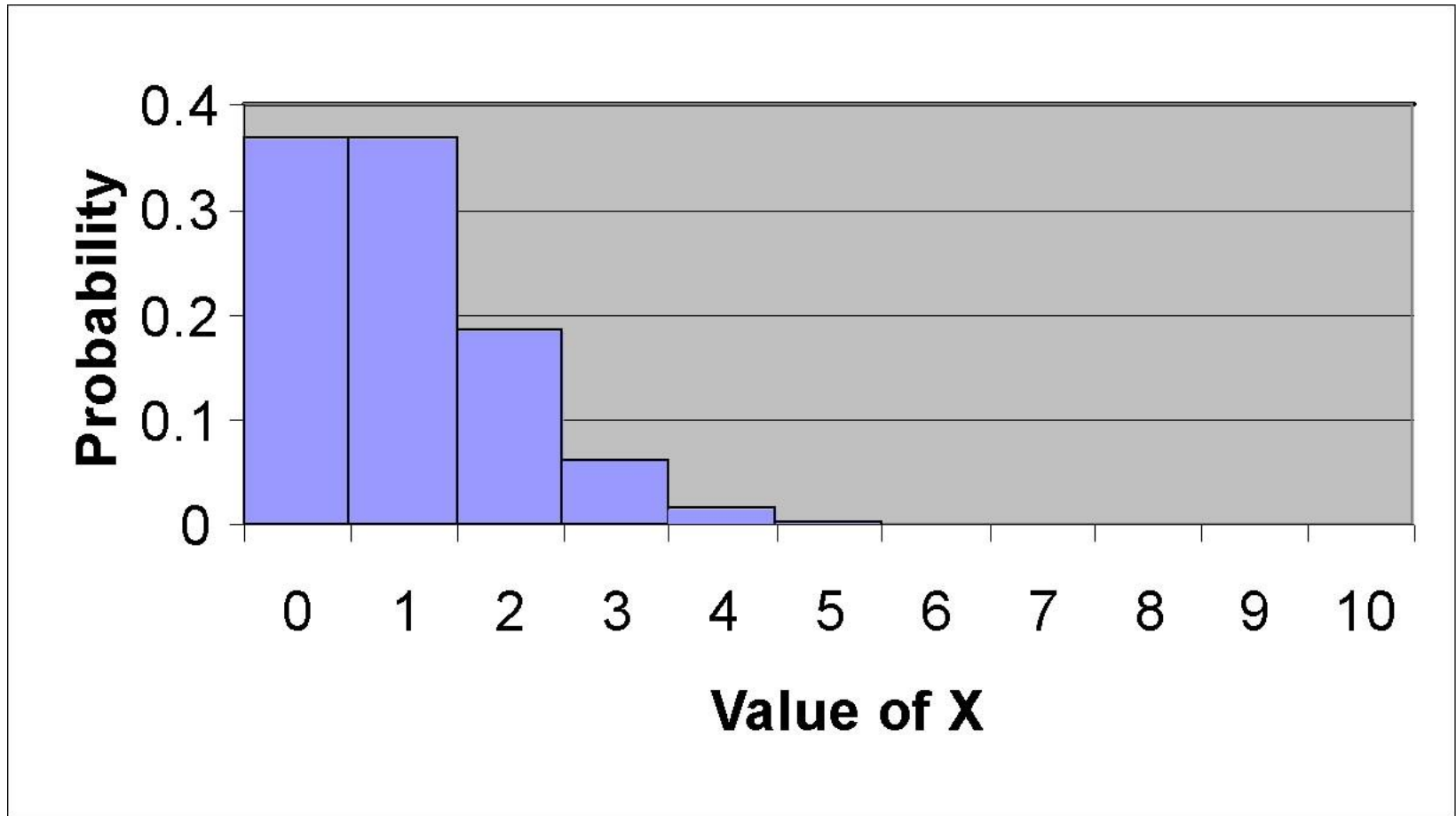
Poisson Probability Distribution Function

If X is the number of successes in an interval of fixed length t , then the probability formula for X is

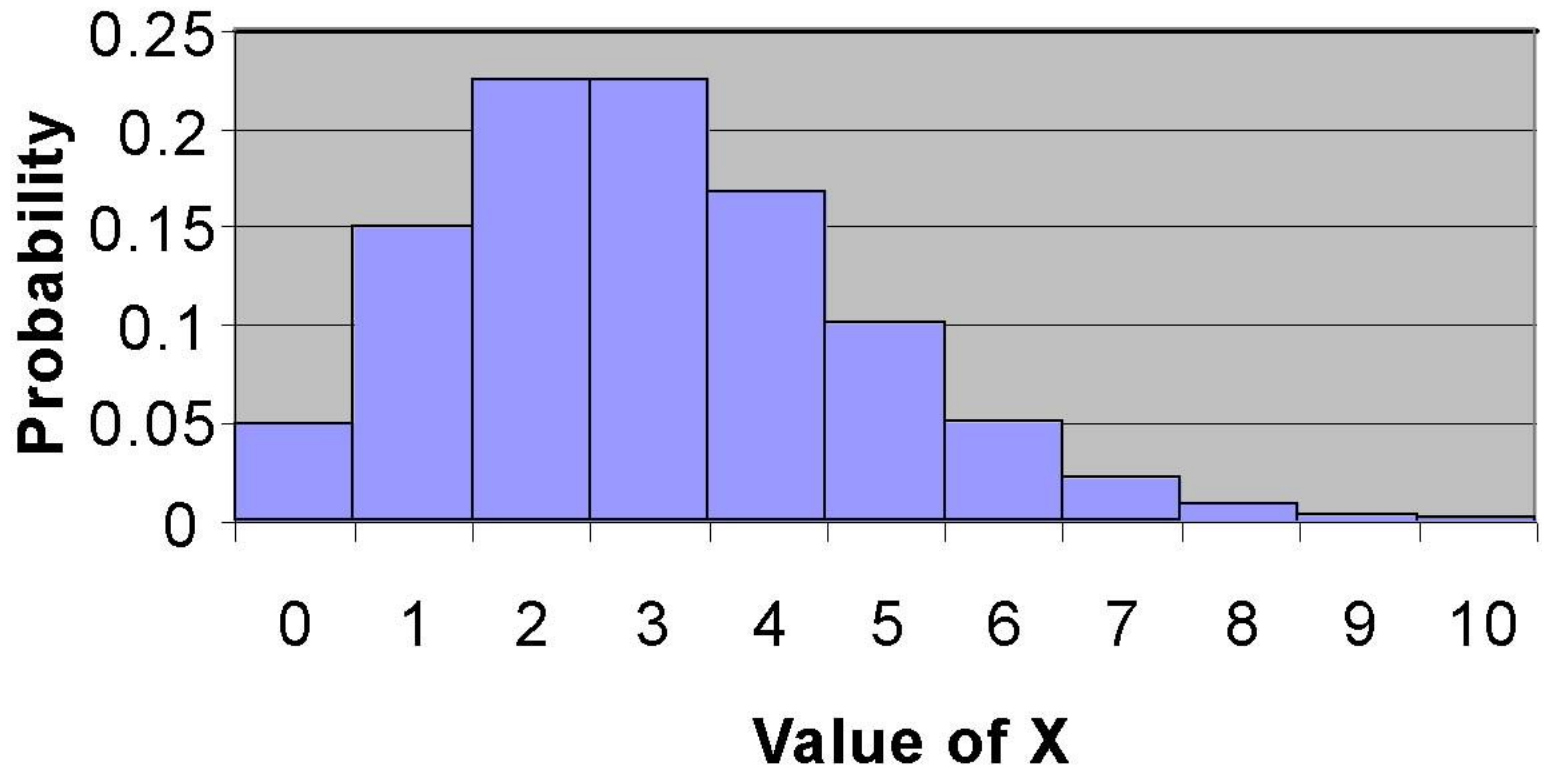
$$P(X = x) = \frac{(\lambda t)^x}{x!} e^{-\lambda t} \quad x = 0, 1, 2, 3, \dots$$

where λ represents the average number of occurrences of the event in some interval of length 1.

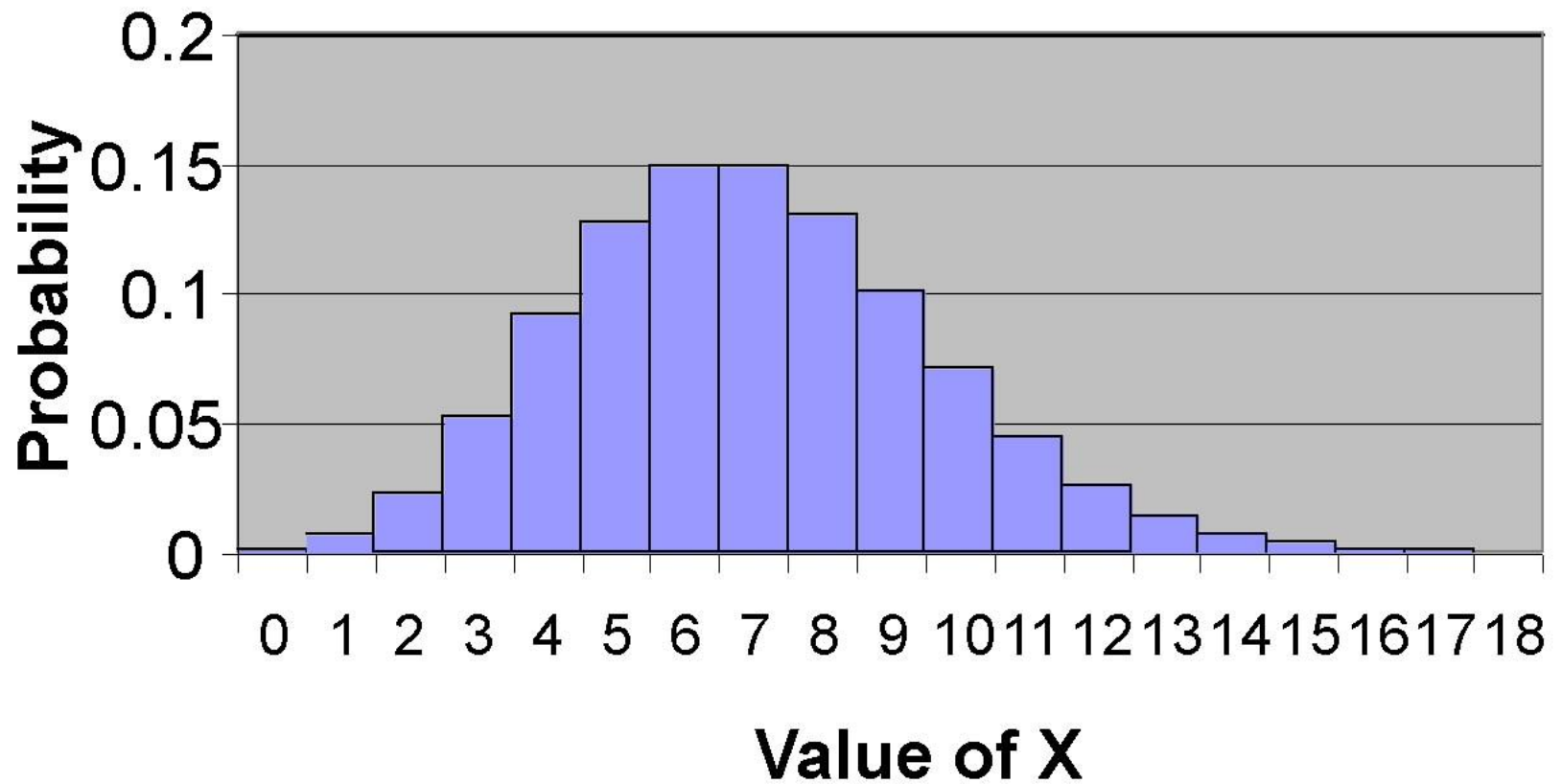
Probability Histogram of a Poisson Distribution with $\mu = 1$



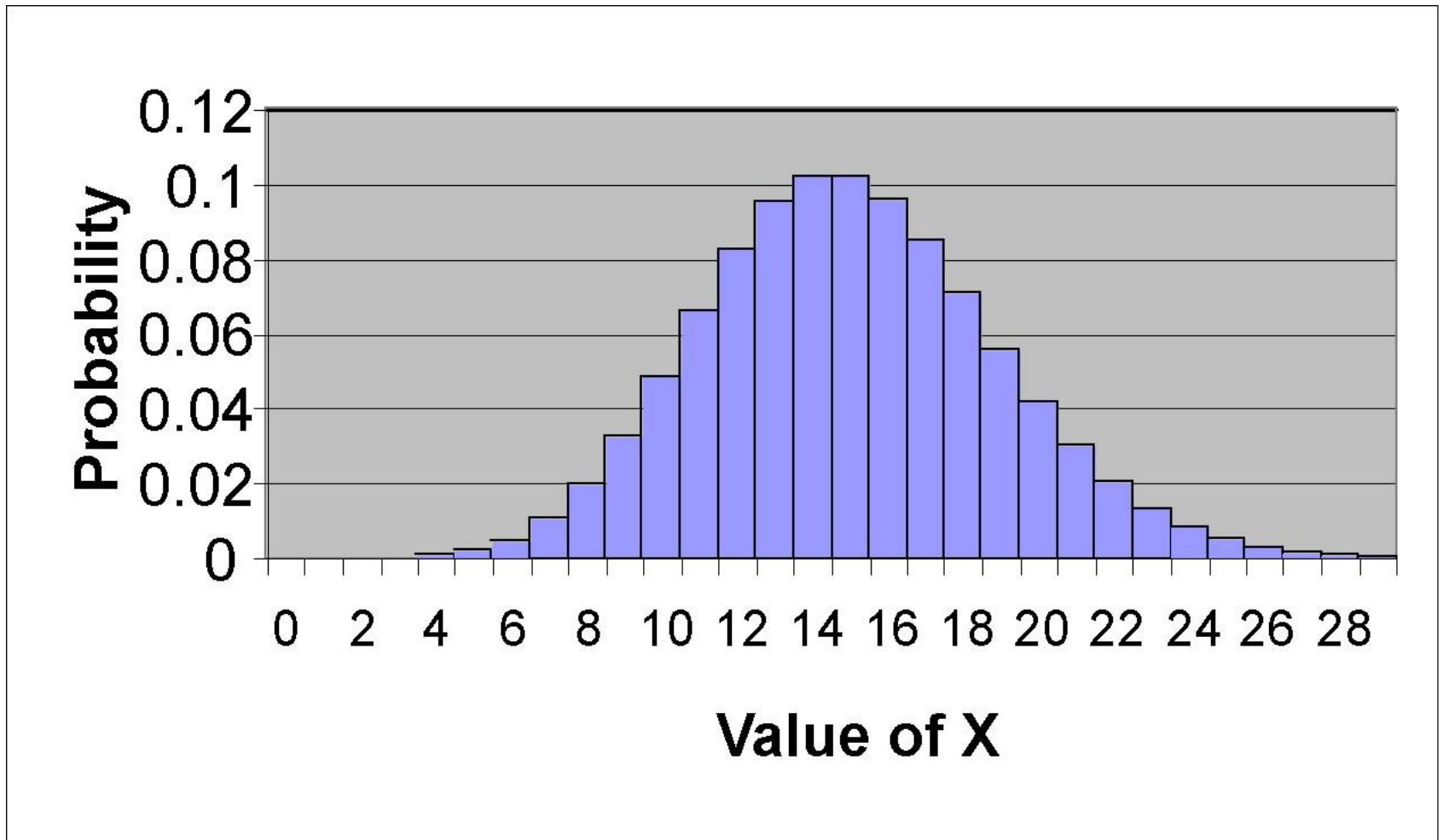
Probability Histogram of a Poisson Distribution with $\mu = 3$



Probability Histogram of a Poisson Distribution with $\mu = 7$



Probability Histogram of a Poisson Distribution with $\mu = 15$



EXAMPLE

Poisson Particles

In 1910, Ernest Rutherford and Hans Geiger recorded the number of α -particles emitted from a polonium source in eighth-minute (7.5 second) intervals. The results are reported in the table on the next slide. Does a Poisson probability function accurately describe the number of α -particles emitted?

Source: Rutherford, Sir Ernest; Chadwick, James; and Ellis, C.D.. Radiations from Radioactive Substances. London, Cambridge University Press, 1951, p. 172.

Number Detected

Frequency

0

57

1

203

2

383

3

525

4

532

5

408

6

273

7

139

8

45

9

27

10

10

11

6

x	Frequency, <i>f</i>	<i>x</i><i>f</i>	Expected Proportion, if Poisson	Expected Frequency, if Poisson
0	57	0	0.020866	54.4
1	203	203	0.080744	210.6
2	383	766	0.156225	407.4
3	525	1575	0.201511	525.5
4	532	2128	0.194943	508.4
5	408	2040	0.150872	393.5
6	273	1638	0.097303	253.8
7	139	973	0.053789	140.3
8	45	360	0.026018	67.9
9	27	243	0.011187	29.2
10	10	100	0.004329	11.3
11	6	66	0.001523	4.0
Sum: 2608		10092		

$$\mu = 3.869632$$

The Poisson probability distribution function can be used to approximate binomial probabilities provided the number of trials $n \geq 100$ and $np \leq 10$. In other words, the number of independent trials of the binomial experiment should be large and the probability of success should be small.

EXAMPLE *Using the Poisson Distribution to Approximate Binomial Probabilities*

According to the U.S. National Center for Health Statistics, 7.6% of male children under the age of 15 years have been diagnosed with Attention Deficit Disorder (ADD). In a random sample of 120 male children under the age of 15 years, what is the probability that at least 4 of the children have ADD?