Chapter 6—The Binomial Probability Distribution

6.1 Probability Distributions

Definition (p. 285): A **random variable** is a numerical measure of the outcome of 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 result from counting.
- A **continuous random variable** is a random variable that results from measurement.

EXAMPLE—*Distinguishing Between Discrete and Continuous Random Variables*

Determine whether the following random variables are discrete or continuous. To help answer the question, 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.

Definition (p. 286): The **probability distribution** of a random variable X 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.

x	P(X=x)
-10	0.20
0	0.25
+10	0.40
+20	0.15

Rules for a Discrete Probability Distribution (p. 287)

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

1. $\Sigma P(X=x) = 1$
2. $0 \le P(X=x) \le 1$

Example: Business Return and Risk

Suppose you plan on starting a new business and would like to know the **expected return** for the business. You develop a subjective probability distribution of the returns and their associated probabilities. Note: X=return in $1,000/year (e.g., -10 means a loss of $10,000).

x ($1,000/yr)	**P(X=x)**
-10	0.20
0	0.25
+10	0.40
+20	0.15

What is the expected return for this business?

Mean of a Discrete Random Variable (p. 289)

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

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

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

$$\mu_x = \Sigma\, x \cdot P(X=x) = x_1P(x_1) + x_2P(x_2) + x_3P(x_3) + x_4P(x_4)$$

$$= -10(0.20) + 0(0.25) + 10(0.40) + 20(0.15)$$

$$= -2 + 0 + 4 + 3$$

$= 5$ >>> Expected return is $5,000

Interpretation of the Mean of a Discrete Random Variable

Suppose an experiment is repeated n independent times and the values of the random variable X are recorded. As the number of repetitions of the experiment, n, increases the 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 value of the random variable X after the second experiment, and so on. Then

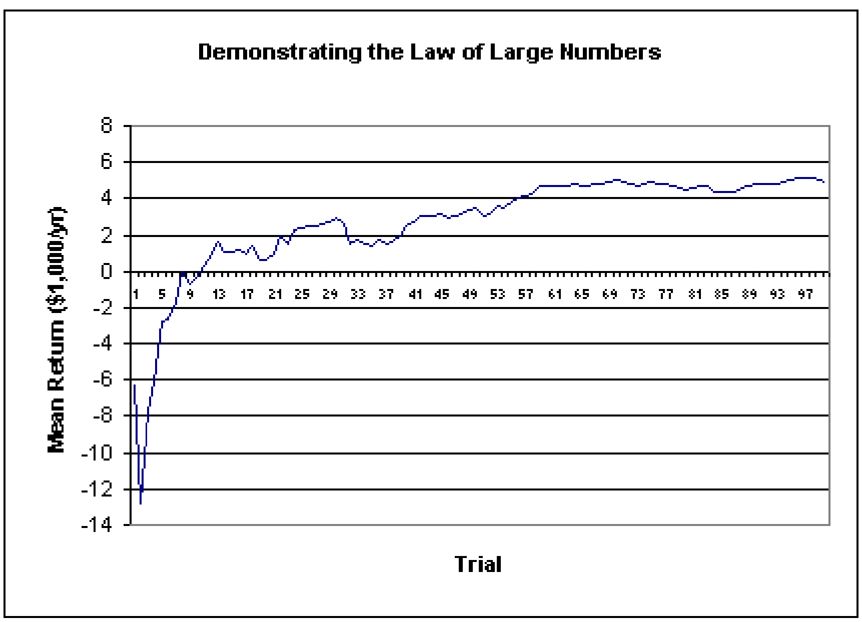
$$\overline{X} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

The difference between $\overline{X}$ *and* μ_x will get closer to 0 as n increases.

The following data represent the business net return ($1,000/yr) for 100 years.

-6.3	7.7	5.8	-8.3	9.3	-17.4	9.0	-6.0	14.3	3.8
-19.2	5.4	22.7	-30.8	16.1	8.0	1.8	-1.0	12.1	8.1
2.3	11.8	-7.5	8.8	7.4	25.7	2.4	-3.8	-1.7	7.4
-1.1	-6.4	21.1	-6.7	-0.7	2.5	17.0	17.0	-15.4	16.8
10.2	1.4	4.8	-2.9	7.9	17.1	-6.3	6.6	0.1	14.3
-2.0	2.9	5.0	12.3	-7.7	18.9	5.7	-3.5	2.1	12.1
4.2	-3.1	0.8	-5.6	10.0	7.6	9.7	11.4	5.5	5.4
11.5	9.3	5.5	7.5	11.1	14.6	6.9	-5.6	21.1	6.8
-6.3	-12.8	7.9	13.9	13.3	**28.0**	14.0	-1.5	15.5	-8.0
2.7	1.4	9.0	23.7	4.5	2.9	12.1	-4.8	14.3	-5.6

$$\overline{X} = \frac{x_1 + x_2 + \cdots + x_{100}}{100} = 4.901$$



As n increases, the value of $\overline{X}$ approaches μ_x = $5(000)

What is the risk in this business? Risk can be represented by the standard deviation of return, with a higher standard deviation indicating a higher degree of risk.

Variance and Standard Deviation of a Discrete Random Variable (p. 292)

The variance of a discrete random variable is given by

$$\sigma^2_x = \Sigma\,[(x-\mu_x)^2 \cdot P(X=x)]$$

Where x is the value of the random variable, μ_x is the mean of the random variable

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

To find the standard deviation of the discrete random variable, take the square root of the variance. That is, $\sigma_x = \sqrt{\sigma_x^2}$

x	P(X=x)	$(x-\mu_x)^2 \cdot P(X=x)$
-10	0.20	$(-10-5)^2 \cdot 0.20 = 45.00$
0	0.25	$(0-5)^2 \cdot 0.25 = 6.25$
+10	0.40	$(10-5)^2 \cdot 0.40 = 10.00$
+20	0.15	$(20-5)^2 \cdot 0.15 = 33.75$

$\mu_x = 5$ $\sigma^2_x = \Sigma\,(x-\mu_x)^2 \cdot P(X=x) = 95$

$\sigma_x = \sqrt{\sigma_x^2} = 9.75$

Example 7, p. 291—Finding Expected Value

Problem: 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 an 18-year-old male for $350. According to the *National Vital Statistics Report*, Vol. 47, No. 28, the probability the male will survive the year is 0.998789. Compute the expected value of this policy to the insurance company.

x=payoff to insurance company	**P(X=x)**
$350 (survives)	0.998789
-$249,650 (dies)	0.001211

In-Class Activity: Expected Value (p. 292)
Consider the following game of chance. A player pays $1 and rolls a pair of fair dice. If the player rolls a 2, 3, 4, 10, 11, or 12, the player loses the $1 bet. If the player rolls 5, 6, 8, or 9, there is a "push" and the player gets his or her dollar back. If the player rolls a 7, the player wins $1. **(a)** Construct a probability distribution that describes the game. **(b)** Compute the expected value of the game from the player's point of view.

Oil Tanker Insurance:
An oil tanker is making a trip from Iraq's offshore southern oil terminals to Europe. The value of the tanker and oil is $80 million. The probability distribution for damage (X) to the oil tanker is shown below:

x=loss to insurance company (mil)	P(X=x)
$0	0.982
$10	0.010
$20	0.005
$50	0.002
$80	0.001

(a) What is the expected damage payout for an insurer (such as Lloyd's of London) who writes this policy?

(b) Assume the premium for the policy is $0.4 million. What does the insurer expect to make on each policy?

Heart Attack Insurance:
Mr. Smith would like to buy catastrophic insurance against a heart attack. The insurance will pay Mr. Smith $100,000 if he has a heart attack. The probability that Mr. Smith will have a heart attack in a given year is 0.01, so the probability that he will not have a heart attack is 0.99. The insurance premium for the heart attack policy is $800 per year.

(a) What is the expected return for the insurance company?

(b) Because the insurance company is losing money on heart attack insurance, a company officer asks the premium manager to determine the annual premium required for the insurance company to make $150 on each $100,000 policy. Perform the calculations to determine the annual premium.

6.2 The Binomial Probability Distribution

Criteria for a Binomial Probability Experiment

1. Experiment is performed a fixed number of times, with each repetition called a **trial**.
2. Trials are independent, which means that the outcome of one trial will not affect the outcome of another trial.
3. For each trial, there are two mutually exclusive outcomes—success or failure.
4. Probability of success is fixed for each trial of the experiment.

Notation Used in the Binomial Probability Distribution

- n=number of independent trials of the experiment.
- Let p = probability of success and (1-p) = probability of failure.
- X=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?

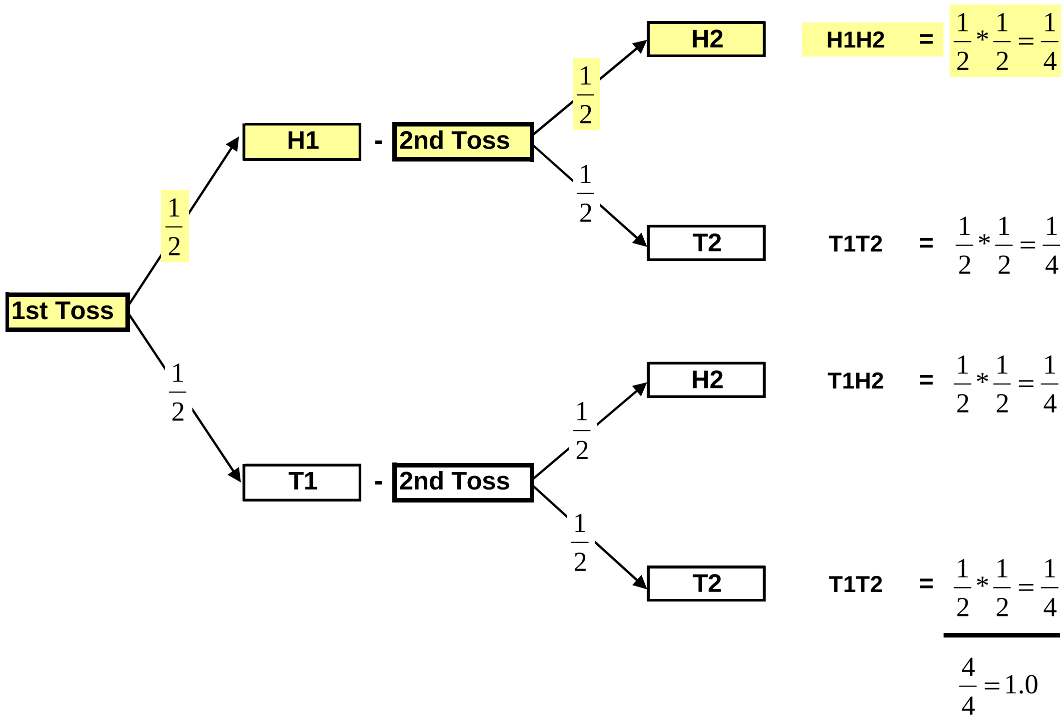
(a) A player rolls a pair of fair die 10 times. The number *X* of 7's rolled is recorded.

(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: Binomial Probability Distribution for Coin Toss

Problem: A fair coin is tossed twice and the number of heads, X, is recorded. Construct a probability distribution for the random variable X.



Let X=number of heads.

P(X=0) = P(T1T2)=P(T1)*P(T2) = ½ * ½ = ¼ = 0.25 ← Multiplication Rule for Independent Events

P(X=1) = P[(H1T2) or (T1H2)] = P(H1T2) + P(T1H2) ← Addition Rule for Mutually Exclusive Events

= P(H1)*P(T2) + P(T1)*P(H2) ← Multiplication Rule for Independent Events

= ½ * ½ + ½ * ½

= 2(0.25) = 0.50

P(X=2) = P(H1H2)=P(H1)*P(H2) = ½ * ½ = ¼ = 0.25

Probability Distribution of X:

x	P(X=x)
0	0.25
1	0.50
2	0.25

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

$$P(X=x) = {}_nC_x\, p^x\, (1-p)^{n-x} \qquad x=0, 1, 2, \ldots, n$$

Apply the Binomial Probability Distribution Function to the coin toss problem on the previous page with n=2 and p=1/2.

x	${}_nC_x = \frac{2!}{x!(2-x)!}$	p^x		$(1-p)^{2-x}$		$P(X=x)={}_nC_x p^x (1-p)^x$	
0	**1**	$0.50^0 =$	**1**	$(1-0.50)^2 =$	**0.25**	1*1*0.25	**=0.25**
1	**2**	$0.50^1 =$	**0.50**	$(1-0.50)^1 =$	**0.50**	2*0.50*0.50	**=0.50**
2	**1**	$0.50^2 =$	**0.25**	$(1-0.50)^0 =$	**1**	1*0.25*1	**=0.25**
							1.00

Phase	Math Symbol
"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?

		# of successes	**P(X=x)**	**Cumulative Sum**
n=	20	0	0.017557	0.01755663
p=	0.183	1	0.07865	0.09620689
		2	0.16736	0.26356732
		3	0.224923	0.48848992
		4	0.214117	0.70260687
		5	0.153472	0.85607918

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:

$\mu_x = np$ and $\sigma_x = \sqrt{np(1-p)}$

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.

Using Excel to Compute Binomial Probabilities (see p. 305):

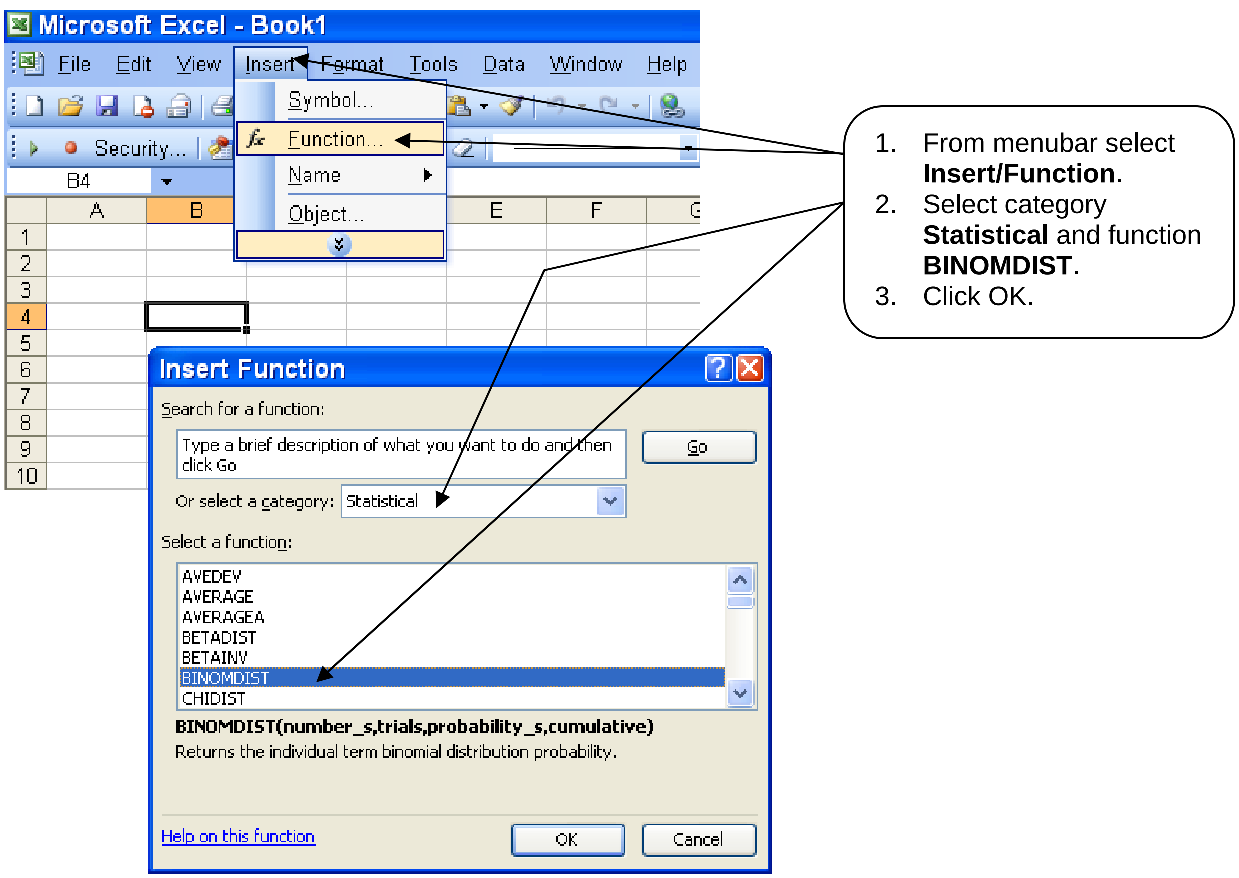
According to the United States Census Bureau, 18.3% of all households have 3 or more cars. In a random sample of 20 households, what is the probability that less than 4 have 3 or more cars?

Computing P(X = x):

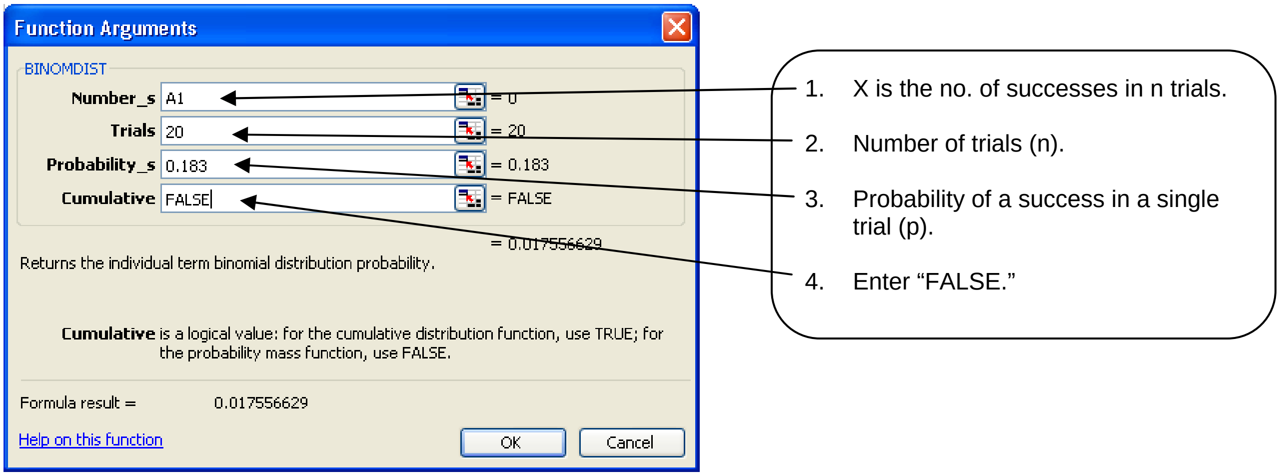
Step 1: Enter the possible values of the random variable x in column A, e.g., with n=20, p=0.183, enter 0, 1, 2, 3 into A.

	A	B
1	0	0.017557
2	1	0.078650
3	2	0.167360
4	3	0.224923

Step 2: With the cursor in cell B1, select **Insert/Function** (*f_x*) from the Windows Menubar. Highlight **Statistical** in the Function category window. Highlight **BINOMDIST** in the Function name window.



Step 3: Fill in the window as follows:



Click OK.

Step 4: Copy the contents in cell B1 to the remaining cells.

Computing P(X $\leq$ x):

Follow the same steps as those presented for computing P(X=x). In the BINOMDIST window, type "TRUE" in the cumulative cell.

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:

$\mu_x = np$ and $\sigma_x = \sqrt{np(1-p)}$

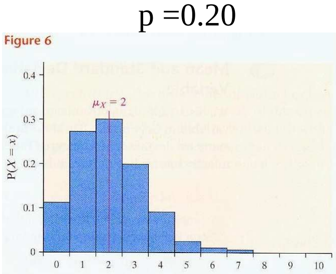
Example 7: Constructing Binomial Probability Histograms (p. 308):

n=10 and p=0.20 $\mu_x = 10(0.20) = 2$ $\sigma_x = \sqrt{10(0.20)(1-0.20)} = 1.26$

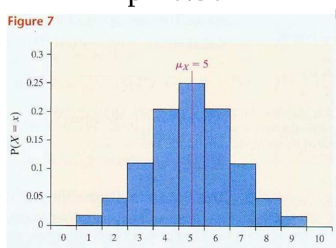
n=10 and p=0.50 $\mu_x = 10(0.50) = 5$ $\sigma_x = \sqrt{10(0.50)(1-0.50)} = 1.58$

n=10 and p=0.80 $\mu_x = 10(0.80) = 8$ $\sigma_x = \sqrt{10(0.80)(1-0.80)} = 1.26$

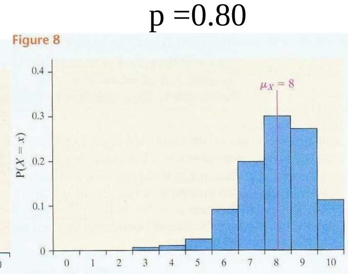
p =0.20



p =0.50

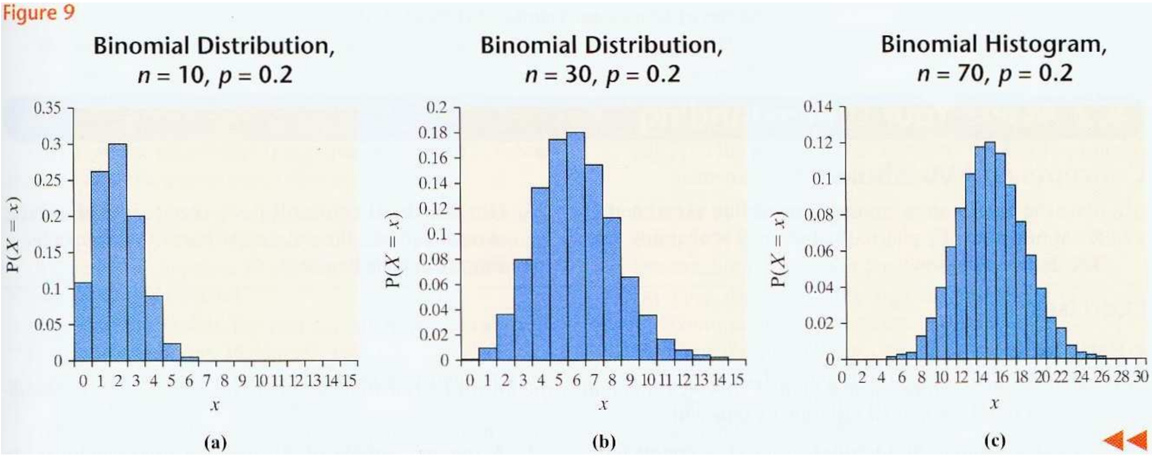


p =0.80



What happens to the shape of the binomial distribution as “p” changes from 0.20 to 0.50 to 0.80?

Exploration (p. 308)—Binomial Distributions for p=0.20 with n=10, 30, and 70.



What happens to the shape of the binomial distribution as n increases from 10 to 30 to 70? (Use the formula on next page to determine whether the distribution is bell-shaped.)

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 (p. 226) 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 finds 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 finds 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 finds 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%?

Heart Attack Insurance:
Mr. Smith would like to buy catastrophic insurance against a heart attack. The insurance will pay Mr. Smith $100,000 if he has a heart attack. The probability that Mr. Smith will have a heart attack in a given year is 0.01, so the probability that he will not have a heart attack is 0.99. The insurance premium for the heart attack policy is $800 per year.

(a) What is the expected return for the insurance company?

$$E(X) = \sum xP(X = x)$$
$$= 0.99(800) + 0.01(-99{,}200)$$
$$= \quad +792 \qquad -992$$
$$= -200$$

(b) Because the insurance company is losing money on heart attack insurance, a company officer asks the premium manager to determine the annual premium required for the insurance company to make $150 on each $100,000 policy. Perform the calculations to determine the annual premium.

$$E(X) = \sum xP(X = x)$$
$$150 = 0.99(premium) + 0.01(-100{,}000 + premium)$$
$$Solve\ for\ premium = 1{,}150.00$$

Example 6: Interpretation of the Mean of a Discrete Random Variable (p. 290)

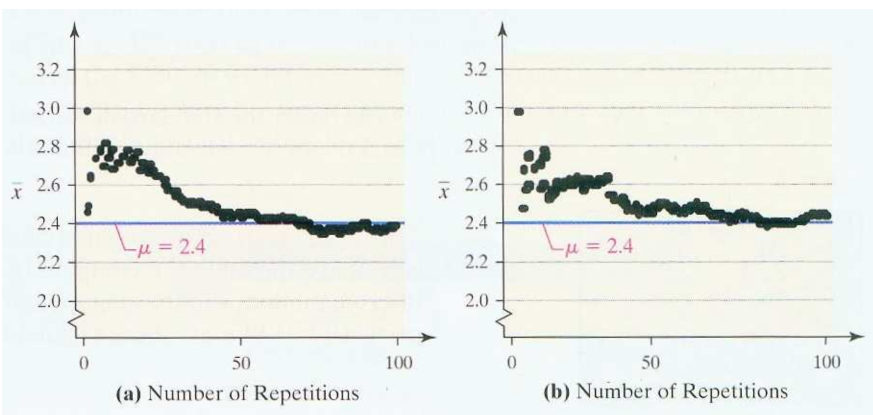
Problem: A basketball player historically makes 80% of her free throws. She is asked to shoot three free throws 100 times. Computer the mean number of free throws made.

TABLE 4

3	3	2	3	3	3	1	2	3	2
2	3	3	1	2	2	2	2	2	3
3	3	2	2	3	2	3	2	2	2
3	3	2	3	2	3	3	2	3	1
3	2	2	2	2	0	2	3	1	2
3	3	2	3	2	3	2	1	3	2
2	3	3	3	1	3	3	1	3	3
3	2	2	1	3	2	2	2	3	2
3	2	2	2	3	3	2	2	3	3
2	3	2	1	2	3	3	2	3	3

$$\overline{X} = \frac{3+2+\ldots 3}{100} = 2.35$$

Note that the mean, $\overline{X}$, approaches the expected value ($\mu_x = 2.4$) as the number of repetitions increases.



Example 2: A Discrete Probability Distribution

Suppose a basketball player historically makes 80% of her free throw attempts and free throw attempts are independent. The table below shows the probability distribution for the random variable X, where X represents the number of successful attempts in three shots.

X	P(X=x)
0	0.008
1	0.096
2	0.384
3	0.512

Mean of a Discrete Random Variable (p. 289)

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

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

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

Example 5: Computing the Mean of a Discrete Random Variable (p. 289)

$$\mu_x = \Sigma\, x \cdot P(X=x) = x_1P(x_1) + x_2P(x_2) + x_3P(x_3) + x_4P(x_4)$$

$$= 0(0.008) + 1(0.096) + 2(0.384) + 3(0.512) = 2.4$$

Example 7: Computing the Variance and Standard Deviation of a Discrete Random Variable (p. 292)

Variance and Standard Deviation of a Discrete Random Variable (p. 292)

The variance of a discrete random variable is given by

$$\sigma^2_x = \Sigma\,[(x - \mu_x)^2 \cdot P(X=x)]$$

Where x is the value of the random variable, μ_x is the mean of the random variable

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

To find the standard deviation of the discrete random variable, take the square root of the variance. That is, $\sigma_x = \sqrt{\sigma_x^2}$

x	**P(X=x)**	$\mathbf{(x-\mu_x)^2 \cdot P(X=x)}$
0	0.008	$(0 - 2.4)^2 \cdot 0.008 = 0.04608$
1	0.096	$(1 - 2.4)^2 \cdot 0.096 = 0.18816$
2	0.384	$(3 - 2.4)^2 \cdot 0.384 = 0.06144$
3	0.512	$(4 - 2.4)^2 \cdot 0.512 = 0.18432$

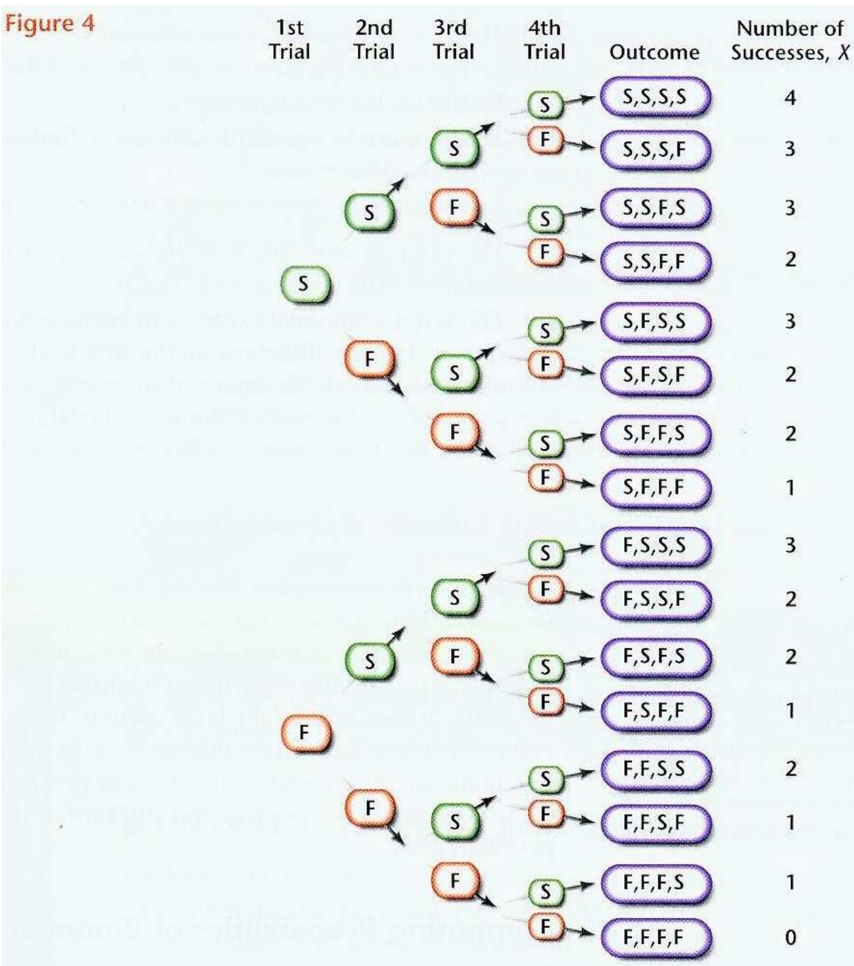
$\sigma^2_x = \Sigma\,(x - \mu_x)^2 \cdot P(X=x) = 0.483$

$\sigma_x = \sqrt{\sigma_x^2} = 0.695$

Example 2: Binomial Probability Distribution for Blood Type O-Negative.

Problem: According to the *Information Please* almanac, 6% of the human population is blood type O-negative. A simple random sample of size four is taken and the number of people X with blood type O-negative is recorded. Construct a probability distribution for the random variable X.

Success is a person with blood type O-negative:	p=0.06.
Failure is a person with any other blood type:	(1-p)=0.94
Number of people in the random:	n=4
Number of successes:	0, 1, 2, 3, or 4



$P(X=0) = P(FFFF) = P(F)*P(F)*P(F)*P(F)$

$= (0.94)(0.94)(0.94)(0.94)$

$= (0.94)^4$

$= 0.78075$

Multiplication Rule for Independent Events—P(E and F)=P(E)*P(F)

$P(X=1) = P(SFFF \text{ or } FSFF \text{ or } FFSF \text{ or } FFFS))$

$= P(SFFF) + P(FSFF) + P(FFSF) + P(FFFS)$ Addition Rule for Mutually Exclusive Events --P(E or F)=P(E)+P(F)

$=(0.06)(0.94)^3 +(0.06)(0.94)^3+(0.06)(0.94)^3 +(0.06)(0.94)^3$

$= 4(0.06)(0.94)^3$ Multiplication Rule for Independent Events

$=0.19934$

$P(X=2) = P(SSFF \text{ or } SFSF \text{ or } SFFS \text{ or } FSSF \text{ or } FSFS \text{ or } FFSS)$

$= P(SSFF) + P(SFSF) + P(SFFS) + P(FSSF) + P(FSFS) + P(FFSS)$

$=(0.06)^2(0.94)^2 +(0.06)^2(0.94)^2+(0.06)^2(0.94)^2 +(0.06)^2(0.94)^2+(0.06)^2(0.94)^2$
$+(0.06)^2(0.94)^2$

$= 6(0.06)^2(0.94)^2$

$=0.01909$

P(X=3) and P(X=4) are computed similarly.

Probability distribution of X:

x	**P(X = x)**
0	0.78075
1	0.19934
2	0.01909
3	0.00081
4	0.00001

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
>
> $P(X=x) = {}_nC_x\, p^x\, (1-p)^{n-x}$ x=0, 1, 2, …, n

Question: A simple random sample of size four is taken and the number of people X with blood type O-negative is recorded. What is the probability that 0, 1, 2, 3, or all 4 will have blood type O-negative?

Success is a person with blood type O-negative—p=0.06.
Failure is a person with any other blood type—(1-p)=0.94
Number of people in the random—n=4
Number of successes—0, 1, 2, 3, or 4

The following table helps keep all the factors straight:

x	${}_nC_x = \frac{4!}{x!(4-x)!}$	p^x		$(1-p)^{4-x}$		$P(X=x)={}_nC_x p^x(1-p)^x$
0	1	0.06^0 =	1	$(1-0.06)^4$ =	0.780749	**0.78074896**
1	4	0.06^1 =	0.06	$(1-0.06)^3$ =	0.830584	**0.19934016**
2	6	0.06^2 =	0.0036	$(1-0.06)^2$ =	0.8836	**0.01908576**
3	4	0.06^3 =	0.000216	$(1-0.06)^1$ =	0.94	**0.00081216**
4	1	0.06^4 =	0.00001296	$(1-0.06)^0$ =	1	**0.00001296**
						1