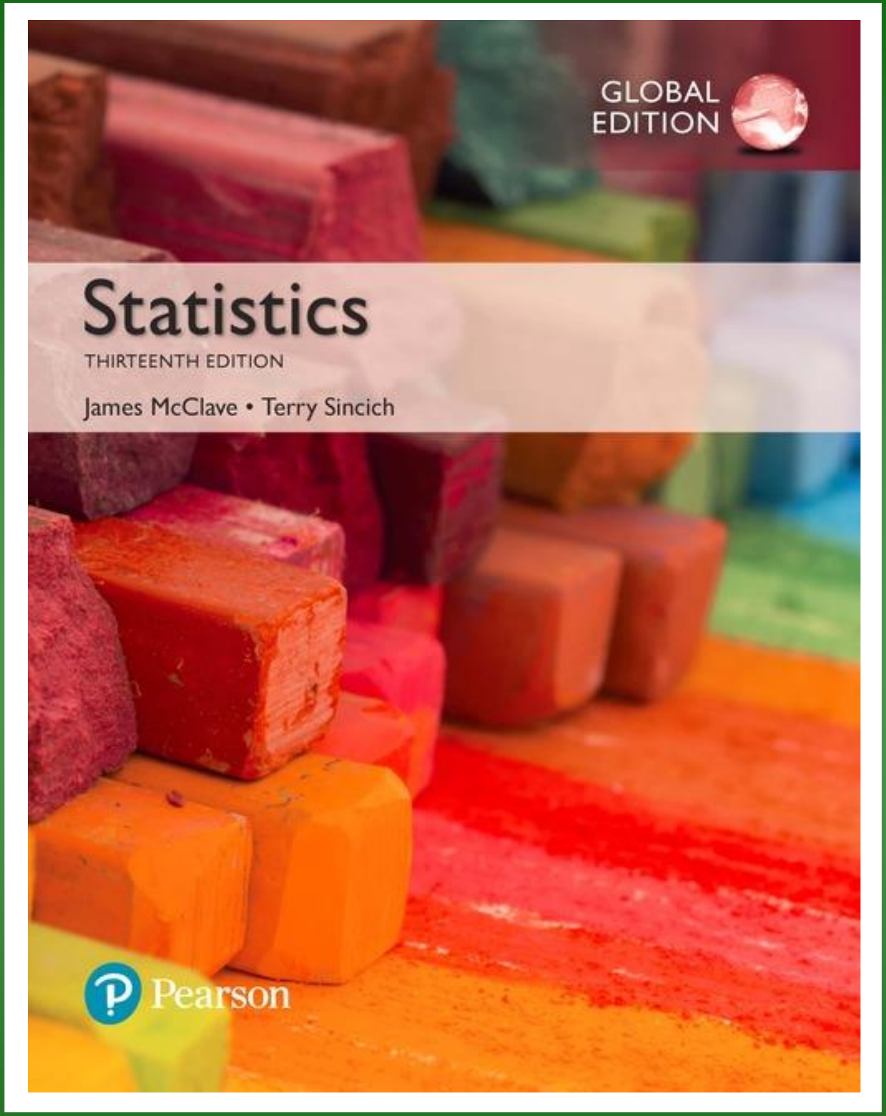
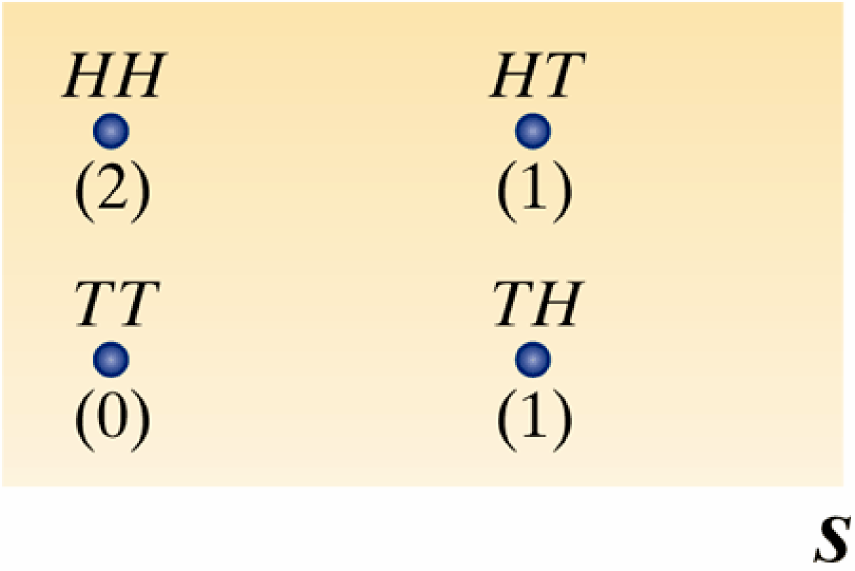


Chapter 4

Discrete Random Variables

Figure 4.1 Venn diagram for coin-tossing experiment

- Many of the examples of experiments last week generated quantitative (numerical) observations.
- One possible numerical outcome in coin tossing experiment is the total number of heads observed.



Definition

- In the jargon of probability, the variable "**total number of heads observed in two tosses of a coin**" is called a random variable.

A **random variable** is a variable that assumes numerical values associated with the random outcomes of an experiment, where one (and only one) numerical value is assigned to each sample point.

4.1

Two Types of Random Variables

Infinite sample points problem

- Sample-point probabilities corresponding to an experiment **must sum to 1**.
- If the number of sample points can be completely listed, the job is straightforward.
- But if the experiment results in an infinite number of sample points, the task of assigning probabilities to the sample points is impossible without the aid of a **probability model**.

Problem

- A panel of **10 experts** for the Wine Spectator is asked to taste a new white wine and assign it a rating of **0**, **1**, **2**, or **3**. A score is then obtained by adding together the ratings of the 10 experts. How many values can this random variable assume?

PEARSON

Solution

- A sample point is a sequence of 10 numbers associated with the rating of each expert.
- For example, one sample point is $\{1, 0, 0, 1, 2, 0, 0, 3, 1, 0\} \rightarrow x = 8$
- A total of $4^{10} = 1{,}048{,}576$ sample points.
- Smallest score is 0 (if all 10 ratings are 0).
- Largest score is 30 (if all 10 ratings are 3).
- x can assume **31 values**.

Problem

- Suppose the Environmental Protection Agency (EPA) takes readings once a month on the amount of pesticide in the discharge water of a chemical company. If the amount of pesticide exceeds the maximum level set by the EPA, the company is forced to take corrective action and may be subject to penalty. Consider the random variable x to be the number of months before the company's discharge exceeds the EPA's maximum level. What values can x assume?

Solution

- The company's discharge of pesticide may exceed the maximum allowable level on the first month of testing, the second month of testing, etc.
- It is possible that the company's discharge will **never** exceed the maximum level. Thus, the set of possible values for the number of months until the level is first exceeded is the set of all positive integers $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \cdots$.

Problem

- Refer to previous example. A second random variable of interest is the amount x of pesticide (in milligrams per liter) found in the monthly sample of discharge waters from the same chemical company. What values can this random variable assume?

Solution

- Some possible values of x are 1.7, 28.42, and 100.987 milligrams per liter.
- Unlike the number of months before the company's discharge exceeds the EPA's maximum level, the set of all possible values for the amount of discharge **cannot be listed** (i.e., is **not countable**).
- The possible values for the amount x of pesticide would correspond to the points on the interval **between 0 and the largest possible value the amount of the discharge could attain**, the maximum number of milligrams that could occupy 1 liter of volume.

Definition

Random variables that can assume a *countable* number of values are called **discrete.**

The following are examples of **discrete random variables**:

1. The number of seizures an epileptic patient has in a given week: $\boldsymbol{x = 0, 1, 2, \cdots}$
2. The number of voters in a sample of 500 who favor impeachment of the president: $\boldsymbol{x = 0, 1, 2, \cdots, 500}$
3. The change received for paying a bill: $\boldsymbol{x = 1¢, 2¢, 3¢, \cdots, \$1, \$1.01, \$1.02, \cdots}$
4. The number of customers waiting to be served in a restaurant at a particular time: $\boldsymbol{x = 0, 1, 2, \cdots}$

Definition

Random variables that can assume values corresponding to any of the points contained in an interval are called **continuous**.

The following are examples of **discrete random variables**:

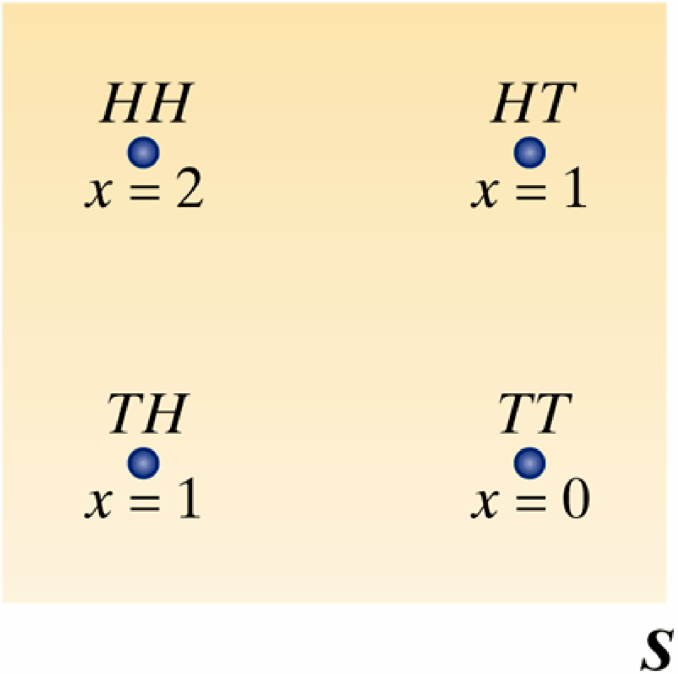
1. The length of time (in seconds) between arrivals at a hospital clinic: $\mathbf{0} \le \boldsymbol{x} \le \infty$
2. The length of time (in minutes) it takes a student to complete a one-hour exam: $\mathbf{0} \le \boldsymbol{x} \le \mathbf{60}$
3. The depth (in feet) at which a successful oil-drilling venture first strikes oil: $\mathbf{0} \le \boldsymbol{x} \le \boldsymbol{c}$, where $\boldsymbol{c}$ is the maximum depth obtainable.

4.2

Probability Distributions for Discrete Random Variables

Problem - Figure 4.2 Venn diagram for the two-coin-toss experiment

- Recall the experiment of tossing two coins, and let x be the number of heads observed. Find the probability associated with each value of the random variable x, assuming that the two coins are fair.
- Let's identify the probabilities of the sample points.



$$P(x=0) = P(TT) = \frac{1}{4}$$

$$P(x=1) = P(TH) + P(HT) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(x=2) = P(HH) = \frac{1}{4}$$

PEARSON

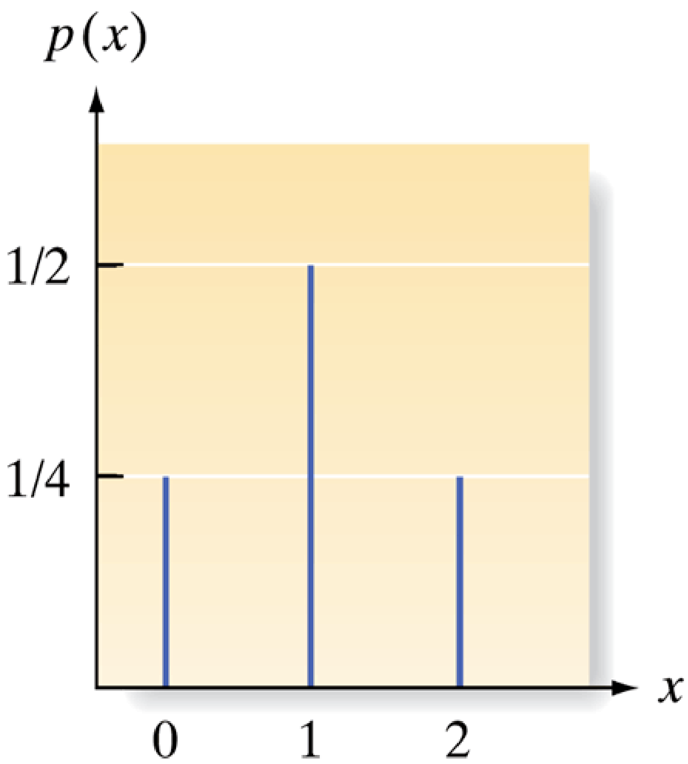
Table 4.1

- Thus, we now know the values the random variable can assume $(0, 1, 2)$ and how the probability is distributed over those values $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.
- This dual specification completely describes the random variable and is referred to as the **probability distribution**, denoted by the symbol $\boldsymbol{p(x)}$.

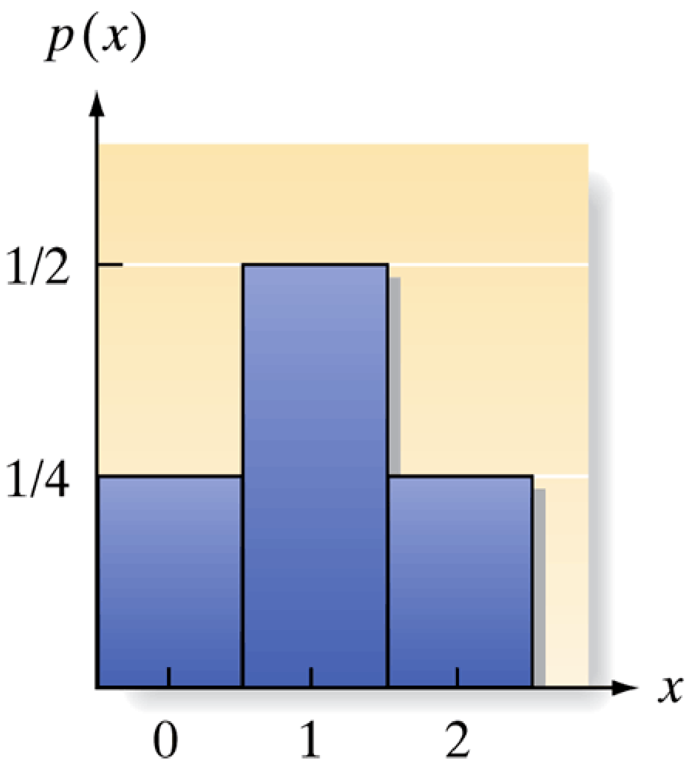
Table 4.1 Probability Distribution for Coin-Toss Experiment: Tabular Form

x	$p(x)$
0	$1/4$
1	$1/2$
2	$1/4$

Figure 4.3 Probability distribution for coin-toss experiment: graphical form



a. Point representation of $p(x)$



b. Histogram representation of $p(x)$

Definition

The **probability distribution of a discrete random variable** is a graph, table, or formula that specifies the probability associated with each possible value that the random variable can assume.

Definition

- Two requirements must be satisfied by all probability distributions for discrete random variables:

Requirements for the Probability Distribution of a Discrete Random Variable x

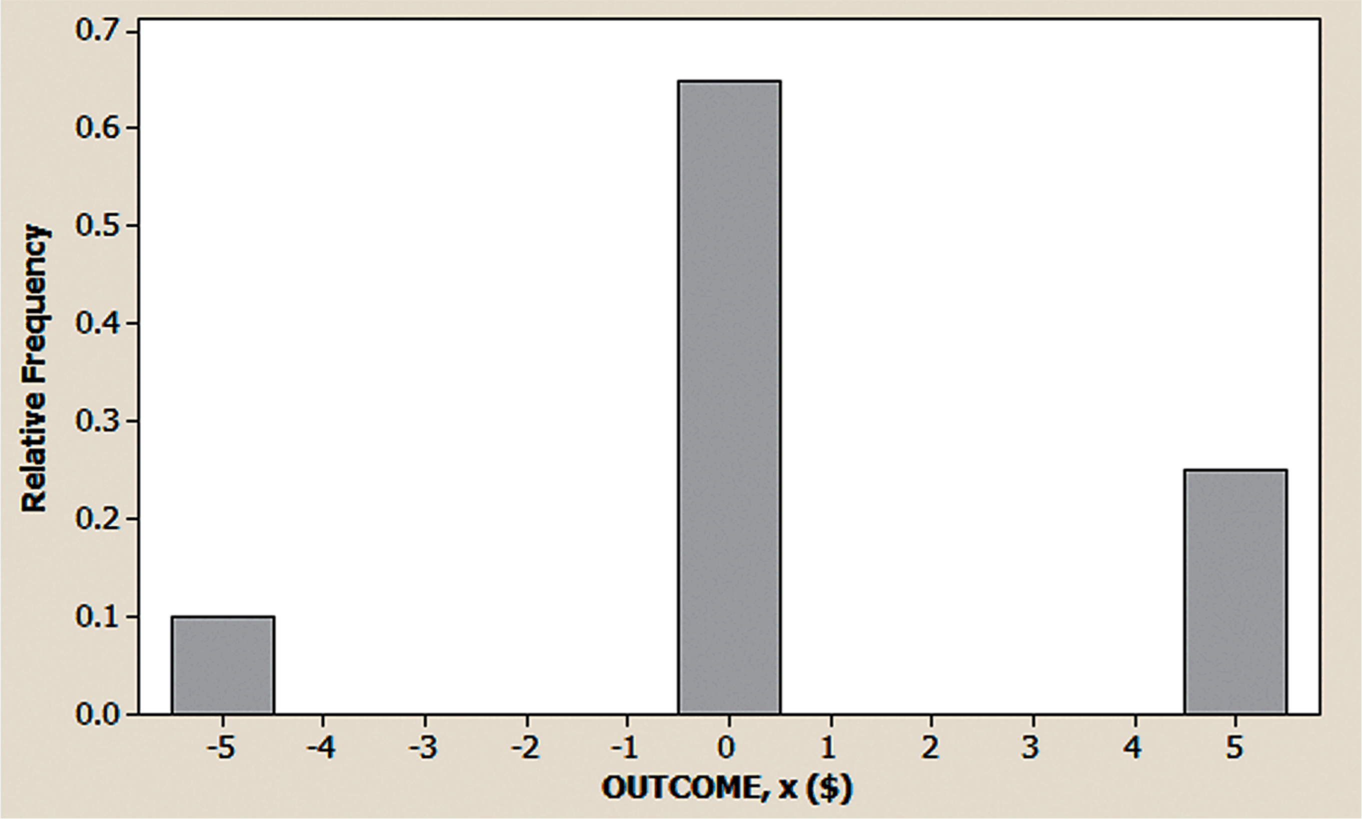
1. $p(x) \geq 0$ for all values of x.
2. $\sum p(x) = 1$

where the summation of $p(x)$ is over all possible values of x.*

Problem

- Craps is a popular casino game in which a player throws two dice and bets on the outcome (the sum total of the dots showing on the upper faces of the two dice).
- Consider a $5 wager. On the first toss (called the come-out roll), if the total is 7 or 11 the roller wins $5.
- If the outcome is a 2, 3, or 12, the roller loses $5.
- For any other outcome (4, 5, 6, 8, 9, or 10), a point is established and no money is lost or won on that roll.
- Use the histogram to find the approximate probability distribution of x.

Figure 4.4 MINITAB Histogram for $5 Wager on Come-out roll in Craps



$$p(-\$5) = .1, p(\$0) = .65, p(\$5) = .25$$

4.3

Expected Values of Discrete Random Variables

Probability distributions and relative frequency distributions

- If a discrete random variable x were observed **a very large number of times** and the data generated were arranged in a relative frequency distribution, the relative frequency distribution would be indistinguishable from the probability distribution for the random variable.
- Thus, the **probability distribution for a random variable is a theoretical model for the relative frequency distribution** of a population.
- The probability distribution for x possesses a mean μ and a variance σ^2 that are identical to the corresponding descriptive measures for the population.

PEARSON

Figure 4.5 Probability distribution for a two-coin toss

- Try to locate the mean of the distribution intuitively.
- We may reason that the mean μ of this distribution is equal to **1** as follows:

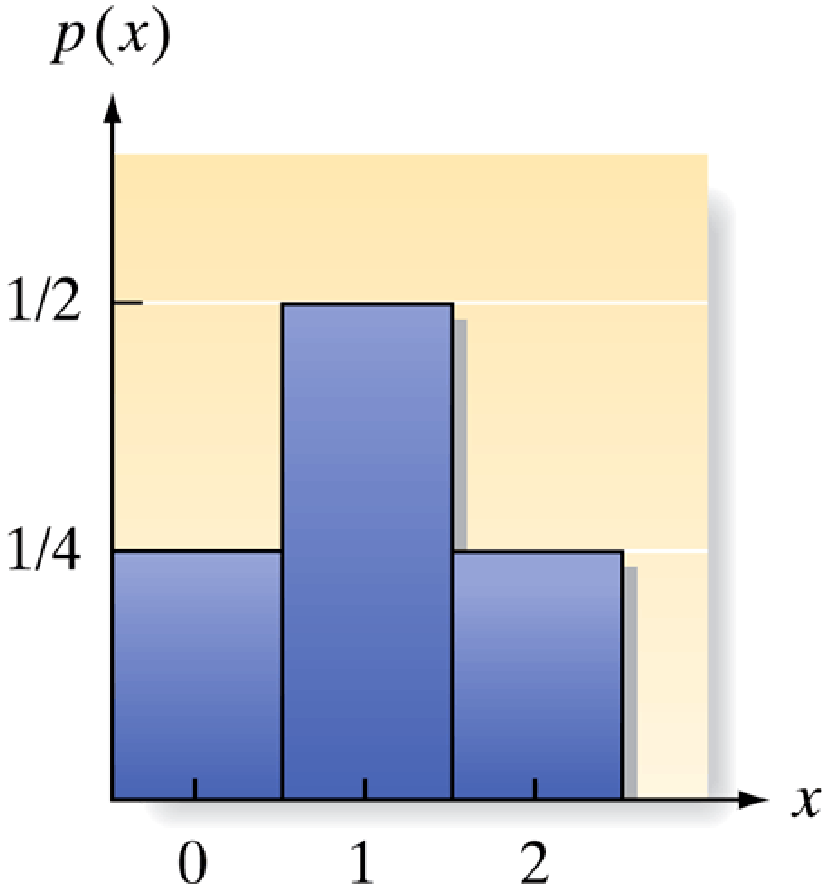
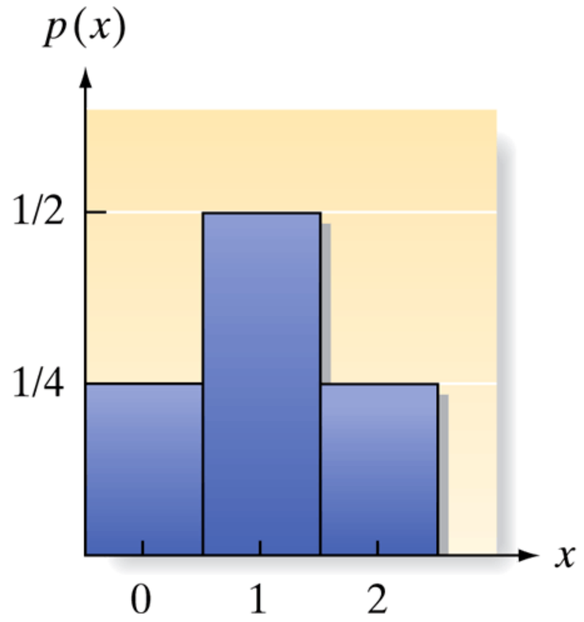


Figure 4.5 Probability distribution for a two-coin toss

- In a large number of experiments—say, 100,000—
  - 1/4 (or 25,000) should result in $x = 0$ heads,
  - 1/2 (or 50,000) in $x = 1$ head,
  - 1/4 (or 25,000) in $x = 2$ heads.



$$\mu = \frac{0(25{,}000) + 1(50{,}000) + 2(25{,}000)}{100{,}000} = 0(^1/_4) + 1(^1/_2) + 2(^1/_4)$$

$$= 0 + {}^1/_2 + {}^1/_2 = 1$$

PEARSON

Definition

- Note that to get the **population mean** of the random variable $\boldsymbol{x}$, we multiply each possible value of x by its probability $\boldsymbol{p}(\boldsymbol{x})$, and then we **sum this product** over all possible values of x.
- The **mean of $\boldsymbol{x}$** is also referred to as the **expected value** of x, denoted $\boldsymbol{E}(\boldsymbol{x})$.

The **mean,** or **expected value,** of a discrete random variable x is

$$\mu = E(x) = \Sigma x p(x)$$

- *Expected* is a mathematical term and should not be interpreted as it is typically used. Specifically, *a random variable might never be equal to its "expected value."*

PEARSON

Problem

- Suppose you work for an insurance company and you sell a **$10,000** one-year term insurance policy at an annual premium of **$290**. Actuarial tables show that the probability of death during the next year for a person of your customer's age, sex, health, etc., is **.001**. What is the expected gain (amount of money made by the company) for a policy of this type?

Solution

- The experiment is to observe whether the customer survives the upcoming year. The probabilities associated with the two sample points, *Live* and *Die*, are **.999** and **.001**, respectively. The random variable you are interested in is the gain x, which can assume the values shown in the following table:

Gain *x*	Sample Point	Probability
\$290	Customer lives	.999
−\$9,710	Customer dies	.001

- If the customer lives, the company gains the \$290 premium as profit. If the customer dies, the gain is negative because the company must pay \$10,000, for a net "gain" of $\$(290 - 10{,}000) = -\$9{,}710$. The expected gain is therefore

$$\boldsymbol{\mu = E(x) = \Sigma x p(x) = (290)(.999) + (-9{,}710)(.001) = \$280}$$

Definition

- We learned that the mean and other measures of central tendency tell only part of the story about a set of data.
- We need to **measure variability** as well.
- Since a probability distribution can be viewed as a representation of a population, we will use the population variance to measure its variability.

The **variance** of a random variable x is

$$\sigma^2 = E[(x-\mu)^2] = \Sigma(x-\mu)^2 p(x) = \Sigma x^2 p(x) - \mu^2$$

Definition

- Variance is also called the ***expected value of the squared distance from the mean***.
- The standard deviation of x is defined as the square root of the variance σ^2.

The **standard deviation** of a discrete random variable is equal to the square root of the variance, or $\sigma = \sqrt{\sigma^2}$.

Procedure

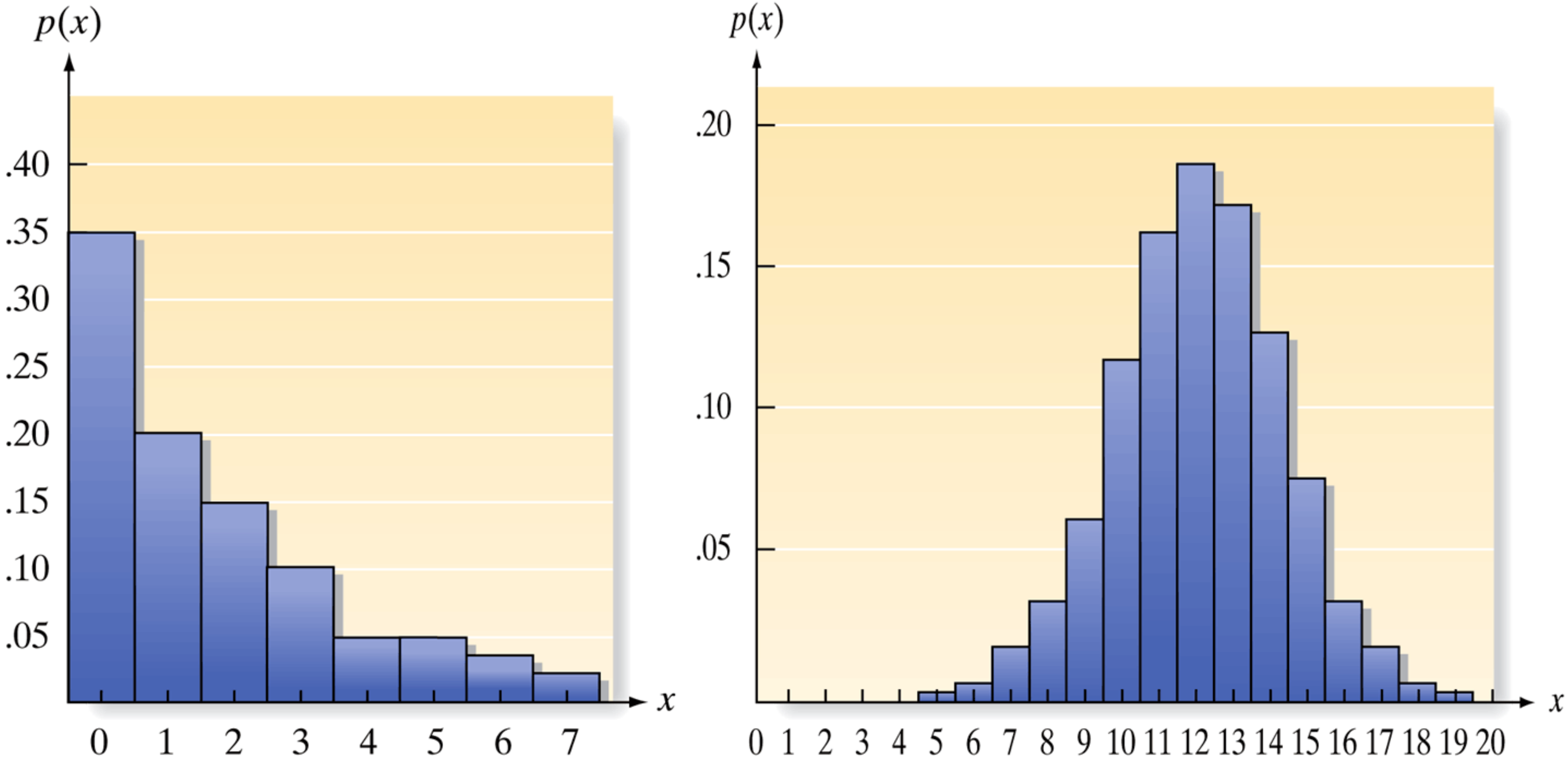
Probability Rules for a Discrete Random Variable

Let x be a discrete random variable with probability distribution $p(x)$, mean μ, and standard deviation σ. Then, depending on the shape of $p(x)$, the following probability statements can be made:

	Chebyshev's Rule	Empirical Rule
	Applies to any probability distribution (see Figure 4.6a)	Applies to probability distributions that are mound shaped and symmetric (see Figure 4.6b)
$P(\mu - \sigma < x < \mu + \sigma)$	≥ 0	$\approx .68$
$P(\mu - 2\sigma < x < \mu + 2\sigma)$	$\geq 3/4$	$\approx .95$
$P(\mu - 3\sigma < x < \mu + 3\sigma)$	$\geq 8/9$	≈ 1.00

PEARSON

Figure 4.6 Shapes of two probability distributions for a discrete random variable *x*



a. Skewed distribution

b. Mound shaped, symmetric

Problem

- Medical research has shown that a certain type of chemotherapy is successful 70% of the time when used to treat skin cancer. Suppose five skin cancer patients are treated with this type of chemotherapy, and let x equal the number of successful cures out of the five. The probability distribution for the number x of successful cures out of five is given in the following table:

x	0	1	2	3	4	5
$p(x)$	.002	.029	.132	.309	.360	.168

Solution

x	0	1	2	3	4	5
$p(x)$	.002	.029	.132	.309	.360	.168

a) Find $\mu = E(x)$. Interpret the result.

- Applying the formula for μ, we obtain

  $\mu = E(x) = \Sigma xp(x)$

  $= 0(.002) + 1(.029) + 2(.132) + 3(.309)$

  $+ 4(.360) + 5(.168) = 3.50$

On average, the number of successful cures out of five skin cancer patients treated with chemotherapy will equal 3.5.

Solution

x	0	1	2	3	4	5
$p(x)$	.002	.029	.132	.309	.360	.168

b) Find $\sigma = \sqrt{E[(x-\mu)^2]}$. Interpret the result.

Now we calculate the variance of x:

$$\begin{aligned}\sigma^2 &= E[(x-\mu)^2] = \Sigma(x-\mu)^2 p(x)\\ &= (0-3.5)^2(.002) + (1-3.5)^2(.029) + (2-3.5)^2(.132)\\ &\quad + (3-3.5)^2(.309) + (4-3.5)^2(.360) + (5-3.5)^2(.168)\\ &= 1.05\end{aligned}$$

Thus, the standard deviation is

$$\sigma = \sqrt{\sigma^2} = \sqrt{1.05} = 1.02$$

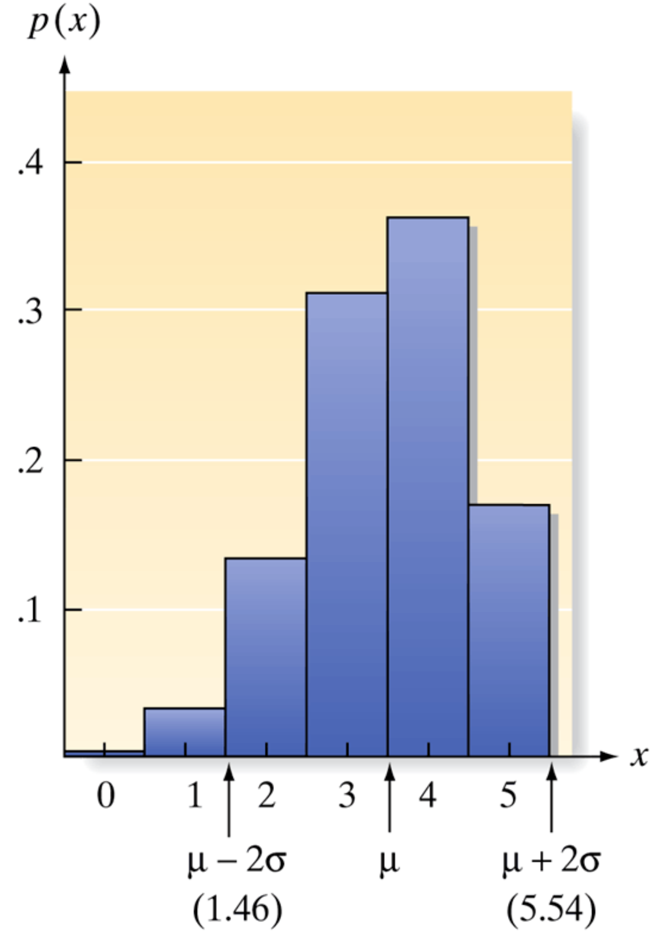
Solution

x	0	1	2	3	4	5
$p(x)$	.002	.029	.132	.309	.360	.168

c) Graph $p(x)$. Locate μ and the interval $\mu \pm 2\sigma$ on the graph. Use either Chebyshev's rule or the empirical rule to approximate the probability that x falls into this interval. Compare your result with the actual probability.

Figure 4.7 Graph of $p(x)$ for Example 4.8

- We expect (from Chebyshev's rule) at least 75% and, more likely (from the empirical rule), approximately 95% of observed x values to fall between 1.46 and 5.54.
- probability that x falls in the interval $\mu \pm 2\sigma$ includes the sum of $p(x)$ for the values $x = 2,\ x = 3,\ x = 4$, and $x = 5$, i.e., **.969**.
- Therefore, 96.9% of the probability distribution lies within two standard deviations of the mean.



PEARSON

Solution

x	0	1	2	3	4	5
$p(x)$	.002	.029	.132	.309	.360	.168

d) Would you expect to observe fewer than two successful cures out of five?

- Fewer than two successful cures out of five implies that $x = 0$ or $x = 1$. Both of these values of x lie outside the interval $\mu \pm 2\sigma$, and the empirical rule tells us that such a result is unlikely (approximate probability of .05).
- The exact probability is **.031**.
- We would not expect to observe fewer than two successful cures.

4.4

The Binomial Random Variable

The Binomial Random Variable

- Many experiments result in dichotomous responses
  - Yes–No, Pass–Fail, or Male–Female
- Analogous to coin toss
  - Two outcomes: Head or Tail

Definition

Characteristics of a Binomial Random Variable

1. The experiment consists of n identical trials.
2. There are only two possible outcomes on each trial. We will denote one outcome by S (for Success) and the other by F (for Failure).
3. The probability of S remains the same from trial to trial. This probability is denoted by p, and the probability of F is denoted by $q = 1 - p$.
4. The trials are independent.
5. The binomial random variable x is the number of S's in n trials.

Problem: For the following examples, decide whether x is a binomial random variable:

a) A university scholarship committee must select two students to receive a scholarship for the next academic year. The committee receives 10 applications for the scholarships—6 from male students and 4 from female students. Suppose the applicants are all equally qualified, so that the selections are randomly made. Let x be the number of female students who receive a scholarship.

Solution

- In checking the binomial characteristics, a problem arises with both characteristic 3 (probabilities the same across trials) and characteristic 4 (independence).
- On the one hand, given that the first student selected is female, the probability that the second chosen is female is $3/9$. On the other hand, given that the first selection is a male student, the probability that the second is female is $4/9$.
- The trials are therefore *dependent* (***not binomial***).

Problem: For the following examples, decide whether x is a binomial random variable:

b) Before marketing a new product on a large scale, many companies conduct a consumer-preference survey to determine whether the product is likely to be successful. Suppose a company develops a new diet soda and then conducts a taste-preference survey in which 100 randomly chosen consumers state their preferences from among the new soda and the two leading sellers. Let x be the number of the 100 who choose the new brand over the two others.

Solution

- Surveys that produce dichotomous responses and use random-sampling techniques are classic examples of binomial experiments.
- In this example, each randomly selected consumer either states a preference for the new diet soda or does not.
- The sample of 100 consumers is a very small proportion of the totality of potential consumers, so the response of one would be, for all practical purposes, independent of another.
- Thus, x **is a binomial random variable**.

Problem: For the following examples, decide whether x is a binomial random variable:

c) Some surveys are conducted by using a method of sampling other than simple random sampling. For example, suppose a television cable company plans to conduct a survey to determine the fraction of households in a certain city that would use its new fiber-optic service (FiOS). The sampling method is to choose a city block at random and then survey every household on that block (called cluster sampling). Suppose 10 blocks are so sampled, producing a total of 124 household responses. Let x be the number of the 124 households that would use FiOS.

Solution

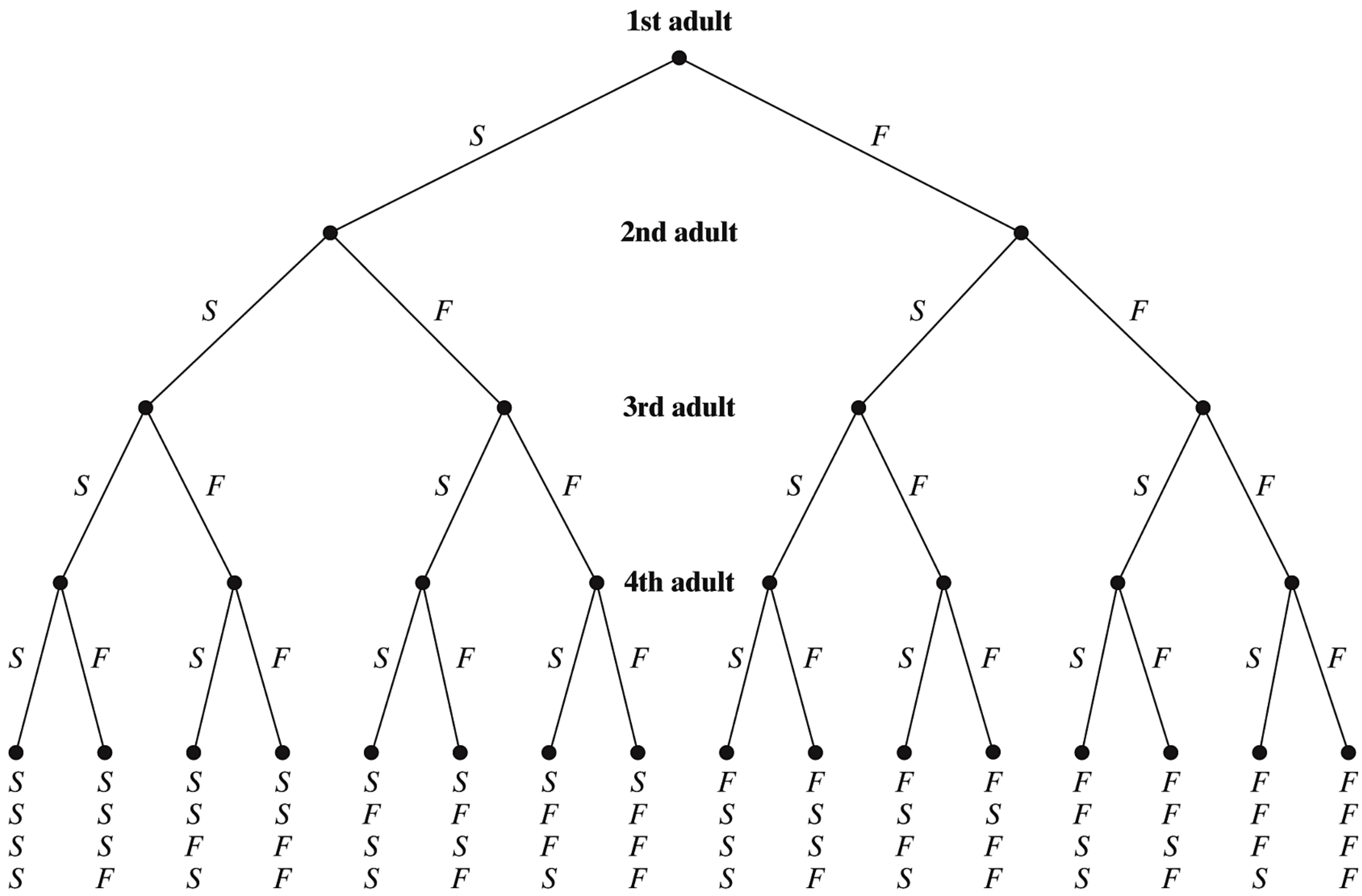
- This example is a survey with dichotomous responses (Yes or No to using FiOS), but the sampling method is not simple random sampling. Again, the binomial characteristic of independent trials would probably not be satisfied. The responses of households within a particular block would be dependent, since households within a block tend to be similar with respect to income, level of education, and general interests.
- Thus, the binomial model would not be satisfactory for x if the cluster sampling technique were employed.

Problem

The Heart Association claims that only 10% of U.S. adults over 30 years of age meet the minimum requirements established by the President's Council on Fitness, Sports, and Nutrition. Suppose four adults are randomly selected and each is given the fitness test.

a) Find the probability that none of the four adults passes the test.
b) Find the probability that three of the four adults pass the test.
c) Let x represent the number of the four adults who pass the fitness test. Explain why x is a binomial random variable.
d) Use the answers to parts a and b to derive a formula for $\mathrm{p}(x)$, the probability distribution of the binomial random variable x.

Solution - Figure 4.8 Tree Diagram Showing Outcomes for Fitness Tests



Solution - Table 4.2

Table 4.2	Sample Points for Fitness Test of Example 4.10			
SSSS	*FSSS*	*FFSS*	*SFFF*	*FFFF*
	SFSS	*FSFS*	*FSFF*	
	SSFS	*FSSF*	*FFSF*	
	SSSF	*SFFS*	*FFFS*	
		SFSF		
		SSFF		

$$P(\text{All four adults fail}) = P(FFFF) = (.9)^4 = .6561$$

$$\begin{aligned} P(\text{3 of 4 adults pass}) &= P(FSSS) + P(SFSS) + P(SSFS) + P(SSSF) \\ &= (.1)^3(.9) + (.1)^3(.9) + (.1)^3(.9) + (.1)^3(.9) \\ &= 4(.1)^3(.9) = .0036 \end{aligned}$$

Solution

c) We can characterize this experiment as consisting of four identical trials: the four test results. There are two possible outcomes to each trial, **S or F**, and the probability of passing, $p = .1$, is the same for each trial. Finally, we are assuming that each adult's test result is independent of all the others', so that the four trials are independent. Then it follows that x, the number of the four adults who pass the fitness test, **is a binomial random variable**.

Solution

d) The event probabilities in parts a and b provide insight into the formula for the probability distribution $p(x)$. Consider the event that three adults pass.

$$P(x = 3) =$$
$$(\text{Number of sample points for which } x = 3) \times (.1)^{\text{Number of successes}} \times (.9)^{\text{Number of failures}}$$
$$= 4(.1)^3(.9)^1$$

In general,

Number of sample points for which $x = 3$

$$= \text{Number of different ways of selecting 3 successes in the 4 trials}$$
$$= \binom{4}{3} = \frac{4!}{3!(4-3)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{(3 \cdot 2 \cdot 1) \cdot 1} = 4$$

The formula that works for any value of x can be deduced as follows:

$$P(x = 3) = \binom{4}{3}(.1)^3(.9)^1 = \binom{4}{x}(.1)^x(.9)^{4-x}$$

Procedure

The Binomial Probability Distribution

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

where

$$\begin{aligned}
p &= \text{Probability of a success on a single trial} \\
q &= 1 - p \\
n &= \text{Number of trials} \\
x &= \text{Number of successes in } n \text{ trials} \\
n - x &= \text{Number of failures in } n \text{ trials} \\
\binom{n}{x} &= \frac{n!}{x!(n-x)!}
\end{aligned}$$

Problem

- Refer to previous example.
- Use the formula for a binomial random variable to find the probability distribution of x, where x is the number of adults who pass the fitness test.
- Graph the distribution.

Solution

For this application, we have $n = 4$ trials. Since a success S is defined as an adult who passes the test, $p = P(S) = .1$ and $q = 1 - p = .9$. Substituting $n = 4$, $p = .1$, and $q = .9$ into the formula for $p(x)$, we obtain

$$p(0) = \frac{4!}{0!(4-0)!}(.1)^0(.9)^{4-0} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{(1)(4 \cdot 3 \cdot 2 \cdot 1)}(.1)^0(.9)^4 = 1(.1)^0(.9)^4 = .6561$$

$$p(1) = \frac{4!}{1!(4-1)!}(.1)^1(.9)^{4-1} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{(1)(3 \cdot 2 \cdot 1)}(.1)^1(.9)^3 = 4(.1)(.9)^3 = .2916$$

$$p(2) = \frac{4!}{2!(4-2)!}(.1)^2(.9)^{4-2} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{(2 \cdot 1)(2 \cdot 1)}(.1)^2(.9)^2 = 6(.1)^2(.9)^2 = .0486$$

$$p(3) = \frac{4!}{3!(4-3)!}(.1)^3(.9)^{4-3} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{(3 \cdot 2 \cdot 1)(1)}(.1)^3(.9)^1 = 4(.1)^3(.9) = .0036$$

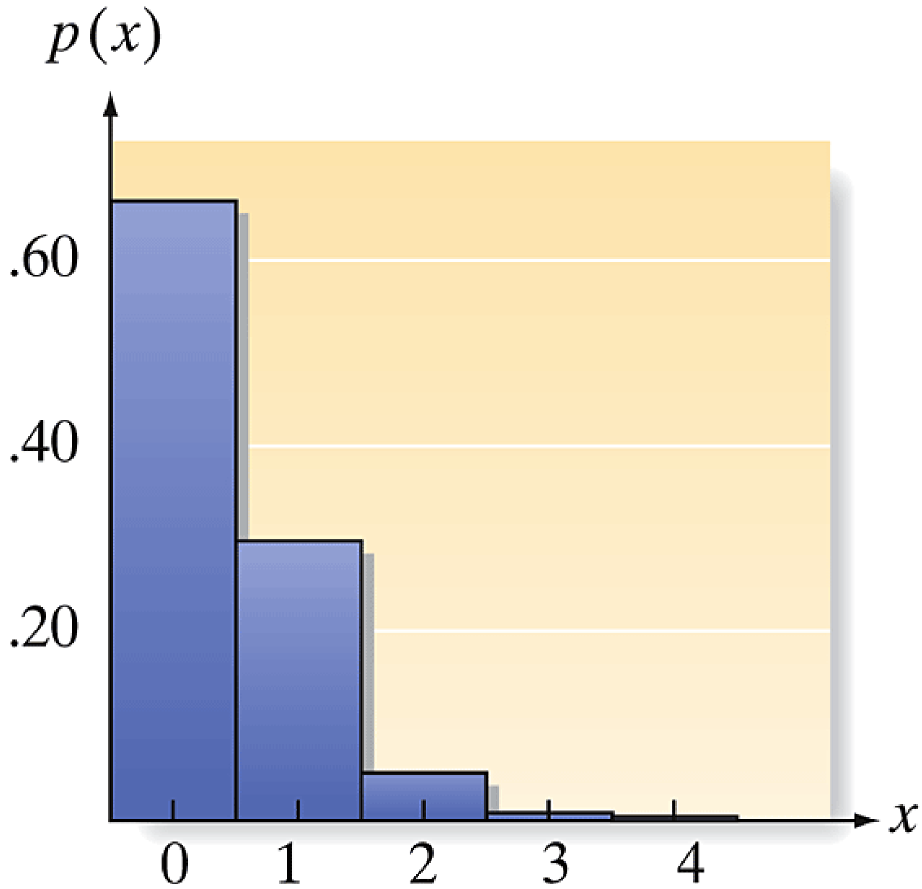
$$p(4) = \frac{4!}{4!(4-4)!}(.1)^4(.9)^{4-4} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{(4 \cdot 3 \cdot 2 \cdot 1)(1)}(.1)^4(.9)^0 = 1(.1)^4(.9) = .0001$$

PEARSON

Table 4.3

Table 4.3	Probability Distribution for Physical Fitness Example: Tabular Form
x	$p(x)$
0	.6561
1	.2916
2	.0486
3	.0036
4	.0001

Figure 4.9 Probability distribution for physical fitness example: Graphical form



Problem

- Refer to previous examples.
- Calculate μ and σ, the mean and standard deviation, respectively, of the number of the four adults who pass the test. Interpret the results.

Solution

- We know that the mean of a discrete probability distribution is $\mu = \sum xp(x)$
- So,

$$\mu = 0(.6561) + 1(.2916) + 2(.0486) + 3(.0036) + 4(.0001) = .4$$
$$= 4(.1) = np$$

- The variance is

$$\begin{aligned}\sigma^2 &= \Sigma(x-\mu)^2 p(x) = \Sigma(x-.4)^2 p(x) \\ &= (0-.4)^2(.6561) + (1-.4)^2(.2916) + (2-.4)^2(.0486) \\ &\quad + (3-.4)^2(.0036) + (4-.4)^2(.0001) \\ &= .104976 + .104976 + .124416 + .024336 + .001296 \\ &= .36 = 4(.1)(.9) = npq\end{aligned}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{.36} = .6$$

Definition

Mean, Variance, and Standard Deviation for a Binomial Random Variable

Mean: $\mu = np$

Variance: $\sigma^2 = npq$

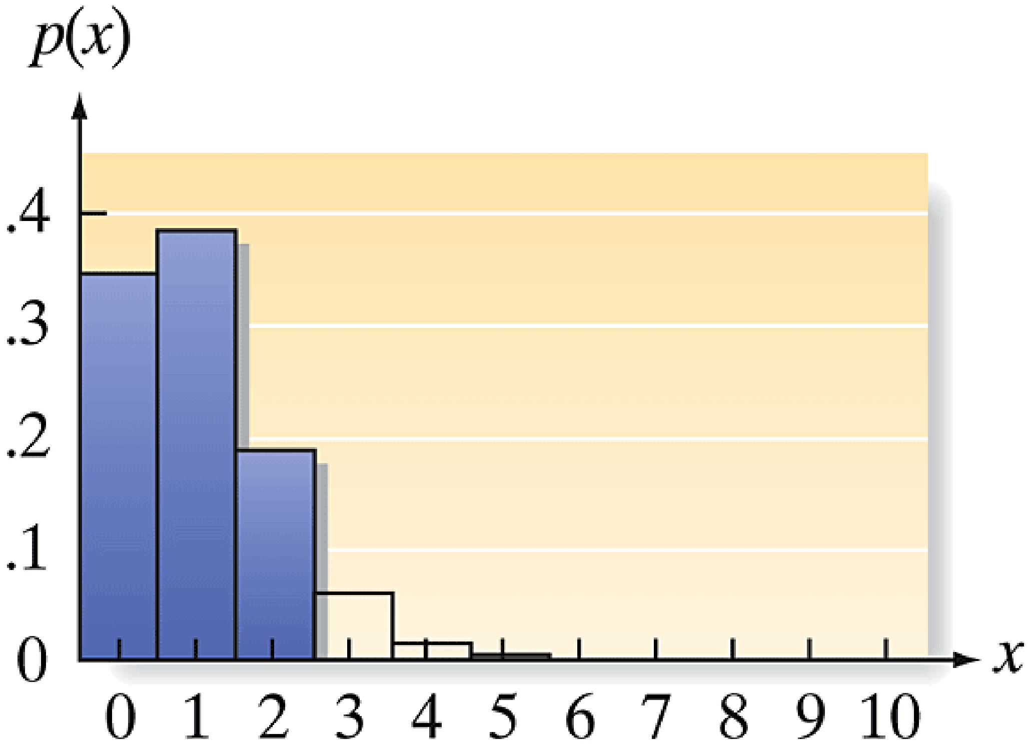
Standard deviation: $\sigma = \sqrt{npq}$

Table 4.4

Table 4.4 Reproduction of Part of Table I of Appendix B: Binomial Probabilities for $n = 10$

k \ p	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0	.904	.599	.349	.107	.028	.006	.001	.000	.000	.000	.000	.000	.000
1	.996	.914	.736	.376	.149	.046	.011	.002	.000	.000	.000	.000	.000
2	1.000	.988	.930	.678	.383	.167	.055	.012	.002	.000	.000	.000	.000
3	1.000	.999	.987	.879	.650	.382	.172	.055	.011	.001	.000	.000	.000
4	1.000	1.000	.998	.967	.850	.633	.377	.166	.047	.006	.000	.000	.000
5	1.000	1.000	1.000	.994	.953	.834	.623	.367	.150	.033	.002	.000	.000
6	1.000	1.000	1.000	.999	.989	.945	.828	.618	.350	.121	.013	.001	.000
7	1.000	1.000	1.000	1.000	.998	.988	.945	.833	.617	.322	.070	.012	.000
8	1.000	1.000	1.000	1.000	1.000	.988	.989	.954	.851	.624	.264	.086	.004
9	1.000	1.000	1.000	1.000	1.000	1.000	.999	.994	.972	.893	.651	.401	.096

Figure 4.10 Binomial probability distribution for n=10 and p=.10, with $P(x \le 2)$ highlighted



4.5

The Poisson Random Variable (Optional)

The Poisson Random Variable

- A type of probability distribution that is often useful in describing the number of rare events that will occur during a specific period or in a specific area or volume
  - The number of traffic accidents per month at a busy intersection
  - The number of noticeable surface defects (scratches, dents, etc.) found by quality inspectors on a new automobile
  - The number of parts per million (ppm) of some toxin found in the water or air emissions from a manufacturing plant
  - The number of diseased trees per acre of a certain woodland
  - The number of death claims received per day by an insurance company
  - The number of unscheduled admissions per day to a hospital

PEARSON

Definition

Characteristics of a Poisson Random Variable

1. The experiment consists of counting the number of times a certain event occurs during a given unit of time or in a given area or volume (or weight, distance, or any other unit of measurement).
2. The probability that an event occurs in a given unit of time, area, or volume is the same for all the units.
3. The number of events that occur in one unit of time, area, or volume is independent of the number that occur in other units.
4. The mean (or expected) number of events in each unit is denoted by the Greek letter lambda (λ).

Procedure

Probability Distribution, Mean, and Variance for a Poisson Random Variable

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (x = 0, 1, 2, \ldots) \ \mu = \lambda \quad \sigma^2 = \lambda$$

where

λ = Mean number of events during a given unit of time, area, volume, etc.

$e = 2.71828\ldots$

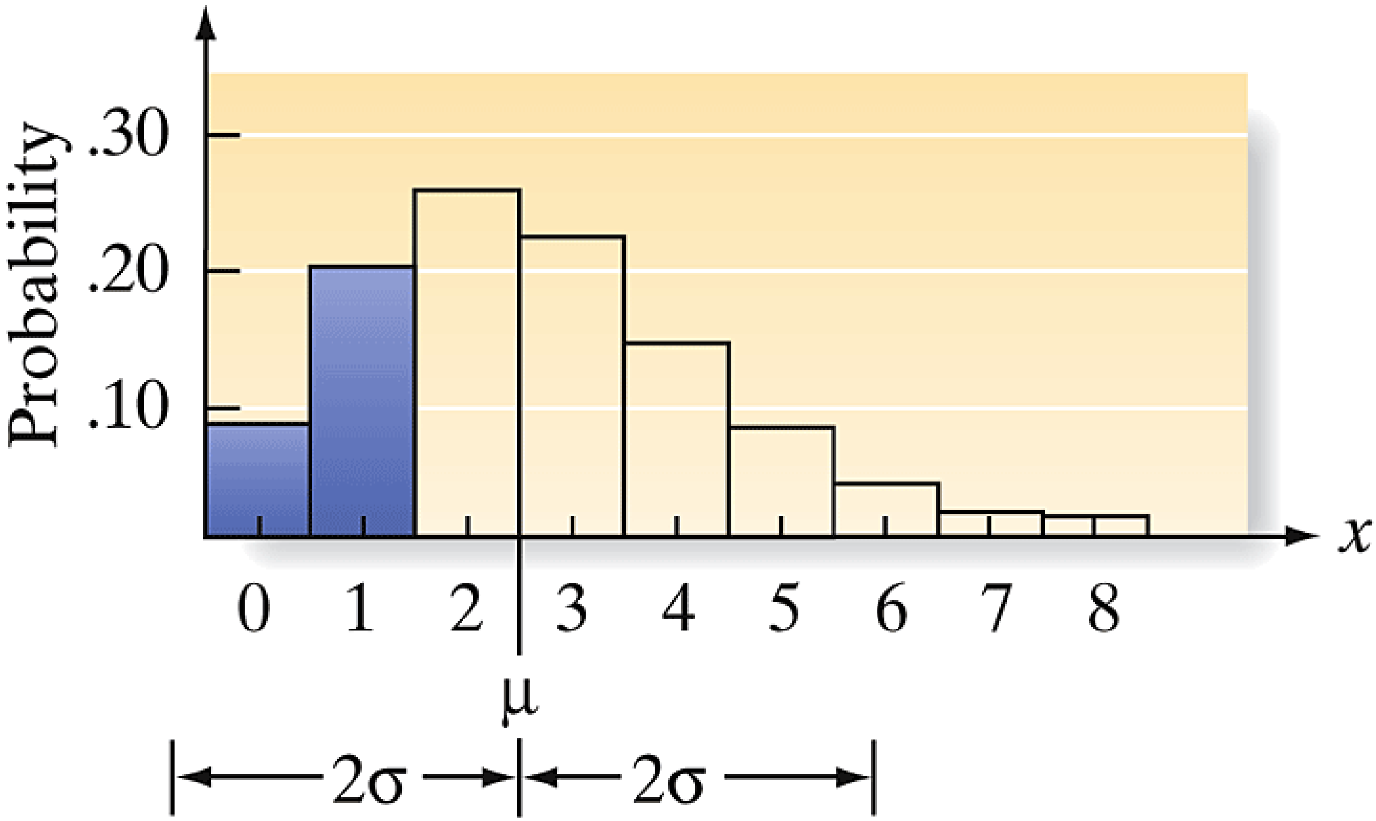
Problem

- Ecologists often use the number of reported sightings of a rare species of animal to estimate the remaining population size. For example, suppose the number x of reported sightings per week of blue whales is recorded. Assume that x has (approximately) a Poisson probability distribution. Furthermore, assume that the average number of weekly sightings is 2.6.
  - Find the mean and standard deviation of x, the number of blue-whale sightings per week. Interpret the results.

Solution

- The mean and variance of a Poisson random variable are both equal to λ.
  $\mu = \lambda = 2.6 \quad \sigma^2 = \lambda = 2.6 \rightarrow \sigma = 1.61$

Figure 4.13 Probability distribution for number of blue-whale sightings



4.6

The Hypergeometric Random Variable (Optional)

Definition

Characteristics of a Hypergeometric Random Variable

1. The experiment consists of randomly drawing n elements without replacement from a set of N elements, r of which are S's (for Success) and $(N - r)$ of which are F's (for Failure).
2. The hypergeometric random variable x is the number of S's in the draw of n elements.

- Hypergeometric and binomial characteristics stipulate that each draw, or trial, results in one of two outcomes.
- The basic difference between these random variables is that the hypergeometric trials are dependent, while the binomial trials are independent.

Procedure

Probability Distribution, Mean, and Variance of the Hypergeometric Random Variable

$$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \left[x = \text{Maximum}\left[0, n-(N-r)\right], \ldots, \text{Minimum}(r, n)\right]$$

$$\mu = \frac{nr}{N} \qquad \sigma^2 = \frac{r(N-r)n(N-n)}{N^2(N-1)}$$

where

N = Total number of elements

r = Number of S's in the N elements

n = Number of elements drawn

x = Number of S's drawn in the n elements

Figure 4.15 MINITAB Output for Example 4.15

Probability Density Function

```
Hypergeometric with N = 10, M = 4, and n = 3

x   P( X = x )
0     0.166667
```

Problem

- Suppose a professor randomly selects three new teaching assistants from a total of 10 applicants—six male and four female students. Let x be the number of females who are hired.
  a) Find the mean and standard deviation of x.
  b) Find the probability that no females are hired.

Solution (a)

- Since x is a hypergeometric random variable with $N = 10$, $n = 3$, and $r = 4$, the mean and variance are

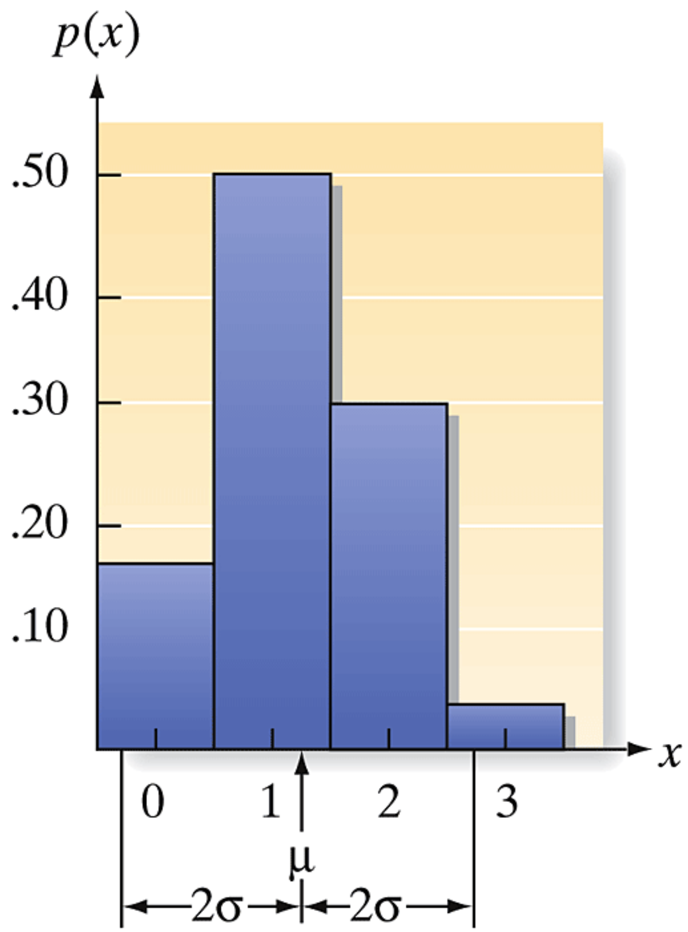
$$\mu = \frac{nr}{N} = \frac{(3)(4)}{10} = 1.2$$

$$\sigma^2 = \frac{r(N-r)n(N-n)}{N^2(N-1)} = \frac{4(10-4)3(10-3)}{(10)^2(10-1)}$$

$$= \frac{(4)(6)(3)(7)}{(100)(9)} = .56$$

$$\sigma = \sqrt{.56} = .75$$

Figure 4.16 Probability distribution for *x* in Example 4.15



Solution (b)

- The probability that no female students are hired by the professor, assuming that the selection is truly random, is

$$P(x = 0) = p(0) = \frac{\binom{4}{0}\binom{10-4}{3-0}}{\binom{10}{3}}$$

$$= \frac{\dfrac{4!}{0!(4-0)!}\dfrac{6!}{3!(6-3)!}}{\dfrac{10!}{3!(10-3)!}} = \frac{(1)(20)}{120} = \frac{1}{6} = .1667$$

Key Formulas

Random Variable	Probability Distribution	Mean	Variance
General Discrete:	Table, formula, or graph for $p(x)$	$\sum_{\text{all } x} x \cdot p(x)$	$\sum_{\text{all } x} (x-\mu)^2 \cdot p(x)$
Binomial:	$p(x) = \binom{n}{x} p^x q^{n-x}$	np	npq
Poisson:	$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$	λ	λ
Hypergeometric:	$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{x}}$	$\frac{nr}{N}$	$\frac{r(N-r)\,n(N-n)}{N^2\,(N-1)}$