Chapter 7: Random Variables

Section 7.1
Discrete and Continuous Random Variables

The Practice of Statistics, 4th edition – For AP*
STARNES, YATES, MOORE

Chapter 7
Random Variables

- 7.1 **Discrete and Continuous Random Variables**
- 7.2 Transforming and Combining Random Variables
- 7.2 Binomial and Geometric Random Variables

Section 7.1
Discrete and Continuous Random Variables

Learning Objectives

After this section, you should be able to…

- ✓ APPLY the concept of discrete random variables to a variety of statistical settings
- ✓ CALCULATE and INTERPRET the mean (expected value) of a discrete random variable
- ✓ CALCULATE and INTERPRET the standard deviation (and variance) of a discrete random variable
- ✓ DESCRIBE continuous random variables

Discrete and Continuous Random Variables

Random Variable and Probability Distribution

A **probability model** describes the possible outcomes of a chance process and the likelihood that those outcomes will occur.

A numerical variable that describes the outcomes of a chance process is called a **random variable**. The probability model for a random variable is its probability distribution

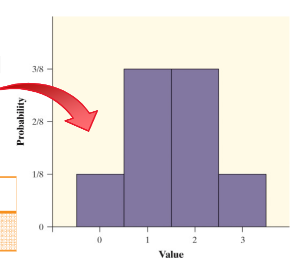
Definition:

A **random variable** takes numerical values that describe the outcomes of some chance process. The **probability distribution** of a random variable gives its possible values and their probabilities.

Example: Consider tossing a fair coin 3 times. Define X = the number of heads obtained

X = 0: TTT
X = 1: HTT THT TTH
X = 2: HHT HTH THH
X = 3: HHH

Value	0	1	2	3
Probability	1/8	3/8	3/8	1/8



RANDOM VARIABLE

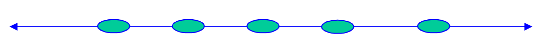
A **random variable** takes numerical values determined by the outcome of a chance process.

For example:

- The number of heads when 4 coins are tossed
- The salary of a randomly selected employee
- The age of a randomly selected math student
- The amount of gas in a randomly selected car
- The market value of a randomly selected home

Types of Random Variables

A random variable is discrete if the number of possible outcomes is finite or countable. Discrete random variables are determined by a count.



A random variable is continuous if it can take on any value within an interval. The possible outcomes cannot be listed. Continuous random variables are determined by a measure.

Larson/Farber Ch. 4

Types of Random Variables

Identify each random variable as discrete or continuous.

x = The number of people in a car
Discrete – *you count the number of people in a car 0, 1, 2, 3… Possible values can be listed.*

x = The gallons of gas bought in a week
Continuous – *you measure the gallons of gas. You cannot list the possible values.*

x = The time it takes to drive from home to school
Continuous – *you measure the amount of time. The possible values cannot be listed.*

x = The number of trips to school you make per week
Discrete – *you count the number of trips you make. The possible numbers can be listed.*

Larson/Farber Ch. 4

Discrete Probability Distributions

A **discrete probability distribution** lists each possible value of the random variable, together with its probability.

A survey asks a sample of families how many vehicles each owns.

number of vehicles → x

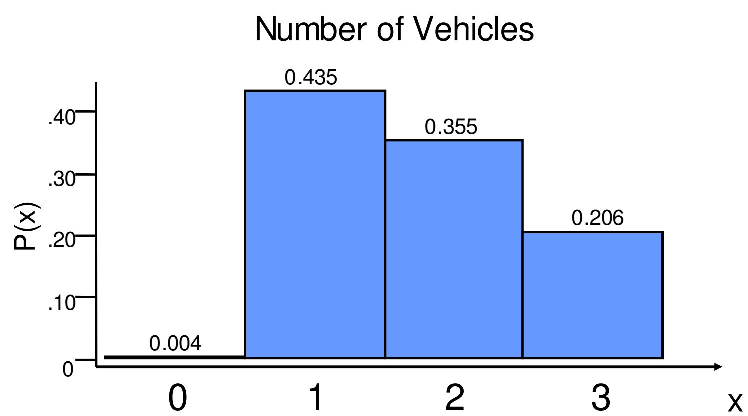
x	P(x)
0	0.004
1	0.435
2	0.355
3	0.206

Properties of a probability distribution

- Each probability must be between 0 and 1, inclusive.
  $$0 \leq P(x) \leq 1$$
- The sum of all probabilities is 1.
  $$\Sigma P(x) = 1$$

Larson/Farber Ch. 4

Probability Histogram



- The height of each bar corresponds to the probability of *x*.
- When the width of the bar is 1, the area of each bar corresponds to the probability the value of *x* will occur.

Larson/Farber Ch. 4

DISCRETE RANDOM VARIABLES

Examples of Discrete Random Variables :

- The number of girls in a random family
- The sum of the dice in a game of Monopoly
- The count of broken eggs in a random dozen
- The number of goals scored in a randomly selected soccer game
- The number of days in a randomly selected week that you went to a fast food restaurant

CONTINOUS RANDOM VARIABLES

- **Continuous Random Variables** take all values in some interval of numbers.
- These variables can take any value and are generally decimals.
- Continuous variables represent **measurable** quantities.
- A continuous probability distribution can be expressed using a density curve.

CONTINUOUS RANDOM VARIABLES

Examples of Continuous Random Variables :

- The GPA of a random college student
- The amount of water in a random bathtub
- The average cost of a gallon of gas on a random day in Simi Valley
- The weight of a randomly selected chihuahua
- The amount of time needed for a random athlete to complete a 100-meter race

WHICH TYPE OF RANDOM VARIABLE?

Are each of the following *discrete* or *continuous*?

- The number of defective light bulbs in a randomly selected box of 10 bulbs
- The amount of sugar in a random orange
- The height of a random 1st-grade boy
- The number of dogs in a random household
- The attendance in a random movie theater
- The length of a randomly selected song

DISCRETE VARIABLE DISTRIBUTIONS

For now, we will focus only on discrete variables. The distribution of a discrete variable is generally done in table form that consists of:

1. A list of the possible values that the variable can take
2. The probability of each of these values
   1. These probabilities must be between 0 and 1
   2. The sum of these probabilities must be 1

DISCRETE VARIABLE DISTRIBUTIONS

An example:

Suppose a random car on a freeway is selected and **X = the number of passengers in the car.**

The distribution could look like this:

X	1	2	3	4
P(X)	.57	.27	.13	.03

Note that each probability is between 0 and 1 and the sum of the probabilities is 1.

EXAMPLE

Apgar Scores: Babies' Health at Birth

Discrete random variables

In 1952, Dr. Virginia Apgar suggested five criteria for measuring a baby's health at birth: skin color, heart rate, muscle tone, breathing, and response when stimulated. She developed a 0-1-2 scale to rate a newborn on each of the five criteria. A baby's Apgar score is the sum of the ratings on each of the five scales, which gives a whole-number value from 0 to 10. Apgar scores are still used today to evaluate the health of newborns.

What Apgar scores are typical? To find out, researchers recorded the Apgar scores of over 2 million newborn babies in a single year.[2] Imagine selecting one of these newborns at random. (That's our chance process.) Define the random variable X = Apgar score of a randomly selected baby one minute after birth. The table below gives the probability distribution for X.

Example: Babies' Health at Birth

Read the example on page 343.

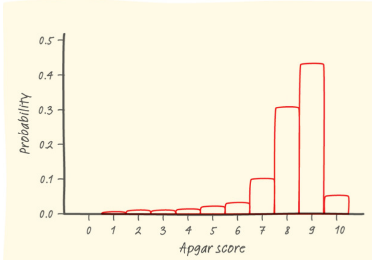
(a) **Show that the probability distribution for X is legitimate.**

(b) **Make a histogram of the probability distribution. Describe what you see.**

(c) **Apgar scores of 7 or higher indicate a healthy baby. What is $P(X \geq 7)$?**

Value:	0	1	2	3	4	5	6	7	8	9	10
Probability:	0.001	0.006	0.007	0.008	0.012	0.020	0.038	0.099	0.319	0.437	0.053

(a) All probabilities are between 0 and 1 and they add up to 1. This is a legitimate probability distribution.



(c) $P(X \geq 7) = .908$
We'd have a 91 % chance of randomly choosing a healthy baby.

(b) The left-skewed shape of the distribution suggests a randomly selected newborn will have an Apgar score at the high end of the scale. There is a small chance of getting a baby with a score of 5 or lower.

1. **Toss 4 times** Suppose you toss a fair coin 4 times. Let X = the number of heads you get.
   (a) Find the probability distribution of X.
   (b) Make a histogram of the probability distribution. Describe what you see.
   (c) Find $P(X \leq 3)$ and interpret the result.

3. **Spell-checking** Spell-checking software catches "nonword errors," which result in a string of letters that is not a word, as when "the" is typed as "teh." When undergraduates are asked to write a 250-word essay (without spell-checking), the number X of nonword errors has the following distribution:

Value of X:	0	1	2	3	4
Probability:	0.1	0.2	0.3	0.3	0.1

(a) Write the event "at least one nonword error" in terms of X. What is the probability of this event?

(b) Describe the event $X \leq 2$ in words. What is its probability? What is the probability that $X < 2$?

4. **Kids and toys** In an experiment on the behavior of young children, each subject is placed in an area with five toys. Past experiments have shown that the probability distribution of the number X of toys played with by a randomly selected subject is as follows:

Number of toys x_i:	0	1	2	3	4	5
Probability p_i:	0.03	0.16	0.30	0.23	0.17	0.11

(a) Write the event "plays with at most two toys" in terms of X. What is the probability of this event?

(b) Describe the event $X > 3$ in words. What is its probability? What is the probability that $X \geq 3$?

6. **Working out** Choose a person aged 19 to 25 years at random and ask, "In the past seven days, how many times did you go to an exercise or fitness center or work out?" Call the response Y for short. Based on a large sample survey, here is a probability model for the answer you will get:[6]

Days:	0	1	2	3	4	5	6	7
Probability:	0.68	0.05	0.07	0.08	0.05	0.04	0.01	0.02

(a) Show that this is a legitimate probability distribution.

(b) Make a histogram of the probability distribution. Describe what you see.

(c) Describe the event $Y < 7$ in words. What is $P(Y < 7)$?

(d) Express the event "worked out at least once" in terms of Y. What is the probability of this event?

9. **Keno** Keno is a favorite game in casinos, and similar games are popular with the states that operate lotteries. Balls numbered 1 to 80 are tumbled in a machine as the bets are placed, then 20 of the balls are chosen at random. Players select numbers by marking a card. The simplest of the many wagers available is "Mark 1 Number." Your payoff is \$3 on a \$1 bet if the number you select is one of those chosen. Because 20 of 80 numbers are chosen, your *probability of winning* is 20/80, or 0.25. Let X = the amount you gain on a single play of the game.

(a) Make a table that shows the probability distribution of X.

(b) Compute the expected value of X. Explain what this result means for the player.

10. **Fire insurance** Suppose a homeowner spends \$300 for a home insurance policy that will pay out \$200,000 if the home is destroyed by fire. Let Y = the profit made by the company on a single policy. From previous data, the probability that a home in this area will be destroyed by fire is 0.0002.

(a) Make a table that shows the probability distribution of Y.

(b) Compute the expected value of Y. Explain what this result means for the insurance company.