Chapter 6: Random Variables

Section 6.1

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.

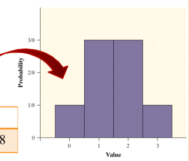
A numerical variable that describes the outcomes of a chance process is called a **random variable**. The probability model for a random variable is its probability distribution

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.

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



Discrete Random Variables

There are two main types of random variables: *discrete* and *continuous*. If we can find a way to list all possible outcomes for a random variable and assign probabilities to each one, we have a **discrete random variable**.

Discrete Random Variables and Their Probability Distributions

A **discrete random variable** X takes a fixed set of possible values with gaps between. The probability distribution of a discrete random variable X lists the values x_i and their probabilities p_i:

Values of X: $x_1 \quad x_2 \quad x_3 \quad \ldots x_k$
Probability: $p_1 \quad p_2 \quad p_3 \quad \ldots p_k$

The probabilities p_i must satisfy two requirements:

1. Every probability p_i is a number between 0 and 1.
2. The sum of the probabilities is 1. $p_1 + p_2 + p_3 + \ldots + p_k = 1$.

To find the probability of any event, add the probabilities p_i of the particular values x_i that make up the event.

Example 1: 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 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. Imagine selecting one of these newborns at random. 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.

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

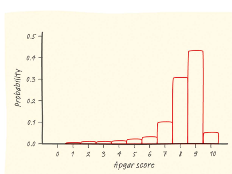
a) Show that the probability distribution for X is legitimate.

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

---

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

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.



---

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

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

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

Mean of a Discrete Random Variable
When analyzing discrete random variables, we'll follow the same strategy we used with quantitative data – describe the shape, center, and spread, and identify any outliers.

The mean of any discrete random variable is an average of the possible outcomes, with each outcome weighted by its probability.

Definition:
Suppose that X is a discrete random variable whose probability distribution is

Values of X:	x_1	x_2	x_3	$\ldots x_k$
Probability:	p_1	p_2	p_3	$\ldots p_k$

To find the **mean (expected value)** of X, multiply each possible value by its probability, then add all the products:

$$\mu_x = E(X) = x_1 p_1 + x_2 p_2 + x_3 p_3 + \ldots x_k p_k$$
$$= \sum x_i p_i$$

Example 2: In our earlier example, we defined the random variable X to be the Apgar score of a randomly selected baby. The table below gives the probability distribution for X once again. Compute the mean of the random variable X and interpret this value in context.

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

$$\mu_x = E(X) = \sum x_i p_i$$
$$= (0)(0.001) + (1)(0.006) + (2)(0.007) + \ldots + (10)(0.053)$$
$$= 8.128$$

The mean Apgar score of a randomly selected newborn is 8.128. This is the long-term average Apgar score of many, many randomly chosen babies.

Note: The expected value does not need to be a possible value of X or an integer! It is a long-term average over many repetitions.

AP EXAM TIP: If the mean of a random variable has a non-integer value, but you report it as an integer, your answer will be marked as incorrect.

Standard Deviation of a Discrete Random Variable
Since we use the mean as the measure of center for a discrete random variable, we'll use the standard deviation as our measure of spread. The definition of the **variance of a random variable** is similar to the definition of the variance for a set of quantitative data.

Definition:
Suppose that X is a discrete random variable whose probability distribution is

Values of X:	x_1	x_2	x_3	$\ldots x_k$
Probability:	p_1	p_2	p_3	$\ldots p_k$

and that μ_X is the mean of X. The **variance** of X is

$$\text{Var}(X) = \sigma_X^2 = (x_1 - \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + (x_3 - \mu_X)^2 p_3 + \ldots (x_k - \mu)^2 p_k$$
$$= \sum (x_i - \mu_X)^2 p_i$$

To get the **standard deviation of a random variable**, take the square root of the variance.

Example 3: In the last example, we calculated the mean Apgar score of a randomly chosen newborn to be $\mu_x = 8.128$. The table below gives the probability distribution for X one more time. Compute the standard deviation of the random variable X and interpret it in context.

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

$$\sigma_X^2 = \sum (x_i - \mu_X)^2 p_i$$
$$= (0-8.128)^2(0.001) + (1-8.128)^2(0.006) + \ldots + (10-8.128)^2(0.053)$$
$$= 2.066$$

Variance

$$\sigma_X = \sqrt{2.066} = 1.437$$

The standard deviation of X is 1.437. On average, a randomly selected baby's Apgar score will differ from the mean 8.128 by about 1.4 units.

Example 2 **Winning (and Losing) at Roulette**
Finding the mean of a discrete random variable

On an American roulette wheel, there are 38 slots numbered 1 through 36, plus 0 and 00. Half of the slots from 1 to 36 are red; the other half are black. Both the 0 and 00 slots are green. Suppose that a player places a simple \$1 bet on red. If the ball lands in a red slot, the player gets the original dollar back, plus an additional dollar for winning the bet. If the ball lands in a different-colored slot, the player loses the dollar bet to the casino.

Let's define the random variable X = net gain from a single \$1 bet on red. The possible values of X are −\$1 and \$1. (The player either gains a dollar or loses a dollar.) What are the corresponding probabilities? The chance that the ball lands in a red slot is 18/38. The chance that the ball lands in a different-colored slot is 20/38. Here is the probability distribution of X:

Value:	−\$1	\$1
Probability:	20/38	18/38

What is the player's average gain? The ordinary average of the two possible outcomes −\$1 and \$1 is \$0. But \$0 isn't the average winnings because the player is less likely to win \$1 than to lose \$1. In the long run, the player gains a dollar 18 times in every 38 games played and loses a dollar on the remaining 20 of 38 bets. The player's long-run average gain for this simple bet is

$$\mu_X = (-\$1)\left(\frac{20}{38}\right) + (\$1)\left(\frac{18}{38}\right) = -\$0.05$$

You see that the player loses (and the casino gains) an average of five cents per \$1 bet in many, many plays of the game.

CHECK YOUR UNDERSTANDING

A large auto dealership keeps track of sales made during each hour of the day. Let X = the number of cars sold during the first hour of business on a randomly selected Friday. Based on previous records, the probability distribution of X is as follows:

Cars sold:	0	1	2	3
Probability:	0.3	0.4	0.2	0.1

1. Compute and interpret the mean of X.

Show Answer

2. Compute and interpret the standard deviation of X.

Show Answer

CHECK YOUR UNDERSTANDING

A large auto dealership keeps track of sales made during each hour of the day. Let X = the number of cars sold during the first hour of business on a randomly selected Friday. Based on previous records, the probability distribution of X is as follows:

Cars sold:	0	1	2	3
Probability:	0.3	0.4	0.2	0.1

1. Compute and interpret the mean of X.

Hide Answer

Correct Answer
$\mu_X = 1.1$. If many, many Fridays are randomly selected, the average number of cars sold will be about 1.1.

2. Compute and interpret the standard deviation of X.

Hide Answer

Correct Answer
$\sigma_X = \sqrt{0.89} = 0.943$. The number of cars sold on a randomly selected Friday will typically vary from the mean (1.1) by about 0.943 cars.

Continuous Random Variables

Discrete random variables commonly arise from situations that involve counting something. Situations that involve measuring something often result in a **continuous random variable**.

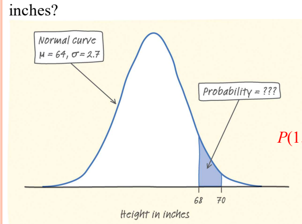
Definition:
A **continuous random variable** X takes on all values in an interval of numbers. The probability distribution of X is described by a **density curve**. The probability of any event is the area under the density curve and above the values of X that make up the event.

The probability model of a discrete random variable X assigns a probability between 0 and 1 to each possible value of X.

A continuous random variable Y has *infinitely many* possible values. All continuous probability models assign probability 0 to every individual outcome. Only *intervals* of values have positive probability.

In many cases, discrete random variables arise from counting something—for instance, the number of siblings that a randomly selected student has. Continuous random variables often arise from measuring something—for instance, height, SAT score, or blood pressure of a randomly selected student.

Example 4: The heights of young women closely follow the Normal distribution with mean $\mu = 64$ inches and standard deviation $\sigma = 2.7$ inches. This is a distribution for a large set of data. Now choose one young woman at random. Call her height Y. If we repeat the random choice very many times, the distribution of values of Y is the same Normal distribution that describes the heights of all young women. Define Y as the height of a randomly chosen young woman. Y is a continuous random variable whose probability distribution is $N(64, 2.7)$. What is the probability that a randomly chosen young woman has height between 68 and 70 inches?



$P(68 \le Y \le 70) = ???$

$z = \frac{68-64}{2.7} = 1.48$ $\qquad z = \frac{70-64}{2.7} = 2.22$

$P(1.48 \le z \le 2.22) = P(z \le 2.22) - P(z \le 1.48)$

$= 0.9868 - 0.9306$

$= 0.0562$

There is about a 5.6% chance that a randomly chosen young woman has a height between 68 and 70 inches.

AP EXAM TIP: When you solve problems involving random variables, start by defining the random variable of interest. For example, let X = the Apgar score of a randomly selected baby or let Y = the height of a randomly selected young woman. Then state the probability you're trying to find in terms of the random variable: $P(68 \le Y \le 70)$ or $P(X \ge 7)$.

Do you remember how to find area under the curve with your graphing calculator?

2nd Vars (DISTR) button
Normalcdf(lower limit, upper limit, mean, standard deviation)

Go back to the previous problems and use your calculator

EXAMPLE 5)
The weights of 3 year old females closely follows a normal distribution with a mean of 30.7 pounds and standard deviation of 3.6 pounds.

What is the probability that a randomly selected 3 year old female weighs at least 30 pounds? Draw a picture, show your work, and then check with calculator.

EXAMPLE 6)

A study of 12,000 able-bodied male students at U of I found that their times for the mile run were approximately normal with a mean of 7.11 minutes and standard deviation of .74 minutes.

Find the probability P(5.5 < M $\le$ 6 or 8.5 < M $\le$ 9.