# Unit 19: Probability Models

## SUMMARY OF VIDEO

Probability is the language of uncertainty. Using statistics, we can better predict the outcomes of random phenomena over the long term – from the very complex, like weather, to the very simple, like a coin flip, or of more interest to gamblers, a dice toss.

In a casino, the house always has the upper hand – even when the edge is very small, less than 1%, the casino makes money. Over thousands and thousands of gambles, that small edge starts to generate big revenues. Unlike most of their guests, casinos are playing the long game and not just hoping for a short-term windfall.

When a player first comes into the casino, his chance of beating the house is slightly less than 50-50. A smart gambler will come in and bet the maximum he can afford to lose on one roll and walk away either a winner or loser because the house edge is only around one half of one percent on that particular bet. So, he's got a forty-nine and a half percent chance of winning and the house has fifty and a half percent chance of winning. But most people won't play this way. While at the table, the more bets placed by the gambler, the more likely he is to lose in the long run.

Casinos count on the fact that over the very long term, they know with near certainty the average result of thousands and thousands of chance outcomes. That's why the house always wins in the end. Statisticians can create probability models to mathematically describe this fact.

Take, for example, the probability model for the popular dice game of craps. Depending on how players bet, money is made or lost based on the 'shooter' rolling certain numbers and not others. Assuming they are playing with fair dice, smart gamblers want to know the probability of any particular roll coming up. Here's where we start building that probability model. First, we define the sample space, $S$, the set of all possible outcomes. As you can see from Figure 19.1, rolling two six-sided dice means we have 36 possible outcomes.
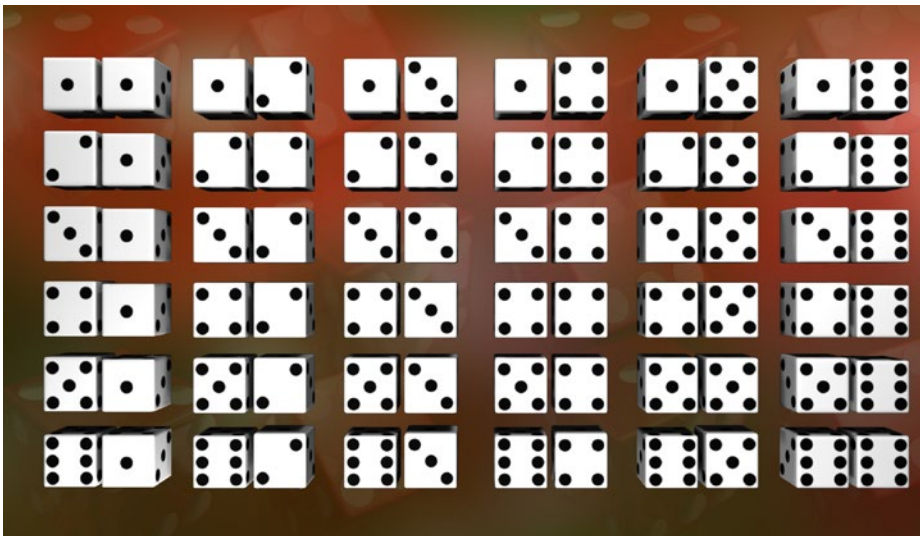
*Figure 19.1. The sample space for rolling two dice.*

Next, we assign probabilities to each of the possible outcomes in our sample space. Each roll is independent, meaning that the occurrence of one doesn't influence the probability of another. If the dice are perfectly balanced, all 36 outcomes are equally likely. In the long run, each of the outcomes would come up 1/36th of the rolls. So, the probability for each outcome is 1/36 or approximately 0.0278 – so, each outcome occurs roughly 3% of the time. Probabilities are always between 0 and 1, with those closer to 0 less likely to happen and those closer to 1 more likely to happen. The sum of the probabilities of all the possible outcomes in a sample space always equals 1.

For games such as craps, gamblers are more worried about the sum of the dice. So, the sample space they are interested in for rolling two dice looks like this:

$S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

While each roll pictured in Figure 19.1 has an equal chance of occurring, that's not true for each sum between 2 and 12. For example, there is only one way to roll a two and only one way to roll a twelve. So, each of those outcomes has a 1/36th chance of occurring. But there are six ways to roll a seven. So, the gambler has a 6/36th or 1/6th chance of rolling a seven, which is about a 17% chance. The probability model for how many spots are going to turn up when a player rolls two dice is given in Table 19.1. A probability model is made up of all the possible outcomes together with the probabilities associated with those outcomes.

| Spots | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Probability | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 6/36 | 5/36 | 4/36 | 3/36 | 2/36 | 1/36 |

*Table 19.1. Probability model for two dice.*

By tweaking the rules for what rolls pay out in what way, the casino can use the probability model to ensure that over the long term, no one will beat the house. For example, in craps the most common roll – a seven – is the one that instantly loses the round once it's underway.

Suppose we wanted to know the probability of rolling anything OTHER than a seven, $P$(not 7). We can use the Complement Rule to figure this out.

---

**Complement Rule**
For any event $A$, $P(A$ does not occur$) = 1 – P(A)$.

---

So, $P$(not 7) $= 1 – P(7) = 1 – 1/6 = 5/6$.

The probability model in Table 19.1 provides plenty of other examples as well. Let's say one gambler placed separate bets on 4 and 5. He wants the next roll to add up to one or the other of those two numbers. How do we determine his chances for winning? First of all, these two events, rolling a 4 or rolling a 5, are what statisticians call mutually exclusive events, which means that these two events have no outcomes in common. Because these events are mutually exclusive, we can use the Addition Rule for Mutually Exclusive Events to figure out the gambler's chances of winning.

---

**Addition Rule**
If $A$ and $B$ are mutually exclusive events, then $P(A$ or $B) = P(A) + P(B)$.

---

In this case, $P(4$ or $5) = P(4) + P(5) = 3/36 + 4/36 = 7/36$. So, our gambler has about a 19% chance of winning.

Craps players can also bet that the shooter will roll a number 'the hard way,' meaning by rolling doubles. Let's say one gambler bets the shooter will roll six the hard way. What are the chances that this bet will pay off on the next roll? We can figure this out using the Multiplication Rule, which says that if two events are independent, we can find the probability that they both happen by multiplying their individual probabilities.

---

**Multiplication Rule**
If $A$ and $B$ are independent, then $P(A$ and $B) = P(A)P(B)$.

---

The roll of one die is our event A and the roll of the other is our event B. The two events are independent because whatever is rolled on one die does not affect what is rolled on the other die.

Hence, we need to calculate

$P(3$ and $3) = P(3)P(3)$.

Rather than the probability model we have been using for rolling two dice at once, we need a new one for the probabilities of each number coming up in the roll of a single die. That model is shown in Table 19.2.

| Spots | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Probability | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

*Table 19.2. Probability model for one die.*

Now, we finish calculating the probability of rolling a six the hard way:

$P(3$ and $3) = P(3)P(3) = (1/6)(1/6) = 1/36$.

By now you should realize that no matter how skilled you get in using probability models, the house will probably win if you keep on playing.

# STUDENT LEARNING OBJECTIVES

A. Be able to list the outcomes in a sample space.

B. Be able to assign probabilities to individual outcomes in a sample space.

C. Know how to check that a proposed probability model is legitimate.

D. Understand the concepts of mutually exclusive and independent events and know the difference between them.

E. Know how to use the Complement, Addition, and Multiplication Rules of probability.

# CONTENT OVERVIEW

In this unit, we develop probability models, and then use those models to determine probabilities that certain events will occur. The video focused on probability models describing games of chance. However, probability models have applications that go way beyond casino gaming.

A probability model has two parts: a description of the sample space and a means of assigning probabilities. The **sample space** is the set of all possible outcomes of some random phenomenon. For example, suppose we want to study traffic patterns of vehicles approaching a particular corner. The vehicles reach the corner and can do one of the three things presented in the sample space below.

*S* = {go straight, turn right, turn left}

After studying the traffic patterns at this corner over a long period of time, we determine that cars go straight 60% of the time, turn right 25% of the time, and turn left 15% of the time. Now, we form a probability model by listing the elements of our sample space, together with their associated probabilities. (See Table 19.3.)

| Vehicle Direction | Straight | Right | Left |
|---|---|---|---|
| Probability | 0.6 | 0.25 | 0.15 |

*Table 19.3. Probability model for traffic patterns.*

Notice two things about the row labeled Probability in Table 19.3. The probabilities are numbers between 0 and 1 and the probabilities add up to 1. When presented with a probability model, always check that these two properties are satisfied.

Any subset of a sample space is called an event. For example, we could let event *A* be the outcome that a vehicle approaches the corner and then goes straight and event *B* that a vehicle turns right or turns left:

*A* = {Straight} and *B* = {Right, Left}

Notice the following about these two events: *B* contains all the outcomes in the sample space that are not in *A*; in other words, *B* = not *A*. Because of this fact, *A* and *B* are said to be **complementary events**. If two events are complementary, then there is a relationship between their probabilities. That relationship is spelled out by the Complement Rule.

---

**Complement Rule**
For any event *C*, *P*(not *C*) = 1 – *P*(*C*).

---

In this case, the probability model specifies that *P*(*A*) = 0.60. We can use the Complement Rule to figure out *P*(*B*):

*P*(*B*) = *P*(not *A*) = 1 – *P*(*A*) = 1 – 0.60 = 0.40.

Next, we work with a probability model for blood type. Human blood comes in different types. Each person has a specific ABO type (A, B, AB, or O) and Rh factor (positive or negative). Hence, if you are O+, your ABO type is O and your Rh factor is positive. Unlike the probabilities associated with rolling a fair die (each number has an equal chance of occurring), blood types are not uniformly distributed. A probability model for blood types in the United States is given in Table 19.4.

| Blood Type | A + | A - | B + | B - | AB + | AB - | O + | O - |
|---|---|---|---|---|---|---|---|---|
| Probability | 0.357 | 0.063 | 0.085 | 0.015 | 0.034 | 0.006 | 0.374 | 0.066 |

Table 19.4. Probability model for blood types in U.S.

Consider the following two events: let *E* be the event that a randomly chosen person has blood type A and *F* be the event that that person has blood type O:

*E* = {A+ or A-} and *F* = {O+ or O-}

Notice that the same person cannot have ABO blood type A and O at the same time. So, events *E* and *F* have no outcomes in common – mathematically, they are disjoint sets. Two events that have no outcomes in common are called **mutually exclusive events**. If we are able to break an event down into mutually exclusive events, then we can use the Addition Rule to figure out its probability.

---

> **Addition Rule**
>
> If *C* and *D* are mutually exclusive events, then *P*(*C* or *D*) = *P*(*C*) + *P*(*D*).

Since the same person can't have both A+ and A- blood type, we can use the Addition Rule to calculate the probability that *E* occurs as follows:

*P*(*E*) = *P*(A+ or A-) = *P*(A+) + *P*(A-) = 0.357 + 0.063 = 0.420.

Hence, roughly 42% of residents in the U.S. have ABO-type A blood.

Next, we tackle the problem of finding the probability that two randomly chosen U.S. residents both have ABO-type A blood. Since the two people were chosen at random, the fact that the first person is type A should not affect the chances that the second person is type A. Whenever the occurrence of one event does not affect the probability of the occurrence of the other, we say the two events are independent. To solve the problem at hand, we use the Multiplication Rule.

> **Multiplication Rule**
> If *C* and *D* are independent events, then *P*(*C* and *D*) = *P*(*C*)*P*(*D*).

Now we are ready to calculate the probability of both people being type A:

*P*(Person 1 is type A and Person 2 is type A)

$$= P(\text{Person 1 is type A})P(\text{Person 2 is type A})$$

$$=(0.420)(0.420)$$

$$= 0.1764.$$

Hence, there is roughly an 18% chance that two randomly selected U.S. residents will both have ABO-type A blood.

With two randomly chosen people, there are four possible outcomes:

1) Person 1 type A and Person 2 type A

2) Person 1 type A and Person 2 not type A

3) Person 1 not type A and Person 2 type A

4) Person 1 not type A and Person 2 not type A

Let event *C* be the outcome that at least one of the two people does not have type A blood. Then *C* consists of outcomes (2) – (4) above. We could calculate the probabilities of outcomes (2), (3), and (4) and then sum them to get the answer. However, an easier approach is to recognize that not *C* is the same as outcome (1) and use the Complement Rule:

*P*(*C*) = 1 – *P*(not *C*) = 1 – 0.1764 = 0.8236.


Therefore, in random samples of size two, we expect at least one of the two people not to be type A about 82% of the time.

# KEY TERMS

Two events are **mutually exclusive** if they have no outcomes in common. Two events are **independent** if the fact that one of the events occurs does not affect the probability that the other occurs. If two events are not independent, they are **dependent**. If events *A* and *B* are mutually exclusive and $P(B) > 0$, then events *A* and *B* are dependent. That's because if A occurs, then B cannot occur. In this case, knowing that *A* occurs changes *B*'s probability to zero.

Two events are **complementary** if they are mutually exclusive and combining their outcomes into a single set gives the entire sample space. *A* is the complement of *B* if *A* consists of all the outcomes in the sample space that are not in *B*; in other words, *A* = not *B*.

This unit covered three rules of probability:

1. Complement Rule

For any event *C*, $P(\text{not } C) = 1 - P(C)$.

2. Addition Rule

If *C* and *D* are mutually exclusive, then $P(C \text{ or } D) = P(C) + P(D)$.

3. Multiplication Rule

If *C* and *D* are independent, then $P(C \text{ and } D) = P(C)P(D)$.

# THE VIDEO

Take out a piece of paper and be ready to write down answers to these questions as you watch the video.

1.  What is a probability model?

2.  Describe the sample space for the sum of two dice.

3.  What is the probability of rolling two dice and getting a sum of seven?

4.  If you know the probability that event *A* occurs, how do you calculate the probability that event *A* does not occur?

5.  What probability can you find using the Addition Rule?

6.  What probability can you find using the Multiplication Rule?

# UNIT ACTIVITY:
## PROBABILITY MODELS AND DATA

The video introduced two probability models based on rolling dice. In this activity, you will compare those models with results gathered from repeatedly rolling the dice.

The probability model for rolling a single die is given in Table 19.2 (see Page 4). Since the outcome, number of spots on the die, is numeric, we can represent this model graphically with the probability histogram in Figure 19.2. In our probability histogram, the possible outcomes of rolling a single die appear on the horizontal axis and the bars are drawn so that their heights are the probabilities. In this case, all bars have equal heights, since each of the numbers 1, 2, 3, 4, 5, and 6 are equally likely to be rolled.
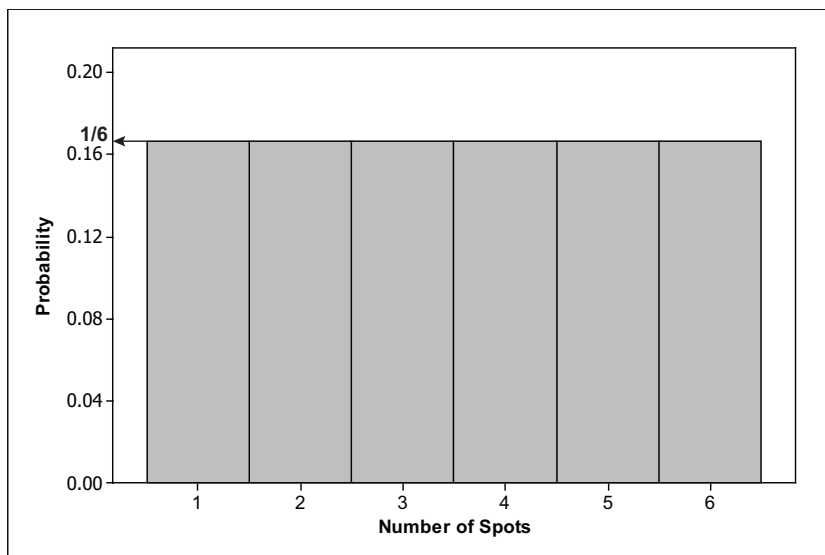


*Figure 19.2. Probability histogram for rolling a single die.*

The probability histogram gives us a model for how we expect histograms of data from a large number of rolls to look. Next, you will roll a die repeatedly and compare data tables to the probability model in Table 1.2 and histograms of data to the probability histogram in Figure 19.2.

1. a. Roll a die 100 times and record your results in a copy of Table 19.5.

| Number of Spots | Frequency | Relative Frequency |
|---|---|---|
| 1 | | |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |
| 6 | | |

*Table 19.5. Distribution of number of spots.*

b. Calculate the relative frequencies (proportions) and record them in your table. Are the relative frequencies close to the probability of 1/6? (Reminder: relative frequency = frequency/(total number of rolls).)

c. Make a histogram of your data using relative frequency (or proportion) for the vertical axis. Use the same scaling for the horizontal axis and vertical axis as was used in Figure 19.2. How closely does your histogram resemble the probability histogram?

2. Probabilities are relative frequencies (proportions) over the very long run, not just over 100 times. Combine the results from the class and redo question 1. Compare the results from the class data to the results from rolling the die 100 times. Are the relative frequencies closer to 1/6th with the class data than for your individual data? Does the histogram for the class data come closer to resembling the probability histogram in Figure 19.2?

3. Next, we work with a pair of dice and consider the sum of the spots on the two dice. The probability model for the sum is given in Table 19.1. Draw a probability histogram for this probability model.

4. a. Roll a pair of dice 100 times. Record the frequencies of the results in a copy of Table 19.6. Enter the probabilities from the probability model (Table 19.1) into the last column.

| Sum of Spots | Frequency | Relative Frequency | Probability |
|---|---|---|---|
| 2 | | | |
| 3 | | | |
| 4 | | | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |
| 10 | | | |
| 11 | | | |
| 12 | | | |
| | | | |

*Table 19.6. Distribution of sum of number of spots.*

b. Calculate the relative frequencies. How close are your relative frequencies to the actual probabilities?

c. Make a histogram of your data using relative frequency (or proportion) for the vertical axis. Use the same scaling on the axes that you used for the probability histogram in question 3. Does your histogram from the data resemble your probability histogram from question 3?

5. Combine the results from the class and redo question 4.

# EXERCISES

1. In the United States, people travel to work in many different ways. Table 19.7 gives the distribution of responses to a survey in which people were asked their means of travel to work.

| Means of Travel | Probability |
|---|---|
| Drive Alone | 0.76 |
| Carpool | 0.12 |
| Public Transportation | 0.05 |
| Walk | 0.03 |
| Work at Home | 0.03 |
| Other | ? |

*Table 19.7. Probability model for transportation to work.*

a. What probability should replace "?" in the probability model?

b. What is the probability that a randomly selected worker does not use public transportation to get to work?

c. What is the probability that a randomly selected worker drives to work, either alone or in a carpool?

d. What is the probability that a randomly selected worker does not drive to work (either alone or in a carpool)?


2. Assume that two U.S. workers are randomly selected. Use the probability model from Table 19.7 to answer the following questions.

a. What is the probability that both workers drive to work (either alone or in a carpool)?

b. What is the probability that neither of the workers drive to work?

c. What is the probability that at least one of the workers drives to work?


3. Use the probability model from Table 19.4 to answer the following questions.

a. What would it mean to say that the chances of having a certain blood type are 50-50?

b. Suppose that a person is selected at random. Compute the probability that the person has Rh-positive blood. Is the chance that the person has Rh-positive blood higher or lower than 50-50? Explain.

c. Any patient with Rh-positive blood can safely receive a transfusion of type O+ blood. What percentage of people in the U.S. can receive a transfusion of type O+ blood?

d. The two most common blood types are O+ and A+. However, many people with O+ and A+ blood do not donate blood. One reason is the belief that because they have a common blood type, their blood is not needed. Is this a valid reason? Support your answer with percentages.


4. Suppose two U.S. residents are randomly selected. Use the probability model from Table 19.4 to find the following probabilities.

a. What is the probability that both have ABO-type O blood?

b. What is the probability that exactly one of the two has ABO-type O blood?

c. What is the probability that neither have type O blood?

# REVIEW QUESTIONS

1. Use the probability model from Table 19.3 to answer the following questions.

a. Let *C* be the event that a vehicle comes to the corner and then either goes straight or turns right. Find *P(C)*.

b. Two vehicles are randomly selected from the traffic study. What is the probability that neither of them goes straight? Which rules of probability did you use in determining your answer?

c. Two vehicles are randomly selected from the traffic study. What is the probability that at least one of them goes straight? Which rules of probability did you use in determining your answer?


2. According to the U.S. Energy Information Administration, about 51% of homes heat with natural gas. Let G represent that a home was heated with gas and N that it was not heated with gas. Suppose three homes were randomly selected.

a. An outcome can be written by a sequence of three letters. For example, GNG represents the outcome that the first home was heated with gas, the second home was not heated with gas, and the third home was heated with gas. List the outcomes in the sample space corresponding to whether the three homes are heated with gas.

b. Consider the following events:
   Event *A*: Exactly one of the three homes heats with gas.

   Event *B*: Exactly two of the three homes heat with gas.

   Event *C*: All three homes heat with gas.

   Event *D*: At least one of the homes heats with gas.

List the outcomes for each event *A, B, C,* and *D.*

c. Which pairs of events *A, B, C,* and *D* are mutually exclusive?

d. Calculate the probabilities of events *A, B, C,* and *D.*

3. Table 19.8 gives a probability model for the distribution of total household income in the U.S.

| Total Household Income | Probability |
|---|---|
| Under $25,000 | 0.223 |
| $25,000 to $49,999 | 0.188 |
| $50,000 to $74,999 | 0.138 |
| $75,000 to $99,999 | 0.179 |
| $100,000 or over | 0.272 |

*Table 19.8. Total household income*
*(from March 2012 Supplement, Current Population Survey)*

a. Check to see whether the probability model in Table 19.8 is legitimate. Explain what you checked.

b. What is the probability that a randomly chosen household will have a total income less than $100,000?

c. What is the probability that a randomly chosen household will have a total income of at least $75,000?

d. What percentage of households has a total income below $75,000?

4. Suppose a random sample of three U.S. households is selected. Use the probability model from Table 19.8 to calculate the following percentages.

a. What is the probability that all three households have total incomes of $25,000 or under? Why is it appropriate to use the Multiplication Rule to calculate this probability?

b. What is the probability that all three households have total incomes of $25,000 or over?

c. What is the probability that at least one of the three households has a total income of $25,000 or over?