3.4: Probability and Compound Events

Learning Objectives

- Compute probability in compound events involving "and"
- Compute probability in compound events involving "or"
- Compute probability in compound events involving "and" and "or"
- Use complements for computing the probability of compound events
- Determine probability from two-way tables
- Extend the use of probability rules to other compound events

Review and Preview Probability and Compound Events

In Chapter 3, we have discussed basic probability and counting concepts to prepare us for our future work with inferential statistics. To use the classical approach for determining the probability of an event A, we need to know the size of the sampe space, n, and we need to know that each of those outcomes in the sample space is equally likely to occur. Then, if event A can occur in x ways in that sample space, $P(A)=\frac{x}{n}$. We also discussed the empirical/experimental approach where we collect data (preferably a large data set to satisfy our Law of Large Numbers more completely) related to our situation of interest. Then, in that data set of size n, if x of the data satisfy the description of event A, we state $P(A)=\frac{x}{n}$. However, we must remember that the empirically based probability value only estimates the actual probability when the data is from a sample set. For example, in Section 2.1, we had a relative frequency table on M&M colors built on data from a package of 55 M&M's, as shown below.

Table $3.4.1$: Frequencies and Relative Frequencies of Sampled M&M's

Color	Relative Frequency
Brown	$\frac{17}{55}\approx 0.309$
Red	$\frac{18}{55}\approx 0.327$
Yellow	$\frac{7}{55}\approx 0.127$
Green	$\frac{7}{55}\approx 0.127$
Blue	$\frac{2}{55}\approx 0.036$
Orange	$\frac{4}{55}\approx 0.073$

From this empirical evidence, we have probability estimates for all M&M colors. If we picked a single M&M from the population of all M&M's ever made, we would estimate, empirically, $P(\text{RED M\&M})\approx 0.327$ and $P(\text{BLUE M\&M})\approx 0.073$.

We have also discussed several counting rules to help us determine the size of an entire sample space and the number of outcomes matching an event description. We established various multiplication rules for counting and an addition rule for counting. Each of these rules has conditions for their use. For example, to use the general permutation rule, we had n distinct objects in which we were interested in how many ways we could select r of those objects where the order of selection mattered. Knowing the conditions allowed us to quickly determine if a given event could be counted by a specific counting method. We now extend the multiplication and addition rules of counting to similar multiplication and addition rules for probabilities of compound events.

Probability with Compound Events Involving "and"

As discussed, some event descriptions can be compound: descriptions involving multiple events. We have discussed two phrases used to combine events: "and" as well as "or." For example, in rolling two dice, the event of $\text{ROLLING A TWO ON THE FIRST DIE and ROLLING A FIVE ON THE SECOND DIE}$ is a compound "and" description. As another example, the event of $\text{PERSON IS OVER SIXTY YEARS OLD or IS UNDER TWELVE YEARS OLD}$ is a compound "or" description. The event of $\text{THREE MANAGERS and SEVEN HOURLY EMPLOYEES or FOUR MANAGERS and SIX HOURLY EMPLOYEES}$ is a compound description involving both "and" as well as "or" descriptions. Although compound event descriptions may get complicated, we aim to simplify determining the probability of such compound events by developing computation methods similar to our counting methods.

We will deal first with the "and" type of event descriptions. Our previous work with probability established that with fair dice, $P(\text{TWO ON FIRST DIE})=\frac{1}{6}$ and $P(\text{FIVE ON SECOND DIE})=\frac{1}{6}$. Our work with the sample space of all outcomes of two fair dice showed $P(\text{TWO ON FIRST DIE and FIVE ON SECOND DIE})=\frac{1}{36}$. We notice that the product of our two event probabilities, $\frac{1}{6}\cdot\frac{1}{6}=\frac{1}{36}$, is the same value as the compound "and" event probability. A multiplication operation happens within probability, just as in counting. Also shown in Section 3.3, we must carefully check if the events separated by "and" depend on each other. Remember the example of the deck of cards $\text{SPADE FIRST and BLACK CARD SECOND}$. When the events are dependent, the count changes for the second event causing the probability to change. $P(\text{SPADE FIRST and BLACK CARD SECOND})=\frac{13}{52}\cdot\frac{25}{51}$. We must notice if the probability of the second event changes when the first event has occurred, as in the card example. With this added condition, we still see multiplication between the two event probabilities. We have a useful general **Probability Multiplication Rule**. Given events A and B,

$$\begin{aligned}P(A\text{ and }B)&=P(A)\cdot P(B|A)\\&=P(B)\cdot P(A|B)\end{aligned}$$

Recall from Section 3.3 that $B|A$ is read as "event B given event A has occurred." Such a probability involving the "given" condition, $P(B|A)$ is called a **Conditional Probability**.

There are events in which $P(B|A)=P(B)$ or equivalently in which $P(A|B)=P(A)$; the occurrence of one event does not affect the occurrence of the other event. For example, the results of the first die roll do not impact the second die roll. In events where $P(B|A)=P(B)$, we say that the two events A and B are **independent** of each other. But, in our example of selecting a spade card first and then a black card second, the occurrence of the first event <u>did</u> impact the occurrence of the second event. In events where $P(B|A)\neq P(B)$, we say that the two events A and B are **dependent** of each other. The conditional probability analysis is not needed if the independence of the events is already known, we need only multiply the two simple event probabilities together, a **Probability Multiplication Rule for Known Independent Events**:

$$\begin{aligned} P(A \text{ and } B) &= P(A) \cdot P(B) \\ &= P(B) \cdot P(A). \end{aligned}$$

We reiterate, use of this rule requires us to know the independence of the events involved before computing our compound probability value. We see in later content that the independence of events can be required to apply specific statistical analyses. Generally, it is best to assume events are dependent and be pleasantly surprised when we can show the events are independent.

? Text Exercise 3.4.1

Answer the following compound probability questions. Recognize that the given event description is a compound event that can be broken down into multiple events.

1. We throw a fair die and randomly select a card from a standard deck. What is the probability of getting a 6 on the die and an ace on the card draw?

Answer

This is a compound "and" event description. We also notice that the two simple events are independent of each other

$$\begin{aligned} P(A \text{ and } B) &= P(\text{SIX ON DIE and ACE ON CARD DRAW} \mid \text{SIX ON DIE}) \\ &= P(\text{SIX ON DIE and ACE ON CARD DRAW}) \\ &= P(\text{SIX ON DIE}) \cdot P(\text{ACE ON CARD DRAW}) \\ &= \frac{1}{6} \cdot \frac{4}{52} \\ &= \frac{1}{78} \approx 1.2821\% \end{aligned}$$

2. We shuffle a standard deck of playing cards so the cards are randomly placed through the deck.
   a. We draw two cards. What is the probability of getting two face cards if the first card is replaced randomly in the deck before drawing the second?
   b. We draw two cards. What is the probability of getting two face cards if the first card is not replaced in the deck before drawing the second?

Answer

a. We do not want to build the sample space in this situation: there are $52 \cdot 52 = 2,704$ different outcomes. We notice that if the first card is replaced randomly in the deck before drawing the second, the two events are independent. We also notice that the deck has 12 face cards. By our multiplication rule:

$$\begin{aligned} P(A \text{ and } B) &= P(\text{FACE CARD ON FIRST DRAW and FACE CARD ON SECOND DRAW}) \\ &= P(\text{FACE ON FIRST FIRST}) \cdot P(\text{FACE ON SECOND} \mid \text{FACE CARD ON FIRST}) \\ &= P(\text{FACE ON FIRST FIRST}) \cdot P(\text{FACE ON SECOND}) \\ & \frac{12}{52} \cdot \frac{12}{52} \\ &= \frac{144}{2,704} = \frac{9}{169} \\ &\approx 5.3254\% \end{aligned}$$

We notice this is not an unusual event.

b. If the first card is not replaced in the deck before drawing the second, the second event's probability depends on the first event occurring because the number of face cards still in the deck will be down to 11. The number of cards in the entire deck will be 51 By our multiplication rule:

$$\begin{aligned} P(A \text{ and } B) &= P(\text{FACE CARD ON FIRST DRAW and FACE CARD ON SECOND DRAW}) \\ &= P(\text{FACE ON FIRST FIRST}) \cdot P(\text{FACE ON SECOND} \mid \text{FACE ON FIRST}) \\ &= \frac{12}{52} \cdot \frac{11}{51} \\ &= \frac{132}{2,652} = \frac{11}{221} \\ &\approx 4.9774\% \end{aligned}$$

Although the probability measure does not differ from the first situation drastically, it is different. The event is a little less likely to occur if the first card is not replaced; this is something we might have expected.

3. Suppose we have our bag of 55 M&M candies with the color distribution as given previously:

Table $3.4.2$: Frequencies and Relative Frequencies of Sampled M&M's

Color	Relative Frequency
Brown	$\frac{17}{55}$
Red	$\frac{18}{55}$
Yellow	$\frac{7}{55}$
Green	$\frac{7}{55}$
Blue	$\frac{2}{55}$

Color	Relative Frequency
Orange	$\frac{4}{55}$

We randomly grab four candies, one at a time, without replacement. What is the probability that we have a red candy first, a green candy second, a brown candy third, and a brown candy fourth?

Answer

These events are not independent. For example, the probability of getting a green candy second will vary depending on whether we had a red candy first. By our multiplication rule:

$$\begin{aligned} P\left(\text{RED } 1^{st} \text{ and GREEN } 2^{nd} \text{ and BROWN } 3^{rd} \text{ and BROWN } 4^{th}\right) &= P(\text{RED}) \cdot P\left(\text{GREEN} \mid \text{RED } 1^{st}\right) \cdot P(\text{BROWN} \mid \text{RED and GREEN previously}) \\ &\quad \cdot P(\text{BROWN} \mid \text{RED and GREEN and BROWN previously}) \\ &= \frac{18}{55} \cdot \frac{7}{54} \cdot \frac{17}{53} \cdot \frac{16}{52} \\ &= \frac{34,272}{8,185,320} = \frac{476}{113,685} \\ &\approx 0.4187\% \end{aligned}$$

4. Senior citizens make up about 18% of the U.S. adult population, according to the Pew Research Center website.
   a. What is the probability of randomly selecting two U.S. adults who are both senior citizens?
   b. What is the probability of randomly selecting two U.S. adults, the first not a senior citizen and the second is a senior citizen?
   c. What is the probability of randomly selecting three U.S. adults such that the first two are not senior citizens and the third is a senior citizen

Answer

a. As the U.S. population of senior citizens is very large, we will assume independence in the events, realizing our probability measures are approximations. By our multiplication rule:

$$\begin{aligned} P(A \text{ and } B) &= P(\text{SENIOR CITIZEN and SENIOR CITIZEN}) \\ &= P(\text{SENIOR CITIZEN}) \cdot P(\text{SENIOR CITIZEN}) \\ &= 18\% \cdot 18\% \\ &\approx 3.24\% \end{aligned}$$

We do emphasize that assumption of independence is not always reasonable, but is used at times when dealing with large population/sample sizes as probability measures do not get impacted significantly in value. We can see this in an example of comparison on, say, a dependence measure of $\frac{2456783}{34595200} \cdot \frac{2456782}{34595199} \approx 0.504314831996\%$ versus assumed independence measure of $\frac{2456783}{34595200} \cdot \frac{2456783}{34595200} \approx 0.50431502269\%$ Although the probability values are technically different, the practical interpretation for real-life application would not usually have practical meaning in the difference. We do note that we reflect carefully before assuming independence as it can lead to drastic consequences.

b. Again, as the U.S. population is very large, we will assume independence in the events. By our complement rule, we note that about 82% of the U.S. adult population is less than 65 years old. By our multiplication rule:

$$\begin{aligned} P(\text{NOT A SENIOR CITIZEN and SENIOR CITIZEN}) &= P(\text{NOT A SENIOR CITIZEN}) \cdot P(\text{SENIOR CITIZEN}) \\ &= 82\% \cdot 18\% \\ &\approx 14.76\% \end{aligned}$$

c. Again, as the U.S. population is very large, we will assume independence in the events. By our complement rule, we note that about 82% of the U.S. adult population is less than 65 years old. By our multiplication rule:

$$\begin{aligned} P(\text{NOT A SENIOR CITIZEN and NOT A SENIOR CITIZEN and SENIOR CITIZEN}) &= 82\% \cdot 82\% \cdot 18\% \\ &\approx 12.10\% \end{aligned}$$

Probability with Compound Events Involving "or"

Now that we have handled probability measures on events separated by "and," we turn to probability and events separated by "or." For example, we might ask, "In rolling a single fair die, what is the probability of rolling a two or a three?" In the case of our M&M's, we might ask, "What is the probability of the next M&M we randomly remove from a bag being a blue or an orange M&M?" As discussed in Section 3.3, the "or" compound description is tied to addition. As a reminder, our developed Addition Rule for Counting was given as

$$\# \text{ of outcomes in } (A \text{ or } B) = \# \text{ of outcomes in } A + \# \text{ of outcomes in } B - \# \text{ of outcomes in } (A \text{ and } B)$$

Our developed rule reminds us to check for the outcomes that matched the descriptions of both events and to subtract the set of the twice-counted outcomes.

Similar to how the Multiplication Rule for Counting extends to the Multiplication Rule for Probability, the Addition Rule for Counting will also extend to an **Addition Rule for Probability**.

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

In answering the question, "In rolling a single fair die, what is the probability of rolling a two or a three?", we can apply this addition rule to produce

$$\begin{aligned}&P\left(\text{ROLLING A TWO or ROLLING A THREE}\right)\\&=P\left(\text{ROLLING A TWO}\right)+P\left(\text{ROLLING A THREE}\right)-P\left(\text{ROLLING A TWO and ROLLING A THREE}\right)\\&=\frac{1}{6}+\frac{1}{6}-\frac{0}{6}\\&=\frac{2}{6}=\frac{1}{3}\approx 33.3333\%\end{aligned}$$

Since the sample space of rolling a single fair die is small in size, one could likely answer this question much faster by knowing that only two of the six outcomes in the sample space meet the description of being "EITHER A TWO or A THREE. But as we see in the text exercises, this addition rule can be very useful when the sample spaces are large.

Text Exercise 3.4.2

Answer the following compound probability questions. Recognize that the given event description is a compound event that can be broken down into multiple events.

1. A standard deck of playing cards is well-shuffled for randomness. Determine the following probabilities about a single card draw.
   a. Find the probability that a randomly drawn card is a king or a queen.
   b. Find the probability that a randomly drawn card is black or an ace.
   c. Find the probability that a randomly drawn card is a spade or a face card.

Answer

a. This is a compound "or" event description, as the event of interest is a king or queen card.

$$\begin{aligned}&P\left(\text{CARD IS A KING or CARD IS A QUEEN}\right)\\&=P\left(\text{KING}\right)+P\left(\text{QUEEN}\right)-P\left(\text{KING and QUEEN}\right)\\&=\frac{4}{52}+\frac{4}{52}-\frac{0}{52}\\&=\frac{8}{52}=\frac{2}{13}\approx 15.3846\%\end{aligned}$$

We notice that, because $P\left(\text{KING and QUEEN}\right)\ =0\%$, these two events are mutually exclusive. There is a little over 15% probability that a random card draw will produce a king or a queen. Although not highly probable, we would not consider this outcome unusual.

b. This is a compound "or" event description, as the event of interest is a black card or an ace.

$$\begin{aligned}&P\left(\text{CARD IS BLACK or CARD IS AN ACE}\right)\\&=P\left(\text{BLACK}\right)+P\left(\text{ACE}\right)-P\left(\text{BLACK and ACE}\right)\\&=\frac{26}{52}+\frac{4}{52}-\frac{2}{52}\\&=\frac{28}{52}=\frac{7}{13}\approx 53.8462\%\end{aligned}$$

We notice that, because $P(\text{BLACK and ACE})\neq 0\%$, these two events are not mutually exclusive. There is almost a 54% chance that a random card draw will produce a black or an ace card; such an outcome is reasonably probable.

c. This is a compound "or" event description, as the event of interest is a card that is a spade or a face card. We note "FACE CARD" is also an "or" event since "FACE CARD" means "$\text{KING or QUEEN or JACK}$ ". Sometimes, event descriptions can be rephrased to make a compound event more evident as an "and" or an "or" type event

$$\begin{aligned}&P\left(\text{CARD IS SPADE or FACE CARD}\right)\\&=P\left(\text{SPADE}\right)+\left(P\left(\text{KING}\right)+P\left(\text{QUEEN}\right)+P\left(\text{JACK}\right)\right)-P\left(\text{SPADE and FACE CARD}\right)\\&=\frac{13}{52}+\left(\frac{4}{52}+\frac{4}{52}+\frac{4}{52}\right)-\frac{3}{52}\\&=\frac{22}{52}\\&=\frac{11}{26}\approx 42.3077\%\end{aligned}$$

We again notice, because $P(\text{SPADE and FACE CARD})\neq 0\%$, these two events are not mutually exclusive. There is a little over 42% probability that a random card draw will produce a spade or a face card.

2. An online clothing store has a liberal return policy since customers cannot try on the items before purchasing. From survey data gathered in the return process from their customers, 21% of all purchased items are returned due to the item being too small/tight, while 4% are returned due to the item being too big/loose. If a customer-purchased item is randomly selected from all purchases, what is the probability that the item will be returned due to being too big/loose or too small/tight?

Answer

This is a compound "or" event description as the purchased item must be returned due to being too big/loose or too small/tight. We notice that these two events are mutually exclusive; we assume that a clothing item cannot simultaneously be too big and too small

$$\begin{aligned}&P\left(\text{TOO BIG/LOOSE or TOO SMALL/TIGHT}\right)\\&=P\left(\text{TOO BIG/LOOSE}\right)+P\left(\text{TOO SMALL/TIGHT}\right)-P\left(\text{TOO BIG/LOOSE and TOO SMALL/TIGHT}\right)\\&=21\%+4\%-0\%=25\%\end{aligned}$$

Twenty-five percent of all customer-purchased items are returned due to being either too big/loose or too small/tight. Such a significant return rate for these two issues will likely require the store to find ways to help customers find a better fit for their orders.

3. We return to our bag of 55 M&M candies.

Table 3.4.3: Frequencies and Relative Frequencies of Sampled M&M's

Color	Relative Frequency
Brown	$\frac{17}{55}$
Red	$\frac{18}{55}$
Yellow	$\frac{7}{55}$
Green	$\frac{7}{55}$
Blue	$\frac{2}{55}$
Orange	$\frac{4}{55}$

a. What is the probability of randomly drawing one candy of a primary color, red, yellow, or blue?
b. What is the probability of randomly drawing one candy that is not of one of those primary colors?

Answer

a. By our addition rule and noticing the events are mutually exclusive, we compute

$$\begin{aligned} P(\text{PRIMARY COLOR M\&M}) &= P(\text{RED M\&M or YELLOW M\&M or BLUE M\&M}) \\ &= P(\text{RED}) + P(\text{YELLOW}) + P(\text{YELLOW}) \\ &= \frac{18}{55} + \frac{7}{55} + \frac{2}{55} \\ &= \frac{27}{55} \approx 49.0909\%. \end{aligned}$$

b. We could apply our addition rule again relative to the colors of brown, green, and orange. But we notice this is also a complement to the previous event.

$$\begin{aligned} P(\text{NOT PRIMARY COLOR M\&M}) &= 1 - P(\text{PRIMARY COLOR M\&M}) \\ &\approx 1 - 0.490909 \\ &= .509091 = 50.9091\% \end{aligned}$$

We notice that the probability of randomly selecting a primary color M&M is practically the same as getting a non-primary color. Stated equivalently, the proportion of M&M's is approximately 50% for each color grouping.

Extended Concepts on Compound Events

As mentioned, we sometimes have event descriptions that include multiple compounding actions. Thankfully, we do not need any extra computation rules for these. However, we do have to read very carefully to properly understand the compound event descriptions and ask for clarification if needed.

We must clarify a small detail about our "AND" rule. In our previous examples using compound "AND" event descriptions, we worked with descriptions involving "sequential trials" of two or more trials: tossing two dice, drawing four M&M candies, and randomly selecting two people. What if we are describing an event with the word "and" but in a single trial: Selecting a single M&M that is both red and brown, drawing a single card that is red and a face card? We use a compound "and," but does our multiplication rule apply in a single trial event? We can often reflect on our sample space and naturally handle the "and" within a single trial. Our general multiplication rule still works, but the two events separated by "and" will often be dependent in a single trial situation. We must carefully consider the conditional probability on the second described event as we apply the multiplication rule.

Let us first examine the probability of selecting a single M&M that is both red and brown from our bag of 55 M&M's. Based on the given information (and our experience eating M&M's), we can reason that there are no M&M's that meet that description, so the probability is $\frac{0}{55} = 0\%$. By reflecting on the sample space, we determine the probability value; no special probability rule is necessary, but let us see that the multiplication rule still works.

$$\begin{aligned} P(\text{RED and BROWN}) &= P(\text{RED}) \cdot P(\text{BROWN} \mid \text{RED}) \\ &= \frac{18}{55} \cdot \frac{0}{18} \\ &= \frac{\cancel{18}}{55} \cdot \frac{0}{\cancel{18}} \\ &= \frac{0}{55} = 0\% \end{aligned}$$

We can see that in handling the conditional probability of selecting a single M&M that is BROWN given that we are restricted to the RED M&M's produces the probability value $\frac{0}{18}$; none of the red M&M's are brown. Our multiplication rule for "AND" still works for even single trial cases. It is acceptable and encouraged to reflect on the sample space and not use our multiplication rule to determine such probability values. We should use sound reasoning, not blind use of formulas, to determine measures.

To see another example, find the probability of drawing a single card that is both red and a face card. By thinking about the sample space of a deck of cards, one might quickly reason that there are six cards in the deck of fifty-two that are red face cards. The probability is $\frac{6}{52} = \frac{3}{26} \approx 11.5385\%$. Let us check that our multiplication rule produces this value as well.

$$\begin{aligned} P(\text{RED and FACE CARD}) &= P(\text{RED}) \cdot P(\text{FACE CARD} \mid \text{RED}) \\ &= \frac{26}{52} \cdot \frac{6}{26} \\ &= \frac{\cancel{26}}{52} \cdot \frac{6}{\cancel{26}} \\ &= \frac{6}{52} \approx 11.5385\% \end{aligned}$$

We can see that in handling the conditional probability of selecting a single FACE CARD given that we are restricted to the RED cards produces the probability value $\frac{6}{52}$ since six of the red cards are face cards. Using conditional probability with the multiplication rule can be handy in more challenging event descriptions.

Let's try a few single trial "and" compound event exercises.

? Text Exercise 3.4.3

Answer the following single-trial compound probability questions. Recognize that the given event description is a compound event that can be broken down into multiple events.

1. We draw a single card from a well-shuffled standard deck of playing cards. What is the probability of getting a black ace?

Answer

This is a compound "and" event description as the card must be both black and an ace.

$$\begin{aligned} P(\text{BLACK CARD and ACE CARD}) &= P(\text{BLACK CARD}) \cdot P(\text{ACE CARD} \mid \text{BLACK CARD}) \\ &= \frac{26}{52} \cdot \frac{2}{26} \\ &= \frac{2}{52} = \frac{1}{26} \\ &\approx 3.8462\% \end{aligned}$$

We could have answered this by reflecting on the sample space of the 52 outcomes, realizing there are 2 black aces, but our multiplication rule does work.

2. Based on $2023-2024$ information from the U.S. Center for Disease Control website, about 70% of senior citizens (65 years and older) in the U.S. get the flu vaccine, whereas about 40 of those adults under 65 years old get vaccinated. Senior citizens make up about 18% of the U.S. adult population, according to the Pew Research Center website.
   a. What is the probability of randomly selecting one U.S. adult who is a senior citizen and has had the flu shot?
   b. What is the probability of randomly selecting one U.S. adult who is not a senior citizen and has not had the flu shot?

Answer

a. We can use our multiplication rule.

$$\begin{aligned} P(\text{SENIOR CITIZEN and FLU SHOT}) &= P(\text{SENIOR CITIZEN}) \cdot P(\text{FLU SHOT} \mid \text{SENIOR CITIZEN}) \\ &= 18\% \cdot 70\% \\ &= 0.18 \cdot 0.70 \\ &= 0.126 \\ &= 12.6\% \end{aligned}$$

In $2023-2024$,about $12.6\%$of the U.S. adult population consisted of senior citizens who had taken the flu shot.

b. We can use our multiplication rule.

$$\begin{aligned} P(\text{NOT A SENIOR CITIZEN and NO FLU SHOT}) &= P(\text{NOT A SENIOR CITIZEN}) \cdot P(\text{NO FLU SHOT} \mid \text{NOT A SENIOR CITIZEN}) \\ &= 82\% \cdot 60\% \\ &= 49.2\% \end{aligned}$$

In $2023-2024$,about $49.2\%$of the U.S. adult population consisted of non-senior citizens who had not taken the flu shot. Notice that this tells health officials about the demographics of those who did not have the flu shot.

Now, we focus on investigating our event descriptions with multiple compound events. No new probability calculation rules are necessary; we must apply our current understanding to slightly new contexts. For example, suppose we wish to find the probability of first throwing a fair die with a result of an even number and then drawing a card from a standard deck of playing cards that is either a red or a face card. This compound event description involves both "and" and "or." Reading carefully, we break down the description: a first event involving throwing a fair die, then a second event involving drawing a card. We notice these two events are independent. In the card draw event, we note that drawing a red card and a face card are not mutually exclusive. We have

$$\begin{aligned}
&P(\text{DIE TOSS OF EVEN and }[\text{CARD DRAW OF RED or FACE}])\\
&= P(\text{DIE TOSS OF EVEN}) \cdot P(\text{CARD DRAW RED or FACE})\\
&\qquad\qquad = P(\text{DIE TOSS OF EVEN})\\
&\cdot (P(\text{CARD DRAW RED}) + P(\text{CARD DRAW FACE}) - P(\text{CARD DRAW RED and FACE}))\\
&= \frac{3}{6} \cdot \left(\frac{26}{52} + \frac{12}{52} - \frac{6}{52}\right)\\
&= \frac{1}{2} \cdot \left(\frac{32}{52}\right)\\
&= \frac{1}{2} \cdot \frac{8}{13} = \frac{4}{13} \approx 30.7692\%.
\end{aligned}$$

Text Exercise 3.4.4

Determine the probabilities of each of the following.

1. We draw five cards from a well-shuffled standard deck of playing cards. What is the probability of getting a flush (all cards share the same suit)?

Answer

This is a compound "and" event description, as the five cards must all be of one suit. We note that the suit does not matter, so the first card drawn can be of any suit; the remaining cards must match the first. The events are not independent.

$$\begin{aligned}
&P(\text{FLUSH})\\
&= P\left(\text{ANY SUIT CARD } 1^{st} \text{ and SAME SUIT CARD } 2^{nd} \text{ and SAME SUIT CARD } 3^{rd} \text{ and SAME SUIT CARD } 4^{th} \text{ and SAME SUIT CARD } 5^{th}\right)\\
&= P\left(\text{ANY SUIT CARD } 1^{st}\right) \cdot P\left(\text{SAME SUIT CARD } 2^{nd} | 1^{st} \text{ CARD SUIT}\right) \cdot P\left(\text{SAME SUIT CARD } 3^{rd} | 1^{st}, 2^{nd} \text{ CARD SUIT}\right)\\
&\qquad \cdot P\left(\text{SAME SUIT CARD } 4^{th} | 1^{st}, 2^{nd}, 3^{rd} \text{ CARD SUIT}\right) \cdot P\left(\text{SAME SUIT CARD } 5^{th} | 1^{st}, 2^{nd}, 3^{rd}, 4^{th} \text{ CARD SUIT}\right)\\
&= \frac{52}{52} \cdot \frac{12}{51} \cdot \frac{11}{50} \cdot \frac{10}{49} \cdot \frac{9}{48}\\
&= \frac{33}{16660} \approx 0.1981\%
\end{aligned}$$

It should not be surprising that flush of five cards from a well-shuffled deck is a very unlikely event.

2. If we toss a fair die and then flip a fair coin, what is the probability that we get either a 6 on the die or a head on the coin (or both)?

Answer

We approach this in two different ways. The first may initially sound easier or more natural, but the second brings us different, valuable insights. Accurate rephrasing of a given situation can make the probability calculation different yet produce the same results.

a. The basic events are a toss of a fair die followed by a flip of a fair coin. Our event description is of the "or" type. Since we can toss a 6 and get a head on the coin in one execution of the situation, we note that our events are not mutually exclusive. By our addition rule, we obtain the following.

$$\begin{aligned}
&P(6 \text{ ON DIE or HEAD ON COIN})\\
&= P(6 \text{ ON DIE}) + P(\text{HEAD ON COIN}) - P(6 \text{ ON DIE and HEAD ON COIN})\\
&= \frac{1}{6} + \frac{1}{2} - \left(\frac{1}{6} \cdot \frac{1}{2}\right)\\
&= \frac{7}{12} \approx 58.3333\%
\end{aligned}$$

b. By examining complement descriptions first, the complement of "either a 6 on the die or a head on the coin (or both)" is "not getting a 6 on the die and also not a head on the coin". Now, we can use our complement and multiplication rule.

$$\begin{aligned}
&P(6 \text{ ON DIE or HEAD ON COIN})\\
&= 1 - (P(\text{NOT A 6 ON DIE and NOT A HEAD ON COIN}))\\
&= 1 - (P(\text{NOT A 6 ON DIE}) \cdot P(\text{NOT A HEAD ON COIN}))\\
&= 1 - \left(\frac{5}{6} \cdot \frac{1}{2}\right)\\
&= 1 - \frac{5}{12} = \frac{7}{12} \approx 58.3333\%
\end{aligned}$$

We do notice we get the same answer although computed through a very different approach.

3. If we roll a fair die three times, what is the probability that one or more throws will come up with a 1?

Answer

On the surface, this sounds easy, but that changes once we think about the possible ways to have one or more of the three approaches produce a 1. We could have the first toss produce a 1 while the other two do not, or we can have the second produce a 1 and the other two not, or we can have the first two tosses both produce a 1 and the third not, or many other possibilities. We must determine many probabilities to use our rule with "or." However, there is an easier way. Notice the complement of "at least one of the throws will be a 1" is given by "none of the three throws produce a 1."

We obtain the following by our complement and multiplication rule, utilizing the independence of the die throws.

$$\begin{aligned}&P\left(\text{AT LEAST ONE - 1 AMONG THREE TOSSES}\right)\\&=1-P\left(\text{NOT 1 ON FIRST DIE and NOT 1 ON SECOND DIE and NOT 1 ON THIRD DIE}\right)\\&=1-\left[P\left(\text{NOT 1 ON FIRST DIE}\right)\cdot P\left(\text{NOT 1 ON SECOND DIE}\right)\cdot P\left(\text{NOT 1 ON THIRD DIE}\right)\right]\\&=1-\left(\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\right)\\&=1-\frac{125}{216}=\frac{91}{216}\approx 42.1296\%\end{aligned}$$

4. Based on $2023-2024$ information from the U.S. Center for Disease Control website, about 70% of senior citizens (65 years and older) in the U.S. get the flu vaccine, whereas about 40 of those adults under 65 years old get vaccinated. Senior citizens make up about 18% of the U.S. adult population, according to the Pew Research Center website. What is the probability of randomly selecting two U.S. adults who are both senior citizens and recipients of the flu vaccine?

Answer

Due to the large population size, we assume the selection of individuals is independent. We believe that if the first selected adult is a senior citizen, the second selected adult still has a 18% chance of being a senior citizen; likewise, we assume that the probability that the second person chosen is vaccinated is the same as the first person. We now use our multiplication rule.

$$\begin{aligned}&P\left(\left[\text{FIRST IS A SENIOR CITIZEN and FIRST HAS FLU SHOT}\right]\text{ and }\left[\ slight\text{SECOND IS A SENIOR CITIZEN and SECOND HAS FLU SHOT}\right]\right)\\&=\left[P\left(\text{FIRST IS A SENIOR CITIZEN}\right)\cdot P\left(\text{FIRST HAS FLU SHOT}\mid\text{FIRST IS A SENIOR CITIZEN}\right)\right]\\&\cdot\left[P\left(\text{SECOND IS A SENIOR CITIZEN}\right)\cdot P\left(\text{SECOND HAS FLU SHOT}\mid\text{SECOND IS A SENIOR CITIZEN}\right)\right]\\&=(18\%\cdot 70\%)\cdot(18\%\cdot 70\%)\\&\approx 1.5876\%\end{aligned}$$

So in regard to this $2023-2024$data, it would be unusual for us to randomly select two U.S adults both of whom were flu vaccinated senior citizens.

Two-Way Tables and Probability

Sometimes, data frequencies for a collected single variable are **disaggregated**, separated into mutually exclusive subgroups. For example, quiz data frequencies for a class may be separated by the frequency of those who passed the quiz versus those who did not. Alternatively, perhaps the data is separated into those students who are in extra-curricular school activities and those who are not. If done well, this allows a comparison between subgroups.

Let us go one step further in this discussion: disaggregating data in two ways. These results are displayed in a **Two-Way Table** or a **Contingency Table**. We take the quiz data of a class of 20 students first given in Section 2.4 and separate by passing, 6 or above, and non-passing scores, below 6, then by student involvement in extra-curricular school activities resulting in the following two-way table.

Disaggregation	Passed Quiz	Failed Quiz
Not Involved in Extra-Curricular School Activities	7	1
Involved in Extra-Curricular School Activities	9	3

When reading the table, we must pay attention to the row and column headings. The value 3 in the table implies that 3 of the 20 students were involved in extra-curricular activities and failed the quiz. We often include row and column totals for each disaggregation when working with two-way tables.

Disaggregation	Passed Quiz	Failed Quiz	Row Totals
Not Involved in Extra-Curricular School Activities	7	1	8
Involved in Extra-Curricular School Activities	9	3	12
Column Totals	16	4	20

We can see a total of 20 student data points. Of those, there were 16 that passed the quiz, while there were 8 not involved in extra-curricular activities.

We can now answer several probability/proportion questions concerning these results. What is the probability that a randomly selected student from the class was involved in extra-curricular school activities? By our classical probability approach, $P(\text{IN EXTRA CURRICULAR})=\frac{12}{20}=\frac{3}{5}=60\%$.

What is the probability that a randomly selected student from the class failed the quiz and was not involved in extra-curricular activities? Since randomly selecting a student from the twenty, $P(\text{FAILED QUIZ and NOT IN EXTRA CURRICULAR})=\frac{1}{20}=5\%$. This is because the table shows only 1 student in both the FAILED QUIZ column as well as the NOT IN EXTRA CURRICULAR row. We could also apply our multiplication rule, noting the events are not independent.

$$\begin{aligned}&P\left(\text{FAILED QUIZ and NOT IN EXTRA CURRICULAR}\right)\\&=P\left(\text{FAILED QUIZ}\right)\cdot P\left(\text{NOT IN EXTRA CURRICULAR}\mid\text{FAILED QUIZ}\right)\\&=\frac{4}{20}\cdot\frac{1}{4}=\frac{1}{20}=5\%\end{aligned}$$

In this situation, the first approach is much easier due to the table information.

Text Exercise 3.4.5

Answer the following probability questions about a given two-way table.

1. A polygraph device is being studied for its accuracy in lie detection. A group of 200 randomly drawn participants were divided into two groups: those instructed not to lie and those who were supposed to lie. The technician running the device did not know each participant's group. Each participant was tested with the device, and a positive or negative detection of a lie was returned. The results are given in the following contingency table.

Disaggregation	Positive Lie Detection Reading	Negative Lie Detection Reading
Participant Did Not Lie	9	71
Participant Did Lie	85	35

As an example, 71 of the 200 participants did not lie, and the detector did not sense a lie occurring.

a. What is the probability that a randomly selected participant had a positive lie detection reading?
b. What proportion of the participants had a false positive reading (the device detected a lie, but the participant did not lie)?
c. What is the probability of randomly selecting two participants in sequence so that both had a positive lie detection reading?
d. What is the probability of randomly selecting a participant who lied but had a negative detection reading or who did not lie but had a positive detection reading? In other words, for what proportion of the trials was the device inaccurate?
e. What is the probability of randomly selecting two participants in sequence so that the first lied with a positive detection reading and the second did not lie with a negative detection reading?

Answer

We first add column and row totals to our table.

Disaggregation	Positive Lie Detection Reading	Negative Lie Detection Reading	Row Totals
Participant Did Not Lie	9	71	80
Participant Did Lie	85	35	120
Column Totals	94	106	200

Now, we answer the probability/proportion questions using appropriate probability rules.

a. From the two-way table, we notice there were 94 participants that had positive lie detection readings. So, $P(\text{POSITIVE DETECTION}) = \frac{94}{200} = \frac{47}{100} = 47\%$.
b. From the table, we can notice there were 9 participants had a false positive reading. So, we can quickly produce results of $P(\text{FALSE POSITIVE DETECTION}) = \frac{9}{200} = 4.5\%$.

We instead might have noticed this was an "and" compound event: DID NOT LIE and $\text{POSITIVE LIE DETECTION}$. Therefore, it would also have been appropriate (though more cumbersome) to measure the proportion using our general multiplication rule.

$$\begin{aligned}&P\left(\text{DID NOT LIE and POSITIVE LIE DETECTION}\right)\\&= P\left(\text{DID NOT LIE}\right)\cdot P\left(\text{POSITIVE LIE DETECTION} \mid \text{DID NOT LIE}\right)\\&= \frac{80}{200}\cdot\frac{9}{80}\\&= \frac{9}{200} = 4.5\%\end{aligned}$$

We mention two items in reflection of this exercise. First, we notice how the table is used for a conditional probability such as $P(\text{POSITIVE LIE DETECTION} \mid \text{DID NOT LIE}) = \frac{9}{80}$. The given condition of DID NOT LIE requires us to use only the information in the table row of "Participants Did Not Lie."

This exercise demonstrates that multiple approaches might be taken to a probability/proportion question, but there is only one correct answer. What is important is to have sound reasoning in our approach. After that, experience is the best teacher in finding the easiest approaches without making mistakes in our reasoning.

c. We notice the event description can be re-worded as $\text{FIRST PARTICIPANT HAD POSITIVE LIE DETECTION}$ and $\text{SECOND PARTICIPANT HAD POSITIVE LIE DETECTION}$, so we use the general multiplication rule.

$$\begin{aligned}&P\left(\text{FIRST PARTICIPANT HAD POSITIVE LIE DETECTION and SECOND PARTICIPANT HAD POSITIVE LIE DETECTION}\right)\\&= P\left(\text{FIRST HAD POSITIVE DETECTION}\right)\cdot P\left(\text{SECOND HAD POSITIVE DETECTION} \mid \text{FIRST HAD POSTIVE DETECTION}\right)\\&= \frac{94}{200}\cdot\frac{93}{199}\\&= \frac{8,742}{39,800} = \frac{4371}{19,900} \approx 21.9648\%\end{aligned}$$

We notice how these sequential events are not independent since we cannot pick the same participant twice.

d. We notice the event description appears to primarily be an "or" compound event. That is, our event of interest requires selection of one participant that $\text{LIED WITH NEGATIVE DETECTION}$ or $\text{DID NOT LIE WITH POSITIVE LIE DETECTION},$ so we use the general addition rule noting these two events are mutually exclusive.

$$\begin{aligned} & P\left(\text{LIED WITH NEGATIVE DETECTION or DID NOT LIE WITH POSITIVE LIE DETECTION}\right) \\ & = P\left(\text{LIED WITH NEGATIVE DETECTION}\right) + P\left(\text{DID NOT LIE WITH POSITIVE LIE DETECTION}\right) \\ & = \frac{35}{200} + \frac{9}{200} \\ & = \frac{44}{200} = \frac{11}{50} = 22\% \end{aligned}$$

We observe that a test with a 22% inaccuracy rate is not very dependable. Also, we excluded the subtraction in the general addition rule for possible double counting of $\text{LIED WITH NEGATIVE DETECTION}$ and $\text{DID NOT LIE WITH POSITIVE LIE DETECTION}$ since we recognized the events were mutually exclusive.

e. The event description is an "and" compound type as we want sequential selection of two participants with $\text{FIRST PARTICIPANT LIED WITH POSITIVE LIE DETECTION}$ and $\text{SECOND PARTICIPANT DID NOT LIE WITH NEGATIVE LIE DETECTION}$. We use the general multiplication rule.

$$\begin{aligned} & P\left(1^{st}\text{ PARTICIPANT LIED WITH POSITIVE LIE DETECTION and }2^{nd}\text{ PARTICIPANT DID NOT LIE WITH NEGATIVE LIE DETECTION}\right) \\ & = P\left(1^{st}\text{ PARTICIPANT LIED WITH POSITIVE DETECTION}\right) \\ & \cdot P\left(2^{nd}\text{ PARTICIPANT DID NOT LIE WITH NEGATIVE DETECTION }|1^{st}\text{ PARTICIPANT LIED WITH POSTIVE DETECTION}\right) \\ & = \frac{85}{200} \cdot \frac{71}{199} \\ & = \frac{6,035}{39,800} = \frac{1,207}{7,960} = 15.1633\% \end{aligned}$$

It would not be unusual, since the probability value is above 5%.