Week 4 - An Introduction to Probability

Ryan Miller

Introduction

- Video #1
  - Connecting study design and probability, introducing basic terminology
- Video #2
  - Sample spaces and complementary events
- Video #3
  - The addition rule
- Video #4
  - Conditional probabilities and the multiplication rule
- Video #5
  - Examples

Randomness

In our discussions of study design, *randomness* arose in two contexts:

- **Random sampling**
  - We used randomness to ensure every case in the population had an equal chance of being sampled
  - This prevented *sampling bias*, but it introduced *sampling variability*

Randomness

In our discussions of study design, *randomness* arose in two contexts:

- **Random sampling**
  - We used randomness to ensure every case in the population had an equal chance of being sampled
  - This prevented *sampling bias*, but it introduced *sampling variability*
- **Random assignment**
  - We used randomness to split our sample to into treatment and control groups
  - This protected us against *confounding variables*, but it introduces variability (you can view group assignment similarly to random sampling)

We'll spend the remainder of the course learning ways to quantify the variability resulting from randomness, a task that requires us to study *probability*

Random Processes

- A **trial** refers to a single instance of a **random process**
  - Every trial results in an **outcome**

Random Processes

- A **trial** refers to a single instance of a **random process**
  - Every trial results in an **outcome**
- As an example, consider the selection of one member of the Xavier faculty (population) into a sample
  - This selection represents one trial of a random process
  - One possible outcome would be "Ryan Miller"
  - If we record our selection different, another outcome might be "Teaches Math"

Random Processes

- A **trial** refers to a single instance of a **random process**
  - Every trial results in an **outcome**
- As an example, consider the selection of one member of the Xavier faculty (population) into a sample
  - This selection represents one trial of a random process
  - One possible outcome would be "Ryan Miller"
  - If we record our selection different, another outcome might be "Teaches Math"
- The collection of *all possible outcomes* of a trial is called the **sample space**
  - For example, the sample space when sampling Xavier faculty would be a list of hundreds of names
  - In the special case of random sampling, each outcome in the sample space is *equally likely*

Events

- Statisticians tend to focus on **events**, which are *combinations of one or more observed outcomes*
- Below are a couple examples of events in our Xavier faculty example:
  - The faculty member is younger than 40 and teaches math
  - The faculty member teaches math or teaches computer science

Probability

- The outcomes that form the basis of these events involve randomness, so they are inherently linked to *probability*, but what exactly do we mean by "probability"?
  - For example, everyone agrees the probability of a fair coin landing "heads" is $1/2$, but why?

Probability

- The outcomes that form the basis of these events involve randomness, so they are inherently linked to *probability*, but what exactly do we mean by "probability"?
  - For example, everyone agrees the probability of a fair coin landing "heads" is $1/2$, but why?
- **Frequentist** statisticians define probability as the *long-run proportion of an event occurring*
  - Thus, $P(\text{Heads}) = 0.5$ means that if we conducted many *trials* (different coin flips) we'd expect the *event* "Heads" to be observed in half of them

Probability

- The outcomes that form the basis of these events involve randomness, so they are inherently linked to *probability*, but what exactly do we mean by “probability”?
  - For example, everyone agrees the probability of a fair coin landing “heads” is $1/2$, but why?
- **Frequentist** statisticians define probability as the *long-run proportion of an event occurring*
  - Thus, $P(\text{Heads}) = 0.5$ means that if we conducted many *trials* (different coin flips) we'd expect the *event* “Heads” to be observed in half of them
- Flipping a coin is also an example of the special type of random process where each possible outcome is equally likely, which provides an alternative justification for $P(\text{Heads}) = 0.5$

Empirical Probability

- Since probabilities are *long-run proportions*, we sometimes will estimate them using data
  - For example, Steph Curry's career free throw percentage is 90.7%, so the next time he's at the free throw line we might estimate $P(\text{Make}) = 0.907$

Empirical Probability

- Since probabilities are *long-run proportions*, we sometimes will estimate them using data
  - For example, Steph Curry's career free throw percentage is 90.7%, so the next time he's at the free throw line we might estimate $P(\text{Make}) = 0.907$
- This is called an *empirical probability*, it is different from a *theoretical probability* like $P(\text{Heads}) = 0.5$
  - Empirical probabilities are *estimated* using a finite amount of data
  - Theoretical probabilities are *derived* based upon the nature of the random process

Unions and Intersections

- As previously mentioned, an *event* is a combinations of *one or more outcomes*
  - There are *two ways* to combine multiple outcomes, *unions* and *intersections*

Unions and Intersections

- As previously mentioned, an *event* is a combinations of *one or more outcomes*
  - There are *two ways* to combine multiple outcomes, *unions* and *intersections*
- **Intersection** refers to two (or more) outcomes simultaneously occurring
  - Intersections are expressed using "and" or the symbol $\cap$
  - Consider rolling a six-sided die,
    $P(\text{Five and Six}) = P(\text{Five} \cap \text{Six}) = 0$
  - Alternatively,
    $P(\text{Five and Odd Number}) = P(\text{Five} \cap \text{Odd Number}) = 1/6$

Unions and Intersections

- As previously mentioned, an *event* is a combinations of *one or more outcomes*
  - There are *two ways* to combine multiple outcomes, *unions* and *intersections*
- **Intersection** refers to two (or more) outcomes simultaneously occurring
  - Intersections are expressed using "and" or the symbol $\cap$
  - Consider rolling a six-sided die, $P(\text{Five and Six}) = P(\text{Five} \cap \text{Six}) = 0$
  - Alternatively, $P(\text{Five and Odd Number}) = P(\text{Five} \cap \text{Odd Number}) = 1/6$
- **Union** refers to at least one of the specified outcomes occurring
  - Union are expressed using "or" or the symbol $\cup$
  - Consider rolling a six-sided die, $P(\text{Five or Six}) = P(\text{Five} \cup \text{Six}) = 2/6 = 1/3$
  - Alternatively, $P(\text{Five or Odd Number}) = P(\text{Five} \cup \text{Odd Number}) = 3/6 = 1/2$

Closing Remarks (Introduction)

- Probability provides a framework for understanding randomness, something is necessary when our data involve sampling or random assignment (or both)

Closing Remarks (Introduction)

- Probability provides a framework for understanding randomness, something is necessary when our data involve sampling or random assignment (or both)
- A *trial* described an instance of a random process that resulted in an *outcome*
  - The collection of all possible outcomes was the *sample space*
- An *event* was a combination of one or more outcomes
  - Events can be expressed as *unions* or *intersections* of different outcomes

Valid Probabilities

- By definition, the probability of any *event* must be between 0 and 1
  - So, please never report a probability outside of this range on an exam or homework assignment (if necessary, leave your incorrect work and don't circle a final answer)

Valid Probabilities

- By definition, the probability of any *event* must be between 0 and 1
  - So, please never report a probability outside of this range on an exam or homework assignment (if necessary, leave your incorrect work and don't circle a final answer)
- An event with a probability of 1 will *always* occur
  - For example, if we flip a single coin: $P(\text{Heads or Tails}) = 1$

Valid Probabilities

- By definition, the probability of any *event* must be between 0 and 1
  - So, please never report a probability outside of this range on an exam or homework assignment (if necessary, leave your incorrect work and don't circle a final answer)
- An event with a probability of 1 will *always* occur
  - For example, if we flip a single coin: $P(\text{Heads or Tails}) = 1$
- An event with a probability of 0 will *never* occur
  - For example, if we sample a random adult and measure their height: $P(6'0) = 0$
  - This might seem odd, but there are infinitely many different heights, so the probability of getting someone who's *exactly* 6 feet and 0.00000... inches tall is zero

Sample Spaces and Probability

- The probability of the *union of all outcomes* in a sample space is 1
  - If we flip a single coin: $P(\text{Heads or Tails}) = 1$
  - If we randomly sample letter grades on an exam: $P(\text{A or B or C or D or F}) = 1$

The Complement Rule

- The probabilities of an event occurring, and that event not occurring, must sum to 1
  - For a coin flip: $P(\text{Heads}) + P(\text{Not Heads}) = 1$
  - For a random exam: $P(\text{A}) + P(\text{Not A}) = 1$

The Complement Rule

- The probabilities of an event occurring, and that event not occurring, must sum to 1
  - For a coin flip: $P(\text{Heads}) + P(\text{Not Heads}) = 1$
  - For a random exam: $P(\text{A}) + P(\text{Not A}) = 1$
- We call these **complementary events** using the following notation
  - If A is an event, then A^C denotes the complement of event A
  - So mathematically we can say: $P(A) + P(A^C) = 1$ for any event A
  - It's sometimes useful to rearrange this expression: $P(A^C) = 1 - P(A)$

Example - Complement Rule

- Driving home from work, Professor Miller approaches a traffic light that he knows will be "Green" with probability 0.4, "Yellow" with probability 0.1, or "Red" with probability 0.5
  - What is the probability the light is *not Red*?
  - $P(\text{Red}^C) = 1 - P(\text{Red}) = 0.5$

Example - Complement Rule

- Driving home from work, Professor Miller approaches a traffic light that he knows will be "Green" with probability 0.4, "Yellow" with probability 0.1, or "Red" with probability 0.5
  - What is the probability the light is *not Red*?
  - $P(\text{Red}^C) = 1 - P(\text{Red}) = 0.5$
- Intuitively, you might also have thought to add the probabilities $P(\text{Green}) + P(\text{Yellow}) = 0.5$

Example - Complement Rule

- Driving home from work, Professor Miller approaches a traffic light that he knows will be "Green" with probability 0.4, "Yellow" with probability 0.1, or "Red" with probability 0.5
  - What is the probability the light is *not Red*?
  - $P(\text{Red}^C) = 1 - P(\text{Red}) = 0.5$
- Intuitively, you might also have thought to add the probabilities $P(\text{Green}) + P(\text{Yellow}) = 0.5$
- Be aware that this approach only works for **disjoint events** (events that cannot occur simultaneously)
  - By definition, all *outcomes* in a sample space must be *disjoint*, so it's not a problem here
  - Next we'll see what happens for non-disjoint (dependent) events

Closing Remarks (Sample Spaces and Complements)

- A *sample space* is the collection of all possible *outcomes* of a trial
  - The probability of the intersection of all outcomes in a sample space is 1
- The probabilities of an event occurring, and that event not occurring, must sum to 1 (complementary events)
  - $P(A) + P(A^C) = 1$
  - $P(A^C) = 1 - P(A)$

Disjoint Events

- Two *events* are **disjoint** if they have no *outcomes* in common
  - Consider rolling a six-sided die, the event of rolling a six is disjoint from the event rolling an odd number

Disjoint Events

- Two *events* are **disjoint** if they have no *outcomes* in common
  - Consider rolling a six-sided die, the event of rolling a six is disjoint from the event rolling an odd number
- For two disjoint events, we can find the probability of *unions* by addition
  - $P(A \text{ or } B) = P(A) + P(B)$
  - For a six-sided die, $P(\text{Six or Odd Number}) = P(\text{Six}) + P(\text{Odd Number}) = 1/6 + 3/6 = 2/3$

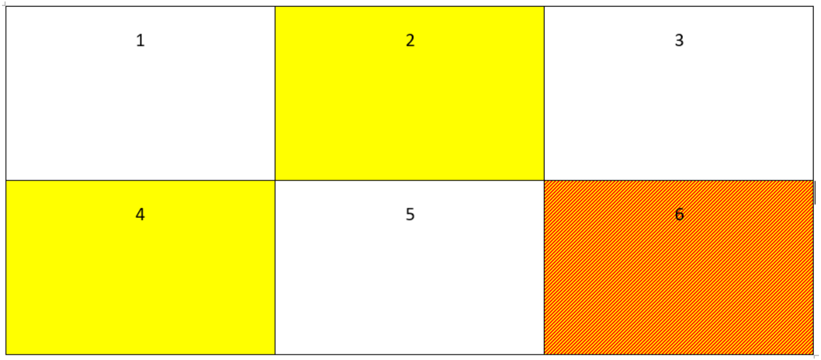
Disjoint Events

It's easy to visually confirm this example by looking at a simple representation of the sample space:

1	2	3
4	5	6

Non-disjoint events

In contrast, consider $P(\text{Six or Even Number})$, clearly these events are *not disjoint*, so adding their probabilities would be a mistake



The Addition Rule

- In general, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
  - This is known as the **addition rule**
  - In the special case where events A and B are *disjoint*, $P(A \text{ and } B) = 0$

The Addition Rule

- In general, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
  - This is known as the **addition rule**
  - In the special case where events A and B are *disjoint*, $P(A \text{ and } B) = 0$
- In our previous examples:

$P(\text{Six or Odd Number}) = P(\text{Six}) + P(\text{Odd Number}) -$
$P(\text{Six and Odd Number}) = 1/6 + 3/6 - 0 = 2/3$

The Addition Rule

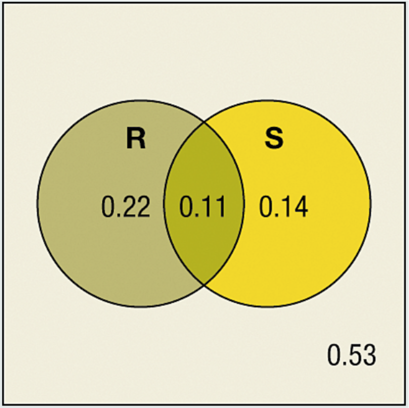
- In general, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
  - This is known as the **addition rule**
  - In the special case where events A and B are *disjoint*, $P(A \text{ and } B) = 0$
- In our previous examples:

$P(\text{Six or Odd Number}) = P(\text{Six}) + P(\text{Odd Number}) - P(\text{Six and Odd Number}) = 1/6 + 3/6 - 0 = 2/3$

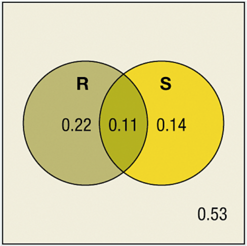
$P(\text{Six or Even Number}) = P(\text{Six}) + P(\text{Even Number}) - P(\text{Six and Even Number}) = 1/6 + 3/6 - 1/6 = 1/2$

Venn Diagrams

- Venn diagrams are frequently used as a visual aid when learning the addition and complement rules
- The diagram below depicts survey results where 33% of college students were in a relationship (R), 25% were involved in sports (S), and 11% were in both

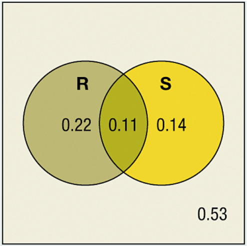


Venn Diagrams - Example



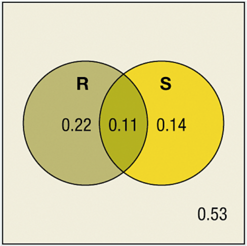
1) $P(R \text{ or } S) = 0.22 + 0.11 + 0.14 = 0.47$ (direct calculation)

Venn Diagrams - Example



1) $P(R \text{ or } S) = 0.22 + 0.11 + 0.14 = 0.47$ (direct calculation)
2) $P(R \text{ or } S) = 0.33 + 0.25 - 0.11 = 0.47$ (addition rule)

Venn Diagrams - Example



1) $P(R \text{ or } S) = 0.22 + 0.11 + 0.14 = 0.47$ (direct calculation)
2) $P(R \text{ or } S) = 0.33 + 0.25 - 0.11 = 0.47$ (addition rule)
3) $P(R \text{ or } S) = 1 - P(\text{Neither}) = 1 - 0.53 = 0.47$ (complement rule)

Closing Remarks (Addition Rule)

- We use the *addition rule* to find the probability of the union of any two events:

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

- If the events are *disjoint*, we know their intersection is zero, or $P(A \text{ and } B) = 0$
  - In this special case, the union of the events is simply the sum of their individual probabilities

Estimating Probabilities

- As we've previously discussed, probabilities are defined as *long-run proportions*
  - The probability of a coin landing heads is 0.5 because it's what we'd expect after many tosses

Estimating Probabilities

- As we've previously discussed, probabilities are defined as *long-run proportions*
  - The probability of a coin landing heads is 0.5 because it's what we'd expect after many tosses
- So, it makes sense to use proportions from a sample to *estimate* probabilities
  - For example, the probability of Steph Curry making a free throw is 0.907
  - Obviously Steph hasn't shot infinitely many free throws, but he's taken enough for us to get a good estimate (recall this was called an *empirical probability*)

Conditional Probability

- The connection between probabilities and proportions will greatly expand the ways in which can apply probability
  - Analogous to *conditional proportions* (row and column percentages), is the concept of a **conditional probability**

Conditional Probability

- The connection between probabilities and proportions will greatly expand the ways in which can apply probability
  - Analogous to *conditional proportions* (row and column percentages), is the concept of a **conditional probability**
- Conditional probability is used in scenarios involving *dependent events*
  - For example, the probability that Steph Curry makes a free throw could depend on whether he's playing in a home game or an away game

Conditional Probability

- The connection between probabilities and proportions will greatly expand the ways in which can apply probability
  - Analogous to *conditional proportions* (row and column percentages), is the concept of a **conditional probability**
- Conditional probability is used in scenarios involving *dependent events*
  - For example, the probability that Steph Curry makes a free throw could depend on whether he's playing in a home game or an away game
- We use a vertical bar to denote conditional probabilities: $P(A|B)$
  - In this example, we might define "A" to be making the free throw and "B" to be playing at home
  - As you'd expect, conditional probabilities can be *estimated* from a contingency table

Example - Conditional Probabilities

ACTN3 is known as the fast twitch gene, everyone has one of three genotypes (XX, RR, or RX). The table below summarizes a sample of 301 elite athletes:

	RR	RX	XX	Total
Sprint/power	53	48	6	107
Endurance	60	88	46	194
Total	113	136	52	301

From this table, let's estimate a few different probabilities:

1) An endurance athlete has genotype XX?

Example - Conditional Probabilities

ACTN3 is known as the fast twitch gene, everyone has one of three genotypes (XX, RR, or RX). The table below summarizes a sample of 301 elite athletes:

	RR	RX	XX	Total
Sprint/power	53	48	6	107
Endurance	60	88	46	194
Total	113	136	52	301

From this table, let's estimate a few different probabilities:

1) An endurance athlete has genotype XX? $46/194 = 0.237$
2) An athlete with the XX genotype is an endurance athlete?

Example - Conditional Probabilities

ACTN3 is known as the fast twitch gene, everyone has one of three genotypes (XX, RR, or RX). The table below summarizes a sample of 301 elite athletes:

	RR	RX	XX	Total
Sprint/power	53	48	6	107
Endurance	60	88	46	194
Total	113	136	52	301

From this table, let's estimate a few different probabilities:

1) An endurance athlete has genotype XX? $46/194 = 0.237$
2) An athlete with the XX genotype is an endurance athlete? $46/52 = 0.885$
3) An athlete has the XX genotype *and* is an endurance athlete?

Example - Conditional Probabilities

ACTN3 is known as the fast twitch gene, everyone has one of three genotypes (XX, RR, or RX). The table below summarizes a sample of 301 elite athletes:

	RR	RX	XX	Total
Sprint/power	53	48	6	107
Endurance	60	88	46	194
Total	113	136	52	301

From this table, let's estimate a few different probabilities:

1) An endurance athlete has genotype XX? $46/194 = 0.237$
2) An athlete with the XX genotype is an endurance athlete? $46/52 = 0.885$
3) An athlete has the XX genotype *and* is an endurance athlete? $46/301 = 0.153$

The Multiplication Rule

The relationship between these probabilities motivates the **multiplication rule**, which states:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

In our previous example, notice:

1) $P(\text{XX}|\text{End}) = 46/194 = 0.237$
2) $P(\text{End}) = 194/301 = 0.645$
3) $P(\text{XX and End}) = 46/301 = 0.153$

	RR	RX	XX	Total
Sprint/power	53	48	6	107
Endurance	60	88	46	194
Total	113	136	52	301

It's easy verify the multiplication rule: $46/194 = \frac{46/301}{194/301}$

Independence

- Two events are considered *independent* when $P(A \text{ and } B) = P(A) * P(B)$

Independence

- Two events are considered *independent* when $P(A \text{ and } B) = P(A) * P(B)$
  - Notice that independence does not mean the events are *disjoint*
  - $P(A \text{ and } B) = 0$ for disjoint events

Independence

- Two events are considered *independent* when $P(A \text{ and } B) = P(A) * P(B)$
  - Notice that independence does not mean the events are *disjoint*
  - $P(A \text{ and } B) = 0$ for disjoint events
- Independence tends to greatly simplify probability calculations
  - Consider 3 coin flips:

$$P(H_1 \text{ and } H_2 \text{ and } H_3) = P(H_1) * P(H_2) * P(H_3) = (1/2)^3 = 1/8$$

- This is a much easier calculation to think about compared to:

$$P(H_1 \text{ and } H_2 \text{ and } H_3) = P(H_3) * P(H_2|H_1) * P(H_3|H_1 \text{ and } H_2)$$

Closing Remarks (Probability Rules)

We've now covered three different probability rules:

1) The addition rule, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$, allows us to calculate the probability of *unions* of events
2) The multiplication rule, $P(A \text{ and } B) = P(A|B) * P(B)$, allows us to calculate the probability of *intersections* of events
3) The complement rule, $P(A) + P(A^C) = 1$, allows simpler calculations for large sample spaces

Our final video for this week will work through a few examples illustrating how to apply these rules to different situations

Example #1 - part 1

A local hospital has 22 patients staying overnight, 15 are adults and 7 are children. Among the adults, this is the first ever hospital stay for 4 of them. Among the children, this is the first ever hospital stay for 5 of them. Use this information to calculate the following probabilities:

1) A randomly selected patient is an adult
2) A randomly selected patient is an adult, given it's their first ever hospital stay
3) A randomly selected patient is in their first ever hospital stay, given they are a child
4) A randomly selected patient is in their first ever hospital stay, *or* they are a child

Example #1 - part 1 (solution)

1) $P(\text{Adult}) = 15/22$, there are 22 patients and 15 are adults

Example #1 - part 1 (solution)

1) $P(\text{Adult}) = 15/22$, there are 22 patients and 15 are adults
2) $P(\text{Adult}|\text{First}) = 4/9$, there are 9 first-time patients and 4 are adults

Example #1 - part 1 (solution)

1) $P(\text{Adult}) = 15/22$, there are 22 patients and 15 are adults
2) $P(\text{Adult}|\text{First}) = 4/9$, there are 9 first-time patients and 4 are adults
3) $P(\text{First}|\text{Child}) = 5/7$, there are 7 children and 5 of them are first-time patients

Example #1 - part 1 (solution)

1) $P(\text{Adult}) = 15/22$, there are 22 patients and 15 are adults
2) $P(\text{Adult}|\text{First}) = 4/9$, there are 9 first-time patients and 4 are adults
3) $P(\text{First}|\text{Child}) = 5/7$, there are 7 children and 5 of them are first-time patients
4) $P(\text{First or Child}) = P(\text{First}) + P(\text{Child}) - P(\text{First and Child}) = 9/22 + 7/22 - 5/22 = 0.5$, notice we could have calculated this *directly* by realizing there are 7 children and 4 first-time adults (totaling 11 of 22 patients)

Example #1 - part 2

A local hospital has 22 patients, 15 are adults and 7 are children. Among the adults, this is the first ever hospital stay for 4 of them. Among the children, this is the first ever hospital stay for 5 of them. Now let's consider randomly selecting two patients sequentially:

1) What is the probability that *both* selections are adults?
2) What is the probability that *at least one* of the selections is an adult?

Example #1 - part 2 (solution)

1) Let A_1 and A_2 denote the selection of adults, then $P(A_1 \text{ and } A_2) = P(A_2|A_1) * P(A_1) = \frac{14}{21} * \frac{15}{22} = 0.45$; notice these events are not independent

Example #1 - part 2 (solution)

1) Let A_1 and A_2 denote the selection of adults, then $P(A_1 \text{ and } A_2) = P(A_2|A_1) * P(A_1) = \frac{14}{21} * \frac{15}{22} = 0.45$; notice these events are not independent
2) Using the additional rule could get complicated here because the events are not independent. Instead, let C_1 and C_2 denote the selection of children and consider $P(A_1 \text{ or } A_2) = 1 - P(\text{Neither}) = P(C_2|C_1) * P(C_1) = \frac{6}{21} * \frac{7}{22} = 1 - 0.09 = 0.91$

Example #2

Consider a well-shuffled deck of 52 playing cards and the random selection of two cards, a "top" card and a "bottom" card

1) The following line of reasoning is incorrect: "Because of the addition rule, the probability that the top card is the jack of clubs *and* the bottom card is the jack of hearts is $2/52$." Point out the flaw in this argument.
2) The following line of reasoning is also incorrect: "Because of the addition rule, the probability that the top card is the jack of clubs *or* the bottom card is the jack of hearts is $2/52$." Point out the flaw in this argument.
3) The statements in 1 and 2 both contain flaws, but these mistakes are not equally bad. Which approach will result in an answer closer to the truth (for the situation it describes)?

Example #2 (solution)

1) The addition rule pertains to intersections or “or” statements, so it shouldn’t be applied here

Example #2 (solution)

1) The addition rule pertains to intersections or “or” statements, so it shouldn’t be applied here
2) The events involved are not disjoint, it is possible for the top card to be the jack of clubs and the bottom card to be the jack of hearts.

Example #2 (solution)

1) The addition rule pertains to intersections or "or" statements, so it shouldn't be applied here
2) The events involved are not disjoint, it is possible for the top card to be the jack of clubs and the bottom card to be the jack of hearts.
3) The second statement is much closer to the truth, because the possibility for both is very small ($\frac{1}{52} * \frac{1}{51}$ by the multiplication rule)