Basic Probability

5CS037

Contents - 1

- Calculate basic probabilities
- Identify independent and dependent variables
- Understand and calculate permutations
- Calculate probabilities using permutations and combinations

Unit Overview

- Inferential Statistics is the practice of using mathematical analysis to make inferences about a population from a sample
- The mathematics which underlie inferential statistics are largely based on probability theory

- First, this unit will cover independence; then it will move forward and apply the concepts from set theory in terms of probability calculations. Finally, the unit will cover statistical counting concepts like permutations and combinations, and apply these concepts in the calculation of probabilities.

Calculating Probability

Calculating probability is attempting to figure out the likelihood of a specific event happening, given some number of attempts. The most fundamental and important probability calcululation is defined as:

The probability of some event A occuring is the number of possible outcomes in that event, divided by the total number of possible outcomes in the sample space. That is,

Number of outcomes in $A = |A| =$ "The Cardinality of A"
Number of outcomes in $S = |S| =$ "The Cardinality of S"

$$P(A) = \frac{|A|}{|S|}$$

The $|A|$ notation may seem a little strange at first. But actually, this is a very common mathematical notational style to specify size, or magnitude.

Example

Given a fair six-sided die, what is the probability of rolling a 5?

$$Event\ A = \text{Rolling a five}$$

$$P(A) = \frac{1}{6} = .166667$$

Solution

The total number of possible outcomes is six, in other words the cardinality of the sample space is six. There is only one outcome in which our die will show five pips, so the cardinality of our event A is 1. Hence, our probability is $\frac{1}{6}$.

Notation	Meaning
$P(A)$	Probability of A
$P(A^c)$	Probability of A complement
$P(AB)$	Probability of A intersect B
$P(A \cup B)$	Probability of A union B
$P(A \mid B)$	Probability of A given B

Dependence and Independence

- Identifying dependence between events is often a first step in developing insightful models.
- The real-world events that a data scientist might study tend to be interrelated:
- Does the probability that a person will get Lyme disease change if you know that person shops at REI? Are states with populations over 10 million more likely to experience voter fraud?

⚠ Correlation does not imply causation:

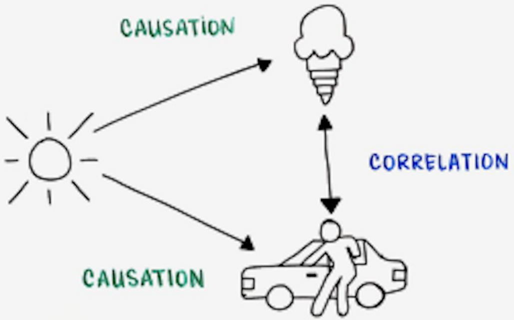
Just remember, there is ***no implication of causation*** when it comes to dependent events. This should be clear from the first example above; though it may be true that REI shoppers get Lyme disease at a higher rate than the general population, the act of shopping at REI does not necessarily change one's likelihood of getting Lyme disease.

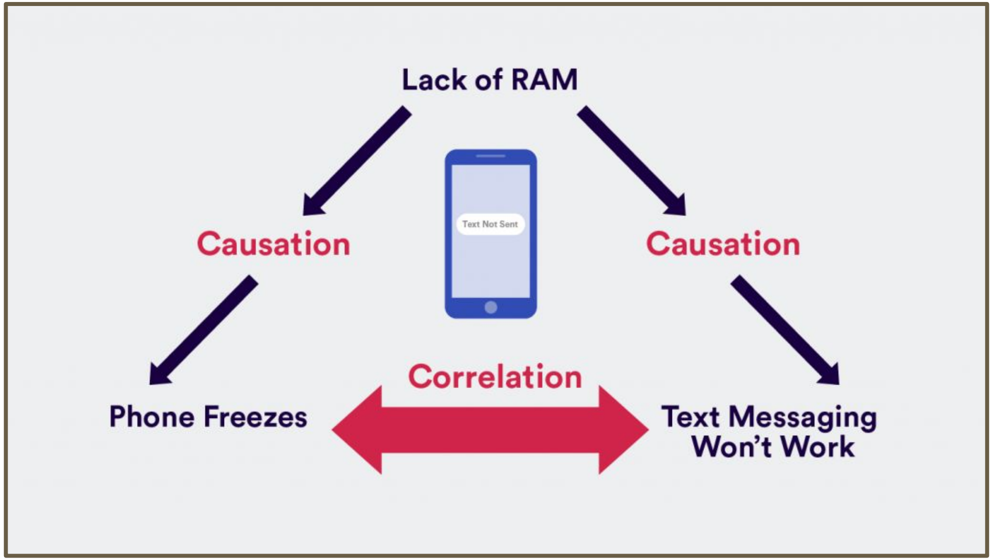
Correlation and causation are terms often confused when relationships between variables are described.

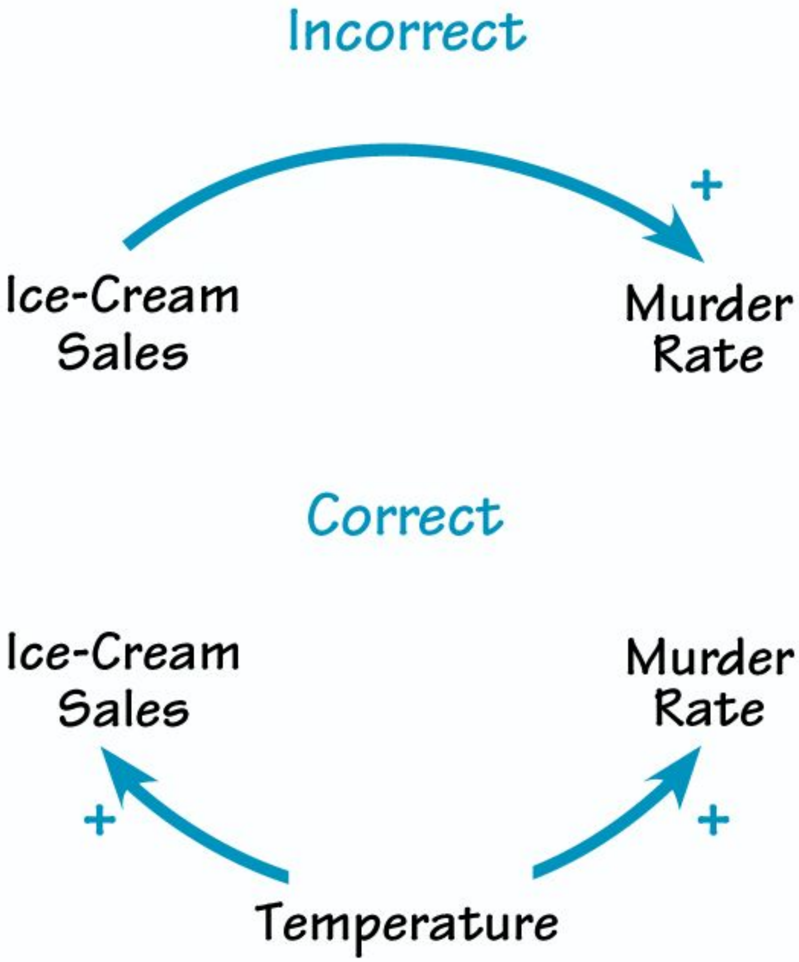
Correlation is an evaluation of how variables move in relation to one another. A positive correlation indicates variables move in the same direction and a negative correlation means they move in opposite directions.

Causation means one event is the direct cause of another event.

An example illustrating the difference is ice cream sales and car thefts. There's a positive correlation between ice cream sales and car thefts, but no causal relationship: eating ice cream doesn't cause people to steal cars. Instead, a third variable (the weather) explains the correlation. Both ice cream sales and car thefts increase in warmer months.




Lack of RAM
Causation
Text Not Sent
Causation
Correlation
Phone Freezes
Text Messaging
Won't Work


Incorrect
Ice-Cream Sales
+
Murder Rate
Correct
Ice-Cream Sales
Murder Rate
+
+
Temperature

Dependent & Independent events

- **DE** : There are fewer possible event during the 2nd selection vs. the 1st selection. If you picked a card out of a standard deck of cards without replacing the card, on the 2nd pick there would be 51 instead of 52.

- **IE:** The total possible outcomes never changes. Each time you flip a coin, there is always 2 possible outcomes

Which is an example of dependent events?

1. Flipping a fair coin twice and getting tails both times.
2. Choosing the starting player line-up for a basketball game.
3. Choosing two cards from a stack of colored cards, with replacement, and both cards are blue.
4. Rolling a 6-sided die 2 times and getting a 5 both times.

Which is an example of dependent events?

1. Flipping a fair coin twice and getting tails both times.
2. **Choosing the starting player line-up for a basketball game.**
3. Choosing two cards from a stack of colored cards, with replacement, and both cards are blue.
4. Rolling a 6-sided die 2 times and getting a 5 both times.

Which is an example of independent events?

1. Picking which two 6-sided die to roll out of 6 given to you
2. Determining who will be the 2 captains on the football team
3. Picking which 3 outfits to take with you on your weekend vacation
4. Rolling a 6-sided die 2 times and getting a 5 both times

Which is an example of independent events?

1. Picking which two 6-sided die to roll out of 6 given to you
2. Determining who will be the 2 captains on the football team
3. Picking which 3 outfits to take with you on your weekend vacation
4. **Rolling a 6-sided die 2 times and getting a 5 both times**

Which is not an example of dependent events?

1. Choosing the starting lineup for the baseball team
2. Picking an ace out of a deck of cards, then picking another ace out of the same deck without replacing the first
3. Picking an ace out of a deck of cards, then picking another ace out of the same deck after replacing the first card
4. Picking which 2 friends to take with you to a concert

Which is not an example of dependent events?

1. Choosing the starting lineup for the baseball team
2. Picking an ace out of a deck of cards, then picking another ace out of the same deck without replacing the first
3. **Picking an ace out of a deck of cards, then picking another ace out of the same deck after replacing the first card**
4. Picking which 2 friends to take with you to a concert

Which is not an example of independent events?

1. Picking the person who scored the most points at the basketball game, and the person who got the most rebounds
2. Picking a president and vice president
3. Picking an ace out of a deck of cards, then picking another ace out of the same deck after replacing the first card
4. Picking two people to get food for lunch

Which is not an example of independent events?

1. Picking the person who scored the most points at the basketball game, and the person who got the most rebounds
2. **Picking a president and vice president**
3. Picking an ace out of a deck of cards, then picking another ace out of the same deck after replacing the first card
4. Picking two people to get food for lunch

The events above are defined as dependent because each event is related to the others. That is, the probability of event E's occurence changes when events A, B, C, or D occur. For example, if it is known that event A occurs, the probability of event E will be much greater. That is to say, $P(E)$ is dependent on event A; the probability of A is dependent on event B, etc..

Notation

Conditional Probability is notated by using the pipe symbol or `|`. The example below represents the probability of event E occuring, given that event A has occurred or will occur based on some assumption, presumption, or previous knowledge:

$$P(E \mid A)$$

Definition: Independence

Two events are independent, statistically independent, or stochastically independent if the occurrence of one does not affect the probability of occurrence of the other.

$$P(A) = P(A \mid B) = P(A \mid B^C)$$

Example #1

Determine whether the events "drawing a heart" and "drawing a six" are independent.

- Step 1: Find $P(\text{"six"} \mid \text{"heart"})$
  - $P(\text{"six"} \mid \text{"heart"}) = \frac{1}{13}$
- Step 2: Find $P(\text{"six"})$
  - $P(\text{"six"}) = \frac{4}{52} = \frac{1}{13}$
- Step 3: Check for equivalence
  - $P(\text{"six"}) = P(\text{"six"} \mid \text{"heart"})$

Solution

$$P(\text{"six"}) = P(\text{"six"} \mid \text{"heart"}) = \frac{1}{13}$$

Example #2

Unlike the example above, color and suit are not independent. Without any additional information, the probability that a randomly selected card will be a diamond is 1/4. However, if it is known that the card is red, the probability that it is also a diamond is 1/2 (half of the red cards in the deck are diamonds). Furthermore, if it is known that the card is black, the probability that it is also a diamond is 0, or has zero probability.

- Step 1: Find $P(\text{"diamond"} \mid \text{"red"})$
  - P(\text{"diamond" | "red"}) = \frac{13}{26} = \frac{1}{2})
- Step 2: Find $P(\text{"diamond"})$
  - $P(\text{"diamond"}) = \frac{13}{52} = \frac{1}{4}$
- Step 3: Check for equivalence
  - $\frac{1}{2} \neq \frac{1}{4}$

Solution

Being that the probabilities are not equal, it can be determined that these two events are not independent

$$P(\text{"diamond"} \mid \text{"red"}) \neq (P(\text{"diamond"})$$

Multiplication Rule for Independent Events

The multiplication rule for independent events relates the individual probabilities of multiple events to the probability that they all occur. In order to apply this rule, one must first determine that the events are independent. Secondly, it is necessary to have the probability of each event to complete the calculation.

Assuming independence, the multiplication rule is defiend as below:

ⓘ Definition: Multiplication rule for independent events

$$P(AB) = P(A) * P(B)$$

Note:

This is different than the general multiplication rule, which can be applied to find the probability of intersection of two events regardless of whether they are independent. The general multiplication rule is introduced in the conditional probability section.

Mutual Exclusivity

- Mutually exclusive events are related in such a way that one event, makes all other events impossible
- In the random experiment defined by flipping a single coin, the only possible results are heads and tails. If the result is heads, there is no way for it also to be tails, and vice-versa.

Definition: Mutual Exclusivity

In logic and probability theory, two events (or propositions) are mutually exclusive, or disjoint, if they cannot both occur at the same time.

Two events, A and B are characterized as *mutually exclusive* when $AB = \emptyset$, that is the intersection of two mutually exclusive events is the empty set, and therefore impossible.

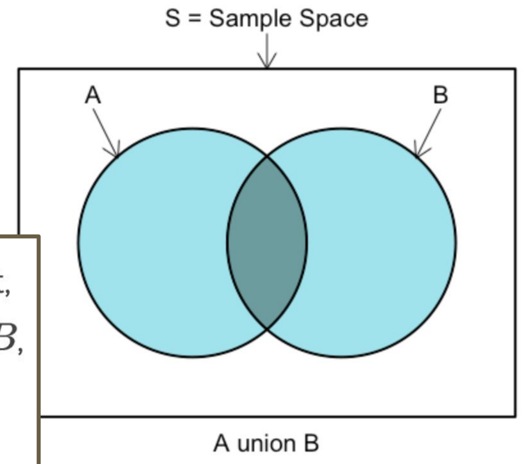
Given the random experiment of flipping a coin, an event A can be interpretted as getting a heads on the coin flip, and an event B can be considered the coin landing on tails. The intersection of the two events results in the empty set, that is, there is no way that both events can occur as the result of the same random experiment.

Inclusion-Exclusion Principle

- The inclusion-exclusion principle provides a way to calculate the union of two (or more) events. This is an especially helpful tool when it comes to calculating probabilities of events

Inclusion-Exclusion for Two Events

Given two events A and B, and their respective probabilities, they cannot be simply added to calculate the union, because the items in the intersection of the two events will be counted twice. See diagram below for explanation:

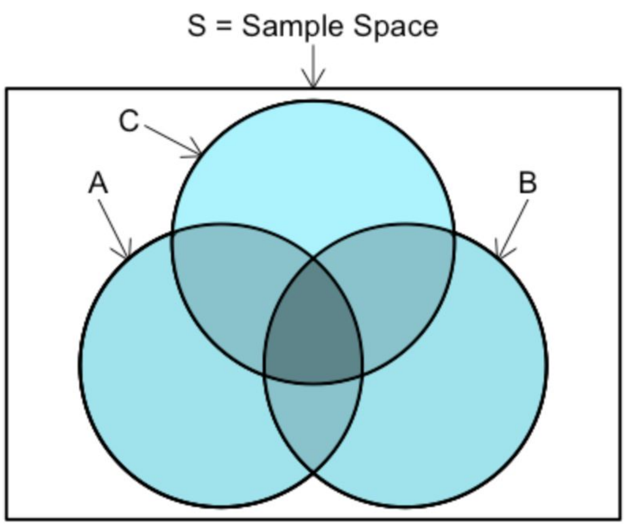


The darker shaded region on the venn diagram visualizes the overlap of each event, portraying the area to be counted twice. To account for this, when calculating $A \cup B$, simply substract AB from $P(A) + P(B)$. See the definition of the inclusion-exclusion principle for two events below:

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

Inclusion-Exclusion for Three Events

When applying the inclusion-exclusion principle for three events, a similar approach must be taken. However, there is a difference when we extend the principle. See the venn diagram below.



A union B union C

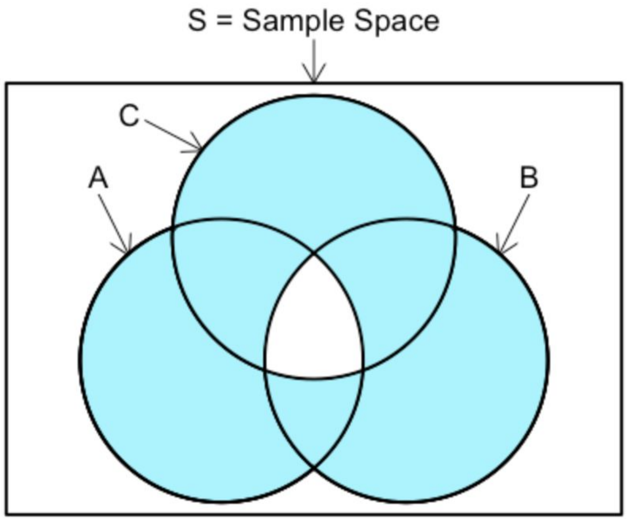
Notice, there areas which are counted twice (each of the two-way intersections - similar to what is described in the union of two events), and the intersection of all three events in counted three times. First, we will apply the subtractive method similar to the two-way inclusion-exclusion principle is applied.

$$P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC)$$

While this is effective in removing the areas which are double-counted, the threeway intersection is now not being counted at all. See venn Diagram below.

To account for this, the value of the threeway intersection can simply be added back in to find the union of the three events. See the extended version of the inclusion-exclusion principle below.

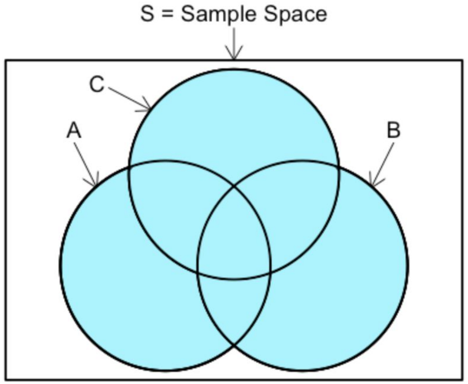
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$



A union B union C

As described above, the pattern shows that first all of the singular probabilities are added together. Second, all of the two-way intersections are subtracted; lastly the three-way intersection is added. This pattern for inclusion-exclusion continues.

That is, for **any** number of events, the probability of all intersections which involve an odd number of events (including the singular probabilities) are added, and the probabilities of all intersections which involve an even number of events are subtracted.



A union B union C

Cartesian Product and Factorial

Combinatorics

Combinatorics: a field of mathematics concerned with counting the possible orderings and combinations of elements within a set, has many applications in data science, including database architecture and graph theory. Here, we will be using it in the service of our probability studies to calculate the number of possible events within a sample space.

Cartesian Product

When dealing with two sets, let's say S represents the set of shirts, and P represents the set of pants that a Galvanize student, Carl, owns. The cartesian product represents all possible combinations of pants and shirts that Carl could wear on a given day (assuming all the shirts and pants are clean, and that only one of each can be worn at a time). See below for an example:

$$P = \{\text{Blue Jeans, Slacks}\}$$

$$S = \{\text{Black T-shirt, Blue collared Shirt}\}$$

Now, the list of all possible combinations using one item from each set and expressing them as an ordered pair, the cartesian product of S and P, can be expressed as:

$$\boxed{P \times S = \{(\text{Jeans, Black shirt}), (\text{Jeans, Blue shirt}), (\text{Slacks, Black shirt}), (\text{Slacks, Blue shirt})\}}$$

Notice that the result is in and of itself a new set, specifically, a set of ordered pairs. You can also see that there are four objects in this new set. That is equal to the cardinality (size) of each set multiplied together, that is:

$$|P| = 2 \text{ and } |S| = 2$$

$$|P \times S| = |P| \times |S| = 2 \times 2 = 4$$

Factorial

Many mathematicians interpret a factorial simply as a specific multiplication problem, however, in reality a *factorial* represents the number of ways a set objects can be arranged.

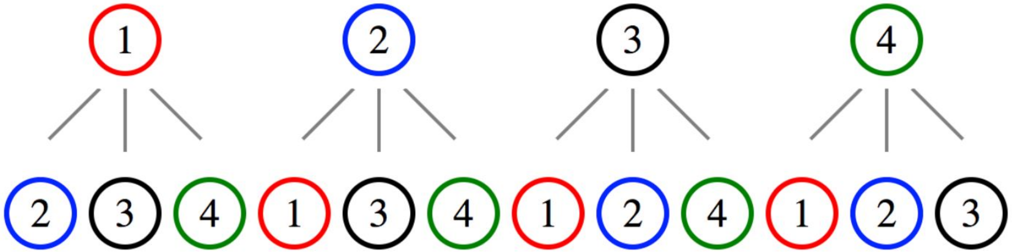
Notations: A *factorial* is represented by an integer followed the exclamation mark, **!**.

The factorial of a positive integer n is computed by taking the product of n and all positive integers less than n. For example:

$$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

If there are four balls in a jar and one is randomly selected, there are four possible outcomes:

If a second ball is subsequently selected, then, for each of the four options in the first draw, there are three remaining possibilities for the second ball (assuming the first ball has not been returned to the jar). This is called selection ***without replacement***:



There will then be two remaining possibilities for the third draw, and one remaining possibility for the fourth, for a total of 24 possible orderings. To determine the total possible number of outcomes, factorial can be applied.

Solution:

$$4 \cdot 3 \cdot 2 \cdot 1 = 4! = 24$$

The factorial of four, is twenty four. That is, there are twenty four different ways that four balls can be sequentially chosen, without replacement.

Zero Factorial

Zero is a special case of factorial, and only through the understanding that factorial represents the possible number of arrangements does it make intuitive sense. That is, how many ***different*** ways can one arrange zero objects? Simple, there is one way, and that is by not arranging them at all!

$$0! = 1$$

Permutations

- A permutation is one of several possible variations in which a set or number of objects can be ordered or arranged

$n =$ the number of possible objects to choose from

AND

$r =$ the number of balls being chosen

Definition: Number of Possible Permutations

$$nPr = \frac{n!}{(n-r)!}$$

Combinations

- A combination is much like a permutation, though when dealing with combinations, order doesn't matter.
- When finding the number of possible combinations, it is akin to asking: "How many different subsets of a specified size can be made from the original set".

Definition: Number of Possible Combinations

$$combinations(n, r) = nCr = \frac{n!}{(n-r)!\,r!}$$

Notation:

It is common to see the following common notations:

- nCr
- n choose r
- C_r^n
- $\binom{n}{r}$

Regardless of which notation is used, these symbols are always read as "n choose r."

Contents - 2

- Use conditional probability to confirm independence
- Use the multiplication rule to confirm independence
- Calculate basic conditional probabilities
- Apply the Law of Total Probability
- Apply Bayes' Theorem to calculate conditional probabilities

Conditional Probability

Conditional probability is a measure of the probability of an event occuring given that some other (different) event has or will occur. The expression $P(A \mid B)$, can be interpretted as "The probability of A given B." Implicitly there is a need to presume, assume, assert, or see evidence that event B has occured.

Conditional probability can be defined as the likelihood of an event or outcome occuring, based on the occurence of some previous (or concurrent) event(s) also occuring. Conditional probability is fundamental to one's knowledge of probability theory; in our complex and interconnected world, calculating the probability of a given event often relies upon the knowledge of an occurence of some other previous or concurrent event(s). When conditioning on an event, the first and most common definition used to calculate a probability is known as Kolmogorov's definition of conditional probability, see a definition below:

ⓘ Definition: Kolmogorov's Definition of Conditional Probability

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

That is, the conditional probability of A given B is the quotient of the probability A intersect B, divided by the probability of B. This can be visualized as limiting the sample space to only outcomes in which B occurs, and considering it our new sample space. The numerator accounts for all outcomes of event A which exist in our new, restricted sample space; this numerator is simply the intersection of A and B.

Example #1

A computer science professor has given their class two exams; eighty percent of the class passed the first exam, and 77% percent of the class passed both exams. What is the probability that a randomly selected student passes the second exam given they passed the first?

- Step 1. Convert the information given into probabilities
  - $P(E_1) = \text{the probability that the first exam is passed} = 0.8$
  - $P(E_1E_2) = \text{the probability that both exams are passed} = 0.77$

- Step 2. Apply the definition of conditional probability to determine the probability
  - $P(E_2 \mid E_1) = \frac{P(E_1E_2)}{P(E_1)}$
  - $P(E_2 \mid E_1) = \frac{0.77}{.8} = .9625$

Solution:

$$P(E_2 \mid E_1) = 0.9625$$

By using the definition, it can be determined that the probability of a student passing the second exam given they passed the first is 0.9625 or 96.25%

Conditional Probability as an Axiom

Most probabilists approach conditional probability as an axiom, or fundamental truth, of probability theory. This common axiom can be derived from Kolmogorov's definition with an algebraic rearrangement, or in other words by solving for $P(AB)$, shown below:

Definition: General multiplication rule

$$P(AB) = P(A \mid B) * P(B)$$

This interpretation of conditional probability is fundamental to more advanced conditional probability laws such as the law of total probability and Bayes Theorem. Also this axiom is also known as the "General Multiplication Rule," as it provides a means of calculating the probability of an intersection regardless of whether the involved events are independent or non-independent. See an example application of this axiom to calculate the probability of an intersection:

Two individuals are each dealt a single card from a well-shuffled standard 52-card deck. Let the event A be the event that the first person has an Ace, and event B be the event that the second has an Ace. Note that the two cards are drawn in order, with the first person drawing their card first. Calculate the probability that each player gets an ace using the general multiplication rule:

- Step 1. Compute the probabilities of each event
  - $P(A) = \frac{4}{52} = 0.077$
  - $P(B \mid A) = 3/51 = 0.059$
- Step 2. Apply the general rule of multiplication
  - $P(AB) = P(B \mid A) * P(A)$
  - $P(AB) = 0.059 * 0.077 = 0.0045$

Solution:

$$P(AB) = 0.0045$$

The result of our calculation is intuitive, in that is relatively unlikely that either player get an ace; however, it can be seen that both players drawing an ace is even less likely.

Using Conditional Probability to Test for Independence

Being able to compute conditional probabilities, there is now an ability to explicitly test any two events for independence. See an example of this application below. Although in this example one could easily determine independence using intuition, applying the explicit test is good practice for events which might be more difficult to determine independence through intuition. To show that two events are independent it must be shown that $P(A) = P(A \mid B) = P(A \mid B^C)$. In words, this mathematical expression says, the probability of the event A occuring remains the same whether or not the event B has occured.

Consider two sequential flips of a fair coin. Let A represent the event that the coin lands on heads on the first toss; let B represent the event that the coin lands on heads on the second toss.

$$P(A) = 0.5 \text{ and } P(B) = 0.5$$

Consider the the probability of $P(B \mid A)$; being that the second coin flip is not affected by the first, we can conclude that $P(B \mid A) = 0.5$ and $P(B \mid A^C) = 0.5$.

Solution:

$$P(B) = P(B \mid A) = P(B| \; A^C) = 0.5$$

By the result above, it can be confirmed that these two events are indeed independent.

Multiplication Rule and iid

- While the multiplication rule itself may seem intuitive to many, it introduces the concept of independent and identically distributed (iid) experiments
- This condition of iid is critical to the validity of many studies; that is, researchers must be able to perform an experiment repeatedly under identical conditions in order to have enough information to make statistical inferences

All of the previous counting procedures: cartesian product, factorial, permutations, and combinations, had a kind of "sequential" and "order of process" feel to them. To really accentuate this characteristic consider the following question: "How many four-letter strings can you make out of the 26 letters of the English alphabet?"

The answer is that there are twenty-six options for the first character, twenty-six options for the second character, twenty-six options for the third, and twenty-six for the last. That's 26^4. This phenomenon is known as the ***multiplication rule***. Notice how for the multiplication rule, unlike for factorials, permutations, and combinations -- the number of options didn't change as the process progressed along. Why is that?

Well, it has to do with the notion of independence, which has been previously defined as a characteristic of two events E_1 and E_2 indicating that $P(E_2 \mid E_1) = P(E_2)$ and $P(E_1 \mid E_2) = P(E_1)$ or equivalently, that $P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$. But the characteristic shown here is actually a little bit stronger than just independence. It is actually the notion that randomly choosing a letter from the alphabet is unaffected by the previous or subsequent letters drawn (this is indeed the independence) but also that the options available from one step to the next do not change. This means the randomness in question is *identical* as well as being independent. This is a very central and powerful notion in statistics and such a process is referred to as being ***independently and identically distributed***, or ***iid***, which is the common shorthand for this property.

What is actually being shown with the application of the multiplication rule is a sequence of iid events. Whereas in the case of factorials, permutations, and combinations there is actually a single joint event that can be expressed using the chain rule as a series of conditional events. I.e., distinction between

$$P(A_1 \cap A_2 \cap \cdots \cap A_m) = P(A_1) \cdot P(A_2) \cdots P(A_n)$$

and

$$P(A_1 \cap A_2 \cap \cdots \cap A_m) = P(A_1)P(A_2 \mid A_1) \cdots P(A_m \mid A_{n-1}, A_{n-2}, \cdots, A_1)$$

or independence and the chain rule, where the former is actually made up of **n** i.i.d. experiments and the latter is actually just one single complicated experiment with lots of dependencies between the different components (events).

Here the discussion revolves around probabilities, where previously the discussion was largely talking about different counting methods. How do these notions share this connection? Well, it turns out that all probability boils down to is the notion of set size, which is a count. The probability is calculated by dividing your set size in question (the number of possible outcomes that comprise an event of interest or cardinality) by the total number of outcomes in the whole sample space S (which is of course also a set -- but just the set of all possible sample points x of the random experiment X). Because counting is so fundamental to probability, it should be no surprise to find that there is a special notation to designate the size, or the ***cardinality*** (i.e., number of elements) of a set; namely,

$$card(A_1) = |A_1|$$

Suppose that an engineering firm has made three bids on three different construction projects. Let A represent the event that the firm is selected for the first project, B for the second, and C for the third. Given the probabilities below, find the probability that the firm is selected for all three projects assuming that recieving any one contract has no bearing on recieving the others:

$$P(A) = 0.83, \text{ and } P(B) = 0.79, \text{ and } P(C) = 0.41$$

- Step 1: Devise a strategy to make the calculation
  - Because the three events are explicitly independent, the multiplication rule can be applied.
  - $P(ABC) = P(A) \cdot P(B) \cdot P(C)$
- Step 2: Complete the calculation
  - $P(ABC) = P(A) \cdot P(B) \cdot P(C) = 0.83 \cdot 0.79 \cdot 0.41 = 0.269$

Solution:

$$0.269$$

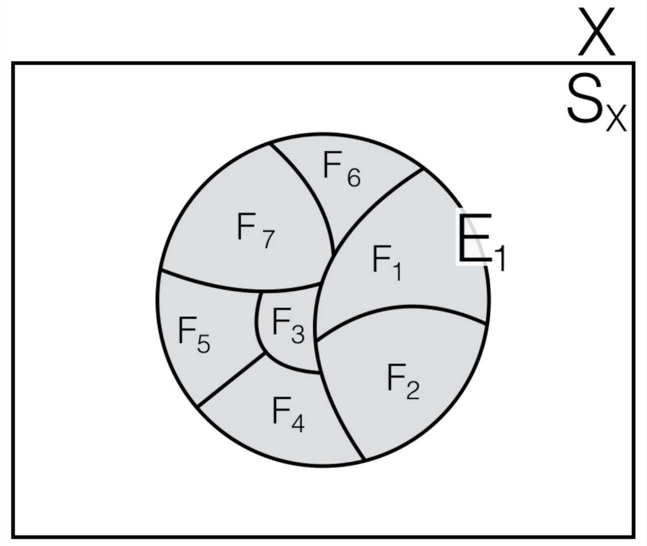
Through the application of the multiplication rule for independent events, it can be shown that the probability of the firm getting all three contracts is approximately 0.269 or 26.9.

The Law of Total Probability

You cannot always measure the probability of the precise event that is of interest to you. Rules of formal probability, including the Law of Total Probability, allow you to use the information you have to draw conclusions about these less accessible events.

See below for the use of a venn diagram to describe the law of total probability, where the F_i events represent mutually exclusive partitions of a greater event E_1:

Figure #1: A set of F events comprise a greater event E

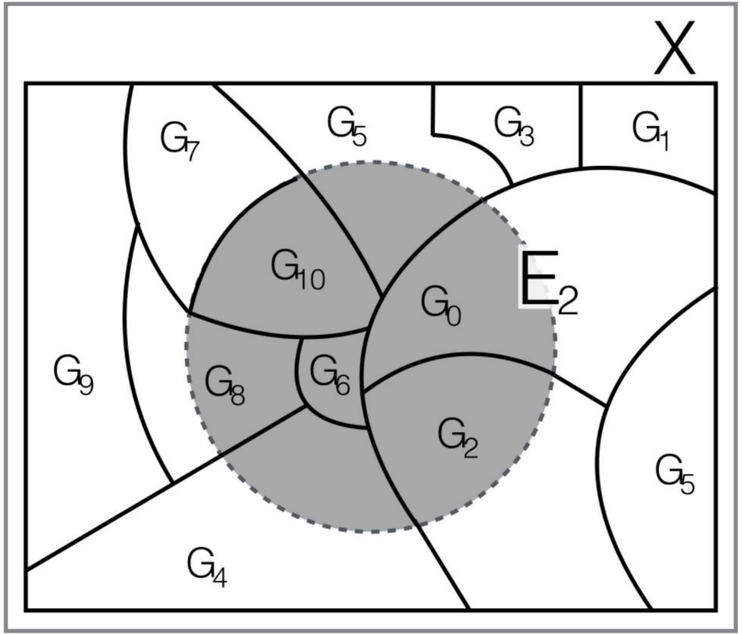


This figure shows that the law of total probability simply calculates the probability of event E_1 based on the sum of probabilities for several underlying disjoint events, in this case, $F_1, F_2, F_3, \cdots, F_n$. With this understanding, it's pretty easy to see why this is called the law of total probability. Now, the trick here that makes the law of total probability work is that the set of disjoint events $F_1, F_2, F_3, \cdots, F_n$ are mutually exclusive (i.e., $F_i \cap F_j = \emptyset$ for all $i \neq j \text{ and } i \text{ and } j \text{ between } 1 \text{ and } n$) and when unioned together perfectly construct the event E_1 (i.e, $F_1 \cup F_2 \cup F_3 \cup \cdots \cup F_n = E_1$). Such a set of F's is known as a_partition_, or it can be said that these events partition the sample space.

It turns out the law of total probability works for any partition of the entire sample space S_X, so long as that partition gets intersected with the event in question. So, the general form of the law of total probability is, for any partition of the sample space S_X, for some partitioning set of disjoint events $G_1, G_2, G_3, \cdots, G_n$. It can be see in the diagram below that the total area, or probability for the event E_2 can also be expressed as shown below:

Figure #2: Law of total probability, shown as the sum of intersections

$$P(E_2) = P(E_2 \cap G_0) + P(E_2 \cap G_1) + P(E_2 \cap G_2) + \cdots + P(E_2 \cap G_n)$$



$$P(E_2) = P(E_2|G_0) * P(G_0) + P(E_2|G_1) * P(G_1) + P(E_2|G_2) * P(G_2) + \cdots + P(E_2|G_n) * P(G_n)$$

Definition: Law of Total Probability

$$P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + \cdots + P(A|B_n) \cdot P(B_n)$$

Where all B events partition the sample space

Suppose that a school will be closed 90% of the time when it snows, and 5% of the time when it rains, and 1% of the time when it neither snows nor rains. Suppose you are interested in calculating the probability that the school will be closed on any given day. You can use the law of total probability to calculate that

$$P(\text{"closed"}) = P(\text{"snow"}) * P(\text{"closed"} \mid \text{"snow"})$$

$$+P(\text{"rain"}) * P(\text{"closed"} \mid \text{"rain"})$$

$$+P(\text{"not rain or snow"}) * P(\text{"closed"} \mid \text{"not rain or snow"})$$

But of course this doesn't get you all the way there -- you also need to know the probability that it will rain or snow on any given day. Say that:

- $P(\text{"snow"}) = 0.02$
- $P(\text{"rain"}) = 0.15$
- $P(\text{"not rain or snow"}) = 0.83$

Notice here that the sum $P(\text{"snow"}) + P(\text{"rain"}) + P(\text{"not rain or snow"})$ is equal to one. This must be the case because snow, rain, and not rain or snow must be a partition of the sample space (and hence sum to one). Now, with this information you can compute your desired probability

$$P(\text{"closed"}) = 0.02 * 0.90 + 0.15 * 0.05 + 0.83 * 0.01$$

$$= 0.018 + 0.0075 + 0.0083 = 0.0338$$

Law of total probability

Example: Choosing a Fruit
Suppose you have a box of fruits containing apples, oranges, and bananas. The probabilities of selecting each fruit are as follows:
P(Apple) = 0.4, P(Orange) = 0.3 , P(Banana) = 0.3

Now, let's say you want to calculate the probability of randomly selecting a ripe fruit. You know that the probability of a fruit being ripe depends on its type. The probabilities of a fruit being ripe given its type are:
P(Ripe|Apple) = 0.8
P(Ripe|Orange) = 0.6
P(Ripe|Banana) = 0.5

Using the law of total probability, we can calculate the probability of selecting a ripe fruit:
P(Ripe) = P(Ripe|Apple) * P(Apple) + P(Ripe|Orange) * P(Orange) + P(Ripe|Banana) * P(Banana)
= 0.8 * 0.4 + 0.6 * 0.3 + 0.5 * 0.3 = 0.32 + 0.18 + 0.15 = 0.65

Therefore, the probability of randomly selecting a ripe fruit from the box is 0.65.

Bayes' Theorem

Bayes' theorem describes the probability of a **posterior** event, based on previous conditions related to the event. For example, given the probabilities $P(B)$ and $P(B \mid A)$, Bayes' theorem can be applied to calculate $P(A \mid B)$.

ⓘ **Definition: Bayes' theorem**

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

There are two bowls of marbles: bowl #1 contains 30 blue marbles and 10 red marbles, and bowl #2 contains 20 blue marbles and 20 red marbles. If you choose a bowl at random, and without looking pick a blue marble, what is the probability that this blue marble came from bowl #1?

How to approach a solution

1. Translate the question from words to math: "what is the probability of having picked from bowl #1, given that we grabbed a blue marble" can be rewritten as

$$P(bowl_1 \mid blue)$$

2. Use Bayes' theorem from above by simply plugging in values for events A and B where event $A = bowl_1$, and event $B = blue$:

$$P(bowl_1 \mid blue) = \frac{P(blue \mid bowl_1) \cdot P(bowl_1)}{P(blue)}$$

This is an approach that will yield an answer, but may not grant much understanding. That said, let us continue forward to a solution.

Immediately observable probabilities

Write down the information we already have:

- $P(blue \mid bowl_1) = \frac{30}{40} = 0.75$
  - This expression is asking "What is the probability of choosing a blue marble from bowl 1?"
  - 30 of the 40 marbles in bowl #1 are blue
- $P(bowl_1) = P(bowl_2) = 0.5$
  - Since we have not stated this explicitly, we can assume that there is an equal chance of choosing bowl #1 or bowl #2.
- $P(blue \mid bowl_2) = \frac{20}{40} = 0.5$
  - This expression is asking "What is the probability of choosing a blue marble from bowl #2?"
  - 20 of the 40 marbles in bowl #1 are blue

Calculating $P(blue)$

This is not an immediately observable probability, so we must apply the **law of total probability,** mathematically written as

$$P(B) = P(A) \cdot P(B \mid A) + P(\text{not } A) \cdot P(B \mid \text{not } A)$$

Plugging our values in, we get

$$\begin{aligned} P(blue) &= P(bowl_1) \cdot P(blue \mid bowl_1) + P(bowl_2) \cdot P(blue \mid bowl_2) \\ &= 0.5 \cdot 0.75 + 0.5 \cdot 0.5 \\ &= 0.625 \end{aligned}$$

Plug in our observations and calculations

Finally, we can plug all the pieces into the Bayes' theorem formula from above:

$$P(bowl_1 \mid blue) = \frac{P(blue \mid bowl_1) \cdot P(bowl_1)}{P(blue)}$$

$$= \frac{0.5 \cdot 0.75}{0.625}$$

$$= 0.6$$

And there we have it: given that you drew a blue marble, there is a probability of 0.6 that you pulled the marble from bowl #1.

But Why?

Since each bowl is picked from with equal probability, intuition may lead one to believe that the probability of the blue marble coming from either bowl would likewise be equal. However, it is important to note that the proportion of blue marbles in each bowl differs, which affects our probability calculation. Intuition may also lead one to approach this problem with simple arithmetic, but this is not always the best approach.

For a moment, let's approach this problem using simple arithmetic. Count the blue marbles in the first bowl and divide that by the total count of all of the blue marbles in both bowls:

1. Total number of blue marbles: $30 + 20 = 50$
2. Blue marbles in bowl #1: 30
3. Blue marbles in bowl #1 divided by total number of blue marbles: $\frac{30}{50} = 0.6$

The same conclusion is reached when approaching this problem through the use of arithmetic, however this approach is less robust. If the probability of picking a marble from any given bowl changes (i.e., no longer equal probability), our simple arithmetic approach becomes tedious and untenable. Though it involves more work, using Bayes' theorem in any case is the preferred approach.

Thank You