Contents

1 Counting and Probability

In this chapter our primary goal is to discuss probability, which quantifies how likely it is that a particular event will occur. We will begin with a brief review of various basic properties of sets and set operations, and then introduce basic counting principles, permutations, combinations, and binomial coefficients.

We then develop the fundamentals of discrete probability (which arose historically from the study of games of chance) using the results on sets and counting techniques we have developed. We then treat conditional probability (which allows us to compute probabilities based on the addition of new information) and the related topic of independence (which describes whether two events are related) and then use these principles to make more general probability calculations. We close with a discussion of some classic probability puzzles such as the Monty Hall problem and the birthday problem, along with some applications of Bayes' formula.

1.1 Sets and Set Operations

- In order to discuss probability, we will require some basic counting techniques, which are in turn ultimately grounded in properties of sets.
- We begin by reviewing basic set operations such as union and intersection, along with ways of visualizing sets, particularly Venn diagrams. We then discuss a number of techniques for solving various counting problems that serve as a prelude to discrete probability, where we will frequently need to enumerate the set of outcomes of an event such as rolling a pair of dice.

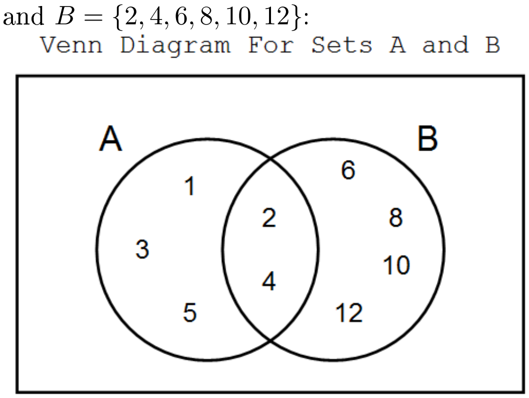
1.1.1 Sets, Subsets, and Cardinality

- Definition: A set is a well-defined collection of distinct elements.
  - The elements of a set can be essentially anything: integers, real numbers, other sets, people.
  - Sets are generally denoted by capital or script letters, and when listing the elements of a set, curly brackets $\{\cdot\}$ are used.
  - Sets do not have to have any elements: the empty set $\emptyset = \{\ \}$ is the set with no elements at all.
  - Two sets are the same precisely if all of their elements are the same. The elements in a set are not ordered, and no element can appear in a set more than once: thus the sets $\{1, 4\}$ and $\{4, 1\}$ are the same.
- There are two primary ways to describe a set.
  - One way is to list all the elements: for example, $A = \{1, 2, 4, 5\}$ is the set containing the four numbers 1, 2, 4, and 5.
  - The other way to define a set is to describe properties of its elements[1]: for example, the set S of one-letter words in English has two elements: $S = \{\text{a, I}\}$.
- We often employ "set-builder" notation for sets: for example, the set S of real numbers between 0 and 5 is denoted $S = \{x : x \text{ is a real number and } 0 \le x \le 5\}$.
  - Some authors use a vertical pipe $|$ instead of a colon $:$ but this distinction is irrelevant.
  - Example: $S = \{n : n \text{ is a positive integer less than } 6\} = \{1, 2, 3, 4, 5\}$.
- In practice, to save time we will often not give a totally explicit description of a set when we are describing a pattern that is clear from the context.
  - For example, if we write $A = \{1, 2, 3, 4, \dots, 10\}$, we mean that A is the set of positive integers from 1 to 10 (so that, explicitly, $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$).
- Notation: If S is a set, $x \in S$ means "x is an element of S", and $x \notin S$ means "x is not an element of S".
  - Example: For $S = \{1, 2, 5\}$ we have $1 \in S$ and $5 \in S$ but $3 \notin S$ and $\pi \notin S$.
  - Example: For S equal to the set of English words starting with the letter A, we have apple $\in S$ and antlers $\in S$, while potatoes $\notin S$.
- Definition: If A and B are two sets with the property that every element of A is also an element of B, we say A is a subset of B (or that A is contained in B) and write $A \subseteq B$.
  - Example: If $A = \{1, 2, 3\}$, $B = \{1, 4, 5\}$, and $C = \{1, 2, 3, 4, 5\}$, then $A \subseteq C$ and $B \subseteq C$ but neither A nor B is a subset of the other.
  - Example: If S is the set of all English words and T is the set of all English words starting with the letter P, then $T \subseteq S$.
  - Example: If A is any set, then the empty set $\emptyset$ is contained in A.
- Warning: Subset notation is not universally agreed-upon: the notation $A \subset B$ is also commonly used to say that A is a subset of B.
  - The difference is not terribly relevant except for when A can be equal to B: some authors allow $A \subset B$ to include the possibility that A could be equal to B, while others insist that $A \subset B$ means that A is a subset of B which cannot be all of B.
  - We will always use the notation $A \subseteq B$ to include the possibility that $A = B$, and if we wish to say that A is a subset of B that is not equal to B, we will write $A \subsetneq B$.
- Definition: If A is any set, the cardinality of A, denoted $\#A$ or $|A|$, is the number of elements of A.
  - Example: For $A = \{1, 2, 3\}$ and $B = \{2, 4, 6, 8, 10, \dots, 100\}$, then $\#A = 3$ and $\#B = 50$.
  - Example: The cardinality of the empty set $\emptyset$ is 0.
  - Example: The cardinality of the set $\{1, 2, 3, 4, \dots\}$ of positive integers is ∞.

[1] It is possible to run into trouble by trying to define sets in this "naive" way of specifying qualities of their elements. In general, one must be more careful when defining arbitrary sets, although we will not worry about this.

1.1.2 Intersections, Unions, and Complements

- Definition: If A and B are two sets, then the intersection $A \cap B$ is the set of all elements contained in both A and B. The union $A \cup B$ is the set of all elements contained in either A or B (or both).
  - Example: If $A = \{1, 2, 3\}$ and $B = \{1, 4, 5\}$, then $A \cap B = \{1\}$ and $A \cup B = \{1, 2, 3, 4, 5\}$.
  - Example: If $E = \{2, 4, 6, 8, \ldots\}$ is the set of all positive even integers and $O = \{1, 3, 5, 7, \ldots\}$ is the set of all positive odd integers, then $O \cap E = \emptyset$ is the empty set (no integer is both even and odd), while $O \cup E = \{1, 2, 3, 4, \ldots\}$ is the set of all positive integers.
  - Example: If $E = \{2, 4, 6, 8, \ldots\}$ is the set of all positive even integers and $S = \{1, 4, 9, 16, \ldots\}$ is the set of all positive perfect squares, then $E \cap S = \{4, 16, 36, 64, \ldots\}$ is the set of all even perfect squares.
  - It is not hard to see that if $A \subseteq B$, then $A \cap B = A$ and $A \cup B = B$: explicitly, if every element of A is contained in B, then the elements common to both are simply the elements of A, while the elements in at least one of the two are simply the elements of B.
- We can also take unions and intersections of more than two sets at a time: the intersection of any collection of sets is the set of elements contained in all of them, while the union of any collection of sets is the set of elements contained in at least one of them.
  - Example: If $A = \{1, 2, 3\}$, $B = \{1, 3, 4\}$, and $C = \{1, 3, 9\}$, then $A \cap B \cap C = \{1, 3\}$ while $A \cup B \cup C = \{1, 2, 3, 4, 9\}$.
  - We will note in passing that $(A \cap B) \cap C = A \cap B \cap C = A \cap (B \cap C)$, since each of these sets represents the elements in all three of A, B, C, and the analogous fact holds for unions. Thus, we do not need to specify the order in which intersections are taken.
  - On the other hand, we cannot mix unions and intersections without specifying the order of operations.
  - Example: If $A = \{1, 2, 3\}$, $B = \{1, 3, 4\}$, and $C = \{1, 3, 9\}$, then $(A \cap B) \cup C = \{1, 3\} \cup \{1, 3, 9\} = \{1, 3, 9\}$ while $A \cap (B \cup C) = \{1, 2, 3\} \cap \{1, 3, 4, 9\} = \{1, 3\}$.
  - We can see that the expression "$A \cap B \cup C$" therefore does not make sense[2], because it is not immediately clear which of the two expressions $(A \cap B) \cup C$ and $A \cap (B \cup C)$ it is supposed to mean.
- A very useful tool for visualizing unions and intersections of sets are Venn diagrams, in which we represent each set as a region, with overlaps of regions corresponding to intersections of the sets in such a way that any possible combination of intersections corresponds to a portion of the diagram.
  - An example makes the idea clearer than a description in words; here is a Venn diagram corresponding to the sets $A = \{1, 2, 3, 4, 5\}$ and $B = \{2, 4, 6, 8, 10, 12\}$:



  - In the Venn diagram above, we place all of the elements of A into the region (in this case, a circle) labeled A, and likewise we place all the elements of B into the region labeled B, with elements in both sets (i.e., in $A \cap B$) in the overlap between the two regions.

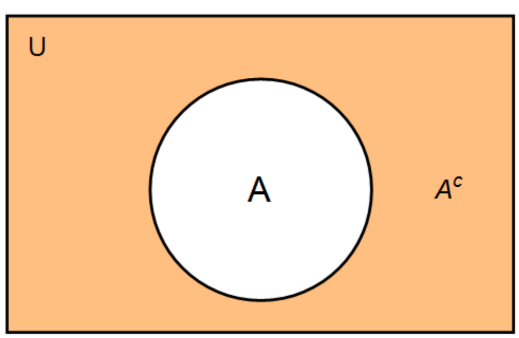
[2]It is a moderately common convention that intersections are always performed before unions (much like in the order of operations for arithmetic, in which multiplications are performed before additions), in which case we would always interpret $A \cap B \cup C$ to mean $(A \cap B) \cup C$.

- Using a Venn diagram, we can see that there is a relationship between the cardinalities of A, B, $A \cup B$, and $A \cap B$:
- Theorem (Sizes of Unions and Intersections): If A and B are any sets, then $\#(A \cup B) + \#(A \cap B) = \#A + \#B$.
  - Proof: Observe that if we count the total number of elements in A and add it to the total number of elements in B, then we have counted every element in the union $A \cup B$ (since every element in $A \cup B$ is either in A or in B) but we have double-counted the elements in the intersection $A \cap B$.
  - Therefore, both expressions $\#(A \cup B) + \#(A \cap B)$ and $\#A + \#B$ count every element in $A \cap B$ twice and every other element once, so they are equal.
- Example: A survey of 100 pet owners shows that 55 own a cat and 61 own a dog, and none have any other pets. How many owners have both a cat and a dog?
  - If we let A denote the set of cat owners in the survey and B denote the set of dog owners in the survey, then the given information says that $\#A = 55$, $\#B = 61$, and $\#(A \cup B) = 100$.
  - Therefore, we see that $\#(A \cap B) = \#A + \#B - \#(A \cup B) = 55 + 61 - 100 = 16$, meaning that $\boxed{16}$ owners have both a cat and a dog.

1.1.3 Complements, Universal Sets, and Cartesian Products

- In many contexts, it is useful to think of all the sets we are discussing as being subsets of some particular larger set S, which we refer to as a "universal set" of elements under consideration.
  - In general, we must always specify precisely what this universal set U is, unless it is clear from context. For example, if we are discussing sets of integers, a sensible choice is to take U to be the set of integers, but there is no reason we couldn't instead take U to be the set of all real numbers.
  - It might seem to be convenient to use the same universal set in all contexts, but it turns out that assuming the existence of a general "universal set" of all possible elements leads to logical contradictions[3].
  - If we have chosen a suitable universal set U and A is a subset of U, then we may speak of the elements of A not in U.
- Definition: If U is a universal set and $A \subseteq U$, then the complement of A (as a subset of U), denoted as A^c, is the set of elements of U not in A.
  - Notation: Other notations used for the complement of A as a subset of U include A', $\overline{A}$, $U \backslash A$, and $U - A$.
  - Example: With universal set $U = \{1, 2, 3, 4, 5, 6\}$, if $A = \{1, 3, 4\}$ and $B = \{1, 2, 3, 4, 5, 6\}$, then $A^c = \{2, 5, 6\}$ and $B^c = \emptyset$.
  - In a Venn diagram, we typically identify the universal set U in the diagram, and represent A^c as the area outside the region marked as A:

Venn Diagram For A and A^c



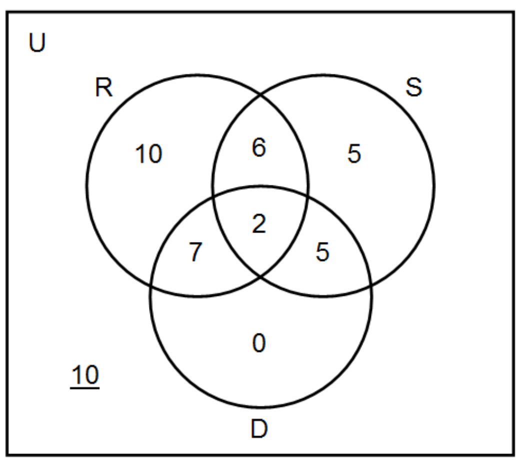
[3] As first noted by Bertrand Russell, if we assume there exists a universal set U containing every other set as a subset, we could define T to be the subset of U consisting of all sets (such as $\{1, 2\}$) that do not contain themselves as an element. Now we ask: is T an element of T? If so, then T would be a set that contains itself, but this contradicts the definition of T. If not, then T is not a set that contains itself, but this would mean that T is an element of T. Either case leads to a logical contradiction, so there cannot exist a universal "set of all sets" U.

- Observe that for any set A, we have $A \cup A^c = U$ (every element is either in A or not in A), $A \cap A^c = \emptyset$ (no element is both in A and not in A), and $(A^c)^c = A$.
- Then by the cardinality formula for unions and intersections, we see that $\#A + \#A^c = \#(A \cup A^c) + \#(A \cap A^c) = \#U$.
- In particular, if A is a finite set, we obtain the formula $\#A^c = \#U - \#A$.

- By using all of this information in tandem with a Venn diagram, we can solve problems involving overlapping categories:

- Example: In a literature class, a total of 45 short stories are read. Of these, 25 are romantic, 18 are science fiction, 14 are dystopian. Furthermore, 8 of the science fiction stories are romantic, 2 of which are also dystopian; also, every dystopian story is either romantic or science fiction, and there are 7 dystopian science fiction stories. Determine the number of short stories that are (i) romantic or dystopian, (ii) non-dystopian science fiction, and (iii) none of the three categories.
  - If we let U be the set of the 45 short stories, R be the romantic stories, S be the science-fiction stories, and D be the dystopian stories, we can make a Venn diagram and label various regions with the corresponding number of stories.
  - From the given information, we know that $\#U = 45$, $\#R = 25$, $\#S = 18$, $\#D = 14$, $\#(R \cap S) = 8$, $\#(R \cap S \cap D) = 2$, $\#(S \cap D) = 7$, and that $\#(D \cap R^c \cap S^c) = 0$ (because there are no stories that are dystopian, but not romantic and not science fiction).
  - Using these values we can fill in all the regions in the Venn diagram. For example, since $\#(R \cap S \cap D) = 2$ and $\#(R \cap S) = 8$, this means that the number of elements in $R \cap S$ not in D must be $8 - 2 = 6$. Similarly, since $\#(S \cap D) = 7$, this means that the number of elements in $S \cap D$ not in R must be $7 - 2 = 5$.
  - Then since the total number of elements in S is equal to 18, we see that the number of elements in S not in R or D must be $18 - 6 - 2 - 5 = 5$. Continuing in this way, we can fill in all of the remaining entries inside the R, S, and D regions:

Venn Diagram For Short Stories



  - Then we see that the number of short stories that are romantic or dystopian is $10+6+7+2+5+0 = \boxed{30}$, the number of non-dystopian science fiction is $6 + 5 = \boxed{11}$, and the number of stories outside the three categories is $\boxed{10}$.

- We will mention one additional set construction that is very useful; namely, the Cartesian product:

- Definition: If A and B are any sets, the Cartesian product is the set $A \times B$ consisting of all ordered pairs (a, b) where $a \in A$ and $b \in B$.
  - Example: If $A = \{1, 2\}$ and $B = \{1, 3, 5\}$, then $A \times B = \{(1, 1), (1, 3), (1, 5), (2, 1), (2, 3), (2, 5)\}$.
  - Example: If $A = \{H, T\}$ then $A \times A = \{(H, H), (H, T), (T, H), (T, T)\}$.

- A common use of the Cartesian product is to list all possible outcomes when one event is followed by another. The second example above indicates the possible outcomes of flipping one coin followed by flipping another coin.

- As suggested by the examples above, there is a simple formula for the cardinality of a Cartesian product: for any sets A and B we have $\#(A \times B) = \#A \cdot \#B$.

1.2 Counting Principles

- Our goal now is to use some of our results on sets and cardinality to solve counting problems.

1.2.1 Addition and Multiplication Principles

- Two fundamental counting principles are as follows:
  - ("Addition Principle") When choosing among n disjoint options labeled 1 through n, if option i has a_i possible outcomes for each $1 \leq i \leq n$, then the total number of possible outcomes is $a_1 + a_2 + \cdots + a_n$.
  - To illustrate the addition principle, if a restaurant offers 5 main courses with chicken, 6 main courses with beef, and 12 vegetarian main courses, then (presuming no course is counted twice) the total possible number of main courses is $5 + 6 + 12 = 23$.
  - The addition principle can be justified using our results about cardinalities of unions of sets: if A_i corresponds to the set of outcomes of option i, then because all of the different options are disjoint, $\#(A_1 \cup A_2 \cup \cdots \cup A_n) = \#A_1 + \#A_2 + \cdots + \#A_n$.
  - ("Multiplication Principle") When making a sequence of n independent choices, if step i has b_i possible outcomes for each $1 \leq i \leq n$, then the total number of possible collections of choices is $b_1 \cdot b_2 \cdot \cdots \cdot b_n$.
  - To illustrate the multiplication principle, if a fair coin is tossed (2 possible outcomes) and then a fair 6-sided die is rolled (6 possible outcomes), the total number of possible results of flipping a coin and then rolling a die is $2 \cdot 6 = 12$.
  - The multiplication principle can be justified using our results about cardinalities of Cartesian products: if B_i corresponds to the set of outcomes of choice i, then by $\#(B_1 \times B_2 \times \cdots \times B_n) = \#B_1 \cdot \#B_2 \cdot \cdots \cdot \#B_n$.

- By combining these principles appropriately, we can solve a wide array of counting problems.

- Example: Determine the number of possible outcomes from rolling a 6-sided die 5 times in a row.
  - Each individual roll has 6 possible outcomes. Thus, by the multiplication principle, the number of possible sequences of 5 rolls is $6^5 = \boxed{7776}$.

- Example: An ice creamery offers 25 different flavors. Each order of ice cream may be served in either a sugar cone, a waffle cone, or a dish, and may have 2 or 3 scoops (which must be the same flavor). Also, any order may come with a cherry or nuts (or neither), but not both. How many different orders are possible?
  - We tabulate all of the possible choices separately.
  - First, we choose an ice cream flavor: there are 25 options.
  - Then we choose a sugar cone, waffle cone, or dish: there are 3 options.
  - Next we choose the number of scoops: there are 2 options.
  - Finally, we choose either a cherry, nuts, or neither: there are 3 options.
  - By the multiplication principle, the total number of possible orders is $25 \cdot 3 \cdot 2 \cdot 3 = \boxed{450}$.

- Example: In the Unicode family of character encodings, each character is represented by a string of n bits, each of which is either a 0 or 1 (where n depends on the particular implementation). If it is necessary to be able to encode at least 150,000 different characters, what is the smallest possible value of n that will suffice?
  - If we have a string of n bits each of which is 0 or 1, then by the multiplication principle the total number of possible strings is 2^n.

- Thus, we want $2^n \geq 150000$. Taking logarithms, we need $n \geq \log_2(150000) \approx 17.194$, so the smallest integer value of n that will work is $n = \boxed{18}$.

- Example: Determine the number of subsets of the set $\{1, 2, \ldots, n\}$.
  - We may characterize a subset S of $\{1, 2, \ldots, n\}$ by listing, for each $k \in \{1, 2, \ldots, n\}$, whether $k \in S$ or $k \notin S$.
  - By the multiplication principle, the number of possible ways of making this sequence of n choices is $\boxed{2^n}$.

- In many counting problems, we must break into several cases and tabulate possibilities separately.

- Example: At a car dealership, Brand X sells 11 different models of cars each of which comes in 20 different colors, while Brand Y sells 6 different models of cars each of which comes in 5 different colors. How many different possible car options (including brand, model, and color) can be purchased at the dealership?
  - If a Brand X car is purchased, there are 11 choices for the model and 20 choices for the color, so by the multiplication principle there are $11 \cdot 20 = 220$ possible options in this case.
  - If a Brand Y car is purchased, there are 6 choices for the model and 5 choices for the color, so by the multiplication principle there are $6 \cdot 5 = 30$ possible options in this case.
  - Since these two cases are disjoint, in total there are $220 + 30 = \boxed{250}$ possible car options.

- In other cases, we may use "complementary counting": count possibilities and then subtract ones that are not allowed to occur, or that have been double-counted.

- Example: A local United States telephone number has 7 digits and cannot start with 0, 1, or the three digits 555. How many such telephone numbers are possible?
  - The first digit has 8 possibilities (namely, the digits 2 through 9 inclusive) and the other six digits each have 10 possibilities. Thus, by the multiplication principle, there are $8 \cdot 10^6 = 8\,000\,000$ total telephone numbers.
  - However, we have included the numbers starting with 555: each of these has 10 choices for each of the last 4 digits, for a total of $10^4 = 10\,000$ telephone numbers.
  - Subtracting the disallowed numbers yields a total of $8\,000\,000 - 10\,000 = \boxed{7\,990\,000}$ local telephone numbers.
  - Remark: Another method is to count all 10^7 possible 7-digit numbers, and then subtract the 10^6 starting with 0, the 10^6 starting with 1, and the 10^4 starting with 555.

1.2.2 Permutations and Combinations

- Certain problem types involving rearrangements of distinct objects ("permutations"), or ways to select subsets of a particular size ("combinations"), arise frequently in counting problems.

- Example: Determine the number of permutations (i.e., ways to rearrange) the six letters ABCDEF.
  - There are 6 letters to be arranged into 6 locations.
  - For the first letter, there are 6 choices (any of ABCDEF).
  - For the second letter, there are only 5 choices (any letter except the one we have already chosen).
  - For the third letter, there are only 4 choices (any letter except the first two).
  - Continuing in this way, we see that there are 3 choices for the fourth letter, 2 choices for the fifth letter, and only 1 choice for the last letter.
  - By the multiplication principle, the total number of permutations is therefore $6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = \boxed{720}$.

- Example: A new company logo has four design elements, which must all be different colors chosen from red, orange, yellow, green, blue, and purple. How many different logos are possible?

- There are 6 possible colors. The first design element has 6 possible colors, the second has 5 possible colors (any of the 6 except the one already used), the third has 4 possible colors, and the fourth has 3 possible colors.
- Thus, the total number of logos is $6 \cdot 5 \cdot 4 \cdot 3 = \boxed{360}$.

- Both of the problems above are example of computing permutations, where we choose k distinct items from a list of n possibilities, and where the order of our choices matters.
  - We can give a general formula for solving problems of this type in terms of factorials.

- Definition: If n is a positive integer, we define the number $n!$ (read "n factorial") as $n! = n \cdot (n-1) \cdot \cdots \cdot 2 \cdot 1$, the product of the positive integers from 1 to n inclusive. We also set $0! = 1$.
  - Some small values are $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$, $5! = 120$, and $6! = 720$.
  - The factorial function grows very fast: to 4 significant figures, we have $10! = 3.629 \cdot 10^6$, $100! = 9.333 \cdot 10^{157}$, and $1000! = 4.024 \cdot 10^{2567}$.
  - A useful approximation known as Stirling's formula says that $n! \approx n^n e^{-n}\sqrt{2\pi n}$ for large n (in the sense that the ratio between the two quantities approaches 1 as n grows). In particular, $n!$ grows faster than any exponential function of the form A^n for any positive A.

- Proposition (Permutations): The number of ways of choosing k ordered items from a list of n distinct possibilities (where the order of the k items matters) is equal to $\dfrac{n!}{(n-k)!} = n \cdot (n-1) \cdot \cdots \cdot (n-k+1)$. In particular, the number of ways of rearranging n distinct items is $n!$.
  - Proof: There are n possibilities for the first item, $n-1$ for the second item (any possibility but the one already chosen), $n-2$ for the third item (any possibility but the two already chosen), ... , and $n-k+1$ possibilities for the kth item.
  - This yields a total number of possibilities of $n \cdot (n-1) \cdot \cdots \cdot (n-k+1)$.
  - For the formula, we have $n \cdot (n-1) \cdot \cdots \cdot (n-k+1) \cdot (n-k)! = n \cdot (n-1) \cdot \cdots \cdot (n-k+1) \cdot (n-k) \cdot \cdots \cdot 1 = n!$.
  - Thus, $n \cdot (n-1) \cdot \cdots \cdot (n-k+1) = \dfrac{n!}{(n-k)!}$.

- Example: A sports league has 31 teams in total. How many ways are there to choose 16 teams that make the playoffs, assuming that the ranking of the playoff teams matters and there are no ties?
  - We are choosing $k = 16$ teams from a list of $n = 31$, where the order matters. From our result on permutations, the total number of choices is $\boxed{\dfrac{31!}{15!}} = 31 \cdot 30 \cdot \cdots \cdot 16$.

- In certain other types of counting problems, the order of the list of the k items we choose from the list of n does not matter. We can also give a formula for counting in this way:

- Proposition (Combinations): The number of ways of choosing k unordered items from a list of n distinct possibilities is equal to $\dbinom{n}{k} = {}_nC_k = \dfrac{n!}{k!(n-k)!} = \dfrac{n \cdot (n-1) \cdot \cdots \cdot (n-k+1)}{k \cdot (k-1) \cdot \cdots \cdot 1}$.
  - Remark: The symbols $\binom{n}{k}$ and ${}_nC_k$ are both typically read as "n choose k".
  - Proof: From our calculation above, we know that the number of ways to choose k ordered items from a list of n distinct possibilities is $\dfrac{n!}{(n-k)!}$.
  - If instead we want to count unordered lists, we can simply observe that for any unordered list, there are $k!$ ways to rearrange the k elements on the list.
  - Thus we have counted each unordered list $k!$ times, so the number of unordered lists is $\dfrac{1}{k!} \cdot \dfrac{n!}{(n-k)!} = \dfrac{n!}{k!(n-k)!}$.

- The numbers $\binom{n}{k}$ are called binomial coefficients because they arise as coefficients of binomial expansions.
  - Specifically, in the expansion of $(x+y)^n$, the coefficient of $x^k y^{n-k}$ is equal to $\binom{n}{k}$.
  - This follows by observing that in expanding the product $(x+y)\cdot(x+y)\cdot\dots\cdot(x+y)$, we may choose an x or a y from each of n terms. The term $x^k y^{n-k}$ will arise from products that choose exactly k terms equal to x: thus, from our discussion above, there are precisely $\binom{n}{k}$ such terms.
  - For example, we can compute that $(x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4$, and the middle term is indeed equal to $\binom{4}{2} = \dfrac{4!}{2!2!} = 6$.
  - Binomial coefficients show up in many different places, and satisfy many interesting algebraic identities, such as the "reflection identity" $\binom{n}{k} = \binom{n}{n-k}$ along with the recurrence $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$.
- In general, expanding the products of factorials is not the most efficient way to evaluate binomial coefficients.
  - Instead, the formula $\binom{n}{k} = \dfrac{n\cdot(n-1)\cdot\dots\cdot(n-k+1)}{k\cdot(k-1)\cdot\dots\cdot 1}$ is typically the most efficient.
  - For example, computing $\binom{13}{4}$ as $\dfrac{13!}{4!9!}$ requires computing both 13! and 4!9!, and then evaluating the quotient, which is rather painful to do by hand.
  - On the other hand, the formula above gives $\binom{13}{4} = \dfrac{13\cdot 12\cdot 11\cdot 10}{4\cdot 3\cdot 2\cdot 1} = 13\cdot 11\cdot 5 = 715$, which is easy to evaluate by hand.
- Example: How many 3-element subsets of $\{1,2,3,4,5,6,7,8,9\}$ are there?
  - Since subsets are not ordered, we are simply counting the number of ways to choose 3 unordered elements from the given set of 9.
  - From our discussion of combinations, the number of such subsets is $\binom{9}{3} = \dfrac{9\cdot 8\cdot 7}{3\cdot 2\cdot 1} = \boxed{84}$.
- Example: At a conference with 30 mathematicians, every pair of attendees shakes hands once. How many total handshakes occur?
  - Since pairs of people are not ordered, we are counting the number of ways to choose 2 attendees from a total of 30, which is $\binom{30}{2} = \dfrac{30\cdot 29}{2\cdot 1} = \boxed{435}$.
- Example: A pizza parlor offers 13 different possible toppings on a pizza. A pizza may have from 0 up to 3 different toppings. How many different pizza topping combinations are possible?
  - In general, there are $\binom{13}{k}$ possible pizzas that have exactly k toppings.
  - Thus, the number of pizzas with at most 3 toppings is $\binom{13}{0} + \binom{13}{1} + \binom{13}{2} + \binom{13}{3} = 1 + 13 + 78 + 286 = \boxed{378}$.
- Example: Determine the number of different full-house hands, consisting of 3 cards of one rank, and a pair of cards in another rank, that can be dealt from a standard 52-card deck.
  - Note that there are 13 possible card ranks, and 4 cards of each rank.
  - First, there are 13 ways to choose the rank of the 3-of-a-kind, and then there are 12 ways to choose the rank of the pair.
  - Once we have chosen the ranks, there are $\binom{4}{3} = 4$ ways to choose the three cards forming the 3-of-a-kind, and there are $\binom{4}{2} = 6$ ways to choose the two cards forming the pair.
  - Thus, in total there are $13\cdot 12\cdot 4\cdot 6 = \boxed{3744}$ possible full houses.
- Example: Determine the number of possible ways of permuting the letters in the word MISSISSIPPI.
  - Since there are 11 letters, it might seem as if there are 11! permutations of the letters.

- However, not all of these permutations yield different words: for example, if we swap two of the Ss, the resulting words are the same.
- There are 4 Ss, 4 Is, 2 Ps, and 1 M, which we will arrange in that order.
- First, we place the 4 Ss: since there are 11 possible locations, there are $\binom{11}{4}$ ways to place them (since the 4 Ss are identical).
- Next we place the 4 Is: there are 7 remaining locations, so there are $\binom{7}{4}$ ways to place them.
- After this, there are 3 remaining locations in which we may place the 2 Ps, yielding $\binom{3}{2}$ choices. Finally, there is only 1 location for the M.
- In total, there are $\binom{11}{4} \cdot \binom{7}{4} \cdot \binom{3}{2} = 330 \cdot 35 \cdot 3 = \boxed{34650}$ ways of permuting the letters.
- Remark: Another way to perform the count is to determine the number of times each word shows up in the 11! permutations of the letters. Since there are 4! ways of permuting the 4 Ss among themselves, 4! ways of permuting the 4 Is, and 2! ways of permuting the 2 Ps, each word shows up $4! \cdot 4! \cdot 2!$ times. Thus, the number of different words is $\dfrac{11!}{4!4!2!} = 34650$.

- Example: Determine the number of possible ways of permuting the letters in the word BOSTONIANS that contain the word BOOS.
  - The number of such permutations is the number of permutations of the six letters T, N, I, A, N, S and the string BOOS (which we can think of as being a single string).
  - There are 2 Ns, and 1 each of T, I, A, S, and BOOS to arrange.
  - First, we place the 2 Ns: since there are 7 possible locations, there are $\binom{7}{2}$ ways to place them. The remaining 5 strings can be permuted arbitrarily, so there are 5! ways to arrange them.
  - In total, there are $\binom{7}{2} \cdot 5! = 42 \cdot 120 = \boxed{2520}$ ways of permuting the letters.
  - Remark: As above, another way to perform the count is by observing that there are 7! ways to arrange the 7 given strings, but each arrangement is counted twice because of the two Ns, so there are only $7!/2 = 2520$ different arrangements.

1.3 Probability and Probability Distributions

- Now that we have discussed the basics of sets and counting, we can develop discrete probability. We begin by discussing sample spaces and methods for computing probabilities of events.

1.3.1 Sample Spaces and Events

- The fundamental idea of "probability" arises from performing an experiment or observation, and tabulating how often particular outcomes arise.
  - For any experiment or observation, the set of possible outcomes is called the sample space, and an event is a subset of the sample space.
- Example: Consider the experiment of "rolling a standard 6-sided die once".
  - There are 6 possible outcomes to this experiment, namely, rolling a 1, a 2, a 3, a 4, a 5, or a 6, so the sample space is the set $S = \{1, 2, 3, 4, 5, 6\}$.
  - One event is "rolling a 3", which would correspond to the subset $\{3\}$.
  - Another event is "rolling an even number", which would correspond to the subset $\{2, 4, 6\}$.
  - A third event is "rolling a number bigger than 2", which would correspond to the subset $\{3, 4, 5, 6\}$.
  - A fourth event is "rolling a negative number", which would correspond to the empty subset $\emptyset = \{\}$ because there are no outcomes in the sample space that make this event occur.
- Example: Consider the experiment of "flipping a coin once".

- There are 2 possible outcomes to this experiment: heads and tails. Thus, the sample space is the set $S = \{\text{heads, tails}\}$, which we typically abbreviate as $S = \{H, T\}$.
- One event is "obtaining heads", corresponding to the subset $\{H\}$.
- Another event is "obtaining tails", corresponding to the subset $\{T\}$.

- Example: Consider the experiment of "flipping a coin four times".
  - By the multiplication principle, there are $2^4 = 16$ possible outcomes to this experiment (namely, the 16 possible strings of 4 characters each of which is either H or T).
  - One event is "exactly one head is obtained", corresponding to the subset $\{HTTT, THTT, TTHT, TTTH\}$.
  - Another event is "the first three flips are tails", corresponding to the subset $\{TTTH, TTTT\}$.

- Example: Consider the experiment of "shuffling a standard 52-card deck".
  - There are many possible outcomes to this experiment, namely, all 52! ways of arranging the 52 cards in sequence. The sample space S is the (quite large!) set of all 52! of these sequences.
  - One event is "the first card is an ace of clubs", while another event is "the entire deck alternates red-black-red-black". It would be infeasible to write out the exact subsets corresponding to these events, but in principle it would be possible to list them all.

- Example: Consider the experiment of "measuring the lifetime of a refrigerator in years".
  - There are many possible outcomes to this experiment, including 0, 5, 28, 3.2, and 100.
  - The sample space would (at least in principle) be the set of nonnegative real numbers: $S = [0, \infty)$.
  - One event is "the refrigerator stops working after at most 3 years", corresponding to the subset $S = [0, 3]$.
  - Another event is "the refrigerator works for at least 6 years", corresponding to the subset $S = [6, \infty)$.

- Example: Consider the experiment of "measuring the temperature in degrees Fahrenheit outside".
  - There are many possible outcomes to this experiment, including 50°, 87.4°, and 120°.
  - For this experiment, depending on how accurately the temperature is measured, and what the possible outside temperatures are, the sample space could be very large or even infinite.
  - If the temperature is measured to the nearest whole degree, and it is known that the temperature is never colder than 0° nor hotter than 120°, then the sample space would be $S = \{0^\circ, 1^\circ, 2^\circ, \dots, 120^\circ\}$.
  - One event is "the temperature is above freezing", corresponding to the subset $\{33^\circ, 34^\circ, \dots, 120^\circ\}$.
  - Another event is "the temperature is a multiple of 10°", corresponding to the subset $\{0^\circ, 10^\circ, 20^\circ, \dots, 120^\circ\}$.

- Since we view events as subsets of the sample space, we define the union (or intersection) of two events as the union (respectively, the intersection) of their corresponding subsets, and we define the complement of an event to be the complement of the corresponding subset inside the sample space.
  - For example, with sample space $S = \{1, 2, 3, 4, 5, 6\}$ obtained by rolling a standard 6-sided die once, and take $A = \{2, 4, 6\}$ to be the event of rolling an even number and $B = \{3, 4, 5, 6\}$ to be the event of rolling a number larger than 2.
  - Then the complement $A^c = \{1, 3, 5\}$ is the event of not rolling an even number (i.e., rolling an odd number), and the complement $B^c = \{1, 2\}$ is the event of not rolling a number larger than 2 (i.e., rolling a number less than or equal to 2).
  - Also, the union $A \cup B = \{2, 3, 4, 5, 6\}$ is the event of rolling a number that is even or larger than 2, while the intersection $A \cap B = \{4, 6\}$ is the event of rolling a number that is even and larger than 2.

- In the case where the events A and B have $A \cap B = \emptyset$, we say that A and B are mutually exclusive, since they cannot both occur at the same time.

1.3.2 Probabilities of Events

- We would now like to analyze probabilities of events, which measure how likely particular events are to occur.
  - If we have an experiment with corresponding sample space S, and E is an event (which we consider as being a subset of S), we would like to define the probability of E, written $P(E)$, to be the frequency with which E occurs if we repeat the experiment many times independently.
  - Specifically, if we repeat the experiment n times and the event occurs e_n times, then the relative frequency that E occurs is the ratio e_n/n. If we let n grow very large (more formally, if we take the limit as $n \to \infty$) then the ratios e_n/n should approach a fixed number, which we call the probability of the event E.
  - Since for each n we have $0 \le e_n \le n$, and thus $0 \le e_n/n \le 1$, we see that the limit of the ratios e_n/n must be in the closed interval $[0, 1]$. This tells us that $P(E)$ should always lie in this interval.
  - For the event S (consisting of the entire sample space), we clearly have $P(S) = 1$, because if we perform the experiment n times, the event S always occurs n times, so $e_n = n$ for every n.
  - Also, if E_1 and E_2 are mutually exclusive events, then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$: this follows by observing that because both events cannot occur simultaneously, then the total number of times E_1 or E_2 occurs in n experiments is equal to the total number of times E_1 occurs plus the total number of times E_2 occurs.
  - In particular, if $S = \{s_1, s_2, \dots, s_k\}$ is a finite sample space, then by applying the above observation repeatedly, we can see that $P(\{s_1\}) + P(\{s_2\}) + \cdots + P(\{s_k\}) = P(\{s_1, s_2, \dots, s_k\}) = P(S) = 1$.
  - This means that the sum of all the probabilities of the events in the sample space is equal to 1.
- We now give a more formal definition of a probability distribution on a sample space, using the properties we have worked out above as motivation:
- Definition: If S is a sample space, a probability distribution on S is any function P defined on events (i.e., subsets of S) with the following properties:

  [P1] For any event E, the probability $P(E)$ satisfies $0 \le P(E) \le 1$.

  [P2] The probability $P(S) = 1$.

  [P3] If $E_1, E_2, \dots, E_k$ are mutually exclusive events[4] (meaning that $E_i \cap E_j = \emptyset$ for $i \ne j$), then $P(E_1 \cup E_2 \cup \cdots \cup E_k) = P(E_1) + P(E_2) + \cdots + P(E_k)$.

  - We may interpret the probability $P(E)$ as the relative frequency that the event E occurs, if we repeat the experiment a large number of times.
- There are a few fundamental properties of probabilities that can be derived from these basic assumptions [P1]-[P3]:
- Proposition (Basic Properties of Probability): For any events E, E_1, E_2 inside a sample space S with probability distribution P, the following properties hold:
  1. $P(\emptyset) = 0$.
  2. $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.
  3. $P(E^c) = 1 - P(E)$.

  - Proof: For (1), apply [P3] to $E_1 = E_2 = \emptyset$, so that $E_1 \cup E_2 = \emptyset$ also, to see that $P(\emptyset) = P(\emptyset) + P(\emptyset)$, so $P(\emptyset) = 0$.
  - For (2), observe (e.g., via a Venn diagram) that if we define the events $A = E_1 \cap E_2^c$, $B = E_1 \cap E_2$, and $C = E_1^c \cap E_2$, then A, B, C are mutually disjoint with $A \cup B = E_1$, $B \cup C = E_2$, and $A \cup B \cup C = E_1 \cup E_2$.
  - Thus by [P3] applied repeatedly, we have $P(E_1 \cup E_2) = P(A \cup B \cup C) = P(A) + P(B) + P(C)$, while $P(E_1) = P(A) + P(B)$, $P(E_2) = P(B) + P(C)$, and $P(E_1 \cap E_2) = P(B)$.

[4] If S is an infinite sample space, this property also applies to infinite collections of mutually exclusive events: explicitly, we require $P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$.

- Hence we see $P(E_1 \cup E_2) = P(A) + P(B) + P(C) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$, as claimed.
- For (3), apply the result of (2) with $E_1 = E$ and $E_2 = E^c$: since $E \cup E^c = S$ and $E \cap E^c = \emptyset$ the statement of (1) reduces to $1 = P(S) = P(E) + P(E^c) - P(\emptyset)$, and thus by (1) and [P2] we see that $P(E^c) = 1 - P(E)$ as claimed.

- In the particular case where $S = \{s_1, s_2, \dots, s_k\}$ is a finite sample space, we can describe what a probability distribution is more explicitly:
  - Explicitly, by properties [P2] and [P3], we see that $P(\{s_1\})+P(\{s_2\})+\cdots+P(\{s_k\}) = P(\{s_1, s_2, \dots, s_k\}) = P(S) = 1$.
  - Thus, by [P1] the numbers $P(\{s_1\})$, ... , $P(\{s_k\})$ are all between 0 and 1 inclusive, and have sum 1.
  - If we make any selection for these values satisfying these conditions, then we may compute the probability of an arbitrary event $E = \{t_1, \dots, t_d\}$ by using property [P3] again: $P(E) = P(\{t_1\}) + \cdots + P(\{t_d\})$.
  - It is not hard to check that this assignment of $P(E)$ for each subset E of S satisfies all three of [P1]-[P3].
  - Therefore, a probability distribution on $S = \{s_1, s_2, \dots, s_k\}$ is simply an assignment of probabilities between 0 and 1 inclusive to each of the individual outcomes in the sample space, such that the sum of all the probabilities is 1.

- Example: Consider the sample space $S = \{H, T\}$ corresponding to flipping a coin.
  - By our discussion above, a probability distribution on S is determined by assigning values to $P(H)$ and $P(T)$ such that $0 \le P(H), P(T) \le 1$ and $P(H) + P(T) = 1$.
  - One probability distribution arises from assuming that a head is equally likely to appear as a tail: then we would have $P(H) = P(T) = 1/2$.
  - A different probability distribution arises from assuming that a tail is twice as likely to appear as a head: then we would have $P(H) = 1/3$ while $P(T) = 2/3$.
  - Another probability distribution, for an even more unfair coin, would have $P(H) = 0.1$ with $P(T) = 0.9$: this represents a coin that lands heads 10% of the time and tails the other 90% of the time.
  - We even have a probability distribution for a coin that never comes up heads; namely, the one with $P(H) = 0$ and $P(T) = 1$.

- In the example above, we can see that the choice of probability distribution for this sample space affects our interpretation of how fair the coin is.
  - We view the probability of an event on a sliding scale from 0 to 1: a probability near 0 means that the event happens rarely, while a probability near 1 means the event happens often.
  - If the sample space is finite, then an event with probability 0 never occurs, while an event with probability 1 always occurs.
  - However, if the sample space is infinite, there are situations where events of probability 0 can (perhaps surprisingly) occur. One example of this is the experiment of choosing a real number uniformly at random from the interval $[0, 1]$. The assumption that we choose "uniformly at random" means that every real number has the same probability of being chosen, but since there are infinitely many numbers in the interval, the probability of choosing any particular one is equal to 0. Nonetheless, each time this experiment is performed, there is some real number chosen as the outcome, even though that particular outcome has probability 0 of occurring!

1.3.3 Computing Probabilities

- So far, our discussion of probability has been relatively abstract, since we have primarily spoken about probability distributions.
  - In order to compute probabilities of particular events, we must make assumptions about the corresponding probability distributions, such as assuming that a coin is fair (i.e., that heads and tails are equally likely).

- In the particular case where all of the outcomes $s_1, s_2, \ldots, s_k$ in the sample space are equally likely, we would have $P(s_1) = P(s_2) = \cdots = P(s_k) = \dfrac{1}{k}$.
- Then the probability of any event E is then simply $\#(E)/k$, which only depends on the number of outcomes in E. Thus, we may find $P(E)$ simply by counting all of the outcomes in E; in particular, when the sample space is small, we can simply list them all.

- Example: If a fair coin is flipped 3 times, determine the respective probabilities of obtaining (i) no heads, (ii) 1 head, (iii) 2 heads, and (iv) 3 heads.
  - The sample space has $2^3 = 8$ outcomes: explicitly, $S = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$.
  - Under the assumption that the coin is fair, each of the 8 outcomes is equally likely, so by our analysis above, the probability of each of the 8 outcomes is 1/8.
  - Now we simply count to determine each of the respective probabilities.
  - There is only 1 outcome with no heads (namely, TTT), so the probability of obtaining no heads is $\boxed{1/8}$.
  - There are 3 outcomes (namely TTH, THT, and HTT) with 1 head, so the probability of obtaining 1 head is $\boxed{3/8}$.
  - There are 3 outcomes (namely THH, HTH, and HHT) with 2 heads, so the probability of obtaining 2 heads is $\boxed{3/8}$.
  - There is 1 outcome (namely, HHH) with 3 heads, so the probability of obtaining 3 heads is $\boxed{1/8}$.

- Example: If two fair 6-sided dice are rolled, determine the probabilities of the respective events (i) the two dice read 6 and 3 in some order, (ii) the sum of the two rolls is equal to 4, (iii) both rolls are equal, (iv) neither roll is a 2, and (v) at least one roll is a 2.
  - In this case, the sample space consists of $6^2 = 36$ outcomes representing the 36 ordered pairs (R_1, R_2) where R_1 is the outcome of the first die roll and R_2 is the outcome of the second die roll.
  - Under the assumption that both dice are fair, all 36 outcomes are equally likely, so the probability of any one of these individual outcomes is 1/36.
  - Now we simply count to determine each of the respective probabilities.
  - For event (i), there are two possible outcomes, namely $(3, 6)$ and $(6, 3)$. Thus the probability of this event is $2/36 = \boxed{1/18}$.
  - For event (ii), there are three possible outcomes, namely $(1, 3)$, $(2, 2)$, and $(3, 1)$. Thus the probability of this event is $3/36 = \boxed{1/12}$.
  - For event (iii), there are six possible outcomes, namely $(1, 1)$, $(2, 2)$, $(3, 3)$, $(4, 4)$, $(5, 5)$, and $(6, 6)$. Thus the probability of this event is $6/36 = \boxed{1/6}$.
  - For event (iv), there are 25 possible outcomes, consisting of the $5 \cdot 5$ ordered pairs (a, b) where each of a and b is one of the 5 numbers $1, 3, 4, 5, 6$. Thus the probability of this event is $\boxed{25/36}$.
  - For event (v), one way to compute the probability is simply to list all 11 possible outcomes, namely $(2, 1)$, $(2, 2)$, $(2, 3)$, $(2, 4)$, $(2, 5)$, $(2, 6)$, $(1, 2)$, $(3, 2)$, $(4, 2)$, $(5, 2)$, and $(6, 2)$, to see that the probability of this event is $\boxed{11/36}$.
  - Another way to find the probability of event (v) is to observe that it is the complement of event (iv), and thus its probability must be $1 - 25/36 = \boxed{11/36}$.

- When the sample space is large, it can be unreasonable to enumerate all the outcomes in particular events explicitly. Nonetheless, by counting appropriately, we may still compute probabilities:

- Example: Three cards are randomly dealt from a standard 52-card deck. Determine the probabilities of the respective events (i) all three cards are aces, (ii) all three cards are the same suit, (iii) one of the cards is the 3 of diamonds, and (iv) at least one card is a jack.

- Let us choose the sample space S to be the set of possible triples (C_1, C_2, C_3) of the three cards dealt in order from the deck. By the multiplication principle, there are $52 \cdot 51 \cdot 50$ such triples, so $\#S = 52 \cdot 51 \cdot 50$.
- As with the previous examples, under the assumption that the cards are drawn randomly, each of the outcomes in S is equally likely.
- For event (i), there are 4 choices for the first card (any of the 4 aces), 3 choices for the second card (any of the 3 remaining aces), and 2 choices for the third card. Thus there are $4 \cdot 3 \cdot 2$ outcomes that yield this event, and so the probability that this event occurs is $\dfrac{4 \cdot 3 \cdot 2}{52 \cdot 51 \cdot 50} = \boxed{\dfrac{1}{5525}}$.
- For event (ii), there are 4 choices for the common suit. Once we have chosen the suit, there are 13 choices for the first card, 12 for the second, and 11 for the third, yielding a total number of $4 \cdot 13 \cdot 12 \cdot 11$ outcomes. Thus the desired probability is $\dfrac{4 \cdot 13 \cdot 12 \cdot 11}{52 \cdot 51 \cdot 50} = \boxed{\dfrac{22}{425}} \approx 0.0518$.
- For event (iii), there are 3 choices for which card is the 3 of diamonds. Once this choice is made, there are 51 possibilities for the first remaining card (any card except the 3 of diamonds) and 50 possibilities for the second remaining card (any card except the 3 of diamonds or the card just chosen), yielding a total number of $3 \cdot 51 \cdot 50$ outcomes. Thus the desired probability is $\dfrac{3 \cdot 51 \cdot 50}{52 \cdot 51 \cdot 50} = \boxed{\dfrac{3}{52}} \approx 0.0577$. Intuitively, since we are choosing 3 cards out of 52, and the 3 of diamonds is equally likely to be any one of these 52 cards, it is natural to say that the probability that it is among the 3 cards we have chosen should be $3/52$.
- For event (iv), we first find the probability that none of the cards is a jack. If none of the cards is a jack, then there are $48 \cdot 47 \cdot 46$ possible ways to choose the cards (the first card can be any card except the 4 jacks, and so forth), and so the probability that none of the cards is a jack is $\dfrac{48 \cdot 47 \cdot 46}{52 \cdot 51 \cdot 50} = \dfrac{4324}{5525}$. Then the probability that at least one card is a jack is the probability of the complementary event, and is therefore equal to $1 - \dfrac{4324}{5525} = \boxed{\dfrac{1201}{5525}} \approx 0.2174$.

- <u>Example</u>: A fair coin is flipped 15 times. Determine the probabilities of the respective events (i) exactly 7 heads are obtained, (ii) at most 2 tails are obtained, (iii) at least 3 tails are obtained, and (iv) the number of heads is a multiple of 6.
  - Here, the sample space is the set of $2^{15} = 32768$ possible strings of 15 heads or tails, where (as above) each of the possible outcomes is equally likely.
  - From our results on combinations, there are exactly $\binom{15}{n}$ possible strings that have n heads and $15 - n$ tails, and an equal number that have $15 - n$ heads and n tails. We therefore only need to tally the various possibilities in each case.
  - For event (i), there are $\binom{15}{7} = \dfrac{15 \cdot 14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9}{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 6435$ possible outcomes, and thus the probability of this event is $\boxed{\dfrac{6435}{32768}} \approx 0.1964$.
  - For event (ii), a string with at most 2 tails can have 2 tails, 1 tail, or 0 tails, so there are $\binom{15}{2} + \binom{15}{1} + \binom{15}{0} = \dfrac{15 \cdot 14}{2} + 15 + 1 = 121$ possible outcomes. Thus the probability of this event is $\boxed{\dfrac{121}{32768}} \approx 0.0037$.
  - For event (iii), observe that this event is the complement of event (ii), and therefore its probability is $1 - \dfrac{121}{32768} = \boxed{\dfrac{32647}{32768}} \approx 0.9963$. This approach is simpler than the more direct method of observing that this event occurs if there are 3, 4, 5, ... , or 15 tails, and so the total number of possibilities is $\binom{15}{3} + \binom{15}{4} + \cdots + \binom{15}{15} = 32647$.
  - For event (iv), such a string can have 0 heads, 6 heads, or 12 heads, so there are $\binom{15}{0} + \binom{15}{6} + \binom{15}{12} = 1 + 5005 + 455 = 5461$ possible outcomes. Thus the probability of this event is $\boxed{\dfrac{5461}{32768}} \approx 0.1667$.

- Example: A random five-card poker hand is dealt from a standard 52-card deck. Determine the probabilities of the respective events (i) the hand is a straight flush, (ii) the hand is a four-of-a-kind, (iii) the hand is a full house (three cards of one rank and a pair in another), (iv) the hand is a flush (all cards are the same suit) that is not a straight flush, (v) the hand is a straight (five cards in sequence) that is not a straight flush, (vi) the hand is a three-of-a-kind, (vii) the hand is two pair (two pairs of different ranks), (viii) the hand is one pair, and (ix) the hand is none of the above.
    - First observe that there are $\binom{52}{5} = 2598960$ ways to deal a five-card hand, where the order of the cards does not matter. In some cases, it is slightly easier to count hands where we do pay attention the ordering, in which case there are $52 \cdot 51 \cdot 50 \cdot 49 \cdot 48$ total hands instead.
    - For (i), there are 4 possible suits for a straight flush and 10 possible card sequences (A-2-3-4-5 through 10-J-Q-K-A) yielding a straight flush. Thus, the probability of obtaining a straight flush is $\frac{4 \cdot 10}{2598960} = \boxed{\frac{1}{64974}} \approx 0.0015\%$.
    - For (ii), there are 13 possible ranks for the 4-of-a-kind and 48 possibilities for the remaining single card. Thus, the probability of obtaining a 4-of-a-kind is $\frac{13 \cdot 48}{2598960} = \boxed{\frac{1}{4165}} \approx 0.0240\%$.
    - For (iii), there are 13 possible ranks for the 3-of-a-kind and $\binom{4}{3} = 4$ ways to choose the corresponding cards, and also 12 possible remaining ranks for the pair and $\binom{4}{2} = 6$ ways to choose those cards. Thus, the probability of obtaining a full house is $\frac{13 \cdot 4 \cdot 12 \cdot 6}{2598960} = \boxed{\frac{6}{4165}} \approx 0.1981\%$.
    - For (iv), there are 4 possible choices for the suit and then $\binom{13}{5} = 1287$ choices for the 5 cards in that suit. Of these, 10 yield straight flushes by the analysis in (i), so the remaining 1277 do not. Therefore, the probability of obtaining a flush that is not a straight flush is $\frac{4 \cdot 1277}{2598960} = \boxed{\frac{1277}{649740}} \approx 0.1965\%$.
    - For (v), there are 10 possible card sequences yielding a straight, and each of the 5 cards can be any of the four suits, yielding a total of $10 \cdot 4^5 = 10240$ straights. Of these, 40 are straight flushes, and the remaining 10200 are not. Therefore, the probability of obtaining a straight that is not a straight flush is $\frac{10200}{2598960} = \boxed{\frac{5}{1274}} \approx 0.3925\%$.
    - For (vi), we use ordered hands. There are $\binom{5}{3} = 10$ ways to choose the locations of the 3-of-a-kind cards, and there are 13 possible ranks for the 3-of-a-kind with $4 \cdot 3 \cdot 2 = 24$ ways to choose the corresponding cards. Once these choices are made, there are 48 possibilities for the first remaining card and 44 for the second remaining card (since it must be a different rank). Therefore, the probability of obtaining a three-of-a-kind is $\frac{10 \cdot 13 \cdot 24 \cdot 48 \cdot 44}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48} = \boxed{\frac{88}{4165}} \approx 2.1128\%$.
    - For (vii), there are $\binom{13}{2} = 78$ possible ranks for the two pairs and $\binom{4}{2} \cdot \binom{4}{2} = 36$ ways to choose the cards in each pair. Once these are chosen, there are 44 possibilities for the last card (since it must be a different rank than the pairs). Therefore, the probability of obtaining two pair is $\frac{78 \cdot 36 \cdot 44}{2598960} = \boxed{\frac{198}{4165}} \approx 4.7539\%$.
    - For (viii), we use ordered hands. There are $\binom{5}{2} = 10$ ways to choose the locations of the pair cards, 13 possible ranks for the pair, and $4 \cdot 3 = 12$ ways to choose the corresponding cards. Once these choices are made, there are $48 \cdot 44 \cdot 40$ possibilities for the remaining three cards (since they must all be different ranks). Therefore, the probability of obtaining one pair is $\frac{10 \cdot 13 \cdot 12 \cdot 48 \cdot 44 \cdot 40}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48} = \boxed{\frac{352}{833}} \approx 42.2569\%$.
    - For (ix), since we have chosen possibilities above that are all mutually exclusive, the probability of obtaining none of these events is simply 1 minus the sum of all of them, which ends up simplifying to $\boxed{\frac{1277}{2548}} \approx 50.1177\%$.

- Our general definition of probability distribution also allows us to analyze situations where the outcomes in the sample space are not equally likely.
  - However, this situation carries the added difficulty that we must now also assign a probability to each individual outcome in the events we are studying.
  - We will be able to do this more easily once we discuss conditional probabilities.

1.4 Conditional Probability and Independence

- We now discuss conditional probability, which provides us a way to compute the probability that one event occurs, given that another event also occurred.

1.4.1 Definition of Conditional Probability

- As a motivating example, we calculated earlier that if two fair 6-sided dice are rolled, then the probability that neither roll is a 2 is equal to 25/36. Now suppose we are given the additional information that the first die was a 5, and we ask again for the probability that neither roll was a 2.
  - There are now only six possible outcomes of this experiment that are consistent with the given information, namely, $(5,1)$, $(5,2)$, $(5,3)$, $(5,4)$, $(5,5)$, and $(5,6)$.
  - Since each of these outcomes was equally likely to occur originally, it is reasonable to say that they should still be equally likely. Our desired event (of not obtaining a 2) occurs in 5 of the 6 cases, so the probability of the event is 5/6.
  - Notice that by providing additional information (namely, that the first roll was a 5), the probability that neither roll was a 2 changes: this is the essential idea of conditional probability.
- In some cases, we can compute conditional probabilities using counting arguments like in our earlier examples.
- Example: Suppose that 100 students in a course have grades and academic standing as given in the table below.

Category	A	B	C	Total
Sophomore	4	8	6	18
Junior	14	11	10	35
Senior	38	4	5	47
Total	56	23	21	100

  Compute the probabilities that (i) a randomly-chosen student is getting an A, (ii) a randomly-chosen student is a junior, (iii) a randomly-chosen junior is getting a A, and (iv) a randomly-chosen A student is a junior.
  - For event (i), there are 56 A-students out of 100 total students, so the probability is $\boxed{56/100} = 0.56$.
  - For event (ii), there are 35 juniors out of 100 total students, so the probability is $\boxed{35/100} = 0.35$.
  - For event (iii), there are 14 A-student juniors out of 35 total juniors, so the probability is $\boxed{14/35} = 0.4$.
  - For event (iv), there are 14 A-student juniors out of 56 total A-students, so the probability is $\boxed{14/56} = 0.25$.
  - Event (iii) is an example of a conditional probability, namely, the probability that a student is getting an A given that the student is a junior. In order to compute this conditional probability, notice that we restrict our attention only to the set of juniors, and perform our computations as if this is our entire sample space.
  - Likewise, event (iv) is also a conditional probability, namely, the probability that a student is a junior, given that the student is getting an A, and like with event (iii) we restrict our attention only to the set of A-students and perform our computations as if this is our entire sample space.
  - More abstractly, let S be the full sample space of all 100 students, and let J represent the event that a student is a junior and let A represent the event that a student receives an A.

- If we write $P(A|J)$ for the probability that a student is receiving an A given that they are a junior, which is event (iii), then as we calculated above, $P(A|J)$ is equal to the ratio of juniors receiving an A to the total number of juniors, or, symbolically, $P(A|J) = \dfrac{\#(J \cap A)}{\#(J)}$.
- Now observe that $P(A|J) = \dfrac{\#(J \cap A)}{\#(J)} = \dfrac{\#(J \cap A)/\#(S)}{\#(J)/\#(S)} = \dfrac{P(J \cap A)}{P(J)}$, which gives us a way to write $P(A|J)$ in terms of probabilities of events in the sample space (specifically, the events $J \cap A$ and J).

- The analysis in the example above motivates a way to define conditional probabilities in general:
- Definition: If A and B are events and $P(B) > 0$, we define the conditional probability $P(A|B)$ that A occurs given that B occurred by $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$.
  - Remark: If $P(B) = 0$, then the conditional probability $P(A|B)$ cannot be computed using this definition. If the sample space is finite, then the conditional probability $P(A|B)$ does not make sense if $P(B) = 0$, because the event B can never occur.
- Example: Suppose two fair 6-sided dice are rolled. If A is the event "neither roll is a 2" and B is the event "the first roll is a 5", find $P(A|B)$ and $P(B|A)$ and describe what these two probabilities mean.
  - By definition, $P(A|B)$ is the probability that A occurs given that B occurred, so in this case it is the probability that neither roll is a 2, given that the first roll is a 5.
  - Inversely, $P(B|A)$ is the probability that B occurs given that A occurred, so in this case it is the probability that the first roll is a 5 given that neither roll is a 2.
  - According to the definition, we have $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ and $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$, so we need only calculate $P(A)$, $P(B)$, and $P(A \cap B)$.
  - As we have already calculated by enumerating all of the possible outcomes, $P(A) = 25/36$ (there are $5 \cdot 5$ possible outcomes where neither roll is a 2), $P(B) = 1/6$ (there are 6 outcomes where the first roll is a 5), and $P(A \cap B) = 5/36$ (there are 5 outcomes where the first roll is a 5 and neither roll is a 2).
  - Hence the formulas give $P(A|B) = \dfrac{5/36}{1/6} = \boxed{\dfrac{5}{6}}$, which agrees with the calculation we made earlier, and also $P(B|A) = \dfrac{5/36}{25/36} = \boxed{\dfrac{1}{5}}$.
- Example: Suppose four fair coins are flipped. Determine the probabilities of the respective events (i) there are exactly two heads, given that the first flip is a tail, (ii) the first flip is a tail, given that there are exactly two heads, and (iii) all four flips are tails, given that there is at least one tail.
  - For event (i), if we let A be the event that there are exactly two heads and B be the event that the first flip is a tail, then we wish to find $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$.
  - We can see that $A \cap B$ is the event that the first flip is a tail and there are exactly two heads, which occurs in three ways: $\{THHT, THTH, TTHH\}$, so $P(A \cap B) = \dfrac{3}{16}$.
  - Furthermore, we can see that there are 8 ways in which the first flip is a tail (since the remaining 3 flips can be either heads or tails) so $P(B) = \dfrac{1}{2}$.
  - Thus, we see $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{3/16}{1/2} = \boxed{\dfrac{3}{8}}$.
  - For event (ii), using the same events as above, we now wish to find $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$.
  - We can see that there are $\binom{4}{2} = 6$ ways in which we can obtain exactly two heads, so $P(A) = \dfrac{6}{16} = \dfrac{3}{8}$.

    - Thus, since $P(A \cap B) = \frac{3}{16}$ as computed above, we have $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{3/16}{3/8} = \boxed{\frac{1}{2}}$.
    - For event (iii), if we let C be the event that all four flips are tails and D be the event that there is at least one tail, then we wish to find $P(C|D) = \frac{P(C \cap D)}{P(D)}$.
    - It is easy to see that $C \cap D$ is simply the event C (since if all four flips are tails, then there is certainly at least one tail), so $P(C \cap D) = P(C) = \frac{1}{16}$.
    - Also, D will occur in every outcome except the one where all four flips are heads (which occurs with probability 1/16), so $P(D) = 1 - \frac{1}{16} = \frac{15}{16}$.
    - Thus, $P(C|D) = \frac{P(C \cap D)}{P(D)} = \frac{1/16}{15/16} = \boxed{\frac{1}{15}}$.

1.4.2 Independence

- It is natural to say that two events are independent of one another if the knowledge that one occurs does not give any additional information about whether the other occurs. We can easily phrase this in terms of conditional probability:
- Definition: We say that two events A and B are independent if $P(A|B) = P(A)$, or equivalently, if $P(B|A) = P(B)$. Two events that are not independent are said to be dependent.
    - From the definition of the conditional probability we know that $P(A|B) = \frac{P(A \cap B)}{P(B)}$, so if we rearrange the definition above, we obtain an equivalent definition of independence, namely, that A and B are independent if $P(A \cap B) = P(A) \cdot P(B)$.
    - This observation also allows us to see that the two statements of independence are equivalent to each other.
- Example: Suppose 2 fair 6-sided dice are rolled. Let A be the event "the first roll is a 3", let B be the event "the second roll is a 6", and let C be the event "the sum of the two rolls is 6". Determine whether each of the pairs of events is independent.
    - Of the 36 possible outcomes of rolling the two dice, 6 have the first roll equal to 3, 6 have the second roll equal to 6, and 5 have the sum of the two rolls equal to 6. Thus, $P(A) = P(B) = 1/6$ and $P(C) = 5/36$.
    - First, $A \cap B$ is the event that the first roll is a 3 and the second roll is a 6, which can happen in only 1 way, so $P(A \cap B) = 1/36$. Since $P(A \cap B) = 1/36 = (1/6) \cdot (1/6) = P(A) \cdot P(B)$, we see that $\boxed{A \text{ and } B \text{ are independent}}$.
    - Second, $A \cap C$ is the event that the first roll is a 3 and the sum of the two rolls is 6, which can happen in only 1 way (namely, both rolls are 3s), so $P(A \cap C) = 1/36$. Since $P(A) \cdot P(C) = 5/216 \neq 1/36$, the events $\boxed{A \text{ and } C \text{ are not independent}}$.
    - Third, $B \cap C$ is the event that the first roll is a 6 and the sum of the two rolls is 6, which can never occur (since the second die roll cannot be 0), so $P(B \cap C) = 0$. Since $P(B) \cdot P(C) = 5/216 \neq 0$, the events $\boxed{B \text{ and } C \text{ are not independent}}$.
    - Intuitively, we should expect that A and B are independent because the two rolls of the die do not affect one another. On the other hand, C is not independent from A and B because knowing that one roll is a 3 means it is possible for the sum of the dice to be 6, while knowing that one roll is a 6 means that a sum of 6 is impossible.
- Example: A single card is randomly dealt from a standard 52-card deck. Let A be the event "the card is a spade", let B be the event "the card is an ace", and let C be the event "the card is an ace, 2, 3, or 4". Determine whether each of the pairs of events is independent.

- Since there are 13 spades, 4 aces, and 16 cards total that are a 2, 3, 4, or 5, we see $P(A) = 1/4$, $P(B) = 1/13$, and $P(C) = 4/13$.
- First, $A \cap B$ is the event that the card is the ace of spades, so $P(A \cap B) = 1/52 = P(A) \cdot P(B)$, meaning that $\boxed{A \text{ and } B \text{ are independent}}$.
- Second, $A \cap C$ is the event that the card is a spade and also an ace, 2, 3, or 4, so it must be the ace, 2, 3, or 4 of spades. Thus $P(A \cap C) = 1/13 = P(A) \cdot P(C)$, meaning that $\boxed{A \text{ and } C \text{ are independent}}$.
- Third, $B \cap C$ is the event that the card is an ace and also an ace, 2, 3, or 4, which is the same as saying it is an ace. Thus $P(B \cap C) = 1/13 \neq 4/169 = P(B) \cdot P(C)$, meaning that $\boxed{B \text{ and } C \text{ are not independent}}$.
- Remark: This example shows that independence is not transitive: A and B are independent, as are A and C, yet B and C are not independent.

- If A and B are independent events, then knowing that A occurs does not affect the probability that B occurs. Under this interpretation, it is reasonable to say that A^c and B are also independent events:

- Proposition (Independence of Complements): If A and B are independent events, then so are A^c and B.
  - Proof: First observe that $P(B) = P(A \cap B) + P(A^c \cap B)$, since the events $A \cap B$ and $A^c \cap B$ are mutually disjoint and have union B.
  - Then if A and B are independent, we have $P(A \cap B) = P(A)P(B)$, so we may rearrange the equation above to see that $P(A^c \cap B) = P(B) - P(A)P(B) = [1 - P(A)] \cdot P(B) = P(A^c) \cdot P(B)$.
  - Hence A^c and B are also independent, as claimed.

- We may also define independence of more than two events at once:

- Definition: We say that the events $E_1, E_2, \dots, E_n$ are collectively independent if $P(F_1 \cap F_2 \cap \cdots \cap F_k) = P(F_1) \cdot P(F_2) \cdot \cdots \cdot P(F_n)$ for any subset $F_1, F_2, \dots, F_k$ of $E_1, E_2, \dots, E_n$.
  - By rearranging this definition, one can see that it is equivalent to saying that knowledge of whether some of the events have occurred does not affect the probability of any of the others.
  - Example: A fair coin is flipped 3 times. If E_1 is the event that the first flip is heads, E_2 is the event that the second flip is heads, and E_3 is the event that the third flip is heads, then $P(E_1) = P(E_2) = P(E_3) = \frac{1}{2}$, while $P(E_1 \cap E_2) = \frac{1}{4} = P(E_1)P(E_2)$, $P(E_1 \cap E_3) = \frac{1}{4} = P(E_1)P(E_3)$, $P(E_2 \cap E_3) = \frac{1}{4} = P(E_2)P(E_3)$, and $P(E_1 \cap E_2 \cap E_3) = \frac{1}{8} = P(E_1)P(E_2)P(E_3)$. Thus, these three events are collectively independent.

- If any pair of the events is not independent, then the entire collection of events cannot be collectively independent. On the other hand, it is possible for all of the pairs to be individually independent, but for the entire set not to be independent:

- Example: A fair coin is flipped 3 times. If E_{12} is the event that flips 1 and 2 are the same, E_{13} is the event that flips 1 and 3 are the same, and E_{23} is the event that flips 2 and 3 are the same, show that each pair of these events is independent, but the three events together are not collectively independent.
  - Observe that E_{12} occurs in 4 of the 8 possible flip sequences (TTT, TTH, HHT, HHH), as does E_{13} (TTT, THT, HTH, HHH) and E_{23} (TTT, HTT, THH, HHH), so $P(E_{12}) = P(E_{13}) = P(E_{23}) = \frac{1}{2}$.
  - Also, $E_{12} \cap E_{13}$ is the event where all 3 flips are the same, as is $E_{12} \cap E_{23}$ and $E_{13} \cap E_{23}$, so each of these events has probability $\frac{1}{4}$. Thus $P(E_{12} \cap E_{13}) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(E_{12}) \cdot P(E_{13})$ so these two events are independent, and similarly for the other two pairs.
  - On the other hand, $E_{12} \cap E_{13} \cap E_{23}$ is also the event where all 3 flips are the same, so $P(E_{12} \cap E_{13} \cap E_{23}) = \frac{1}{4} \neq \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = P(E_{12}) \cdot P(E_{13}) \cdot P(E_{23})$.

- Collective independence allows us to calculate probabilities for sequences of independent events:

- Example: A baseball player's batting average is 0.378, meaning that she has a probability of 0.378 of getting a hit on any given at-bat, independently of any other at-bat. If she bats 4 times during a game, compute the probabilities that she gets (i) four hits during the game, (ii) a hit in her first two at-bats but no other hits, (iii) exactly two hits during the game, and (iv) at least one hit during the game.
  - For each of the four at-bats, the player has an independent probability 0.378 of getting a hit. Let E_i be the event of getting a hit on the ith at-bat: then $P(E_i) = 0.378$ and $P(E_i^c) = 1 - 0.378 = 0.622$.
  - Event (i) is $E_1 \cap E_2 \cap E_3 \cap E_4$. Since the four at-bats are independent, the probability of getting a hit during all four at-bats is therefore $P(E_1 \cap E_2 \cap E_3 \cap E_4) = P(E_1) \cdot P(E_2) \cdot P(E_3) \cdot P(E_4) = \boxed{0.378^4} \approx 2.04\%$.
  - Event (ii) is $E_1 \cap E_2 \cap E_3^c \cap E_4^c$. Since the four at-bats are independent (and independent events also have independent complements), the probability of this event is therefore $P(E_1 \cap E_2 \cap E_3^c \cap E_4^c) = P(E_1) \cdot P(E_2) \cdot P(E_3^c) \cdot P(E_4^c) = \boxed{0.378^2 \cdot 0.622^2} \approx 5.53\%$.
  - Event (iii) is the union of $\binom{4}{2} = 6$ possible events (one of which is event (ii)), each of which has 2 hits and 2 non-hits in the 4 at-bats. By independence and the calculation for event (ii), each of these events has probability $0.378^2 \cdot 0.622^2$ and they are all mutually exclusive. Thus, the probability of their union is simply the sum of their individual probabilities, which is $\boxed{6 \cdot 0.378^2 \cdot 0.622^2} \approx 33.17\%$.
  - Event (iv) is the complement of the event of not getting any hits in the game, which is $E_1^c \cap E_2^c \cap E_3^c \cap E_4^c$. By independence, we have $P(E_1^c \cap E_2^c \cap E_3^c \cap E_4^c) = P(E_1^c) \cdot P(E_2^c) \cdot P(E_3^c) \cdot P(E_4^c) = 0.622^4$, and therefore the probability of getting at least one hit in the game is $\boxed{1 - 0.622^4} \approx 85.03\%$.

1.4.3 Computing Probabilities, Bayes' Formula

- By rearranging the definition of conditional probability, we obtain a useful formula for the probability of an intersection of two events, namely, $P(A \cap B) = P(B|A) \cdot P(A)$.
  - Intuitively, one may interpret this formula as saying that if we wish to find the probability that both A and B occur, then first we compute the probability that A occurs, and then we multiply this by the probability that B also occurs (given that A occurred).
  - In many situations, it is much easier to compute the probabilities $P(A)$ and $P(B|A)$ separately by viewing the events of "choosing A" and then "choosing B given A" as being choices made in a sequence.
  - We also remark that this formula is the probabilistic version of the "multiplication formula" from our discussion of counting principles, and we can invoke it in much the same manner.

- Example: Suppose an urn contains 7 red balls and 11 purple balls. If two balls are randomly drawn from the urn without replacement[5], determine the probabilities of the respective events (i) the first ball is red, (ii) the second ball is red given that the first ball is red, (iii) both balls are red, and (iv) the first ball is purple and the second ball is red.
  - Let R_1 be the event that the first ball is red, R_2 be the event that the second ball is red, and P_1 be the event that the first ball is purple.
  - For event (i), we want to compute $P(R_1)$. If we use the sample space consisting only of the first ball drawn from the urn, then each of the 18 outcomes is equally likely, and 7 of them yield a red ball drawn, so we have $P(R_1) = \boxed{\frac{7}{18}} \approx 0.3889$.
  - For event (ii), we want to compute $P(R_2|R_1)$. If we draw the first red ball and then ignore it, drawing the second ball is the same as drawing one ball from an urn containing 6 red balls and 11 purple balls. By the same logic as above, the probability that a red ball is drawn now is $P(R_2|R_1) = \boxed{\frac{6}{17}} \approx 0.3529$.
  - For event (iii), we want to compute $P(R_1 \cap R_2)$. By using the intersection formula, this is equal to $P(R_2|R_1) \cdot P(R_1) = \frac{7}{18} \cdot \frac{6}{17} = \boxed{\frac{7}{51}} \approx 0.1373$.

[5]"Without replacement" means that when a ball is drawn, it is not placed back into the urn. The opposite is "with replacement", where each ball is placed back into the urn after it is drawn.

- For event (iv), we want to compute $P(P_1 \cap R_2) = P(R_2|P_1) \cdot P(P_1)$.
- By the same logic as for events (i) and (ii), we see that $P(P_1) = \frac{11}{18}$ since when we draw the first ball, there are 11 purple balls out of 18 total, and also $P(R_2|P_1) = \frac{7}{17}$ since when we draw the second ball under the assumption that the first one was purple, there are 7 red balls out of 17 total.
- Therefore we see that $P(P_1 \cap R_2) = P(R_2|P_1) \cdot P(P_1) = \frac{7}{17} \cdot \frac{11}{18} = \boxed{\frac{77}{306}} \approx 0.2516$.

- We may also iteratively apply the result above to obtain formulas for intersections of more than two events.
  - For example, for three events A, B, C, we have $P(A \cap B \cap C) = P(C|A \cap B) \cdot P(A \cap B) = P(C|A \cap B) \cdot P(B|A) \cdot P(A)$.
  - If we view these probabilities as a sequence of choices, then this formula tells us that we can compute the probability of $A \cap B \cap C$ by "choosing A", then "choosing B given A", then "choosing C given both A and B".
  - The same idea extends to intersections of four or more events: the probability of the intersection can be found by computing the probabilities the events one at a time (each conditional on all of the previous events listed), and then multiply all of the results.
  - The advantage to this approach is that it allows us to break complicated events down into simpler ones whose probabilities can often be computed quickly by working with a small sample space (like in the example above with the urns where we only considered a single ball being drawn from the urn).

- By combining these ideas with our results on counting, we can solve a wide array of probability problems.

- Example (Monty Hall Problem): On a game show, a contestant chooses one of three doors: behind one door is a car and behind the other two are goats. The host opens one of the two unchosen doors to reveal a goat, and then offers the contestant the option of switching their choice from their original door to the remaining unopened door in the hopes of winning the car[6]. Should the contestant accept the offer to switch doors?
  - We want to compute the probability that the car is behind the contestant's door. Suppose we label the contestant's door number 1, the door opened by the host number 2, and the remaining door number 3.
  - Let P_1, P_2, P_3 be the events in which the prize is behind door 1, 2, or 3 respectively, and H_2 and H_3 be the events in which the host opens door 2 and door 3 respectively. We wish to compute the conditional probability $P(P_1|H_2) = \frac{P(P_1 \cap H_2)}{P(H_2)}$.
  - We know that $P(P_1) = P(P_2) = P(P_3) = 1/3$ since the car is equally likely to be behind any of the doors at the start of the game, and also $P(H_2) = P(H_3) = 1/2$ since by symmetry the host is equally likely to open door 2 or door 3.
  - So we need only compute $P(P_1 \cap H_2) = P(H_2|P_1) \cdot P(P_1)$. But $P(H_2|P_1) = P(H_3|P_1) = 1/2$ since if the prize is behind door 1, the host is equally likely to open door 2 or door 3.
  - Therefore, we get $P(P_1 \cap H_2) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$, and then $P(P_1|H_2) = \frac{P(P_1 \cap H_2)}{P(H_2)} = \frac{1/6}{1/2} = \frac{1}{3}$.
  - This means that there is a 1/3 probability that the prize is behind door 1 (the door chosen by the contestant), and thus there is a 2/3 probability that the prize is behind door 3 (the remaining unchosen door), so it is better to switch doors.
  - Remark: This is a famous probability puzzle originally popularized in this form by vos Savant in 1990. Although she gave the correct answer to the puzzle, she evidently received many thousands of letters from readers who disagreed with the answer! (It is a very common mistake to argue that because there are now only 2 doors remaining to choose between, the probability that the prize is behind each of them must be 1/2. As we have seen, this is not correct!)

[6] For this particular scenario, we also assume that the car is randomly hidden behind one of the doors, that the host knows what is behind each door, that the host always opens a door that the contestant has not chosen that reveals a goat (randomly selecting between the two if the contestant has chosen the car), and that the host always offers the contestant the option to switch doors.

- Example (Two-Children Problem): Suppose that children are equally likely to be male or female[7] and that their sexes are independent. Mr. Jones has two children, and the older child is a girl: what is the probability that both children are girls? Mr. Smith has two children, and at least one of them is a boy: what is the probability that both children are boys?
  - Let BB, BG, GB, and GG be the four possible outcomes for the two children (with the younger sibling listed first): then we have $P(BB) = P(BG) = P(GB) = P(GG) = 1/4$.
  - If $G2$ is the event that the older child is a girl, then $P(G2) = 1/2$, and so the probability that the Jones family has two girls is $P(GG|G2) = \dfrac{P(GG \cap G2)}{P(G2)} = \dfrac{1/4}{1/2} = 1/2$.
  - If B is the event of having at least one boy, then $P(B) = 3/4$, and so the probability that the Smith family has two boys is $P(BB|B) = \dfrac{P(BB \cap B)}{P(B)} = \dfrac{P(BB)}{P(B)} = \dfrac{1/4}{3/4} = 1/3$, not $1/2$.
  - We can reason this out more intuitively by noting that for the Jones family the probability that both children are girls is the same as the probability that the younger child is a girl, which is 1/2. On the other hand, for the Smith family, there are three scenarios in which the Smith family has at least one boy, but in only one of these do they have two boys, so the probability should be 1/3.
  - Remark: This example illustrates that the value of a probability can change substantially with small alterations to the given information. It is natural to assume that these two probabilities should be the same, since the additional information of specifying which of the children is a boy or a girl seems like it should not make a difference, but it does!

- Example (Birthday Problem): Assuming that birthdays are randomly distributed among the 365 days in a non-leap year, and ignoring February 29th, what is the probability p_n that in a group of n people, some pair have the same birthday? Also determine the smallest number of people for which there is at least a 50% chance of having two people with the same birthday.
  - We compute the probability of the complementary event by choosing the birthdays for the members of the group one at a time and ensuring that no two have the same birthday.
  - The first person may have any birthday. The second person may have any birthday except the first person's birthday, and the probability that this occurs is 364/365.
  - The third person may have any birthday except that of the first and second person, and the probability that this occurs is 363/365.
  - We continue in this way, observing that the kth person may have any of $366 - k$ possible birthdays, and that the probability of this event (conditioned on the previous ones) is $(366 - k)/365$.
  - Then by our formula for the probability of the intersection of events, the desired probability that there is a pair of people with the same birthday is $p_n = \boxed{1 - \dfrac{365}{365} \cdot \dfrac{364}{365} \cdot \dfrac{363}{365} \cdot \cdots \cdot \dfrac{366-n}{365}}$.
  - Here is a table of values of this probability for various n:

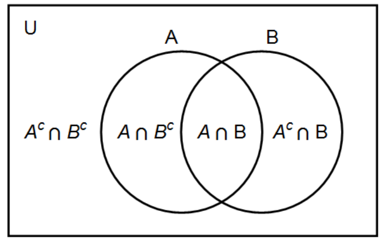
n	5	10	15	20	22	23	25	30	40	50
p_n	2.7%	11.7%	25.3%	41.1%	47.6%	50.7%	56.9%	70.6%	89.1%	97.0%

  - From the table we can see that the minimal n such that there is a greater-than-50% chance of having at least one pair of people with the same birthday is $n = \boxed{23}$.
  - Remark: The fact that the number of people required for a decent probability of having 2 with the same birthday is only 23 is generally found to be surprising, as many people typically guess that a much larger number (e.g., around half of 365) would be needed to have a high probability of getting a match.
  - One way to estimate the correct answer is to observe that the desired event is the union of the events that one of the pairs of people in the group has the same birthday. The probability that any given pair has the same birthday is 1/365, and although these events are not disjoint, they are moderately close. With k people, an estimate for the probability of getting at least one match is then $\binom{k}{2}/365$, and setting this equal to 50% leads to an estimate $k \approx \sqrt{365} \approx 19.1$, not too far away from the actual answer of 23.

[7] This is not actually the case for humans; at birth the ratio is approximately 1.03-1.06 males per female, with somewhere between 0.02%-1.7% of live births being intersex, depending on the criterion used. The sexes of siblings are also not completely independent.

- Example: Suppose A and B are events such that $P(A) = 0.6$, $P(B|A) = 0.7$, and $P(B|A^c) = 0.2$. Find (i) $P(A \cap B)$, (ii) $P(A \cap B^c)$, (iii) $P(A^c \cap B)$, (iv) $P(B)$, (v) $P(A|B)$, (vi) $P(A \cup B)$, and (vii) $P(A^c \cup B^c)$.
    - In computing probabilities of this form, it is useful to label the results on a Venn diagram like this one:

Venn: Unions and Intersections



    - For (i), we know that $P(A \cap B) = P(B|A) \cdot P(A) = 0.7 \cdot 0.6 = \boxed{0.42}$.
    - For (ii), we know that $P(A) = 0.6$ and from above we also have $P(A \cap B) = 0.42$, so per a Venn diagram we see that $P(A \cap B^c) = 0.6 - 0.42 = \boxed{0.18}$.
    - For (iii), we may write $P(A^c \cap B) = P(B|A^c) \cdot P(A^c) = 0.2 \cdot (1 - 0.6) = \boxed{0.08}$.
    - For (iv), we know that $P(A \cap B) = 0.42$ and from above we also have $P(A^c \cap B) = 0.08$, so per a Venn diagram we see that $P(B) = P(A \cap B) + P(A^c \cap B) = 0.42 + 0.08 = \boxed{0.5}$.
    - For (v), we have $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0.42}{0.5} = \boxed{0.84}$ using the results we found above.
    - For (vi), we have $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.5 - 0.42 = \boxed{0.68}$.
    - For (vii), by a Venn diagram we have $P(A^c \cup B^c) = 1 - P(A \cap B) = \boxed{0.58}$.

- Example: A medical test for a rare disease will detect the disease in 99% of test samples from patients who have the disease, but it also has a false positive rate of 0.5%, meaning that it also detects the disease in 0.5% of test samples from patients who do not have the disease. If 0.1% of the population actually has the disease, what is the probability that a randomly chosen patient who tests positive actually has the disease?
    - Let D be the event that a person has the disease and $+$ be the event that they test positive for the disease: we wish to compute the conditional probability $P(D|+) = \dfrac{P(D \cap +)}{P(+)}$.
    - We are given that $P(D) = 0.1\% = 0.001$ and thus $P(D^c) = 0.999$, and also that $P(+|D) = 99\% = 0.99$ and $P(+|D^c) = 0.5\% = 0.005$.
    - Therefore, we have $P(D \cap +) = P(+|D) \cdot P(D) = 0.99 \cdot 0.001 = 0.00099$.
    - We also have $P(+) = P(+ \cap D) + P(+ \cap D^c)$, and we can compute the latter as $P(+ \cap D^c) = P(+|D^c) \cdot P(D^c) = 0.005 \cdot 0.999 = 0.004995$.
    - Therefore, $P(+) = P(+ \cap D) + P(+ \cap D^c) = 0.005985$.
    - Hence the desired conditional probability is $\dfrac{0.00099}{0.005985} \approx \boxed{16.54\%}$.
    - We thus obtain the (perhaps surprising!) result that a patient who tests positive for the disease only has about a 1-in-6 chance of actually having the disease, despite the fact that the test yields the correct result 99% of the time for positive patients and 99.5% of the time for negative patients.
    - Ultimately, the reason for this disparity is that a typical person is very unlikely to have the disease, and thus there are far more false positives than actual cases of the disease. Thus, even if a person tests positive once, it is still not especially likely that they have the disease, unless there is some reason to think that the person was not a random member of the population.

- In the last two examples above, we ended up computing a conditional probability using the values of the conditional probabilities "in the other order": in the first example we used $P(B|A)$ and $P(B|A^c)$ to find $P(A|B)$, while in the second example we used $P(+|D)$ and $P(+|D^c)$ to find $P(D|+)$. We can in fact write down a general formula for this calculation:

- Theorem (Bayes' Formula): If A and B are any events, then $P(B|A) = \dfrac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)}$. More generally, if events $B_1, B_2, \dots, B_k$ are mutually exclusive and have union the entire sample space, then $P(B_i|A) = \dfrac{P(A|B_i) \cdot P(B_i)}{P(A|B_1) \cdot P(B_1) + \cdots + P(A|B_k) \cdot P(B_k)}$.
  - Proof: By definition, $P(B|A) = P(A \cap B)/P(A)$, and we also have $P(A \cap B) = P(A|B) \cdot P(B)$.
  - We also have $P(A) = P(A \cap B) + P(A \cap B^c)$ since these events are mutually exclusive and have union A.
  - Then $P(A \cap B) = P(A|B) \cdot P(B)$ and $P(A \cap B^c) = P(A|B^c) \cdot P(B^c)$ from the definition of conditional probability; plugging all of these values in yields the formula immediately.
  - The second formula follows in the same way by writing $P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k)$.

- Example: Alex has two urns, one labeled H which has 5 black and 4 red balls, and one labeled T which has 4 black and 11 red balls. Alex flips an unfair coin that has a 2/3 probability of landing heads, and then draws one ball at random from the correspondingly-labeled urn (H for heads, T for tails). If Alex draws a red ball, what is the probability that the coin flip was tails?
  - Let R, H, and T denote the events of drawing a red ball, flipping heads, and flipping tails.
  - We want $P(H|R)$, which we may get via Bayes' formula: $P(H|R) = \dfrac{P(R|H) \cdot P(H)}{P(R|H) \cdot P(H) + P(R|T) \cdot P(T)}$
  - We have $P(H) = 2/3$, $P(T) = 1/3$, and also $P(R|H) = 4/9$ and $P(R|T) = 11/15$.
  - Therefore, $P(H|R) = \dfrac{P(R|H) \cdot P(H)}{P(R|H) \cdot P(H) + P(R|T) \cdot P(T)} = \dfrac{4/9 \cdot 2/3}{4/9 \cdot 2/3 + 11/15 \cdot 1/3} = \boxed{\dfrac{40}{73}} \approx 54.8\%$.

- Example (Prosecutor's Fallacy): A DNA sample from a minor crime is compared to a state forensic database containing 100,000 records and a single suspect is identified on this basis alone, with no other evidence suggesting guilt or innocence. From analysis of human genetic variation, it is determined that the probability that a randomly-selected innocent person would match the DNA sample is 1 in 10000. At the trial, the prosecutor states that the probability that the suspect is innocent is only 1 in 10000, and observes that this figure means that it is overwhelmingly likely the suspect is guilty. Critique this statement.
  - Suppose M is the event that there is a DNA match, and I is the event that the suspect is innocent.
  - The conditional probability $P(I|M)$ is the probability that the suspect is innocent given that there is a DNA match, which is what the prosecutor is claiming is equal to 1/10000.
  - However, the 1-in-10000 figure is actually the probability that there is a match given that the suspect is innocent: this is the conditional probability $P(M|I)$, which as is quite different from $P(I|M)$!
  - In this case we may use Bayes' formula to write $P(I|M) = \dfrac{P(M|I) \cdot P(I)}{P(M|I) \cdot P(I) + P(M|I^c) \cdot P(I^c)}$.
  - A priori, there is no reason to believe that the given suspect is any more likely to be guilty than any other person in the database, so we will take $P(I^c) = 1/100000 = 0.00001$ and then $P(I) = 0.99999$.
  - We also take $P(M|I) = 1/10000 = 0.0001$ as indicated, and $P(M|I^c) = 1$ since (we presume) the DNA analysis will always identify a guilty suspect.
  - Then we obtain $P(I|M) = \dfrac{0.0001 \cdot 0.99999}{0.0001 \cdot 0.99999 + 1 \cdot 0.00001} \approx 90.9\%$.
  - Our calculation shows there is a 90.9% chance that the suspect is innocent, given the existence of a positive match and no other evidence: quite a far cry from the prosecutor's claim of 99.999% guilt!
  - Remark: The confusion of the probability of innocence given a positive match with the probability of a positive match given innocence is called the "prosecutor's fallacy", and (as shown with this dramatic example) it is a very serious error. In this particular example, one would expect approximately $100000/10000 = 10$ DNA matches to come from the database, and given the lack of evidence to say otherwise, it is no more likely that the given suspect is guilty than any of these 10 people.

Well, you're at the end of my handout. Hope it was helpful.