9/15/2022

Definition of probability

Probability is a proportion: the number of times the event you are interested in occurs divided by the total number of possible events. Probabilities are often described as the chance or the likelihood. Since it is a proportion, it can be expressed either as a fraction, a percent or a decimal.

Probabilities can come in several flavors:

- Classical or theoretical
- Experimental or empirical
- Subjective or personal

Classical or theoretical probabilities are derived from mathematics. Listing out all the possible outcomes, determining which of those outcomes are in the event in question and then dividing under the assumption that all outcomes are equally likely, or some other theoretical derivation. An example would be the probability of a fair coin coming up heads.

Experimental or empirical probabilities are derived from data, from experiments. A random event is sampled many times and then the ratio of the successes to the total trials is calculated. Empirical probability and classical probabilities are related in that the Law of Large Numbers applies: the larger the number of experiments tried, the closer the empirical probability will become to the true or classical probability. We will say more about the mathematics of this (how close is close?) later on in the course when we talk about the Central Limit Theorem. And example would be rolling a dice over and over to determine if the dice is weighted or fair.

Subjective or personal probability are for statements of chance obtained from methods other than the two above. One might consider it sometimes an expression of confidence. It may refer to probabilities that are too rare to calculate experimentally. Other non-empirical factors may be taken into consideration. An example might be determining the probability of another 9/11. Since that event (or anything like it) has only occurred once, the best we can do is make an educated guess about the likelihood of a repeat. It might be an educated guess, but still, a guess. Another example might be saying that you are 95% sure you passed your exam before you see the score. Since you can't repeat the exam under the same circumstances, this is not empirical or classical. It, too, is a best guess, or an expression of confidence about the outcome, rather than a true probability.

Set Notation

In order to talk about probability, we need to talk about sets and set notation.

- A set is an unordered list of objects that does not have any repeated elements. Sets are often listed using set notation as a list in {}, such as the outcomes of the roll of a standard dice {1, 2, 3, 4, 5, 6}. If the list is too long, you can use ellipses. It can also be written using mathematical set notation such as $\{x|condition\}$, for example $\{x|x\ is\ an\ integer\}$, which is equivalent to $\{\ldots -2, -1, 0, 1, 2, \ldots\}$. Sets do not need to have only numbers as elements.
- Typically, sets are given capital letters as names, such as A. Elements in sets are given lower case letters, like a.

- Relatedly, a random variable is typically given a capital letter, like X. The value of a random variable is given a lower case letter, like x.
- To say that an element belongs to a set, we use the symbol ∈. The statement $a \in A$ is read "a is an element of the set A", or "a in A" for short.
- A subset is a set for which all the elements in that set are also elements in another (usually larger) set. To say that set A is a subset of set B, we write $A \subset B$. $A \subseteq B$ is often used interchangeably. Another way to put this is that if $a \in A, and\ A \subset B, then\ a \in B$.
- The union of two sets is the set where any element that is in either A or B is in the union. This is written as $A \cup B$.
- The intersection of two sets is the set such that only the elements in both A and B are in the intersection. We write the intersection as $A \cap B$.
- The complement of a set is written as A', but may also be written as $\bar{A}$, $\sim A$, or several other notations, depending the author of the text. All these notations can be read as "not A". If $a \in A, then\ a \notin A'$.
- The universal set is the set of all possible elements (usually limited by context).
- A convenient but less common notation is $A - B$ read as "A minus B". This is the set where any element of the intersection of A and B is removed from A. Put another way, it is the set of all elements of A that are not shared with B.
- To describe the number of events in a set we can write $|A|$, or sometimes $n(A)$. The number of elements in a set can be described as its cardinality.
- The empty set is a set with no elements. Its cardinality is 0. It can be written ∅ or $\{\ \}$.

Sample space – is like the universal set described above. It is the set of all possible outcome of a given random variable. Typically, the same space is broken down into a list of elements that are as simple as possible so that each element in the set has equally likely outcomes, but this is not required.

Events

- A simple event is an event that can't be broken down any further, for instance, the outcome of a roll of a die that contains only one possible result.
- A compound event is a collection of simple events. For instance, the outcome of any even numbered face when rolling a standard die. This event is made up of the simple events {2, 4, 6}.

Probability notation:
The probability of event A = $P(A)$. The probability that a random variable X takes the value 3 is written $P(X = 3)$.

To find the classical probability of find the probability of rolling an even number on a standard die: $P(even) = \frac{3}{6} = \frac{1}{2}$. The number of events in the sample space of a standard die roll is $|\{1, 2, 3, 4, 5, 6\}| = 6$. The number of outcomes in the event "rolls an even number" is $|\{2, 4, 6\}| = 3$. So the probability is $\frac{3}{6}$, which reduces to one half.

If the probability of an event is 0, then the event is impossible.
If the probability of an event is 1, then the event is certain (it must occur).
All other probabilities must be between 0 and 1, and the sum of all possible non-overlapping probabilities must add to 1.

Since all probabilities must add to 1, if $P(A) = p$, then $P(A') = 1 - p$.

Events that cannot co-occur (non-overlapping) are mutually exclusive. Simple events in sample space should be mutually exclusive events.

The expression $P(A|B)$ is read "the probability of A given B" is the probability of A once you know that B has occurred. This is a conditional probability.

The definition of independent probability is that $P(A|B) = P(A)$. In other words, knowing that B has occurred has not changed the probability of A occurring. If knowing that B has occurred means that the probability A changes, then these events are dependent.

An example of two events that are independent are flipping a coin and rolling a die. The outcome of the coin toss does not have any impact on the outcome of the die roll. An example of two events that are dependent are gender and baldness. In generally, men are more likely to suffer from hair loss than women are, so if you know the gender already, you know the probability is different (higher or lower) than the whole population taken together.

The odds of an event is the probability of the event divided by the probability of its complement, i.e. the odds of event A is $\frac{P(A)}{P(A')} = \frac{p}{1-p}$. The odds against an event A are $\frac{P(A')}{P(A)} = \frac{(1-p)}{p}$.

Examples. If the odds of event A are 4:3, then the total number of events in space are 4+3=7, and $P(A) = \frac{4}{7}$, and $P(A') = \frac{3}{7}$. (Confirm that this satisfies the complement rule above.)

If $P(A) = \frac{4}{11}$, what are the odds against A? First, we find $P(A') = 1 - \frac{4}{11} = \frac{7}{11}$. The odds against A are $\frac{P(A')}{P(A)} = \frac{\frac{7}{11}}{\frac{4}{11}} = \frac{7}{4}$ or 7:4. Odds can be greater than 1.

Odds will not come up often, but they will come up next semester when we talk about logistic regression.

If two sets are independent, then $P(A \cap B) = P(A)P(B)$. If the events are dependent, then $P(A \cap B) = P(A|B)P(B)$. We will use this relationship later to derive Bayes' Rule.

If two sets are mutually exclusive (i.e. $P(A \cap B) = 0$, then $P(A \cup B) = P(A) + P(B)$. If the intersection is not empty, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. The extra subtraction is necessary because the elements in the intersection are counted in both sets and so we need to subtract off one count to avoid this double-counting.

The extension of this rule to three sets is provided in the Devore book. There is even more subtracting off, and then adding back in again.

Some probability problems, especially those that intersection and union, can be done without the use of these formulas, particularly those expressed in two-way (contingency) tables.

Handed-ness / Sex	Right-handed	Left-handed	Total
Male	43	9	52
Female	44	4	48
Total	87	13	100

- What is the probability that a randomly selected person from the sample described in the table is left-handed? $\frac{13}{100} = 0.13$.
- What is the probability that a randomly selected person from the sample described in the table is female? $\frac{48}{100} = 0.48$
- What is the probability that a randomly selected person from the sample described in the table is both female and left-handed? $\frac{4}{100} = 0.04$
- What is the probability that a randomly selected person from the sample described in the table is female given that they are left-handed? $\frac{4}{13} \approx 0.3077$.
- What is the probability that a randomly selected person from the sample described in the table is female or left-handed? $\frac{13}{100} + \frac{48}{100} - \frac{4}{100} = \frac{57}{100} = 0.57$
- In this data set, are gender and handedness independent? $P(F) = 0.48, P(F|left) \approx 0.3077$. These probabilities are not equal, so they are dependent.

Calculating classical probabilities can be a problem when the events and sample spaces are very large. In order to calculate these probabilities, we need methods for calculating the number of events in the sets without having to list them. We are going to look at three special counting rules that will be most useful.

Multiplication rule (be careful not to confuse this with the Multiplication rules for independent probabilities). Use this rule when you are combining multiple smaller events, or when repetition is allowed.

Examples. How many different outcomes are there when you flip a coin ten times? Each flip has 2 outcomes so the total number of outcomes is $2^{10} = 1024$.

How many outcomes are there if you roll three dice (one blue, one red and one green)? There are 3 dice and if each one is a standard six-sided die then that is $6^3 = 216$. But if one die is a standard die, one is a octahedral die (8-sided) and one is a dodecahedral die (12-sided), then the number of outcomes is $6 \times 8 \times 12 = 576$.

How many different license plates are there is there is three letters (without O) and then three numbers? Each letter had 25 possibilities (since we are excluding O) and each number has 10 possibilities (0,1,2,…,9) so we have $25 \times 25 \times 25 \times 10 \times 10 \times 10 = 15{,}625{,}000$.

Permutations are used when selecting repeatedly from the same set, repetition is not allowed, but the order of selection matters. The formula for the number of permutations uses n for the number of things being selected from and r to be the number of things selected. $P(n,r) = nPr = P_{n,r} = \frac{n!}{(n-r)!}$. The ! is a factorial. $7! = 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 5040$. $0! = 1$.

Examples. Suppose you want to pick three officers (president, vice-president and treasurer) for a club from 7 members. How many different combinations of officers are there? Since the order matters (president is a different thing than vice-president), and no one can have two jobs (no repetition), this is a permutation. Picking from 7 people and selecting 3 of them gives us $P(7,3) = \frac{7!}{(7-3)!} = \frac{7!}{4!} = \frac{7\times6\times5\times4\times3\times2\times1}{4\times3\times2\times1} = 7 \times 6 \times 5 = 210$.

Suppose that you pick 9 people out of 14 kids to play the field in a t-ball game. Different positions are different, and the same kid can't play two spots, so this is a permutation. $P(14,9) = \frac{14!}{5!} = 726{,}485{,}760$. Permutations get big fast.

Combinations are used when selecting repeatedly from the same set, repetition is not allowed, but the order of selection does not matter. The formula is $C(n,r) = nCr = C_{n,r} = \binom{n}{r} = \frac{n!}{(n-r)!r!} = \frac{P_{n,r}}{r!}$. Where n is the number being chosen from and r is the number chosen. The last notation $\binom{n}{r}$ is important because this is usually the way combinations are expressed in formulas we will encounter in the next class.

Examples. In the club example earlier with 7 members, what if we were choosing three people to be on a committee together instead of officers? Since the committee membership is all the same, the order doesn't matter, but one person can't serve in two positions on the committee, so this is still no repetition. Use $C(7,3) = \binom{7}{3} = \frac{7!}{4!3!} = \frac{210}{3!} = \frac{210}{3\times2\times1} = 35$.

What if a company was drawing a raffle to give away four vacations as a holiday bonus chosen from among 20 employees? This is also a combination since the prizes are all the same and one employee can't win more than one vacation. So this is $\binom{20}{4} = \frac{20!}{16!4!} = 4845$.

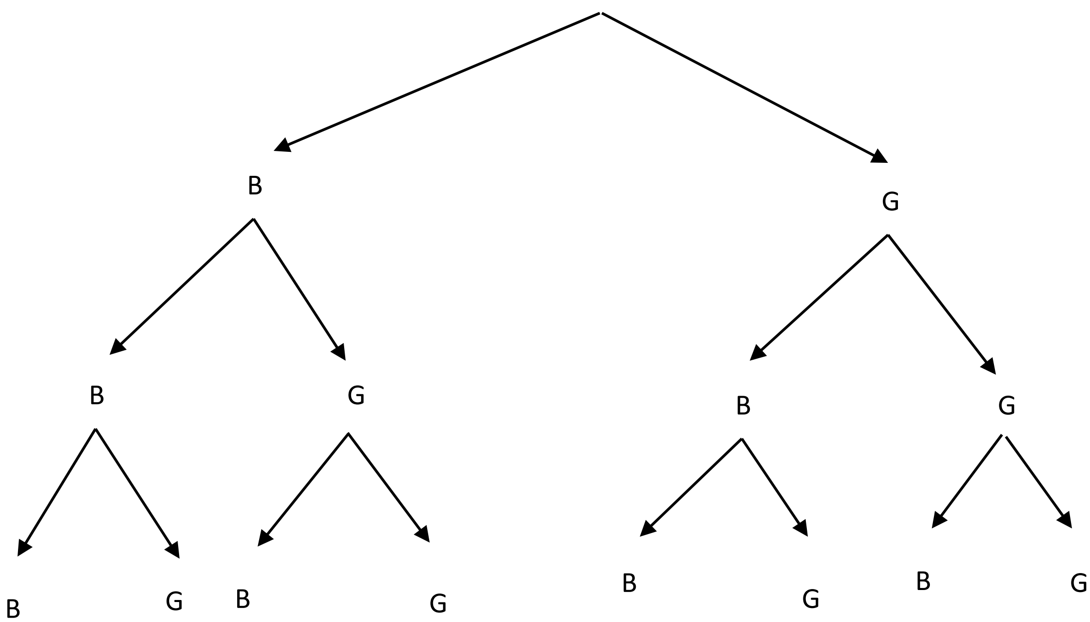
Hands of poker also run by combinations in the simplest scenarios.

Other special rules exist for when some elements can't be told apart, or when we don't care about the order but we can repeat. We won't use these in this course, but a discrete math textbook will cover them. It turns out that when have two classes out outcomes we can't tell apart the special rule becomes equivalent to the combination formula. Such is the case when we flip coins. All the heads look the same, and all the tails look the same, so if we want to know how many ways we can get 6 heads in 10 flips, the count of that is a combination also: $\binom{10}{6} = \frac{10!}{4!6!} = 210$. If we have three classes or more classes, we need the special formula, so this only works when we have two outcomes for each flip.

What if we wanted to find the probability of getting 6 heads if we flipped a coin 10 times? If the coin is fair, then we just count the number of ways to get 6 heads out of 10 flips and divide by the total number of outcomes of 10 flips: $\frac{210}{1024} \approx 0.205$. If the coin is not fair, we'll tackle that problem in the next lecture.

Tree Diagrams

To help visualize the multiplication rule, if there aren't that many outcomes, a tree diagram can come in handy. Suppose I want to list all the possible outcomes in the sample space of having three children (by their gender) or flipping three fair coins.



Each level is one flip of the coin or one birth in this case. Similar to a decision tree, but it's chance that is deciding. The first child can either be a boy or a girl. The next level are the possible outcomes from that birth and the third level are the outcomes for the third child. If you read down the branches of the tree, you can find all the possible outcomes. You could have all boys: BBB. Or you could have two boys and then a girl: BBG. Or you could have a boy, then a girl, then a boy: BGB, and so on. $\{BBB, BBG, BGB, BGG, GBB, GBG, GGB, GGG\}$, which is 8 outcomes, and if you look at the end of the tree, you see there are 8 outcomes at the bottom level: $2 \times 2 \times 2 = 8$, one 2 for each level.

This works for small problems, but it does get unwieldy if you have a lot of options at each level or many more levels.

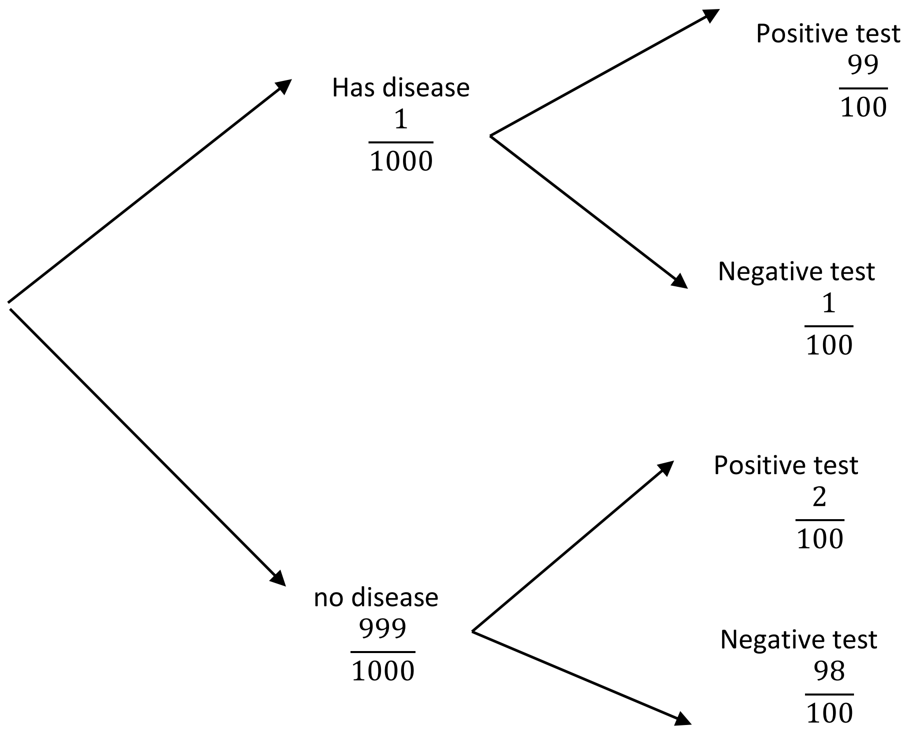
Tree diagrams can also be useful for thinking about Bayes' Rule. Bayes' Rule itself is just a way of rearranging the conditional probability formula.

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

Where we usually apply this is in problems where the probability of B is given in pieces, or for where we might have $P(B|A)$ instead of $P(A|B)$. These aren't interchangeable.

Suppose there is a test for a rare disease. Only 1 in 1000 adults has the disease. A test is developed where if the person actually has the disease, then 99% of the time, the patient will be correctly diagnosed from a positive test. But an individual without the disease will also test positive 2% of the time. What is the probability that a person with a positive test actually has the disease?

Building a tree will help here.



We can multiply the probabilities going down the tree branches since the second level are the conditional probabilities from the first level.

$$P(disease \cap positive) = \frac{1}{1000} \times \frac{99}{100}$$
$$P(disease \cap negative) = \frac{1}{1000} \times \frac{1}{100}$$
$$P(no\ disease \cap positive) = \frac{999}{1000} \times \frac{2}{100}$$
$$P(no\ disease \cap negative) = \frac{999}{1000} \times \frac{98}{100}$$

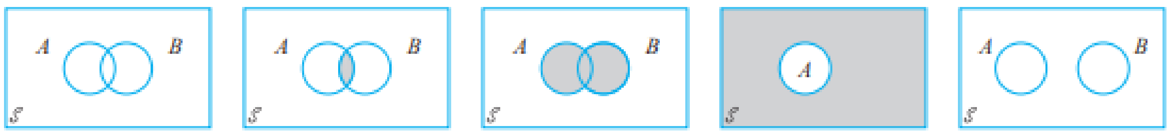
Using Bayes' Rule with A = disease and B = positive, the numerator is $P(disease \cap positive)$, the denominator is the sum of all the positive cases so $P(disease \cap positive) + P(no\ disease \cap positive)$, and if we divide we will get the conditional probability we desire.

$$P(disease|positive) = \frac{\frac{1}{1000} \times \frac{99}{100}}{\frac{1}{1000} \times \frac{99}{100} + \frac{999}{1000} \times \frac{2}{100}} = \frac{9.9 \times 10^{-4}}{0.02097} \approx 0.0472$$

This means that because the disease is so rare, the false positives usually overwhelm the true positives. You need a much lower rate of false positives for the test to be useful. Put another way, if you get a positive test for this disease, there is better than 95% chance that it's a false alarm.

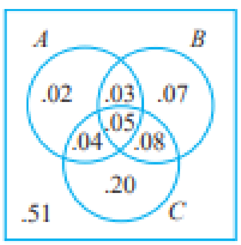
Venn Diagrams

Sometimes its useful for thinking about probabilities using pictures and Venn diagrams help us do that.



(a) Venn diagram of events A and B (b) Shaded region is $A \cap B$ (c) Shaded region is $A \cup B$ (d) Shaded region is A' (e) Mutually exclusive events

It can be helpful to see relations between sets. Venn diagrams can also be helpful to organize data in a problem.



For more on Venn diagrams, I've written two handouts on Venn diagrams and set notation, and Venn diagrams and probability. I've linked them both in the reference list below. They are also linked on my Archive Site. You may find them useful for completing some of the written homework problems.

References:

1. https://faculty.ksu.edu.sa/sites/default/files/probability_and_statistics_for_engineering_and_the_sciences.pdf
2. https://assets.openstax.org/oscms-prodcms/media/documents/IntroductoryStatistics-OP_i6tAI7e.pdf
3. http://betsymccall.net/prof/courses/mathnotes/handouts/Venn%20Diagrams%20and%20Set%20Notation.pdf
4. http://betsymccall.net/prof/courses/mathnotes/handouts/statistics/Venn%20Diagrams%20and%20Probability.pdf
5. http://betsymccall.net/prof/courses/mathnotes/handouts/statistics/A%20Brief%20Survey%20of%20Statistical%20Symbols.pdf
6. https://community.rstudio.com/t/how-do-you-convert-a-tidy-data-frame-to-a-contingency-table/99562