

Why probability?

Basic probability concepts

Approaches to probability calculation

- Another major approach to probability is the Bayesian approach (pronounced “bay-zee-an”).
- The Bayesian approach is based on the idea that probability is a measure of the degree of belief that an event will occur.
 - Hence, it is like a measure of how “good” our information is.
- We won’t focus Bayesian approaches on this module, but it is good to know that it exists, and researchers have devoted a lot of time to it.
 - Typically, we will want to understand frequentist approaches first and gain mathematical intuition before we move to Bayesian approaches.

Probability: the world of board games

- Two typical examples often emerge when discussing probability:
 - Flipping a “fair” coin
 - Rolling a “fair” die
- A fair coin is a coin that has an equal probability of landing on heads or tails.
- So, because there is 1 head and 1 tail, the probability of getting a head is $1/2$, and the probability of getting a tail is $1/2$.

$$(P(H)) = \frac{\text{Number of heads}}{\text{Total number of outcomes}} = \frac{1}{2}$$

$$(P(T)) = \frac{\text{Number of tails}}{\text{Total number of outcomes}} = \frac{1}{2}$$

The lingo: experiments, events, outcomes and sample spaces

- There is specific terminology and definitions in probability that must be understood before we move forward.
- An **experiment** is a process that generates well-defined outcomes, and we're interested in probabilities associated with these outcomes.
 - For example, flipping a coin, rolling a die, or drawing a card from a deck.
- An **event** is a subset of the outcomes of an experiment.
 - For example, in flipping a coin, the event of getting a head is a subset of the outcomes of the experiment.

Set theory and probability

- Probability is closely related to set theory.
 - Sets are collections of objects, and in probability, we often think of events as sets of outcomes.
- Thus, it is common to see notation from set theory in probability.
 - A set is denoted by curly braces, e.g. $\{1, 2, 3, 4, 5, 6\}$.
 - We use letters to denote sets, e.g. A, B, C, etc.
 - We use set operations to calculate probabilities.
 - We use Venn diagrams to visualize probabilities.
- Set theory operators, such as belongs to (\in), union (\cup), intersection (\cap), and complement ($'$) are used in probability.
 - These often also come in terms of logical operators, such as “and” and “or”.

Basic operations in probability

- There are three basic operations in probability:
 - Union
 - Intersection
 - Complement

Union

- The union of two events A and B is the event that *either* A or B or both occur.
 - Meaning that any two of them can occur, but not necessarily both.
 - The union of A and B is denoted by $A \cup B$ (set theory notation).
 - The probability of the union of A and B is denoted by $P(A \cup B)$.
- Example: what is the probability of getting one or two when rolling a die?
 - Notice the use of “or” in the question. Union is often associated with the “OR” logical operator.
 - This can be written as $P(A \cup B)$, where A is the event of getting a 1 and B is the event of getting a 2.

Venn Diagrams

- A Venn diagram is a visual representation of sets and their relationships.
- In probability, we often use Venn diagrams to visualize the union of two events.
- The union of two events is the shaded area in the Venn diagram that covers both events.

Union in a Venn diagram

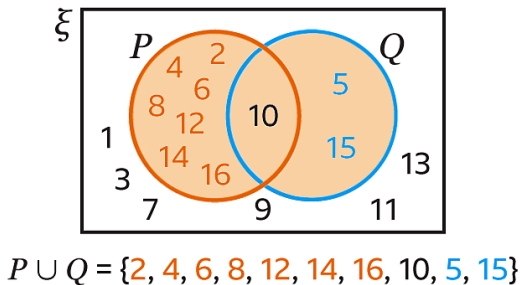


Figure 1: Union in a Venn diagram. Source: BBC Bitesize

Intersection

- The intersection of two events A and B is the event that *both* A and B occur.
 - Meaning that both events must occur.
 - The intersection of A and B is denoted by $A \cap B$ (set theory notation).
 - The probability of the intersection of A and B is denoted by $P(A \cap B)$.
- Example: what is the probability of getting an even number **and** a number less than 4 when rolling a die?
 - Intersection is often associated with the “AND” logical operator.

Intersection in a Venn diagram

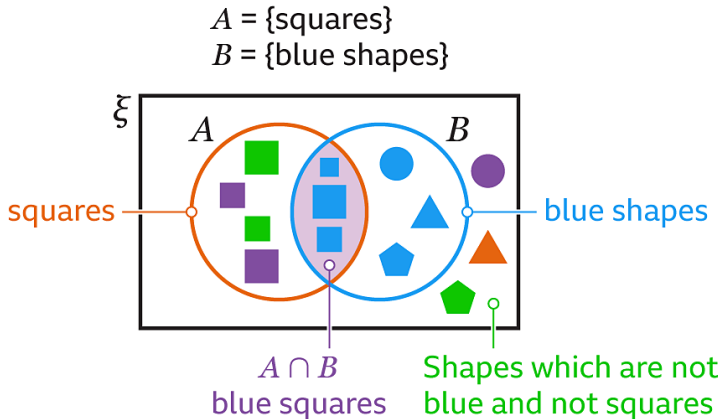


Figure 2: Intersection in a Venn diagram. Source: BBC Bitesize

Complement

- The complement of an event A is the event that A does *not* occur.
 - Meaning that the event does not happen.
 - The complement of A is denoted by A' (set theory notation).
 - The probability of the complement of A is denoted by $P(A')$.
- Example: what is the probability of not getting a 1 when rolling a die?
 - The complement of getting a 1 is not getting a 1.
 - The complement of an event is often associated with the “NOT” logical operator.
- It is often very easy to calculate the complement of an event, as it is the *whole* minus the probability of the event.

$$P(A') = 1 - P(A)$$

Complement in a Venn diagram

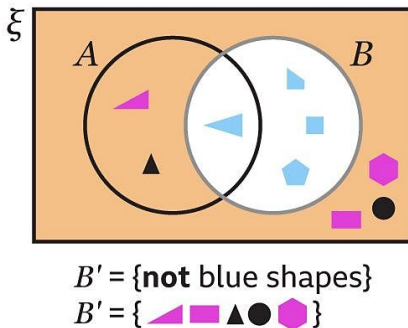


Figure 3: Complement in a Venn diagram. Source: BBC Bitesize

Computing the probability of union: mutually exclusive events and the addition rule

- To actually calculate a probability associated with the union of two events, we need to know a little bit more about the nature of the events.
- For instance, in the die example, the events of getting a 1 and getting a 2 are **mutually exclusive**
 - This means that the events *cannot happen at the same time*.
 - The probability of the union of mutually exclusive events is the sum of the probabilities of the events.

Union under mutually exclusive events

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

where A and B are mutually exclusive events.

Computing the probability of union: non-mutually exclusive events and the addition rule

- However, if the events are not mutually exclusive, we need to consider the probability of the intersection of the events.
- Events that might not be mutually exclusive happen very often. For instance, what is the probability of going to the beach and getting a sunburn?
 - These events are not mutually exclusive, as you can go to the beach and get a sunburn at the same time.

Union under non-mutually exclusive events

- If we apply the “formula” above, we will overcount the probability of the intersection of the events.
- Example: the probability of finding a company that is both in Pichincha, and is listed as active.
 - If we know the probability of finding companies in Pichincha, it will include those listed as active and otherwise.
 - If we know the probability of finding companies listed as active, it will include those in Pichincha and otherwise.
- Summing these two without doing anything else will overcount the probability of finding a company that is both in Pichincha and is listed as active.
 - Solution: subtract the probability of the intersection of the events.

Union under non-mutually exclusive events: the addition rule

- The probability of the union of two events is given by the addition rule:

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

- This is a broad formula that can be applied to any two events, whether they are mutually exclusive or not.
 - The intersection of two mutually exclusive events is 0, so the formula simplifies to the one we saw before.

Computing probabilities of intersection

- There is no direct way to calculate the probability of the intersection of two events unless we know more.
- Some cases emerge:
 - If we know the union of two events and their associated probabilities, we can calculate the probability of the intersection by rearranging the addition rule.
 - If we know the events are mutually exclusive, the probability of the intersection is 0.
 - If we know events are *independent*, there is a direct way to calculate the probability of the intersection.

Independence

- Two events are independent if the occurrence of one event does not affect the occurrence of the other event.
- For example, the probability of getting a head when flipping a coin is independent of the probability of getting a 1 when rolling a die.
 - Basically means the events are not related.
- If two events are independent, the probability of the intersection of the events is the product of the probabilities of the events.
 - This is called the multiplication rule.

Independence: the multiplication rule

- If two events are independent, the probability of the intersection of the events is given by the multiplication rule:

$$P(A \cap B) = P(A) \times P(B)$$

when A and B are independent events.

Conditional probability

- Conditional probability is the probability of an event given that another event has occurred.
- We are interested in this because once an event has occurred, the probability of another event might change.
 - This is somewhat of a Bayesian idea: once we obtain new information, our beliefs might change.
- For instance, we are interested in knowing the probability of finding an active company, *given* that the company is in Pichincha.
 - This is different from the probability of finding an active company in general.
- This is a conditional probability: the condition is that the company is in Pichincha.
 - This is denoted by $P(A|B)$, where A is the event of finding an active company and B is the event of finding a company in Pichincha.

Conditional probability formula

- The conditional probability of event A given event B is denoted by $P(A|B)$.
- The formula for conditional probability is:

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

- Following our above example:
 - A is the event of finding an active company.
 - B is the event of finding a company in Pichincha.
- We find active companies in Pichincha (the intersection of the events) and divide by the probability of finding a company in Pichincha.
- Conditionality is the theoretical probability concept of “subsetting” a sample space based on the occurrence of another event.

Conditional probability: independence

- Conditional probability is a key concept for independence: if two events are truly independent, the conditional probability of one event given the other is the same as the probability of the event (without the condition).
 - This is since conditioning on the other event does not change the probability of the event.
- Example: if we know that a die roll gives an even number, what is the probability of getting a 2 in a second roll?
 - Two die rolls are independent events (from our logical understanding of the experiment)
 - So, the probability of getting a 2 in the second roll is the same as the probability of getting a 2 any time we roll a die.

Conditional probability: independence

- The general rule is that for any two independent events A and B:

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

The multiplication rule for independent events

- If two events are independent, the probability of the intersection of the events is the product of the probabilities of the events.
 - This is the multiplication rule for independent events.
 - It will only apply if the events are independent!

$$P(A \cap B) = P(A) \times P(B)$$

The multiplication rule for independent events

- For instance, if we know the probability of getting a head when flipping a coin is 0.5, and the probability of getting a 1 when rolling a die is 1/6, we can calculate the probability of getting a head and a 1.

$$P(\text{Head} \cap 1) = P(\text{Head}) \times P(1) = 0.5 \times \frac{1}{6} = \frac{1}{12}$$

- This is a very powerful rule that can be used to calculate the probability of the intersection of independent events.
- Notice that I don't use my companies' example here, as the events are likely not independent.
 - Active companies may tend to be in Pichincha as it is the capital
 - This underscores the need to actually understand the nature of the events, or the *data generating process*.

Random variables and probability distributions

Probability distributions

- Probability distributions are a way to describe the likelihood of different outcomes in an experiment.
- This means that we can use our laws of probability to construct more sophisticated models.
 - These models will describe the likelihood of different outcomes.

Random variables

- Random variables (RVs) numerically describe the outcomes of an experiment.
 - Because they are random, the numerical value of the random variable is uncertain, hence the name.
- Two major types of random variables:
 - Discrete random variables (e.g. the number of heads when flipping a coin)
 - Continuous random variables (e.g. the height of a person)
- Random variables are denoted by capital letters, e.g. X , Y , Z .

Probability distributions of RVs

- The fact that these are random does not mean that we cannot describe them or predict their behavior.
- RVs will have a probability distribution, since numerical outcomes that they might take have associated probabilities.
- The probability distribution of a random variable is a *function* that describes the likelihood of different outcomes.
 - This function will assign a probability to each possible outcome of the random variable.

Discrete random variables

- Discrete random variables are random variables that can take on a finite number of values.
 - It is finite because we can count the number of values the random variable can take (even if it is a large number).
- For example, the number of heads when flipping a coin is a discrete random variable.
 - It can take on values 0 or 1.

The PDF of RVs (love the jargon or perish!)

- A probability density function (PDF) is a function that describes the likelihood of different outcomes of a random variable.
- Denoted by $f(x)$, where x is the value of the random variable, and the output of the function f is the probability of the random variable taking on that value.
- For discrete random variables, the PDF is often called the probability mass function (PMF), but the idea is the same.
- We won't always have a function that takes an equation form.
 - Discrete RVs will often have a table of probabilities which is called the PDF (f)
 - Continuous RVs will often have a function that describes the likelihood of different outcomes.

The Cumulative Distribution Function (CDF)

- The cumulative distribution function (CDF) is a function that describes the probability that a random variable is **less than or equal** to a certain value.
- Denoted by $F(x)$, where x is the value of the random variable, and the output of the function F is the probability that the random variable is less than or equal to x .
- The CDF is a very useful function that can be used to calculate probabilities of ranges of values of the random variable.

Continuous random variables

- There is a specific detail about the PDF of continuous random variables: the probability of the random variable taking on a specific value is 0.
- This is because continuous random variables can take on an infinite number of values, and the probability of any one value is 0.
 - Why infinite? Because we can always find a value between two values (i.e. with a lot of decimals).
- Intuitively, think that whenever we have a continuous random variable, it is extremely unlikely that we will get a specific value.
 - Measurements are never perfect. We can never measure a person's height to the exact millimeter, hence, we can't even know the exact value of the random variable.

Laboratorio de Investigación para el Desarrollo del Ecuador