Chapter 3
Probability

Objectives

After completing this chapter, you will be able to:

- To Understand what Probability is
- To perform simple Probability calculations
- Understand the limitations of formulas used in Probability
- Know when more complex Probability methods are required
- Summary statistics for continuous and discrete data
- Different formula to use on different scenarios in Probability

Introduction

The world is an uncertain place. Making predictions about something as seemingly mundane as tomorrow's weather, for example, is actually quite a difficult task. Even with the most advanced computers and models of the modern era, weather forecasters still cannot say with absolute certainty whether it will rain tomorrow. The best they can do is to report their best estimate of the chance that it will rain tomorrow. For example, if the forecasters are fairly confident that it will rain tomorrow, they might say that there is a 90% chance of rain. You have probably heard statements like this your entire life, but have you ever asked yourself what exactly it means to say that there is a 90% chance of rain?

Let us consider an even more basic example: tossing a coin. If the coin is fair, then it is just as likely to come up heads as it is to come up tails. In other words, if we were to repeatedly toss the coin many times, we would expect about about half of the tosses to be heads and and half to be tails. In this case, we say that the probability of getting a head is 1/2 or 0.5.

Note that when we say the probability of a head is 1/2, we are not claiming that any sequence of coin tosses will consist of exactly 50% heads. If we toss a fair coin ten times, it would not be surprising to observe 6 heads and 4 tails, or even 3 heads and 7 tails. But as we continue to toss the coin over and over again, we expect the long-run frequency of heads to get ever closer to 50%. In general, it is important in statistics to understand the distinction between theoretical and empirical quantities. Here, the true (theoretical) probability of a head was 1/2, but any realized (empirical) sequence of coin tosses may have more or less than exactly 50% heads.

Now suppose instead that we were to toss an unusual coin with heads on both of its faces. Then every time we flip this coin we will observe a head — we say that the probability of a head is 1. The probability of a tail, on the other hand, is 0. Note that there is no way we can further modify the coin to make flipping a head even more likely. Thus, a probability is always a number between 0 and 1 inclusive.

First Concepts

Terminology

When we later discuss examples that are more complicated than flipping a coin, it will be useful to have an established vocabulary for working with probabilities. A probabilistic experiment (such as tossing a coin or rolling a die) has several components. The sample space is the set of all possible outcomes in the experiment. We usually denote the sample space by Ω, the Greek capital letter "Omega." So in a coin toss experiment, the sample space is

$$\Omega = \{\mathrm{H}, \mathrm{T}\},$$

since there are only two possible outcomes: heads (H) or tails (T). Different experiments have different sample spaces. So if we instead consider an experiment in which we roll a standard six-sided die, the sample space is

$$\Omega = \{1,2,3,4,5,6\}.$$

Collections of outcomes in the sample space Ω are called events, and we often use capital Roman letters to denote these collections. We might be interested in the event that we roll an even number, for example. If we call this event E, then

$$E = \{2,4,6\}.$$

Any subset of Ω is a valid event. In particular, one-element subsets are allowed, so we can speak of the event F of rolling a 4, F = {4}.

Assigning probabilities to dice rolls and coin flips

In a random experiment, every event gets assigned a probability. Notationally, if A is some event of interest, then P(A) is the probability that A occurs. The probabilities in an experiment are not arbitrary; they must satisfy a set of rules or axioms. We first require that all probabilities be nonnegative. In other words, in an experiment with sample space Ω, it must be the case that

$$P(A) \geq 0$$

for any event $A \subseteq \Omega$. This should make sense given that we've already said that a probability of 0 is assigned to an impossible event, basic probability 7 and there is no way for something to be less likely than something that is impossible!
The next axiom is that the sum of the probabilities of all the outcomes in Ω must be 1. We can restate this requirement by the equation

$$\sum_{\omega \in \Omega} P(\omega) = 1.$$

This rule can sometimes be used to deduce the probability of an outcome in certain experiments. Consider an experiment in which we roll a fair die, for example. Then each outcome (i.e. each face of the die) is equally likely. That is,

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = a,$$

for some number a. Equation (2) now allows us to conclude

$$1 = \sum_{k=1}^{6} P(k) = \sum_{k=1}^{6} a = 6a,$$

so a = 1/6. In this example, we were able to use the symmetry of the experiment along with one of the probability axioms to determine the probability of rolling any number. Once we know the probabilties of the outcomes in an experiment, we can compute the probability of any event. This is because the probability of an event is the sum of the probabilities of the outcomes it comprises. In other words, for an event $A \subseteq \Omega$, the probability of A is

$$P(A) = \sum_{\omega \in A} P(\omega).$$

To illustrate this equation, let us find the probability of rolling an even number, an event which we will denote by E. Since E = {2, 4, 6}, we simply add the probabilities of these three outcomes to obtain

$$\begin{aligned} P(E) &= \sum_{\omega \in E} P(\omega) \\ &= P(2) + P(4) + P(6) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \\ &= \frac{1}{2}. \end{aligned}$$

What is the probability that we get at least one H? Solution. One way to solve this problem is to add up the probabilities of all outcomes that have at least one H. We would get

$$\begin{aligned} P(\text{flip at least one } \mathtt{H}) &= P(\mathtt{HH}) + P(\mathtt{HT}) + P(\mathtt{TH}) \\ &= p^2 + p \cdot (1-p) + (1-p) \cdot p \\ &= p^2 + 2 \cdot (p - p^2) \\ &= 2p - p^2 \\ &= p \cdot (2-p). \end{aligned}$$

Another way to do this is to find the probability that we don't flip at least one H, and subtract that probability from 1. This would give us the probability that we do flip at least one H. The only outcome in which we don't flip at least one H is if we flip T both times. We would then compute

$$P(\text{don't flip at least one } \mathtt{H}) = P(\mathtt{TT}) = (1-p)^2$$

Then to get the complement of this event, i.e. the event where we do flip at least one H, we subtract the above probability from 1. This gives us

$$
\begin{aligned}
P(\text{flip at least one } \mathtt{H}) &= 1 - P(\text{don't flip at least one } \mathtt{H}) \\
&= 1 - (1-p)^2 \\
&= 1 - (1 - 2p + p^2) \\
&= 2p - p^2 \\
&= p \cdot (2 - p).
\end{aligned}
$$

Wowee! Both methods for solving this problem gave the same answer. Notice that in the second calculation, we had to sum up fewer probabilities to get the answer. It can often be the case that computing the probability of the complement of an event and subtracting that from 1 to find the probability of the original event requires less work.

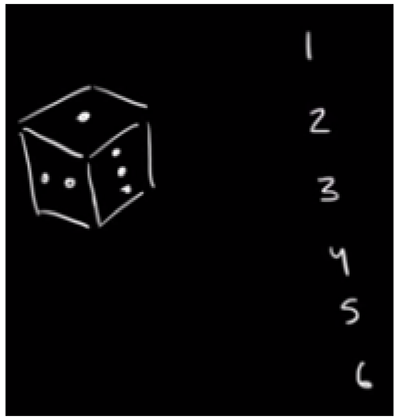
Probability is simply how likely something is to happen.
Whenever we're unsure about the outcome of an event, we can talk about the probabilities of certain outcomes—how likely they are. The analysis of events governed by probability is called statistics.
The best example for understanding probability is flipping a coin:
There are two possible outcomes—heads or tails.
What's the probability of the coin landing on Heads? We can find out using the equation P(H) = ?*P*(*H*)=?P, left parenthesis, H, right parenthesis, equals, question mark.You might intuitively know that the likelihood is half/half, or 50%. But how do we work that out? Probability =

$$\frac{\#\text{ of possibilities that meet my condition}}{\#\text{ of equally likely possibilities}}$$

Formula for calculating the probability of certain outcomes for an event
In this case:

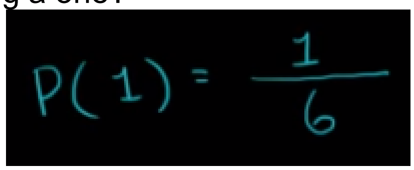
$$P(H) = \frac{1}{2} = 50\%$$

Probability of a coin landing on heads
Probability of an event = (# of ways it can happen) / (total number of outcomes)
P(A) = (# of ways A can happen) / (Total number of outcomes)
Example 1
There are six different outcomes.



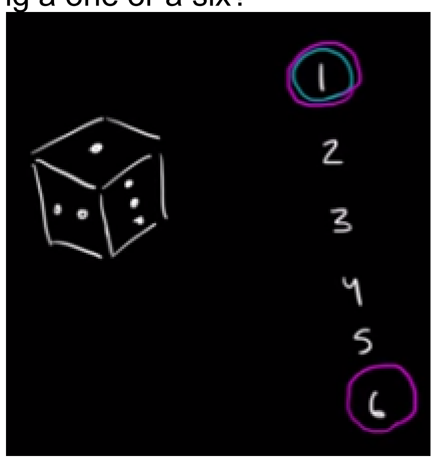
Different outcomes rolling a die

What's the probability of rolling a one?



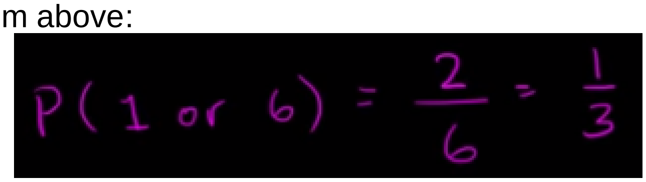
Probability formula for rolling a '1' on a die

What's the probability of rolling a one or a six?



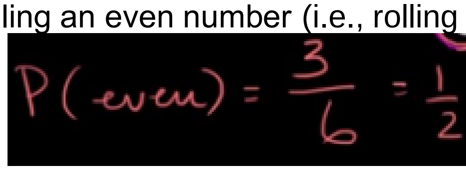
Probability of a 1 or a 6 outcome when rolling a die

Using the formula from above:



Probability formula applied

What's the probability of rolling an even number (i.e., rolling a two, four or a six)?



Probability of rolling an even number? The formula and solution

Tips

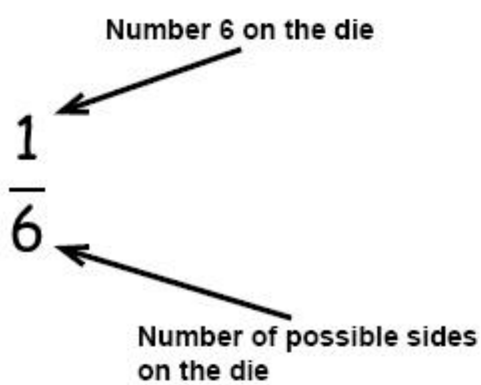
- The probability of an event can only be between 0 and 1 and can also be written as a percentage.
- The probability of event AAA is often written as P(A)P(A)P, left parenthesis, A, right parenthesis.
- If P(A) > P(B)P(A)>P(B)P, left parenthesis, A, right parenthesis, is greater than, P, left parenthesis, B, right parenthesis, then event AAA has a higher chance of occurring than event BBB.
- If P(A) = P(B)P(A)=P(B)P, left parenthesis, A, right parenthesis, equals, P, left parenthesis, B, right parenthesis, then events AAA and BBB are equally likely to occur.

Probability of events

Probability = The number of wanted outcomes Thenumber of possible outcomes

Example

What is the probability to get a 6 when you roll a die?
A die has 6 sides, 1 side contain the number 6 that give us 1 wanted outcome in 6 possible outcomes.



Independent events: Two events are independent when the outcome of the first event does not influence the outcome of the second event.
When we determine the probability of two independent events we multiply the probability of the first event by the probability of the second event.

P(XandY)=P(X)·P(Y)P(XandY)=P(X)·P(Y)

To find the probability of an independent event we are using this rule:

Example

If one has three dice what is the probability of getting three 4s?
The probability of getting a 4 on one die is 1/6
The probability of getting 3 4s is:

P(4and4and4)=16·16·16=1216P(4and4and4)=16·16·16=1216

When the outcome affects the second outcome, which is what we called dependent events.

Dependent events: Two events are dependent when the outcome of the first event influences the outcome of the second event. The probability of two dependent events is the product of the probability of X and the probability of Y **AFTER** X occurs.

P(XandY)=P(X)·P(Yafterx)P(XandY)=P(X)·P(Yafterx)

Example

What is the probability for you to choose two red cards in a deck of cards?
A deck of cards has 26 black and 26 red cards. The probability of choosing a red card randomly is:

P(red)=2652=12P(red)=2652=12

The probability of choosing a second red card from the deck is now:

P(red)=2551P(red)=2551

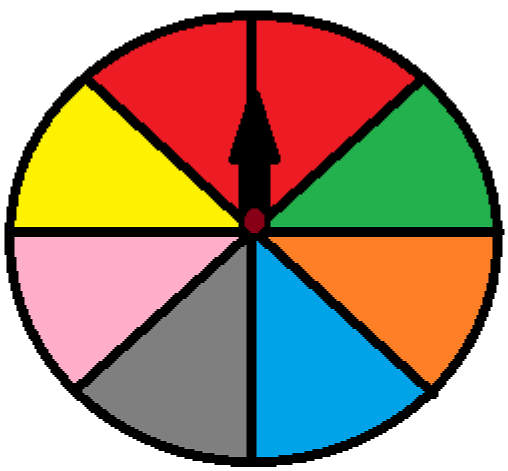
The probability:

P(2red)=12·2551=25102P(2red)=12·2551=25102

Two events are mutually exclusive when two events cannot happen at the same time. The probability that one of the mutually exclusive events occur is the sum of their individual probabilities.

P(XorY)=P(X)+P(Y)P(XorY)=P(X)+P(Y)

An example of two mutually exclusive events is a wheel of fortune. Let's say you win a bar of chocolate if you end up in a red or a pink field.



What is the probability that the wheel stops at red or pink?

P(red or pink)=P(red)+P(pink)

P(red)=28=14P(red)=28=14

P(pink)=18P(pink)=18

P(redorpink)=18+28=38P(redorpink)=18+28=38

Inclusive events are events that can happen at the same time. To find the probability of an inclusive event we first add the probabilities of the individual events and then subtract the probability of the two events happening at the same time.

P(XorY)=P(X)+P(Y)−P(XandY)P(XorY)=P(X)+P(Y)−P(XandY)

Example

What is the probability of drawing a black card or a ten in a deck of cards?

There are 4 tens in a deck of cards P(10) = 4/52
There are 26 black cards P(black) = 26/52
There are 2 black tens P(black and 10) = 2/52

P(blackorten)=452+2652−252=3052−252=2852=713

Probability

Probability is the branch of mathematics that studies the possible outcomes of given events together with the outcomes' relative likelihoods and distributions. In common usage, the word "probability" is used to mean the chance that a particular event (or set of events) will occur expressed on a linear scale from 0 (impossibility) to 1 (certainty), also expressed as a percentage between 0 and 100%. The analysis of events governed by probability is called statistics.

There are several competing interpretations of the actual "meaning" of probabilities. Frequentists view probability simply as a measure of the frequency of outcomes (the more conventional interpretation), while Bayesians treat probability more subjectively as a statistical procedure that endeavors to estimate parameters of an underlying distribution based on the observed distribution.

A properly normalized function that assigns a probability "density" to each possible outcome within some interval is called a probability density function (or probability distribution function), and its cumulative value (integral for a continuous distribution or sum for a discrete distribution) is called a distribution function (or cumulative distribution function).

A variate is defined as the set of all random variables that obey a given probabilistic law. It is common practice to denote a variate with a capital letter (most commonly X). The set of all values that X can take is then called the range, denoted $\mathbb{R}_X$ (Evans *et al.* 2000, p. 5). Specific elements in the range of X are called quantiles and denoted x, and the probability that a variate X assumes the element x is denoted $P(X=x)$.

Probabilities are defined to obey certain assumptions, called the probability axioms. Let a sample space contain the union ($\bigcup$) of all possible events E_i, so

$$S \equiv \left(\bigcup_{i=1}^{N} E_i\right), \qquad (1)$$

and let E and F denote subsets of S. Further, let $F' = \text{not} - F$ be the complement of F, so that

$$F \bigcup F' = S. \qquad (2)$$

Then the set E can be written as

$$E = E \bigcap S = E \bigcap (F \bigcup F') = (E \bigcap F) \bigcup (E \bigcap F'), \qquad (3)$$

where $\bigcap$ denotes the intersection. Then

$$P(E) = P(E \bigcap F) + P(E \bigcap F') - P[(E \bigcap F) \bigcap (E \bigcap F')] \qquad (4)$$
$$= P(E \bigcap F) + P(E \bigcap F') - P[(F \bigcap F') \bigcap (E \bigcap E)] \qquad (5)$$
$$= P(E \bigcap F) + P(E \bigcap F') - P(\emptyset \bigcap E) \qquad (6)$$
$$= P(E \bigcap F) + P(E \bigcap F') - P(\emptyset) \qquad (7)$$
$$= P(E \bigcap F) + P(E \bigcap F'), \qquad (8)$$

where $\emptyset$ is the empty set.

Let $P(E \mid F)$ denote the conditional probability of E given that F has already occurred, then

$$P(E) = P(E \mid F)P(F) + P(E \mid F')P(F') \tag{9}$$

$$= P(E \mid F)P(F) + P(E \mid F')[1 - P(F)] \tag{10}$$

$$P(A \cap B) = P(A)P(B \mid A) \tag{11}$$

$$= P(B)P(A \mid B) \tag{12}$$

$$P(A' \cap B) = P(A')P(B \mid A') \tag{13}$$

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}. \tag{14}$$

The relationship

$$P(A \cap B) = P(A)P(B) \tag{15}$$

holds if A and B are independent events. A very important result states that

$$P(E \cup F) = P(E) + P(F) - P(E \cap F), \tag{16}$$

which can be generalized to

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i P(A_i) - \sum_{ij}' P\left(A_i \cap A_j\right) + \sum_{ijk}'' P\left(A_i \cap A_j \cap A_k\right) - \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right). \tag{17}$$

Lecture Notes

Below are listed the terms that usually used in Statistics and Analysis and definition.

Definitions

Average - The average is a number that is one way to find the typical value of a set of numbers. You find the average by adding up all the numbers and then dividing the total by the number of numbers in the set.

Example:
To find the average of the data set (1, 3, 3, 4, 4, 5, 8)
Add all the values together 1+3+3+4+4+5+8 = 28
Then divide by the total number of values 28 ÷ 7 = 4
The average value is 4.

Correlation - A measurement of how closely related two variables are.

Dependent event - Events are dependent if the occurrence of either event affects the probability of the occurrence of the other event. In other words, one event depends on the other.

Event - A collection of outcomes from an experiment.

Extrapolate - Extrapolation is a way to estimate values beyond the known data. You can use patterns and graphs to determine other possible data points that were not actually measured.

Frequency - The frequency is how often an event occurs during a specific amount of time.

Interpolate - Interpolation is a way to estimate data. When you interpolate you estimate the data between two known points on a graph. This can be done by drawing a curve or line between the two points.

Interval - The set of numbers between two other numbers in a data set. It often refers to a period of time between two events.

Percent - A percent is a special type of fraction where the denominator is 100. It can be written using the % sign.

Example: 50%, this is the same as ½ or 50/100

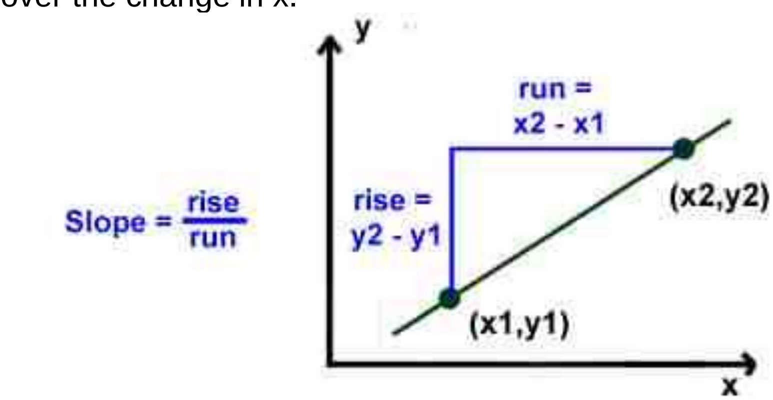
Probability - The probability is the chance that an event will or will not occur.

Random - If something is random, then all possible events have an equal chance of occurring.

Range - The range is the difference between the largest number and the smallest number in a data set.

Example: The range of the data set (2, 2, 7, 8, 12, 7, 2, 14) is 14 -2 = 12.

Ratio - A ratio is a comparison of two numbers. It can be written a few different ways.
Example: The following are all a way to write the same ratio: 1/2 , 1:2, 1 of 2
Slope - A number that indicates the incline or steepness of a line on a graph. Slope equals the "rise" over the "run" on a graph. This can also be written as the change in y over the change in x.



Example: If two points on a line are (x1, y1) and (x2, y2), then the slope = (y2 - y1) ÷ (x2-x1).

Video Links:

Probability

- https://www.youtube.com/watch?v=yUaI0JriZtY&vl=en
- https://www.youtube.com/watch?v=KzfWUEJjG18
- https://www.youtube.com/watch?v=uzkc-qNVoOk

References

- https://seeing-theory.brown.edu/doc/basic-probability.pdf
- https://mathworld.wolfram.com/Probability.html
- https://www.ducksters.com/kidsmath/probability_statistics_glossary.php#:~:text=Probability%20%2D%20The%20probability%20is%20the,number%20in%20a%20data%20set.
- https://www.khanacademy.org/math/statistics-probability/probability-library/basic-theoretical-probability/a/probability-the-basics#:~:text=Probability%20is%20simply%20how%20likely,by%20probability%20is%20called%20statistics.