Probability

Owen G. Ward

2021-05-05

Introduction

Probability is the key component of all statistical concepts. It is probability which is used to develop statistical tests and, more generally, to answer questions where there is uncertainty.

Suppose you and a friend both roll a die once. What is the chance you get a number larger than your friend?

Suppose you roll a die and get a 6. What is the chance you will get a six if you roll the die again?

These are uncertain outcomes. We know you will get some number between 1 and 6 if you roll the die again, but you can't say which.

Probability gives you a way to quantify how likely each of the possible outcomes actually are.

Historic Foundations

While these dice examples might seem irrelevant, the field of probability originally arose from gamblers trying to maximize their profits on problems like this.

In the mid 1600's, a French nobleman (Chevalier De Mere) asked Blaise Pascal to investigate how he should bet on a popular game. He wanted to know whether the chance of rolling a 6 in 4 rolls of a die was higher than getting a pair of 6's in 24 rolls of a pair of die.

We'll be able to answer this problem shortly.

We want to quantify the chance of specific outcomes for some activity where the outcome is uncertain.

If we toss a coin 3 times, we don't know in advance how many heads we will get, or what side of a die will land face up.

We call an experiment a process where there outcome is uncertain or random.

An experiment has a list of possible outcomes. If we roll a die, we can see any of the numbers 1,2,3,4,5,6. This is the sample space of the experiment.

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

An event is any subset of this sample space. For example, rolling a 2 is an event. So is rolling an odd number. Or rolling a number greater than 4.

If there is only one outcome in an event it is simple event. Otherwise it is called composite.

A **random variable** X is any function/rule which associates a number to each outcome of an experiment.

We can use set notation to describe these events. If A is the event of rolling an odd number then $A = \{1, 3, 5\}$.

We want to be able to assign a number, $P(A)$, which gives the **probability** of any event A.

Axioms of Probability

We have several fundamental properties for probability.

- $P(\Omega) = 1$.
- $P(\emptyset) = 0$.
- For any event A we have $0 \leq P(A) \leq 1$.

We can think of the probability as the proportion of times an event would occur if you observed the random process an infinite number of times.

If we rolled a fair die forever, we would roll an odd number half the time.

For the initial examples we will see, we will generally assume that each outcome in an experiment is equally likely. Each side of a die is as likely as any other.

This gives $P(A) = \frac{|A|}{|\Omega|}$.

In that case, we can compute probabilities by simply counting the outcomes in an event and dividing them by the number of all possible outcomes.

For $A = \{1, 2, 3\}$ and $B = \{2, 4\}$ that means $P(A) = 1/2$ and $P(B) = 1/3$.

Combining Events

We can think about the probability of multiple events happening.

If $A = \{1, 2, 3\}$ and $B = \{2, 4\}$ we can think about $P(A \cup B)$ and $P(A \cap B)$.

Similarly, we can also think about $P(A^c)$, the complement of $P(A)$

If events A and B are **disjoint** or **mutually exclusive** then they cannot happen at the same time. We say that $A \cap B = \emptyset$

If $A = \{1\}$ and $B = \{2\}$, you can't roll a die and get both of these outcomes.

In this case we have

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

If A and B are not mutually exclusive we can still compute $P(A \cup B)$, we just need to account for their overlap,

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

Thinking again about set theory, we can use this to compute the probability of the **complement** of an event.

We have

$$A^c = \Omega \backslash A$$

so in fact

$$P(A^c) = P(\Omega \backslash A) = P(\Omega) - P(A)$$

which means

$$P(A^c) = 1 - P(A).$$

Will see an example using this in a little while.

Examples

Suppose we draw a single card from a standard deck of playing cards. What is the probability we draw an Ace?

What is the probability we draw a diamond?

What is the probability we draw an ace **or** a diamond?

Examples

You roll a fair dice twice. What is the probability that

- The sum of the two rolls is 1?
- The sum of the two rolls is 5?
- The sum of the two rolls is 12?

Counting

A slightly more complicated version of these problems is where it can be challenging to even count the possible outcomes.

Thankfully, there are several tools available for this.

Multiplication rule

Suppose you have a password which consists of 4 single digit numbers. How many possible different combinations are there?

We can have any of the 10 digits 0-9 for the first number, and the second, and so on. So for each number there are 10 possible choices.

We can get the first two digits $10 \times 10 = 100$ different ways (01-99 and 00).

So we have $10 \times 10 \times 10 \times 10 = 10000$ different possible combinations.

Product Rule

If we can choose X n ways and choose Y m ways then we can choose pairs of X, Y in nm ways.

Suppose there are 5 plumbers and 2 electricians in your area and you need one of each. How many different combinations can you choose? What is the probability you end up with a specific pair?

We can choose the plumber 5 different ways, and for each plumber there are 2 choices for the electrician. So in total there are $5 \times 2 = 10$ possible choices.

For problems like this it is important to note that how you choose the plumber has nothing to do with how you choose the electrician. They are not related.

Independence

The probability you pick a specific plumber has nothing to do with how you choose an electrician. They are **independent**.

Suppose the plumbers are $\{P_1, P_2, P_3, P_4, P_5\}$ and the electricians are $\{E_1, E_2\}$. What is the probability you pick P_3 and E_2?

$$P(P_3 \cap E_2) = P(P_3)P(E_2),$$

where $P(P_3) = 1/5$ and $P(E_2) = 1/2$.

Example

In a multiple choice exam, there are 5 questions and 4 choices for each question (a, b, c, d). Bob has not studied for the exam at all and decides to randomly guess the answers. What is the probability that

- The first question he gets right is the 5th
- He gets all the questions right
- He gets at least one question right?

We can now solve the original problem.

> Whether the chance of rolling at least one 6 in 4 rolls of a die was higher than getting at least a pair of 6's in 24 rolls of a pair of die.

We first do the probability of at least one 6. Either we can get 1 (or more) 6's, or no 6's in 4 rolls.

So we can use the complement probability from above.

$$P(\text{At least 1 6 in 4 rolls}) = 1 - P(\text{no 6 in 4 rolls}).$$

The probability of getting no 6 in 4 rolls is the probability we don't get one in the first, and don't get one in the second, and the third, and the fourth.

Each roll, this probability is 5/6.

So the probability of no 6 in 4 rolls is

$$\frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = \left(\frac{5}{6}\right)^4$$

Therefore the probability of at least 1 6 is

$$1 - \left(\frac{5}{6}\right)^4.$$

Similarly, for at least one pair of 6's in 24 rolls of a pair of dice. The probability of getting a pair of 6's when we roll a pair is 1/36.

So the probability of no pair of 6's is 35/36.

The probability of at least one pair in 24 tries is 1 minus the probability we don't get a pair of 6's in any of the 24 tries.

This is

$$1 - \left(\frac{35}{36}\right)^{24}.$$

It turns out that the first outcome has a probability just above 0.5 while the second is slightly below 0.5

Permutations

Consider there is some prize in the local newspaper for the top 2 local plumbers. How many ways can you choose the first and second ranked plumber, if there are 5 in the area?

For the top plumber, if each is equally likely to win, there are five different choices.

After picking the top plumber, each of the four remaining are equally likely to be second. So there are 4 choices.

By the multiplication rule we can choose the first and second in $5 \times 4 = 20$ different ways.

We can actually write this as

$$20 = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1} = {}^5P_2,$$

where

$${}^nP_k = \frac{n!}{(n-k)!},$$

where

$$n! = n \cdot (n-1) \cdot (n-2) \cdot ... \cdot 2 \cdot 1.$$

We use ${}^nP_k$ when we want to choose k items from n possible choices and **the order matters.**

How many 4 number passwords can you choose when you cannot repeat a digit?

We can do this using permutations, to get ${}^{10}P_4 = \frac{10!}{6!}$.

Here we want to make sure we count both 3456 and 4356 as separate passwords.

If the order doesn't matter we do something else.

Example

Suppose a Spotify playlist contains 100 songs, 10 of which are by The Beatles. If the shuffling is truly random, what is the probability the first Beatles song is the fifth song played?

Combinations

Suppose we are playing a card game where a hand has five cards. Then the order of the cards doesn't matter.

As such, we want to count each possible ordering of the same five cards as only one hand.

If there are 5 cards in a hand there are

$$5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5!$$

different orderings, but they all result in the same hand.

So we can count the number of **unordered** combinations by counting the number of ordered combinations and dividing by the number of different orderings.

This gives us the formula

$${}^nC_k = \binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{{}^nP_k}{k!}$$

Note that

$$\binom{n}{k} = \binom{n}{n-k}$$

We'll see some examples of how to use this.

Suppose there is a committee being formed between an Econ department and a History department, which will contain 3 professors from each.

If there are 6 Econ professors and 4 History professors, how many committees can be formed?

We can pick the Econ members in $\binom{6}{3}$ ways and the History members in $\binom{4}{3}$ ways.

So there are $\binom{6}{3}\binom{4}{3}$ possible committees which can be formed.

Suppose you draw a hand of 5 cards from a deck of 52 cards. What is the probability your hand contains a red Ace and 4 Spades?

We first need to figure out the number of possible hands. We want any 5 cards from 52, which can be done in $\binom{52}{5}$ ways.

For the red Ace, we want one of the two possible cards, which is $\binom{2}{1}$. For the 4 spades, we want 4 of the 13 spades, which is $\binom{13}{4}$.

So the probability of such a hand is given by

$$\frac{\binom{2}{1}\binom{13}{4}}{\binom{52}{5}}.$$

Recap

Conditional Probability

Conditional probability is when we want to compute the probability of an event, but we already know "something" out the outcome.

> Suppose I roll a die and tell you the outcome was an odd number. What is the probability it was a 1?
>
> Suppose I roll 2 dice and the sum of their rolls was 2. What is the probability the two individual dice both came up 1?

Let A and B be two events in the sample space and $P(B) \neq 0$.

Then, we define the **conditional probability** of A, **given** B as

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

Let A be the event you roll a 1 and B be the event you roll an odd number.

Then $P(A \cap B)$ is the probability you roll an odd number and you roll a 1, which is 1/6.

$P(B)$ is 1/2, the probability you roll an odd number.

So $P(A|B) = \frac{1/6}{1/2} = \frac{1}{3}$, which you would expect.

Suppose you flip a coin 3 times and get 2 heads. What is the probability the first flip was a head?

There are 8 possible outcomes from 3 coin flips, 3 of which give 2 heads.

Of those 3 outcomes, 2 include the first flip being a head.

So $P(A \cap B) = 2/8$ and $P(B) = 3/8$ giving

$$P(A|B) = \frac{2/8}{3/8} = \frac{2}{3},$$

which seems reasonable?

Going further

We have

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

Similarly, we can replace the order of A, B to get

$$P(A \cap B) = P(B \cap A) = P(B|A)P(A).$$

Putting these together gives

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

or, rearranging,

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

We will see problems where we will have the numerators of the right hand side. But we need to deal with the denominator.

Law of Total Probability

Let's assume that A is a binary event, so that $P(A) + P(A^c) = 1$.

We actually want to be able to write $P(B)$ in terms of conditional probabilities given A, A^c that we know.

We have

$$B = (B \cap A) \cup (B \cap A^c)\,,$$

where these two sets are disjoint.

That means we can write

$$P(B) = P(B \cap A) + P(B \cap A^c).$$

Then, for these two probabilities, we can use conditional probability to write

$$P(B \cap A) = P(B|A)P(A),\ P(B \cap A^c) = P(B|A^c)P(A^c).$$

Together these give

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c).$$

Bayes Rule

We go all the way back and use these in

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

We get

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}.$$

This is known as **Bayes Rule**.

This can be done more generally, for example a partition over 3 disjoint events which cover the whole sample space. If A_1, A_2, A_3 are disjoint with $A_1 \cup A_2 \cup A_3 = \Omega$ then

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B).$$

Can similarly compute $P(A_1|B)$.

Note that we have A and B are **independent** if and only if

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

Then

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

Example

Suppose you throw 2 dice with outcomes X and Y. Let A be the event that X is odd, B the event that $X + Y$ is odd. Are A and B independent?

We need to compute $P(A), P(B)$ and see if the product equals $P(A \cap B)$.

$P(A) = 1/2$ and $P(A \cap B) = 1/4$.

Compute $P(B)$ using Law of Total Probability.

Putting this together we see that.

Example

Suppose you throw 2 dice with outcomes X and Y. Let A be the event that X is odd, B the event that XY is odd. Are A and B independent?

If XY is odd then X must be odd (odd times odd equals odd).

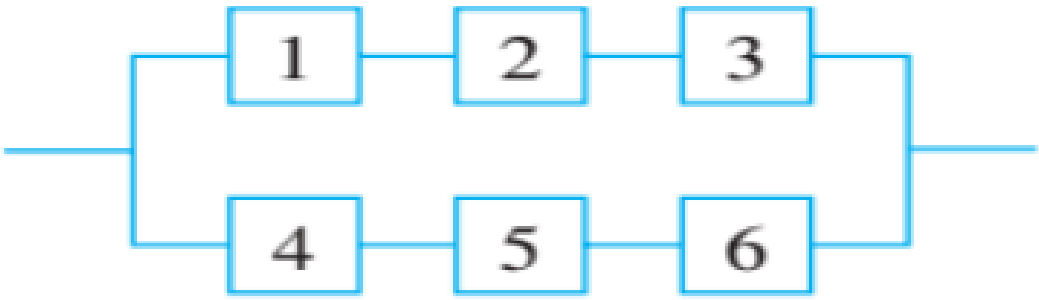
So if B happens A must happen, which means $B \subset A$.

This means $A \cap B = B$.

Which means $P(A \cap B) = P(B) \neq P(A)P(B)$ because that would require $P(B) = 0$ or $P(A) = 1$.

Example

Suppose you have the following electrical system, where the probability of each component working is 0.9. What is the probability electricity can flow through the system?



Disease Testing

Bayes rule is extremely useful, and is particularly useful in the context of disease testing.

If you have a test for a disease, two things can go wrong:

- The test says you have the disease when you don't.
- The test says you don't have the disease when you do.

Both of these cause issues, in different ways. We can get around some of these using Bayes rule.

Suppose you have a test which tells you whether you have a disease or not. Given that you test positive, what is the probability you actually have the disease?

To do this, we need to know

- The sensitivity of the test
- The specificity of the test
- The true underlying proportion of the population who have the disease.

Will explain these in a moment.

Let D be the event you have the disease and Pos the event you test positive. Then what we want to compute is

$$P(D|Pos).$$

By Bayes rule, we can write

$$P(D|Pos) = \frac{P(D \cap Pos)}{P(Pos)} = \frac{P(Pos|D)P(D)}{P(Pos|D)P(D) + P(Pos|D^c)P(D^c)}.$$

To solve this we need to know

- $P(D)$, the underlying proportion of people who have the disease. This gives $P(D^c) = 1 - P(D)$.
- $P(Pos|D)$, the probability you test positive if you do have the disease.
- $P(Pos|D^c)$, the probability you test positive if you do not have the disease.

$P(Pos|D)$ is the **sensitivity** of the test while $1 - P(Pos|D^c)$ is the **specificity** of the test.

Sensitivity and Specificity

- Sensitivity is the probability the test is positive for the disease, if you truly do have it.
- Obviously, we want this to be as high as possible.
- Specificity is the probability the test correctly gives you a negative result when you don't have the disease.
- We also want this to be high. But it maybe isn't as important as the sensitivity.

- Suppose we have a test for some non serious disease, which 1 in 20 people have (so $P(D) = 0.05$)
- Let the sensitivity be 90%. So if you have the disease and get tested 10 times, you expect to be positive 9 times.
- Let the specificity be 80%. So, if you don't have the disease and you get tested 10 times, you expect that 2 of the times you will have a positive test.
- Suppose you test positive. What is the probability you have the disease?

We have

$$P(D|Pos) = \frac{P(Pos|D)P(D)}{P(Pos|D)P(D) + P(P|D^c)P(D^c)}$$

where

- $P(Pos|D) = 0.9$
- $P(Pos|D^c) = 0.2$
- $P(D) = 0.05$.

This gives

$$P(D|Pos) = \frac{(0.9)(0.05)}{(0.9)(0.05) + (0.2)(0.95)} \approx 0.19.$$

So if you test positive, you maybe shouldn't worry too much.

An example

- If we test 1000 people then we expect 50 of them to have the disease. We will correctly get positive tests for ≈ 45 of them.
- Of the 950 who do not have the disease we expect 2 in 10 false positives, giving 190 positives.
- In total, we would expect 235 positive tests, but only 45 of them will have the disease.
- Is this good? Is this good enough?

Increasing the specificity

If we increase the specificity to 0.95 then we can repeat these calculations and we get

- $P(D|P) \approx 0.49$
- If we test 1000 people now we will only expect to get 48 false positives.

What this means for Covid testing

- In a much publicised Santa Clara study done at Stanford in early 2020, they test 3330 people.
- They get 50 positives, so raw $P(D) \approx 0.015$.
- They estimate the prevalence, after reweighting, is ≈ 0.03.
- They estimate the sensitivity is between 84% and 97%.
- They estimate the specificity is between 90% and 100%.

Covid Antibody Testing

Here, when we are looking at a rare disease, the specificity is particularly important. We saw in the example that most positives were false positives, unless the specificity is high.

- If the specificity is 98.5% and there are no true Covid 19 cases in the data, you would expect 50 false positives.
- It is difficult to know the true specificity, so any uncertainty in its value makes it plausible that the true disease prevalence is actually 0.