

Induction

Induction is an extremely powerful tool in mathematics. It is a way of proving propositions that hold for all natural numbers, such as these:

- $\forall k \in \mathbb{N} . 0 + 1 + 2 + 3 + \dots + k = k(k + 1)/2$.
- For every $k \in \mathbb{N}$, the sum of the first k odd numbers is a perfect square.

Each of these propositions is of the form $\forall k \in \mathbb{N} . P(k)$. For example, in the first proposition, $P(k)$ is the statement $0 + 1 + \dots + k = k(k + 1)/2$, $P(0)$ says $0 = 0 \cdot (0 + 1)/2$, $P(1)$ says $0 + 1 = 1 \cdot (1 + 1)/2$, etc. The *principle of induction* asserts that you can prove $P(k)$ is true for all $k \in \mathbb{N}$ by proving two facts:

Base case: Prove that $P(0)$ is true.

Induction step: Prove that $P(k) \implies P(k + 1)$ is true for all $k \in \mathbb{N}$.

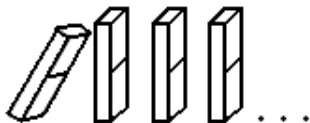
The implication $P(k) \implies P(k + 1)$ is in turn normally proven by assuming that $P(k)$ is true and showing that $P(k + 1)$ follows as a consequence. So, in practice, induction proofs follow these three steps:

Base case: Prove that $P(0)$ is true.

Inductive hypothesis: Assume that $P(k)$ is true, where k is arbitrary.

Inductive step: Prove that $P(k + 1)$ follows from the inductive hypothesis.

The principle of induction formally says that if $P(0)$ and $\forall n \in \mathbb{N} . (P(n) \implies P(n + 1))$ are both true, then $\forall n \in \mathbb{N} . P(n)$ is true. Intuitively, the base case says that $P(0)$ holds, while the inductive step says that $P(0) \implies P(1)$ and $P(1) \implies P(2)$ and $P(2) \implies P(3)$ and so on. Intuitively, it makes sense that this is enough to prove that $P(n)$ is true for every n : from $P(0)$ and $P(0) \implies P(1)$, we can deduce $P(1)$; from $P(1)$ and $P(1) \implies P(2)$, we can deduce $P(2)$; and so on. The principle of induction says that this domino effect eventually shows that $P(n)$ is true for every $n \in \mathbb{N}$. In fact, dominoes are a wonderful analogy: we have a domino for each proposition $P(k)$. The dominoes are lined up so that if the k^{th} domino is knocked over, then it in turn knocks over the $k + 1^{\text{st}}$. Knocking over the k^{th} domino corresponds to proving $P(k)$ is true. So the induction step corresponds to the statement that the k^{th} domino knocks over the $k + 1^{\text{st}}$ domino, for every k , or in other words, that there are no gaps in the line of dominoes. Now, if we knock over the first domino (the one numbered 0), and if there are no gaps in the line of dominoes, then this sets off a chain reaction that knocks down all the dominoes.



Let's see some examples.

Theorem: $\forall k \in \mathbb{N}. \sum_{i=0}^k i = \frac{k(k+1)}{2}$.

Proof: (by induction on k)

Base case: $P(0)$ states that $\sum_{i=0}^0 i = \frac{0(0+1)}{2}$. This clearly is true, since the left and right hand sides both equal 0. Therefore, $P(0)$ is true.

Inductive hypothesis: Assume $P(k)$ is true. That is, we assume $\sum_{i=0}^k i = \frac{k(k+1)}{2}$.

Inductive step: We must show that $P(k+1)$ follows from $P(k)$. That is, we must show that $\sum_{i=0}^{k+1} i = \frac{(k+1)(k+2)}{2}$. That can be done as follows:

$$\begin{aligned} \sum_{i=0}^{k+1} i &= \left(\sum_{i=0}^k i \right) + (k+1) \\ &= \frac{k(k+1)}{2} + (k+1) && \text{(by the inductive hypothesis)} \\ &= (k+1) \left(\frac{k}{2} + 1 \right) \\ &= \frac{(k+1)(k+2)}{2}. \end{aligned}$$

Notice that we didn't assume anything about the value of k , other than that $P(k)$ is true. Therefore, $\forall k \in \mathbb{N}. P(k) \implies P(k+1)$ must be true. Hence, the theorem follows from the principle of induction. \square

Note the structure of the inductive step. You try to show $P(k+1)$ starting only from the assumption that $P(k)$ is true. The idea is that $P(k+1)$ by itself is a difficult proposition to prove. Many difficult problems in computer science are solved by breaking the problem into smaller, easier ones. This is precisely what we did in the inductive step: $P(k+1)$ is difficult to prove, but we were able to recursively define it in terms of $P(k)$.

We will now look at another proof by induction, but first we will introduce some notation and a definition for divisibility. We say that integer a divides b (or b is divisible by a), written as $a|b$, if and only if for some integer q , $b = aq$.

Theorem: $n^3 - n$ is divisible by 3 for all $n \in \mathbb{N}$.

Proof: (by induction over n)

Base case: $P(0)$ asserts that $3|(0^3 - 0)$ or $3|0$, which is clearly true (since $0 = 3 \cdot 0$).

Inductive hypothesis: Assume $P(n)$ is true. That is, $3|(n^3 - n)$. In particular, there is some integer q such that $n^3 - n = 3q$.

Inductive step: We must show that $P(n+1)$ is true. In other words, we must prove that $3|((n+1)^3 - (n+1))$. We can calculate:

$$\begin{aligned} (n+1)^3 - (n+1) &= n^3 + 3n^2 + 3n + 1 - (n+1) \\ &= (n^3 - n) + 3n^2 + 3n \\ &= 3q + 3n^2 + 3n && \text{(by the inductive hypothesis)} \\ &= 3(q + n^2 + n). \end{aligned}$$

The latter expression is clearly divisible by 3, so $(n + 1)^3 - (n + 1)$ must also be divisible by 3. The theorem follows, by the principle of induction. \square

The next example we will look at is an inequality between two functions of n . Such inequalities are useful in computer science when showing that one algorithm is more efficient than another. Notice that for this example, we have chosen as our base case $n = 2$, which is natural because the claim we are aiming to prove holds for all natural numbers greater than or equal to 2. If you think about the underlying induction principle, it should be clear that this is perfectly valid, for the same reason that standard induction starting at $n = 0$ is valid (think back again to the domino analogy, where now the first domino is domino number 2).¹

Theorem: $\forall n \in \mathbb{N} . n > 1 \implies n! < n^n$.

Proof: (by induction over n)

Base case: $P(2)$ asserts that $2! < 2^2$, or $2 < 4$, which is clearly true.

Inductive hypothesis: Assume $P(n)$ is true (i.e., $n! < n^n$).

Inductive step: We must show $P(n + 1)$, which states that $(n + 1)! < (n + 1)^{n+1}$. We can calculate:

$$\begin{aligned}(n + 1)! &= (n + 1) \cdot n! \\ &< (n + 1) \cdot n^n && \text{(by the inductive hypothesis)} \\ &< (n + 1) \cdot (n + 1)^n \\ &= (n + 1)^{n+1}\end{aligned}$$

Hence, by the induction principle, $n! < n^n$ for all $n > 1$. \square

In the middle of the last century, a colloquial expression in common use was “that is a horse of a different color,” referring to something that is quite different from normal or common expectation. The famous mathematician George Polya (who was also a great expositor of mathematics for the lay public) gave the following proof to show that there is no horse of a different color.

Theorem: All horses are the same color.

Proof: The proof will be by induction on the number of horses. Let $P(n)$ denote the statement that for every set of n horses, all n horses in the set have the same color.

Base case: $P(1)$ is certainly true, since with just one horse, all horses in the set have the same color.

Inductive hypothesis: Assume $P(n)$, which is the statement that for any set of n horses, all n horses have the same color.

Inductive step: Given a set of $n + 1$ horses $\{h_1, h_2, \dots, h_{n+1}\}$, we can exclude the last horse in the set and apply the inductive hypothesis just to the first n horses $\{h_1, \dots, h_n\}$, deducing that they all have the same color. Similarly, we can conclude that the last n horses $\{h_2, \dots, h_{n+1}\}$ all have the same color. But now the “middle” horses $\{h_2, \dots, h_n\}$ (i.e., all but the first and the last) belong to both of these sets, so they have the same color as horse h_1 and horse h_{n+1} . It follows, therefore, that all $n + 1$ horses have the same color. Thus, by the principle of induction, all horses have the same color. \square

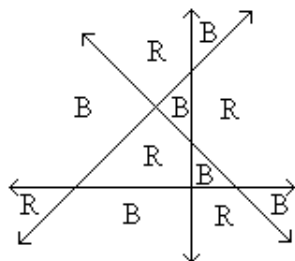
Clearly, it is not true that all horses are of the same color, so where did we go wrong in our induction proof? It is tempting to blame the induction hypothesis. But even though the induction hypothesis is false (for $n \geq 2$), that is not the flaw in the reasoning! Before reading on, think about this and see if you can understand why, and figure out the real flaw in the proof.

¹Alternatively, we could insist on making the base case $n = 0$ (which holds vacuously here because $0 > 1$ is false, so therefore $0 > 1 \implies 0! < 0^0$ is true). Then we would assert that $P(0) \implies P(1)$, since $P(1)$ holds (vacuously again, since $1 > 1$ is false, so $1 > 1 \implies 1! < 1^1$ is true), and that $P(1) \implies P(2)$ since $P(2)$ holds (as we show in the base case below). Then we would proceed as in the inductive step of the proof below. But this is all rather tedious.

What makes the flaw in this proof a little tricky to pinpoint is that the induction step *is* valid for a “typical” value of n , say, $n = 3$. The flaw, however, is in the induction step when $n = 1$. In this case, for $n + 1 = 2$ horses, there are *no* “middle” horses, and so the argument completely breaks down!

Some of you might still not feel completely convinced. Why is the above flaw more convincing than simply saying that the induction hypothesis is false? Saying that the induction hypothesis is false is like saying that the statement of the theorem is false, and so there is definitely a flaw in the proof. True, but our task was to pinpoint exactly where *in the proof* the flaw occurs. The point is that a valid induction proof involves only showing the base case, say $P(0)$, and that $\forall n . P(n) \implies P(n + 1)$. One way of showing $P(n) \implies P(n + 1)$ is to assume $P(n)$ is true and then show that $P(n + 1)$ is true. If $P(n)$ is false, then $P(n) \implies P(n + 1)$ is vacuously true. So just saying that the induction hypothesis $P(n)$ is false does not pinpoint the flaw in the proof. In this case, the proposition $P(n) \implies P(n + 1)$ happens to be true for every n , *except* for $n = 1$, in which case the proposition is false. This makes the flaw in the proof quite difficult to spot.

Two Color Theorem: There is a famous theorem called the four color theorem. It states that any map can be colored with four colors such that any two adjacent² countries must have different colors. The four color theorem is very difficult to prove, and several bogus proofs were claimed since the problem was first posed in 1852. It was not until 1976 that the theorem was finally proved (with the aid of a computer) by Appel and Haken. (For an interesting history of the problem, and a state-of-the-art proof, which is nonetheless still very challenging, see <http://www.math.gatech.edu/~thomas/FC/fourcolor.html>.) We consider a simpler scenario, where we divide the plane into regions by drawing straight lines. We want to know if we can color this map using no more than two colors (say, red and blue) such that no two regions that share a boundary have the same color. Here is an example of a two-colored map:



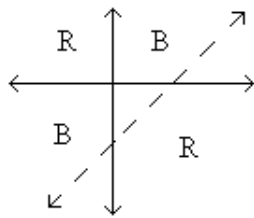
It turns out that every such map can indeed be two-colored. We will prove this “two color theorem” by induction on n , the number of lines.

Base case: $P(0)$ is the proposition that a map with $n = 0$ lines can be colored using no more than two colors. But this is easy, since we can just color the entire plane using one color. Therefore $P(0)$ is indeed true.

Inductive hypothesis: Assume $P(n)$, that is, every map with n lines can be two-colored.

Inductive step: We want to prove $P(n + 1)$. We are given a map with $n + 1$ lines and wish to show that it can be two-colored. Let’s see what happens if we remove a line. With only n lines on the plane, we know we can two-color the map (by the inductive hypothesis). Let us make the following observation: if we swap red \leftrightarrow blue, we still have a two-coloring. With this in mind, let us place back the line we removed, and leave colors on one side of the line unchanged. On the other side of the line, swap red \leftrightarrow blue. We claim that this is a valid two-coloring for the map with $n + 1$ lines.

²Two countries are considered adjacent if they share a border—not just a point.



Why does this work? We can say with certainty that regions which do not border the line are properly two-colored. But what about regions that do share a border with the line? We must be certain that any two such regions have opposite coloring. But any two regions that border the line must have been the same region when the line was removed, so the reversal of color on one side of the line guarantees an opposite coloring.

Strengthening the Inductive Hypothesis

Let us prove by induction the following proposition:

Theorem: For all $n \geq 1$, the sum of the first n odd numbers is a perfect square.

Attempted proof: By induction on n .

Base case: For $n = 1$, the first odd number is 1, which is a perfect square.

Inductive hypothesis: Assume that the sum of the first n odd numbers is a perfect square, say k^2 .

Inductive step: The $n + 1$ -th odd number is $2n + 1$, so the sum of the first $n + 1$ odd numbers is $k^2 + 2n + 1$. But now we are stuck. Why should $k^2 + 2n + 1$ be a perfect square? This attempt at a proof is a failure.

Here is an idea: let us show something stronger!

Theorem: For all $n \geq 1$, the sum of the first n odd numbers is n^2 .

Proof: By induction on n .

Base case: For $n = 1$, the first odd number is 1, which is 1^2 .

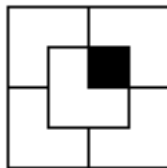
Inductive hypothesis: Assume that the sum of the first n odd numbers is n^2 .

Inductive step: The $(n + 1)$ -th odd number is $2n + 1$, so the sum of the first $n + 1$ odd numbers is $n^2 + (2n + 1) = (n + 1)^2$. Thus by the principle of induction the theorem holds. \square

See if you can understand what happened here. We could not prove a proposition, so we proved a harder proposition instead! Can you see why that can sometimes be easier when you are doing a proof by induction? When you are trying to prove a stronger statement by induction, you have to show something harder in the induction step, but you also get to assume something stronger in the induction hypothesis. Sometimes the stronger assumption helps you reach just that much further...

Here is another example:

Imagine that we are given L-shaped tiles (i.e., a 2×2 square tile with a missing 1×1 square), and we want to know if we can tile a $2^n \times 2^n$ courtyard that is missing a 1×1 square in the middle. Here is an example of a successful tiling in the case that $n = 2$:



It turns out that such a tiling always exists. Let us try to prove this fact by induction on n .

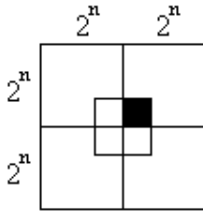
Base case: $P(1)$ is the proposition that a 2×2 courtyard can be tiled with L-shaped tiles with a missing 1×1 square in the middle. But this is easy:



So $P(1)$ is indeed true.

Inductive hypothesis: Assume $P(n)$ is true, that is, we can tile a $2^n \times 2^n$ courtyard that's missing a 1×1 square in the middle.

Inductive step: We want to show that we can tile a $2^{n+1} \times 2^{n+1}$ courtyard with a missing 1×1 square in the middle. Let's try to reduce this problem so we can apply our inductive hypothesis. A $2^{n+1} \times 2^{n+1}$ courtyard can be broken up into four smaller courtyards of size $2^n \times 2^n$, each with a missing 1×1 square as follows:



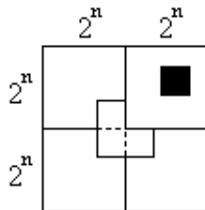
But the holes are not in the middle of each $2^n \times 2^n$ courtyard, so the inductive hypothesis does not help. Once again, we are stuck. This attempted proof is a failure.

How should we proceed? We should strengthen our inductive hypothesis!

What we are about to do is completely counter-intuitive. It's like attempting to lift 100 pounds, failing, and then saying "I couldn't lift 100 pounds. Let me try to lift 200," and then succeeding! Instead of proving that we can tile a $2^n \times 2^n$ courtyard with a hole in the middle, we will try to prove something stronger: that we can tile the courtyard with a hole *no matter where the hole is located*. It is a trade-off: we have to prove more, but we also get to assume a stronger hypothesis. The base case is the same, so we will just work on the inductive hypothesis and step.

Inductive hypothesis: (second attempt) Assume $P(n)$ is true, so that we can tile a $2^n \times 2^n$ courtyard with a missing 1×1 square anywhere.

Inductive step: (second attempt) Suppose we are given a $2^{n+1} \times 2^{n+1}$ that is missing a single 1×1 square somewhere—say, in the upper-right corner. As before, we can break up the $2^{n+1} \times 2^{n+1}$ courtyard as follows.



By placing the first tile in the middle as shown, we get four $2^n \times 2^n$ courtyards, each with a 1×1 hole. Three of these courtyards have the hole in one corner, while the fourth has the hole in a position determined by the hole in the $2^{n+1} \times 2^{n+1}$ courtyard. The stronger inductive hypothesis now applies to each of these four courtyards, so that each of them can be successfully tiled. Thus, we have proven that we can tile a $2^{n+1} \times 2^{n+1}$ courtyard with a hole anywhere! Hence, by the induction principle, we have proved the (stronger) theorem.