One Should Always Generalize

Jonathan Tsai

University of Hong Kong

December 1, 2018

What is generalization?

Mathematician Carl Jacobi (1804-1851) once said about mathematics: "One should always generalize".

But what is "generalization"?

Definition

Generalization is the process of making a general statement by inferring from a specific case.

It is one of the key ideas in advanced (university level) mathematics.

What is generalization?

Idea:

1. Start with a statement which we know is true.
2. Make a statement which includes the original statement as a special case (the general statement).
3. Prove the general statement.
4. Repeat this process to generalize further.

We consider a very simple exaggerated example:
"The number ways of obtaining a total of 7 when rolling two dice is 6"

How can we generalize this statement?

The simple example

One might generalize this statement as follows:

- The number ways of obtaining a total of 7 when rolling two 6-sided dice is 6
- The number ways of obtaining a total of k when rolling two 6-sided dice is $6 - |k - 7|$
  The previous statement is a special case ($k = 7$).
- The number ways of obtaining a total of k when rolling n 6-sided dice is ... [Exercise]
  The previous statements are special cases.

Can we generalize this further?

WARNING

There is a problem!

Abstraction

We are limited by the fact that the statement only makes sense when the numbers are positive whole numbers.

The solution: **Abstraction**

Definition

Abstraction is the process of extracting the underlying essence of a concept by removing the original (real-world) context.

By removing the context, we are free to make further generalizations.

An example

(Euclidean) Geometry (angles in triangles add up to 180°). $\rightarrow$

- Spherical geometry (angles in triangles add up to more than 180°).
- Hyperbolic geometry (angles in triangles add up to less than 180°).

There are spherical and hyperbolic variants of familiar geometric results e.g. Spherical/Hyperbolic cosine rule.

Spherical geometry "makes sense". But later Einstein's general relativity showed that space-time has hyperbolic geometry.

Investigation

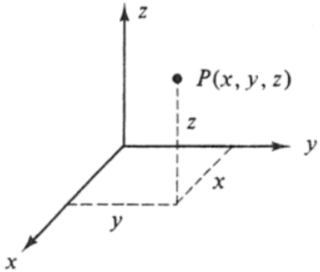
We will now look at two situations in detail:

1. Generalizing the concept of **dimension**.
2. Generalizing the concept of **the real numbers**.

First Case: Dimension

We know that 2 dimensions corresponds to "flat". A 2 dimensional object can be expressed with (cartesian) coordinates (x, y).

Similar, we can express 3 dimensions using three coordinates (x, y, z)



We want to define the concept of higher dimensions.

Higher Dimensions

We first perform an abstraction: remove the "visual" context.

We can generalize to $4, 5, 6 \ldots$ dimensions by adding coordinates.
In other words, an n-dimensional object can be expressed with n coordinates $(x_1, x_2, \ldots, x_n)$. Notice how the generalization is accompanied by a "new" notation - this is very common.

Lines (2 dimensions) and planes (3 dimensions) are generalized to n-dimensional hyperplanes. Other concepts can also be generalized to higher dimensions.

Higher Dimensions - Applications

In 2 dimensions, the solution to a system of equations in x, y can be expressed as the intersection of lines.

A system of equations with n unknowns can be expressed as the intersection of n-dimensional hyperplanes.

Similarly for inequalities i.e. linear programming.

Applications:

- Physics e.g. "space-time"
- Economics - different variables correspond to different economic quantities (cost, price, quantity etc.). System of inequalities and linear programming are very important (e.g. maximize profit given certain constraints).
- ...

Problem!

A plane in 3-dimensional space should be considered as 2 dimensional - but we don't have any way to define dimension in this case!
Our original definition is not good enough. We need to generalize further.

Generalizing further

We can generalize further. Points in two dimensions can be represented by vectors:

$$(x, y) \to \begin{pmatrix} x \\ y \end{pmatrix}$$

We can define addition of vectors and also multiplication by scalars (numbers) by

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} \text{ and } k \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} kx \\ ky \end{pmatrix}$$

These also have geometric interpretations. Thus:

$$x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$$

This can be generalized to n-dimensions in the obvious way.

Vector spaces

We abstract by discarding the geometric information (e.g. coordinate axes) and just retaining the concept of addition and scalar multiplication (algebraic information).
[This is a very common concept in "abstract algebra"].

Any set in which an "addition" and "scalar multiplication" is defined is called a **vector space**. By "defined" we mean they satisfy certain properties that we would expect these operations to satisfy.

The original n-dimensional coordinates are the particular vector space which is called $\mathbb{R}^n$

Dimension of a vector space

In $\mathbb{R}^n$, every vector can be written (uniquely) as

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

This collection of n vectors is called a **basis**. In general, a basis of a vector space is a collection of vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}$ so that any vector $\mathbf{v}$ can be written as:

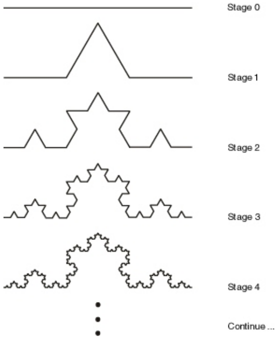
$$\mathbf{v} = x_1\mathbf{v_1} + x_2\mathbf{v_2} + \cdots + x_n\mathbf{v_n}$$

in a unique way. The **dimension** of a vector space is the number of vectors in the basis.

Question: What problems might arise in this definition?

Start again!

It's a natural idea that 1 dimensional objects have length; 2 dimensional objects have area; 3 dimensional objects have volume. Consider:



If we continue this procedure forever ("Stage ∞"), we will get a curve called the Koch snowflake. It can be calculated (easily) that the Koch snowflake has infinite length. Should we consider it one dimensional?

Fractional dimension

The Koch snowflake is an example of a fractal. All fractal curves have infinite length. Can we distinguish them by considering a generalization of dimension?

We abstract by focusing on one property of length: If a curve has finite length l and we decompose it to pieces of length $1/k$, the number of such pieces is proportional to k^1. Doing this with the Koch snowflake results in a number of pieces which is proportional to $k^{4/3}$.

We define the dimension of the Koch snowflake to be $4/3$. This fractional dimension allows one to distinguish between different fractals

Case 2: Real numbers

The integers $\mathbb{Z}$ is the set of all whole numbers; the rational numbers $\mathbb{Q}$ is the set of all fractions.

However, not all numbers can be expressed as fractions - these are the irrational numbers e.g. $\sqrt{2}$, π. The set of "all" numbers (rational and irrational) are the real numbers $\mathbb{R}$

Problem: This definition is circular! i.e. defined in terms of itself

You may also have heard the definition "a real number represents a distance on the line" or "a real number has a decimal expansion".

Leopold Kronecker (1823-1891) once said: "God made the integers, all else is the work of man"

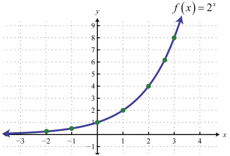
An interesting example

You should know how 2^a is defined when a is a whole number. For a rational number $\frac{m}{n}$, you also know that

$$2^{\frac{m}{n}} = \sqrt[n]{2^m}$$

Question: What is 2^a when a is an irrational number?

You also may have met exponential functions and their graphs e.g. $y = 2^x$.



This graph shows us that we actually do know the value of 2^a when a is irrational. It gives us a hint to our approach.

The real numbers vary continuously

We expect the graph of an exponential function to be "continuous" because we expect the real numbers to "vary continuously".

In other words, when two numbers x and y are close, we expect 2^x and 2^y to be close. We can plot an exponential graph because for each real number x and no matter how small a distance specified, there is a rational number that is within this distance to x.

In other words we can fill in the gaps just by knowing the value at the rational numbers.

This method of filling in the gaps in the numbers is called **completion**; we define the real numbers to be the completion of the rational numbers. Here, the distance from x to y refers to $|x - y|$ and this relies on the concept of absolute value $|\cdot|$.

Generalizing absolute values and distances

We can abstract the real numbers by generalizing the notion of absolute values and distance.

Any concept of absolute value of a number x, denoted by $\|\cdot\|$ should satisfy certain expected properties like:

1. $\|x\| \geq 0$ and $\|x\| = 0$ only when $x = 0$.
2. $\|xy\| = \|x\|\|y\|$.
3. $\|x + y\| \leq \|x\| + \|y\|$ ("Triangle Inequality")

Any absolute value $\|\cdot\|$ defines a concept of a distance between two numbers x and y by $\|x - y\|$. Using this distance when considering closeness gives us a different completion and a generalization of the real numbers.

The p-adic numbers

Consider a rational number x and a prime number p. Since every integer can be uniquely factorized into the product of prime numbers, we can write

$$x = \frac{a}{b} p^n$$

so that a and b have no common factors with p. We then define the p-adic absolute value of a to be $|a|_p = p^{-n}$.

It can be shown that this satisfies the required properties of an absolute value. Then $|x - y|_p$ defines a distance between x and y called the p-adic distance.

If we consider the completion of the rational numbers using the p-adic distance, we get the p-adic numbers.

Is this useful?

Number theory can be thought of as the study of the properties of integers. In this context, it is not immediately apparent that the real numbers should play any role in this.

However, calculus is an immensely powerful tool in number theory but to define the tools of calculus, we need the real numbers.

From a purely number theory viewpoint, p-adic distance is a more natural concept of distance than the standard definition of distance. So p-adic numbers are arguably a more natural construction than the real numbers!

They also allow us to define a new set of calculus tools.
This "p-adic calculus" is a key tool in modern day number theory research e.g. in the proof of Fermat's Last Theorem by Andrew Wiles.

Also...

Theorem (Ostrowski's Theorem)

Every absolute value that can be defined on the rational numbers is equivalent to either $|\cdot|$, the standard absolute value, or $|\cdot|_p$, the p-adic absolute value.

There is no other completion of the rational numbers besides the real numbers and p-adic numbers that can be defined in this way!

Food for thought

Several areas of mathematics are created with the procedure of abstraction followed by generalization.

- What are some motivations for generalizing a concept?
- What considerations does one have to make when generalizing a concept?
- What does a mathematician gain from generalization?
- What potential pitfalls are there?

The end

Thanks for listening!

Any questions?