Lecture 1: Review of Sets, Sequences and Functions

Anup Rao

January 3, 2018

We discuss some basic terminology related to sets, sequences and functions. We talk about the difference between continuous and discrete sets, and how many important problems in computer science can be cast as understanding the size of sets.

Sets

WE BEGIN BY REVIEWING THE CONCEPT OF SETS. A *set* is a collection of distinct objects. The objects that are included in the set are called its *elements*. The identity of the set is determined by the objects it contains. The order in which the objects is presented is not relevant. For example the set $A = \{1, 2, 3\}$ is a set containing the numbers $1, 2$ and 3. It is identical to the set $B = \{2, 1, 3\}$.

The empty set is often denoted $\emptyset$.

Sets can be finite, or even infinite, discrete or continuous. Some well known examples of sets of numbers include:

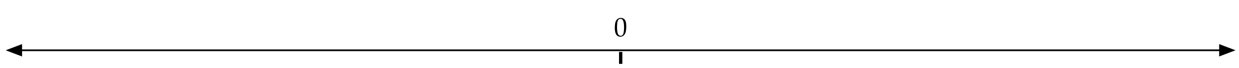
integers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$.

natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$.

the first n natural numbers $[n] = \{1, 2, 3, \ldots, n\}$.

rational numbers all numbers that can be expressed as ratios of integers: $\mathbb{Q} = \left\{\frac{p}{q} : p, q \in \mathbb{Z}, q \neq 0\right\}$.

reals the set of all numbers $\mathbb{R}$ that are on the real number line.



There are several operations one can perform on sets, to construct new sets from old sets:

cartesian product Given a set S, the cartesian product $S \times S$ is a new set whose elements are ordered pairs of elements from S.

$$S \times S = \{(a, b) : a \in S, b \in S\}.$$

For a natural number k, we write S^k to denote the k-fold cartesian product of the set S with itself. For example, $\mathbb{R}^2$ represents all points on the plane. $\mathbb{R}^3$ represents all points in 3 dimensional space.

union The union of two sets A, B is the set that contains elements that are either in A *or* in B.

$$A \cup B = \{c : c \in A \text{ or } c \in B\}.$$

intersection The intersection of two sets A, B is the set that contains elements that are in both A *and* in B.

$$A \cap B = \{c : c \in A \text{ and } c \in B\}.$$

complement When a set S is contained[1] in a set U, the complement of S is the set of elements that are not included in S.

[1] namely, all elements of S are also elements of U

$$S^c = \{x \in U : x \notin S\}.$$

power set Given any set S, its power set $2^S = \{T : T \subseteq S\}$ is the set whose elements are the subsets of S. For example, $2^{[n]}$ is the set containing all subsets of $[n]$.

The power set is always a lot bigger than the original set. In fact, even when S is infinite, you can show that there is no injective function mapping 2^S to S, which corresponds to the fact that 2^S is a lot bigger than S.

It is often helpful to visualize the interaction of sets using Venn diagrams.

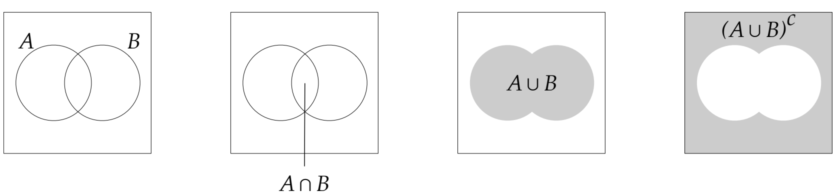


Figure 1: A Venn diagram showing some of the parts of the universe that can be generated using two sets.

We write $S \subset U$ to denote that S is a subset of U. For example, we have $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$. The sets $\mathbb{Z}, \mathbb{N}$ and all finite sets are usually considered to be *discrete*. The sets $\mathbb{R}, \mathbb{R}^2$ are examples of *continuous* sets.

Later, we shall use diagrams similar to Venn diagrams to visualize probability spaces.

Strictly speaking, the set $\mathbb{Q}$ can be viewed as discrete, because it is countable, but it is not really very important to rigorously define which sets are discrete and which are not.

Sets are often easy to define, but hard to understand. Some of the hardest algorithmic problems in computer science have to do with telling whether a well defined set is empty or not. For example, in the famous NP-complete problem 3-SAT, we are given a 3-SAT formula ϕ and want to know whether or not this formula has an assignment that satisfies it. Given any candidate assignment, it is easy to check whether or not the assignment satisfies the formula. However, we do not know of any method to count the number of satisfying assignments that is better than trying each potential assignment.

Another interesting example is the set of prime numbers. This is the set of natural numbers that do not have non-trivial factorizations:

$$P = \{2, 3, 5, 7, 11, 13, 17, 19, 23, \dots\}.$$

We can define $P_n = P \cap [n]$ to be the prime numbers that are at most n. We do not have a concise formula for the number of elements $|P_n|$, and coming up with such an estimate is related to one of the most famous open problems in mathematics—the Riemann hypothesis.

Sequences

NEXT, LET US TALK ABOUT SEQUENCES. A *sequence* is a list of elements coming from a set. For example:

$$1, 3, 2, -2, 1, -4$$

is a sequence of integers. Here the *order* of the elements is crucial. The sequence $1, 2, 3$ is not the same as the sequence $2, 1, 3$.

Another way to view the set S^k is as the set of sequences of length k from S.

Functions

FUNCTIONS ARE CLOSELY RELATED TO BOTH sets and sequences. A *function* is an object mapping elements of one set to elements of another. We write $f : D \to R$ if the function assigns an element $f(x) \in R$ to every element $x \in D$. Both sets and sequences can be thought of as functions, so all three of these concepts are closely related to each other.

There are many other words used to describe concepts similar or identical to sequences and functions. These include *tuple, array, vector, map*. We do not discuss the differences between these concepts here, but you can look it up if you are curious.

For example, suppose $A \subseteq [n]$ is a subset of the first n natural numbers. Then A is also defined by its indicator function $1_A : [n] \to \{0, 1\}$:

$$1_A(i) = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{otherwise.} \end{cases}$$

The set can also be encoded by the indicator vector $v \in \{0, 1\}^n$, where

$$v_i = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{otherwise.} \end{cases}$$

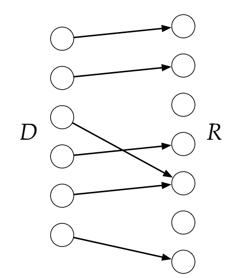


Figure 2: It is natural to visualize a function as a directed graph where every vertex in the domain has an outgoing edge going to some vertex in the range.

Refresh your memory on the following concepts:

- *one to one correspondence* or *bijection*
- *injective* function or *one to one* function
- *onto* function or *surjective function*

Union Closed Conjecture

Let us make a digression to discuss a really cute open problem that has confounded mathematicians for decades.

Given a collection of sets $S_1, S_2, \ldots, S_m \subseteq [n]$, we say they are *closed under union* if the union of two of the sets is also a set in the list. In other words, for every i, j, there is a k such that $S_k = S_i \cup S_j$. The union closed conjecture states that given any family of sets closed under union, there must be an element that is in at least half of the sets.

Conjecture 1. *If $S_1, \ldots, S_m \subseteq [n]$ are closed under union, then there must be an element $i \in [n]$ such that i is an element of at least $m/2$ of the sets.*

The fact that mathematicians cannot resolve the conjecture is evidence that the final resolution will involve a really creative and original idea.

If you try to play around with some examples of union closed families of sets, you will see that the conjecture always seems to hold. Computer search has been used to verify that the conjecture holds for small values of n. But no one has been able to prove that the conjecture is true for every n.