Topic 2: Sets, Operators, and Function

02-680: Essentials of Mathematics and Statistics

September 6, 2024

1 Sets

A ***set*** is an unordered collection of unique elements. We have two standard ways of explicitly defining a set. First is simple (exhaustive) enumeration:

$$A = \{\text{“Welcome”}, \text{“to”}, \text{“02-680”}\}.$$

In this case the set contains 3 elements, each of which is a string. The second way to define a set by abstraction:

$$B = \left\{x^2 \mid x = 2 \vee x = 3\right\}.$$

Note here that the right hand side of the definition (after the $|$) is a statement written in ***propositional logic*** (we talked about this last topic).

Set Membership. Likely the most important thing we can do with sets is check if something is in the set, or we would say it is an element of the set. We write that mathematically as $x \in S$ where x is an element and S is a set. So going back to the examples above we would say $4 \in B$, but $16 \notin B$.

Set Cardinality. We may also want to consider the size of a set, we call this the ***cardinality***. In the examples above $|A| = 3$ and $|B| = 2$. Note that if we define another set:

$$C = \left\{(2+2), (2-2), (2 \times 2), (2 \div 2), \left(2^2\right)\right\}$$

the cardinality $|C| = 3$. Why is that? Shouldn't it be 5? Try to rewrite the set using exhaustive enumeration on single integers.

1.1 Special Sets

There are some sets that we use that are so important we define them on their own, they also usually have special symbols.

The Empty Set. Its often useful to define the set that contains no elements, we call this the ***empty set*** or alternately the ***null set***.

$$\emptyset = \{\}$$

Since there are no elements in $\emptyset$ (which we can write as $\forall x : x \notin \emptyset$, or $\nexists x : x \in \emptyset$), $|\emptyset| = 0$.

Common Sets. In mathematics we are often talking about a defined set of numbers (think integers, real numbers, etc.), and we use them so often that they get special notation as well. In this course we will mainly restrict ourselves to the *real* numbers, we may encounter any of these:

Numbers	Symbol	Definition
Naturals	$\mathbb{N}$	Positive Whole Numbers
Integers	$\mathbb{Z}$	All Whole Numbers
Rational Numbers	$\mathbb{Q}$	Numbers Representable as Fractions
Real Numbers	$\mathbb{R}$	Decimal Numbers
Booleans	$\{0, 1\}$	No special symbol, but still important
Characters	Σ	Sometimes "Alphabet"[1], useful here since this course is in Computational Biology

(Note in LaTeX we use the command `\mathbb` command to create these symbols.)

So we can easily expand B from above to be all squares:

$$B' = \left\{x^2 \mid x \in \mathbb{Z}\right\}$$

We some times also want to slightly restrict these sets, so you may see something like $\mathbb{Z}^+$ which is equivalent to $x \in \mathbb{Z} \mid x > 0$, or $\mathbb{Z}^{\geq 0}$ which is equivalent to $x \in \mathbb{Z} \mid x \geq 0$. Note that technically $\mathbb{Z}^+$ does not include 0 and $\mathbb{Z}^{\geq 0}$ does, but be careful because some authors (hopefully not including myself) use the former when they mean the latter.

Given that we can further redefine B' above

$$B'' = \left\{x^2 \mid x \in \mathbb{Z}^{\geq 0}\right\}$$

though B' and B'' are equivalent.

2 Operators

Just like we do with numbers, we often need to compare sets, or do computations on them. We call these set operators.

[1] While the other sets on this list are always the same size, the size of an alphabet is context dependent. Sometimes it's DNA, sometimes its capital latin (english) letters, other times its kanji.

Union ($\cup$) and Intersection ($\cap$). Both of these operations result in the creation of a new set, you can think of these like the logical operators *and*($\wedge$) and *or*($\vee$). (We will talk about logical operators next week.) For two sets S and T we define the following:

$$S \cup T := \{x \mid x \in S \vee x \in T\} \text{ and}$$

$$S \cap T := \{x \mid x \in S \wedge x \in T\}$$

That is to say and element is the union if it's in *either* of the two sets, but only in the intersection if its in both.

Lets look at an example:

$$B \cap C = \{4\},$$

4 is the only element thats in both B and C. On the other hand

$$B \cup C = \{0, 1, 4, 9\},$$

since all of the elements of both sets are included.

This leads to two interesting properties that we may make use of later, for arbitrary sets S and T:

$$\begin{array}{ccccc} \max\{|S|, |T|\} & \leq & |S \cup T| & \leq & |S| + |T| \\ 0 & \leq & |S \cap T| & \leq & \min\{|S|, |T|\} \end{array}$$

Set Difference ($\setminus$). Similar to numerical subtraction, the ordering of the operators matters. In fact difference is so similar to subtraction that you may sometimes see the minus sign used in place of the standard backlash.

$$S \setminus T := \{x \mid x \in S \wedge x \notin T\}$$

You can think if it as the elements that are unique to S.

Going back to our example:

$$B \setminus C = \{9\} \text{ and}$$

$$C \setminus B = \{0, 1\}$$

This means we can play around a little and say that

$$S \cup T = (S \cap T) \cup (S \setminus T) \cup (T \setminus S)$$

Complement ($\sim$). While we can write something like $\sim S$, you will normally see set complements written as $\overline{S}$. We also need to assume a ***universe*** for this operator to be properly defined. This is normally understood from context, but the universe is the set of all possible elements the set S is a part of. Just like we did earlier with S and T being arbitrary sets, let U be an arbitrary universe. Once we do that we can define

$$\overline{S} := \{x \in U \mid x \notin S\}$$

So if we assume that the universe for B is $\mathbb{Z}$, then

$$\overline{B} = \{-\infty, ..., -1, 0, 1, 2, 3, \;\; 5, 6, 7, 8, \;\; 10, 11...., \infty\}$$

Using our previously defined items we can also say that

$$C \setminus B = C \cap \overline{B}$$

Subset ($\subseteq$), Superset ($\supseteq$), and equality ($=$). We often need to know if one set *is contained* within another (notice while similar this is different from saying is larger than another). To describe this relationship we say one set is an ***subset*** of another if all of it's elements are in both sets. That is

$$S \subseteq T \iff \forall x \in S : x \in T$$

Similarly, if a set *contains* another set, we can say the original is a ***superset*** of the other.

$$S \supseteq T \iff \forall x \in T : x \in S$$

We already know that $B \nsubseteq C$ and $B \nsupseteq C$, but $B \subseteq B'$ and $B' \supseteq B$.

But what are those little lines under the $\subseteq$ symbol? We will need to explain equality first. ***Set equality*** is defined as follows:

$$S = T \iff (\forall x \in S : x \in T) \wedge (\forall y \in T : y \in S)$$

or another way

$$S = T \iff (S \subseteq T) \wedge (S \supseteq T)$$

So if we define $C' = \{0, 1, 2, 4\}$, then $C = C'$.

So then coming back to $\subseteq$ (vs. $\subset$); we define a ***proper subset*** as

$$S \subset T \iff (\forall x \in S : x \in T) \wedge (\exists y \in T : y \notin S)$$

In other words for S to be a proper subset of T it first needs to be a subset, but there needs to be at least one item in T thats not in S. A similar condition exists for ***proper superset*** which we define as

$$S \supset T \iff (\forall x \in T : x \in S) \wedge (\exists y \in S : y \notin T)$$

So going back to our previous examples we can be more precise $B \subset B'$ and $B' \supset B$; but also $B'' \subseteq B'$ but $B'' \not\subset B'$.

Power Set ($\mathcal{P}(\cdot)$) So, can we find all the subsets of a set? The ***Power Set*** is actually the collection of all the subsets of a given set.

$$\mathcal{P}(S) := \{T \mid T \subseteq S\}$$

It may be easiest to look at an example:

$$\mathcal{P}(B) = \left\{ \begin{matrix} \emptyset, \\ \{4\}, \\ \{9\}, \\ \{4, 9\} \end{matrix} \right\}$$

So the power set is a set of sets, and each element of the power set *is a* set. (That means a the power set of a power set of a set, i.e. $\mathcal{P}(\mathcal{P}(B))$, is a set of sets of sets and an element is a set of sets....)

An interesting fact is that the size of the power set of a set is

$$|\mathcal{P}(S)| = 2^{|S|}$$

Why? Think of each element of $\mathcal{P}(S)$ as a binary number of length $|S|$, if the i-th bit (or position) is 1 then the i-th element[2] from S is in that subset. Then each binary number represents a *unique* subset of S.

3 Functions

We've all seen functions in algebra, and looked at them, say, as a graph:

$$y = 3x^2 + 2$$

But we want to make sure we're on the same page when it comes to the generalized function we're going to talk about in this course. We will say that a ***function*** provides a mapping

[2] In this particular example we would impose an ordering on the elements which is not in the definition of sets

from one set onto another; formally we say that a function f that maps from a set S to a set T is written as

$$f : S \mapsto T$$

In most algebra classes you were likely working with functions g in the form $g : \mathbb{R} \mapsto \mathbb{R}$ (which while not defined above is implied), but that doesn't have to be the case. For instance in Computational Biology, we often want a function to map strings (which we will use the generic notation Σ^* for now without explanation, wait till next week) to other strings; something like $h : \Sigma^* \mapsto \Sigma^*$.

Domain, Codomain, and Range. When writing $f : S \mapsto T$ we call S the ***domain*** set, and T the ***codomain***. Note that the codomain is slightly different from the ***range*** of a function; the range is the subset of T that is reachable from an input in S. Formally the range is

$$\{y \in T \mid \exists x \in S : f(x) = y\}$$

So thinking about the function at the start of this section ($y = 3x^2 + 2$), in algebra class we would have said that the domain is $\mathbb{R}$, as is the codomain, but the range (you may have heard it called the image) is the set $\{y \in \mathbb{R} \mid y \geq 2\}$ (notice this is always a subset of the codomain).

Extra details

Note that the specification of the function *includes* both the domain and codomain. But if those three elements are not compatible, it's not a function.

Thinking again about the function we used earlier, it is a function if the domain and codomain are both $\mathbb{R}$, and it is a function if they are both $\mathbb{Z}$.

But it is ***not*** a function if the domain is $\mathbb{R}$ and the $\mathbb{Z}$; in this case there are elements in the domain (lets say $d = 1.5 \in \mathbb{R}$) for which the result of the equation part ($f(d) = (1.5)^2 + 3 = 2.25 + 3 = 5.25$) is not in the codomain ($5.25 \notin \mathbb{Z}$). Thus the specification is not a function ($f(x) = 3x^2 + 2$, $f : \mathbb{R} \mapsto \mathbb{Z}$).

Function Composition ($\circ$). We may find ourselves later in the semester needing to apply one function to the output of another, this it known as ***function composition*** defined formally as

$$g \circ f : S \mapsto U \text{ assuming } f : S \mapsto T \wedge g : T \mapsto U.$$

In use it may look something like $g\,(f\,(x))$, so x would need to be from S since thats the input to f and since the input to g needs to be from T, that also needs to be the codomain of f then the codomain of g ends up as the codomain of the composite function.

Useful References

Liben-Nowell, “Connecting Discrete Mathematics and Computer Science, 2e”. §2.3,2.5