Chapter 15

Sets of Sets

So far, most of our sets have contained atomic elements (such as numbers or strings) or tuples (e.g. pairs of numbers). Sets can also contain other sets. For example, $\{\mathbb{Z}, \mathbb{Q}\}$ is a set containing two infinite sets. $\{\{a, b\}, \{c\}\}$ is a set containing two finite sets. In this chapter, we'll see a variety of examples involving sets that contain other sets.

15.1 Sets containing sets

Sets containing sets arise naturally when an application needs to consider some or all of the subsets of a base set A. For example, suppose that we have a set of 6 students:

$$A = \{\text{Ian}, \text{Chen}, \text{Michelle}, \text{Emily}, \text{Jose}, \text{Anne}\}$$

We might divide A up into non-overlapping groups based on what dorm they live in:

$$B = \{\{\text{Ian}, \text{Chen}, \text{Jose}\}, \{\text{Anne}\}, \{\text{Michelle}, \text{Emily}\}\}$$

We could also construct a set of overlapping groups, each containing students who play a common musical instrument (e.g. perhaps Michelle and

Chen both play the oboe). The set of these groups might look like:

$$D = \{\{\text{Ian}, \text{Emily}, \text{Jose}\}, \{\text{Anne}, \text{Chen}, \text{Ian}\}, \{\text{Michelle}, \text{Chen}\}, \{\text{Ian}\}\}$$

Or we could try to list all ways that we could choose a 3-person committee from this set of students, which would be a rather large set containing elements such as $\{\text{Ian}, \text{Emily}, \text{Jose}\}$ and $\{\text{Ian}, \text{Emily}, \text{Michelle}\}$.

When a set like B is the domain of a function, the function maps an entire subset to an output value. For example, suppose we have a function $f : B \to \{\text{dorms}\}$. Then f would map each set of students to a dorm. E.g. $f(\{\text{Michelle}, \text{Emily}\}) = \text{Babcock}$.

The value of a function on a subset can depend in various ways on whatever is in the subset. For example, suppose that we have a set

$$D = \{\{-12, 7, 9, 2\}, \{2, 3, 7\}, \{-10, -3, 10, 4\}, \{1, 2, 3, 6, 8\}\}$$

We might have a function $g : D \to \mathbb{R}$ which maps each subset to some descriptive statistic. For example, g might map each subset to its mean value. And then we would have $g(\{-12, 7, 9, 2\}) = 1.5$ and $g(\{1, 2, 3, 6, 9\}) = 4.2$.

When manipulating sets of sets, it's easy to get confused and "lose" a layer of structure. To avoid this, imagine each set as a box. Then $F = \{\{a, b\}, \{c\}, \{a, p, q\}\}$ is a box containing three boxes. One of the inside boxes contains a and b, the other contains c, and the third contains a, p, and q. So the cardinality of B is three.

The empty set, like any other set, can be put into another set. So $\{\emptyset\}$ is a set containing the empty set. Think of it as a box containing an empty box. The set $\{\emptyset, \{3, 4\}\}$ has two elements: the empty set and the set $\{3, 4\}$.

15.2 Powersets and set-valued functions

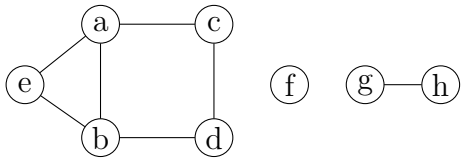
If A is a set, the powerset of A (written $\mathbb{P}(A)$ is the set containing all subsets of A. For example, suppose that $A = \{1, 2, 3\}$. Then

$$\mathbb{P}(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$$

Suppose A is finite and contains n elements. When forming a subset, we have two choices for each element x: include x in the subset or don't include it. The choice for each element is independent of the choice we make for the other elements. So we have 2^n ways to form a subset and, thus, the powerset $\mathbb{P}(A)$ contains 2^n elements.

Notice that the powerset of A always contains the empty set, regardless of what's in A. As a consequence, $\mathbb{P}(\emptyset) = \{\emptyset\}$.

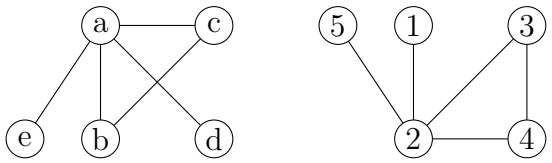
Powersets often appear as the co-domain of functions which need to return a set of values rather than just a single value. For example, suppose that we have the following graph whose set of nodes is $V = \{a, b, c, d, e, f, g, h\}$.



Now, let's define the function N so that it takes a node as input and returns the neighbors of that node. A node might have one neighbor, but it could have several, and it might have no neighbors. So the outputs of N can't be individual nodes. They must be sets of nodes. For example, $N(a) = \{b, c, e\}$ and $N(f) = \emptyset$. It's important to be consistent about the output type of N: it always returns a set. So $N(g) = \{h\}$, **not** $N(g) = h$.

Formally, the domain of N is V and the co-domain is $\mathbb{P}(V)$. So the type signature of N would be $N : V \to \mathbb{P}(V)$.

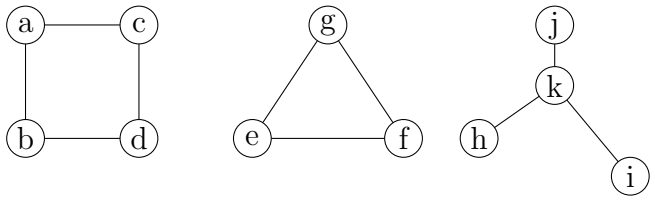
Suppose we have the two graphs shown below, with sets of nodes $X = \{a, b, c, d, e\}$ and $Y = \{1, 2, 3, 4, 5\}$. And suppose that we're trying to find all the possible isomorphisms between the two graphs. We might want a function f that retrieves likely corresponding nodes. For example, if p is a node in X, then $f(p)$ might be the set of nodes in Y with the same degree as p.



f can't return a single node, because there might be more than one node in Y with the same degree. Or, if the two graphs aren't isomorphic, no nodes in Y with the same degree. So we'll have f return a set of nodes. For example, $f(e) = \{1, 5\}$ and $f(a) = \{2\}$. The co-domain of f will need to be $\mathbb{P}(Y)$. So, to declare f, we'd write $f : X \to \mathbb{P}(Y)$.

15.3 Partitions

When we divide a base set A into non-overlapping subsets which include every element of A, the result is called a *partition* of A. For example, suppose that A is the set of nodes in the following graph. The partition $\{\{a, b, c, d\}, \{e, f, g\}, \{h, i, j, k\}\}$ groups nodes into the same subset if they belong to the same connected component.

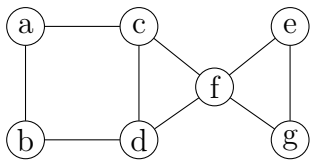


Notice that being in the same connected component is an equivalence relation on the nodes of a graph. In general, each equivalence relation corresponds to a partition of its base set, and vice versa. Each set in the partition is exactly one of the equivalence classes of the relation. For example, congruence mod 4 corresponds to the following partition of the integers:

$$\{\{0, 4, -4, 8, -8, \ldots\}, \{1, 5, -3, 9, -7, \ldots\},$$
$$\{2, 6, -2, 10, -6, \ldots\}, \{3, 7, -1, 11, -5, \ldots\}\}$$

We could also write this partition as $\{[0], [1], [2], [3]\}$ since each equivalence class is a set of numbers.

Collections of subsets don't always form partitions. For example, consider the following graph G.



Suppose we collect sets of nodes in G that form a cycle. We'll get the following set of subsets. This isn't a partition because some of the subsets overlap.

$$\{\{f, c, d\}, \{a, b, c, d\}, \{a, b, c, d, f\}, \{f, e, g\}\}$$

Formally, a partition of a set A is a collection of non-empty subsets of A which cover all the elements of A and which don't overlap. So, if the subsets in the partition are $A_1, A_2, \ldots A_n$, then they must satisfy three conditions:

1. covers all of A: $A_1 \cup A_2 \cup \ldots \cup A_n = A$
2. non-empty: $A_i \neq \emptyset$ for all i
3. no overlap: $A_i \cap A_j = \emptyset$ for all $i \neq j$.

It's possible for a partition of an infinite set A to contain infinitely many subsets. For example, we can partition the integers into subsets each of which contains integers with the same magnitude:

$$\{\{0\}, \{1, -1\}, \{2, -2\}, \{3, -3\}, \ldots\}$$

We need more general notation to cover the possibility of an infinite partition. Suppose that P is a partition of A. Then P must satisfy the following conditions:

1. covers all of A: $\bigcup_{X \in P} X = A$
2. non-empty: $X \neq \emptyset$ for all $X \in P$
3. no overlap: $X \cap Y = \emptyset$ for all $X, Y \in P$, $X \neq Y$

The three defining conditions of an equivalence relation (reflexive, symmetric, and transitive) were chosen so as to force the equivalence classes to be a partition. Relations without one of these properties would generate "equivalence classes" that might be empty, have partial overlaps, and so forth.

15.4 Combinations

In many applications, we have an n-element set and need to count all subsets of a particular size k. A subset of size k is called a k-combination. Notice the difference between a permutation and a combination: we care about the order of elements in a permutation but not in a combination.

For example, how many ways can I select a 7-card hand from a 60-card deck of Magic cards (assuming no two cards are identical)?[1]

One way to analyze this problem is to figure out how many ways we can select an ordered list of 7 cards, which is $P(60, 7)$. This over-counts the number of possibilities, so we have to divide by the number of different orders in which the same 7-cards might appear. That's just 7!. So our total number

[1] Ok, ok, for those of you who actually play Magic, decks do tend to contain identical land cards. But maybe we are using lots of special lands or perhaps we'll treat cards with different artwork as different.

of hands is $\frac{P(60,7)}{7!}$ This is $\frac{60 \cdot 59 \cdot 58 \cdot 57 \cdot 56 \cdot 55 \cdot 54}{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}$. Probably not worth simplifying or multiplying this out unless you really have to. (Get a computer to do it.)

In general, suppose that we have a set S with n elements and we want to choose an unordered subset of k elements. We have $\frac{n!}{(n-k)!}$ ways to choose k elements in some particular order. Since there are $k!$ ways to put each subset into an order, we need to divide by $k!$ so that we will only count each subset once. So the general formula for the number of possible subsets is $\frac{n!}{k!(n-k)!}$.

The expression $\frac{n!}{k!(n-k)!}$ is often written $C(n,k)$ or $\binom{n}{k}$. This is pronounced "n choose r." It is also sometimes called a "binomial coefficient," for reasons that will become obvious shortly. So the shorthand answer to our question about magic cards would be $\binom{60}{7}$.

Notice that $\binom{n}{r}$ is only defined when $n \geq r \geq 0$. What is $\binom{0}{0}$? This is $\frac{0!}{0!0!} = \frac{1}{1 \cdot 1} = 1$.

15.5 Applying the combinations formula

The combinations formula is often used when we want to select a set of locations or positions to contain a specific value. For example, suppose we want to figure out how many 16-digit bit strings contain exactly 5 zeros. Let's think of the string as having 16 positions. We need to choose 5 of these to be the positions containing the zeros. We can how apply the combinations formula: we have $\binom{16}{5}$ ways to select these 5 positions.

To take a slightly harder example, let's figure out how many 10-character strings from the 26-letter ASCII alphabet contain no more than 3 A's. Such strings have to contain 0, 1, 2, or 3 A's. To find the number of strings containing exactly three A's, we first pick three of the 10 positions to contain the A's. There are $\binom{10}{3}$ ways to do this. Then, we have seven positions to fill with our choice of any character except A. We have 25^7 ways to do that. So our total number of strings with 3 A's is $\binom{10}{3}25^7$.

To get the total number of strings, we do a similar analysis to count the strings with 0, 1, and 2 A's. We then add up the counts for the four possibilities to get a somewhat messy final answer for the number of strings with no 3 or fewer A's:

$\binom{10}{3}25^7 + \binom{10}{2}25^8 + \binom{10}{1}25^9 + 25^{10}$

15.6 Combinations with repetition

Suppose I have a set S and I want to select a group of objects of the types listed in S, but I'm allowed to pick more than one of each type of object. For example, suppose I want to pick 6 plants for my garden and the set of available plants is $S = \{\text{thyme}, \text{oregano}, \text{mint}\}$. The garden store can supply as many as I want of any type of plant. I could pick 3 thyme and 3 mint. Or I could pick 2 thyme, 1 oregano, and 3 mint.

There's a clever way to count the possibilities here. Let's draw a picture of a selection as follows. We'll group all our thymes together, then our oreganos, then our mints. Between each pair of groups, we'll put a cardboard separator #. So 2 thyme, 1 oregano, and 3 mint looks like

T T # O # M M M

And 3 thyme and 3 mint looks like

T T T ## M M M

But this picture is redundant, since the items before the first separator are always thymes, the ones between the separators are oreganos, and the last group are mints. So we can simplify the diagram by using a star for each object and remembering their types implicitly. Then 2 thyme, 1 oregano, and 3 mint looks like

** # * # ***

And 3 thyme and 3 mint looks like

*** ## ***

To count these pictures, we need to count the number of ways to arrange 6 stars and two #'s. That is, we have 8 positions and need to choose 2 to fill with #'s. In other words, $\left(\begin{array}{c} 8 \\ 2 \end{array} \right)$.

In general, suppose we are picking a group of k objects (with possible duplicates) from a list of n types. Then our picture will contain k stars and $n-1$ #'s. So we have $k+n-1$ positions in the picture and need to choose $n-1$ positions to contain the #'s. So the number of possible pictures is $\left(\begin{array}{c} k+n-1 \\ n-1 \end{array} \right)$.

Notice that this is equal to $\left(\begin{array}{c} k+n-1 \\ k \end{array} \right)$ because we have an identity that says so (see above). We could have done our counting by picking a subset of k positions in the diagram that we would fill with stars (and then the rest of the positions will get the #'s).

If wanted to pick 20 plants and there were five types available, I would have $\left(\begin{array}{c} 24 \\ 4 \end{array} \right) = \left(\begin{array}{c} 24 \\ 20 \end{array} \right)$ options for how to make my selection. $\left(\begin{array}{c} 24 \\ 4 \end{array} \right) = \frac{24 \cdot 23 \cdot 22 \cdot 21}{4 \cdot 3 \cdot 3} = 23 \cdot 22 \cdot 21$.

15.7 Identities for binomial coefficients

There are a large number of useful identities involving binomial coefficients, most of which you can look up as you need them. Two really basic ones are worth memorizing. First, a simple consequence of the definition of $\binom{n}{k}$ is that

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

Pascal's identity also shows up frequently. It states that

$$\text{(Pascal's identity) } \tbinom{n+1}{k} = \tbinom{n}{k} + \tbinom{n}{k-1}$$

This is not hard to prove from the definition of $\binom{n}{k}$. To remember it, suppose

that S is a set with $n+1$ elements. The lefthand side of the equation is the number of k-element subsets of S.

Now, fix some element a in S. There are two kinds of k-element subsets: (1) those that don't contain a and (2) those that do contain a. The first term on the righthand side counts the subsets in group (1): all k-element subsets of $S - \{a\}$. The second term on the righthand side counts the $k-1$-element subsets of $S - \{a\}$. We then add a to each of these to get the subsets in group (2).

If we have Pascal's identity, we can give a recursive definition for the binomial coefficients, for all natural numbers n and k with $k \leq n$.

Base: For any natural number k, $\binom{n}{0} = \binom{n}{n} = 1$.

Induction: $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$, whenever $k < n$

15.8 Binomial Theorem

Remember that a *binomial* is a sum of two terms, e.g. $(x + y)$. Binomial coefficients get their name from the following useful theorem about raising a binomial to an integer power:

Claim 54 *(Binomial Theorem) Let x and y be variables and let n be any natural number. Then*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$

When we expand the product $(x + y)^n$, each term is the product of n variables, some x's and the rest y's. For example, if $n = 5$, one term is $yxyyx$. So each term is an ordered list of x's and y's.

We can think of our large set of terms as partitioned into subsets, each containing terms with the same number of x's. For example, the set of terms with two x's would be

$$\begin{aligned}[xxyyy] &= \{xxyyy, xyxyy, xyyxy, xyyyx, yxxyy,\\ &\quad yxyxy, yxyyx, yyxxy, yyxyx, yyyxx\}\end{aligned}$$

When we collect terms, the coefficient for each term will be the size of this set of equivalent terms. E.g. the coefficient for x^2y^3 is 10, because $[xxyyy]$ contains 10 elements. To find the coefficient for $x^{n-k}y^k$, we need to count how many ways we can make a sequence of n variable names that contains k y's and $n-k$ x's. This amounts to picking a subset of k elements from a set of n positions in the sequence. In other words, there are $\binom{n}{k}$ such terms.

15.9 Variation in notation

We've used the notation $\mathbb{P}(A)$ for the powerset of A. Another common notation is 2^A.