Permutations Continued

Remember that $S_n$ represents the set of all permutations of \{1, 2, ... n\}

There are some basic facts about $S_n$ that we need to have in hand:

1. If $\pi \in S_n$ and $\sigma \in S_n$ then $\pi \circ \sigma \in S_n$

2. If $\pi \in S_n$ and $\sigma \in S_n$ and $\tau \in S_n$ then $\pi \circ (\sigma \circ \tau) = (\pi \circ \sigma) \circ \tau$

3. If $\pi \in S_n$ then $\pi \circ \iota = \iota \circ \pi = \pi$

4. If $\pi \in S_n$ then $\pi^{-1} \in S_n$ and $\pi \circ \pi^{-1} = \pi^{-1} \circ \pi = \iota$

Property 1 says that the composition of two permutations is another permutation. This sounds trivial but it is our first look at a very important concept: **closure**. When we apply an operation to two elements of a set and **always** get another element of the same set, we say that set is **closed** under that operation.

Not all sets are closed under all operations. For example, $\mathbb{N}$ is not closed under the operation of subtraction (for instance, $3 \in \mathbb{N}$ and $4 \in \mathbb{N}$, but $3 - 4 = -1$, which is not in $\mathbb{N}$). However $\mathbb{N}$ is closed under addition and multiplication. This concept is vital to us as computer scientists because we frequently work with strongly typed programming languages, where each variable has a specific type that cannot change. If we are dealing with integer variables, we need to be sure the operations we perform will only produce integer values.

Property 2 is called the associative property. It says that if we are composing a sequence of permutations we can group them with parentheses in different ways without changing the result. We will see an application of this later in these notes.

Property 3 asserts that the identity permutation $\iota$ can be composed with any permutation
without changing it. Once again, we can draw a parallel to other sets and operations. For example, in the set \( \mathbb{N} \) and the operation of multiplication, we know that 
\[
x \cdot 1 = 1 \cdot x = x \quad \forall \ x
\]

Property 4 asserts that every permutation has an inverse. This property is also true for some sets and operations but not all. For example, in the set \( \mathbb{Z} \) and the operation of addition, the identity element is 0 and every element has an inverse (for instance, the inverse of 7 is -7). However in the set \( \mathbb{N} \) and the operation of addition, the identity element is still 0 but the non-zero values do not have inverses (for instance there is no integer \( x \in \mathbb{N} \) such that \( 8 + x = 0 \)).

Property 4 is particularly important when we use permutations in cryptography – there’s not much point encoding information with a permutation if there is not some other permutation that will do the decoding.

Each of these properties follows from the definition of permutations and the properties of functions. I recommend that you do some examples and convince yourself that these are true.

Note that there is one property that is possessed by many operations that is not true of permutations: commutativity. Commutativity holds when we can switch the left-to-right order of the operands without changing the result. For example, when we are multiplying integers, we know that \( x \cdot y = y \cdot x \) and the same is true for addition. But subtraction is not commutative: \( x - y \neq y - x \) except when \( x = y \).

Composition of permutations is not commutative. In general, \( \pi \circ \sigma \neq \sigma \circ \pi \) although we will see some special cases where they are equal.

In class I posed a challenge: given permutations \( \pi \) and \( \sigma \) in \( S_n \), can you always find a permutation \( \alpha \) such that \( \pi \circ \alpha = \sigma \)?

The answer is yes. Here’s how: we can solve this equation for \( \alpha \) in much the same way as we would solve an equation involving numbers ... we just try to get \( \alpha \) by itself on one side. But here the only operation we have is composition. So what can we do with composition to get rid of the \( \pi \) on the left side? Well, remember that \( \pi^{-1} \circ \pi = l \) ... and \( l \circ \alpha = \alpha \).
So we can start with $\pi \circ \alpha = \sigma$ and apply the following operations that maintain equality

\[
\pi \circ \alpha = \sigma
\]
\[
\pi^{-1} \circ (\pi \circ \alpha) = \pi^{-1} \circ \sigma
\]
\[
(\pi^{-1} \circ \pi) \circ \alpha = \pi^{-1} \circ \sigma
\]
\[
\iota \circ \alpha = \pi^{-1} \circ \sigma
\]
\[
\alpha = \pi^{-1} \circ \sigma
\]

and we are done!

It’s easy to check that this is correct. If we take $\pi \circ \alpha$ and replace $\alpha$ by $\pi^{-1} \circ \sigma$ we get $\pi \circ (\pi^{-1} \circ \sigma)$ which equals $(\pi \circ \pi^{-1}) \circ \sigma$ which equals $\iota \circ \sigma$ which equals $\sigma$

---

**Cycle Notation for Permutations**

Now we introduce another representation for permutations ... one that makes it possible to work with permutations very easily.

Consider this permutation:

\[
\pi = \begin{bmatrix} 4 & 1 & 5 & 2 & 7 & 3 & 6 \end{bmatrix}
\]

What happens if we imagine composing $\pi$ with itself? Let’s trace what happens to the element 1. We are going to apply $\pi$ twice: the first application maps 1 to 4, and the second application maps that 4 to 2. If we compose with $\pi$ again, that 2 is mapped back to 1. Treating $\pi$ as a function, we see $\pi(1) = 4$, $(\pi \circ \pi)(1) = 2$, and $(\pi \circ \pi \circ \pi)(1) = 1$

Composing with $\pi$ even more times will cycle through 4 then 2 then 1 then 4 then 2 then 1 etc. We can write this behaviour as $1 \rightarrow 4 \rightarrow 2 \rightarrow 1 \rightarrow 4 \rightarrow 2 \rightarrow 1$ etc.

If we trace what happens to 2 when we repeatedly compose $\pi$ with itself and apply the resulting function to 2, we see exactly the same pattern: $2 \rightarrow 1 \rightarrow 4 \rightarrow 2 \rightarrow 1 \rightarrow 4 \rightarrow 2$ etc. The same thing happens if we start with 4 and trace what happens to it when we
repeatedly compose $\pi$ with itself and apply the resulting function to 4: we get the pattern
$4 \rightarrow 2 \rightarrow 1 \rightarrow 4 \rightarrow 2 \rightarrow 1 \rightarrow 4 \text{ etc}$

So in this sense, 1, 4 and 2 form a cycle: 1 goes to 4, 4 goes to 2, and 2 goes to 1. We write this
cycle as $(1, 4, 2)$ - it is a notational device that describes the three ordered pairs $(1,4),(4,2),
(2,1)$ which belong to $\pi$

What about the rest of $\pi$? Since we have dealt with 1, 2 and 4, let’s see what happens to 3.
Following the same analysis as we did for 1, 2 and 4 (but skipping over some of the details)
we see this pattern: $3 \rightarrow 5 \rightarrow 7 \rightarrow 6 \rightarrow 3 \rightarrow 5 \rightarrow 7 \rightarrow 6 \rightarrow 3 \text{ etc.}$ which we write as
$(3, 5, 7, 6)$. Again we can see that this is a non-ambiguous shorthand way to represent the
four ordered pairs $(3,5), (5,7), (7,6), (6,3)$ that make up the rest of $\pi$

Thus we can express the entire definition of $\pi$ with the two cycles $(1,4,2)(3,5,7,6)$ - we call this
the cycle notation for $\pi$. All of the information that defines $\pi$ is there, expressed in a different
way (in other words, we can reconstruct the standard representation from the cycle notation).

Notice that from each permutation, we can only get one cycle notation version. (We saw this
for $\pi$ above: no matter which elements we start with, we always get the same repeating
patterns for 1, 4 and 2, and for 3, 5, 7, and 6.) Similarly, from any cycle notation
representation, we can only reconstruct one permutation in standard notation. This means
that the cycle notation for each permutation is unique (up to changing the order of the
cycles, because $(1,4,2)(3,5,7,6)$ gives the same information as $(3,5,7,6)(1,4,2)$ and up to rotating
the elements within each cycle, since $(1,4,2)$ and $(4,2,1)$ and $(2,1,4)$ all represent the same
information).

We can also extract the cycle notation for a permutation $\pi$ without going through the effort of
composing $\pi$ with itself over and over. We can just build the cycles directly from $\pi$ by
observing “1 goes to 4, 4 goes to 2, and 2 goes to 1” to get the cycle $(1, 4, 2)$. Then we can say
“What happens to 3?” and observe “3 goes to 5, 5 goes to 7, 7 goes to 6, and 6 goes to 3” to get
the cycle $(3, 5, 7, 6)$.

We can also represent the permutation by a diagram with each ordered pair represented by
an arrow. This is what we did in class. The diagram for $\pi$ looks like this:
So what can we do with a permutation expressed in cycle notation?

**Computing the Inverse of a Permutation in Cycle Notation**

Suppose we have a permutation \( \pi \) and we need to compute \( \pi^{-1} \). We could do it from standard representation ...

For example, consider \( \pi = [4 \ 1 \ 5 \ 2 \ 7 \ 3 \ 6] \). To compute \( \pi^{-1} \), we could see that “4” is in the first position, so the ordered pair (1,4) must be in \( \pi \) ... which means the ordered pair (4,1) must be in \( \pi^{-1} \). Similarly, “2” is in the fourth position, so the ordered pair (4,2) is in \( \pi \), so (2,4) must be in \( \pi^{-1} \), and so on ... it’s not that hard but it’s a bit tedious.

But with the permutation expressed in cycle notation, computing \( \pi^{-1} \) is trivially easy: we just reverse each cycle. So the inverse of \( (1,4,2)(3,5,7,6) \) is simply \( (2,4,1)(6,7,5,3) \). You can check the details of this example to confirm that it works, but the logic of it is pretty straightforward: \( \pi \) contains the ordered pair (1,4) – this is encoded in the first cycle, so \( \pi^{-1} \) must contain the ordered pair (4,1) – and this is encoded in the reverse of the first cycle.

Note that **reversing** a cycle is very different from **rotating** a cycle: the cycles \( (a, b, c, d) \) and \( (c, d, a, b) \) represent exactly the same information, but \( (d, c, b, a) \) represents the inverse.
Composition of Two Permutations in Cycle Notation

Now suppose we have two permutations \( \pi \) and \( \sigma \) and we want to compute \( \pi \circ \sigma \).
(Remember, this means “the permutation that results when we apply \( \sigma \), then apply \( \pi \)”)
Once again cycle notation makes this very easy, and an example will show how this is done.

Let’s use \( \pi = \begin{bmatrix} 4 & 1 & 5 & 2 & 7 & 3 & 6 \end{bmatrix} \) and \( \sigma = \begin{bmatrix} 4 & 7 & 6 & 1 & 5 & 3 & 2 \end{bmatrix} \).

In cycle notation, \( \pi = (1,4,2)(3,5,6,7) \) and \( \sigma = (1,4)(2,7)(3,6)(5) \) - you should check this.

(Why is 5 all by itself in \( \sigma \)? Because \( \sigma \) maps 5 to itself ... 5 forms a cycle of length 1.)

We can build the cycle notation for \( \pi \circ \sigma \) as follows:

Start with 1. Apply \( \sigma \) to it, giving 4 (that is to say \( \sigma(1) = 4 \)). Then apply \( \pi \) to that 4, giving 2 (that is, \( \pi(4) = 2 \)). So in \( \pi \circ \sigma \), we see that \( 1 \to 2 \) (that is, \( (\pi \circ \sigma)(1) = \pi(\sigma(1)) = \pi(4) = 2 \)).

So our first cycle in \( \pi \circ \sigma \) starts (1, 2 ...)

Now let’s see what \( \pi \circ \sigma \) does to 2. \( \sigma \) takes 2 to 7, and \( \pi \) takes 7 to 6. So \( (\pi \circ \sigma)(2) = 6 \).

So our cycle in \( \pi \circ \sigma \) now looks like (1, 2, 6 ...)

Now let’s see what \( \pi \circ \sigma \) does with 6. \( \sigma \) takes 6 to 3 and \( \pi \) takes 3 to 5, so \( (\pi \circ \sigma)(6) = 5 \).

The cycle in \( \pi \circ \sigma \) we are building now looks like (1, 2, 6, 5 ...)

Following the same steps we see that \( (\pi \circ \sigma)(5) = 7 \). Then we discover that \( (\pi \circ \sigma)(7) = 1 \).

So we have discovered that in \( \pi \circ \sigma \), \( 1 \to 2 \to 6 \to 5 \to 7 \to 1 \) is a cycle, and in cycle notation this is \((1,2,6,5,7)\)

We still haven’t dealt with 3 and 4. It doesn’t matter which we start with, so let’s start with 4. \( \sigma(4) = 1 \) and \( \pi(1) = 4 \), so \( (\pi \circ \sigma)(4) = 4 \). In cycle notation this is just (4)

You can verify for yourself that \( (\pi \circ \sigma)(3) = 3 \), giving (3) as the last cycle in \( \pi \circ \sigma \)

Putting all the cycles together, we get \( \pi \circ \sigma = (1,2,6,5,7)(4)(3) \)

You should check this too!
Shorthand Notation for Cycles

When using cycle notation, sometimes we leave out the cycles of length 1. This can be ambiguous unless we know we are dealing with permutations from a particular $S_n$.

In the previous example, if we specify that $\pi$ and $\sigma$ are in $S_7$ then we can write $\pi \circ \sigma = (1, 2, 6, 5, 7)$ and just leave out the (4) and (3). The reader will know they are cycles of length 1 because if they weren’t we would have included them.

Transpositions

A transposition is a permutation that, when written in cycle notation, has exactly one cycle of length 2, and $n-2$ cycles of length 1.

For example $\pi = (1)(2)(3,7)(4)(5)(6)(8)$ is a transposition.

If we know $\pi \in S_8$, we can simply write it as (3,7) using the shorthand notation mentioned above. This is the normal way of writing a transposition.

Note that the effect of a transposition is to simply swap the positions of two numbers and leave all the others where they are. Using $\pi$ from the example just given, $\pi(3) = 7$ and $\pi(7) = 3$ and $\forall$ other $x$, $\pi(x) = x$.

One interesting property of transpositions is that every transposition is its own inverse! This is pretty easy to see – if $\pi$ just swaps the values of $x$ and $y$, then swapping them again will put them back where they started. So $\pi \circ \pi = I$ … which means $\pi^{-1} = \pi$.

Decomposing any Permutation into the Composition of a Sequence of Transpositions

This is pretty much our last topic in this introduction to permutations. I’m going to introduce a technique that creates yet another representation of a permutation – you may feel that we already have more than we need, but I promise you they are all useful in different scenarios.

I’m not going to prove that this technique always works (I hope you will trust me for now, and perhaps try to prove it on your own when you have a few spare minutes). When you see how it works, you will probably see how it could be proved.
Let’s consider a permutation $\pi \in S_4$ with $\pi = (1, 3, 2, 4)$

Now consider the permutation $\sigma = (1, 3) \circ (3, 2) \circ (2, 4)$

What the heck is that? Well, each bracketed pair is a transposition, and we are composing them. Since each transposition is a permutation, composing them together gives a new permutation. Note that each transposition is created from two of the consecutive elements in the cycle. Remember that cycle notation is just shorthand for standard notation, so the definition of $\sigma$ is equivalent to $\sigma = [3 \ 2 \ 1 \ 4] \circ [1 \ 3 \ 2 \ 4] \circ [1 \ 4 \ 3 \ 2]$

I claim that $\sigma = \pi$

Let’s just check it out.

Remember that if a transposition is applied to a number which is not one of the two in the transposition, it leaves it alone.

\[
\begin{align*}
\pi(1) &= 3 \quad \text{and} \quad \sigma(1) = [3 \ 2 \ 1 \ 4] \circ [1 \ 3 \ 2 \ 4] \circ [1 \ 4 \ 3 \ 2] \times (1) \\
&= [3 \ 2 \ 1 \ 4] \circ [1 \ 3 \ 2 \ 4] \times (1) \\
&= [3 \ 2 \ 1 \ 4] \times (1) \\
&= 3 \\

\pi(2) &= 4 \quad \text{and} \quad \sigma(2) = [3 \ 2 \ 1 \ 4] \circ [1 \ 3 \ 2 \ 4] \circ [1 \ 4 \ 3 \ 2] \times (2) \\
&= [3 \ 2 \ 1 \ 4] \circ [1 \ 3 \ 2 \ 4] \times (4) \\
&= [3 \ 2 \ 1 \ 4] \times (4) \\
&= 4 \\

\pi(3) &= 2 \quad \text{and} \quad \sigma(3) = [3 \ 2 \ 1 \ 4] \circ [1 \ 3 \ 2 \ 4] \circ [1 \ 4 \ 3 \ 2] \times (3) \\
&= [3 \ 2 \ 1 \ 4] \circ [1 \ 3 \ 2 \ 4] \times (3) \\
&= [3 \ 2 \ 1 \ 4] \times (2) \\
&= 2 \\

\pi(4) &= 1 \quad \text{and} \quad \sigma(4) = [3 \ 2 \ 1 \ 4] \circ [1 \ 3 \ 2 \ 4] \circ [1 \ 4 \ 3 \ 2] \times (4) \\
&= [3 \ 2 \ 1 \ 4] \circ [1 \ 3 \ 2 \ 4] \times (2) \\
&= [3 \ 2 \ 1 \ 4] \times (3) \\
&= 1
\end{align*}
\]
So by examining all cases, we see $\sigma = \pi$

Now if $\pi$ has more cycles, they operate completely independently of each other so we can decompose each of them into transpositions, and then just compose the results together.

For example, if $\pi = (1, 6, 5) (2, 4, 3)$ then

$$\pi = (1, 6) \circ (6, 5) \circ (2, 4) \circ (4, 3)$$

For most of the representations of permutations that we have seen, the representation of each permutation is unique (within some constraints, at least). But for the transposition decomposition of a permutation, this is not true. For many permutations, we can decompose them into different sequences of transpositions. For example, the permutation $\pi$ in the last example can also be expressed as $\pi = (1, 5) \circ (1, 6) \circ (2, 3) \circ (2, 4)$

In fact the number of transpositions in the decomposition of a permutation is not even fixed. We can have a permutation that can be decomposed into 4 transpositions, and also decomposed into 6 transpositions, or 8 or 12. What we do know is that for any permutation, either all of its decompositions will have an even number of transpositions, or all of its decompositions will have an odd number of transpositions.
Introduction to Probability

In casual conversation, people tend to use the concepts of probability quite loosely. We are going to give precise definitions and formulas to give probability theory a firm footing in discrete mathematics.

Sample Spaces

Probability theory is based on the idea of observing the outcome of an experiment that has different possible outcomes. We call this process of experimenting and observing sampling. The type of experiment we are usually talking about here is one with a fixed set of possible outcomes, such as flipping a coin, tossing a six-sided die, or picking one ball out of a box of different coloured balls.

A sample space consists of the set of possible outcomes of an experiment, and a function \( P(\cdot) \) that assigns a value to each outcome. \( P(\cdot) \) must have the following properties:

\[
0 \leq P(x) \leq 1 \quad \text{for each outcome } x
\]

\[
\sum_{x \in S} P(x) = 1
\]

We call \( P(\cdot) \) a probability function.

We will sometimes write \((S,P)\) to identify the sample space where \( S \) is the set of outcomes and \( P \) is the probability function.

Any function that satisfies the requirements is a valid probability function, but we usually want our probability functions to correspond to the “likelihood” of the different outcomes occurring. But that is dangerously close to a circular definition, since “likelihood” is often used as a synonym for “probability”.

We can get a concrete sense of what we want our probability functions to do by considering an experiment which we sample many, many times. If we count the number of times a particular outcome occurs and divide that by the number of samples, we expect that this ratio will change less and less as the number of samples increases. The limit of this ratio as the number of samples goes to \( \infty \) is what we want as the probability of that particular outcome. However we want to want to do this without working out limits because that’s calculus!

For example, consider a box containing a red ball, a yellow ball and a blue ball. If the balls are all identical in size and weight, we expect that if we take out one ball, record its colour and return it to the box, over and over again, the number of times we withdraw the red ball divided by the total number of samples, will get closer and closer to \( 1/3 \) ... as will the ratios for the yellow ball and the blue ball.

Now suppose the box contains 2 red balls and 1 yellow ball. The possible outcomes from our sampling experiment are \{red, yellow\}. Again assuming that the balls are identical except for their colours, we expect that the ratio for red (the ratio of occurrences / samples) will approach \( 2/3 \), while the ratio for yellow approaches \( 1/3 \).
So in general, we want $P(x)$ to be the ratio “occurrences of $x$”/”number of samples” as the number of samples goes towards $\infty$.

**Events**

If $(S,P)$ is a sample space, we use the word *event* to describe any subset of $S$. For example, if $S = \{1,2,3,4,5\}$ then $A = \{1,3,5\}$ is an event. $\emptyset$ is also an event, and so is $S$.

Suppose we sample $(S,P)$ by conducting the experiment once. If the outcome of the sample is an element of $A$, we say that $A$ has occurred.

Now we can define the **probability of an event**: the probability of event $A$ is the sum of the probabilities of the elements of $A$. In notation

$$P(A) = \sum_{a \in A} P(a)$$

For example, given $S$ as above, suppose $P(1) = P(2) = P(3) = P(4) = P(5) = 0.2$

Then $P(\{1,3,5\}) = 0.6$

But suppose $P(1) = 0.4$, $P(2) = 0.1$, $P(3) = 0.3$, $P(4) = 0.1$, $P(5) = 0.1$

Then $P(\{1,3,5\}) = 0.8$

In casual discussions of probability, people often forget that the probability of an event depends on the probability function – they assume (without cause) that all outcomes are equally probable. A very simple example is when the experiment consists of rolling two fair 6-sided dice and adding the numbers that come up. The possible outcomes are $\{2,3,4,...,12\}$ but the probabilities are not all equal. So the probability of the event $\{6,7,8\}$ is very different from the probability of the event $\{2,3,4\}$ even though both contain the same number of outcomes.

To be honest, the (unjustified) assumption that all outcomes of an experiment are equally probable sometimes shows up in scientific discussions as well.
Combinations of Events

Let (S,P) be a sample space, and let A and B be events in that sample space.

What can we say about $A \cup B$? More precisely, can we compute $P(A \cup B)$ from $P(A)$ and $P(B)$?

Unfortunately $P(A)$ and $P(B)$ do not give enough information to compute $P(A \cup B)$. Consider this example. Let the experiment be tossing a single 6-sided die. We will assume that all outcomes are equally probable (i.e. $P(i) = \frac{1}{6} \ \forall \ i \in \{1,2,3,4,5,6\}$)

Let $A = \{1,2,3\}$ $B = \{2,3,4\}$ $C = \{4,5,6\}$

$P(A) = P(B) = P(C) = \frac{1}{2}$, but $P(A \cup B) = P(\{1,2,3,4\}) = \frac{2}{3}$ while $P(A \cup C) = 1$

In the first case two events with individual probabilities = 1/2 have a combined probability = 2/3, and in the second case two events with individual probabilities = 1/2 have a combined probability = 1

The difference of course is that A and B overlap (i.e. they have non-empty intersection) while A and C do not overlap ... and we can’t tell that just by looking at their probabilities. Fortunately the solution is obvious as soon as we recognize the problem. We just apply our old friend the Principle of Inclusion/Exclusion, and arrive at this formula:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Note that this means the only time $P(A \cup B) = P(A) + P(B)$ is when $P(A \cap B) = 0$

Also observe that using this equation, if we know any three of the terms we can deduce the fourth. For example if we know $P(A \cup B) = 0.7$, $P(A) = 0.3$, $P(A \cap B) = 0.2$, then we know $P(B) = 0.6$

Here are some more useful facts about the probabilities of events

$P(\emptyset) = 0$

$P(S) = 1$

$P(A \cup B) \leq P(A) + P(B)$

$P(\overline{A}) = 1 - P(A)$ where $\overline{A}$ represents the complement of A