Lecture outline: strings and permutations

We introduce the formal definitions of strings and permutations and various important related concepts such as prefix, suffix, substring, ...

We review the basic counting results that should already be known and that involve three kinds of numbers: exponential numbers, factorial numbers and binomial numbers (and the binomial theorem).

We see how strings can be used to model/encode other kinds of combinatorial objects such as lattice paths or integer compositions.

Strings

Definition. A **string** S, of size n, over an alphabet $\mathcal{A}$ is a linearly ordered list of n elements (**letters**) taken from $\mathcal{A}$, possibly with repetition.

Notation. For a string S of length n, we denote by $S[1]$ its first letter, $S[2]$ its second letter, ..., $S[n]$ its last letter.

Remark. Strings are sometimes called sequences or words. We will always use strings. The size of a string is also called its **length**.

Remark. The important property of strings is the total/linear order on the atoms (symbols from the alphabet): in a string, there is a first symbol, a second symbol, ..., a last symbol.

Definition. Let S be a string of size n.

For any $1 \leq i \leq j \leq n$, the string $S[i]\ S[i+1] \ldots S[j-1]\ S[j]$ is called a **substring** of S. It is denoted by $S[i,j]$

If $i = 1$, then $S[1,j]$ is a **prefix** of S.

If $j = n$, then $S[i,n]$ is a **suffix** of S.

A substring that is neither a prefix nor a suffix is a **proper substring**.

1 2 3 4 5 6 7 8 9

Example. $S = \underbrace{0\ 1\ 1\ 0}_{\text{prefix}}\ 0\ \underbrace{0\ 1\ 0\ 1}\ 0\ \underbrace{1\ 1\ 0}_{\text{suffix}}$

$S[6,9] = 0101$

$S[3] = 1$

We now review the first, basic counting result.

Theorem (review). If $\mathcal{A}$ is an alphabet of k symbols, there are

$$k^n$$

strings of size n over $\mathcal{A}$, for $n \geq 0$.

Notation. The set of all strings over $\mathcal{A}$ is often denoted by $\mathcal{A}^*$. The set of all strings of size n is denoted by $\mathcal{A}^*_n$.

Remark. The result above implies there is one string of size 0. This is the **empty string**, that contains no symbol, and is often denoted by ϵ.

Proof. If we want to generate a arbitrary ~~random~~ string S of length n over $\mathcal{A}$, we need to chose one symbol for each of the n positions of S.
There are k choices for each $S[i]$, $i \in [1, n]$ as the choice made for a given position does not influence the choices that can be made for the other positions.

Examples. $\mathcal{A} = \{0, 1\}$ (binary strings)

0110111110101

$\mathcal{A} = \{a, b\}$ (binary strings again, the alphabet is of size 2)

ababbba

$\mathcal{A} = \{A, C, G, T\}$ (DNA strings)

$\mathcal{A} = \{0, 1, 2, \ldots, 9\}$ (how many PIN)

Factorial numbers, Binomial numbers and the Binomial Theorem

Factorial numbers (review). For $n \geq 1$,

$$n! = n(n-1)(n-2)\cdots 2 \cdot 1$$

For $n = 0$, we define $0! = 1$

Remark. $n!$ is the number of ways to arrange n pair-wise distinct elements (permutations as we will soon see).

Example.

$A = \{a, b, c\}$

perm of A : abc, acb
bac, bca
cab, cba

$\left.\begin{matrix} abc, acb \\ bac, bca \\ cab, cba \end{matrix}\right\} 6 = 3!$

Binomial numbers (review). For $n \geq 0,\ n \geq k \geq 0$,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

If $k < 0$ or $k > n$,

$$\binom{n}{k} = 0$$

Remark. $\binom{n}{k}$ is the number of ways to select a subsets of k elements out of a set of n pair-wisely distinct elements.

Remark. It should be obvious, both from the formula and from the combinatorial interpretation that

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

Example.

A : {a, b, c, d, e}

choose subset of size 2

ab, ac, ad, ae
bc, bd, be
cd, ce
de

$\left.\right\} 10 = \binom{5}{2}$

We use the binomial numbers to count strings with some prescribed properties.

Theorem. Let $\mathcal{A}$ and $\mathcal{B}$ be two alphabets with no common letters, of respective sizes a and b.
Let $\mathcal{C} = \mathcal{A} \cup \mathcal{B}$ be the alphabet obtained by the union of $\mathcal{A}$ and $\mathcal{B}$.
The number of strings of size n over $\mathcal{C}$ with exactly k symbols from $\mathcal{A}$ and $n-k$ symbols form $\mathcal{B}$ is given by

$$\binom{n}{k} a^k b^{n-k}$$

Proof. To construct a string of size n over $\mathcal{C}$, one needs to

- chose which positions will receive symbols from $\mathcal{A}$, and there are $\binom{n}{k}$ choices,
- for each of these k chosen positions we need to choose one of the k symbols of $\mathcal{A}$,
- for each of the remaining $n-k$ chosen positions we need to choose one of the ℓ symbols of $\mathcal{B}$.

Example. $\mathcal{A} = \{0, 1\}$, $\mathcal{B} = \{a, b, c\}$, $n = 4$ and $k = 3$

$C = \{0,1,a,b,c\}$

want strings of size 4 with 3 elts from A
1 elt from B

thm says: the number of strings of this type is $\binom{4}{3}\cdot 2^3 \cdot 3^1 = 96$

8 ways to fill in

a _ _ _	_ a _ _	_ _ a _	_ _ _ a
b _ _ _	_ b _ _	_ _ b _	_ _ _ b
c _ _ _	_ c _ _	_ _ c _	_ _ _ c

Using the previous counting result, we can understand (almost prove) combinatorially a classical result you probably know and that will be useful later.

Binomial Theorem. If x and y are two variables and n is a positive integer:

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

Proof idea. If x and y are assumed to be positive integers, then $(x+y)^n$ is the number of strings of size n over an alphabet $\mathcal{C}$ obtained by the union of two completely distinct alphabets $\mathcal{A}$ and $\mathcal{B}$ of respective sizes x and y.

For any such string, denote by k its number of symbols from $\mathcal{A}$ (so it has $n-k$ symbols from $\mathcal{B}$). k ranges from 0 (no symbol from $\mathcal{A}$) to n (no symbol from $\mathcal{B}$).

In other words, combinatorially, this identity just states the obvious fact that the set of all strings of size n over $\mathcal{C}$ is the union, for $k=0$ to n, of the sets of all strings of size n over $\mathcal{C}$ that contains exactly k symbols from $\mathcal{A}$.

This proves, combinatorially, the result in the case of x and y being integers, which will be enough for us here. The general proof is available on pages 21-22 of your textbook (**please, read it**).

$$(x+y)^n = \underbrace{(x+y)(x+y)(x+y)\cdots(x+y)}_{n}$$

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

Correspondence with sequences over $\{x,y\}$ of length n

Example. For binary strings, this gives

$$2^n = (1+1)^n = \sum_{k=0}^{n} \binom{n}{k} = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n}$$

2^0 1

2^1 1 1

2^2 1 2 1 $\binom{2}{0}$ $\binom{2}{1}$ $\binom{2}{2}$

2^3 1 3 3 1

2^4 1 4 6 4 1

1 5 10 10 5 1

Example. What is $\binom{n}{0} - \binom{n}{1} + \binom{n}{2} + \cdots + (-1)^n \binom{n}{n} = \sum_{k=0}^{n} (-1)^k \binom{n}{k}$?

$0 = (1-1)^n$ //