

Math 55 - Spring 2004 - Lecture notes # 15 - March 11 (Thursday)

Finish reading Chapter 4

Read sections 1-5 of Lenstra's notes on the web

Goals for Today: Finish counting principles

Comments on Homework questions 3.4-50 and 3.4-62:

Where do

Ackermann's Function  $A(m,n)$

iterated logarithm function  $\log^*(n)$

arise? Both come up in the analysis of an important algorithm for building a search tree (the Union-Find algorithm) discussed in CS170. Here is some background.

The most prominent property of  $A(n,n)$  (both arguments the same) is that it grows incredibly fast. For example

$$A(2,2) = 4, A(3,3) = 2^{16},$$

$$A(3,4) = 2^{(2^{(\dots)})} \text{ with a tower of } 2^{16} \text{ 2s (question 3.4-52).}$$

$$A(4,4) \text{ is much larger still than } A(3,4).$$

This means that  $A(n,n)$ 's inverse function

$$a(n) = \text{smallest } m \text{ such that } n \leq A(m,m)$$

grows incredibly slowly, for example

$$a(\# \text{ atoms in the universe}) = a(\sim 10^{80}) = 4$$

but it does go to infinity as  $n$  grows.

Similarly,  $\log^*(n)$  also grows incredibly slowly:

$$\log^*(\# \text{ atoms in the universe}) = 5$$

Both these functions arise in the study of how fast the Union-Find algorithm runs. This algorithm takes  $n$  data items and search requests, and builds a search tree to find the items quickly. It is clear that any algorithm must take at least  $n$  steps, which is needed just to look at each input once. It can be shown that building and searching the Union-Find search tree takes  $O(n \cdot a(n))$  time, i.e. just barely more than linear time  $O(n)$ . But since  $a(n)$  grows so slowly, as long as your tree has fewer items in it than there are atoms in the universe (this seems likely, unless you borrowed atoms from some other universe to build your computer :)), then  $n \cdot a(n) \leq 4 \cdot n$ .

But the proof that Union-Find algorithm take  $O(n \cdot a(n))$  time is very difficult (due to Tarjan). Instead in CS 170 one proves the easier result that the time is bounded by  $O(n \cdot \log^*(n))$ . This is still pretty good, since  $n \cdot \log^*(n) \leq 5 \cdot n$  for  $n$  no larger than the number of atoms in the universe.

Comment on Homework: To get a closed form solution of

$$g(1)=7, g(2)=8, g(n) = 2 \cdot g(n-1) - 2 \cdot g(n-2)$$

a good answer is

$$g(n) = .5 \cdot ( (3-4i) \cdot (1+i)^n + (3+4i) \cdot (1-i)^n )$$

where  $i = \sqrt{-1}$ . If you evaluate this complex number, the imaginary part cancels. It is possible to get a different expression without complex numbers, but this is not necessary.

EX: How many ways can a class of  $n$  students be divided into  $m$  person teams? We assume  $n = q \cdot m$ , so there are  $q$  teams.

We will first solve this problem in a straightforward way, getting a complicated answer as a product of combinatorial coefficients

$$C(i,j) = i! / (j! (i-j)!)$$

Then we will simplify this expression: most of the factorials will "miraculously" cancel after a great deal of algebra, leaving a very simple answer. This very simple answer will suggest a new, clever way to derive it in just a few lines.

The point here is that many counting problems often have multiple ways to get the answer, and there is sometimes a clever short solution.

Solution 1: This is straightforward, but leads to a complicated looking answers that we will have to simplify.

Let  $T(n,m)$  = number of ways to divide  $n$  students into  $m$  person teams.

Here is a recursive way to write down all possible  $m$  person teams:

- 1) Fix student #1 to be a member of team #1
- 2) From the remaining  $n-1$  students, choose  $m-1$  of them to also be members of team #1. There are  $C(n-1, m-1)$  ways to do this.
- 3) This leaves  $n-m$  students, to be divided into teams of  $m$  students.

Applying the product rule to this recursive way of generating teams yields

$$\begin{aligned} T(n,m) &= (\# \text{ ways to do steps 2) } * (\# \text{ ways to do step 3) } \\ &= C(n-1, m-1) * T(n-m, m) \end{aligned}$$

This can be applied recursively, yielding

$$\begin{aligned}
 T(n,m) &= C(n-1,m-1)*T(n-m,m) \\
 &= C(n-1,m-1)*C(n-m-1,m-1)*T(n-2*m,m) \\
 &= C(n-1,m-1)*C(n-m-1,m-1)*C(n-2*m-1,m-1)*T(n-3*m,m) \\
 &= \dots \\
 &= C(n-1,m-1)*C(n-m-1,m-1)*C(n-2*m-1,m-1)*C(n-3*m-1,m-1)*\dots \\
 &\quad *C(m-1,m-1) \\
 &= \text{prod}_{\{i=0 \text{ to } q-1\}} C(n-i*m-1,m-1)
 \end{aligned}$$

Substituting in the definition of  $C(j,k)$  yields a large number of factorials, but the numerator and denominator of adjacent terms turn out to cancel: Writing

$$T(n,m) = \frac{(n-1)!}{(m-1)!} \frac{(n-m-1)!}{(n-m)!} \frac{(n-2*m-1)!}{(m-1)!} \frac{(n-3*m-1)!}{(n-3*m)!} \dots \frac{(m-(q-1)*m-1)!}{(m-1)!} \frac{0!}{0!}$$

we we see that

- the  $(n-m)!$  in the 1st denominator and  $(n-m-1)!$  in the 2nd numerator mostly cancel
- the  $(n-2*m)!$  in the 2nd denominator and  $(n-2*m-1)!$  in the 3rd numerator mostly cancel
- the  $(n-3*m)!$  in the 3rd denominator and  $(n-3*m-1)!$  in the 4th numerator mostly cancel
- ...

leaving

$$\begin{aligned}
 T(n,m) &= \frac{(n-1)!}{(m-1)!} \frac{1}{(n-m)} \frac{1}{(m-1)!} \frac{1}{(n-2*m)} \frac{1}{(m-1)!} \frac{1}{(n-3*m)} \dots \frac{1}{(m-1)!} \frac{1}{1} \\
 &= \frac{(n-1)!}{((m-1)!)^q (n-m)(n-2*m)(n-3*m)\dots(n-(q-1)*m)} \\
 &= \frac{(n-1)!}{((m-1)!)^q (q*m-m)(q*m-2*m)(q*m-3*m)\dots(q*m-(q-1)*m)} \\
 &= \frac{(n-1)!}{((m-1)!)^q (q-1)! m^{(q-1)}}
 \end{aligned}$$

$$\begin{aligned}
&= \frac{(n-1)!}{((m-1)!)^q (q-1)! m^{(q-1)}} * \frac{n}{q^m} \\
&= \frac{n!}{(m!)^q q!}
\end{aligned}$$

which is a much simpler expression. So simple in fact that it calls out for a simpler way to get it, which we now describe.

Solution 2: Let us start to write down all the ways of dividing all the students into  $q$  teams of  $m$  students each by writing down all  $n!$  ways of ordering  $n$  students, and just saying the first  $m$  students are the first team, the 2nd  $m$  students are the second team and so on.

But it is clear that we have counted the same sets of teams too many times. Let us try to divide out  $n!$  by the number of multiple copies of the same set of teams.

First, it is clear that no matter what the order of the first  $m$  students in the list is, we get the same team. Since there are  $m!$  such orders, we have counted sets of teams which differ only in the order of the first teams  $m!$  times too often, so we should divide by  $m!$ .

Similarly, the order of the 2nd group of  $m$  students does not matter, so we should divide by  $m!$  again. The same argument applies to each of the  $q$  groups of  $m$  students, so we should divide by  $m!^q$  times.

But we are still not done, because the team consisting of the first  $m$  students could appear anywhere in the  $q$  possible positions, as could the second group of  $m$  students, and so on. In other words, we have still counted the same set of teams  $q!$  times too often, because the teams can appear in  $q!$  possible orders, and still represent the same set of teams. So we have to divide by  $q!$  also.

All in all, we get  $n! / ((m!)^q q!)$ , which is the same answer we got before (whew!).

EX: How many different desserts can you make out of 4 scoops of ice cream, each of which may be chocolate (C), vanilla (V) or strawberry (S)?

Here are the 15 possibilities:

CCCC VVVV SSSS

CCCV VVVC SSSC  
 CCCS VVVS SSSV  
 CCVS VVCS SSCV  
 CCVV VVSS  
 CCSS

Here is a more systematic way to get the answers: we will represent each dessert by a sequence of 4 stars (representing the 4 scoops) and 2 bars (dividing the stars into 3 groups: C, V and S). Here are some examples:

\*\*|\*|\* represents 2 Cs, 1 V and 1 S  
 \*|\*\*|\* represents 1 C, 2 V's and 1 S  
 \*|\*\*\*| represents 1 C, 3 V's and 0 S's  
 |\*\*\*\*| represents 0 C, 4 V's and 0 S's  
 ||\*\*\*\* represents 0 C, 0 V's and 4 S's etc

The idea is that every sequence of 4 stars and 2 bars represents exactly one dessert. How many such sequences are there? The idea is that we take 6 possible possible positions (for 4 stars and 2 bars) and choose 2 of them for bars. There are  $C(6,2) = 6!/(2! 4!) = 15$  ways to do this.

Here is the general result:

Theorem: Suppose I have  $n$  types of objects ("flavors"). How many different sets ("desserts") consisting of  $r$  objects ("scoops") are there?

The answer is  $C(n+r-1, n-1)$ .

Proof: The idea is the same as before: each sequence of  $r$  stars ("scoops") and  $(n-1)$  bars represents a possible set. There are  $C(n+r-1, n-1)$  ways to pick  $n-1$  places out of  $r+n-1$  locations to put the bars.

Ex: If I have  $n=3$  flavors of ice cream, and make desserts of  $r=4$  scoops, there are  $C(n+r-1, n-1) = C(3+4-1, 3-1) = C(6, 2) = 15$  different desserts.

EX: How many anagrams are there of the word "mammal"?

Recall that an anagram is a distinct ordering of the letters.

Here are some smaller examples:

the word "the": The 6 anagrams are the, teh, eth, eht, het, hte  
 the word "see": The 3 anagrams are see, ese, ees

Here are different ways to try to solve this problem for the word "mammal", followed by the general result:

Solution 1: Pick 3 locations for the m's  
           Pick 2 of the remaining locations for the 2 a's  
           Pick the remaining location for l

By the product rule, the number of ways to pick locations is

$$\begin{aligned}
& C(6,3) \quad \dots \text{ for the m's} \\
& * C(3,2) \quad \dots \text{ for the a's} \\
& * C(1,1) \quad \dots \text{ for the l} \\
& = 20*3*1 = 60
\end{aligned}$$

Solution 2: Pick 1 location for the l  
Pick 3 of the remaining locations for the m's  
Pick the remaining 2 locations for the a's

By the product rule, the number of ways to pick locations is

$$\begin{aligned}
& C(6,1) \quad \dots \text{ for the l} \\
& * C(5,3) \quad \dots \text{ for the m's} \\
& * C(2,2) \quad \dots \text{ for the a's} \\
& = 6*10*1 = 60, \text{ the same answer (whew!)}
\end{aligned}$$

Solution 3: Let us start by labeling the m's as m1,m2 and m3, and and the a's as a1 and a2, so we can distinguish them. So now we have 6 distinct symbols, m1,a1,m2,m3,a2,l, and the number of ways to order them is 6!. But clearly we have counted some ordering as distinct that we should not, so let's try to divide out by the number of multiple copies.

For example, consider all the orderings where the first 3 characters are m's, and the last three are a1,a2,l. There are clearly 3! = 6 such orderings, since m1,m2,m3 can appear in the first three positions in any order, but yield the same anagram. This argument that we are counting each anagram 3! times works no matter where the 3 m's appear, so we should divide the number of orderings by 3! to account for the 3 m's.

Similarly, we should divide by 2! to account for the two a's.

This yields  $6! / (3! 2!) = 60$ , the same answer (whew!)

Solution 3 is the one that generalizes to arbitrary anagrams:

Theorem: Suppose we have  
 $n(1)$  copies of symbol 1

$n(2)$  copies of symbol 2  
 ...  
 $n(k)$  copies of symbol  $k$   
 Let  $n = n(1) + n(2) + \dots + n(k)$ . Then the number of  
 distinct anagrams of these symbols is

$$\frac{n!}{n(1)! n(2)! n(3)! \dots n(k)!}$$

Proof 1: (hard way): We generate the anagrams as follows:

Choose  $n(1)$  positions out of  $n$  for symbol 1  
 (there are  $C(n, n(1))$  ways to do this)  
 Choose  $n(2)$  positions of the remaining  $n-n(1)$  for symbol 2  
 (there are  $C(n-n(1), n(2))$  ways to do this)  
 Choose  $n(3)$  positions of the remaining  $n-n(1)-n(2)$  for symbol 3  
 (there are  $C(n-n(1)-n(2), n(3))$  ways to do this)  
 ...  
 Choose  $n(k)$  positions of the remaining  $n-n(1)-n(2)-\dots-n(k-1)$   
 for symbol  $k$   
 (there is  $C(n-n(1)-n(2)-\dots-n(k-1), n(k)) = C(n(k), n(k))=1$   
 way to do this)

By the product rule we multiply these together to get

$$C(n, n(1)) * C(n-n(1), n(2)) * C(n-n(1)-n(2), n(3)) * \dots * C(n(k), n(k))$$

or

$$\frac{n!}{n(1)! (n-n(1))!} * \frac{(n-n(1))!}{n(2)! (n-n(1)-n(2))!} * \frac{(n-n(1)-n(2))!}{n(3)! (n-n(1)-n(2)-n(3))!} * \dots$$

We can cancel the numerator in the  $i$ -th term with a factor of the denominators from the  $(i-1)$ -st term, leaving

$$\frac{n!}{n(1)!} * \frac{1}{n(2)!} * \frac{1}{n(3)!} * \dots$$

as desired.

Proof 2 (easy way): Consider all  $n!$  permutations of the  $n$  symbols.

Some of these are identical:

Given a permutation, all  $n(1)!$  permutations with symbol 1 in the same positions are identical

Given a permutation, all  $n(2)!$  permutations with symbol 2 in the same positions are identical

...

Given a permutation, all  $n(k)!$  permutations with symbol  $k$  in the same positions are identical

Therefore, we need to divide  $n!$  by  $n(1)! * n(2)! * \dots * n(k)!$  to get the correct number of anagrams.