Problem Set 6 Solutions

Problems:

Problem 1 Determine the flaw in the following reasoning and explain your answer.

$$\left(\sum_{j=1}^{n} a_j\right)\left(\sum_{k=1}^{n} 1/a_k\right) = \sum_{j=1}^{n}\sum_{k=1}^{n} a_j/a_k = \sum_{k=1}^{n}\sum_{k=1}^{n} a_k/a_k = \sum_{k=1}^{n}\sum_{k=1}^{n} 1 = \sum_{k=1}^{n} n = n^2$$

Solution.

Here, the problem is taking the second step. It is incorrect to change the variable of the first summation to k since the second summation used that variable. By doing this change of variables, we change the meaning of the double summation.

□

Problem 2 Find a closed-form expression for the summation $\sum_{i=1}^{n} i^2 x^i$.

Solution.

We first come up with a formula for $\sum_{i=1}^{n} i^2 x^i$ and then prove it using induction. From class notes we know that:

$$\sum_{i=1}^{n} i x^i = \frac{x - (n+1)x^{n+1} + nx^{n+2}}{(1-x)^2}.$$

Differentiating (with respect to x) the above expression and multiplying with x we find that

$$\sum_{i=1}^{n} i^2 x^i = \frac{x + x^2 - (n+1)^2 x^{n+1} + (2n^2 + 2n - 1)x^{n+2} - n^2 x^{n+3}}{(1-x)^3}$$

□

Problem 3 Show that $\sum_{k=1}^{\infty} 1/k^{3/2}$ is bounded above by a constant.

Solution.

We use the integral method for bounding the infinite sum. Since $f(k) = \frac{1}{k^{3/2}}$ is monotonically decreasing for $k > 1$, we have:

$$\begin{aligned}
\sum_{k=1}^{\infty} 1/k^{3/2} &= 1 + \sum_{k=2}^{\infty} 1/k^{3/2} \\
&\leq 1 + \int_{1}^{\infty} \frac{1}{x^{3/2}} dx \\
&= 1 + \left[\frac{-2}{\sqrt{x}}\right]_1^{\infty} \\
&= 1 + (0 + 2) \\
&= 3
\end{aligned}$$

Hence, $\sum_{k=1}^{\infty} 1/k^{3/2} \leq 3$.

□

Problem 4 In this problem, we will prove that the "average" number of divisors of a number of size about n is "about" $\ln n$. More formally, let $t(n)$ be the number of divisors of the number n, and let

$$T(n) = \frac{1}{n} \sum_{j=1}^{n} t(j)$$

be the average number of divisors of all numbers less than or equal to n. Show that $T(n) \sim \ln n$.

Hint: Let the function $f(i, j) = 1$ if i is a divisor of j, and 0 otherwise. What do the row and column sums of $f(i, j)$ mean?

Solution.

Consider the image of the function $f(i, j)$ as a matrix whose rows and columns correspond to values of i and j respectively; that is, element (i, j) of the matrix denotes whether or not i divides j.

				i				
	$f(i,j)$	1	2	3	4	5	6	
	1	1	1	1	1	1	1	
j	2	0	1	0	1	0	1	$\cdots$
	3	0	0	1	0	0	1	
	4	0	0	0	1	0	0	
	$\vdots$			$\vdots$				

Notice that the sum of each row column j corresponds to the number of divisors of the number j and the sum of each row i corresponds to the number of multiples of the number j. In fact, the sum of the first n elements of a row i correspond to the number of multiples of i up to the number n, which can be compactly represented as $\lfloor \frac{n}{i} \rfloor$.

It follows that:

$$\begin{aligned} T(n) &= \frac{1}{n}\sum_{j=1}^{n} t(j) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} f(i,j) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\lfloor \frac{n}{i} \right\rfloor \end{aligned}$$

$T(n)$ can then be bounded from below and above as follows:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}\left(\frac{n}{i}-1\right) &\le T(n) \le \frac{1}{n}\sum_{i=1}^{n}\frac{n}{i} \\ \frac{1}{n}\left(\sum_{i=1}^{n}\frac{n}{i}-\sum_{i=1}^{n}1\right) &\le T(n) \le \sum_{i=1}^{n}\frac{1}{i} \\ \sum_{i=1}^{n}\frac{1}{i}-1 &\le T(n) \le \sum_{i=1}^{n}\frac{1}{i} \end{aligned}$$

In class, we determined that $H_n = \sum_{i=1}^{n}\frac{1}{i} \sim \ln n$. Thus, it follows that $T(n) \sim \ln n$.

□

Problem 5 Asymptotic notation

For each pair of functions $(f(n), g(n))$ in the table below, indicate whether the statement at the top of the column is true (e.g., whether $f(n) = O(g(n))$). Assume $k \ge 1$, $\epsilon > 0$, and $c > 1$ are constants. Your answer should be in the form of a table with "yes" or "no" in each box.

$f(n)$	$g(n)$	$f(n) = O(g(n))$	$g(n) = O(f(n))$	$f(n) = \Theta(g(n))$
$\log^k n$	n^ϵ			
n^k	c^n			
$\sqrt{n}$	$n^{\sin n}$			
2^n	$2^{n/2}$			
$\log(n!)$	$\log(n^n)$			

Solution.

$f(n)$	$g(n)$	$f(n) = O(g(n))$	$g(n) = O(f(n))$	$f(n) = \Theta(g(n))$
$\log^k n$	n^ϵ	yes	no	no
n^k	c^n	yes	no	no
$\sqrt{n}$	$n^{\sin n}$	no	no	no
2^n	$2^{n/2}$	no	yes	no
$\log(n!)$	$\log(n^n)$	yes	yes	yes

Following are some hints on deriving the table above:
(a) See notes on O-notation. (to appear)
(b) Polynomial growth versus exponential growth. (See Theorem 3.6, Lecture 9, Fall97.)
(c) When n is a multiple of π, $n^{\sin n} = 1$, so no constant times 1 will bound $\sqrt{n}$ as n ranges over multiples of π. When n is of the form $2k'\pi + \pi/2$, for some integer k', then $n^{\sin n} = n$ and $\frac{n}{\sqrt{n}} = \sqrt{n}$ grows without bound as k' increases.
(d) $\frac{2^n}{2^{n/2}} = 2^{n/2}$ grows without bound as n grows – it is not bounded by a constant.
(e) See Rosen §1.8 Ex. 61,62

$\square$

Problem 6 Recurrences

Consider a set of record items, each carrying some data and a key and the following definition of a *majority key.*

Definition. *The* majority key *of a set of n records is a key that occurs more than $n/2$ times.*

For example, 2 is the majority key in a set of 3 records which are tagged, consecutively, by keys 1, 2, and 2. On the other hand, in the set of 4 records tagged by keys 1, 2, 3, and 3, there is no majority key.

Consider the following algorithm for finding the majority key of a set of records: If there is an even number of records, we pair adjacent items up (first with second, third with fourth, etc.) and compare their keys. If they have the same key, we keep one of them, otherwise we throw them both away. We run this algorithm recursively on the resulting set. If there is an odd number of items in our set, we find the majority key of the set containing all but the last item. If there was no majority key than we check if the key of the last record is the majority key by explicitly counting how many times this key appears in the whole set. We continue in this fashion until we either get one record, in which case we check if its key is in fact the majority key — again by counting how many times this key appears in the original set — or we end up with no items, in which case we conclude that there was no majority key.

(a) Prove that if there is a majority key in the initial set of records then this algorithm terminates with a record item carrying that key. (Hint: Find an invariant that is preserved at every stage. Consider the even and odd cases separately.)

Solution.

Case 1: Even number of elements $= 2k$. Element t occurs at least $2k+1$ times. After we pair the elements into k pairs, suppose l pairs have two different elements and $k-l$ pairs have two same elements. Of the l different pairs, there is at most 1 t in each pair, therefore at most l t's are taken away. Hence, the same elements have at least $k-l+1$ t's out of $2k-2l$ elements. So, t is a majority element now and it still will be after dividing the number of elements by 2.

Case 2: Odd number of elements $= 2k+1$. If the last element is not t, then t occurs $k+1$ times in the first $2k$, i.e. t is a majority element of the first $2k$ and we can see case 1.

If the last element is t, than t may not be a majority element of the first $2k$, but the first $2k$ can have no other majority element so the algorithm will return either t or no majority, and in the latter case we discover t is the majority element in the checking phase.

$\square$

(b) Write down a recurrence for worst-case time this algorithm takes. For this part assume that the size of the working set at every stage is even.

Solution.

A recurrence for the worst-case time for this algorithm is $T(n) = T(n/2) + n/2$.

$\square$

(c) Solve the recurrence you wrote down in part (b) to obtain an asymptotic time bound (theta) for worst-case running time. (We are interested in the *order* of growth, not the exact constant.)

Solution.

The asymptotic time bound for the worst-case running time is $T(n) = T(n/2) + n/2 = T(n/4) + n/4 + n/2 = ... = \theta(n)$.

$\square$

(d) Argue that this time bound also holds without the assumption that the working set at each stage is even.

Solution.

If the set is odd, then $T(n) = T(n/2) + n/2 + n = T(n/2) + 3/2n$ which comes down to $T(n) = 3n = \theta(n)$.

$\square$

Problem 7 Linear Recurrences

Find closed-form solutions to the following linear recurrences.

(a) $x_n = 5x_{n-1} - 6x_{n-2} \quad (x_0 = 0, x_1 = 1)$

Solution.

The characteristic equation is $r^2 - 5r + 6 = 0$. The roots can be found with the quadratic formula:

$$\begin{aligned} r_1 &= 2 \\ r_2 &= 3 \end{aligned}$$

This implies a solution of the form

$$x_n = A2^n + B3^n.$$

Substituting the initial conditions into this general form gives a system of linear equations:

$$\begin{aligned} 0 &= A + B \\ 1 &= 2A + 3B \end{aligned}$$

The solution to this system is $A = -1$ and $B = 1$. Therefore, the final solution to the recurrence is

$$x_n = 3^n - 2^n.$$

□

(b) $x_n = 5x_{n-1} - 8x_{n-2} + 4x_{n-3} \quad (x_0 = 1, x_1 = 2, x_2 = 3)$

Solution.

The characteristic equation is $r^3 - 5r^2 + 8r - 4 = 0$. Solving a cubic equation can be messy process, but in this case the roots are easy to find:

$$\begin{aligned} r_1 &= 2 \\ r_2 &= 2 \\ r_3 &= 1 \end{aligned}$$

The general form of the solution is therefore

$$x_n = A2^n + Bn2^n + C.$$

Substituting the initial conditions into this general form gives a system of linear equations.

$$\begin{aligned} 1 &= A + C \\ 2 &= 2A + 2B + C \\ 3 &= 4A + 8B + C \end{aligned}$$

The solution to this linear system is $A = 2$, $B = -\frac{1}{2}$, and $C = -1$. The complete solution to the recurrence is therefore

$$x_n = 2^{n+1} - n2^{n-1} - 1.$$

$\square$

(c) (Fibonacci Plus) $x_n = x_{n-1} + x_{n-2} + 1 \quad (x_0 = 0, x_1 = 1)$

Solution.

First, we find the general solution to the homogeneous recurrence. The characteristic equation is $r^2 - r - 1 = 0$. The roots of this equation are:

$$\begin{aligned} r_1 &= \frac{1+\sqrt{5}}{2} \\ r_2 &= \frac{1-\sqrt{5}}{2} \end{aligned}$$

Therefore, the solution to the homogeneous recurrence is of the form

$$x_n = A\left(\frac{1+\sqrt{5}}{2}\right)^n + B\left(\frac{1-\sqrt{5}}{2}\right)^n.$$

Next, we need a particular solution to the inhomogeneous recurrence. Since the inhomogeneous term is constant, we guess a linear solution.

$$\begin{aligned} an + b &= a(n-1) + b + a(n-2) + b + 1 \\ &= 2an - 3a + 2b + 1 \\ 0 &= an - 3a + b + 1 \end{aligned}$$

This equation holds if we set $a = 0$ and $b = -1$. Therefore, the inhomogeneous recurrence has the simple particular solution $x_n = -1$. The complete solution to the recurrence is the homogeneous solution plus the particular solution:

$$x_n = A\left(\frac{1+\sqrt{5}}{2}\right)^n + B\left(\frac{1-\sqrt{5}}{2}\right)^n - 1$$

All that remains is to find the constants A and B. Substituting the initial conditions gives a system of linear equations.

$$\begin{aligned} 0 &= A + B - 1 \\ 1 &= A\left(\frac{1+\sqrt{5}}{2}\right) + B\left(\frac{1-\sqrt{5}}{2}\right) - 1 \end{aligned}$$

The solution to this linear system is:

$$\begin{aligned} A &= \frac{5+3\sqrt{5}}{10} \\ B &= \frac{5-3\sqrt{5}}{10} \end{aligned}$$

Therefore, the complete solution to the recurrence is

$$x_n = \left(\frac{5+3\sqrt{5}}{10}\right) \cdot \left(\frac{1+\sqrt{5}}{2}\right)^n + \left(\frac{5-3\sqrt{5}}{10}\right) \cdot B \left(\frac{1-\sqrt{5}}{2}\right)^n - 1.$$

□

Problem 8 Injections, Surjections, and Bijections

Let f be a function from A to B, and g a function from B to C. The composition $g \circ f$ is defined by: $g \circ f(x) = g(f(x))$.

Determine whether each of the following statements is true or false. If it's true, prove it. If it's false, give a counterexample.

(a) If f and g are both injective then $g \circ f$ is injective.

Solution.

□

(b) If f and g are both surjective then $g \circ f$ is surjective.

Solution.

□

(c) If f is surjective and $g \circ f$ is injective then g is injective.

Solution.

□

(d) If g is injective and $g \circ f$ is surjective then f is surjective.

Solution.

Problem 9 Counting Using Functions

Determine the cardinality of the following:

(a) The number of binary relations on a set A of n elements.

Solution.

Thinking of a binary relation as a binary matrix whose elements indicate whether the pair of elements in A are related wrt to the given relation, the number such relations is given by $2^{(n^2)}$.

(b) The number of different digraphs without self-loops for a set V of n vertices.

Solution.

Again, think of the adjacency matrix of a digraph without self-loops for a set V of n vertices. Since the digraph is restricted to have no self-loops, elements on the diagonal of the adjacency matrix are restricted to equal 0. However, the remaining elements of the adjacency matrix can be either 0 or 1. It follows that the number of digraphs without self-loops for a set V of n vertices is equal to $2^{(n^2-n)}$.

(c) The number of functions from A to B, where $|A| = n$ and $|B| = m$.

Solution.

Since the image of any element of A is unique, it follows that each element of A could map to any one element of B. Thus, the number of functions from A to B, where $|A| = n$ and $|B| = m$, is equal to m^n.

Problem 10 Pigeonhole Principle

(a) Prove that among every set of 18 integers, there exists a pair with sum or difference divisible by 32.

Solution.

The solution must be adjusted from 52 and 100 to 18 and 32, respectively.

We assign each of the 52 numbers to one of 51 bins, which are defined as follows.

First, we define 49 "ordinary" bins. Let bin 1 contain numbers congruent to 1 or 99 modulo 100. Let bin 2 contain numbers congruent to 2 to 98 modulo 100. Continuing in this vein, let bin 49 contain numbers congruent to 49 or 51 modulo 100. Now we define two "special" bins. Let bin 0 contain numbers congruent to 0 modulo 100, and let bin 50 contain numbers congruent to 50 modulo 100.

By the Pigeonhole Principle, at least two numbers are assigned to the same bin. If two numbers are assigned to an ordinary bin and are congruent modulo 100, then their difference is divisible by

100. If two numbers are assigned to an ordinary bin and are not congruent modulo 100, then their sum is divisible by 100. If two numbers are assigned to a special bin, then both their sum and their difference is divisible by 100.

$\square$

(b) One of the 6.042 TAs has decided to conduct his next tutorial in the following manner: he will arrange his m students in alphabetical order and hand out tutorial problems to each of the students such that each student ends up with one or more problems. The TA claims that using such an arrangement, there is always a group of (one or more) students that are consecutive in the alphabetical order and collectively receive a number of problems that is a multiple of m.

As an example, consider a 6.042 tutorial with $m = 10$ students in which the students are arranged in alphabetical order and each of which receives $1, 2, 4, 8, 16, 16, 8, 4, 2, 1$ problems. Then, since the fourth, fifth, and sixth students have collectively received 40 problems, the TA's claim holds.

Prove this claim for all m.

Solution.

Let P_i be the total number of problems given out to the first i students. There are two cases. In the first case, suppose that at least one of $P_1, P_2, \ldots, P_m$ is congruent to 0 modulo m. In particular, suppose that P_j is congruent to 0 modulo m. This implies that the number of problems received by the first j pigs is a multiple of m, and so the claim holds. In the second case, suppose that none of $P_1, P_2, \ldots, P_m$ is congruent to 0 modulo m. Then each of the m terms $P_1, P_2, \ldots, P_m$ is congruent to one of the $m - 1$ values $1, 2, \ldots, m - 1$ modulo m. By the Pigeonhole Principle, there are at least two terms P_j and P_k that are congruent to the same value modulo m. Thus students $j + 1, j + 2, \ldots, k$ received a total of $P_k - P_j$ problems, which is a multiple of m.

$\square$