Solutions to Review for Quiz 2

Problem 1. Prove that among any set of 150 natural numbers, there must be three numbers, a, b, and c, such that all the pairwise differences, $a-b, a-c, b-c, b-a, c-a, c-b$, are multiples of 70.

Proof. Use Pigeonhole.
Pigeons The 150 numbers.
Holes The natural numbers mod 70.
There is some hole with $\left\lceil \frac{150}{70} \right\rceil = 3$ pigeons, that is, three numbers a, b, c that are equivalent modulo 70. Then their differences are equivalent to 0 modulo 70, therefore they are divisible by 70. ∎

Problem 2. For each of the following sets, decide whether the set is countable or uncountable and prove your answer. (If an answer to one part of the question follows from an answer you have already given to another part, then you need only show that is follows – you don't need to prove the second part from scratch.)

(a) The set E of all the even natural numbers.

This set is countable. The function $f(n) = 2n$ is a bijection from $\mathbb{N}$ to E.

(b) The set of all subsets of E.
This set is uncountable. The power set of any set has bigger cardinality than the original set.

(c) The set of all finite subsets of E.
Let $S_i =$ The set of all subsets of E with i elements.

Each of the S_i is countable and the set under question is the countable union of all the S_i, therefore it is countable.

(d) The set of all even-size finite subsets of E.
This set is countable, being a subset of the set in part (c).

Problem 3. The CS department has just completed a revamping of its curriculum. The following 10 courses are now required: 6.001, 6.002, 6.003, ..., 6.009, and 6.042.
In addition, for every i, j, 6.00i is a prerequisite of 6.00j or 6.0j, if $i < j$ and $gcd(i, j) > 1$.

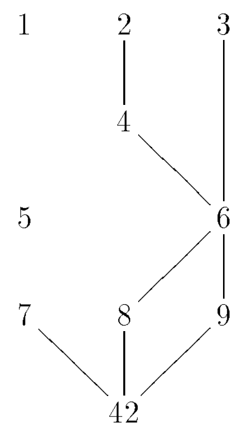
(a) Write down all the elements of the *prerequisite* relation.
prerequisite = { (6.002,6.004) , (6.002,6.006) , (6.002,6.008) , (6.002,6.0042)

(6.003,6.006) , (6.003,6.009) , (6.003,6.042) , (6.004,6.006)
(6.004,6.008) , (6.004,6.042) , (6.006,6.008) , (6.006,6.042)
(6.007,6.042) , (6.008,6.042) , (6.009,6.042) , (6.006,6.009) }

(b) State whether this relation is reflexive, symmetric, transitive, antisymmetric. Is it an equivalence relation?

The relation is antisymmetric, but nothing else.

(c) Now consider the *taken before* relation which is defined to be the reflexive and transitive closure of the *prerequisite* relation. Draw the Hasse Diagram for the *taken before* relation.



(d) The new regulations place no limit on how many courses a student can take in one term. What is the minimum number of terms required to complete all the

required courses? Why?
The minimum number of terms is 5, which is the length of the critical path in the above Hasse diagram.

(e) Write down a maximum-size antichain in the partial order. What is the meaning of an antichain in terms of a student's schedule?

6.001, 6.002, 6.003, 6.005, 6.007.
An antichain is a list of subjects that a student can take at the same term.

Problem 4. An *independent set* in a graph is a subset S of the vertices such that no edge connects two vertices in S. Consider the following algorithm on a graph with vertices $v_1, \ldots, v_n$:

initially no vertex is marked
for $i = 1$ to n
  if no neighbor of v_i is marked **then**
    mark v_i

(a) Prove that when this algorithm terminates, the set of marked vertices is an independent set.

By induction on i:

The inductive hypothesis is:

$P(i)$: after the completion of the ith loop, the marked nodes form and independent set.

Base Case:

$P(1)$: Since only v_1 is marked, it forms an independent set since it isn't connected to any other marked nodes.

Inductive Step: $P(i) \rightarrow P(i+1)$

At the beginning of the $i+1$st iteration, we know by the inductive hypothesis that the marked nodes form an independent set of $v_1 \ldots v_i$. We add v_{i+1}, only if it is not connected to any other v_k that's marked, so the marked vertices of $v_1 \ldots v_{i+1}$ remain an independent set.

Hence after completion of the nth iteration of the for loop, the marked vertices of $v_1 \ldots v_n$ form an independent set.

(b) (Optional) Prove that the set of marked vertices is a *maximal* (under containment) independent set, in that adding any other vertex will make it not-independent.

Assume for contradiction that there exists some vertex, v_k that is not marked, yet is not connected to any marked vertex. When $i = k$ in the for loop, however, this vertex would have to have been marked since it wasn't connected to any marked vertex. So we have a contradiction, and there be no unmarked vertices with all unmarked neighbors.

Therefore the marked vertices are a maximal independent set (note maximal, not maximum since we're not claiming that this set is the largest independent set).

Problem 5. We know that trees are bipartite because they contain no odd cycles. Give a direct proof that trees are bipartite by induction on the number of tree vertices.

Base case: clearly a 1 vertex tree is bipartite.

Inductive step: Suppose all n-vertex trees are bipartite—that is, can be legally colored with two colors red and green. Consider any $n+1$ vertex tree T. It has a leaf vertex v. If we remove v, we are left with an n-vertex tree that is bipartite by the IH. So we can legally color the remaining n tree vertices red and green. Now put v back. It is adjacent to only 1 node. If that node is red, make v green, otherwise make v red. This extends the legal coloring to v, so we have successfully colored T with 2 colors, proving it is bipartite.

Problem 6. Sums.

(a) Give tight (Θ) bounds for

$$\sum_{i=1}^{n} i^{1/2}.$$

$$\begin{aligned}
\int_1^n \frac{1}{\sqrt{x}}dx \quad &\le \sum_{i=1}^{n} \frac{1}{\sqrt{i}} \quad \le 1 + \int_1^n \frac{1}{\sqrt{x}}dx \\
2x^{\frac{1}{2}}|_1^n \quad &\le \sum_{i=1}^{n} \frac{1}{\sqrt{i}} \quad \le 1 + 2x^{\frac{1}{2}}|_1^n \\
\sqrt{n} \le 2\sqrt{n} - 2 \quad &\le \sum_{i=1}^{n} \frac{1}{\sqrt{i}} \quad \le 2\sqrt{n} - 1 \le 2\sqrt{n}
\end{aligned}$$

for all $n \ge 4$, that is

$$\sum_{i=1}^{n} \frac{1}{\sqrt{i}} \in \Theta(\sqrt{n}).$$

(b) Express the following in closed form

$$\sum_{p=1}^{\infty}\sum_{q=p}^{\infty} x^p y^q.$$

(Note that the lower indices are not constants.)

$$\begin{aligned}
\sum_{p=1}^{\infty}\sum_{q=p}^{\infty} x^p y^q &= \sum_{p=1}^{\infty} x^p y^p \sum_{q=p}^{\infty} y^{q-p} \\
&= \sum_{p=1}^{\infty} x^p y^p \sum_{r=0}^{\infty} y^r \\
&= \left(\sum_{p=1}^{\infty} x^p y^p\right)\left(\sum_{r=0}^{\infty} y^r\right) \\
&= \left(\tfrac{xy}{1-xy}\right)\left(\tfrac{1}{1-y}\right).
\end{aligned}$$

Problem 7. Consider the following recurrence:

$$T(n) = 2T(\lfloor n/2 \rfloor) + n \log\log n, \qquad T(1) = 1$$

(a) Assume that n is a power of 2 so that the floor operation can be ignored. Give tight (Θ) bounds for $T(n)$. Show your derivation, but don't bother proving the result correct. (Hint: you might use Stirling's approximation to asymptotically bound $\log((\log n)!)$.

Plug and chug gives

$$\begin{aligned}
T(n) &= \sum_{k=0}^{i-1} 2^k (n/2^k) \log\log(n/2^k) + 2^i T(n/2^i) \\
&= n \sum_{k=0}^{\log n - 1} \log\log n/2^k + 2^{\log n} T(1) \\
&= n \sum_{k=0}^{\log n - 1} \log(\log n - k) + n \\
&= n \sum_{k=1}^{\log n} \log k + n \\
&= n \log \prod_{k=1}^{\log n} k + n
\end{aligned}$$

$$\begin{aligned} &= n(\log((\log n)!) + 1) \\ &\in \Theta(n \log n \log\log n) \end{aligned}$$

(b) Prove that $T(n)$ is an increasing function of n (that is, that $T(n+1) \geq T(n)$).

By strong induction. We know that $T(n+1) = 2T(\lfloor (n+1)/2 \rfloor) + (n+1)\log\log(n+1)$. We also know that $\lfloor (n+1)/2 \rfloor \geq \lfloor n/2 \rfloor$. It follow from the inductive hypothesis that $T(\lfloor (n+1)/2 \rfloor) \geq T(\lfloor n/2 \rfloor)$. Thus,

$$\begin{aligned} T(n+1) &= 2T(\lfloor (n+1)/2 \rfloor) + (n+1)\log\log(n+1) \\ &\geq 2T(\lfloor n/2 \rfloor) + n\log\log n \\ &= T(n) \end{aligned}$$

(c) From the previous two parts, deduce tight (Θ) bounds on $T(n)$ when n is an arbitrary natural number.

Let $k = \lceil \log_2 n \rceil$. It follows that $2^{k-1} \leq n \leq 2^k$. Therefore, $T(2^{k-1}) \leq T(n) \leq T(2^k)$. In other words,

$$2^{k-1}(\log(k-1)! + 1) \leq T(n) \leq 2^k(\log k! + 1)$$

However,

$$\begin{aligned} 2^{k-1}(\log(k-1)! + 1) &= \frac{1}{2}2^k((\log k)! - \log k + 1) \\ &= \Omega(n(\log n \log\log n - \log\log n)) \\ &= \Omega(n \log n \log\log n) \end{aligned}$$

Also, $2^k(\log k! + 1) = O(n \log n \log\log n)$, which implies that $T(n) = \Theta(n \log n \log\log n)$ for all n.

Problem 8. Rank the following functions by order of growth; that is find an arrangement $g_1 = \alpha(g_2) = \alpha(g_3) = \ldots = \alpha(g_6)$ and state whether α is o or Θ for each step.

- $n(\log_2 n)^5$
- n^3
- $n!$
- $(n!)^2$
- n^n
- $3 \cdot 8^{\log_2 n} + 4^{\log_2 n}$

$$n(\log_2 n)^5 < n^3 \leq 3 \cdot 8^{\log_2 n} + 4^{\log_2 n} < n! < n^n < (n!)^2$$

where $<$ means "is $o(\cdot)$" and $\leq$ means "is $\Theta(\cdot)$".

Problem 9. Recurrences.

(a) Suppose that each pair of a genetically engineered species of rabbits left on an island produces two new pairs of rabbits at the age of one month and six new pairs of rabbits at the age of two months and every month afterwards. None of the rabbits ever die or leave the island. Find a recurrence relation for the number of *pairs* of rabbits on the island n months after one newborn pair is left on the island. Do not solve the recurrence relation.

Let $N(t)$ be the number of pairs of rabbits after t months. Then $N(0) = 1$, $N(1) = 3$, and we have that, for $t \geq 2$:

$$N(t) = 7N(t-2) + 3(N(t-1) - N(t-2)) = 3N(t-1) + 4N(t-2)$$

(b) Solve the following inhomogeneous recurrence relation:

$$f(n) = 6f(n-1) - 9f(n-2) + 2^n, n \geq 2$$

The initial conditions are $f(0) = 5, f(1) = 26$.

The characteristic equation $r^2 - 6r + 9$ has a repeated root $\alpha = 3$. Hence the homogeneous solution has the form $f_h(n) = A3^n + Bn3^n$, for some constants A and B. To find a particular solution, we try $f_p(n) = K2^n$. Plugging into the difference equation gives:

$K2^n = 6K2^{n-1} - 9K2^{n-2} + 2^n \Rightarrow K = \frac{6K}{2} - \frac{9K}{4} + 1$
$\Rightarrow 4K = 12K - 9K + 4 \Rightarrow K = 4$

Hence the complete solution has the form $f(n) = A3^n + Bn3^n + 2^{n+2}$. We now use the initial conditions to solve for A and B:

$f(0) = 5 = A + 4 \Rightarrow A = 1$
$f(1) = 26 = 3 + 3B + 8 \Rightarrow B = 5$

Hence the solution is $f(n) = 3^n + 5n3^n + 2^n, n \geq 2$.

Problem 10. Consider the following nuclear reaction. There is one particle of type 1 at time instant 1. Each particle of type i existing at time t undergoes fission to produce one particle of type $i+1$ and one of type $i+2$ at time instant $t+1$.

(a) Write and justify a recurrence for $f(i,t)$, the number of particles of type i at time t. (Make sure to include the base cases for your recurrence.)

Particles of types $i-2$ and $i-1$ at time $t-1$ give rise to particles of type i at time t and this is the only way that particles of type i are formed at time t. Hence we get the recurrence

$$f(i,t) = f(i-1,t-1) + f(i-2,t-1), t > 1.$$

The base case is

- $f(1,1) = 1$.
- $f(i,1) = 0, -\infty < i < \infty, i \neq 1$.

(b) Prove by induction that $f(i,t) = \binom{t-1}{i-t}$. (You do not need to reprove identities that were proven in class.)

(Note: Let a be any nonnegative integer and b be any integer. In what follows, we use the definition that $\binom{a}{b} = \frac{a!}{b!(a-b)!}$ when $a \geq b \geq 0$ and otherwise $\binom{a}{b} = 0$. Pascal's identity holds for this definition.)
The proof is by induction on t. Let $P(t)$ be

$$f(i,t) = \binom{t-1}{i-t}, \forall i.$$

Base Case. $P(1)$ is true since

- $f(1,1) = \binom{0}{0} = 1$.
- $f(i,1) = \binom{0}{i-1} = 0, -\infty < i < \infty, i \neq 1$.

Inductive step. Assume $P(t)$ is true. Consider $P(t+1)$.

$$\begin{aligned} f(i,t+1) &= f(i-1,t) + f(i-2,t) && \text{by the recurrence from part a)} \\ &= \tbinom{t-1}{i-1-t} + \tbinom{t-1}{i-2-t} && \text{by inductive hypothesis} \\ &= \tbinom{t}{i-1-t} && \text{by Pascal's identity} \end{aligned}$$

Hence $P(t+1)$ is true. Thus $P(t)$ is true for all t and we are done.

Problem 11. How many 6-digit decimal numbers are there that do not contain "123" or "456" as a subsequence? (for purposes of this problem numbers may have leading 0's)

Let

- A be the set of all six-digit strings, using the digits $\{0, 1, \ldots, 9\}$.
- B be the set of six-digit strings that contain the subsequence 123.
- C be the set of all six-digit strings that contain the subsequence 456.
- D be the set of all six-digit strings that do not contain 123 or 456.

We're interested in figuring out $|D|$. Since

$$D = A - (B \cup C),$$

and $B \cup C \subseteq A$, we can use the inclusion-exclusion formula:

$$|D| = |A| - |B \cup C| = |A| - |B| - |C| + |B \cap C|.$$

First, $|A| = 10^6$, since we can put any of ten digits in each of six places.

Now, a string can contain 123 in any of four positions; once this position is chosen, we have 1000 choices for the remaining positions. However, this approach counts the string 123123 twice, so the total size of B is

$$|B| = 4(1000) - 1 = 3999.$$

By the same reasoning, we get

$$|C| = 3999.$$

Finally, there are only two strings in $B \cap C$, namely 123456 and 456123. Thus,

$$|B \cap C| = 2.$$

Plugging this into our formula, we get

$$|D| = 10^6 - 3999 - 3999 + 2 = 992004.$$

(Common Mistakes: A lot of people forgot about the double-counting of 123123 *and* 456456*, and so had* $|B| = |C| = 4000$*. Some people also argued that since* 123 *can appear in one of only four places in the string,* $|B| = 4$*.)*

Problem 12. Binomial Coefficients.

(a) Prove, by direct calculation, that the following binomial identity is true for all l, m, n such that $0 \leq l \leq m \leq n$:

$$\binom{n}{m}\binom{m}{l} = \binom{n}{l}\binom{n-l}{m-l}$$

We just expand the binomial coefficients and cancel:

$$\binom{n}{m}\binom{m}{l} = \frac{n!}{m!(n-m)!}\frac{m!}{l!(m-l)!} = \frac{n!}{(n-m)!}\frac{1}{l!(m-l)!} = \frac{n!}{l!(n-l)!}\frac{(n-l)!}{(m-l)!(n-m)!} = \binom{n}{l}\binom{n-l}{m-l}$$

(b) Now prove the same identity combinatorially. Let A be an n-element set. Explain how the left-hand and right-hand sides of the identity in part a) may be interpreted as the number of ways of choosing two subsets B and C of A, where $|B| = m$, $|C| = l$, and $C \subseteq B$.

The left hand side corresponds to first choosing B, then C. The number of ways of doing this is, using the product rule, $\binom{n}{m}\binom{n}{l}$. The right hand side corresponds to first choosing C, and then choosing $B - C$. The number of ways of doing this is, using the product rule, $\binom{n}{l}\binom{n-l}{m-l}$. Since both sides are equivalent to the number of ways of choosing subsets B and C such that $|B| = m$, $|C| = l$, and $C \subseteq B$, the identity holds.