Computer Science 341
Discrete Mathematics

Final Exam Solution

Problem 1: (30 points)
Let G be a simple path of length n. A valid coloring of the path is an assignment of colors to the vertices such that no edge is monochromatic (ie. has both end points of the same color). The goal is to compute the number of ways to color the path with five colors (red, green, blue, yellow, violet) in three different scenarios:

1. (5 points) There are no more restrictions.

   Solution:
   For the first vertex 5 options, after that 4 for the next and so on... altogether $5 \cdot 4^{n-1}$.

   bigskip

2. (10 points) For every color there is at least one node colored with that color.

   Solution:
   We use the inclusion-exclusion principle. The number of solutions is the colorings in 5 colors, minus those with 4, plus those with 3... Altogether:

   $$(5 \cdot 4^{n-1}) - \binom{5}{4}(4 \cdot 3^{n-1}) + \binom{5}{3}(3 \cdot 2^{n-1}) - \binom{5}{2}(2 \cdot 1^{n-1})$$

   More precisely, let A be the set of all valid colorings of the path and let A_i denote the set of colorings that exclude the ith color. The required number is $|A| - |\cup_{i=1}^{5} A_i|$. We calculate $|\cup_{i=1}^{5} A_i|$ by the inclusion-exclusion formula. For this, we need the sizes of intersections of the sets A_i. This can be calculated easily since the intersection of t such sets are precisely the colorings that exclude t of the 5 available colors, i.e. colorings that use at most $5-t$ colors. When you write down the expression you get from using inclusion-exclusion, you should get the formula written above.

3. (15 points) The colors red and green do not appear one next to the other. (Note that the constraint in the previous part does not apply).

   Solution:
   We write a recurrence formula for the number of options to color n vertices such that there are no red and green adjacent. Let g_n, r_n, c_n be the number of such colorings such that the n'th vertex is green/red/anything else. Then

   $$g_n = c_{n-1} + g_{n-1} \quad g_1 = 1$$
   $$r_n = c_{n-1} + r_{n-1} \quad r_1 = 1$$
   $$c_n = 3 \cdot (c_{n-1} + g_{n-1} + r_{n-1}) \quad c_1 = 3$$

   Notice that from symmetry $r_n = g_n$. Also, replacing terms we get that:

   $$g_{n+1} = 4g_n + 3g_{n-1}$$

   The characteristic equation gives $r \in \{2 \pm \sqrt{7}\}$. Then solving with the initial conditions we get that:

   $$g_n = r_n = \frac{1}{2\sqrt{7}}\left[(2+\sqrt{7})^n - (2-\sqrt{7})^n\right]$$

   And then extracting c_n we get:

   $$c_n = g_{n+1} - g_n = \frac{1}{2\sqrt{7}}\left[(1+\sqrt{7})(2+\sqrt{7})^n - (1-\sqrt{7})(2-\sqrt{7})^n\right]$$

Problem 2 (20 points):

A bin contains $\frac{n}{3}$ white balls and $\frac{2n}{3}$ red balls. Consider the following process: a blindfolded person picks a ball from the bin, then check the color. If the ball was white, he returns the ball to the bin and adds c more white balls. If the ball was red, he returns the ball and this time adds c red balls.

If this process is repeated k times, what is the probability of taking out a red ball in the $k+1$ iteration ?

Solution:

The probability is 2/3. There are several ways to solve this. Here is one possible solution.

If we start with r red balls and w white balls, we prove by induction that the probability is exactly $\frac{r}{r+w}$. More precisely, the induction hypothesis is that if we begin with K_1 red balls and K_2 white balls, and let x_k be the ball drawn at the k'th trial, then:

$$\Pr[x_k = R] = \frac{K_1}{K_1 + K_2}$$

Now we can write (for the process beginning with r, w):

$$\begin{aligned}\Pr[x_k = R] &= \Pr[x_k = R | x_1 = R] \cdot \Pr[x_1 = R] + \Pr[x_k = R | x_1 = W] \cdot \Pr[x_1 = W] \\ &= \Pr[y_{k-1} = R]\frac{r}{r+w} + \Pr[z_{k-1} = R]\frac{w}{r+w}\end{aligned}$$

Where y_{k-1}, z_{k-1} are r.v. of a sequence starting with a different number of red/blue balls. But this is a shorter sequence so we can apply the induction hypothesis:

$$\begin{aligned}\Pr[x_k = R] &= \Pr[y_{k-1} = R]\frac{r}{r+w} + \Pr[z_{k-1} = R]\frac{w}{r+w} \\ &= \frac{r+c}{r+w+c} \cdot \frac{r}{r+w} + \frac{r}{r+w+c} \cdot \frac{w}{r+w} \\ &= \frac{r}{r+w}\end{aligned}$$

Problem 3: (15 points)

Scientists at the Princeton Genomics Institute have discovered the following process: Given a test tube filled with DNA strands, it is possible to insert an enzyme that will connect two end points of stands (possibly of the same strand), and then dissolve. Experiments show every pair of endpoints is equally likely to be joined. If two endpoints of the same strand connect, then a "DNA cycle" is formed.

Suppose that N enzymes are inserted into a test tube with N strands. What is the expected number of DNA cycles that will be formed?

Solution:

The answer is $\sum_{i=1}^{n} \frac{1}{2i-1}$.

Problem 4:

Prove that for every $k \geq 3$ there is a number $N(k)$ such that for every set $|S| \geq N$ of points on the plane there is a subset $T \subseteq S$ of size $|T| = k$, such that all points in T lie on line, or no three points in T lie on the same line.

Solution:

We will show that we can pick $N(k) = k^k$, i.e. for any set S of k^k points, either there are k points ino S that all lie on the same line, or there are k points in S such that no three points lie on the same line.

Consider a set S of k^k points. If there is a line containing k points in S we are done. Suppose no such line line exists. Then we perform the following precedure.

1. Let $S_1 = S$
2. For $i = 1$ to $k - 1$ repeat steps 3-7
3. Let x_i be an arbitrary point in S_i.
4. Let $S'_i = S_i \setminus \{x_i\}$.
5. Let L_i be the set of lines that pass through x_i and at least one point in S'_i.
6. Note that each line in L_1 contains at most $k-1$ points from S_i, i.e. at most $k-2$ points from S'_i (since every line contains x_i). For each line in L_1, arbitrarily delete points on the line so that only one point from S'_i remains. Note that we delete at most a $(k-3)/(k-2)$ fraction of the points in S'_i.
7. Let S_{i+1} denote the set of remaining points.
8. Let x_k be an arbitrary point in S_k.

The choice of $N(k) = k^k$ ensures that S_k is non empty in the execution of the above procedure. We claim that the set of points $\{x_1, \ldots, x_k\}$ picked by the algorithm satisfy the property that no three of them lie on a straight line. Note that none of these k points could have been deleted in Step 6 of the procedure. Suppose x_i, x_j, x_k, $i < j < k$ lie on a straight line. Consider the iteration of the above procedure when x_i was picked. At that point, x_j and x_k must have been elements of S'_i. Since x_i, x_j, x_k lie on a straight line, x_j or x_k must have been deleted in Step 6. This is a contradiction.