Tutorial 4

Pigeonhole Principle and Counting problems with recursion

Learning Objectives

In this tutorial you will be using the pigeonhole principle and find recursive formulas for genetic sequences.

These problems relate to the following course learning objectives:

1. Select and justify appropriate tools (induction, graphs, recurrences, complexity theory, generating functions, probability) to analyze a counting problem.
2. Analyze a counting problem by proving an exact or approximate enumeration, or a method to compute one efficiently.

Pigeonhole Principle

1. Show that among 4 numbers one can find 2 numbers such that their difference is divisible by 3.
2. Show that among $n+1$ numbers one can find 2 numbers such that their difference is divisible by n.
3. Show that for any natural number n there is a number composed of digits 5 and 0 only and divisible by n.

Coupled Recurrences

4. A **genetic sequence** is a string over the 4-letter alphabet $\{A, T, C, G\}$. Show that the number of genetic sequences of length n with no two consecutive repeated letters (AA, TT, CC, or GG) is $4 \cdot 3^{n-1}$.
5. Give a recurrence relation and initial conditions for the number of genetic sequences with no consecutive Gs, that is, strings that do not contain GG as a substring.
6. Find a recurrence relation for genetic sequences that do not contain AA or GG.

   Hint: Try to introduce more than one function (this is called a coupled recurrence)

1. There are 3 possible remainders when considering division by 4, so by the pigeonhole principle two of the 4 numbers must have the same remainder.

2. There are n possible remainders when considering division by $n+1$, so by the pigeonhole principle two of the $n+1$ numbers must have the same remainder.

3. We will use the previous problem. We want to find a number divisible by n; the previous problem tells us that given any set of $n+1$ numbers, some two of them have a difference that's divisible by n. So we should try to find a set of $n+1$ numbers with the property that for any two of them, the difference is a number composed of digits 5 and 0 only. One possibility is the sequence of numbers $5, 55, 555, 5555, 55555, \dots$, since the difference of any two of these will be some number of 5s followed by some number of 0s. So we can take the first $n+1$ numbers whose only digits are 5, and there must be some pair whose difference is composed of only 5s and 0s, and divisible by n.

4. This is clear as every letter except the first one has to be different from the one immediately preceding it

5. We have the recurrence $b(n) = 3b(n-1) + 3b(n-2)$, we can end a string with either A, C, T, or AG, CG, TG

6. (a) We clearly have $s(n) = q(n) + a(n) + g(n)$. We also have the recurrences $q(n) = 2s(n-1)$, $a(n) = q(n-1) + g(n-1)$, $g(n) = q(n-1) + a(n-1)$. Then we have

$$\begin{aligned} s(n) &= q(n) + a(n) + g(n) \\ &= 2s(n-1) + (q(n-1) + g(n-1)) + (q(n-1) + a(n-1)) \\ &= 2s(n-1) + q(n-1) + (q(n-1) + g(n-1) + a(n-1)) \\ &= 2s(n-1) + 2s(n-2) + s(n-1) \\ &= 3s(n-1) + 2s(n-2) \end{aligned}$$