Solutions to Problem Set 3

1. Define C to be all strings consisting of some positive number of 0's, followed by some string twice, followed again by some positive number of 0. For example 1100 is not in C, since it does not start with at least one 0. However 0001011010000000 is in C since it is three 0's, followed by 101 twice, followed by seven 0's. Prove that C is not regular.
   [10 points]

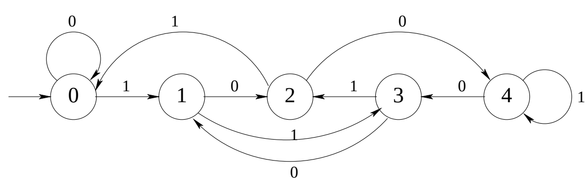
   SOLUTION: We will show that there are infinitely many strings, any two of which are distinguishable with respect to C. This will mean there are infinitely many indistinguishability classes. By the Myhill-Nerode Theorem, we can then conclude that C is not regular.

   Our strings will be 01^k0 for each natural number k. Let k_1 and k_2 be distinct natural numbers. $01^{k_1}01^{k_1}00$ is in L. If $01^{k_1}01^{k_2}00$ were in L, then it must be $0ss0$ or $0ss00$ for some string s. So s must contain at least one zero. Thus $01^{k_1}01^{k_2}00$ must be $0ss0$. So s must end with a 0, and that is the only 0 in s But then must s must be both $1^{k_1}0$ and $1^{k_2}0$. This is impossible since those strings have different lengths. So each 01^k0 is in a different indistinguishability class and C is not regular.

   (4 points for giving the correct strings, 6 points for arguing distinguishability)

2. Let A be the set of all binary strings which, when interpreted as a number with the most significant bit on the left, are divisible by 5. We know the language is regular from a previous homework. Construct an optimal DFA for A and prove its optimality by giving pairwise distinguishable strings, equal in number to the number of states in your DFA.
   [10 points]

   SOLUTION: The following DFA with 5 states recognizes the language (as proved in HW 1). We claim that any two binary strings, which when interpreted as numbers have different remainders modulo 5, are distinguishable. This would imply that there must be at least 5 equivalence classes for the indistinguishability relation and hence the DFA here is optimal.



   To prove the claim, consider any two strings (thinking of them as numbers) x and y. Let $x = r_1 \mod 5$ and $y = r_2 \mod 5$, with $r_1 \neq r_2$. Let w be the number $5 - r_1$, written using *four* bits. Then

$$xw \equiv (2^4 x + 5 - r_1) \equiv [(2^4 \mod 5)(x \mod 5) + 5 - r_1] \equiv 0 \mod 5$$

   On the other hand,

$$yw \equiv [(2^4 \mod 5)(y \mod 5) + 5 - r_1] \equiv r_2 - r_1 \not\equiv 0 \mod 5$$

   Hence, the two strings can be distinguished.

   (3 points for giving the DFA, 2 points for giving the equivalence classes and 5 points for showing the distinguishability.)

3. Consider the language $F = \{a^i b^j c^k \mid i, j, k \geq 0 \text{ and if } i = 1 \text{ then } j = k\}$.
   [4 + 4 + 2 = 10 points]

   (a) Show that F acts like a regular language in the pumping lemma i.e. give a pumping length p and show that F satisfies the conditions of the lemma for this p.

   SOLUTION: The pumping lemma says that for any string s in the language, with length greater than the pumping length p, we can write $s = xyz$ with $|xy| \leq p$, such that xy^iz is also in the language for every $i \geq 0$.

   For the given language, we can take $p = 2$. Consider any string $a^i b^j c^k$ in the language. If $i = 1$ or $i > 2$, we take $x = \epsilon$ and $y = a$. If $i = 1$, we must have $j = k$ and adding any number of a's still preserves the membership in the language. For $i > 2$, all strings obtained by pumping y as defined above, have two or more a's and hence are always in the language.

   For $i = 2$, we can take $x = \epsilon$ and $y = aa$. Since the strings obtained by pumping in this case always have an even number of a's, they are all in the language. Finally, for the case $i = 0$, we take $x = \epsilon$, and $y = b$ if $j > 0$ and $y = c$ otherwise. Since strings of the form $b^j c^k$ are always in the language, we satisfy the conditions of the pumping lemma in this case as well.

   (1 point for handling each of the cases.)

   (b) Show that F is not regular.

   SOLUTION: We claim all strings of the form ab^i must be in distinct equivalence classes for all $i \geq 0$. This is because any two strings ab^{i_1} and ab^{i_2} can be distinguished by c^{i_1}, since $ab^{i_1}c^{i_1} \in F$, while $ab^{i_2}c^{i_1} \notin F$. Since there are infinitely many equivalence classes of the indistinguishability relation, we conclude by the Myhill-Nerode theorem that no DFA can recognize F.

   (2 points for giving a set of distinguishable strings and 2 points for arguing the distinguishability.)

   (c) Why is this not a contradiction?

   SOLUTION: The pumping lemma only says that if a language is regular, then it must satisfy the conditions of the lemma. However, this does not necessarily mean that no non-regular language can satisfy these conditions. (2 points)

4. Show that for any positive integer m, there exists a language A_m such that:

   (a) There is a DFA with m states which recognizes A_m.

   (b) No DFA with less than m states recognizes A_m.

   [10 points]

   SOLUTION 1: Consider the language $A_m = \{1^{m-2}\}$ over the alphabet $\Sigma = \{0, 1\}$. A DFA with m states which has states $0, \ldots, m-2$ for counting the number of 1s seen so far, and an additional "crash state" into which it enters on ever seeing a 0 or more than $m-2$ 1s, is a DFA with m states which recognizes this language. On the other hand, any two strings of the form 1^i, 1^j for $0 \leq i < j \leq m-1$ are distinguished by 1^{m-2-i} and hence, any DFA must have at least m states.

   SOLUTION 2: Let $A_m = \{1^k \mid m \text{ divides } k)$ over $\Sigma = \{1\}$. A DFA with m states which simply stores the number of 1s seen so far, modulo m recognizes this language. Also, for any two strings 1^{k_1} and 1^{k_2} such that $k_1 \not\equiv k_2 \mod m$, the string $1^{m-(k_1 \mod m)}$ distinguishes the two. Hence, any two strings in which the number of 1s is different modulo m must be in different equivalence classes, showing that no DFA with less than m states can recognize this language.

(5 points for exhibiting the first condition and 5 points for the second.)