Solutions for Problem Set 7/week 10
cs172 Allison Coates

1. Prove with the pigeonhole principle that there are infinitely many natural numbers $N \in \{0, 1, \ldots\}$ that have $K(N) \geq \log N$. To encode an integer N, we need a binary string of length $\log N$. Fixing N, there are 2^N strings of length n, and $\sum i = 0^{n-1} 2^i = 2^n - 1$ of strings less than N. There is at least one string for every length N that cannot be compressed. However, since the integers are countably infinite, then there are an infinite number of strings with this property.

   Another way to look at it is as a proof by contradiction. Assume that there are finitely many natural numbers N that have $K(N) \geq \log N$. Then there is some number N_0 which is the last (in the natural ordering) of the numbers with large $K()$. All numbers $> N_0$ have $K(N) < \log N$. At most, every number before N_0 could have had large K. So, at most, N_0 numbers had large $K()$. Now consider the next N_0 numbers, each of which must have small K complexity: they can all be shrunk to smaller descriptions than their simple bit string representation. But there are only $2^{\log N_0} - 1$ shorter descriptions. Therefore, the rest of the numbers cannot be fit into this set of descriptions: therefore not every number above N_0 has small $K()$. Hence, contradiction.

2. Let S be a finite set of bit strings. We know that there is an $x \in S$ with $K(x) \geq \log |S|$. But, what can you say about the average $K(s)$? Prove that for $|S| = 2^n - 1$,

$$\frac{1}{|S|} \sum_{X \in S} K(x) \geq n - 2$$

   The universe is of size $|2^n - 1|$. Using bit strings, there are $2^(l)$ words can be written as a bit string of length l. In order to cover our universe, we need all words of length $< n$, since $\sum_{i=0}^{n-1} 2^i = 2^n - 1$.

   If we use the smallest words of the set $\{0, 1\}*$ for our descriptions, then the the total length L of all descriptions will be at least $\sum_{i=0}^{n-1} i * 2^i = 1*0 + 2*1 + 4*2 + 8*3 + ... + 2^(n-1) * (n-1)$. Therefore, the average will be

$$\frac{\sum_{i=0}^{n-1} i * 2^i = 1*0 + 2*1 + 4*2 + 8*3 + ... + 2^(n-1) * (n-1)}{2^n - 1}.$$

   which is always bigger than n-2.

3. 6.13 Show how to compute the descriptive complexity of strings, K(x), with an oracle for A_{TM}.

   Given a string x, the descriptive complexity of x is the minimum description length $|d(x)|$; that is, the shortest string $< M, w >$ where TM M on input w halts with x on its tape.

   Come up with some binary strings representing $< M, w >$ in lexicographic-length ordering (shortest first.) Ask Oracle $O(A_{TM})$ if $< M, w > \in A_{TM}$. If the oracle says yes, then we can run M on w. It is sure to halt (and accept), so we can check if what is written on the tape is x. Since the strings are in lexicographic ordering, the first string for which the oracle answers yes will be the shortest of the possible machine descriptions.

4. 6.17

   Show that the set of incompressible strings contains an infinite Turing recognizable subset. Proof by contradiction:

   Assume that the set of incompressible strings contains an infinite Turing recognizable subset. If there is a Turing recognizable subset, then there is a recognizer R and an enumerator E for that infinite subset. Then the enumerator E is a compression for the set of infinite strings, since the description of the enumerator E is finite. Hence, an infinite subset of the set of incompressible strings are compressible. Contradiction.