CS687 2017. Homework 1

1. Let $\mathbb{R}^{\mathbb{N}}$ be the set of sequences $(r_0, r_1, \dots)$ of real numbers. Note that each member of this set is an infinite sequence of reals. Alternatively, it can be defined as a cross product of the set $\mathbb{R}$ with itself countably infinitely many times:
$$\mathbb{R}^{\mathbb{N}} = \bigotimes_{i \in \mathbb{N}} \mathbb{R}.$$
Prove or disprove: $\mathbb{R}^{\mathbb{N}}$ has the same cardinality as $\mathbb{R}$.

(20 points)

2. (a) Prove that a language L is decidable if and only if there is a Turing machine which accepts a string x if x is in L, and rejects and halts otherwise. (We defined L to be decidable if both L and L^c are Turing-acceptable.)

(10 points)

(b) Prove that every infinite acceptable language has an infinite decidable subset.

(20 points)

(c) A language $L \subseteq \Sigma^*$ is said to be *immune* if it has no infinite c.e. subset. Construct an immune language. (Hint: You may want to submit this at the end of the course.)

(*)

3. Lossless data compression has to be invertible - for every compressed word, there should be a unique decompressed word. Consider the binary alphabet $\Sigma = \{0, 1\}$. Let $C : \Sigma^* \to \Sigma^*$ be a compressor. Then the above constraint forces it to be one-to-one (an injection).

Prove that lossless data compression is ineffective: for all long enough lengths n and constant $c \in \mathbb{N}$, the number of strings of length n which have compressed words less than length $n - c$, is at most 2^{n-c}. Discuss why it still makes sense to `gzip` a file.

(10 points)

4. (a) Prove or disprove: There is a prefix encoding $\langle , \rangle$ of pairs of strings with the following property. There is a constant c such that for all pairs of strings (x, y),
$$|\langle x, y \rangle| \leq |x| + |y| + c.$$

(15 points)

5. Let n be a non-negative number.

   (a) Consider the concatenation of the first n binary strings in the standard enumeration
   $$c_n = s_0 s_1 \ldots s_{n-1}.$$
   Prove that $C(c_n) \leq \log_2 n + O(1)$. (10 points)

   (b) Recall that there is a computable enumeration $T_0, T_1, \ldots$ of Turing machines. The "diagonal" halting language
   $$H = \{x : T_x(x) \text{ halts}\},$$
   that is, the set of strings x such that the Turing machine T_x halts on input x, is undecidable.
   Let
   $$h_n = b_0 b_1 \ldots b_{n-1}$$
   be the concatenation of n bits, where b_i is 1 if $s_i \in H$, and b_i is 0 otherwise. Prove that if n is large enough,
   $$C(h_n) \leq \log_2 n + O(1),$$
   that is, even though the diagonal halting problem is uncomputable, almost all the prefixes of its characteristic sequence have very low complexity. (30 points)

6. Recall that $m : \Sigma^* \to \mathbb{N}$ is defined as
   $$m(x) = \min_{y \geq x} C(y),$$
   that is, $m(x)$ is the minimum complexity of all strings beyond x in the standard enumeration of strings.

   Prove: Let $F : \Sigma^* \to \mathbb{N}$ be a partial computable function monotone increasing from some x_0 onwards. Then for every large enough x, $m(x) < F(x)$ when $F(x)$ is defined. (20 points)

7. Prove that self-delimiting Kolmogorov complexity is not invariant with respect to cyclic shifts. That is, there is a string $x_0 x_1 \ldots x_{n-1}$ and an m, where $0 \leq m \leq n-2$ such that the Kolmogorov complexity of $x_{m+1} \ldots x_{n-1} x_0 \ldots x_m$ differs from that of the first by more than an additive constant.[Source: Li and Vitanyi, 2$^{\text{nd}}$ ed, pg 204.]

   (20 points)

8. (a) (Data Processing Inequality) Let ϕ be a total computable function. Show that there is a constant c_ϕ depending only on ϕ such that for any string x,
   $$K(\phi(x)) \leq K(x) + c_\phi \qquad \text{and} \qquad C(\phi(x)) \leq C(x) + c_\phi.$$
   Discuss the implication of this inequality.

   (10 points)

   (b) A physicist friend of yours argues in the following way: "I will place a laptop on a stick of dynamite, and light the fuse. Very soon, you will have a tremendous increase in the entropy - intuitively, a physical process has started with a low complexity configuration and resulted in a high complexity configuration." Reconcile this with the data processing inequality.

   (5 points)

(c) Let $\phi(x, y)$ be a total computable function. Show that there is a constant c_ϕ such that for all strings x and y,

$$K(\phi(x, y)) \leq K(x) + K(y) + c_\phi.$$

(10 points)

(d) Show that the above inequality does not hold for C.

(10 points)

9. Let P be the set of all Turing machines. The halting probability (or Chaitin's Ω) is the following number:

$$\Omega = \sum_{\substack{p \in P \\ p \text{ halts}}} \frac{1}{2^{|p|}}.$$

Prove that a Turing machine cannot decide, for all large enough i, whether the i^{th} bit of Ω is 1 or not.

(5 points)

Assume that you know the following fact: Let n be a given arbitrary large enough number. Given the first n bits of Ω, it is possible to *decide* whether any Turing machine i, $0 \leq i \leq n - 1$ halts or not. Using this fact, prove that there is a constant $c \in \mathbb{N}$, such that for all n,

$$K(\Omega[0 \ldots n - 1]) \geq n - c.$$

(That is, prefixes of Chatin's Ω, unlike the prefixes of characteristic sequence of the halting language, are incompressible - Kolmogorov incompressibility is a stronger requirement than uncomputability.)

(25 points)