CS687 2013. Homework 2

1. (a) Define a prefix encoding $\langle , \rangle : \Sigma^* \times \Sigma^* \to \Sigma^*$. (5 points)

   (b) 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)

2. Let n be a non-negative number.

   (a) Consider the concatenation of the first n binary strings in the standard enumeration

   $$c_n = s_0 s_1 \dots s_{n-1}.$$

   Prove that $C(c_n) \leq \log_2 n + O(1)$. (10 points)

   (b) Recall that there is a computable enumeration $T_0, T_1, \dots$ 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

   $$h_n = b_0 b_1 \dots b_{n-1}$$

   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)

3. 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)

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

(20 points)

5. (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)