Homework #3 RELEASE DATE: 12/11/2019

DUE DATE: 01/14/2019, BEFORE 17:00 on GRADESCOPE

QUESTIONS ABOUT HOMEWORK MATERIALS ARE WELCOMED ON THE FACEBOOK FORUM.

Unless granted by the instructor in advance, you must upload your solution to Gradescope as instructed by the TA.

Any form of cheating, lying, or plagiarism will not be tolerated. Students can get zero scores and/or fail the class and/or be kicked out of school and/or receive other punishments for those kinds of misconducts.

Discussions on course materials and homework solutions are encouraged. But you should write the final solutions alone and understand them fully. Books, notes, and Internet resources can be consulted, but not copied from.

Since everyone needs to write the final solutions alone, there is absolutely no need to lend your homework solutions and/or source codes to your classmates at any time. In order to maximize the level of fairness in this class, lending and borrowing homework solutions are both regarded as dishonest behaviors and will be punished according to the honesty policy.

You should write your solutions in English or Chinese with the common math notations introduced in class or in the problems. We do not accept solutions written in any other languages.

This homework set comes with 200 points and 20 bonus points. In general, every homework set would come with a full credit of 200 points, with some possible bonus points.

1. For Questions 1–6, we will deal with binary symmetric channels with crossover probability ϵ , as shown in Figure 3.2(a) of the IAC notes. The channel can be characterized by

$$p(y = 1 | x = 0) = p(y = 0 | x = 1) = \epsilon.$$

An ensemble X with four equiprobable elements is represented by the four codewords: 000, 011, 101, 110, for sequential transmission over a binary symmetric channel with $\epsilon = 0.1$. Upon receiving the first (leftmost) digit y_1 , calculate the information provided about X. That is, let Y_1 be a binary ensemble with $p(y_1)$ defined by $p(\mathbf{x})$ (or more specifically, $p(x_1)$) and the channel p(y|x). Compute $I(X; Y_1)$.

- **2.** Following the previous question, compute $I(X; Y_2|Y_1)$, where Y_2 is the binary ensemble for the second received digit.
- **3.** Let C denote a binary symmetric channel with crossover probability $\epsilon = 0.4$. Let C^N denote the channel obtained by cascading N statistically independent channels C. Show that C^N is a binary symmetric channel and derive its crossover probability in terms of N.
- 4. Following the previous question, show that the information capacity of \mathcal{C}^N is strictly positive for all positive integers N, and find the smallest N for which the capacity is less than 0.001 bits.
- 5. Consider a channel coding scheme where only two codewords 000 and 111 are sent over a binary symmetric channel with crossover probability $\epsilon = 0.1$, and employ a 'majority vote' decoding (i.e. received binary string with more 0's than 1's is decoded to 000; otherwise decoded to 111). Calculate the rate of the scheme as well as its probability of error (the maximum over both codewords).
- 6. Consider a channel coding scheme where only two codewords 0^N and 1^N are sent over a binary symmetric channel with crossover probability $\epsilon = 0.1$, and employ a 'majority vote' decoding (i.e. received binary string with more 0's than 1's is decoded to 0^N ; otherwise decoded to 1^N). For simplicity, assume that N is an odd integer. Calculate the maximum rate of such a scheme (i.e. by minimizing N) such that the probability of error is ≤ 0.0001 .

- 7. Let \oplus denote the exclusive-or operation on binary strings. Show that for any two binary strings **s** and **t** of length N, $K(\mathbf{s} \oplus \mathbf{t}) \leq K(\mathbf{s}) + K(\mathbf{t}) + c_1 \log N + c_2$, where c_1, c_2 are positive constants.
- 8. Following the previous question, prove or disprove that there exists \mathbf{s} and \mathbf{t} such that $K(\mathbf{s}) \ge N$, $K(\mathbf{t}) \ge N$, but $K(\mathbf{s} \oplus \mathbf{t}) \le \frac{N}{1126}$.
- **9.** Prove that $K(\mathbf{s}^N) \leq K(\mathbf{s}) + c_1 \log N + c_2$, where \mathbf{s}^N denotes concatenating \mathbf{s} for N times, and c_1, c_2 are positive constants.
- **10.** For length-N binary strings **s** with k zeros, show that $K(\mathbf{s}) \leq c_1 N \cdot \mathcal{H}(k/N) + c_2 \log N + c_3$ for some positive constants c_1, c_2, c_3 , where $\mathcal{H}(\theta)$ is the entropy of the binary ensemble with $p(0) = \theta$ and $p(1) = 1 \theta$.

Bonus: Incompressible Strings

11. (Bonus 20 points) A Kolmogorov-incompressible string satisfies $K(\mathbf{s}) \ge |\mathbf{s}|$. If we call any string such that $K(\mathbf{s}) \ge \frac{1}{1126}|\mathbf{s}|$ to be weakly Kolmogorov-incompressible. Prove or disprove that the 'weakly Kolmogorov-incompressible' function $f(\mathbf{s}) = \llbracket K(\mathbf{s}) \ge \frac{1}{1126} |\mathbf{s}| \rrbracket$ is computable.