EE 276 - Information Theory
Midterm
Feb 16, 2024

1. There are a total of 5 questions: 3 shorter questions, and 2 longer ones. You have 2 hours to take the exam. Questions have different numbers of points as indicated before each sub-problem. There are a total of 100 points, with 2 additional bonus points.
2. Please write all answers in the designated area underneath the question. If you need more room for your answer, please indicate as such under the question, and continue your response elsewhere.
3. Scratch paper will be provided and collected at the end of the exam, but will **not** be graded.
4. All answers should be justified, unless otherwise stated.
5. The exam is closed book but you are allowed one double-sided sheet of notes. No other materials are allowed.
6. Calculators are not allowed.

Good luck!

Name:

SUID:

1. Let X and Y be random variables with finite alphabet and non-zero entropy. Consider the quantity

$$\rho := \frac{I(X;Y)}{\min\{H(X), H(Y)\}}.$$

   (a) *(8 points)* Show that $0 \leq \rho \leq 1$.

   (b) *(8 points)* Interpret the extreme cases of $\rho = 0$ and $\rho = 1$: what can you say about the relationship between X and Y under each of these cases?

2. Let the discrete random variables X, Y, Z be such that

$$Z = X + Y.$$

(a) *(8 points)* Show that $H(Z) \leq H(X) + H(Y)$.

(b) *(6 points)* Under what conditions does the inequality hold with equality?

3. *(10 points)* A memoryless source $X_1, X_2, \ldots$ is drawn from $\mathcal{X} = \{a, b, c\}$ with respective probabilities $(1/2, 1/4, 1/4)$. Consider the typical set $\mathcal{A}_\epsilon^{(n)}$ for $n = 8$ and $\epsilon = 1/10$. How many occurrences of the letter "a" are there in a sequence belonging to this set? (i.e., does this set contain sequences with one "a", two "a"s, $\ldots$, etc.?). *You don't need to perform long and difficult arithmetic computations to answer this question.*

4. (*Each part of this problem can be attempted independently of the other parts, using only what is stated in the prompts of the preceding parts.*)

   In this problem, we will study the entropy of the English language as Shannon did. In what follows, we will model an infinitely long string of English text as a random process

$$X_1, X_2, X_3 \dots$$

   where X_i takes values in $\mathcal{X} := \{a, \dots, z\}, |\mathcal{X}| = 26$. Here, X_i represents the ith letter in a long English string. Note that the X_is are *not independent*, but for simplicity, we'll assume that the process is *stationary*, meaning that

$$p(x_1, \dots, x_n) = p(x_{1+k}, \dots, x_{n+k})$$

   for all $n, k \geq 0$ and symbols $x_j \in \mathcal{X}$.

   For $n \geq 1$, let us define the *n-gram entropy* as

$$\begin{aligned} \mathcal{H}_1 &:= H(X_1) \\ \mathcal{H}_n &:= H(X_n | X_{n-1}, X_{n-2}, \dots, X_1), \quad \text{for} \quad n \geq 2, \end{aligned}$$

   which can be thought of as the entropy of the next letter in position n given the previous $n-1$ letters. A natural definition of the entropy of English is then

$$\mathcal{H}_{\text{Eng}} := \lim_{n \to \infty} \mathcal{H}_n.$$

   (a) *(8 points)* Show that $\mathcal{H}_n \geq \mathcal{H}_{n+1}$.

(b) *(4 points)* Using results from class, or otherwise, argue that $\mathcal{H}_n$ converges, i.e., that $\mathcal{H}_{\text{Eng}}$ exists and is finite.

For the sake of simplicity, in what follows, we'll only consider approximating $\mathcal{H}_{\text{Eng}}$ using $n = 1$ and $n = 2$.

(c) *(6 points)* Show that the maximum values that $\mathcal{H}_1$ and $\mathcal{H}_2$ can take on are $\mathcal{H}_1^\star$ and $\mathcal{H}_2^\star$, respectively, which are given by

$$\mathcal{H}_1^\star = \mathcal{H}_2^\star = \log 26 \ \ (\approx 4.7 \text{ bits}).$$

When Shannon and his wife Mary estimated $\mathcal{H}_1$ and $\mathcal{H}_2$ from English text, they found that

$$\mathcal{H}_1 \approx 4.14 \text{ bits} \quad \text{and} \quad \mathcal{H}_2 \approx 3.56 \text{ bits}.$$

(d) *(5 points)* Give a reason why $\mathcal{H}_1$ for English is lower than $\mathcal{H}_1^\star$.

(e) *(5 points)* In general, you showed in part (a) that $\mathcal{H}_2 \leq \mathcal{H}_1$. What does the fact that $\mathcal{H}_2$ is considerably smaller than $\mathcal{H}_1$ tell us about letter pairs in English?

5. (*Each part of this problem can be attempted independently of the other parts, using only what is stated in the prompts of the preceding parts.*)

Consider a channel given by

$$Y = X + Z \tag{1}$$

where the input X, output Y and noise Z are nonnegative. Furthermore, Z and X are independent. Let Z have an exponential distribution with parameter $\lambda > 0$, which we denote by writing $Z \sim \texttt{exp}(\lambda)$. More precisely, this means that the density of Z is given by

$$f(z) = \begin{cases} 0 & z < 0 \\ \lambda e^{-\lambda z} & z \geq 0. \end{cases}$$

It may also be useful to recall that this implies that $\mathbb{E}[Z] = 1/\lambda$.

In this problem, you will find the capacity of this channel, under the constraint on the input

$$\mathbb{E}[X] \leq P. \tag{2}$$

(a) *(8 points)* Show that the differential entropy of Z is given by

$$h(Z) = -\log \lambda + \frac{1}{\ln 2}.$$

(b) *(8 points)* Show that for any random variable $W \geq 0$ with $\mathbb{E}[W] \leq M$, we have

$$h(W) \leq -\log \frac{1}{M} + \frac{1}{\ln 2}.$$

(c) *(8 points)* Use the claims of previous two parts to show that

$$I(X;Y) \leq \log(1 + P\lambda)$$

for Y as in (1) under the constraint (2) on X.

In what follows, you may use the following fact: *Let* $U \sim \texttt{exp}\,(\lambda/(1+P\lambda))$. *If* X *is the random variable defined as*

$$X = \begin{cases} 0 & \text{with probability } \frac{1}{1+P\lambda} \\ U & \text{with probability } \frac{P\lambda}{1+P\lambda} \end{cases}, \tag{3}$$

then $Y = X + Z$ *has the distribution* $\texttt{exp}(\lambda/(1+P\lambda))$.

(d) *(8 points)* Use this fact with the claims of the previous parts to find the capacity of the channel in (1) under the constraint given in (2).

(e) **(Bonus 2 points)** Prove the fact above about the distribution of Y. Namely, show that if X follows the distribution given in (3), then Y has the distribution $\texttt{exp}(\lambda/(1+P\lambda))$. *You may use that the moment generating function of an exponential random variable with parameter* λ *is given by* $\phi(t) = \lambda/(\lambda - t) \quad \text{for} \quad t < \lambda$.