

EE376A - Information Theory
Final, Monday March 14th 2016

Instructions:

- You have **three hours**, 3.30PM - 6.30PM
- The exam has 4 questions, totaling 120 points.
- Please start answering each question on a new page of the answer booklet.
- You are allowed to carry the textbook, your own notes and other course related material with you. Electronic reading devices [including kindles, laptops, ipads, etc.] are allowed, provided they are used solely for reading pdf files already stored on them and not for any other form of communication or information retrieval.
- You are required to provide a detailed explanation of how you arrived at your answers.
- You can use previous parts of a problem even if you did not solve them.
- As throughout the course, entropy (H) and Mutual Information (I) are specified in bits.
- \log is taken in base 2.
- Throughout the exam ‘prefix code’ refers to a variable length code satisfying the prefix condition.
- Good Luck!

1. **Problem 1 : Exponential Noise Channel and Exponential Source** (40 points)

Recall that $X \sim \text{Exp}(\lambda)$ is to say that X is a continuous non-negative random variable with density

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

or, equivalently, that X is a random variable with characteristic function

$$\varphi_X(t) \triangleq E[e^{itX}] = \frac{1}{1 - it/\lambda}.$$

Recall also that in this case $EX = 1/\lambda$.

- (a) Find the differential entropy of $X \sim \text{Exp}(\lambda)$.
- (b) Prove that $\text{Exp}(\lambda)$ uniquely maximizes the differential entropy among all non-negative random variables confined to $EX \leq 1/\lambda$.

Hint: Recall our proof of an analogous fact for the Gaussian distribution.

Fix positive scalars a and b . Let \bar{X} be the non-negative random variable of mean a formed by taking $\bar{X} = 0$ with probability $\frac{b}{a+b}$ and, with probability $\frac{a}{a+b}$, drawing from an exponential distribution $\text{Exp}(1/(a+b))$. Equivalently stated, \bar{X} is the random variable with characteristic function

$$\varphi_{\bar{X}}(t) = \frac{b}{a+b} + \frac{a}{a+b} \cdot \frac{1}{1 - it(a+b)}.$$

Let $N \sim \text{Exp}(1/b)$ and independent of \bar{X} .

- (c) What is the distribution of $\bar{X} + N$?
Tip: simplest would be to compute the characteristic function of $\bar{X} + N$ by recalling the relation $\varphi_{\bar{X}+N}(t) = \varphi_{\bar{X}}(t) \cdot \varphi_N(t)$.
- (d) Find $I(\bar{X}; \bar{X} + N)$.
- (e) Consider the problem of communication over the additive exponential noise channel $Y = X + N$, where $N \sim \text{Exp}(1/b)$, independent of the channel input X , which is confined to being non-negative and satisfying the moment constraint $EX \leq a$. Find $C(a) = \max I(X; X + N)$, where the maximization is over all non-negative X satisfying $EX \leq a$. What is the capacity-achieving distribution?
Hint: Using findings from previous parts, show that for any non-negative random variable X , independent of N , with $EX \leq a$, we have $I(X; X + N) \leq I(\bar{X}; \bar{X} + N)$.
- (f) Let now $U \sim \text{Exp}(1/m)$ and consider the rate distortion problem where the expected reconstruction error is not to exceed D and, in addition, the reconstruction is not allowed to exceed the source value. The associated rate distortion function is thus

$$R(D) = \min I(U; V),$$

where the minimum is over joint distributions respecting the given distribution of U and such that $U - V \geq 0$ while $E[U - V] \leq D$. Explain why for any U, V in the feasible set of this minimization, the following equality and inequalities hold:

$$I(U; V) = h(U) - h(U - V|V) \quad (1)$$

$$\geq h(U) - h(U - V) \quad (2)$$

$$\geq \log(m/D). \quad (3)$$

(g) Show that $R(D)$ from the previous part is given by

$$R(D) = \begin{cases} \log(m/D) & \text{if } 0 < D \leq m \\ 0 & \text{if } D > m. \end{cases}$$

Hint: Using findings from previous parts, for $0 < D \leq m$ establish existence of (U, V) in the feasible set for which the inequalities in (2) and (3) hold with equality.

2. Problem 2: Modulo 8 Channel and a Binary Source with Erasure Distortion
(35 points)

Consider the memoryless channel described by

$$Y = (X + Z) \bmod 8,$$

where the channel input X , output Y , and the noise Z are real valued. For example, if $x = 6.4$ and $z = 1.8$ then $y = 0.2$. The noise is uniformly distributed on $[-B, B]$, i.e. $Z \sim \text{Uniform}[-B, B]$ and is independent of the input X . Assume that $0 < B \leq 4$ is a known channel parameter.

- (a) Find the channel capacity as a function of B .
- (b) A communication system is suggested where the permitted channel input values are restricted to the set $\{1, 3, 5, 7\}$. Show that, when $B = 1$, it is still possible to achieve the channel capacity. In fact, to achieve this capacity with zero probability of error and with a very simple scheme.

Consider now a Bernoulli(1/2) source under erasure distortion. I.e., the reconstruction alphabet is $\{0, 1, e\}$ and

$$d(u, v) = \begin{cases} 0 & \text{if } u = v \\ 1 & \text{if } v = e \\ \infty & \text{otherwise} \end{cases}$$

- (c) Show that the rate distortion function of this source is given by $R(D) = 1 - D$ for $0 \leq D \leq 1$.
- (d) Consider now a joint-source-channel-coding setting for communicating the bit source of part (c) through the modulo 8 channel defined in part (a). For every two source symbols we are allotted one channel use, i.e., the encoder translates a block of $2n$ source bits into n channel inputs.
Find the minimal achievable distortion (when $n \rightarrow \infty$) as a function of B .
- (e) The following communication system is suggested for the setting of the previous part: every two source bits are mapped into a channel input in the following way:

$$\begin{array}{ll} 00 & \rightarrow 1 \\ 01 & \rightarrow 3 \\ 11 & \rightarrow 5 \\ 10 & \rightarrow 7 \end{array}$$

This mapping is known as a Gray Code. Show that this system (with the corresponding optimal decoder) achieves the minimal distortion in the case where $B = 2$.

3. Problem 3: Removing bias from the world (25 points)

Computer simulations require unbiased and independent random samples. However, in most cases, the data obtained contains biased and dependent samples. In this problem we will try to understand aspects of this problem.

Consider a biased coin with unknown bias $0 < p < 1$. The biased coin generates independent coin tosses X_1, X_2, \dots, X_n . We wish to obtain a sequence of unbiased and independent bits $Z_1, Z_2, Z_3, \dots, Z_K$ from the biased coin toss experiment. Specifically, we want a mapping $f(X_1, X_2, \dots, X_n) = (Z_1, Z_2, \dots, Z_K)$, where K may depend upon (X_1, X_2, \dots, X_n) and, conditioned on $K = k$, Z_1, Z_2, \dots, Z_k are fair coin flips.

We define the rate of a scheme to be:

$$R = \frac{E[K]}{n} \quad (4)$$

- (a) Von Neumann gave a simple procedure of generating an unbiased coin, where he considered 2 coin tosses, and mapped the output as follows:

$$f(01) = 0$$

$$f(10) = 1$$

$$f(00) = f(11) = \lambda$$

Here, λ corresponds to a null output (i.e., a string of length 0 where we don't output anything). What is the rate achieved by Von-Neumann's scheme?

- (b) In case of Von-Neumann's method, we assign a bit $\{0, 1\}$ to an outcome in the 'equiprobable set' of $\{01, 10\}$, and assign λ otherwise. Suggest a simple generalization of Von-Neumann's method where you consider n biased coin tosses at a time. (Hint: All sequences within a type have the same probability, regardless of the bias of the coin)
- (c) Show that for any scheme $R \leq H_b(p)$. Can you suggest a sequence of schemes, for increasing n , with rates approaching $H_b(p)$?
- (d) We now consider the opposite problem: generating random samples using the unbiased and independent bits. Specifically, we want to generate independent samples (W_1, W_2, \dots, W_K) with a specific probability mass function, using unbiased and independent bits (Z_1, Z_2, \dots, Z_n) .
- Give a procedure for generating i.i.d samples of a random variable W with probability mass function $\{0.5, 0.25, 0.25\}$
(Hint: A technique similar to Huffman coding might be useful here)
 - Is your procedure optimal?

4. **Problem 4 : Modulo Channel** (20 points)

- (a) Consider the DMC defined as follows: Output $Y = X \oplus_2 Z$ where X , taking values in $\{0, 1\}$, is the channel input, \oplus_2 is the modulo-2 summation operation, and Z is binary channel noise uniformly distributed over $\{0, 1\}$ and independent of X . What is the capacity of this channel?
- (b) Consider the channel of the previous part, but suppose that instead of modulo-2 addition $Y = X \oplus_2 Z$, we perform modulo-3 addition $Y = X \oplus_3 Z$. Now what is the capacity?
- (c) Now suppose the noise Z is no longer independent of the input X , but is instead described by the following conditional distribution:

$$p(Z = z|X = 0) = \begin{cases} 1/4 & \text{if } z = 0 \\ 3/4 & \text{if } z = 1, \end{cases}$$

and

$$p(Z = z|X = 1) = 1/2 \quad \text{both for } z = 0 \text{ and } z = 1.$$

A random code of size 2^{nR} is generated uniformly (that is all codewords are drawn i.i.d. $X \sim \text{Bern}(0.5)$). Find the value V such that if $R < V$ then the average probability of decoding error (average both across the messages and the randomness in the codebook) vanishes with increasing blocklength while if $R > V$ then it does not.