

EE276: Homework #4 Solutions

Due on Friday Feb 13, 6pm - Gradescope entry code: E6VP4X

1. Maximum Differential Entropy

- (a) Show that among all distributions supported in an interval $[a, b]$, the uniform distribution maximizes differential entropy.
- (b) Let X be a continuous random variable with $\mathcal{E}[X^4] \leq \sigma^4$ and Y be a continuous random variable with a probability density function $g(y) = c \exp\left(-\frac{y^4}{4\sigma^4}\right)$ where $c = \frac{1}{\int_{-\infty}^{\infty} \exp\left(-\frac{y^4}{4\sigma^4}\right) dy}$. Show that

$$h(X) \leq h(Y)$$

with equality if and only if X is distributed as Y .

[Hint: you can use the fact that $\mathcal{E}[Y^4] = \sigma^4$.]

Solution:

- (a) Denote by $u(x)$ the uniform distribution, with $x \in [a, b]$, such that $u(x) = \frac{1}{b-a}$ if $x \in [a, b]$, and 0 otherwise. Let $g(x)$ be any distribution supported in the interval $[a, b]$. Then, we have

$$0 \leq D(g||u) \tag{1}$$

$$= \int g(x) \log \frac{g(x)}{u(x)} \tag{2}$$

$$= \int g(x) \log ((b-a)g(x)) \tag{3}$$

$$= \log(b-a) + \int g(x) \log g(x) \tag{4}$$

$$= \log(b-a) - H(X), \tag{5}$$

which implies $H(X) \leq \log(b-a)$.

On the other hand, note that if x is uniformly distributed in the interval $[a, b]$, we have

$$H(X) = \int u(x) \log \frac{1}{u(x)}$$

$$= \int u(x) \log(b-a)$$

$$= \log(b-a),$$

which finishes the proof.

(b) Since $\mathcal{E}[X^4] \leq \sigma^4 = \mathcal{E}[Y^4]$, we have

$$\begin{aligned}
 D(f_X||g) &= \mathcal{E} \left[\log \frac{f_X(X)}{g(X)} \right] \\
 &= -h(X) + \mathcal{E} [-\log g(X)] \\
 &= -h(X) + \mathcal{E} \left[-\log c + \frac{X^4}{4\sigma^4} \log e \right] \\
 &\leq -h(X) + \mathcal{E} \left[-\log c + \frac{Y^4}{4\sigma^4} \log e \right] \\
 &= -h(X) + \mathcal{E} [-\log g(Y)] \\
 &= -h(X) + h(Y).
 \end{aligned}$$

Therefore, $h(Y) \geq h(X) + D(f_X||g) \geq h(X)$.

2. Channel capacity.

Find the capacity of the following channels with given probability transition matrices, where the element p_{ij} of the matrix represents $p(y = j|x = i)$:

(a) $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$

$$p(y|x) = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

(b) $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$

$$p(y|x) = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix}$$

(c) $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ (the Z-channel)

$$p(y|x) = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$$

Solution:

(a) $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$

$$p(y|x) = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

This is a symmetric channel and by the known result for symmetric channel (please read section 6.1.2 in Cover & Thomas & El Gamal to understand this properly. The basic idea is similar to that used for the BSC):

$$C = \log |\mathcal{Y}| - H(\mathbf{r}) = \log 3 - \log 3 = 0. \quad (6)$$

In this case, the output is independent of the input.

(b) $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$

$$p(y|x) = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix}$$

Again the channel is symmetric:

$$C = \log |\mathcal{Y}| - H(\mathbf{r}) = \log 3 - \log 2 = 0.58 \text{ bits}$$

(c) First we express $I(X; Y)$, the mutual information between the input and output of the Z-channel, as a function of $\alpha = \Pr(X = 1)$:

$$\begin{aligned} H(Y|X) &= \Pr(X = 0) \cdot 0 + \Pr(X = 1) \cdot 1 = \alpha \\ H(Y) &= h_2(\Pr(Y = 1)) = h_2(\alpha/2) \\ I(X; Y) &= H(Y) - H(Y|X) = h_2(\alpha/2) - \alpha \end{aligned}$$

Since $I(X; Y)$ is strictly concave on α (why?) and $I(X; Y) = 0$ when $\alpha = 0$ and $\alpha = 1$, the maximum mutual information is obtained for some value of α such that $0 < \alpha < 1$.

Using elementary calculus, we determine that

$$\frac{d}{d\alpha} I(X; Y) = \frac{1}{2} \log_2 \frac{1 - \alpha/2}{\alpha/2} - 1,$$

which is equal to zero for $\alpha = 2/5$. (It is reasonable that $\Pr(X = 1) < 1/2$ since $X = 1$ is the noisy input to the channel.) So the capacity of the Z-channel in bits is $h_2(1/5) - 2/5 = 0.722 - 0.4 = 0.322$.

3. Choice of channels.

Let $\mathcal{C}_1 \equiv (\mathcal{X}_1, p_1(y_1|x_1), \mathcal{Y}_1)$ and $\mathcal{C}_2 \equiv (\mathcal{X}_2, p_2(y_2|x_2), \mathcal{Y}_2)$ be two channels with capacities C_1, C_2 respectively. Assume the input and output alphabets for the two channels are disjoint, i.e., $\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$ and $\mathcal{Y}_1 \cap \mathcal{Y}_2 = \emptyset$. Consider a channel \mathcal{C} , which is a union of the two channels $\mathcal{C}_1, \mathcal{C}_2$, where at each time, one can send a symbol over \mathcal{C}_1 or over \mathcal{C}_2 but not both. In this problem, we calculate the capacity of \mathcal{C} in terms of the C_1 and C_2 .

Define θ as an indicator random variable denoting whether channel \mathcal{C}_1 or \mathcal{C}_2 is used in a particular transmission. Let X and Y denote the input and output for the channel \mathcal{C} . Note that X follows a mixture distribution over the (disjoint) alphabets \mathcal{X}_1 and \mathcal{X}_2 :

$$X = \begin{cases} X_1 & \text{with probability } \alpha \\ X_2 & \text{with probability } 1 - \alpha \end{cases}$$

where X_1 and X_2 are random variables taking values in \mathcal{X}_1 and \mathcal{X}_2 , respectively.

(a) Argue that $H(\theta|X) = H(\theta|Y) = 0$.

(b) Show that

$$I(X; Y) = h_2(\alpha) + \alpha I(X_1; Y_1) + (1 - \alpha) I(X_2; Y_2)$$

where Y_1 and Y_2 are the channel outputs when X_1 and X_2 are transmitted through \mathcal{C}_1 and \mathcal{C}_2 , respectively.

Hint: Start with $I(Y; X, \theta) = I(Y; \theta) + I(Y; X|\theta) = I(Y; X) + I(Y; \theta|X)$.

- (c) Let C be the capacity of the channel \mathcal{C} . Maximize the expression in (b) over α, P_{X_1}, P_{X_2} to show that $2^C = 2^{C_1} + 2^{C_2}$.
- (d) Let $C_1 = C_2$. Then show that $C = C_1 + 1$ and give an intuitive explanation.

Solution:

- (a) Since the output alphabets \mathcal{Y}_1 and \mathcal{Y}_2 are disjoint, θ is a function of Y and so $H(\theta|Y) = 0$. Similarly for $H(\theta|X) = 0$
- (b) Consider,

$$\begin{aligned} I(Y; X, \theta) &= I(Y; \theta) + I(Y; X|\theta) \\ &= I(Y; X) + I(Y; \theta|X) \end{aligned}$$

Now $I(Y; \theta|X) = H(\theta|X) - H(\theta|X, Y) = 0$ because given X , θ is deterministic. Therefore,

$$\begin{aligned} I(X; Y) &= I(Y; \theta) + I(X; Y|\theta) \\ &= H(\theta) - H(\theta|Y) + \alpha I(X_1; Y_1) + (1 - \alpha) I(X_2; Y_2) \\ &= h_2(\alpha) + \alpha I(X_1; Y_1) + (1 - \alpha) I(X_2; Y_2) \end{aligned}$$

(c) It follows from (b) that

$$C = \sup_{\alpha} \{h_2(\alpha) + \alpha C_1 + (1 - \alpha) C_2\}.$$

Maximizing over α one gets the desired result. The maximum occurs for $h'_2(\alpha) + C_1 - C_2 = 0$, with an optimal $\alpha^* = 2^{C_1} / (2^{C_1} + 2^{C_2})$.

Evaluating $I(X; Y)$ at α^* yields the channel capacity:

$$C = \frac{C_2 + \log_2(1 + 2^{C_1 - C_2}) + C_1 + \log_2(1 + 2^{-C_1 + C_2})}{\log_2(2^{C_1} + 2^{C_2})}$$

Thus,

$$2^C = 2^{C_1} + 2^{C_2}$$

- (d) The result follows from (c). Intuitively, if we have two identical channels, then in each transmission, we can transmit one extra bit through our choice of the channel. As an extreme example, suppose both channels are BSC(0.5) channels. Then $C_1 = C_2 = 0$, but $C = 1$. This is because we have two possible channels and we can communicate 1 bit/transmission by sending through channel 1 if the input bit is 0 and sending through channel 2 if the input bit is 1.

4. Cascading an Erasure channel

Let $\{\mathcal{X}, p(y|x), \mathcal{Y}\}$ be a discrete memoryless channel with capacity C . Suppose this channel is immediately cascaded with an erasure channel $\{\mathcal{Y}, p(s|y), \mathcal{S}\}$ that erases α of its symbols.

Specifically, $\mathcal{S} = \{y_1, y_2, \dots, y_m, e\}$, and

$$\begin{aligned}\Pr\{S = y|X = x\} &= (1 - \alpha)p(y|x), \quad y \in \mathcal{Y}, \\ \Pr\{S = e|X = x\} &= \alpha.\end{aligned}$$

Determine the capacity of this channel.

Solution: The capacity of the channel is

$$C = \max_{p(x)} I(X; S) \quad (7)$$

Define a new random variable Z , a function of S , where $Z = 1$ if $S = e$ and $Z = 0$ otherwise. Note that $p(Z = 1) = \alpha$ independent of X . Expanding the mutual information,

$$I(X; S) = H(S) - H(S|X) \quad (8)$$

$$= (H(S, Z) - H(Z|S)) - (H(S, Z|X) - H(Z|S, X)) \quad (9)$$

$$= H(S, Z) - H(S, Z|X) \quad (10)$$

$$= H(Z) + H(S|Z) - H(Z|X) - H(S|X, Z) \quad (11)$$

$$= I(X; Z) + I(S; X|Z) \quad (12)$$

$$= 0 + \alpha I(X; S|Z = 1) + (1 - \alpha)I(X; S|Z = 0) \quad (13)$$

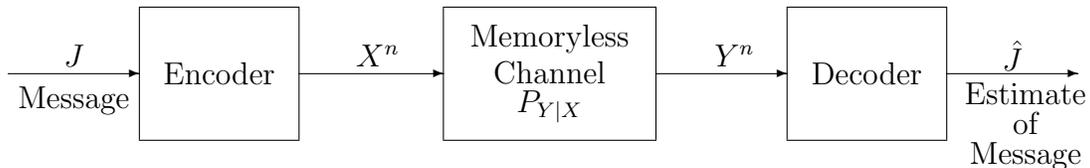
When $Z = 1$, $S = e$ and $H(S|Z = 1) = H(S|X, Z = 1) = 0$. When $Z = 0$, $S = Y$, and $I(X; S|Z = 0) = I(X; Y)$. Thus

$$I(X; S) = (1 - \alpha)I(X; Y) \quad (14)$$

and therefore the capacity of the cascade of a channel with an erasure channel is $(1 - \alpha)$ times the capacity of the original channel.

5. Minimizing Channel Probability of Error.

Below, we are given a communication setting as seen in lecture.



J is a message uniformly distributed on $\{1, 2, \dots, M\}$ passed into the system. The encoder maps message J onto its corresponding n -length codeword X^n from codebook

$c_n = \{X^n(1), X^n(2), \dots, X^n(M)\}$. The encoded message is sent through a memoryless channel characterized by $P_{Y|X}$, and we receive Y^n as output.

The decoder is responsible for estimating J from Y^n ; it is a function \hat{J} that maps Y^n to one of the symbols in $\{1, 2, \dots, M, \text{error}\}$. We define the probability of error $P_e = P(\hat{J}(Y^n) \neq J)$. Show that P_e , for a fixed codebook c_n , is minimized by:

$$\hat{J}(y^n) = \operatorname{argmax}_{1 \leq j \leq M} P(J = j | Y^n = y^n).$$

Solution: This is the MAP (Maximum A Posteriori) decision rule, which ensures that the decoder selects the message with the highest posterior probability given the received sequence.

$$P_e = P(\hat{J}(Y^n) \neq J) = 1 - P(\hat{J}(Y^n) = J)$$

$$\begin{aligned} P_e &= \sum_{y^n} P(Y^n = y^n) P(\hat{J}(y^n) \neq J | Y^n = y^n) && \because \text{Law of Total Probability} \\ &= \sum_{y^n} P(Y^n = y^n) \left(1 - P(\hat{J}(y^n) = J | Y^n = y^n)\right) \end{aligned}$$

We observe that the \hat{J} that minimizes P_e maximizes $P(\hat{J}(y^n) = J | Y^n = y^n)$:

$$\operatorname{argmax}_{1 \leq j \leq M} P(J = j | Y^n = y^n).$$

6. **Output power constraint.** Consider an additive white Gaussian noise channel with an expected *output* power constraint P . Thus $Y = X + Z$, $Z \sim N(0, \sigma^2)$, Z is independent of X , and $EY^2 \leq P$. Find the channel capacity.

Solution: *Output power constraint*

$$C = \max_{f(X): E(X+Z)^2 \leq P} I(X; Y) \tag{15}$$

$$= \max_{f(X): E(X+Z)^2 \leq P} (h(Y) - h(Y|X)) \tag{16}$$

$$= \max_{f(X): E(X+Z)^2 \leq P} (h(Y) - h(Z)) \tag{17}$$

$$\tag{18}$$

Given a constraint on the output power of Y , the maximum differential entropy is achieved by a normal distribution, and we can achieve this by have $X \sim \mathcal{N}(0, P - \sigma^2)$, and in this case,

$$C = \frac{1}{2} \log 2\pi e P - \frac{1}{2} \log 2\pi e \sigma^2 = \frac{1}{2} \log \frac{P}{\sigma^2}. \tag{19}$$