ETH Zürich
Dept. Computer Science

Spring Semester 2016

Information theory

Final exam

18 August 2016

Time limit: 120 minutes
Total points: 100 points

- Please remember to try all the questions before the exam ends. If you are stuck in any part of a problem do not dwell on it, try to move on and attempt it later.
- Collaboration on the exam is strictly forbidden.

Best of luck!

Please fill in your fist name, LAST NAME, and student ID:

Checklist of Questions

Questions	**Points**
1. Secrecy Scenario	/30 points
2. The Sum Channel	/30 points
3. MAP Decoding Over BEC	/40 points
Total	**/100 points**

1 [30 points] Secrecy Scenario

In the simplest secrecy scenario, we have a set of messages to send to a friend of us, but we want to make our message very secure (i.e., encrypt it) so that if an adversary sees the message he can not figure out what our message is.

We can model our message as a random variable X, and in order to make X secure, we use a "secret key" K. In other words, we use the random variable K to encrypt X. The encrypted message Y has the form $Y = f(X, K)$, where f is a deterministic function.

Our friend, who receives the encrypted message Y and also knows secret key K, can decode the message X. In other words, $X = g(Y, K)$ where g is some known deterministic function. Let $\mathcal{X}, \mathcal{Y}$ and $\mathcal{K}$ denote the alphabets of X, Y and K.

Note: In the following, in order to answer each part, it might be helpful to use the result of the previous parts.

a) [5 points] What are the values of $H(Y \mid X, K)$ and $H(X \mid Y, K)$?

b) [5 points] Show that $H(Y \mid K) = H(X \mid K)$.

c) [3 points] Assume additionally that the key K is independent of X. Show that $H(Y) \geq H(X)$.

d) [5 points] Under the same assumption show that $H(Y \mid X) \leq H(K)$.

An encryption system is secure if $I(X;Y) = 0$, i.e., an adversary who observes Y (but does not know K) learns nothing about X.

e) [5 points] Assume that the system is secure and suppose that K is independent of X. Show that $H(K) \geq H(X)$.

f) [2 points] Suppose that the functions f, g and the secret key K are chosen so that the system is secure regardless of the distribution of X (but still assuming that X and K are independent). Show that $H(K) \geq \log |\mathcal{X}|$.

g) [5 points] Show that for $\mathcal{X} = \mathcal{Y} = \mathcal{K} = \{0, \cdots, m-1\}$, the choice: K uniform on $\mathcal{K}$, $f(x,k) = (x+k) \bmod m$, and $g(y,k) = (y-k) \bmod m$ satisfies the assumptions of question f).

2 [30 points] Channel Capacity: The Sum Channel

Let $\mathcal{X} = \mathcal{Y} = \{0, 1, 2, 3\}$ be the input and output alphabets of a discrete memoryless channel with a transition probability matrix

$$W = [\Pr(y \mid x)] = \begin{bmatrix} 1-\varepsilon & \varepsilon & 0 & 0 \\ \varepsilon & 1-\varepsilon & 0 & 0 \\ 0 & 0 & 1-\delta & \delta \\ 0 & 0 & \delta & 1-\delta \end{bmatrix}$$

Such a channel is called a *sum channel*, because it may be thought as the "sum" or union of two parallel (sub-) channels:

$$W_1 = \begin{bmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}, \quad \text{and} \quad W_2 = \begin{bmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{bmatrix}$$

with alphabets $\mathcal{X}_1 = \mathcal{Y}_1 = \{0, 1\}$, and $\mathcal{X}_2 = \mathcal{Y}_2 = \{2, 3\}$ respectively. This problem aims to find the capacity of a sum channel.

a) [3 points] Draw the channel transition diagram for W.

b) [5 points] For the special case of $\varepsilon = \delta = \frac{1}{2}$, show that the capacity is equal to 1 bit.
Hint: You don't need to do any tedious calculations.

c) [4 points] Let $\Pr(x)$ be a probability mass function on $\mathcal{X}$. Let $\alpha = \Pr(X = 0) + \Pr(X = 1)$. Show that the mutual information between the input X and the output Y of the channel W may be written as

$$I(X;Y) = h_2(\alpha) + \alpha I(X;Y \mid X \in \mathcal{X}_1) + (1-\alpha) I(X;Y \mid X \in \mathcal{X}_2),$$

Here, $h_2(\cdot)$ denotes the binary entropy function, i.e., $h_2(\alpha) \triangleq -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$.

d) [4 points] Let C_1 and C_2 denote the capacity of the subchannels W_1 and W_2, respectively. Explain why

$$\max_{\Pr(x)} I(X;Y) = \max_{\alpha} \left[h_2(\alpha) + \alpha C_1 + (1-\alpha) C_2 \right].$$

e) [6 points] Show that the capacity C of the sum channel W is given by

$$C = \log(2^{C_1} + 2^{C_2}),$$

where C_1 and C_2 are the capacities of the subchannels W_1 and W_2.

f) [6 points] Compute the capacity of a channel with input alphabet $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$, output alphabet $\mathcal{Y} = \{1, 2, 3, 4, 5, 6, 7\}$, and channel transition probabilities $\Pr(y \mid x)$ given by

$$[\Pr(y \mid x)] = [w_{xy}] = \begin{bmatrix} \frac{2}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} & 0 & 0 & 0 \\ 0 & 0 & \frac{2}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 \end{bmatrix}$$

Hint: Draw the channel transition diagram.

3 [40 points] MAP Decoding Over BEC

Let $\mathcal{C} = \{\vec{x} \in \{0,1\}^n : H\vec{x} = 0\}$, where H is a given binary matrix. Let $M \triangleq |\mathcal{C}|$ and further assume $\mathcal{C}$ has an (arbitrary) ordering $\mathcal{C} = \{\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_M\}$.

Suppose that we are communicating over a binary erasure channel with erasure probability ε (denoted by BEC (ε)). That is, we have a message set $\mathcal{W} = \{1, \cdots, M\}$, and a message $W \in \mathcal{W}$ is chosen uniformly at random (i.e. $\Pr(W = i) = 1/M$ for $i \in \mathcal{W}$). Once the massage $W = i$ is chosen, its corresponding codeword, $\vec{x}_i$, is sent over the BEC(ε) and the channel outputs $\vec{y} = (y_1, y_2, \cdots, y_n)$. We consider the memoryless setting.

Note that as transmission takes place over a BEC, we have $y_j \in \{0, 1, ?\}$ for $j \in \{1, \cdots, n\}$. Also, the output vector gives some immediate information about the transmitted codeword. E.g. if we have $y_j = 0$ then we immediately know that the transmitted codeword has a 0 at its j-th entry.

In order to estimate the transmitted message, we use the MAP decoder. That is

$$\hat{W}_{\text{MAP}}(\vec{y}) = \arg\max_{i \in \mathcal{W}} \Pr(i \mid \vec{y}) = \arg\max_{i \in \mathcal{W}} \Pr(\vec{x}_i \mid \vec{y}).$$

In this exercise, we intend to derive some properties of the MAP decoder.

a) [3 points] Show that $\mathcal{C}$ is a linear code.

b) [2 points] What is the rate of the code?

c) [3 points] Show that $\arg\max_{i \in \mathcal{W}} \Pr(\vec{y} \mid \vec{x}_i) = \arg\max_{i \in \mathcal{W}} \Pr(\vec{x}_i \mid \vec{y})$.

d) [10 points] Define $e(\vec{y})$ to be the number of entries in $\vec{y}$ which are erased. For example, if $\vec{y} = (0, ?, 1, ?, 0)$ then $e(\vec{y}) = 2$. For each $i \in \mathcal{W}$ simplify the expression $\Pr(\vec{y} \mid \vec{x}_i)$ as much as you can.
Hint: your answer should have ε and $e(\vec{y})$ in it.

e) [2 points] For which indices $i \in \mathcal{W}$ do we have $\Pr(\vec{y} \mid \vec{x}_i) = 0$?

f) [8 points] Prove that finding $\hat{W}_{\text{MAP}}$ is equivalent to solving a linear system of equations.

g) [2 points] What condition should this system of equations have in order to ensure successful decoding (i.e. $\hat{W}_{\text{MAP}}(\vec{y}) = W$)?

h) [10 points] Assume $\mathcal{C}$ has minimum distance d. Prove that MAP decoding is successful if the output vector $\vec{y}$ has at most d erasures.