TUHH
Technische Universität Hamburg-Harburg

INSTITUT FÜR
NACHRICHTENTECHNIK
PROF. DR.-ING. GERHARD BAUCH

Exam WS 2016/2017

SOLUTION

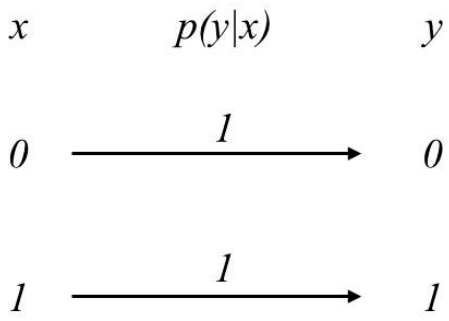
Information Theory and Coding

Name: ……Max Mustermann……..	Immatriculation number: ……123456789….	
	Points	of
Problem 1		18
Problem 2		16
Problem 3		29
Total points		
Grade		

- The following aids are allowed in this exam:
  - 2 Din A4 sheets, handwritten on both sides (4 pages in total)
  - calculator (non-programmable)
  - Pens
- Other aids are not allowed.
- Please use a separate solution sheet for each task.
- Write your name and matriculation number on each solution sheet.
- An arrow next to a question means that this part of the task can be solved independently of the previous questions.
- Please do not write with pencils and do not use a red pen.
- The duration of the exam is 90 minutes.
- The exam consists of **4** pages (including this cover page).
- Switch off your cell phones!

Problem 1: Capacity of the Noiseless Binary Channel

The *noiseless binary channel* is given by the following state transition diagram:



⟹ a) Determine the channel capacity by intuition. Explain your solution clearly and answer in complete sentences.

Answer: The channel causes neither errors nor erasures. Each symbol is received correctly. As the channel input is binary, each symbol carries 1 bit hence only 1 bit can be transmitted per channel use. Therefore the channel capacity is 1 bit/channel use.

⟹ b) Derive the channel capacity formally. Make sure that all steps in your derivation are clearly given.

Answer: The channel capacity, C, is given as:

$$C = \max_{p_X(x)} I(X;Y)$$

Where $I(X;Y)$ is the mutual information between the input and output random variables of the channel and $p_X(x)$ is the probability distribution of the channel input. Then

$$I(X;Y) = \sum_x \sum_y p_{X,Y}(x,y) \log_2 \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}$$

$$= \sum_x \sum_y p_{Y|X}(y|x)p_X(x) \log_2 \frac{p_{Y|X}(y|x)p_X(x)}{p_X(x)p_Y(y)}$$

$$= \sum_x \sum_y p_{Y|X}(y|x)p_X(x) \log_2 \frac{p_{Y|X}(y|x)}{p_Y(y)}$$

The transition probabilities of the channel can be written as:

$$p_{Y|X}(y|x) = \begin{cases} 1 & x = y \\ 0 & x \neq y \end{cases}$$

$p_Y(y)$ can be computed as:

$$p_Y(y=0) = \sum_x p_{Y|X}(y|x)p_X(x)$$

$$= p_{Y|X}(y=0|x=0)p_X(x=0) + p_{Y|X}(y=0|x=1)p_X(x=1)$$
$$= 1.p_X(x=0) + 0.p_X(x=1)$$
$$= p_X(x=0)$$

Similarly, it can be shown

$$p_Y(y=1) = p_X(x=1)$$

Thus we get:

$$p_Y(y) = p_X(x)$$

The mutual information simplifies into:

$$I(X;Y) = \sum_x p_X(x)\log_2\frac{1}{p_Y(x)}$$
$$= p_X(x=0)\log_2\frac{1}{p_X(x=0)} + p_X(x=1)\log_2\frac{1}{p_X(x=1)}$$
$$= H_b(p_X(x=0))$$

So the mutual information between the input and output of the channel is a binary entropy function [This result can also be obtained by using $I(X;Y) = H(Y) - H(Y|X)$ and showing $H(Y) = H_b(p_X(x=0))$ and $H(Y|X) = 0$]. We know that the maximum value of a binary entropy function is 1. Therefore,

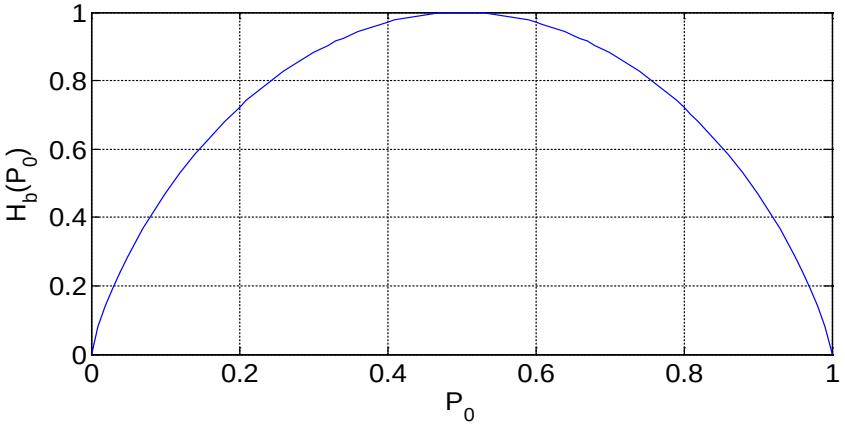
$$C = \max_{p_X(x)} H_b(p_X(x=0)) = 1$$

c) Determine the capacity achieving probability distribution *p(x)* of the channel input symbols.

Answer: As seen in task b), the mutual information in this case is a binary entropy function. The maximum value of the binary entropy function is 1 which is attained for $p_X(x=0) = \frac{1}{2}$. This can be seen in the figure below where $p_X(x=0)$ is labeled as P_0.

Therefore, the capacity achieving probability distribution is:

$$p_X(x=0) = p_X(x=1) = \frac{1}{2}.$$



Problem 2: Maximum Entropy

In this problem, we will prove, that the entropy of a discrete random variable X with N different realizations x and probability distribution $p_X(x)$ is upper bounded by

$$H(X) \leq \log_2 N.$$

We will make use of the *relative entropy* (also called the *Kullback Leibler distance*)

$$0 \leq D(p\|q) = \sum_x p_X(x) \log_2 \frac{p_X(x)}{q_X(x)},$$

which is a measure for the distance of two probability distributions $p_X(x)$ and $q_X(x)$. The relative entropy $D(p||q)$ is always non-negative.

⇨ a) Determine the probability distribution $q_X(x)$ of a uniformly distributed random variable X, which can take N different realizations x.

Answer: Probability distribution of a discrete random variable with N realizations is given as:

$$q_X(x) = \frac{1}{N}$$

b) Determine the entropy $H_u(X)$ of a uniformly distributed random variable X, which can take N different realizations x.

Answer: Entropy is given as:

$$H(X) = \sum_x p_X(x) \log_2 \frac{1}{p_X(x)}$$

For the uniformly distributed random variable:

$$\begin{aligned} H_u(X) &= \sum_x q_X(x) \log_2 \frac{1}{q_X(x)} \\ &= \sum_x \frac{1}{N} \log_2 N \\ &= N \frac{1}{N} \log_2 N \\ &= \log_2 N \end{aligned}$$

c) Determine the relative entropy $D(p||q)$ between a probability distribution $p_X(x)$ and a uniform probability distribution $q_X(x)$ depending on the entropy $H(X)$ of a random variable with probability distribution $p_X(x)$.

Answer: As given in the task

$$D(p||q) = \sum_x p_X(x) \log_2 \frac{p_X(x)}{q_X(x)}$$
$$= \sum_x p_X(x) \left\{ \log_2 p_X(x) + \log_2 \frac{1}{q_X(x)} \right\}$$

Using previous result:

$$D(p||q) = \sum_x p_X(x) \log_2 p_X(x) + \sum_x p_X(x) \log_2 N$$
$$= \sum_x p_X(x) \log_2 p_X(x) + \log_2 N \sum_x p_X(x)$$
$$= -H(X) + \log_2 N \,.\, 1$$
$$= \log_2 N - H(X)$$

d) Use your result from c) in order to prove that $H(X) \leq \log_2 N$.

Answer: As the relative entropy is non-negative, we know that:

$$D(p||q) \geq 0$$

Using previous result:

$$\log_2 N - H(X) \geq 0$$
$$\log_2 N \geq H(X)$$

e) Which probability distribution delivers the maximum possible entropy of a random variable X with N different realizations?

Answer: From task d) we know that the entropy of the discrete random variable X with N realizations is upper bounded by $\log_2 N$. Task b) showed that this maximum entropy is achieved by the uniform distribution i.e.

$$H_u(X) = \max H(X) = \log_2 N$$

Problem 3: Random Codes

When Shannon published his famous paper in 1948, channel codes did not exist. Shannon's theorems have been proven based on the idea of random codes. In this problem, we will address binary random codes, which map information words of length K bits to codeword of length N bits. A random code is constructed by randomly choosing the codewords among all possible binary sequences of length N.

⇨ a) Construct a random code of rate $R=1/2$ with codeword length $N=6$. State all the codewords and the mapping of information words to codewords in a table.

Answer: For given N=6 and R=1/2, we know that

$$K = RN = 3$$

Thus we need to assign randomly selected 6 bit sequences for each of $2^3 = 8$ possible 3-bit information bits sequences. This assignment should be unique. One possible random code is given as:

Information word	Codeword	Codeword label
000	010100	c_0
001	111011	c_1
010	001000	c_2
011	111101	c_3
100	110000	c_4
101	011010	c_5
110	100101	c_6
111	001111	c_7

⇨ b) How many different binary sequences of length N exist?

Answer: There exist 2^N different binary N-bit sequences.

⇨ c) How many different binary sequences of length N with Hamming weight w exist?

Answer: in a binary sequence of Hamming weight w, there are w out of the N bits are 1's. Therefore, the number of such sequences is $\binom{N}{w}$ or ${}^{N}C_{r}$.

⇨ d) How many codewords exist for a binary $(N, K)_2$ code ?

Answer: with a single codeword for each information word, there are 2^K valid codewords.

e) Assume, that a random code has been constructed. We then randomly choose a binary sequence of length N. Determine the probability, that this randomly chosen sequence is a codeword.

Answer: Only 2^K out of the 2^N possible sequences are the codewords. Therefore, for a randomly chosen sequence **x**:

$$P(\mathbf{x}\ is\ a\ codeword) = \frac{\{\#\ of\ valid\ codewords\}}{\{\#\ of\ possible\ sequences\}} = \frac{2^K}{2^N} = 2^{K-N}$$

f) Determine the expected number $E\{A_w\}$ of codewords with Hamming weight w for a random $(N, K)_2$ code.

Answer: From task c) we know that the number of sequences with Hamming weight w is: $\binom{N}{w}$

From task e), probability of a sequence being a codeword is: 2^{K-N}

Therefore, the expected number of weight w codewords is:

$$E\{A_w\} = \binom{N}{w} 2^{K-N}$$

⇨ g) Is a random code in general a linear code ? Give reasons!

Answer: For a code to be linear, sum of any two valid codewords is also a valid codeword. This property cannot be guaranteed if a code is generated randomly. Therefore, a random code is not linear in general.

⇨ h) Which minimum Hamming distance d_{min} can be guaranteed by the random code construction?

Answer: For any code, the codewords need to be unique. Thus any two codewords of a code differ at least in one bit position. Therefore, only a minimum Hamming distance of 1 is guaranteed for randomly generated codes.

i) Determine the minimum Hamming weight w_{min} of your code from a).

Answer: Minimum Hamming weight of the code is 1, as shown in following table:

Information word	Code word	w_H
000	010100	2
001	111011	5
010	001000	1
011	111101	5
100	110000	2
101	011010	3
110	100101	3
111	001111	4

j) Determine the minimum Hamming distance $d_{\min}$ of your code from a).

Answer: Comparison of all possible codeword pairs show that the minimum Hamming distance of the code is $d_H(c_0, c_4) = 2$.