TUHH
Technische Universität Hamburg

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

Exam SS 2017

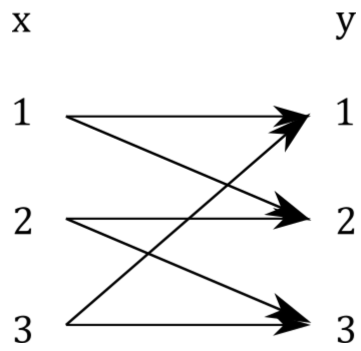
Information Theory and Coding

Name: Sample solution	Student ID:	
	Points	From
Task 1		
Task 2		
Task 3		
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, not graphical, not capable of communication)
  - Pens
  - Dictionary
- 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 rest of the task.
- For calculations the approach as well as the steps must be specified.
- 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 9 pages (including this cover page).
- Switch off your cell phones!

Task 1: Capacity of a noisy three-symbol channel

The figure below shows a noisy three-symbol channel for which the transition probability is given as $p(y|x) = 1/2$, for all possible transitions.



a) Is the channel memoryless? Give reasons and answer in complete sentences!

Yes, the channel is memoryless because the actual output at time k is only influenced by the current input sample at time k, i.e.

$$p(\mathbf{y}|\mathbf{x}) = \prod_k p(y_k|x_k).$$

b) Compute the probability distribution of the received symbols, i.e. $p(y)$.

$$\begin{aligned}
p(y) &= \sum_x p(y|x)p(x) \\
p(y=1) &= \frac{1}{2}p(x=1) + \frac{1}{2}p(x=3) \\
p(y=2) &= \frac{1}{2}p(x=1) + \frac{1}{2}p(x=2) \\
p(y=3) &= \frac{1}{2}p(x=2) + \frac{1}{2}p(x=3)
\end{aligned}$$

c) Show that $p(y)$ will be uniformly distributed if $p(x)$ is uniformly distributed.

$$\begin{aligned}
p(y) &= \sum_x p(y|x)p(x) \\
&= \frac{1}{3} \cdot \sum_x p(y|x)
\end{aligned}$$

The transition probability is either 0.5 if there is a connection or 0 if there is no connection. Each x has two outgoing connections and consequently each y has two incoming connections. Thus $\sum_x p(y|x)$ is always 1.

Hence, $p(y)$ is uniformly distributed.

d) Which probability distribution $p(x)$ maximizes the entropy of $H(Y)$?

The entropy of Y is maximized if $p(y)$ is a uniform distribution. From the previous subtask it is known that $p(y)$ is a uniform distribution if $p(x)$ is a uniform distribution. Hence, $p(x)$ has to be a uniform distribution.

e) Compute the maximum entropy $H_{\max}(Y)$!

$$\begin{aligned}
H_{\max}(Y) &= -\sum_y p(y)\log_2 p(y) \\
&= -\sum_y \frac{1}{3}\log_2\frac{1}{3} \\
&= 3\cdot\frac{1}{3}\log_2 3 \\
&= \log_2 3 \text{ bit}
\end{aligned}$$

f) Determine the conditional entropy $H(Y|X)$!

$$\begin{aligned}
H(Y|X) &= -\sum_x\sum_y p(x,y)\log_2 p(y|x) \\
&= -\sum_x p(x)\underbrace{\sum_y p(y|x)\log_2 p(y|x)}_{\text{either}1/2\log_2\frac{1}{2}\text{ or }0} \\
&= -\sum_x p(x)2\cdot\frac{1}{2}\log_2\frac{1}{2} \\
&= \log_2 2\sum_x p(x) \\
&= 1\text{bit}
\end{aligned}$$

g) Compute the capacity C of the given noisy three-symbol channel.

$$
\begin{aligned}
C &= \max_{p(x)} I(X;Y) \\
&= \max_{p(x)} H(Y) - H(Y|X) \\
&= \max_{p(x)} H(Y) - 1 \\
& H(Y) \text{ is maximized if } p(x) \text{ is a uniform distribution.} \\
&= H_{\text{max}}(Y) - 1 \\
&= \log_2 3 - 1 \\
&= 0.5850 \frac{\text{bit}}{\text{channel use}}
\end{aligned}
$$

Task 2: Analysis of Channel Codes

Assume that we have a channel whose capacity is C = 0.585 bits/channel use. We have two candidate codes.
The first channel code which is investigated in a)-e) is a $(63, 51, 5)_2$ BCH code.

a) How many errors can this code correct?

From the code tuple given in the task description we find $d_{\text{min}} = 5$

$$\left\lfloor \frac{d_{min} - 1}{2} \right\rfloor = \left\lfloor \frac{5 - 1}{2} \right\rfloor$$
$$= 2 \text{ errors can be corrected.}$$

b) How many errors can this code detect?

From the code tuple given in the task description we find $d_{\text{min}} = 5$

$$d_{min} - 1 = 5 - 1$$
$$= 4 \text{ errors can be detected.}$$

c) How many information bits are needed to generate a codeword?

From the code tuple given in the task description we find $K = 51$

d) Compute the code rate R.

From the code tuple given in the task description we find $K = 51$ and $N = 63$

$$R = \frac{K}{N} = \frac{51}{63} = 0.8095 \tag{1}$$

e) In theory, is error-free transmission for the given channel and a code with the rate found in d) possible? Give reasons and answer in complete sentences.

In the channel coding theorem, Shannon tells us that error-free transmission is possible if $R < C$ where C is the channel capacity . From the task description we know that $C = 0.585$. The rate for the BCH code is $R = 0.8095$. Thus error-free transmission is not possible .

The second candidate is a $(171, 133)_8$ convolutional code which was already used during the Voyager mission in 1977.

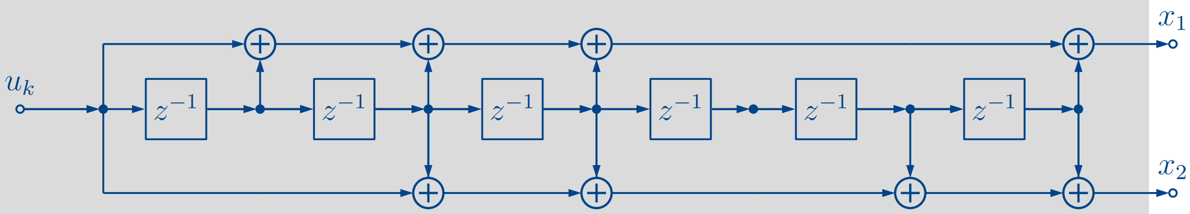
f) What is the rate of this convolutional code?

The given code has two generator polynomials. One information bit is assigned to two code bits. Hence,

$$R = \frac{K}{N} = \frac{1}{2}$$

g) State the generator polynomials and sketch the encoder (shift register).

$$171_8 = 1 + D + D^2 + D^3 + D^6$$
$$133_8 = 1 + D^2 + D^3 + D^5 + D^6$$



h) Theoretically, could you transmit error-free over the given channel with the code rate determined in f)? Give reasons and answer in complete sentences.

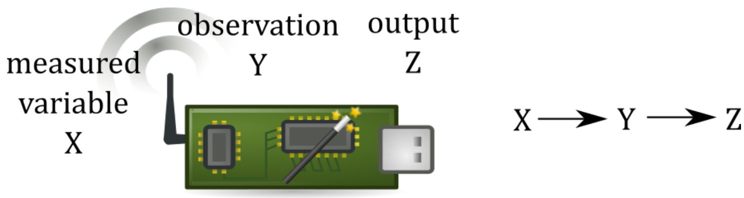
Shannon's channel coding theorem tells us that error-free transmission is possible if $R < C$ where C is the channel capacity . From the task description we know that $C = 0.585$. The rate for the Convolutional Code code is $R = 0.5$. In theory, error-free transmission using a code with this rate would be possible .

i) Which efficient decoding algorithm for a convolutional code exists?

An efficient decoding algorithm is the Viterbi algorithm .

Task 3: Data Processing Inequality

Consider the sensor node depicted below. The observations made by the sensor are denoted as Y and X denotes the quantity which is measured by the sensor node.
You are working in a company and a competitor came up with a product that produces more reliable estimates than your sensor. Your boss has the suspicion that the other company has developed a clever algorithm for the sensor node to gain more information about the measured quantity than contained in Y. The output of this postprocessing is called Z. As indicated below the random variables X, Y, Z form a Markov chain. The Markov chain is characterized by the fact that the next state depends **only** on the current state.



Using the chain rule for mutual information, the mutual information $I(Y, Z; X)$ can be written as

$$I(Y, Z; X) = I(Y; X|Z) + I(Z; X)$$
$$I(Y, Z; X) = I(Z; X|Y) + I(Y; X)$$

a) Show that the joint probability distribution $p(x, y, z)$ is given by

$$p(x, y, z) = p(z|y)p(y|x)p(x)$$

Using the chain rule we rewrite $p(x, y, z)$ as

$$p(x, y, z) = p(z|y, x)p(y|x)p(x).$$

Using the Markov Chain property, that Z is only depend on Y we simpifiy the expression and obtain

$$p(x, y, z) = p(z|y)p(y|x)p(x).$$

Hint: The general chain rule for probability distributions is defined as

$$p(x_1, x_2, \dots, x_n) = p(x_1|x_2, \dots, x_n) \cdot p(x_2|x_3 \dots, x_n) \cdot \dots \cdot p(x_{n-1}|x_n)p(x_n)$$

b) Show, that X and Z are conditionally independent given Y, i.e. $p(x, z|y) = p(x|y)p(z|y)$.

$$p(x,z|y) = \frac{p(x,y,z)}{p(y)}$$

Using the factorization derived in a) we obtain

$$\begin{aligned}
p(x,z|y) &= \frac{p(z|y)p(y|x)p(x)}{p(y)} \\
p(x,z|y) &= \frac{p(z|y)p(y,x)}{p(y)} \\
p(x,z|y) &= p(z|y)p(x|y)
\end{aligned}$$

Thus, $p(x,z|y) = p(x|y)p(z|y)$

c) Use the result from b) to show that $I(Z;X|Y) = 0$.

$$\begin{aligned}
I(Z;X|Y) &= \sum_x \sum_y \sum_z p(x,z|y) \log_2 \left(\frac{p(x,z|y)}{p(x|y)p(z|y)} \right) \\
&= \sum_x \sum_y \sum_z p(x,z|y) \log_2 \left(\frac{p(x|y)p(z|y)}{p(x|y)p(z|y)} \right) \\
&= 0
\end{aligned}$$

d) Show that $I(Z;X) \leq I(Y;X)$ holds. Use the equations for the mutual information from the task description and the results from c). Explain!

From the task description we know that

$$I(Y,Z;X) = I(Y;X|Z) + I(Z;X) = I(Z;X|Y) + I(Y;X)$$

In task c) we showed that $I(Z;X|Y) = 0$. Since, the mutual information is always non-negative

$$I(Z;X) \leq I(Y;X)$$

e) Remember the suspicion of your boss that the mutual information between X and Z will be larger than the information Y carries about X. Is he right? Argue as an information theorist in the light of your results from c) and d). Give reasons and answer in complete sentences!

In the previous subtasks we showed that $I(Z;X) \leq I(Y;X)$. The data processing inequality states, that Z cannot contain more info about X, than Y contains about X. Hence, applying smart signal processing or post-processing cannot increase information. The boss is wrong.