Lund University
Electrical and Information Technology

Home Exam in
Information Theory, EIT080

May 18 – June 1, 2012

Name: ______________________________

Id Number: ______________________________

Programme: ____________

Nbr of sheets: ____________

Mark with a cross the problems you solved.				
1	2	3	4	5

Signature: ______________________________

Control protocol						
1	2	3	4	5	$\sum$	Grade

LUND UNIVERSITY
Electrical and Information Technology

Home Exam in Information Theory, EIT080

May 18 – June 1, 2012

- Write your name on each paper.
- Start a new solution on a new sheet of paper. Use only one side of the paper.
- Solutions should clearly show the line of reasoning.
- Complementing program code, e.g. MATLAB, should be handed in together with the solution. (Normal command line calculations, e.g. entropy, is not required to hand in, if the solution show the line of reasoning.)

Good luck!

Problem 1

The symbols in a binary sequence $\boldsymbol{x} = x_0 x_1 x_2 x_3 \dots$ are considered to be i.i.d. and generated according to the probabilities $P(X=0) = \frac{1}{4}$ and $P(X=1) = \frac{3}{4}$. To compress the sequence, bits are blocked in 3-tuples and coded. A suggested code is given by

$x_i x_{i+1} x_{i+2}$	$\boldsymbol{y}$
000	00000
001	00001
010	0001
011	0101
100	0100
101	001
110	011
111	1

Is the suggested code an optimal prefix-free code for the source? If not, construct an optimal prefix-free code. Encode the sequence

$$\boldsymbol{x} = 111\ 011\ 101\ 101\ 111\ 110\ 011\ 011 \dots$$

(10p)

Problem 2

A text has been encoded with the LZ78 algorithm and the following sequence of codewords was obtained,

Index	Codeword
1 :	$(0, t)$
2 :	$(0, i)$
3 :	$(0, m)$
4 :	$(0, ␣)$
5 :	$(1, h)$
6 :	$(0, e)$
7 :	$(4, t)$
8 :	$(0, h)$
9 :	$(2, n)$
10 :	$(7, w)$
11 :	$(9, ␣)$
12 :	$(1, i)$
13 :	$(0, n)$
14 :	$(0, s)$
15 :	$(3, i)$
16 :	$(5, .)$

Decode to get the text back.

(10p)

Problem 3

On the next page there is a table containing the statistics for the first order approximation of English text, i.e. the probability for the next letter letter being x_j conditioned that the current letter is x_i. There is also one column for the unconditioned letter probability for the same statistics.

The table can also be found as a matrix for MATLAB in the file `TableProb.mat` on the same place as the exam load with `load('TableProb.mat')`. If you want it in some other format, e.g. text or Excel, send a mail to `stefan.host@eit.lth.se`.

Assuming that the source is compressed by two encoders,

- An optimal binary code that is based on the conditional probabilities.
- An optimal binary code that is based on the unconditional probabilities.

What is the gain in average codeword length by using the conditional probabilities compared to the unconditional probabilities?

(10p)

$P(x_j \mid x_i)$ (x_i \ x_j)	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	␣	$P(x_i)$
a	0.0018	0.0163	0.0490	0.0220	0.0025	0.0065	0.0274	0.0020	0.0210	0.0019	0.0064	0.1099	0.0450	0.1778	0.0004	0.0199	0.0005	0.1544	0.0913	0.1351	0.0159	0.0153	0.0041	0.0066	0.0220	0.0021	0.0430	0.0802
b	0.0682	0.0116	0.0252	0.0023	0.1794	0.0002	0.0012	0.0019	0.0907	0.0072	0.0001	0.1387	0.0008	0.0232	0.1151	0.0001	0	0.0454	0.0671	0.0031	0.1256	0.0003	0.0003	0	0.0788	0	0.0134	0.0147
c	0.1098	0.0002	0.0352	0.0009	0.1787	0.0008	0.0004	0.1081	0.1084	0.0001	0.0250	0.0507	0.0014	0.0006	0.1561	0.0002	0.0004	0.0311	0.0135	0.0945	0.0270	0.0003	0.0001	0.0001	0.0087	0.0006	0.0474	0.0311
d	0.0832	0.0026	0.0017	0.0098	0.1627	0.0039	0.0088	0.0018	0.1213	0.0009	0.0001	0.0058	0.0035	0.0011	0.0413	0.0003	0.0013	0.0184	0.0241	0.0009	0.0378	0.0016	0.0105	0	0.0082	0.0002	0.4480	0.0278
e	0.0510	0.0093	0.0297	0.0672	0.0181	0.0488	0.0090	0.0023	0.0084	0.0004	0.0023	0.0392	0.0203	0.0824	0.0079	0.0082	0.0025	0.1433	0.0928	0.0324	0.0030	0.0129	0.0097	0.0097	0.0099	0.0005	0.2790	0.1024
f	0.0468	0.0003	0.0024	0	0.0489	0.0335	0.0002	0.0001	0.1529	0	0	0.0207	0.0010	0.0014	0.1106	0	0	0.0707	0.0018	0.0338	0.0226	0.0008	0.0002	0.0001	0.0014	0.0001	0.4498	0.0193
g	0.0600	0.0017	0.0003	0.0074	0.2076	0.0011	0.0077	0.1098	0.0715	0	0	0.0623	0.0036	0.0468	0.0531	0.0005	0	0.0718	0.0156	0.0081	0.0350	0.0009	0.0002	0.0001	0.0246	0.0001	0.2101	0.0161
h	0.1079	0.0009	0.0008	0.0004	0.4879	0.0001	0.0002	0.0000	0.1005	0	0.0007	0.0039	0.0023	0.0048	0.0586	0.0018	0.0007	0.0135	0.0031	0.0595	0.0142	0.0002	0.0010	0.0000	0.0168	0	0.1202	0.0395
i	0.0328	0.0123	0.0786	0.0285	0.0427	0.0107	0.0289	0.0010	0.0007	0.0002	0.0046	0.0540	0.0267	0.2036	0.0920	0.0094	0.0013	0.0485	0.1351	0.1425	0.0065	0.0263	0.0001	0.0016	0.0011	0.0057	0.0047	0.0689
j	0.0697	0.0039	0.0020	0.0020	0.1356	0.0010	0.0059	0	0.0039	0.0039	0.0010	0.0010	0	0.0010	0.5982	0.0884	0.0010	0.0020	0.0157	0.0010	0.0550	0.0049	0	0	0	0.0010	0.0020	0.0017
k	0.0493	0.0045	0.0012	0.0015	0.2221	0.0015	0.0185	0.0066	0.1411	0	0.0036	0.0164	0.0520	0.0427	0.0164	0.0054	0	0.0036	0.0798	0.0146	0.0066	0.0015	0.0054	0	0.1249	0	0.1808	0.0054
l	0.1839	0.0026	0.0036	0.0347	0.1736	0.0049	0.0016	0.0004	0.1285	0.0002	0.0109	0.0928	0.0042	0.0008	0.0663	0.0019	0.0000	0.0006	0.0252	0.0286	0.0414	0.0034	0.0018	0	0.0570	0	0.1307	0.0432
m	0.1855	0.0446	0.0009	0.0007	0.2720	0.0013	0.0015	0.0007	0.1004	0.0002	0.0009	0.0249	0.0229	0.0026	0.1129	0.0675	0	0.0007	0.0236	0.0004	0.0205	0.0001	0.0014	0.0001	0.0142	0.0002	0.0991	0.0222
n	0.0947	0.0282	0.0466	0.1289	0.0766	0.0051	0.0954	0.0015	0.0409	0.0008	0.0073	0.0062	0.0028	0.0126	0.0410	0.0004	0.0008	0.0012	0.0534	0.1023	0.0109	0.0058	0.0007	0.0002	0.0079	0.0012	0.2267	0.0617
o	0.0058	0.0123	0.0205	0.0278	0.0024	0.1077	0.0159	0.0027	0.0105	0.0007	0.0104	0.0610	0.0630	0.1993	0.0334	0.0308	0.0002	0.1228	0.0376	0.0376	0.0756	0.0223	0.0283	0.0048	0.0027	0.0007	0.0632	0.0659
p	0.1528	0.0010	0.0031	0.0058	0.1501	0.0005	0.0100	0.0876	0.0506	0.0002	0.0016	0.0855	0.0111	0.0027	0.1137	0.0381	0	0.1202	0.0200	0.0306	0.0843	0	0.0005	0.0059	0.0018	0	0.0223	0.0186
q	0.0064	0	0.0021	0	0	0.0021	0	0	0.0042	0.0021	0	0	0.0064	0.0042	0	0.0021	0.0021	0	0.0021	0	0.9597	0	0	0	0	0	0.0064	0.0008
r	0.0860	0.0097	0.0199	0.0164	0.2265	0.0059	0.0174	0.0011	0.0876	0.0001	0.0182	0.0287	0.0222	0.0313	0.0951	0.0046	0.0002	0.0121	0.0714	0.0550	0.0133	0.0146	0.0012	0.0011	0.0329	0.0002	0.1274	0.0631
s	0.0184	0.0058	0.0203	0.0136	0.0786	0.0021	0.0005	0.0502	0.0691	0.0001	0.0139	0.0060	0.0101	0.0011	0.0357	0.0487	0.0017	0.0012	0.0577	0.2176	0.0457	0.0007	0.0009	0	0.0049	0.0005	0.2951	0.0564
t	0.0630	0.0008	0.0024	0.0001	0.1458	0.0001	0.0003	0.2501	0.1363	0	0.0002	0.0272	0.0095	0.0004	0.0722	0.0129	0.0001	0.0455	0.0289	0.0255	0.0236	0.0001	0.0071	0.0001	0.0238	0.0007	0.1232	0.0791
u	0.0310	0.0545	0.0421	0.0247	0.0440	0.0026	0.0285	0.0006	0.0204	0	0.0032	0.1032	0.0634	0.1294	0.0023	0.0343	0.0001	0.2168	0.0953	0.0935	0.0018	0.0006	0.0003	0.0008	0.0011	0.0012	0.0041	0.0286
v	0.1084	0.0010	0	0	0.5567	0.0002	0.0031	0	0.1876	0	0.0002	0.0014	0.0004	0.0227	0.0975	0	0	0.0025	0.0008	0	0.0017	0.0004	0.0004	0	0.0060	0.0002	0.0089	0.0084
w	0.1543	0.0008	0.0020	0.0010	0.1826	0.0007	0.0005	0.1311	0.1637	0	0.0007	0.0060	0.0014	0.0358	0.0837	0.0015	0	0.0301	0.0352	0.0044	0.0007	0	0.0996	0.0001	0.0007	0.0003	0.0631	0.0119
x	0.0614	0.0015	0.0422	0	0.0362	0.0333	0	0.0074	0.2308	0	0	0.0007	0.0059	0	0.0111	0.1731	0	0	0.0022	0.1827	0.0022	0.0022	0.0007	0.0007	0.1435	0	0.0621	0.0022
y	0.0173	0.0038	0.0139	0.0125	0.1246	0.0024	0.0058	0.0004	0.0087	0.0001	0.0009	0.0245	0.0149	0.0145	0.0174	0.0140	0	0.0047	0.0709	0.0058	0.0017	0.0005	0.0010	0	0.0005	0.0078	0.6313	0.0131
z	0.2568	0	0.0061	0.0015	0.3116	0	0.0061	0.0304	0.0729	0	0.0030	0.0137	0.0091	0.0091	0.1429	0.0046	0.0046	0	0.0030	0.0076	0.0076	0	0	0	0.0198	0.0122	0.0775	0.0011
␣	0.1268	0.0425	0.0498	0.0303	0.0259	0.0402	0.0118	0.0291	0.0720	0.0049	0.0034	0.0268	0.0358	0.0301	0.0905	0.0395	0.0008	0.0276	0.0588	0.1659	0.0106	0.0083	0.0501	0	0.0049	0.0005	0.0130	0.1167

Problem 4

The two dimensional continuous random variable has the density function according to

$$f(x,y) = \begin{cases} A, & x > 0, \quad y > 0, \quad ax + by < ab \\ 0, & \text{otherwise} \end{cases}$$

The area in $\mathbb{R}^2$ where $f(x,y) = A$ can be seen to be the shaded area in Figure 4.1.

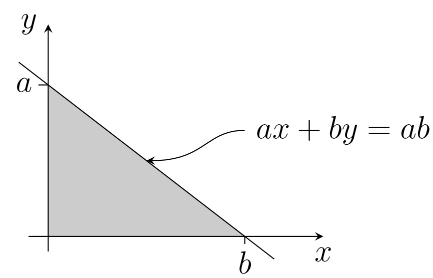


Figure 4.1: The area where $f(x,y) = A$.

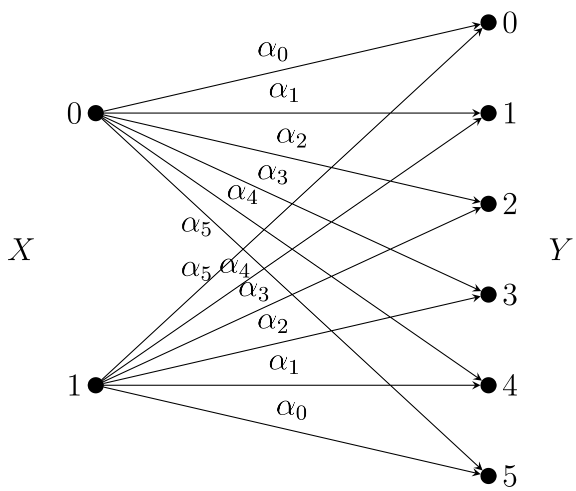
(a) Determine the constant A.
(b) Calculate the joint differential entropy $H(X,Y)$.
(c) Calculate the mutual information $I(X;Y)$.

> **Hint:** For a random variable Z distributed according to a normalized triangular distribution
>
> $$f(z) = 2 - 2z, \quad 0 \leqslant z \leqslant 1$$
>
> the entropy is
>
> $$H(Z) = -\int_0^1 (2-2z)\log(2-2z)dz = \frac{1}{2\ln 2} - 1$$
>
> Furthermore, a scaling of the random variable gives
>
> $$H(rZ) = H(Z) + \log r$$
>
> where r is a positive constant.

(3+3+4=10p)

Problem 5

Consider the discrete memory-less channel (DMC) given in Figure 5.1.



$P(Y\|X)$		Y 0	1	2	3	4	5
X	0	α_0	α_1	α_2	α_3	α_4	α_5
	1	α_5	α_4	α_3	α_2	α_1	α_0

Figure 5.1: A channel and its probability function.

(a) Show that the maximizing distribution giving the capacity

$$C_6 = \max_{p(x)} I(X;Y)$$

is given by $P(X=0) = \frac{1}{2}$ and $P(X=1) = \frac{1}{2}$.
Verify that the capacity is given by

$$C_6 = 1 + H(\alpha_0 + \alpha_5, \alpha_1 + \alpha_4, \alpha_2 + \alpha_3) - H(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$$

(b) Split the outputs in two sets, $\mathcal{Y}_0 = \{0,1,2\}$ and $\mathcal{Y}_1 = \{3,4,5\}$, and construct a binary symmetric channel, i.e. a BSC with error probability $p = \alpha_3 + \alpha_4 + \alpha_5$. Denote the capacity of the corresponding BSC as C_{BSC} and show that

$$C_{\text{BSC}} \leqslant C_6 \leqslant 1$$

where C_6 is the capacity of the channel in Figure 5.1.

(5+5=10p)