Problem 1

Encoding according to

S buffer	B buffer	Codeword
’I screem, you sc’	’reem, we’	(12,6,’w’)
’m, you screem, w’	’e all sc’	(6,1,’ ’)
’ you screem, we ’	’all scre’	(7,1,’l’)
’ou screem, we al’	’l scream’	(1,1,’ ’)
’ screem, we all ’	’scream f’	(15,6,’ ’)
’, we all scream ’	’for icec’	(0,0,’f’)
’ we all scream f’	’or icecr’	(0,0,’o’)
’we all scream fo’	’r icecre’	(7,1,’ ’)
’ all scream for ’	’icecream’	(0,0,’i’)
’all scream for i’	’cecream.’	(11,1,’e’)
’l scream for ice’	’cream.’	(13,5,’.’)

There are 11 codewords and an initialisation vector of 16 letters, giving $11(5+4+8)+16\cdot 8 = 315$ bits. (The codeword length can also be argued to be $4+3+8=15$ bits, but according to the course book it should be $\lceil 16+1\rceil + \lceil 8+1\rceil + 8 = 5+4+8 = 17$). The uncoded length is $49\cdot 8 = 392$ bits. Then the compression ratio is $R = 392/315 = 1.24$.

Problem 2

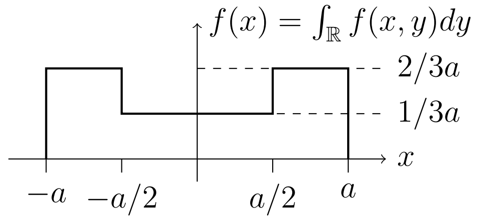
(a) To simplify notations, let $\mathcal{B}$ denote the shaded region in the figure. Then, since the area of $\mathcal{B}$ is $3ab$, the density function is

$$f(x,y) = \begin{cases} \frac{1}{3ab}, & x,y \in \mathcal{B} \\ 0, & x,y \notin \mathcal{B} \end{cases}$$

The entropy is

$$H(X,Y) = -\int_{\mathcal{B}} \frac{1}{3ab}\log\frac{1}{3ab}dxdy = \log\frac{3}{a}b\int_{\mathcal{B}}\frac{1}{3ab}dxdy = \log 3ab$$

(b) To get $f(x)$, integrate $f(x,y)$ over y, to get



Then the entropy of X can be derived as

$$\begin{aligned} H(X) &= -\int_{-a}^{-a/2}\frac{2}{3a}\log\frac{2}{3a}dx - \int_{-a/2}^{a/2}\frac{1}{3a}\log\frac{1}{3a}dx - \int_{a/2}^{a}\frac{2}{3a}\log\frac{2}{3a}dx \\ &= a\frac{2}{3a}\log\frac{3a}{2} + a\frac{1}{3a}\log 3a = \log 3a - \frac{2}{3} \end{aligned}$$

Similarly, $H(Y) = \log 3b - \frac{2}{3}$.

(c) The mutual information is

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$
$$= \log 3a - \frac{2}{3} + \log 3b - \frac{2}{3} - \log 3ab = \log 3 - \frac{4}{3}$$

(d) Since $I(X;Y) = H(X) - H(X|Y)$ we get

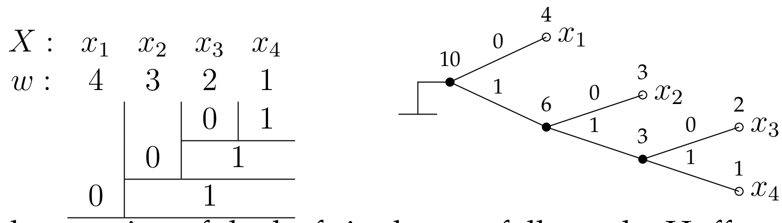
$$H(X|Y) = H(X) - I(X;Y) = \log 3a - \frac{2}{3} - \log 3 + \frac{4}{3} = \frac{2}{3} + \log a$$

Similarly, $H(Y|X) = \frac{2}{3} - \log b$.

Problem 3

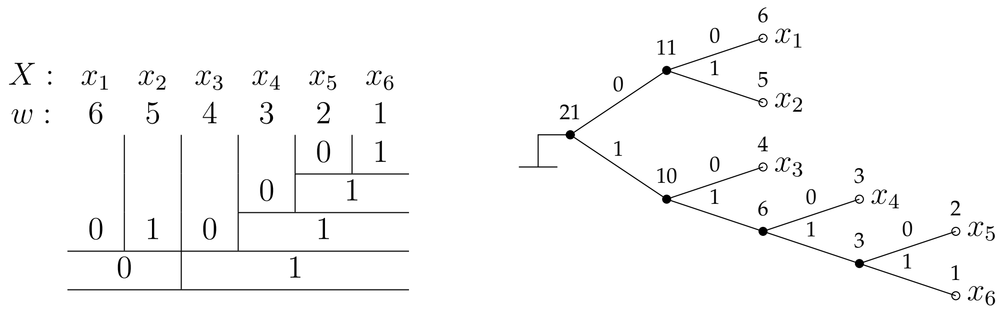
For simpicity, for each sub-problem the common denominator in the probabilities, is dropped and the numerator is used as weight in the algorithm.

(a) In the first example the weights for the outcomes are $w(x_1) = 4$, $w(x_2) = 3$, $w(x_3) = 2$ and $w(x_4) = 1$. The first split separates $\{x_1\}$ in one part and $\{x_2, x_3, x_4\}$ in the other. The first set is marked with 0 and the second with 1. The second set is split again into $\{x_2\}$ and $\{x_3, x_4\}$. Finally the last part is split into $\{x_3\}$ and $\{x_4\}$. Since all sets now contain only one outcome each there is no more splitting. By marking the subsets in each split by 0 and 1, a code is obtained. Below, to the left, the propcedure is shown. To the right the corresponding code tree is shown.



Since the merging of the leafs in the tree follows the Huffman procedure it is a Huffman code, and hence optimal.

(b) In the second example the weights are $w(x_1, x_2, x_3, x_4, x_5, x_6) = (6, 5, 4, 3, 2, 1)$. Following the same procedure as in (a), we get

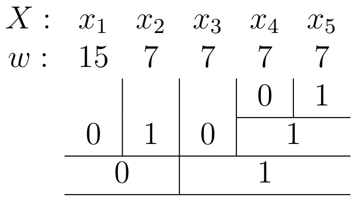


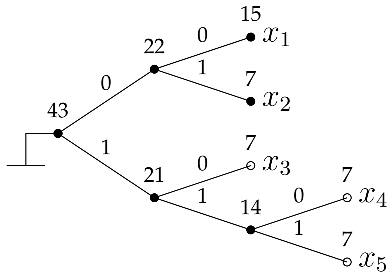
When constructing a Huffman code, first the leaves x_5 and x_6 are merged, then $\{x_5x_6\}$ and x_4 are merged. After this the nodes in the algorithm are $(x_1, x_2, x_3, \{x_4x_5x_6\})$ with weights $(6, 5, 4, 6)$. So in the next step in the Huffman procedure the nodes x_2 and x_3 are merged. This is not the case in the tree above where x_3 and $\{x_4x_5x_6\}$ are merged, and hence the code is not a Huffman code. Continuing the Huffman procedude results in the code tabulated below.

X	w	Y
x_1	6	10
x_2	5	01
x_3	4	00
x_4	3	110
x_5	2	1110
x_6	1	1111

The average codeword length for the Huffman code is 51/21, and according to the path length lemma the codeword length for the Fano code is $L_F = \frac{21+11+10+6+3}{21} = \frac{51}{21}$. Hence the code is optimal.

(c) Following the same structure for the third code gives the following.





The average codeword length is $L_F = \frac{43+22+21+14}{43} = \frac{100}{43}$. When constructing a Huffman code the nodes x_4 and x_5 are merged in the first step. In the second step x_2 and x_3 are merged, which is not the case in the tree for the Fano code. Hence the obtained code is not a Huffman code. In the following table a Huffman code is shown.

X	w	Y
x_1	15	0
x_2	7	100
x_3	7	101
x_4	7	110
x_5	7	111

The average codeword length is $L_H = \frac{99}{43}$. Hence, the Fano code is neither a Huffman code nor optimal.

Problem 4

In the solution the following data is used

$M = m$	J	F	M	A	M	J	J	A	S	O	N	D
DH_m	241	269	369	426	505	522	521	463	381	321	249	220
SH_m	37	64	105	166	231	235	223	212	141	94	52	32
MH_m	744	672	744	720	744	720	744	744	720	744	720	744

where DH_m is the day hours per month, SH_m the sun hours and MH_m the total hours per month. There are three random variables; the month M, time of day T that can take the values $\{D, N\}$ for Day and Night, and weather W that can take the values $\{S, C\}$ for Sunny and Cloudy.

(a) The obtained information about M by observing T can be derived as

$$I(M;T) = H(T) - H(Y|M)$$

For this we need the following probabilities, estimated from the number of positive outcomes divided by the total number of outcomes.

$$P(T = D) = \frac{\sum_i DH_i}{\sum_i SH_i} \approx 0.51$$
$$P(T = D|M = m) = \frac{DH_m}{SH_m}$$
$$P(M = m) = \frac{MH_m}{\sum_i MH_i}$$

The two latter is shown in the following table

$M = m$	J	F	M	A	M	J	J	A	S	O	N	D
$P(T = D\|M = m)$	0.3239	0.4003	0.4960	0.5917	0.6788	0.7250	0.7003	0.6223	0.5292	0.4315	0.3458	0.2957
$P(M = m)$	0.0849	0.0767	0.0849	0.0822	0.0849	0.0822	0.0849	0.0849	0.0822	0.0849	0.0822	0.0849

The entropies are

$$H(T) = h\big(P(T = D)\big) \approx 0.9996$$
$$H(T|M) = \sum_m h\big(P(T = D|M = m)\big)P(M = m) \approx 0.9358$$

Hence, the mutual information is $I(M;T) = 0.0634$ bit

(b) In this part the function $I(W; M|T = D) = H(W|T = D) - H(W|M, T = D)$ is derived. The following probabilities are estimated.

$$P(W = S|T = D) = \frac{\sum_i SH_i}{\sum_i DH_i} \approx 0.3550$$
$$P(W = S|M = m, T = D) = \frac{SH_m}{DH_m}$$
$$P(M = m|T = D) = \frac{DH_m}{\sum_i DH_i}$$

which gives the following table

$M = m$	J	F	M	A	M	J	J	A	S	O	N	D
$P(W = S\|M = m, T = D)$	0.1535	0.2379	0.2846	0.3897	0.4574	0.4502	0.4280	0.4579	0.3701	0.2928	0.2088	0.1455
$P(M = m\|T = D)$	0.0537	0.0600	0.0822	0.0949	0.1125	0.1163	0.1161	0.1032	0.0849	0.0715	0.0555	0.0490

The entropies are derived as

$$H(W|T = D) = h\big(P(W = S|T = D)\big) \approx 0.9385$$
$$H(W|M, T = D) = \sum_m h\big(P(W = S|M = m, T = D)\big)P(M = m|T = D) \approx 0.9015$$

Hence, the information is $I(W; M|T = D) \approx 0.0370$ bit.

Problem 5

(a) The maximum amount of information per transmission is given by the mutual information, $I(X; Y) = H(Y) - H(Y|X)$. Here

$$H(Y|X) = H(Z) = -\int_{-1.5}^{-0.5} \frac{1}{4} \log \frac{1}{4} dz - \int_{-0.5}^{0.5} \frac{1}{2} \log \frac{1}{2} dz - \int_{0.5}^{1.5} \frac{1}{4} \log \frac{1}{4} dz = \frac{3}{2}$$

Since the outcomes for X are equally likely, in this case the density function for Y becomes

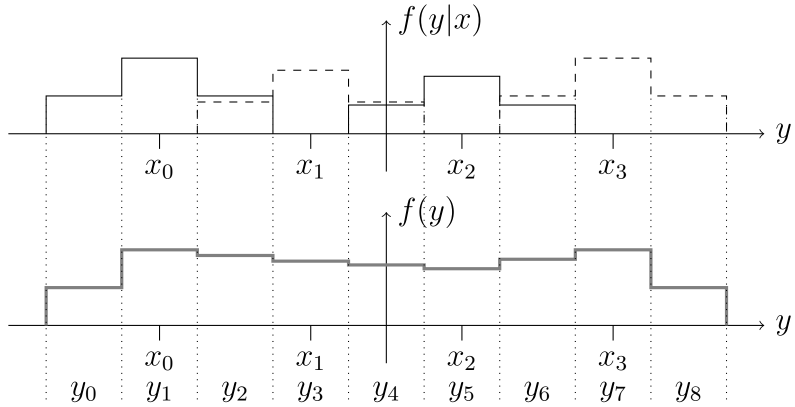
$$f(y) = \begin{cases} \frac{1}{8}, & -3.5 \leqslant y \leqslant 3.5 \\ \frac{1}{16}, & -4.5 \leqslant y < -3.5 \text{ and } 3.5 < y \leqslant 4.5 \\ 0, & \text{o.w.} \end{cases}$$

which gives the entropy

$$H(Y) = -\int_{-4.5}^{-3.5} \frac{1}{16} \log \frac{1}{16} dy - \int_{-3.5}^{3.5} \frac{1}{8} \log \frac{1}{8} dy - \int_{3.5}^{4.5} \frac{1}{16} \log \frac{1}{16} dy = \frac{25}{8}$$

and thus, $I(X;Y) = \frac{25}{8} - \frac{3}{2} = \frac{13}{8} \approx 1.625$ bit/transmission

(b) When assuming that X is not equally distributed, still $H(Y|X) = H(Z) = \frac{3}{2}$. So to maximise $I(X;Y)$ we need to maximise $H(Y)$. In the upper figure below the density functions for $\{Z|X\}$ are shown. These sum up to give the density function for Y. Since it is composed of flat areas, i.e. intervalls in which $f(y)$ is constant, it is possible to construct an equivalent DMC with symbols $\{y_0, y_1, \ldots, y_8\}$ corrsponding to intervals.



To maximise over all distributions on X we can by symmetry reasons set $P(x_0) = P(x_3) = p$ and $P(x_1) = P(x_2) = \frac{1}{2} - p$. Then the distributions on the intervals becomes

$$P(y_i) = \begin{cases} \frac{p}{4}, & i = 0, 8 \\ \frac{p}{2}, & i = 1, 7 \\ \frac{1}{8}, & i = 2, 6 \\ \frac{1}{4} - \frac{p}{2}, & i = 3, 4, 5 \end{cases}$$

and the entropy

$$\begin{aligned} H(Y) &= -2\frac{p}{4} \log \frac{p}{4} - 2\frac{p}{2} \log \frac{p}{2} - 2\frac{1}{8} \log \frac{1}{8} - 3(\frac{1}{4} - \frac{p}{2}) \log(\frac{1}{4} - \frac{p}{2}) \\ &= \cdots = \frac{3}{4} h(2p) + \frac{p}{2} + \frac{9}{4} \end{aligned}$$

Setting the derivative equal to zero gives

$$\frac{\partial}{\partial p} H(Y) = \frac{3}{4}\Big(2\log(1-2p) - 2\log 2p\Big) + \frac{1}{2} = \frac{3}{2} \log \frac{1-2p}{2p} + \frac{1}{2} = 0$$

or, equivalently,

$$p = \frac{1}{2^{2/3} + 2}$$

which gives $H(Y) \approx 3.1322$ and $I(X;Y) \approx 1.6322$ bit/transmission.

The average power is increased from $E[X^2] = 5$ for equaly distribution to $E[X^2] = 2 \cdot 3^2 p + 2(\frac{1}{2} - p) \approx 5.46$ for the optimal distribution.