Problem 1

First derive the estimated length of the given code. The probabilities and the lengths for the 3-tuples are given in the following table. This gives an average length of $E[L] = 2.52$bit/symbol. To check the optimality, construct a Huffman code, as in the table (maked $\boldsymbol{y}_{opt}$ and L_{opt}). This gives an optimal length of $E[L_{opt}] = 2.47$bit/symbol.

$\boldsymbol{x}$	$\boldsymbol{y}$	$p(\boldsymbol{x})$	$L(\boldsymbol{y})$	$\boldsymbol{y}_{opt}$	L_{opt}
000	00000	$1/64 \approx 0.016$	5	00000	5
001	00001	$3/64 \approx 0.047$	5	00001	5
010	0001	$3/64 \approx 0.047$	4	00010	5
011	0101	$9/64 \approx 0.14$	4	010	3
100	0100	$3/64 \approx 0.047$	4	00011	5
101	001	$9/64 \approx 0.14$	3	011	3
110	011	$9/64 \approx 0.14$	3	001	3
111	1	$27/64 \approx 0.42$	1	1	1

The sequence given is encoded as follows:

$$\boldsymbol{y}_{opt} = 10100110111001010010\ldots$$

Problem 2

The decoding is done in the following table.

Index	Codeword	Dictionary (text)
1 :	$(0, t)$	t
2 :	$(0, i)$	i
3 :	$(0, m)$	m
4 :	$(0, \text{␣})$	␣
5 :	$(1, h)$	th
6 :	$(0, e)$	e
7 :	$(4, t)$	␣t
8 :	$(0, h)$	h
9 :	$(2, n)$	in
10 :	$(7, w)$	␣tw
11 :	$(9, \text{␣})$	in␣
12 :	$(1, i)$	ti
13 :	$(0, n)$	n
14 :	$(0, s)$	s
15 :	$(3, i)$	mi
16 :	$(5, .)$	th.

Hence, the text is "tim the thin twin tinsmith.".

Problem 3

The matrix in this problem is too large for paper and pen derivations, so the actual calculations such as matrix inversion is better done in e.g. MATLAB or Python. For the case when the uncoditional probabilities are used, the optimal codeword length is

$$L = H(X) = H(0.082, 0.0147, \dots, 0.1167) = 4.1975$$

For the conditioned case we should use the entropy rate. Then, we construct a MArkoz process where each letter is a state and the arrows are labeled with the conditional probabilities. That is, the state transition probability matrix,P, is given by the matrix for the conditional probabilities,

$$P = \begin{pmatrix} 0.0018 & \cdots & 0.0021 & 0.0430 \\ \vdots & \ddots & \vdots & \vdots \\ 0.2568 & \cdots & 0.0122 & 0.0775 \\ 0.1268 & \cdots & 0.0005 & 0.0130 \end{pmatrix}$$

To derive the staeady state distribution we let $\boldsymbol{w} = (w_a, \dots, w_{\llcorner\lrcorner})$ and consider the equation

$$\begin{aligned} \boldsymbol{w}(P - I) &= \mathbf{0} \\ \sum_j w_j &= 1 \end{aligned}$$

Since neither $\boldsymbol{w}$ nor $P - I$ are zero, the matrix $P - I$ is not of full rank. Therefore, we replace one of its columns with the sexcond equation (written as $\boldsymbol{w}\mathbf{1}^T = 1$, where $\mathbf{1}$ denotes the all-one vector $\mathbf{1} = (1 \dots 1)$). Then we get the matrix equation

$$\boldsymbol{w} \begin{pmatrix} -0.9982 & \cdots & 0.0021 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ 0.2568 & \cdots & -0.9878 & 1 \\ 0.1268 & \cdots & 0.0005 & 1 \end{pmatrix} = (0 \ \cdots \ 0 \ 1)$$

Solving this yields,

$$\begin{aligned} \boldsymbol{w} = \ &(0.0741 \ 0.0151 \ 0.0284 \ 0.0288 \ 0.1053 \ 0.0204 \ 0.0159 \ 0.0365 \ 0.0659 \ 0.0011 \\ &0.0058 \ 0.0422 \ 0.0221 \ 0.0599 \ 0.0605 \ 0.0183 \ 0.0008 \ 0.0586 \ 0.0576 \ 0.0823 \\ &0.0239 \ 0.0082 \ 0.0125 \ 0.0021 \ 0.0148 \ 0.0011 \ 0.1377) \end{aligned}$$

Then the conditional entropies $H(X_j|X_i = x_i)$ is the entropy for each row in the table, which gives

$$\begin{aligned} H(X_j|X_i = x_i) = \ &(3.6984 \ 3.4710 \ 3.4600 \ 2.6915 \ 3.6035 \ 2.6357 \ 3.3997 \ 2.4892 \ 3.6618 \ 2.0508 \\ &3.3371 \ 3.3981 \ 3.1457 \ 3.5635 \ 3.8884 \ 3.5763 \ 0.3764 \ 3.7364 \ 3.2620 \ 3.3268 \\ &3.5954 \ 1.9417 \ 3.2335 \ 3.0656 \ 2.1685 \ 2.8735 \ 4.0770) \end{aligned}$$

Then the entropy rate becomes

$$H_\infty(X) = \sum_i w_i H(X_j|X_i = x_i) = 3.4970$$

So, the loss made when not considering the consditional statistics is $H(X) - H_\infty(X) = 0.7005$ bit/letter.

In MATLAB this can be done with (where the entropy function from the home problems is used):

```
%% Load matrix from exam
load('TableProb.mat')
%% unconditioned
H1=Entropy(Pc(:,end))
%% conditioned
P=Pc(:,1:end-1); % Remove uncoditional prob
% steady state distr
A=P-eye(27);
A=[A(:,1:end-1) ones(27,1)];
y=[zeros(1,26) 1];
pss=y*inv(A)
% Entropy rate
Hinf=sum(Entropy(P').*pss)
%% Compare
Lgain=H1-Hinf
```

Problem 4

(a) To get A use that the volume of $f(x,y)$ is 1,

$$1 = \int_{\triangle} A dx dy = A \int_{\triangle} 1 dx dy = A\frac{ab}{2} \quad \Rightarrow \quad A = \frac{2}{ab}$$

(b) Calculate the entropy,

$$H(X,Y) = -\int_{\triangle} A \log A dx dy = -\log A \int_{\triangle} A dx dy = \log\frac{1}{A} = \log\frac{ab}{2} = \log ab - 1$$

(c) To get the mutual information we need to derive the entropy for X and Y separatly. Starting with X we get the density function

$$f(x) = \int_{y=0}^{-\frac{a}{b}x+a} \frac{2}{ab} dy = \frac{2}{ab}\Big(-\frac{a}{b}x + a\Big) = -\frac{2}{b^2}x + \frac{2}{b}$$

This is a triangular distribution starting at $f(0) = \frac{2}{b}$ and decreasing linearly to $f(b) = 0$. From the hint we know that the entropy of a variable Z with triangular distribution of width 1 is $H(Z) = \frac{1}{2\ln 2} - 1 = \log\sqrt{e} - 1$. Then, if Z is scaled with the constant b we get

$$H(X) = H(Z) + \log b = \log b\sqrt{e} - 1 = \log b + \frac{1}{2\ln 2} - 1$$

Similarly, the entropy of Y is $H(Y) = \log a\sqrt{e} - 1$. Then the mutual information between X and Y becomes

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &= \log b\sqrt{e} - 1 + \log a\sqrt{e} - 1 - \log ab + 1 \\ &= \log e - 1 = \log\frac{e}{2} = \frac{1}{\ln 2} - 1 \end{aligned}$$

Problem 5

(a) The mutual information between X and Y is

$$I(X;Y) = H(Y) - H(Y|X)$$
$$= H(Y) - \sum_{i=0}^{1} H(Y|x=i)P(x=i) = H(Y) - H(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$$

So, what is left to optimize is $H(Y)$. From the probability table we see that $p(y = j|x = 0) = \alpha_j$ and $p(y = j|x = 1) = \alpha_{5-j}$. If we assume that the probability of X is given by $p(x = 0) = p$ and $p(x = 1) = 1 - p$, then the joint probability is given by $p(y = j, x = 0) = p\alpha_j$ and $p(y = j, x = 1) = (1 - p)\alpha_{5-j}$. Hence, we can write the probability for Y as $p(y = j) = p\alpha_j + (1 - p)\alpha_{5-j}$ and the entropy as

$$H(Y) = \sum_{j=0}^{5} (p\alpha_j + (1-p)\alpha_{5-j}) \log(p\alpha_j + (1-p)\alpha_{5-j})$$

The corresponding derivative with respect to p is

$$\frac{\partial}{\partial p} H(Y) = \sum_{j=0}^{5} (\alpha_j - \alpha_{5-j}) \Big(\log(p\alpha_j + (1-p)\alpha_{5-j}) + \frac{1}{\ln 2} \Big)$$

Then, setting $p = \frac{1}{2}$ and splitting in two sums we get

$$\frac{\partial}{\partial p} H(Y)\Big|_{p=\frac{1}{2}} = \sum_{j=0}^{2} (\alpha_j - \alpha_{5-j}) \Big(\frac{1}{2\ln 2} + \log(\alpha_j + \alpha_{5-j}) \Big) + \sum_{j=3}^{5} (\alpha_j - \alpha_{5-j}) \Big(\frac{1}{2\ln 2} + \log(\alpha_j + \alpha_{5-j}) \Big)$$

In the second sum replace the summation variable with $n = 5 - j$, then

$$\frac{\partial}{\partial p} H(Y)\Big|_{p=\frac{1}{2}} = \sum_{j=0}^{2} (\alpha_j - \alpha_{5-j}) \Big(\frac{1}{2\ln 2} + \log(\alpha_j + \alpha_{5-j}) \Big) + \sum_{n=0}^{2} (\alpha_{5-n} - \alpha_n) \Big(\frac{1}{2\ln 2} + \log(\alpha_{5-n} + \alpha_n) \Big)$$

Since $(\alpha_{5-n} - \alpha_n) = -(\alpha_n - \alpha_{5-n})$ we get two identical sums with different sign,

$$\frac{\partial}{\partial p} H(Y)\Big|_{p=\frac{1}{2}} = \sum_{j=0}^{2} (\alpha_j - \alpha_{5-j}) \Big(\frac{1}{2\ln 2} + \log(\alpha_j + \alpha_{5-j}) \Big) - \sum_{j=0}^{2} (\alpha_j - \alpha_{5-j}) \Big(\frac{1}{2\ln 2} + \log(\alpha_j + \alpha_{5-j}) \Big) = 0$$

and we have seen that $p = \frac{1}{2}$ maximizes $H(Y)$. (Here the maximima follows from the fact that the entropy is a concave function.)

Then, for $p = \frac{1}{2}$, we get

$$\begin{aligned} H(Y) &= -\sum_{j=0}^{5} \frac{1}{2}(\alpha_j + \alpha_{5-j}) \log \frac{1}{2}(\alpha_j + \alpha_{5-j}) \\ &= \frac{1}{2}\sum_{j=0}^{5}(\alpha_j + \alpha_{5-j}) - \frac{1}{2}\sum_{j=0}^{5}(\alpha_j + \alpha_{5-j}) \log(\alpha_j + \alpha_{5-j}) \\ &= 1 - \frac{1}{2}\Big(\sum_{j=0}^{5}(\alpha_j + \alpha_{5-j}) \log(\alpha_j + \alpha_{5-j}) + \sum_{n=0}^{2}(\alpha_n + \alpha_{5-n}) \log(\alpha_n + \alpha_{5-n})\Big) \\ &= 1 - \sum_{j=0}^{2}(\alpha_j + \alpha_{5-j}) \log(\alpha_j + \alpha_{5-j}) = 1 + H(\alpha_0 + \alpha_5, \alpha_1 + \alpha_4, \alpha_2 + \alpha_3) \end{aligned}$$

Hence, the capacity is

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

(b) The right hand inequality is straight forward since

$$C_6 \leqslant I(X;Y) = H(X) - H(X|Y) \leqslant H(X) \leqslant \log|\mathcal{X}| = 1$$

For the left hand inequalty we first derive the capacity for the corresponding BSC. The error probability is $p = \alpha_3 + \alpha_4 + \alpha_5$, hence,

$$C_{\text{BSC}} = 1 - h(p) = 1 - H(\alpha_0 + \alpha_1 + \alpha_2, \alpha_3 + \alpha_4 + \alpha_5)$$

So, to show that $C_{\text{BSC}} \leqslant C_6$ we should show that

$$\begin{aligned} C_6 - C_{\text{BSC}} = 1 &+ H(\alpha_0 + \alpha_5, \alpha_1 + \alpha_4, \alpha_2 + \alpha_3) + H(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) \\ &- 1 + H(\alpha_0 + \alpha_1 + \alpha_2, \alpha_3 + \alpha_4 + \alpha_5) \end{aligned}$$

is non-negative. For this we introduce a new pair of random variables A and B with the joint distribution and marginal distributions according to

$P(A,B)$		$B=0$	$B=1$	$B=2$
A	0	α_0	α_1	α_2
	1	α_5	α_4	α_3

A	$P(A)$
0	$\alpha_0 + \alpha_1 + \alpha_2$
1	$\alpha_3 + \alpha_4 + \alpha_5$

B	$P(B)$
0	$\alpha_0 + \alpha_5$
1	$\alpha_1 + \alpha_4$
2	$\alpha_2 + \alpha_3$

Then we can identify in the capacity formula above

$$\begin{aligned} C_6 - C_{\text{BSC}} &= 1 + H(B) - H(A,B) - 1 + H(A) \\ &= H(A) + H(B) - H(A,B) = I(A;B) \geqslant 0 \end{aligned}$$

which is the desired result. (The above inequality can also be obtained from the IT-inequality).