[E4.40, SO20]

Notation:

(a) $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

(b) $\log x$ denotes the logarithm to base 2.

1. A binary prefix code, *C*, maps input symbol x_i having probability p_i into a bit string $C(x_i)$ of length l_i bits. We define $L_C = \sum p_i l_i$.

(a) Explain what is meant by saying that C is a "prefix code" and show that since this is the case, the lengths l_i must satisfy $\sum 2^{-l_i} \le 1$. [4]

(b) Prove that $L_C \ge H(X)$, the entropy of the input symbols. You may assume without proof that if q_i is any other probability mass vector, then $D(\mathbf{p} \| \mathbf{q}) = \sum p_i \log(p_i / q_i) \ge 0$. [6]

(c) If $\mathbf{p} = [0.65 \quad 0.1 \quad 0.1 \quad 0.05 \quad 0.05 \quad 0.03 \quad 0.02]$, calculate the Shannon code lengths $l_i = \lceil -\log p_i \rceil$ and construct a code using these lengths. Calculate L_C for this code and also the entropy of the source, $H(X)$. [5]

(d) Construct a Huffman code for the probability mass vector given in part (c) and calculate L_C. [5]

2. The random vector $X_{1:n}$ contains n independent identically distributed elements each with alphabet $\boldsymbol{\mathcal{X}} = [a,b,c]$ and probability vector $\mathbf{p} = [0.6 \quad 0.2 \quad 0.2]$.

We define the "typical set" as $T_\varepsilon^{(n)} = \left\{\mathbf{x} \in \boldsymbol{\mathcal{X}}^n : \left| -n^{-1} \log p(\mathbf{x}) - H(X) \right| < \varepsilon \right\}$.

(a) Determine the entropy, $H(X)$ where X refers to a single element of $X_{1:n}$. [3]

(b) Describe the characteristics that distinguish the elements of $T_\varepsilon^{(n)}$ from other vectors and determine the range of values of ε for which the following vectors lie in $T_\varepsilon^{(n)}$: (i) aabacabaca, (ii) aaaaaabbbb, (iii) aaaabbbccc, (iv) bbbbbccccc [5]

(c) Calculate the mean and variance of $Y_n = -n^{-1}\sum_{i=1}^{n} \log p(X_i)$. Using the approximation that Y_n is normally distributed, calculate the value of N_0 so that $\forall n > N_0,\ p(|Y_n - H(X)| > \varepsilon) < \varepsilon$ where $\varepsilon = 10^{-3}$. [6]

Note: If V is Gaussian with zero mean and unit variance then $p(|V| > 3.291) = 10^{-3}$.

(d) When $\varepsilon = 10^{-3}$ prove that the size of $T_\varepsilon^{(n)}$ is bounded by $(1-\varepsilon)2^{n(H(X)-\varepsilon)} < \left|T_\varepsilon^{(n)}\right| \le 2^{n(H(X)+\varepsilon)}$ when n is greater than the N_0 determined in (c) above. For $n = N_0$ determine the $\log_{10}$ of (i) the ratio between these bounds and (ii) the ratio of the upper bound to the total number of strings of length n. [6]

3. For a discrete memoryless channel with input X and output Y, the capacity is $C = \max_{\mathbf{p}_X} I(X;Y)$ where the maximum is taken over all probability mass vectors, $\mathbf{p}_X$ for X.

Throughout this question, the input alphabet $\boldsymbol{\mathcal{X}} = [0,1]$ and $\mathbf{p}_X = [1-u, u]$.

(a) Suppose that the channel is a binary symmetric channel with output alphabet $\boldsymbol{\mathcal{Y}} = [0,1]$ and transition probability matrix $\mathbf{Q}_{Y|X} = \begin{pmatrix} g & f \\ f & g \end{pmatrix}$ where $f + g = 1$. Derive an expression for $I(X;Y)$ and show that this is maximized when $u = ½$. [4]

(b) Suppose that the channel is a binary erasure channel with output alphabet $\boldsymbol{\mathcal{Y}} = [0, e, 1]$ and transition probability matrix $\mathbf{Q}_{Y|X} = \begin{pmatrix} g & f & 0 \\ 0 & f & g \end{pmatrix}$ where $f + g = 1$. Derive an expression for $I(X;Y)$ and show that this is maximized when $u = ½$. [4]

(c) Show that if H is the entropy of X, then $\frac{dH}{du} = \log(u^{-1} - 1)$. [2]

(d) Suppose that the channel is a "Z channel" with output alphabet $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}}$ and transition probability matrix $\mathbf{Q}_{Y|X} = \begin{pmatrix} 1 & 0 \\ f & g \end{pmatrix}$ where $f + g = 1$. Show that $I(X;Y)$ is maximized when $u = \frac{f^{f/g}}{1 + gf^{f/g}}$. [6]

(e) For the channel of part (d), determine C and the value of u that achieves it for (i) $f = 0.99$, (ii) $f = 0.5$ and (iii) $f = 0$. Comment on the significance of these results. [4]

4. The Gaussian random variable X has probability density function $p(x) = (2\pi)^{-½} e^{-½x^2}$.

(a) Prove that the differential entropy $h(X) = ½\log(2\pi e)$. [4]

(b) Show that if the random variable Z, not necessarily Gaussian, has probability density function $f(z)$ and unit variance, then $h(Z) \le ½\log(2\pi e)$. You may assume without proof that $D(f \,\|\, p) \ge 0$. [4]

(c) $\hat{X}$ is a quantized version of X with $E(X - \hat{X})^2 \le Q$. Justify steps (i) to (v) of the following argument and say, for each inequality, the circumstances under which equality will apply:

$$
\begin{aligned}
I(X;\hat{X}) &= h(X) - h(X \mid \hat{X}) \\
&\overset{(i)}{=} ½\log 2\pi e - h(X - \hat{X} \mid \hat{X}) \\
&\overset{(ii)}{\ge} ½\log 2\pi e - h(X - \hat{X}) \\
&\overset{(iii)}{\ge} ½\log 2\pi e - ½\log\left(2\pi e \operatorname{Var}\left(X - \hat{X}\right)\right) \\
&\overset{(iv)}{\ge} ½\log 2\pi e - ½\log 2\pi e Q \\
I(X;\hat{X}) &\overset{(v)}{\ge} \max\left(-½\log Q, 0\right)
\end{aligned}
$$

Sketch a graph showing how this bound on $I(X;\hat{X})$ varies with Q. [5]

(d) For a 1-bit quantizer, $\hat{X} = \pm k$. Determine the value of k that minimizes $Q = E\left(X - \hat{X}\right)^2$ and find the resultant value of Q.

Compare this value of Q with the bound implied by part (c) and outline how it would be possible to come closer to this bound. [7]

5. A power-constrained Gaussian channel has input X and output $Y = X + Z$ where Z is independent of X and is zero-mean Gaussian with variance N. The input power constraint is $E\,X^2 \le P$.

In this question you may assume without proof that if $V_{1:n}$ is an n-dimensional Gaussian random variable with covariance matrix $\mathbf{K}$, then $h(V_{1:n}) = ½ n \log 2\pi e + ½ \log|\mathbf{K}|$

(a) Justify the steps labelled (i) to (iii) in the following equation:

$$h(Y \mid X) \overset{(i)}{=} h(X + Z \mid X) \overset{(ii)}{=} h(Z \mid X) \overset{(iii)}{=} h(Z)$$ [3]

(b) Justify the steps labelled (i) to (iii) in the following inequality and state clearly the conditions under which equality will apply in step (ii):

$$I(X;Y) \overset{(i)}{=} h(Y) - h(Y \mid X)$$
$$\overset{(ii)}{\le} ½ \log 2\pi e(P+N) - ½ \log 2\pi e N$$ [4]
$$\overset{(iii)}{=} ½ \log(1 + PN^{-1})$$

(c) Suppose the same power-constrained input X is transmitted over two Gaussian channels having noise variances N_1 and N_2 and outputs Y_1 and Y_2 respectively. Determine an upper bound for $I(X;Y_1,Y_2)$. [7]

(d) If Y_1 and Y_2 are linearly combined as $Y = aY_1 + (1-a)Y_2$, calculate the value of a that maximizes the capacity of the channel $X \to Y$ and find the resultant capacity. [6]

[E4.40, SO20]

6. Each pixel, x_i, in a 64-pixel image can take one of four intensities 0, 1, 2 or 3 with equal probability where $i = 1, \cdots, 64$. The pixel may be represented as arising from a Markov process with transition probabilities

$$t_{jk} = p(x_{i+1} = k \mid x_i = j) = \begin{cases} 0.97 & j = k \\ 0.01 & j \neq k \end{cases} \text{ for } i = 1:63$$

(a) Show that the vector $[\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]^T$ is an eigenvector of the transposed transition matrix and explain why this is significant. [2]

(b) Pixel values are encoded as 2-bit binary numbers and transmitted over a memoryless binary symmetric channel having an error probability of 0.001. Determine the probability that the entire image is received without error. [2]

(c) To encode the pixels differentially, we form $y_i = x_i - x_{i-1}$ where the subtraction is performed modulo 4 and we take $x_0 = 0$. Determine the entropy of Y_i for (i) $i = 1$ and (ii) $i > 1$. [3]

(d) The pairs (y_i, y_{i+1}) are jointly coded using codeword lengths of $[1,4,4,4,4,4,4,8,8,8,8,8,8,8,8,8]$ bits. Show that it is possible to construct a prefix code with these lengths and explain how they should be assigned to the pairs (y_i, y_{i+1}) to minimize the expected codeword length. With this assignment, determine the expected codeword length for (i) (y_1, y_2) and (ii) (y_i, y_{i+1}) for $i > 2$. [4]

(e) Determine the probability of transmitting an encoded pair (y_i, y_{i+1}) without error over the channel defined in part (b) above. Hence determine the probability of transmitting the entire image over the channel without error. [4]

(f) If each encoded pair (y_i, y_{i+1}) is transmitted three times and is decoded by taking a majority decision at the receiver, determine the expected number of bits used to transmit the image and the probability of receiving the image without error. [5]

1. (a) A prefix code is one in which no codeword is a prefix of another.

If we form a binary tree and place each codeword at its corresponding node, the prefix condition is precisely that no codeword lies on the path to any other. If we label codewords on the tree at depth l with the value 2^{-l}, then we can show recursively that the sum of all codeword values below a particular node never exceeds 2^{-l} and in particular the sum of the codewords below the root node does not exceed 1. [4]

(b) We define a new probability mass function as

$$q_i = c^{-1}2^{-l_i} \text{ where } c = \sum_i 2^{-l_i} \le 1$$

Then

$$L_C - H(X) = \sum_i p_i l_i + \sum_i p_i \log p_i$$

$$= E\left(-\log 2^{-l(X)} + \log p(X)\right) = E_{\mathbf{p}}\left(-\log cq(X) + \log p(X)\right) \qquad [6]$$

$$= E\left(\log \frac{p(X)}{q(X)}\right) - \log c = D(\mathbf{p} \,\|\, \mathbf{q}) - \log c \ge 0$$

Alternative proof using $\ln(x) \le x - 1 \text{ for } x > 0$:

$$H(X) - L_C = \sum_i p_i \log p_i^{-1} 2^{-l_i} = \log e \sum_i p_i \ln p_i^{-1} 2^{-l_i}$$

$$\le \log e \sum_i p_i \left(p_i^{-1} 2^{-l_i} - 1\right) = \log e \left(\sum_i 2^{-l_i} - 1\right) \le 0$$

(c) $\mathbf{l} = \lceil 0.62 \quad 3.3 \quad 3.3 \quad 4.3 \quad 4.3 \quad 5.1 \quad 5.6 \rceil = [1 \quad 4 \quad 4 \quad 5 \quad 5 \quad 6 \quad 6]$

This gives the codewords

$$[0 \quad 1000 \quad 1001 \quad 10100 \quad 10101 \quad 101100 \quad 101100]$$

And the expected length is $\sum p_i l_i = 2.25\,\text{bits}$. The entropy of the source is $\sum -p_i \log p_i = 1.7652\,\text{bits}$ so the Shannon code is not especially good. [5]

(d) The Huffman construction is

0.65	0.65	0.65	0.65	0.65	0.65	1
0.1	0.1	0.1	0.1	0.2	0.35	
0.1	0.1	0.1	0.1			
0.05	0.05	0.05	0.15	0.15		
0.05	0.05	0.1				
0.03	0.05					
0.02						

[E4.40, SO20]

This gives the codewords

$$[0 \quad 100 \quad 101 \quad 110 \quad 1110 \quad 11110 \quad 11111]$$

and an expected length of $\sum p_i l_i = 1.85$ bits. [5]

2. (a) $H(X) = \sum - p_i \log p_i = 1.371$ bits [3]

(b) The typical set consists of strings whose log probability is close to the mean value. Since $n^{-1} \log p(\mathbf{x}) = \frac{n_a}{n} \log 0.6 + \frac{n_b + n_c}{n} \log 0.2$ this will be close to $H(X)$ if the number of "a" characters is close to 60% of the total. Note that the relative numbers of "b" and "c" characters do not matter since they have the same probability. Thus we have (i) $\varepsilon > 0$ (ii) $\varepsilon > 0$, (iii) $\varepsilon > 0.317$ and (iv) $\varepsilon > 0.951$. [5]

(c) $E - \log p(X_i) = H(X) = 1.371$ bits.

$$\text{Var}\left(-\log p(X_i)\right) = \sum_j p_j \log^2 p_j - H^2(X) = 2.4824 - 1.8795 = 0.6029 \text{ bits}^2$$

Hence the mean and variance of Y_n are 1.371 and $0.6029 n^{-1}$. From the information in the question we want

$$\varepsilon = 10^{-3} = 3.891\sigma = 3.291\sqrt{0.6029 n^{-1}}$$

From this we get $n = 6.53 \times 10^6$ [6]

(d) To obtain the lower bound we assume that all elements of $T_\varepsilon^{(n)}$ have the maximum possible probability and that the total probability is as low as possible. Hence for $n > N_0$,

$$1 - \varepsilon < p(\mathbf{x} \in T_\varepsilon^{(n)}) \le \sum_{\mathbf{x} \in T_\varepsilon^{(n)}} 2^{-n(H(X)-\varepsilon)} = 2^{-n(H(X)-\varepsilon)} \left|T_\varepsilon^{(n)}\right|$$

$$\Rightarrow \left|T_\varepsilon^{(n)}\right| \ge (1 - \varepsilon) 2^{n(H(X)-\varepsilon)}$$

For the upper bound, we assume that all elements of $T_\varepsilon^{(n)}$ have the lowest possible probability and that the total probability is as high as possible (i.e. 1):

$$1 = \sum_{\mathbf{x}} p(\mathbf{x}) \ge \sum_{\mathbf{x} \in T_\varepsilon^{(n)}} p(\mathbf{x}) \ge \sum_{\mathbf{x} \in T_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)} = 2^{-n(H(X)+\varepsilon)} \left|T_\varepsilon^{(n)}\right|$$

$$\Rightarrow \left|T_\varepsilon^{(n)}\right| \le 2^{n(H(X)+\varepsilon)}$$

Plugging in the values, the $\log_{10}$ of the ratio between the bounds is

$$n(H + \varepsilon - H + \varepsilon)\log_{10} 2 - \log_{10}(1 - \varepsilon) = 2n\varepsilon \log_{10} 2 - \log_{10}(1 - \varepsilon) = 3931$$

and the $\log_{10}$ of the ratio beween the upper bound and the total number of strings is

$$n(H + \varepsilon)\log_{10} 2 - n \log_{10} 3 = -418710$$ [6]

3. (a) We can calculate the probability vector
$\mathbf{p}_Y = \mathbf{Q}^T \mathbf{p}_X = \left(g + (f-g)u \quad f-(f-g)u\right)^T$

We also see that since the rows of $\mathbf{Q}$ are permutations of each other, $H(Y\,|\,X) = H(f)$. Hence

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y\,|\,X) \\
&= H\big(f-(f-g)u\big) - H(f) \\
&= H\big(f-(2f-1)u\big) - H(f)
\end{aligned}
$$

This is maximized when $f-(2f-1)u = ½ \quad \Rightarrow \quad u = ½$ [4]

(b) This time it is easiest to decompose the other way:

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X\,|Y) \\
&= H(X) - H(X\,|Y=e)\,p(Y=e) \\
&= H\big(X\big) - H(X)f \\
&= (1-f)H(X)
\end{aligned}
$$

This is clearly maximized by maximizing $H(X)$, i.e. when $u = ½$. [4]

Alternatively:
$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y\,|\,X) \\
&= H([(1-u)g \quad f \quad ug]) - H(f)
\end{aligned}
$$

which is maximized when the two terms that depend on u are equal.

(c) We can differentiate the expression for $H(X)$ directly:

$$
\begin{aligned}
H &= -(1-u)\log(1-u) - u\log u \\
\frac{dH}{du} &= \log(1-u) + \log e - \log u - \log e \\
&= \log(1-u) - \log u \\
&= \log(u^{-1} - 1)
\end{aligned}
$$
[2]

(d) $\mathbf{p}_Y = \mathbf{Q}^T \mathbf{p}_X = \left(1-u+fu \quad gu\right)^T = \left(1-gu \quad gu\right)^T$

$H(Y\,|\,X) = (1-u)\times 0 + uH(f)$

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y\,|\,X) \\
&= H\big(gu\big) - uH(f)
\end{aligned}
$$

$$
\frac{dI}{du} = g\log\big(g^{-1}u^{-1}-1\big) - H(f) = g\log\big(g^{-1}u^{-1}-1\big) + f\log f + g\log g
$$

[E4.40, SO20]

$$\frac{dI}{du} = 0 \quad \Rightarrow \quad \log\left(g^{-1}u^{-1} - 1\right) = -\log f^{\,f/g} - \log g$$

$$g^{-1}u^{-1} - 1 = g^{-1} f^{\,-f/g}$$

$$g^{-1}u^{-1} = g^{-1} f^{\,-f/g} + 1 = \frac{1 + gf^{\,f/g}}{gf^{\,f/g}}$$ [6]

$$u = \frac{f^{\,f/g}}{1 + gf^{\,f/g}}$$

Alternatively: $I(X;Y) = H(X) - H(X\,|Y)$

$$= H(u) - (1 - ug)H(\,fu(1 - ug)^{-1})$$

(e) Substituting into the above expressions gives (i) $u = 0.3684, C = 0.0053$, (ii) $u = 0.4, C = 0.3219$. For (ii), we note that the channel is now noise-free so $u = 0.5, C = 1$. As the channel degrades, we send a slightly larger fraction of “0” bits but it is interesting and surprising that u never decreases below e^{-1}. [4]

4. (a) We can integrate directly:

$$h(X) = -(\log e)\int_{-\infty}^{\infty} f(x)\ln f(x)dx$$
$$= -(\log e)\int_{-\infty}^{\infty} f(x)\left(-½x^2 - ½\ln(2\pi)\right)$$
$$= ½(\log e)\left(\ln(2\pi) + E\left(x^2\right)\right)$$
$$= ½(\log e)\left(\ln(2\pi) + 1\right) = ½\log\left(2\pi e\right) \text{ bits}$$

[4]

(b) If the p.d.f. of Z is $q(z)$, then following the hint in the question

$$0 \le D(f \,\|\, p) = -h(Z) - E_f \log p(Z)$$
$$\Rightarrow h(Z) \le -(\log e)E_f\left(-½\ln 2\pi - ½z^2\right)$$
$$\le ½(\log e)(\ln 2\pi + 1)$$
$$= ½\log(2\pi e)$$

[4]

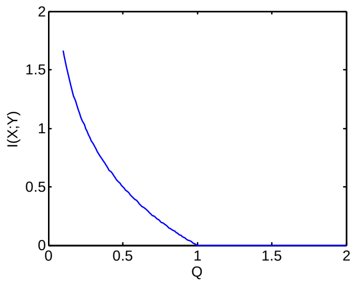
(c) (i) $h(X)$ is known since X is Gaussian. $h(X - \hat{X} \mid \hat{X})$ is unchanged since the translation is by a (conditional) constant.

(ii) Conditioning reduces entropy. Equality when $X - \hat{X}$ is independent of $\hat{X}$.

(iii) Gaussian bound on entropy for a given variance. Equality when $X - \hat{X}$ is Gaussian.

(iv) Variance cannot exceed mean square value and this is less than Q. Equality when $E\left(X - \hat{X}\right) = 0$ and $Var\left(X - \hat{X}\right) = Q$.

(v) Algebra + Mutual information is bounded below by 0.



[5]

(d) The decision boundary is at $X = 0$. If we define $m = E(X \mid x > 0)$. Then for $x > 0$, we have $\hat{X} = +k$ and

$$E\left(X - \hat{X}\right)^2 = E(X - m)^2 + (k - m)^2$$

which is clearly minimized when $k = m = E(X \mid x > 0)$.

[E4.40, SO20]

We have

$$k = 2\int_0^\infty xp(x)dx = 2(2\pi)^{-½}\int_0^\infty xe^{-½x^2}dx = 2(2\pi)^{-½}\left[-e^{-½x^2}\right]_0^\infty = (2/\pi)^{½} = 0.7979$$

$$Q = -k^2 + 2(2\pi)^{-½}\int_0^\infty x^2 e^{-½x^2}dx = -k^2 + \text{Var}(X) = 1-k^2 = 1-2/\pi = 0.3634$$

For a 1-bit quantizer $I(X;\hat{X}) \le 1$ bit so part (c) implies that (i) the minimum rate for this distortion is $-½\log 0.3634 = 0.7302$ bits/s or alternatively that the minimum distortion for a 1-bit quantizer is $Q = 2^{-2} = 0.25$.

To approach these values, we would have to use vector quantization and quantize several bits at a time. [7]

5. (a) (i) $Y = X + Z$

(ii) X is conditionally constant and $h(Z)$ is invariant under translation by a constant.

(iii) Z is independent of X. [3]

(b) (i) Definition of mutual information.

(ii) Since X and Z are independent, the variance of $\operatorname{Var} Y = \operatorname{Var} X + \operatorname{Var} Z \le P + N$. A Gaussian maximizes the entropy for a given variance and so $h(Y) \le ½\log 2\pi e(P+N)$. From the previous part, $h(Y \mid X) = h(Z) = ½\log 2\pi eN$.

(iii) This is just algebra. [4]

(c) **Intended Solution**: The argument is the same as in part (b) but with two output signals. We need to know the following covariances:

$$\mathbf{K}_Z = \operatorname{Cov}\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} N_1 & 0 \\ 0 & N_2 \end{pmatrix} \quad \Rightarrow \quad |\mathbf{K}_Z| = N_1 N_2$$

$$\mathbf{K}_Y = \operatorname{Cov}\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \operatorname{Cov}\begin{pmatrix} X \\ X \end{pmatrix} + \mathbf{K}_N = \begin{pmatrix} P_X + N_1 & P_X \\ P_X & P_X + N_2 \end{pmatrix}$$

$$\Rightarrow \quad |\mathbf{K}_Y| = P_X(N_1 + N_2) + N_1 N_2 \le P(N_1 + N_2) + N_1 N_2$$

$$\begin{aligned} I(X;Y_1,Y_2) &= h(Y_1,Y_2) - h(Y_1,Y_2 \mid X) \\ &\le \log 2\pi e + ½\log\left(P(N_1+N_2) + N_1N_2\right) - \log 2\pi e - ½\log N_1N_2 \\ &= ½\log\left(1 + PN_1^{-1} + PN_2^{-1}\right) \end{aligned}$$ [7]

However the question did not specify which upper bound and a **much easier solution** is:

$$I(X;Y_1,Y_2) = I(X;Y_1) + I(X;Y_2 \mid Y_1) \le I(X;Y_1) + I(X;Y_2)$$
$$\le ½\log\left(1 + PN_1^{-1}\right) + ½\log\left(1 + PN_2^{-1}\right)$$

(d) $Y = aY_1 + (1-a)Y_2 = X + aZ_1 + (1-a)Z_2 = X + Z$

where $Z = aZ_1 + (1-a)Z_2$

The variance of Z is $N = \operatorname{Var}(Z) = a^2 N_1 + (1-a)^2 N_2$

$$\frac{dN}{da} = 0 = 2aN_1 - 2(1-a)N_2 \quad \Rightarrow \quad a = \frac{N_2}{N_1 + N_2} \quad \Rightarrow \quad N = \frac{N_1 N_2}{N_1 + N_2}$$

Using the expression from part (b) now gives

$C = ½\log\left(1 + PN^{-1}\right) = ½\log\left(1 + PN_1^{-1} + PN_2^{-1}\right)$ which interestingly enough is the same as in part (c). [6]

6. (a) The transition matrix is given by

$$\mathbf{T} = \begin{pmatrix} 0.97 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.97 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.97 & 0.01 \\ 0.01 & 0.01 & 0.01 & 0.97 \end{pmatrix}$$

Since each column sums to 1, the vector $[\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]^T$ is an eigenvector with eigenvalue 1. This is therefore the stationary distribution. [2]

(b) The total image contains 128 bits. The probability of no errors is therefore $0.999^{128} = 0.8798$. [2]

(c) (i) Since $y_1 = x_1$, $H(Y_1) = 2$ bits

(ii) For $i > 1$, $H(Y_i) = H([0.97, 0.01, 0.01, 0.01]) = 0.2419$ bits [3]

(d) Verify the Kraft inequality $2^{-1} + 6\times 2^{-4} + 9\times 2^{-8} = 0.9102 \le 1$.

The probabilities of the (y_i, y_{i+1}) pairs are

$$\begin{pmatrix} 9409 & 97 & 97 & 97 \\ 97 & 1 & 1 & 1 \\ 97 & 1 & 1 & 1 \\ 97 & 1 & 1 & 1 \end{pmatrix} \times 10^{-4}$$

so we assign the short codewords to the highest probability pairs.

For the special case of (y_1, y_2), the probabilities are

$$\begin{pmatrix} 2425 & 25 & 25 & 25 \\ 2425 & 25 & 25 & 25 \\ 2425 & 25 & 25 & 25 \\ 2425 & 25 & 25 & 25 \end{pmatrix} \times 10^{-4}$$

The expected code lengths are (i) 3.3625 bits, (ii) 1.1809 bits [4]

(e) The probability of transmitting a pair without error is

$$\begin{pmatrix} 999 & 996 & 996 & 996 \\ 996 & 992 & 992 & 992 \\ 996 & 992 & 992 & 992 \\ 996 & 992 & 992 & 992 \end{pmatrix} \times 10^{-3}$$

[E4.40, SO20]

so the average probability of transmitting a pair without error is (i) 0.9966 for (y_1, y_2) and (ii) 0.9988 otherwise. Hence the probability of transmitting the entire image error free is $0.9966 \times 0.9988^{31} = 0.9608$. [4]

(f) If we transmit three times and take a majority decision, we will get the correct answer if the values are received GGG, GGB, GBG or BGG (where B=bad and G=good). If p is the probability of G, then the total probability of correct decoding is

$p^3 + 3p^2(1-p) = p^2(3-2p)$. For the probabilities above this gives (i) $1-3.374\times10^{-5}$ (ii) $1-4.178\times10^{-6}$

Thus the expected number of bits is $3(3.3625 + 31\times1.1809) = 119.91$ bits and the probability of decoding it without error is 0.9998. [5]