MT341/441/5441 Channels: Problem Sheet 1

Attempt at least questions 1 to 5. Please staple your answers together and remember to write your name or student number. You will get 1.25% of your final mark for a reasonable attempt at this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 11am, Thursday 11.30am, or by appointment.

To be handed in after the lecture on Thursday 10th October.

It is helpful if you indicate questions you did but are uncertain about, or would like seen done in lectures.

Probability reminder. Let Ω be a probability space. Recall that events $A, B \subseteq \Omega$ are *independent* if $\mathbf{P}[A \cap B] = \mathbf{P}[A]\mathbf{P}[B]$. If $\mathbf{P}[B] > 0$ we define the *conditional probability* of A given B by $\mathbf{P}[A|B] = \mathbf{P}[A \cap B]/\mathbf{P}[B]$. A real-valued *random variable* is a function $X : \Omega \to \mathbb{R}$. The *expectation* of X is defined by $\mathbf{E}[X] = \sum_x x\mathbf{P}[X = x]$. See the introduction notes and the probability notes on Moodle for more detail and examples. You could also look up Bayes' Theorem on the web.

1. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ with $p_k = \frac{1}{6}$ for each $k \in \{1, 2, 3, 4, 5, 6\}$ be the probability space for a single roll of a fair die. What is the probability

(a) a 5 is rolled?

(b) the roll is odd?

(c) two consecutive rolls are both 6?

(d) two consecutive rolls add up to 7?

Explain your answers in terms of the probability space, making clear the relevant event as a subset of Ω or Ω^2. [**Sorry, I forgot that the probability space is Ω^2 in (c) and (d).**] For instance, in (b) you should identify the event $\{1, 3, 5\}$ and find its probability. Are the events in (a) and (b) independent?

2. You draw two cards from a deck of 52 cards. As usual it has 4 suits each of 13 cards. The first card is not replaced before drawing the second. What is the probability of drawing two cards of the same suit?

3. A mathematics lecturer gave their class two tests: 48% of the class passed both tests, and 64% of the class passed the first test.

(a) What percentage of those who passed the first test also passed the second test?

(b) Write the result of (a) as a conditional probability $\mathbf{P}[A|B]$, making it clear what are the events A and B.

4. Your friend has a secret number in $\{0, 1, 2, 3, 4, 5, 6, 7\}$. Suppose it is 0 or 4 each with equal probability $\frac{1}{4}$, and each of the other numbers has probability $\frac{1}{12}$. Suggest a good questioning strategy. What is the expectation of the number of questions? What is the corresponding binary code? (It might help to draw the decision tree.)

5. Alice must communicate to Bob a 'Yes'/'No' message using the Binary Symmetric Channel with cross-over probability p. Alice's message is 'No' with probability $\frac{1}{100}$. She sends 000 for 'No' and 111 for 'Yes'. Bob decodes by assuming that the majority symbol in the received word is correct. Explain why $\mathbf{P}[\text{Bob receives } 010 \,|\, \text{Alice's message is 'No'}] = p(1-p)^2$ and find, in terms of p,

(a) $\mathbf{P}[\text{Bob decodes as 'No'} \,|\, \text{Alice's message is 'No'}]$

(b) $\mathbf{P}[\text{Bob receives } 011 \,|\, \text{Alice's message is 'Yes'}]$

(c) $\mathbf{P}[\text{Bob decodes as 'No'} \,|\, \text{Alice's message is 'Yes'}]$

(d) Hence find

$$\frac{\mathbf{P}[\text{Alice's message is 'No'} \,|\, \text{Bob decodes as 'No'}]}{\mathbf{P}[\text{Alice's message is 'Yes'} \,|\, \text{Bob decodes as 'No'}]}$$

in terms of p. [*Hint: using the definition of conditional probability one can express* $\mathbf{P}[A|B]$ *in terms of* $\mathbf{P}[B|A]$, $\mathbf{P}[A]$ *and* $\mathbf{P}[B]$, *for any events* A *and* B *with non-zero probability. You might have seen this as Bayes' Theorem.*]

(e) What is the ratio in (d) when $p = 1/10$? What do you conclude?

(f) Roughly how small does p need to be to make the probability that Bob always get Alice's intended message 99%?

6. (a) Let X be a discrete random variable taking non-negative values. Let $a > 0$. Show that $\mathbf{E}[X] \geq a\mathbf{P}[X \geq a]$ and deduce that $\mathbf{P}[X \geq a] \leq \mathbf{E}[X]/a$. (This is Markov's Inequality.)

(b) Let X be a discrete random variable. By applying (a) to a suitable random variable, show that if $a > 0$ then

$$\mathbf{P}\big[|X - \mathbf{E}X| \geq a\big] \leq \frac{\mathbf{Var}[X]}{a^2}.$$

(This is Chebyshev's Inequality.)

(c) Suppose that $F \sim \text{Bin}(2m+1, \frac{1}{10})$. Use Chebyshev's Inequality to bound $\mathbf{P}[F > m]$. Deduce that the probability of a decoding error in the BSC with cross-over probability $\frac{1}{10}$ can be made arbitrarily small by using a sufficiently long repetition code.

7. People may have disease D. In its early stages, D has no symptoms. A certain test satisfies $\mathbf{P}[\text{test positive}|\text{have } D] = 1 - p$ and $\mathbf{P}[\text{test positive}|\text{do not have } D] = p$. In words: the test works with probability $1 - p$.

Suppose that one in a thousand people have D.

(a) You have just taken the test. Which probability is more important to you: $\mathbf{P}[\text{test positive}|\text{have } D]$ or $\mathbf{P}[\text{have } D|\text{test positive}]$?

(b) Calculate $\mathbf{P}[\text{have } D|\text{test positive}]$.

(c) Comment on your answer to (b) in the cases $p = 0.002$ and $p = 0.0001$. When would you recommend the test be widely used?

MT341/441/5441 Channels: Problem Sheet 2

Attempt at least questions 1 to 5. Question 8 is nothing to do with coding theory, and is for interest only. Please staple your answers together and remember to write your name or student number. You will get 1.25% of your final mark for a reasonable attempt at this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 11am, Thursday 11.30am, or by appointment.

To be handed in after the lecture on Thursday 17th October.

It is helpful if you indicate questions you did but are uncertain about, or would like seen done in lectures.

The entropy $H(p)$ of a probability measure p on $\{1, \ldots, s\}$ is defined in Definition 3.6 to be $-\sum_{i=1}^{s} p_i \log_2 p_i$. You might prefer the formula $\sum_{i=1}^{s} p_i \log_2 \frac{1}{p_i}$ in some cases.

1. Please answer this question without using a calculator or computational assistance. Simplify your answers as much as possible.

(a) Find $\log_2 2^m$ for a general $m \in \mathbb{N}$.

(b) Find $\log_2 192 - \log_2 6$.

(c) Find $\log_{10} 64 / \log_{10} 2$.

(d) Find $\log_2 \sqrt{6} + \frac{1}{2} \log_2 12 - \frac{1}{2} \log_2 72$.

(e) Find $H(\frac{1}{2}, \frac{1}{2})$.

(f) Find $H(\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{32}, \frac{1}{32})$.

2. By working through the proof of the 'if' direction of Proposition 2.9, construct a prefix-free binary code C with codewords of lengths $2, 2, 3, 3, 4, 4, 4, 5, 5$. To show you understand the proof, make clear the number of forbidden prefixes at each step and draw the oriented rooted binary tree corresponding to C.

3. Let $(p_1, \ldots, p_s)$ be a probability measure on $\{1, \ldots, s\}$ and let $\ell_i = \lceil \log_2 \frac{1}{p_i} \rceil$ for each i, as in Proposition 3.8. A prefix-free binary code with codewords of lengths $\ell_1, \ldots, \ell_s$ is said to be a *Shannon code*.

(a) Write down (no detail required) Shannon codes for the probability measures:

(i) $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})$;

(ii) $(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$;

(iii) $(\frac{1}{2^m}, \ldots, \frac{1}{2^m})$ for a general $m \in \mathbb{N}$;

(iv) $(\frac{1}{3}, \frac{1}{3}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9})$.

(b) For each probability measure p in (a), compute the expected codeword length $\sum_{i=1}^{s} p_i \ell_i$ and the entropy $H(p)$, and verify that Proposition 3.8(ii) holds. Do the same for the code C in Question 2 when codewords are sent according to the probability measure in Question 1(f).

Binary form for fractions. Any real number p such that $0 \leq p < 1$ can be written in the form $p = \frac{1}{2}b_1 + \frac{1}{2^2}b_2 + \cdots + \frac{1}{2^\ell}b_\ell + \cdots$, where $b_1, b_2, \ldots, b_\ell, \ldots \in \{0, 1\}$. We say that $0.b_1b_2 \ldots b_\ell \ldots$ is a *binary form* of p and that b_ℓ is bit ℓ of p.

For example $\frac{1}{2} = 0.1$, $\frac{13}{16} = 0.1101$ and $\frac{1}{3} = 0.010101\ldots$ recurs. The form is unique if we agree to write 0.1 rather than $0.0111\ldots$ for $\frac{1}{2}$, and so on. The MATHEMATICA notebook `ShannonHuffman.nb` on Moodle has a function `FractionalBinaryForm` that you are welcome to use.

4. This question uses binary forms to construct Shannon codes. Let $(p_1, \ldots, p_s)$ be a probability measure on $\{1, \ldots, s\}$ with $p_1 \geq \ldots \geq p_s > 0$. Let $\ell_i = \lceil \log_2 \frac{1}{p_i} \rceil$.

For each $j \in \{1, \ldots, s\}$, let $r_j = \sum_{i=1}^{j-1} p_i$ and let $u(j)$ be the first $\ell(j)$ bits in the (unique) binary form of r_j. By convention the empty sum is 0, so $r_1 = 0$ and $u(1)$ is the all-zeros codeword of length ℓ_1.

For example, if $(p_1, p_2, p_3, p_4) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{4}, \frac{1}{12})$ then $(\ell_1, \ell_2, \ell_3, \ell_4) = (2, 2, 2, 4)$ and

$r_1 = 0.00000000\ldots, r_2 = 0.01010101\ldots, r_3 = 0.10101010\ldots, r_4 = 0.11101010\ldots$.

Hence $u(1) = 00, u(2) = 01, u(3) = 10, u(4) = 1110$.

(a) Perform this construction for the probability measures in Questions 3(i) and (iv). Do you get prefix-free codes?

(b) Show that if $0 \leq y - x < 1/2^k$ then the binary forms of x and y agree in their first k positions.

(c) Working in general, show from the definition of ℓ_i that $p_i \geq 2^{-\ell_i}$. Hence show that if $j > i$ then the binary forms of r_i and r_j differ in bit ℓ_i or earlier. [*Hint:* to see what's going on, try doing the special case $j = i + 1$ first.]

(d) Deduce that $\{u(1), \ldots, u(s)\}$ is a prefix-free code, and so is a Shannon code.

5. Use Gibbs' Inequality to show that if $(p_1, \ldots, p_s)$ is a probability measure on $\{1, \ldots, s\}$ then $H(p_1, \ldots, p_s) \leq \log_2 s$. When is equality attained?

6. Show that the expected codeword length of a Shannon code may be arbitrarily close to the upper bound $1 + H(p)$ in Proposition 3.8(ii).

7. Suppose that $q_1 + q_2 + q_3 = 1$. Use Gibbs' Inequality to show that the maximum possible value of $q_1q_2^2q_3^3$ is $1/432$.

8. The Schröder–Bernstein Theorem states that if X and Y are sets and there are injective functions $X \to Y$ and $Y \to X$ then there is a bijection $X \to Y$.

Show that there are bijections between any two of $\{x \in \mathbb{R} : 0 < x < 1\}$, $\mathbb{R}$, $\{(b_1, b_2, b_3, \ldots) : b_i \in \{0, 1\} \text{ for each } i\}$ and the set of all subsets of $\mathbb{N}$. [*Hint:* the binary form seen in Question 4 is useful.]

MT341/441/5441 Channels: Problem Sheet 3

Attempt at least questions 1 to 5. Please staple your answers together and remember to write your name or student number. You will get 1.25% of your final mark for a reasonable attempt at this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 11am, Thursday 11.30am, or by appointment.

To be handed in after the lecture on Thursday 31st October. Note you have a fortnight to do this sheet.

It is helpful if you indicate questions you did but are uncertain about, or would like seen done in lectures.

1. (a) Consider the three encoders for the alphabet $\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}$:

(i) $\mathtt{a} \mapsto 0, \mathtt{b} \mapsto 10, \mathtt{c} \mapsto 110, \mathtt{d} \mapsto 1110, \mathtt{e} \mapsto 1111$;

(ii) $\mathtt{a} \mapsto 1, \mathtt{b} \mapsto 00, \mathtt{c} \mapsto 010, \mathtt{d} \mapsto 0110, \mathtt{e} \mapsto 1110$;

(iii) $\mathtt{a} \mapsto 000, \mathtt{b} \mapsto 111, \mathtt{c} \mapsto 001, \mathtt{d} \mapsto 0001, \mathtt{e} \mapsto 1001$.

For each decide if the corresponding binary code is (1) prefix-free and (2) uniquely decipherable (see after Definition 2.4). Justify your answers.

(b) Give an example of a uniquely decipherable binary code C such that C is not prefix-free and neither is the binary code obtained from C by reversing all its codewords.

2. Let X and Y be two independent rolls of a fair die. Compute $H(X)$ and $H(X, Y)$. Show that if $Z = X + Y$ then $H(Z) < H(X, Y)$. What is $H(X, Z)$?

3. A memoryless source $U_1, U_2, \ldots$ produces symbols from the alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}\}$ so that $\mathbf{P}[U_t = \mathtt{a}] = \mathbf{P}[U_t = \mathtt{b}] = \mathbf{P}[U_t = \mathtt{c}] = \mathbf{P}[U_t = \mathtt{d}] = \mathbf{P}[U_t = \mathtt{e}] = \frac{1}{5}$ for all t.

(a) Find a prefix-free binary code C with five codewords $u(\mathtt{a})$, $u(\mathtt{b})$, $u(\mathtt{c})$, $u(\mathtt{d})$, $u(\mathtt{e})$ so that the expected length

$$\tfrac{1}{5}\big(\ell\big(u(\mathtt{a})\big) + \ell\big(u(\mathtt{b})\big) + \ell\big(u(\mathtt{c})\big) + \ell\big(u(\mathtt{d})\big) + \ell\big(u(\mathtt{e})\big)\big).$$

is a small as possible. [**Corrected typo: `a` was twice, and `e` missing.**]

(b) Since C is prefix-free, it can be used to unambiguously encode pairs of symbols. For instance the concatenation $u(\mathtt{a})u(\mathtt{b})$ encodes $\mathtt{ab}$. What is the expected length when C is used in this way? Compare it with $H(U_1, U_2)$.

(c) Let $m^{(r)}$ be the expected length when concatenations of codewords from C are used to encode r-tuples in $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}\}^r$. You found $m^{(1)}$ and $m^{(2)}$ in (a) and (b). Show that $m^{(r)}/r - H(U_1)$ is constant.

(d) Let $\ell^{(r)}$ be the expected length of the Shannon code on 5^r symbols when each symbol has equal probability $1/5^r$. Find a formula $\ell^{(r)}/r - H(U_1)$ and show that it tends to 0 as $r \to \infty$.

(e) Compare (d) and (e): for what r do we have $m^{(r)} \geq \ell^{(r)}$?

4. Let X and Y be random variables taking values in sets $\mathcal{X}$ and $\mathcal{Y}$. By definition [**Correction: minus sign was omitted**],

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mathbf{P}[X=x, Y=y]\log_2 \mathbf{P}[X=x, Y=y].$$

Show that if X and Y are independent then $H(X,Y) = H(X) + H(Y)$. Please take care to use $\sum$-notation correctly.

5. A source produces symbols from the alphabet $\{1,2,3,4,5,6,7\}$ with the probability distribution $p_1, p_2, p_3, p_4, p_5, p_6, p_7$ shown below.

i	1	2	3	4	5	6	7
p_i	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{3}$

(a) Find a Huffman code for p with a codeword of length 4.

(b) Find a Huffman code for p all of whose codewords have length at most 3.

(c) Compute the expected length of a codeword in each case. What do you notice? Check that the expected length is at least $H(p)$, as required by Theorem 3.10.

(d) Are the Huffman codes strictly better than the Shannon code for this probability distribution?

6. Use Gibbs' Inequality to show that if p is a probability measure on $\{1,\ldots,s\}$ then $H(p) \leq \log_2 s$. When does equality hold?

7. Define two operations on oriented rooted binary trees:

- (Prune) Pick the vertex adjacent to a leaf and delete its adjacent leaves.
- (Grow) Pick a leaf and add two children (new leaves) to it.

A *mutation* is a prune operation followed by a grow operation. Say that a prefix-free binary code C *saturates* Kraft's Inequality if $\sum_{u\in C} 2^{-\ell(u)} = 1$.

(a) Starting with the oriented rooted binary tree having 2^r leaves each distance r from the root, show that applying a sequence of mutations give the tree of a prefix-free binary code saturating Kraft's Inequality.

(b) Can every prefix-free binary code with 2^r codewords that saturates Kraft's Inequality be obtained in this way?

8. Let C be a binary code with s codewords. By comparing every pair of codewords, one can decide if C is prefix-free in $\binom{s}{2} \approx \frac{1}{2}s^2$ comparisons. Is there a significantly faster algorithm?

9. Write a computer program in the language of your choice to find a Huffman code for a given probability distribution and test it on Question 5.

MT341/441/5441 Channels: Problem Sheet 4

Attempt at least questions 1 to 4. If you are not doing MT361/461/5461 Cipher Systems, please also do Question 5. Please staple your answers together and remember to write your name or student number. You will get 1.25% of your final mark for a reasonable attempt at this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 11am, Thursday 11.30am, or by appointment.

To be handed in after the lecture on Thursday 7th December.

It is helpful if you indicate questions you did but are uncertain about, or would like seen done in lectures.

1. Suppose that an optimal (in the sense of Definition 5.6) code for the probability measure $p_1 \leq p_2 \leq \ldots \leq p_s$ has codewords $v(1), v(2), \ldots, v(s)$ of lengths $\ell_1, \ell_2, \ldots, \ell_s$. Prove Lemma 5.7(a) that $\ell_1 \geq \ell_2 \geq \ldots \geq \ell_s$.

2. Alice must guess Bob's secret number X by asking yes/no questions. She knows that X is distributed on $\{0, 1, \ldots, 2^r\}$ according to the probability measure $(\frac{1}{2}, \frac{1}{2^{r+1}}, \ldots, \frac{1}{2^{r+1}})$. Thus $\mathbf{P}[X = 0] = \frac{1}{2}$ and $\mathbf{P}[X = x] = \frac{1}{2^{r+1}}$ if $x \in \{1, \ldots, 2^r\}$.

(a) Find $H(X)$.

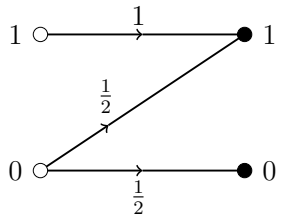
(b) Find, with proof, an optimal prefix-free code for this measure. [*Hint:* you could give a Huffman code, or use Corollary 3.10, or use Theorem 4.6(ii).]

(c) What is the corresponding questioning strategy for Alice?

(d) Let A be the answer to Alice's first question. Find $H(X|A =$ 'yes'$)$, $H(X|A =$ 'no'$)$ and $H(X|A)$. [*Hint:* conditional entropy was defined in Definition 7.4. Exercise 3.7(a) has a useful formula for the entropy when all probabilities are equal.]

(e) Comment on your answers in (d). Is it a surprise to you that the conditional entropy, given a particular answer by Bob, may be higher than $H(X)$?

3. In the binary channel shown below, when 0 is sent, it flips to 1 with probability $\frac{1}{2}$, and when 1 is sent, 1 is always received. [**Sorry, labels on arrows were wrong.**]



Suppose that $\mathbf{P}[X = 0] = q$ and $\mathbf{P}[X = 1] = 1 - q$.

(a) Write down the matrix of channel probabilities, as in Example 7.2.

(b) Show that $\mathbf{P}[Y = 0] = \frac{1}{2}q$ and $\mathbf{P}[Y = 1] = 1 - \frac{1}{2}q$. Hence write down a formula for $H(Y)$.

(c) (i) Find $\mathbf{P}[Y = 0|X = 0]$, $\mathbf{P}[Y = 1|X = 0]$ and hence find $H(Y|X = 0)$.
(ii) Find $\mathbf{P}[Y = 0|X = 1]$, $\mathbf{P}[Y = 1|X = 1]$ and hence find $H(Y|X = 1)$.
(iii) Find $H(Y|X)$ in terms of q.

(d) By Exercise 7.9, the mutual information $I(X;Y)$ is equal to $H(Y) - H(Y|X)$. By differentiating with respect to q, find the value of q that maximizes $I(X;Y)$.

(e) By Definition 7.11, the *capacity* of the channel is the maximum of $I(X;Y)$ as q varies. What is the capacity?

Gibbs' Inequality (Lemma 3.9) states that if p and q are probability measures on $\{1, \ldots, s\}$ then

$$-\sum_{i=1}^{s} p_i \log_2 p_i \leq -\sum_{i=1}^{s} p_i \log_2 q_i.$$

The proof given in lectures shows that equality holds if and only if $p = q$.

4. Let X and Y be random variables taking values in sets $\mathcal{A}$ and $\mathcal{B}$.

(a) For $\alpha \in \mathcal{A}$ and $\beta \in \mathcal{B}$, let $p_{\alpha\beta} = \mathbf{P}[X = \alpha, Y = \beta]$ and $q_{\alpha\beta} = \mathbf{P}[X = \alpha]\mathbf{P}[Y = \beta]$. Show that p and q are probability measures. [*Hint:* you just need to show that the values are non-negative and sum to 1.]

(b) Give an explicit example of random variables X and Y where $p_{\alpha\beta} \neq q_{\alpha\beta}$ for at least one pair (α, β). [*Hint:* you might use coin flips, or a die roll for X.]

(c) Use Gibbs' Inequality to prove that $H(X, Y) \leq H(X) + H(Y)$ with equality if and only if X and Y are independent.

(d) Check that (c) holds for your chosen example in (b).

(e) Deduce from (c) and the Chaining Rule that $H(X) \geq H(X|Y)$. When does equality hold?

5. Let X and Y be random variables taking values in sets $\mathcal{A}$ and $\mathcal{B}$, respectively.

(a) Show that $H(X|Y) = -\sum_{\beta\in\mathcal{B}} \sum_{\alpha\in\mathcal{A}} \mathbf{P}[X = \alpha, Y = \beta] \log_2 \mathbf{P}[X = \alpha|Y = \beta]$, where $\mathbf{P}[X = \alpha|Y = \beta]$ is interpreted as 0 if $\mathbf{P}[Y = \beta] = 0$.

(b) Hence show that

$$H(X|Y) = -\sum_{\beta\in\mathcal{B}} \sum_{\alpha\in\mathcal{A}} \mathbf{P}[X = \alpha, Y = \beta] \log_2 \mathbf{P}[X = \alpha, Y = \beta] + \sum_{\beta\in\mathcal{B}} \Big(\sum_{\alpha\in\mathcal{A}} \mathbf{P}[X = \alpha, Y = \beta] \Big) \log_2 \mathbf{P}[Y = \beta].$$

(c) Show that the second line above is $-H(Y)$. Deduce the Chaining Rule that $H(X|Y) + H(Y) = H(X, Y)$.

6. Verify that $H(X) - H(X|Y) = H(Y) - H(Y|X)$ in Question 3.

MT341/441/5441 Channels: Problem Sheet 5

Attempt at least questions 1 to 4. Please staple your answers together and remember to write your name or student number. You will get 1.25% of your final mark for a reasonable attempt at this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 11am, Thursday 11.30am, or by appointment.

To be handed in after the lecture on Thursday 14th November.

It is helpful if you indicate questions you did but are uncertain about, or would like seen done in lectures.

Throughout $X \in \mathcal{A}$ is the input symbol and $Y \in \mathcal{B}$ is the output symbol in a discrete memoryless channel with input alphabet $\mathcal{A}$ and output alphabet $\mathcal{B}$.

If you are confident, you might start with Question 6, and then apply it to find the capacities needed in the compulsory questions in a quicker way.

1. Consider the binary erasure channel from Example 7.2(1) with input alphabet $\{0,1\}$, output alphabet $\{0,\star,1\}$ and erasure probability p. The channel matrix is

$$\begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & \star & 1 \end{array} \\ \begin{pmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{pmatrix} \end{array}$$

(a) (i) Write down $\mathbf{P}[Y=0|X=0]$, $\mathbf{P}[Y=\star|X=0]$ and $\mathbf{P}[Y=1|X=0]$.
(ii) Use (i) to find $H(Y|X=0)$ in terms of p.
(iii) What is $H(Y|X=1)$?

Now suppose that $q = \mathbf{P}[X=0]$, so $1-q = \mathbf{P}[X=1]$.

(b) Working from Definition 7.4, find $H(Y|X)$.

(c) Find $\mathbf{P}[Y=0]$, $\mathbf{P}[Y=\star]$, $\mathbf{P}[Y=1]$ and hence $H(Y)$ in terms of p and q.

(d) Express $I(X;Y) = H(Y) - H(Y|X)$ in terms of p and q. Hence show that the capacity of the channel is $1-p$. For what value of q is the capacity attained?

[*Hint:* if you differentiate in (d), it may simplify things to multiply $I(X;Y)$ by $\log_e 2$ to replace each $\log_2$ with $\log_e$.]

Entropy Bound. By Question 6 on Sheet 3, the entropy of a probability measure on s symbols is at most $\log_2 s$, with equality if and only if all symbols are equally probable.

2. Consider the variation on the binary erasure channel in which $\mathcal{A} = \{0,1\}^5$, $\mathcal{B} = \{0,\star,1\}^5$ and when $x_1x_2x_3x_4x_5 \in \mathcal{A}$ is sent, one bit is chosen uniformly at random and replaced with $\star$. For example, if 00011 is sent then either $\star$0011, 0$\star$011, 00$\star$11, 000$\star$1 or 0001$\star$ is received, each with probability $\frac{1}{5}$.

Alice and Bob communicate using the binary parity check code

$$P = \left\{(x_1,x_2,x_3,x_4,x_5) \in \{0,1\}^5 : x_1+x_2+x_3+x_4+x_5 = 0 \bmod 2\right\}$$

(a) Show that $|P| = 16$. [*Hint:* you may write down the 16 codewords if you wish.]

(b) Suppose that Bob receives $101\star1$. What codeword in P must Alice have sent? Can Bob receives $101\star\star$ or 10100?

(c) (i) Show that $H(Y|X) = \log_2 5$.

(ii) How many elements of $\mathcal{B}$ may be received? Deduce from the boxed result that $H(Y) \leq 4 + \log_2 5$.

(iii) Hence show that the capacity of the channel is 4. What probability measure(s) on $\mathcal{A}$ attain the capacity?

(d) Specify an encoder $e : \{0, 1, \ldots, 15\} \to P$ and a decoder $d : \{0, \star, 1\}^5 \to \{0, 1, \ldots, 15\}$ such that whenever $e(x)$ is sent through the channel, and $v \in \{0, \star, 1\}^5$ is received, $d(v) = x$.

(e) Show that (a) in Shannon's Noisy Coding Theorem holds when $n = 1$.

3. Consider the lazy typist channel with $2t$ symbols. Show that the capacity of the channel is $\log_2 t$ and prove that (a) in Shannon's Noisy Coding Theorem holds for all $n \in \mathbb{N}$. [*Hint:* generalize the encoding and decoding rules in Example 7.3. When $t = 2$ the capacity was found in Example 7.11.]

Error correcting codes. A binary code $C \subseteq \{0, 1\}^n$ is said to be 1-*error correcting* if whenever $u \in C$, $v \in \{0, 1\}^n$ and $d(u, v) \leq 1$, nearest neighbour decoding of v gives u. For $u \in C$, let $B_1(u) = \big\{v \in \{0, 1\}^n : d(u, v) \leq 1\big\}$, where d is Hamming distance.

4. The binary code $C = \{00000, 11100, 00111, 11011\}$ is used to communicate over the Binary Symmetric Channel with error probability $p < \frac{1}{2}$.

(a) Decode 00111, 10100 and 10001 using nearest neighbour decoding.

(b) Find $\mathbf{P}[10100 \text{ received}|u \text{ sent}]$ for each codeword u and check that nearest neighbour decoding chooses u to maximize this probability.

(c) Write down all elements of $B_1(11100)$. How do these decode using nearest neighbour decoding?

(d) Show that C is 1-error correcting. Can C be extended to a 1-error correcting binary code of size 5?

5. Let $C \subseteq \{0, 1\}^n$ be a 1-error correcting binary code. Show that $B(u) \cap B(w) = \varnothing$ for all $u, w \in C$. Deduce that $|C| \leq 2^n/(1 + n)$.

6. In 1(b) and Question 4 you should have found that $H(Y|X)$ is a constant h, not depending on the probability measure on X.

(a) Show that this is the case whenever the rows of the channel matrix are all equal, up to the order of the entries.

(b) Suppose also that every column in the channel matrix has sum either 0 or 1. Show that then the channel capacity is $\log_2 b - h$, where b is the number of non-zero columns, attained for the uniform probability measure on $\mathcal{A}$.

For instance (a) and (b) both hold for the Binary Symmetric channel, the lazy typist channel, and the Question 2 channel; (a) holds for the Binary Erasure Channel.

MT341/441/5441 Channels: Problem Sheet 6

Attempt at least questions 1 to 5. Please staple your answers together and remember to write your name or student number. You will get 1.25% of your final mark for a reasonable attempt at this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 11am, Thursday 11.30am, or by appointment.

To be handed in after the lecture on Thursday 21st November.

It is helpful if you indicate questions you did but are uncertain about, or would like seen done in lectures.

Throughout $X \in \mathcal{A}$ is the input symbol and $Y \in \mathcal{B}$ is the output symbol in a discrete memoryless channel with input alphabet $\mathcal{A}$ and output alphabet $\mathcal{B}$.

1. Recall that the channel matrix has entries $p_{\alpha\beta} = \mathbf{P}[Y = \beta | X = \alpha]$ for $\alpha \in \mathcal{A}$ and $\beta \in \mathcal{B}$. What properties does the channel have if

(i) P has all an zero column;

(ii) all rows of P are equal;

(iii) in each column of P there is a unique non-zero value;

(iv) in each row of P there is a unique non-zero value;

(v) all columns of P have sum 1 and all rows are equal, up to rearrangement?

(You should consider each property (i), (ii), (iii), (iv), (v) separately. As well as probabilities, you might consider $H(Y|X)$ and the channel capacity.)

2. The Binary Erasure Channel with erasure probability p is used to send codewords from a code $C \subseteq \{0,1\}^n$ [**each with equal probability**]. A received word v is decoded by picking a codeword $u \in C$, *changing only the positions of* v *that are* $\star$.

(a) Suppose C is the binary code 0000000, 1101001, 1100110, 0001111, 1011010, 0110011, 0111100, 1010101. Decode the received words $01\star\star100$, $1\star\star101\star$, $\star\star00\star\star0$ [**corrected**] and $1\star\star\star\star01$. Can $00\star\star10\star\star$ be received?

(b) Prove that this decoder chooses u to maximize $\mathbf{P}[X = u | Y = v]$ and so implements maximum likelihood decoding.

(c) Now suppose C is $\left\{u \in \{0,1\}^4 : u_1 + u_2 + u_3 + u_4 \equiv 0 \bmod 2\right\}$. What is the probability of a decoding error when $0000 \in C$ is sent? Would your answer change for a different codeword?

(d) ($\star$) Show that the code in (a) is closed under addition in $\mathbb{F}_2^7$. Using this to show that $d(u,w) \geq 4$ for all codewords u and w. [*Hint:* use $d(u,w) = d(\mathbf{0}, u+w)$.] What is the least number of errors for which decoding may fail?

3. Suppose that a discrete memoryless channel has capacity c. For $r \in \mathbb{N}$, its r-*extension* is the channel with input alphabet $\mathcal{A}^r$ and output alphabet $\mathcal{B}^r$ defined by sending words of length r through the channel. Show that the capacity of the r-extension is rc.

Proof of Proposition 9.1. Let $0 < r < n/2$, let $p = r/n$ and let $h = H(p, 1-p)$. Proposition 9.1 states that

$$\frac{1}{n+1}2^{hn} \leq \binom{n}{r} \leq |B_{\mathbf{0}}(r)| \leq 2^{hn}$$

where $\mathbf{0}$ is the all zeros word in $\{0,1\}^n$. The lower bound was proved in lectures.

4. Fix $n \in \mathbb{N}$ and let $0 < p < \frac{1}{2}$. Let F be the number of flips in the Binary Symmetric Channel with error probability p when the all-zeros word $\mathbf{0} \in \{0,1\}^n$ is sent. Thus $F \sim \mathrm{Bin}(n,p)$.

(a) What is $\mathbf{P}[F = s]$ for $s \in \{0, 1, \ldots, n\}$?

(b) Prove that $p^s(1-p)^{n-s} \geq p^r(1-p)^{n-r}$ if $s \leq r$. Deduce that

$$1 \geq \sum_{s=0}^{r} \binom{n}{s} p^s(1-p)^{n-s} \geq p^r(1-p)^{n-r} \sum_{s=0}^{r} \binom{n}{s}.$$

(c) Dividing through by $p^r(1-p)^{n-r}$ we get $1/p^r(1-p)^{n-r} \geq \sum_{s=0}^{r} \binom{n}{s}$. Use this inequality to complete the proof of the proposition.

Toy Binary Symmetric Channel. Let $0 < p < \frac{1}{2}$ and $n \in \mathbb{N}$ be such that $pn \in \mathbb{N}$. In the Toy BSC(n,p), the input and output alphabets are $\{0,1\}^n$. When a binary word is sent through the channel *exactly* pn of its bits flip. So

$$\mathbf{P}[Y = v | X = u] = \begin{cases} \frac{1}{\binom{n}{pn}} & \text{if } d(u,v) = pn \\ 0 & \text{otherwise.} \end{cases}$$

5. Let $0 < p < \frac{1}{2}$ and $n \in \mathbb{N}$ be such that $pn \in \mathbb{N}$. Thus for each $r \in \mathbb{N}$, the Toy BSC(rn,p) is defined. Let c_r be its capacity.

(a) Prove that $c_r = nr - \log_2 \binom{nr}{pnr}$. [*Hint:* use the methods from Sheet 5.] [**Binomial coefficient corrected**.]

(b) Let $h = H(p, 1-p)$. Deduce from Proposition 9.1 that $c_r \geq n(1-h)$ and that $\lim_{r\to\infty} c_r/nr = 1 - h$, as stated in Lemma 9.4. [**Number corrected**].

(c) Compare the capacities of the Toy BSC and normal BSC. [*Hint:* Question 3 is relevant.]

6. Consider the Binary Symmetric Channel with error probability p.

(a) Let F_n be the number of flips when the channel is used to send binary words of length n. Use Chebychev's Inequality (see Question 6 on Problem Sheet 1) to show that for any $\varepsilon > 0$, $\mathbf{P}\big[|F_n - pn| > \varepsilon n\big] \to 0$ as $n \to \infty$.

(b) Prove Shannon's Noisy Coding Theorem for this channel. [*Hint:* use the functions $g_i(v)$ in the proof for the Toy version; the expectations in the next step can be bounded using (a). Or see §3.5 of D. Welsh, *Codes and cryptography*, Oxford University Press (1988), 001.5436 WEL for a clear presentation.]

MT341/441/5441 Channels: Problem Sheet 7

Attempt at least questions 1 to 3. Please staple your answers together and remember to write your name or student number. You will get 1.25% of your final mark for a reasonable attempt at this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 11am, Thursday 11.30am, or by appointment.

To be handed in after the lecture on Monday 2nd December. Note you have an extra weekend to do this sheet.

It is helpful if you indicate questions you did but are uncertain about, or would like seen done in lectures.

1. In Step 2 of the proof of the constructive direction of Shannon's Noisy Coding Theorem for the Toy BSC, we chose codewords $U(1), \ldots, U(M)$ uniformly at random from $\{0,1\}^n$. Recall that $X \in \{0,1\}^n$ is the sent codeword and $Y \in \{0,1\}^n$ is the received word.

(a) Let $u, v \in \mathbb{F}_2^n$. Show that $\mathbf{P}[Y = v | X = u] = 1/\binom{n}{pn}$ if $d(u,v) = pn$ and find $\mathbf{P}[Y = v | X = u]$ when $d(u,v) \neq pn$. [*Hint:* refer to the definition of the Toy BSC in Definition 9.3.]

(b) Since $U(i)$ is uniformly distributed,

$$\mathbf{E}_{\text{code}}\big[\mathbf{P}[Y = v | X = U(i)]\big] = \frac{1}{2^n} \sum_{u \in \{0,1\}^n} \mathbf{P}[Y = v | X = u].$$

Deduce from (a) that the left-hand side is $1/2^n$, as was claimed in the proof.

2. Fix $n \in \mathbb{N}$ and $0 < p < 1$ such that $pn \in \mathbb{N}$. The *Toy Binary Erasure Channel* has input alphabet $\{0,1\}^n$ and output alphabet $\{0, \star, 1\}^n$. When a word $u \in \{0,1\}^n$ is sent, pn symbols are chosen uniformly at random, and replaced with $\star$.

For example, if $p = \frac{1}{2}$ and $n = 4$ and 0000 is sent then the received word is one of $\star\star00, \star0\star0, \star00\star, 0\star\star0, 0\star0\star, 00\star\star$ each with equal probability $\frac{1}{6}$. The channel in Question 2 on Sheet 5 is the case $p = \frac{1}{5}$ and $n = 5$.

Let X be the sent word and Y the received word.

(a) Show that precisely $\binom{n}{pn} 2^{(1-p)n}$ words $v \in \{0, \star, 1\}^n$ may be received. Deduce that, for any probability measure on X, we have $H(Y) \leq n(1-p) + \log_2 \binom{n}{pn}$.

(b) Find $H(Y | X = u)$ for each $u \in \{0,1\}^n$ and hence show that $H(Y|X) = \log_2 \binom{n}{pn}$.

(c) Deduce that $I(X;Y) \leq n(1-p)$ and give a case when equality holds.

(d) What is the capacity of the channel?

3. Take the channel from Question 2. Let $M = 2\lceil 2^{rn} \rceil$ where $r < 1 - p$ and let

$$U(1), \ldots, U(M) \in \{0,1\}^n$$

be codewords chosen uniformly and independently at random. To decode a received word v, the receiver finds all the codewords $U(i)$ such that $U(i)$ is equal to v except in the precisely pn positions where v has a $\star$, and then picks one arbitrarily. Let $D(v)$ be the set of such codewords.

For example, if $p = \frac{1}{2}$ and $n = 4$ and the randomly chosen code is 0000, 0011, 1100, 1110, then $D(1\star1\star) = \{1110\}$ and $D(00\star\star) = \{0000, 0011\}$.

Let P_i be the probability, computed using this channel, that when $U(i)$ is sent, the received word v is not decoded as $U(i)$.

(a) (Step 1.) If precisely pn positions of v are $\star$, let $g_i(v)$ be the number of codewords $U(j)$ with $j \neq i$ such that $U(j) \in D(v)$. Otherwise let $g_i(v) = 0$. Show that $P_i \leq \sum_{v \in \{0,\star,1\}^n} \mathbf{P}_{\mathrm{ch}}[Y = v | X = U(i)] g_i(v)$

(b) (Step 2.) Hence compute **an upper bound for** [**correction**] $\mathbf{E}_{\mathrm{code}}[P_i]$.

(c) (Step 3 and Step 4.) Deduce that for every $\varepsilon > 0$, provided n is sufficiently large, there is a code with $\lceil 2^{rn} \rceil$ codewords such that the decoding error probability is $< \varepsilon$ for every codeword.

4. Fix a memoryless channel with input alphabet $\mathcal{A}$, output alphabet $\mathcal{B}$ and capacity c. Suppose that the capacity of the channel is attained for the probability measure q on $\mathcal{A}$ assigning probability q_α to each $\alpha \in \mathcal{A}$.

Its *n-extension* is the channel with input alphabet $\mathcal{A}^n$ and output alphabet $\mathcal{B}^n$ defined by sending n symbols through the channel, one after the other.

Let $X \in \mathcal{A}^n$ be the sent word and let $Y \in \mathcal{B}^n$ be the received word.

(a) Let $\mathbf{P}[X = \alpha_1 \ldots \alpha_n] = q_{\alpha_1} \ldots q_{\alpha_n}$ for each $\alpha_1 \ldots \alpha_n \in \mathcal{A}^n$. Show that, with this distribution on X, we have $I(X;Y) = nc$.

(b) ($\star$) Show that for any probability distribution on X we have $I(X;Y) \leq nc$. [*Hint:* this needs some care, because the probability distribution on X need not be of the special form in (a).]

(c) Deduce from (a) and (b) that the capacity of the n-extension is nc.

5. Let X, Y and Z be random variables taking values in sets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, respectively.

(a) Lemma 10.1 states that $H\big(X|(Y,Z)\big) \leq H\big(X|Z\big)$. Show that equality holds if and only if

$$\mathbf{P}[X = x, Y = y | Z = z] = \mathbf{P}[X = x | Z = z] \mathbf{P}[Y = y | Z = z]$$

for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$. (This states the conditional independence of X and Y given Z; it is equivalent to $X \mapsto Z \mapsto Y$ being a Markov chain.)

(b) The Data Processing Inequality states that if $d : \mathcal{Y} \to \mathcal{Z}$ is a function then $I(X;Y) \geq I\big(X; d(Y)\big)$. Use (a) to show that equality holds **for all probability distributions on X, Y, Z** [**correction**] if and only if d is injective.

MT341/441/5441 Channels: Problem Sheet 8

Attempt at least questions 1 to 3. Please staple your answers together and remember to write your name or student number. You will get 1.25% of your final mark for a reasonable attempt at this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 11am, Thursday 11.30am, or by appointment.

To be handed in after the lecture on Monday 9th December or to McCrea 0-25 by noon on Wednesday 11th December.

It is helpful if you indicate questions you did but are uncertain about, or would like seen done in lectures.

1. The memoryless source in Exercise 1.7 source emits symbols `a`, `t`, `g`, `c` with probabilities $\frac{1}{8}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}$, respectively. Let $S_1 \dots S_r$ be a random word of length r.

(a) Find all the pairs of symbols $s_1 s_2$ such that $\mathbf{P}[S_1 S_2 = s_1 s_2] = \frac{1}{16}$. Hence show that $\mathbf{P}[\log_2 \mathbf{P}[S_1 S_2] = -4] = \frac{5}{16}$.

(b) Calculate $\mathbf{P}[\log_2 \mathbf{P}[S_1 S_2 S_3] = -7]$.

(c) Let $N(S_1 \dots S_r)$ be the number of the symbols in $S_1 \dots S_r$ equal to `a`. Use the Weak Law of Large Numbers to show that, for any $\varepsilon > 0$,

$$\mathbf{P}[\tfrac{1}{8} - \varepsilon \leq \frac{N(S_1 \dots S_r)}{r} \leq \tfrac{1}{8} + \varepsilon] \to 1 \quad \text{as } r \to \infty.$$

(d) What is the conclusion of the AEP for memoryless sources (Lemma 11.5) for this source?

(e) Show that given $\varepsilon > 0$, provided r is sufficiently large, there is a 'very typical' set $\mathcal{V} \subseteq \{\mathtt{a}, \mathtt{t}, \mathtt{g}, \mathtt{c}\}^r$ such that

- $\mathbf{P}[S_1 \dots S_r \in \mathcal{V}] > 1 - \varepsilon$;
- if $s_1 \dots s_r \in \mathcal{V}$ then between $(\frac{1}{8} - \varepsilon)r$ and $(\frac{1}{8} + \varepsilon)$ of the symbols $s_1, \dots, s_r$ are equal to `a`;
- if $s_1 \dots s_r \in \mathcal{V}$ then $2^{-r(\frac{7}{4}+\varepsilon)} \leq \mathbf{P}[S_1 \dots S_r = s_1 \dots s_r] \leq 2^{-r(\frac{7}{4}-\varepsilon)}$.

2. A memoryless source emits symbols $S_1, S_2, \dots$ in an alphabet $\mathcal{A}$. Let $h = H(S_1)$. The constructive part of Shannon's Source Coding Theorem (Theorem 4.7(i)) says that for every $\varepsilon > 0$ there exists $r \in \mathbb{N}$, a prefix-free binary code $C^{(r)}$ and an injective encoder $f^{(r)} : \mathcal{A}^r \to C^{(r)}$ such that

$$\frac{\overline{f}^{(r)}}{r} < h + \varepsilon$$

where $\overline{f}^{(r)}$ is the expected codeword length.

Prove this using the AEP for Memoryless Sources (Lemma 11.5). [*Hint:* generalize Example 12.2.]

3. Many questions about practical channels can be answered using Shannon's Noiseless Coding Theorem (see Theorems 4.6 and 4.7), Shannon's Noisy Coding Theorem (Theorem 7.14) and Proposition 12.5 on lossy source coding.

(a) A Binary Symmetric Channel with error probability 0.05 can transmit 800 bits per second. (This is the raw data rate, with no error correction.) How many bits can it transmit with negligible error probability per second?

(b) A channel with a physical capacity of transmitting 800 bits per second can tramsit at most 500 bits per second with a negligible error probability. What can you deduce about the capacity of the channel?

(c) A memoryless source emits the bits 0 and 1 with equal probability $\frac{1}{2}$ at 600 bits per second. If the symbols must be transmitted through the channel in (b), what, approximately is the minimum probability that a received bit is wrong?

4. Let P be the channel matrix for a noisy channel with input alphabet and output alphabet $\{0, 1, \ldots, s-1\}$, so, writing X for the sent symbol and Y for the received symbol as usual, $P_{ij} = \mathbf{P}[Y = j | X = i]$ for each $i, j \in \{0, 1, \ldots, s-1\}$. Let P^T be the transpose of P.

Suppose that $(PP^T)_{ij} = 0$ for all $i, j \in U \subseteq \{0, 1, \ldots, s-1\}$.

(a) Suppose that only symbols in U are sent. Show that there is a decoder with zero error probability.

(b) Deduce that the channel has capacity has at least $\log_2 |U|$.

(c) Use (b) to obtain the capacity of the lazy typist channel on $2t$ symbols.

(d) Can the capacity of the BSC be achieved by a decoder with zero error probability?

MT341/441/5441 Channels: Problem Sheet 9

Attempt at least questions 1 to 5. If you wish, you may hand in this sheet at the start of next term to McCrea 0-25. Answers will be posted on Moodle on Thursday 19th December.

You are welcome to welcome to email the lecturer, `mark.wildon@rhul.ac.uk`, with any questions over the vacation.

The miniproject for M.Sc. students will be posted to Moodle on Saturday 14th December.

1. Let $S_1, S_2, \ldots$ be a memoryless source taking values in an alphabet $\mathcal{A}$. Thus, by Definition 3.1, the S_t are independent and identically distributed. Show, as claimed in Lemma 13.3, that $H(S_1, \ldots, S_r) = rH(S_1)$. Deduce that the entropy of the source is $H(S_1)$. What are the minimum and maximum possible entropies?

2. The binary source in Example 13.2(3) begins by flipping a coin, biased to land heads with probability p, where $0 < p < 1$. Let T be the toss. If $T =$ heads, then $S_1 = S_2 = \ldots = 1$. If $T =$ tails, then the source now behaves as the memoryless source in Example 13.2(1), so $\mathbf{P}[S_t = 0] = 1 - p$ and $\mathbf{P}[S_t = 1] = p$ for all $t \in \mathbb{N}$.

(a) What is $H(T)$? What is $H(T|S_1S_2S_3 = 111)$?

(b) Find $H(S_1 \ldots S_r|T = \text{heads})$ and $H(S_1 \ldots S_r|T = \text{tails})$. Hence find $H(S_1 \ldots S_r|T)$.

(c) Using the Chaining Rule (Lemma 7.6), find $H(S_1, \ldots, S_r, T)$.

(d) Find $H(T|S_1 \ldots S_r)$. [*Hint:* there is a relevant special case in (a). What can you deduce about T if $S_t = 0$ for some t?]

(e) Hence find $H(S_1 \ldots S_r)$.

(f) Deduce that the entropy of this source is $(1 - p)H(p, 1 - p)$.

3. Let $S_1, S_2, \ldots$ be the source in Question 2.

(a) Is the source memoryless?

(b) Is the source stationary?

(c) Is the source ergodic?

(d) What is the most probable message of length r emitted by the source? Does the source satisfy the AEP?

4. Let $x^{(m)}$ be the binary word obtained by concatenating $01 \in \{0,1\}^2$ with $0011 \in \{0,1\}^4$, and so on, ending with $0\ldots01\ldots1 \in \{0,1\}^{2m}$. For example,

$$x^{(3)} = 010011000111.$$

(a) Perform Lempel–Ziv encoding on $x^{(5)}$. Give the encoded word *and* the final dictionary.

(b) Let $y^{(m)}$ be the Lempel–Ziv encoding of $x^{(m)}$. Find $y^{(m)}$ in general.

(c) Recall that $\ell(x)$ is the length of the word x. What is $\lim_{m\to\infty} \dfrac{\ell(y^{(m)})}{\ell(x^{(m)})}$?

5. In Lempel–Ziv coding, $\lceil \log_2 s \rceil$ bits are used to write the dictionary value at Step s.

(a) Show this sequence $\lceil \log_2 s \rceil$ for $s \in \mathbb{N}$ begins $0, 1, 2, 2, 3, 3, 3$ and find the next five terms.

(b) The word

$$y = 00110010101100011110011011000000010.$$

is the Lempel–Ziv encoding of a binary word x. Splitting the word up as indicated by (a) and Algorithm 13.10, it becomes

$$y = 0, 01, 100, 101, 0110, 0011, 1100, 1101, 10000, 0010$$

where red numbers are the binary form of dictionary values. Find the Lempel–Ziv dictionary and hence find x.

The functions `LempelZiv` and `LempelZivDecode` in the MATHEMATICA notebook `LempelZiv.nb` available from Moodle can be used to check your answer.

6. A memoryless binary source emits 0 with probability $1-p$ and 1 with probability p. If the number of typical words of length r (in the sense of the AEP) is about $(\frac{4}{3})^r$ for large r, [**Corrected typo: $\frac{3}{4}$ should be $\frac{4}{3}$**] what is a good estimate for p?

7. Define a binary source $S_1, S_2, \ldots$ so that if $\lfloor \log_2 t \rfloor$ is even then S_t is uniformly distributed on $\{0,1\}$, and if $\lfloor \log_2 t \rfloor$ is odd then $S_t = 0$.

Thus S_1, S_4, S_5, S_6, S_7, $S_{16}, \ldots, S_{31}$ and so on, are uniformly distributed and all other bits are 0.

(a) Show that $H(S_1, \ldots, S_r)/r$ is $\frac{r}{3}$ for infinitely many r, and more than $\frac{r}{2}$ for infinitely many r.

(b) Deduce that $\lim_{r\to\infty} H(S_1, \ldots, S_r)/r$ does not exist, and so the source does not have an entropy.

8. Fix a memoryless source with alphabet $\mathcal{A}$. Let $\mathbf{P}[S_t = \alpha] = p_\alpha$ for each $\alpha \in \mathcal{A}$. Suppose that $p_\alpha > 0$ for each α. [**Added this assumption to remove trivial cases.**] Let $\varepsilon > 0$ be given and let $\mathcal{T} \subseteq \mathcal{A}^r$ be a set of typical words of length r, in the sense of the AEP. Show, by taking ε small, that the most probable word of length r is in $\mathcal{T}$ if and only if p is the uniform distribution.