PRIFYSGOL ABERYSTWYTH UNIVERSITY

ADRAN MATHEMATEG / DEPARTMENT OF MATHEMATICS

ARHOLIADAU SEMESTER 1 / SEMESTER 1 EXAMINATIONS

IONAWR / JANUARY 2026

MA35810 – Information Theory

The questions on this paper are written in English.

Amser a ganiateir - 2 awr **Time allowed - 2 hours**

- Gellir rhoi cynnig ar bob cwestiwn.
- Rhoddir mwy o ystyriaeth i berfformiad yn rhan B wrth bennu marc dosbarth cyntaf.
- Cyfrifianellau Casio FX-83 neu FX-85 YN UNIG a ganiateir.
- Mae modd i fyfyrwyr gyflwyno atebion i'r papur hwn naill ai yn y Gymraeg neu'r Saesneg.

- All questions may be attempted.
- Performance in section B will be given greater consideration in assigning a first class mark.
- Casio FX-83 or FX-85 calculators ONLY may be used.
- Students may submit answers to this paper in either Welsh or English.

© 2026 Aberystwyth University

Section A

1. (a) The random variable X can take values $x = 1, 2, \ldots, 8$ with the probabilities given by the following table

x	1	2	3	4	5	6	7	8
$p(x)$	$1/2$	$1/8$	$1/8$	$1/8$	$1/32$	$1/32$	$1/32$	$1/32$

What is the value of the entropy $H(X)$? [4 marks]

(b) What is the maximal entropy for a random variable with 8 values? Which distribution achieves this? (No proof required.) [3 marks]

2. Let X and Y be two random variables which both take values from $\{1, 2\}$. The joint distribution $p(x, y)$ is given by $p(1, 1) = 0,\ p(2, 1) = 3/4,\ p(1, 2) = p(2, 2) = 1/8$.

(a) Compute

- the joint entropy $H(X, Y)$,
- the entropies $H(X)$ and $H(Y)$,
- the mutual information $I(X; Y)$,
- the conditional entropies $H(Y|X)$ and $H(X|Y)$. [10 marks]

(b) Draw the entropy Venn diagram for these quantities. [2 marks]

(c) Show that the uncertainty in X is decreased if $Y = 1$ is observed but increased if $Y = 2$ is observed (by computing the corresponding conditional entropies). How does the uncertainty of X change by observation of Y on average? [4 marks]

3. (a) Use a Venn diagram to build a codeword in the Hamming-$(7, 4, 3)$-code for the bitstring 1010. [3 marks]

(b) Suppose a noisy channel flips the third position (of this bitstring) but transmits the other bits correctly. Describe with the Venn diagram how the receiver detects and corrects this error. [4 marks]

4. Suppose you throw a biased coin with $p(\text{head}) = 0.9$ and $p(\text{tail}) = 0.1$ for $n = 10$ times.

(a) Regarding the sequences with at most one occurrence of 'tail' as typical, what is the expected length of the compressed message, using the corresponding naive compression scheme? What is the compression ratio? [10 marks]

(b) What is the theoretical optimum for the possible compression ratios of such a random variable? [4 marks]

5. (a) Find a prefix code with codeword lengths $(1, 2, 3, 4, 5, 5)$. [3 marks]

(b) Find the Huffman code for the probability distribution $(0.4, 0.3, 0.25, 0.05)$. [5 marks]

(c) Compute the expected codeword length for the Huffman code and the entropy in (b). Which inequalities between these are predicted by theory? Verify them in this example. [4 marks]

6. We consider noisy channels $(\mathcal{X}, p(y|x), \mathcal{Y})$ with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and transition matrix $\mathcal{P} = (p(y|x))_{x\in\mathcal{X}, y\in\mathcal{Y}}$.

(a) State the definition of the (informational) capacity of a noisy channel. [2 marks]

(b) What is the condition for a channel to be symmetric? Give a formula for the (informational) capacity in this case (no proof required). [4 marks]

(c) Find the capacity of the following channel.

$\mathcal{X} = \{0, 1, 2\} = \mathcal{Y}$ and $\mathcal{P} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}$ [4 marks]

(d) Is it possible with the channel in (c), by using good codes, to communicate reliably (with errors below any given bound) and with a rate of $1/2$? Give a reason by quoting Shannon's noisy coding theorem. [4 marks]

Section B

7. Let X_1 and X_2 be discrete random variables drawn according to probability distributions p_1 and p_2 with values in alphabets $\mathcal{X}_1 = \{1, \ldots, m\}$ and $\mathcal{X}_2 = \{m+1, \ldots, n\}$. We form a random variable X with values in $\mathcal{X} = \{1, \ldots, n\}$ by doing X_1 with probability α and X_2 with probability $1 - \alpha$.

(a) Find the entropy $H(X)$ in terms of $H(X_1)$, $H(X_2)$ and α. [8 marks]

(b) If $H(X_1) = H(X_2)$, for which α is $H(X)$ maximal? [2 marks]

8. Prove that the relative entropy of two probability distributions on the same finite set is always non-negative. You may use the inequality $\log(t) \leq C(t-1)$ for all $t > 0$ (with a constant $C > 0$) without proving it. [8 marks]

9. Let $X_1, X_2, \ldots$ be a sequence of independent tosses of a biased coin with $p(head) = q > 1/2$. Let H_n be the set of all bitstrings $x_1 \ldots x_n$ produced in this way such that $p(x_1 \ldots x_n) > 2^{-n}$.

(a) Compute $p(H_n)$ for $n = 1, 2, 3$ if $q = 2/3$. [3 marks]

(b) Prove with the law of large numbers that for n large enough the probability of H_n is bigger than 0.99.
Hint: This is a special case of our arguments about the asymptotic equipartition property. Recall these arguments to give a direct proof for the statement above.

[7 marks]

10. We consider binary codes C for the source $\mathcal{X} = \{1, 2, 3, 4\}$, so we have codewords $C(1)$, $C(2)$, $C(3)$, $C(4)$.

(a) State what it means for such a code to be uniquely decodable. [5 marks]

(b) We put $C(1) = 10$, $C(2) = 00$, $C(3) = 11$, $C(4) = 110$. Show that this code is uniquely decodable. [5 marks]

11. Let X and Z be independent $\{0, 1\}$-valued random variables and let α be the probability for $Z = 1$. We consider the channel $Y = XZ$ (from X to Y by multiplying X with Z, so Y is also $\{0, 1\}$-valued). Compute the optimising distribution for X and the capacity of this channel (depending on α).
For the optimisation you can use the fact that the derivative of the binary entropy function $H(p)$ is $\log \frac{1-p}{p}$. [12 marks]