There are four sections (A,B,C,D) worth respectively 20%, 20%, 30% and 30% of the overall score. Individual questions have the marks indicated.

Answer all questions. You can leave numerical results as fractions. When explanations are requested, the quality of the explanation will be taken into account in marking.
SECTION A.

1. In the year 2038 learning enhancement drugs are widely available, but not permitted for use at ANU. It is known that nevertheless 2% of all students use these drugs when sitting exams. However there is a simple breath test that can be easily administered to students as the enter the exam room. The breath test detects the presence of the drug 95% of the time when it is actually present. The test falsely claims presence of the drug when it is in fact absent 1% of the time.

   Albert, an ANU student, has just walked into his information theory exam, is breath tested, and the test says he has the drug in his system.

   (a) What is the prior probability that Albert took the drug? **(1 Marks)**

   (b) What is the conditional probability that the test reports positive given that Albert took the drug? **(1 Marks)**

   (c) What is the conditional probability that the outcome of the test is positive given that Albert did not take the drug? **(1 Marks)**

   (d) What is the posterior probability that Albert took the drug given the test was positive? **(6 Marks)**

2. Suppose X and Y are binary random variables (taking values in $\{0, 1\}$)

   (a) Suppose $P(X = 1|Y = 1) = P(X = 1)P(Y = 1)$. Are X and Y statistically independent? Explain your answer. **(2 Marks)**

   (b) Suppose $P(X = 1|Y = 1) < 0.2$ and $P(X = 1|Y = 0) < 0.6$. What can one say about $P(X = 0)$? **(4 Marks)**

   (c) Supposing none of the probabilities are zero, is it always true that

   $$\frac{P(X = 0, Y = 0)}{P(X = 0, Y = 1)} = \frac{P(Y = 0|X = 0)}{P(Y = 1|X = 0)}?$$

   Show your reasoning. **(5 Marks)**

SECTION B.

3. (a) Consider a random variable X with uniform distribution on $\{1, 2, \ldots, n\}$. Suppose $Y = 2X$ and Z is a random variable with uniform distrubution on $\{1, 2, \ldots, 2n\}$. Compute the entropies of X, Y and Z and explain the relationship between $H(Y)$ and $H(Z)$. **(2 Marks)**

   (b) Consider a random variable X on $\{1, 2, 3, 4\}$ with distribution (p_1, p_2, p_3, p_4). Give an expression for the entropy of X solely in terms of the binary entropy function H_2. **(3 Marks)**

   (c) Consider two random variables X and Y on a common set S with distributions p and q respectively. Answer the following questions, providing a justification for your answer:

      i. Suppose $c > 0$. If $D_{\text{KL}}(p, q) = c$ does $D_{\text{KL}}(q, p) = c$? **(1 Marks)**

      ii. If $D_{\text{KL}}(p, q) = 0$, does $D_{\text{KL}}(q, p) = 0$? **(1 Marks)**

      iii. If $D_{\text{KL}}(p, q) = 0$ does $\Pr(X - Y \neq 0) = 1$? **(1 Marks)**

      iv. If $D_{\text{KL}}(p, q) = 0$ is $X \perp Y$? **(1 Marks)**

      v. If $D_{\text{KL}}(p, q) = 0$ does $I(X; Y) = 0$? **(1 Marks)**

      vi. if $D_{\text{KL}}(p, q) = 0$ does $p = q$? **(1 Marks)**

      vii. If $I(X; Y) = H(X)$ does $Y = g(X)$ for some function $g : S \to S$? **(1 Marks)**

   (d) Suppose X is a positive random variable and we know that $\log(\mathbb{E}(X)) = 1$. What can we say about the value of $\mathbb{E}(\log(X))$? **(2 Marks)**

   (e) Jane's house is connected to the electricity grid, but she has no solar panels. Her average daily electricity consumption is 10kWh. What can be said about the probability that on a given day her electricity consumption exceeds 20kWh? **(3 Marks)**

   (f) Jane measures the standard deviation of her consumption and finds it is $\sqrt{10}$ kWh. With this additional information, what can one say about the probability that on a given day her electricity consumption exceeds 20kWh? **(3 Marks)**

SECTION C.

4. Consider two binary random variables with joint distribution as described below

$p(X,Y)$		X = 1	X = 2
Y	1	0	3/4
	2	1/8	1/8

   (a) Compute $p(X)$, $p(Y)$, $p(X|Y=1)$, $p(X|Y=2)$. **(1 Marks)**

   (b) Show that $H(X|Y=2) > H(X)$ **(1 Marks)**

   (c) Does the result above contradict $H(X|Y) \leq H(X)$? Explain your answer. **(1 Marks)**

5. Let X be an ensemble with alphabet $\{a,b\}$ and probabilities $(\frac{1}{3}, \frac{2}{3})$.

   (a) What is the alphabet of the extended ensemble X^3? **(1 Marks)**

   (b) Give an example of three sequences in the typical set (for $N=3$, $\beta=0.2$). **(1 Marks)**

   (c) What is the smallest δ-sufficient subset of X^3 when $\delta = 1/25$ and when $\delta = 1/10$? **(1 Marks)**

   (d) Suppose $N = 1000$, what fraction of the sequences in X^N are in the typical set (at $\beta = 0.2$)? **(2 Marks)**

   (e) If $N = 1000$, and a sequence in X^N is drawn at random, what is the (approximate) probability that it is in the N,β-typical set? **(2 Marks)**

6. Consider the ensemble X with alphabet $\{a,b,c,d\}$ with probabilities $(1/2, 1/4, 1/8, 1/8)$. Consider the code $C = \{0000, 01, 11, 0\}$ for this ensemble.

   (a) What is the expected length of the code C? **(1 Marks)**

   (b) Is C prefix free? Explain. **(1 Marks)**

   (c) Is C uniquely decodable? Explain. **(1 Marks)**

   (d) Must there exist a code C' that is prefix free and uniquely decodeable and the expected lengths satisfy

$$L(C', X) < L(C, X)?$$

   Explain. **(1 Marks)**

   (e) Can there exist a prefix code C'' with code lengths 1,2,2,2 for this ensemble? Explain. **(1 Marks)**

   (f) Construct a Shannon code for the ensemble X. **(2 Marks)**

   (g) Construct a Huffman code for the ensemble X. **(2 Marks)**

   (h) Suppose that instead the distribution associated with X was (0.527,0.214,0.135, 0.124). Without calculating anything, explain what one would need to do to construct a code C^* such that

$$L(C^*, X) \leq H(X) + 1$$

   (2 Marks)

7. Consider an ensemble $\{a, b, c, d\}$ with distribution (2/9,1/9,1/3,1/3).

(a) Amelia claims to have calculated an SFE code and obtained $\{00, 01, 10, 11\}$. Why must she be mistaken? **(1 Marks)**

(b) Bruce claims that he calculated an SFE code and obtained $\{001, 010, 100, 110\}$. Is this plausible? Explain. **(1 Marks)**

(c) Christie claims that when she computed an SFE code she obtain $\{0001, 0100, 100, 110\}$. Is this plausible? Explain. **(1 Marks)**

(d) David has built a source coder for letters of the English alphabet (i.e. it maps $\{A, B, \ldots, Z, \sqcup\}$ to the same set.) He knows the distribution of such letters (e.g. that 'E' is far more common than 'Q'). He obtains the distribution of such letters and computes the corresponding entropy H. He computes the expected length of his code and is delighted to find $L = H + 1$. He concludes this is about as good as one can get. He applies his source coder to compress the text of his favourite novel (which is oddly enough written in all capital letters and no punctuation), and finds it reduces the length from 1,000,000 uncoded letters to only 500,000 coded letters. However his friend Erin reckons she can do better. She builds a coder that reduces the text to only 200,000 coded letters. How can this be? What must have Erin done? **(3 Marks)**

(e) Explain the idea of arithmetic coding conceptually and explain its advantages over symbol codes. Focus upon the encoding and the advantages, for example, in attacking Erin's problem in the previous part. **(3 Marks)**

SECTION D.

8. A triple of random variables (X, Y, Z) forms a Markov chain if

$$P(X, Y, Z) = P(X)P(Y|X)P(Z|Y).$$

   (a) Show that if (X, Y, Z) forms a Markov chain then so does (Z, Y, X). **(2 Marks)**

   (b) Using the previous part, and properties of Mutual Information, show that if (X, Y, Z) forms a Markov chain, then

$$I(X; Z) \leq \min(I(X; Y), I(Y; Z)).$$

   (4 Marks)

   (c) Let Q^{XY} be a noisy channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, and let Q^{YZ} be a noisy channel with input alphabet $\mathcal{Y}$ and output alphabet $\mathcal{Z}$. Let X be a random varible with distribution p on alphabet $\mathcal{X}$, and let Y denote the random variable obtained by passing X through the channel Q^{XY} and let Z denote the random variable obtained by passing Y through the channel Q^{YZ}.

      i. Write down an expression for the joint probability of (X, Y, Z) in terms of p and the transition matrices Q^{XY} and Q^{YZ}. **(3 Marks)**

      ii. Do the random variables (X, Y, Z) form a Markov chain? **(1 Marks)**

      iii. If the output of Q^{XY} is fed into the input of Q^{YZ} to form a new channel Q^{XZ}, what can be said about the capacity of the new channel in terms of the capacity of Q^{XY} annd Q^{YZ}. Give an intuitive explanation of your answer. **(5 Marks)**

9. A communication channel suitable for transmitting binary messages X taking a sequence of values in $\{-1, +1\}$ works as follows.

When $X = -1$, -5 Volts is transmitted over the wire.

When $X = +1$, +5 Volts is transmitted over the wire.

The receiver receives $Y = X + N$ where N is independent noise distributed uniformly on $[-B, B]$ Volts. The receiver decodes the signal via the operation

$$\hat{X} = \mathrm{sgn}(Y),$$

where $\mathrm{sgn}(y) = -1$ if $y < 0$ and equals $+1$ otherwise.

(a) Model this channel in terms of a transition matrix Q^{XY} that depends upon the value of B. **(3 Marks)**

(b) Compute the capacity C^{XY} of this channel (as a function of B). Show all your working. **(6 Marks)**

(c) Is it possible to construct an encoder and decoder pair such that when combined with this channel, the overall system achieves a probability of a (message symbol) error of 10^{-12} when transmitting at a rate $(1 - 10^{-6})C^{XY}$? **(2 Marks)**

(d) If one further imposes the constraint on the previous part that additionally one requires that the commnication has a latency of less than three input symbols, does your answer change. Explain why. (The "latency" here refers to the delay, in message symbol intervals, from transmission of a message symbol, to the correct decoding of the message symbol. Thus if 1000 message symbols are sent per second, the requirement is that the delay of the entire system is 3ms or less.) **(2 Marks)**

(e) Fearless Frank, who chose not to study Information Theory when he attended university, is asked by his boss to build a reliable communication system. His company has access to a remote site over a noisy channel and it costs \$1 per symbol to send a message, which is often garbled due to transmission errors. Frank did a quick Google search and found some free software that encodes and decodes in blocks of length N managing to send $N/2$ message symbols per block. Frank observes that when $N = 20$ the code runs very fast, but when N is large, it is much slower. When $N = 20$, the code can correct a modest number of errors, improving the error rate of the expensive channel. Frank thinks that fast software is the best thing, so he wants to stick with $N = 20$ and dreams up the following idea to boost the performance of the coding system reasoning as follows:

> *"If I take the output of the coder, and code that again with another copy of the encoder, sequentially encoding the length 10 blocks, then the overall message will be better protected. And I can do this over and over again. I will do it 10 times over — that is, 10 encoders in series, and at the receiver, 10 decoders in series."*

Frank tests the system by sending a message of 1000 symbols and is indeed delighted at its outstanding error correction performance. He proudly presents the results to his boss who fires him on the spot. Why? What should have Frank done? **(2 Marks)**

– – – END OF EXAM – – –