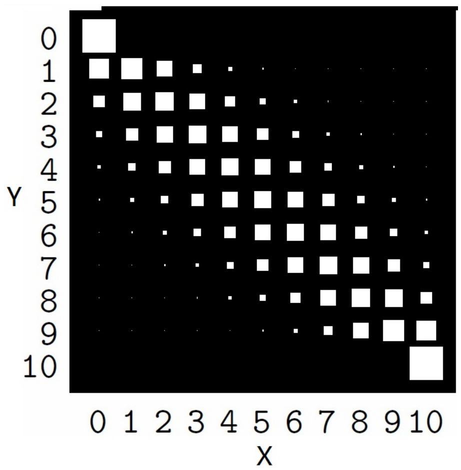
Section A.

Answer each of the following questions [Marks per questions as shown; 25% total]

1. (4 points) Suppose X and Y are random variables with 11 outcomes. The figure below shows a *Hinton* diagram of the joint distribution of X and Y, with the area of the white squares proportional to the corresponding $P(X = x, Y = y)$. (Wherever there is no visible white square at the cell for (x, y), it means $P(X = x, Y = y)$ is close to 0.)

   Is it possible for X and Y to be statistically independent? Justify your answer.



2. (6 points) Let X and Y be two random variables, both with possible outcomes $\{0, 1\}$.

   (a) (2 pts) Does specifying $P(X|Y)$ and $P(Y)$ fully determine the joint distribution $P(X, Y)$?

   (b) (4 pts) Does specifying $P(X|Y)$ and $P(Y|X)$ fully determine the joint distribution $P(X, Y)$?

   For both questions, if your answer is yes, then write down the formula of computing $P(X, Y)$ from the given quantities. Otherwise, give a counter-example.

3. (15 points) The Journal of Information Theory Research reviews all submitted papers in the following process. An editor first assigns the paper to two reviewers, who make the recommendation of either acceptance (1) or rejection (0). Based on their reviews, the editor makes the final decision. Even if the two reviews are consistent, the editor may still make the opposite decision. Let Z be the editor's decision, and X and Y be the reviewers' recommendation. Assume X and Y are independent and $P(X = 1) = P(Y = 1) = 0.5$. The conditional probability $P(Z = 0|X, Y)$ is given by

$P(Z = 0 \mid X, Y)$	$X = 0$	$X = 1$
$Y = 0$	0.9	0.5
$Y = 1$	0.5	0.1

   (a) (6 pts) Compute $P(X = 1|Z = 0)$, showing all your working.

   (b) (2 pts) If from the above you find $P(X = 1|Z = 0) > P(X = 1)$, explain intuitively why the probability increased; else, if you find $P(X = 1|Z = 0) < P(X = 1)$, explain why the probability decreased.

   (c) (5 pts) Compute $P(X = 1|Z = 0, Y = 1)$, showing all your working.

   (d) (2 pts) If you find $P(X = 1|Z = 0, Y = 1) > P(X = 1|Z = 0)$, explain intuitively why the probability increased; else, explain why the probability decreased.

Section B.

Answer each of the following questions [Marks per questions as shown; 25% total]

1. (5 points) Let X, Y be two random variables with the following joint distribution:

$P(X,Y)$		$X=1$	$X=2$
Y	1	1/4	0
	2	1/4	1/2

   (a) (3 pts) Compute $H(Y|X)$.

   (b) (2 pts) Compute $I(X;Y)$. (You may express your answer in terms of $\log_2$ of an appropriate integer.)

2. (6 points) Let X_1 and X_2 be random variables with possible outcomes $\mathcal{X}$. Suppose that these variables are identically distributed, i.e. that $P(X_1 = x) = P(X_2 = x)$ for all $x \in \mathcal{X}$. However, we do *not* assume X_1 and X_2 are independent. Now let

$$\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)}.$$

   (a) (2 pts) Prove that $\rho = \frac{I(X_1;X_2)}{H(X_1)}$.

   (b) (2 pts) Hence, or otherwise, prove that $0 \leq \rho \leq 1$.

   (c) (1 pt) When is $\rho = 0$? Justify your answer.

   (d) (1 pt) When is $\rho = 1$? Justify your answer.

3. (6 points) The post office in Union Court of ANU handles 10,000 letters per day on average.

   (a) (3 pts) Use Markov's inequality to derive an upper bound on the probability that more than 12,000 letters will be handled tomorrow. (You may leave your answer as a fraction.)

   (b) (3 pts) Suppose the variance of letters processed is 2,000. Use Chebyshev's inequality to derive an upper bound on the probability that this post office will handle between 8,000 and 12,000 letters tomorrow. (You may leave your answer as a fraction.)

4. (8 points) Recall that the 3-tuple of random variables (X, Y, Z) form a Markov chain if and only if

$$p(X, Y, Z) = p(X)\ p(Y \mid X)\ p(Z \mid Y).$$

   We can extend it to 4-tuple of random variables (X, Y, Z, W). They form a Markov chain if and only if

$$p(X, Y, Z, W) = p(X)\ p(Y \mid X)\ p(Z \mid Y)\ p(W \mid Z).$$

   In what follows, suppose (X, Y, Z, W) forms a Markov chain.

   (a) (2 pts) Prove that (X, Y, Z) forms a Markov chain.

   (b) (4 pts) Prove that (X, Z, W) forms a Markov chain.

   (c) (2 pts) Prove that $I(X;W) \leq I(Y;Z)$. (You may use without proof the following result from Tutorial 5: if (X, Y, Z) forms a Markov chain, then (Z, Y, X) also forms a Markov chain, and so $I(X;Z) \leq \min(I(X;Y), I(Y;Z))$. You may also use the results of (a) and (b), even if you are unable to prove them.)

Section C.

Answer each of the following questions [Marks per questions as shown; 25% total]

1. (5 pts) Suppose X is an ensemble over three possible outcomes, with probabilities $p_X = (0.4, 0.3, 0.3)$. Recall that X^N denotes an extended ensemble.

   (a) (1 pt) Compute the raw bit content $H_0(X)$. (You may express your answer in terms of $\log_2$ of an appropriate integer.)

   (b) (1 pt) What quantity, if any, does $\frac{1}{N} H_0(X^N)$ converge to as $N \to \infty$? Justify your answer.

   (c) (2 pts) Compute the essential bit content $H_\delta(X)$ when $\delta = 0.35$.

   (d) (1 pt) What quantity, if any, does $\frac{1}{N} H_\delta(X^N)$ converge to as $N \to \infty$ for $\delta = 0.35$? Justify your answer.

2. (12 pts) Suppose X is an ensemble over five possible outcomes, with probabilities $p_X = (0.3, 0.3, 0.2, 0.1, 0.1)$.

   (a) (4 pts) Compute a Huffman code C for X. Show all your working.

   (b) (2 pts) Compute the expected length $L(C, X)$ for your Huffman code.

   (c) (2 pts) Explain the relationship between $L(C, X)$ and the entropy $H(X)$.

   (d) (1 pt) Guffman, a self-trained mathematician, claims he has discovered a new prefix code C' which has expected length $L(C', X)$ strictly smaller than $L(C, X)$. By referencing an appropriate theorem from lectures, explain whether his claim is possible.

   (e) (1 pt) Hoffman, a self-trained quant, claims that he has constructed a new prefix code C'' has codeword lengths $(1, 1, 2, 2, 2)$. By referencing an appropriate theorem from lectures, explain whether his claim is possible.

   (f) (2 pts) Suppose we compute a Shannon code C''' for X. Should we expect $L(C''', X) = H(X)$? Explain why or why not.

3. (6 pts) Suppose X is an ensemble over outcomes $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$. Let the probabilities $p_X = (0.25, 0.5, 0.125, 0.125)$.

   (a) (2 pts) Compute the codeword lengths for all outcomes under a Shannon-Fano-Elias code for X.

   (b) (4 pts) Compute the Shannon-Fano-Elias codewords for `a` and `b`, showing all your working. (You do not need to compute codewords for `c` and `d`.)

4. (2 pts) Briefly describe one potential advantage of arithmetic coding over Huffman coding.

Section D.
Answer each of the following questions [Marks per questions as shown; 25% total]

1. (2 pts) Consider a binary symmetric channel with bit flip probability $f = 0.25$.
   (a) (1 pt) Write down the transition matrix Q for this channel.
   (b) (1 pt) What input distribution achieves the capacity of the channel? (You may refer to a result from lectures.)

2. (5 pts) Let Q denote some channel over input alphabet $\mathcal{X} = \{0, 1\}$ and output alphabet $\mathcal{Y} = \{0, 1\}$.
   (a) (1 pt) Donald computes the mutual information $I(X; Y)$ using an input distribution $p_{\mathcal{X}} = (0.5, 0.5)$. He finds that for this distribution, $I(X; Y) = 0.9$. Based on this fact, what is a lower bound on the capacity of Q? Justify your answer.
   (b) (2 pts) Provide an upper bound on the capacity of Q. Justify your answer.
   (c) (2 pts) Carly claims she can construct a block code that achieves a rate of 1.8 bits per transmission, with arbitrarily small probability of block error. Is her claim possible? Explain why or why not.

3. (14 pts) Consider a channel over inputs $\mathcal{X} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$ and outputs $\mathcal{Y} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$, with transition matrix

$$Q = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

   and input distribution $p_{\mathcal{X}} = (p_{\mathtt{a}}, p_{\mathtt{b}}, p_{\mathtt{c}}, p_{\mathtt{d}})$.
   (a) (2 pts) Is the channel Q symmetric? Explain why or why not.
   (b) (3 pts) Compute the probability distribution $p(Y)$ over outputs. (Express your answer in terms of $p_{\mathtt{a}}, p_{\mathtt{b}}, p_{\mathtt{c}}, p_{\mathtt{d}}$.)
   (c) (3 pts) Compute the conditional entropy $H(Y \mid X)$. (Express your answer in terms of $p_{\mathtt{a}}, p_{\mathtt{b}}, p_{\mathtt{c}}, p_{\mathtt{d}}$.)
   (d) (3 pts) Hence, show that

$$I(X; Y) = -(1 - p_{\mathtt{c}} - p_{\mathtt{d}}) \cdot \log_2(1 - p_{\mathtt{c}} - p_{\mathtt{d}}) - p_{\mathtt{c}} \cdot \log_2 p_{\mathtt{c}} - p_{\mathtt{d}} \cdot \log_2 p_{\mathtt{d}}.$$

   (e) (3 pts) What input distribution achieves the capacity of Q? Explain your answer intuitively.

4. (4 pts) Suppose we use a (7, 4) Hamming code to communicate over a binary symmetric channel with nonzero flip probability.
   (a) (3 pts) Compute the three parity bits for the message `1001`. You may use a diagram to show your working.
   (b) (1 pt) Suppose a receiver sees the bit string $\mathtt{1001b_1b_2b_3}$, where $\mathtt{b_1b_2b_3}$ is the parity bit string you computed above. Is it guaranteed that the sender actually transmitted `1001`? Explain why or why not.
