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

1. [5 points] In what follows, suppose that X, Y are random variables with possible outcomes $\{0, 1\}$. For each of the following statements, write whether they are True (**T**) or False (**F**). In each case, briefly justify your reasoning. (Answers that omit reasoning will *not* be awarded points.)

   i. If $p(X = 1) = p(Y = 1)$, then $p(X = 0) = p(Y = 0)$.

   ii. It is possible for the following inequalities to simultaneously hold:
   $$p(X = 1|Y = 1) < p(X = 1)$$
   $$p(X = 1|Y = 0) < p(X = 1).$$

   iii. If X and Y are statistically independent, then $p(X = 1|Y = 1) + p(X = 0|Y = 0) = 1$.

   iv. If $p(X = 1, Y = 1) = p(X = 1)p(Y = 1)$, then X and Y must be statistically independent.

   v. Assuming none of the involved probabilities are zero, it is always true that
   $$\frac{p(X = 1|Y = 1)}{p(Y = 1|X = 1)} = \frac{p(X = 1)}{p(Y = 1)}.$$

2. [15 points] In cricket, a batsman is considered to fail if he gets out in the first over. Michael is an excellent batsman, and he fails only 5% of the time. However, when Michael plays against the bowler Glenn, he fails 50% of the time. Fortunately for Michael, only 10% of his games are played against Glenn.

   i. What is the probability that Michael fails, given that he does *not play* against Glenn?

   ii. In a particular game, Michael fails. What is the probability that he *was playing* against Glenn that game?

   A bowler is considered to fail if he is hit for a six in the first over. Glenn has a good record: he fails in only 1% of the games he plays. However, of the times Glenn fails, 80% of the games are played against the batsman Sanath. Further, of the games Glenn plays against Sanath, Glenn fails 20% of the time.

   iii. What fraction of the games Glenn plays are against Sanath?

3. [5 points] For his birthday, Carl is given a toy with a button and LED display on the outside, and a two-sided coin and a six-sided die hidden away on the inside. When the button is pressed, the toy flips the coin and rolls the die, and displays the outcome of each item on the LED display.

   Carl is interested in studying whether or not there is a relationship between the outcome of the coin and die. Carl performs $N > 1$ independent trials where he presses the button. For each trial, he records the displayed outcome of the coin and die as $\{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$, where $x^{(i)} \in \{0, 1\}$ and $y^{(i)} \in \{1, 2, \ldots, 6\}$.

i. Let X be the random variable representing the outcome of the coin flip, and Y be the random variable representing the outcome of the die roll. Carl wishes to compute $p(X = a, Y = b)$, the probability that the coin comes up with value a, and the die with value b, for $a \in \{0, 1\}, b \in \{1, 2, \ldots, 6\}$. Write down the maximum likelihood estimate $\hat{p}_{\mathrm{MLE}}(X = a, Y = b)$ for this probability. Explain the intuitive meaning of your answer.

ii. Carl similarly computes the maximum likelihood estimates of $p(X = a)$, the marginal probability that the coin comes up a, and $p(Y = b)$, the marginal probability that the die comes up b, for every a, b. Let these estimates be $\hat{p}_{\mathrm{MLE}}(X = a)$ and $\hat{p}_{\mathrm{MLE}}(Y = b)$ respectively. He finds that for every a, b,

$$\hat{p}_{\mathrm{MLE}}(X = a, Y = b) \neq \hat{p}_{\mathrm{MLE}}(X = a) \cdot \hat{p}_{\mathrm{MLE}}(Y = b).$$

Carl excitedly concludes, "The coin flip and die roll *must* be dependent, because my observations violate the definition of statistical independence". Carl's sister Judith, who watches the experiment, says "No, they *could* be independent; you cannot make a definitive conclusion based on *estimates* of the true probabilities." Who is correct, Carl or Judith? Explain your answer.

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

1. [15 points] Let X be a random variable with possible outcomes $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$, each with probabilities $\mathbf{p} = \{1/2, 1/4, 1/8, 1/8\}$. Similarly, let Y be a random variable with the same possible outcomes, but with probabilities $\mathbf{q} = \{1/4, 1/4, 1/4, 1/4\}$.

   i. Compute $H(X)$ and $H(Y)$.
   ii. Compute $D_{\mathrm{KL}}(\mathbf{p}\|\mathbf{q})$ and $D_{\mathrm{KL}}(\mathbf{q}\|\mathbf{p})$.
   iii. Let $\mathbf{r}$ be a probability vector such that $D_{\mathrm{KL}}(\mathbf{r}\|\mathbf{p}) = 0$. Is it true that $D_{\mathrm{KL}}(\mathbf{p}\|\mathbf{r}) = 0$ also? Justify your answer.
   iv. Let $\mathbf{s}$ be a probability vector such that $D_{\mathrm{KL}}(\mathbf{s}\|\mathbf{p}) = D_{\mathrm{KL}}(\mathbf{p}\|\mathbf{s})$. Is it true that $\mathbf{p} = \mathbf{s}$? Justify your answer.
   v. Consider the coding scheme

$$\begin{aligned} \mathtt{a} &\to \mathtt{0} \\ \mathtt{b} &\to \mathtt{10} \\ \mathtt{c} &\to \mathtt{110} \\ \mathtt{d} &\to \mathtt{111}. \end{aligned}$$

   Explain why this scheme is optimal for encoding the random variable X, but not Y.

2. [5 points] Ruixin selects a number X from $\{1, 2, \ldots, 100\}$ according to some probability distribution $\mathbf{p}$. Gary wants to find the selected number by asking Ruixin questions about it. Fortunately, Gary knows the distribution $\mathbf{p}$ that Ruixin uses.

   i. Let Q_i denote the question "Is $X = i$?", to which the possible responses are "Yes" if $X = i$, or "No" if $X \neq i$. Gary selects questions from $Q_1, Q_2, \ldots, Q_{100}$. What is the best first question he can ask? (Your answer should be based on the distribution $\mathbf{p}$.)
   ii. Suppose Gary determines the optimal series of questions from $Q_1, \ldots, Q_{100}$ to ask. Is it possible that for a *particular* choice of X, Gary asks *fewer* than $H(X)$ questions? Justify your answer.
   iii. Now suppose that the possible responses to Q_i are "Yes" if $X = i$, "Too High" if $X < i$, or "Too Low" if $X > i$. In general, will the best first question to ask in this case be the *same* as that found in part (i)? Explain why or why not.

3. [5 points] Let X and Y be two random variables with the following joint distribution:

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

   i. Compute $p(X = 1)$ and $p(Y = 1)$.
   ii. Hence, or otherwise, show that $I(X;Y) = 0$. Explain the intuitive meaning of this statement.
   iii. Simon says that the data processing inequality implies that for *any* random variable Z, $I(X;Y|Z) = 0$. Explain whether or not Simon is correct.

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

1. [10 points] Let X be an ensemble with alphabet $\mathcal{A}_X = \{\mathtt{a}, \mathtt{b}\}$ with probabilities $p_\mathtt{a} = \frac{1}{3}$ and $p_\mathtt{b} = \frac{2}{3}$. The entropy $H(X) \approx 0.92$.
   i. Write down the alphabet $\mathcal{A}_{X^3}$ for the extended ensemble X^3.
   ii. What is the probability of the sequence `abb` in X^3?
   iii. Which of the following sequences are in the typical set $T_{N\beta}$ for $N = 3$ and $\beta = 0.2$?
      a) `aaa` with log probability $\log_2 P(\mathtt{aaa}) \approx -4.8$
      b) `abb` with log probability $\log_2 P(\mathtt{abb}) \approx -2.8$
      c) `bbb` with log probability $\log_2 P(\mathtt{bbb}) \approx -1.8$
   iv. Write down a smallest δ-sufficient subset S_δ for X^3 when $\delta = \frac{1}{25}$.
   v. What is the essential bit content $H_\delta(X^3)$ for $\delta = \frac{1}{25}$?

2. [8 points]
   Let X be an ensemble with $\mathcal{A}_X = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}\}$ and probabilities $\mathbf{p} = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16})$.
   i. Compute the entropy $H(X)$, expressing your answer as simply as possible.
   ii. Construct a Huffman code C for X and compute its expected code length $L(C, X)$.
   iii. Is it possible to construct a prefix-free code C' for X with smaller expected code length than C? If so, construct such a code. If not, explain why not.
   iv. What is the interval in Shannon-Fano-Elias coding that corresponds to the symbol `b`? Express this interval in decimal and in binary.

3. [3 points] Is it possible for a variable-length code to uniquely decodeable but not prefix-free? If so, give an example and explain why it is uniquely decodeable. If not, provide an argument as to why not.

4. [4 points] While the definitions and interpretations of (prefix) Kolmogorov complexity K and Shannon entropy H differ fundamentally, they share many properties.
   i. Write down the definitions of K and H and their conditional version.
   ii. Write down the five relations (for each, K and H) they formally share (upper bound, extra information, subadditivity, symmetry, information non-increase).
   iii. What are the pros and cons of K versus H?

Section D.

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

1. [10 points] Consider a noisy channel Q that has input symbols $\mathcal{X} = \{0, 1\}$ and output symbols $\mathcal{Y} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$. When the input is 0 is there is a 50% chance of receiving an a and a 50% chance of receiving a b. When the input is 1 all four of the output symbols are equally likely to be received.

   i. Write down the matrix of transition probabilities for Q.

   ii. Is Q a symmetric channel? Explain why or why not.

   iii. Write down the distribution over output symbols assuming that the distribution over inputs is $\mathbf{p}_X = (0.5, 0.5)$.

   iv. Compute the mutual information $I(X;Y)$ between the inputs and outputs assuming the same input distribution as the previous part. Use $\log_2 3 \approx 1.6$ to give an approximate value for $I(X;Y)$.

   v. Using the previous answer, explain whether or not it is possible for Q's channel capacity to be equal to 0.1 bits/usage.

2. [8 points] Suppose $\mathbf{x}^{(1)} = 0001$, $\mathbf{x}^{(2)} = 0010$, $\mathbf{x}^{(3)} = 0100$, $\mathbf{x}^{(4)} = 1000$ is a code for the channel Q with $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ with transition probabilities given by

$$Q = \begin{bmatrix} 1 & 0.25 \\ 0 & 0.75 \end{bmatrix}.$$

   i. What is the rate of the code?

   ii. Give the probability $P(Y = \mathbf{y}|s = 2)$ of seeing each of following output sequences given message $s = 2$ was coded and sent across Q:

      i. $\mathbf{y} = 0010$
      ii. $\mathbf{y} = 1111$
      iii. $\mathbf{y} = 0000$
      iv. $\mathbf{y} = 1000$

   iii. Suppose outputs $\mathbf{y}$ from the channel Q are decoded by the function

$$\hat{s}(\mathbf{y}) = \begin{cases} s & \text{if } \mathbf{y} = \mathbf{x}^{(s)} \\ 0 & \text{otherwise.} \end{cases}$$

   That is, the decoder $\hat{s}$ just finds the message s that codes to to $\mathbf{y}$ if $\mathbf{y} \in \{0001, 0010, 0100, 1000\}$ and returns 0 otherwise. Assuming the probability of sending each input message is equal, what is the probability of error $P(\hat{s} \neq s)$ when using coding/decoding scheme across Q?

3. [3 points] Compute the channel capacity for each of the following:

   (a) A channel Q with $\mathcal{X} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ and a single output $\mathcal{Y} = \{0\}$.

   (b) A channel Q' with transition probability matrix given by

$$Q' = \begin{bmatrix} 0.25 & 0.25 & 0.5 \\ 0.5 & 0.25 & 0.25 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}$$

4. [4 points] For an arbitrary noisy channel Q with N input symbols and and M output symbols, prove that its capacity C satisfies $\mathrm{C} \leq \min\{\log_2 N, \log_2 M\}$.