Question 1. (2 marks)

(Alternative proof of the chain rule for entropy)

Let $X_1, X_2$ and $X_n$ be random variables such that $(X_1, X_2, \ldots, X_n) \sim p(x_1, \ldots, x_n)$. Prove that

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \mid x_{i-1}, \ldots, x_1).$$

Show also that the chain rule for entropy, i.e.,

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i \mid X_{i-1}, \ldots, X_1),$$

is a consequence of the previous formulation of the joint probability.

Question 2. (1 mark)

(From Jensen’s Inequality)

Let $f$ be a strictly convex function and $X$ a discrete random variable. Prove that if $E[f(X)] = f(E[X])$, then $X = E[X]$ with probability 1, i.e., $X$ is a constant.

Question 3. (4 marks)

(Coin flips)

A fair coin is flipped until the first head occurs. Let $X$ denote the number of flips required.

1. Find the entropy $H(X)$ in bits. The following expressions may be useful:

$$\sum_{i=0}^{\infty} r^i = (1 - r)^{-1} \quad \text{and} \quad \sum_{i=0}^{\infty} ir^i = r(1 - r)^{-2}.$$ 

2. A random variable $X$ is drawn according to this distribution. Find an “efficient” sequence of yes-no questions of the form “Is $X$ contained in the set $S$?”. Compare $H(X)$ to the expected number of questions required to determine $X$.

Question 4. (2 marks)

(Zero conditional entropy)

Show that if $H(Y \mid X) = 0$, then $Y$ is a function of $X$, i.e., for all $x$ such that $p(x) > 0$, there exists only one possible value of $y$ such that $p(x, y) > 0$.

Question 5. (6 marks)

(Entropy of a sum)

Let $X$ and $Y$ be random variables that take on values $x_1, x_2, \ldots, x_r$ and $y_1, y_2, \ldots, y_s$, respectively. Let $Z = X + Y$.

1. Show that $H(Z \mid X) = H(Y \mid X)$. Argue that if $X, Y$ are independent, then $H(Y) \leq H(Z)$ and $H(X) \leq H(Z)$. Thus, the addition of independent random variables add uncertainty.

2. Give an example of random variables in which $H(X) > H(Z)$ and $H(Y) > H(Z)$.

3. Under what conditions does $H(Z) = H(X) + H(Y)$?