Required Problems

1. We wish to compress a sequence of independent, identically distributed random variables $X_1, X_2, \ldots$. Each $X_j$ takes on one of $n$ values. The $i$th value occurs with probability $p_i$, where $p_1 \geq p_2 \geq \ldots \geq p_n$. The result is compressed as follows. Set $T_i = \sum_{j=1}^{i-1} p_j$, and let the $i$th codeword be the first $\lceil \log_2(1/p_i) \rceil$ bits of $T_i$. Start with an empty string, and consider $X_j$ in order. If $X_j$ takes on the $i$th value, append the $i$th codeword to the end of the string.

   (A) Show that no codeword is the prefix of any other codeword.
   (B) Let $Z$ be the average number of bits appended for each random variable $X_j$. Show that $H(X_j) \leq z \leq H(X_j) + 1$.

2. Arithmetic coding is a standard compression method. In the case when the string to be compressed is a sequence of biased coin flips, it can be described as follows. Suppose that we have a sequence of bits $X = (X_1, X_2, \ldots, X_n)$, where each $X_i$ is independently 0 with probability $p$ and 1 with probability $1-p$. The sequences can be ordered lexicographically, so for $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$, we say that $x < y$ if $x_i = 0$ and $y_i = 1$ in the first coordinate $i$ such that $x_i \neq y_i$. If $z(x)$ is the number of zeroes in the string $x$, then define $p(x) = p^{z(x)}(1-p)^{n-z(x)}$, and

   $$q(x) = \sum_{y < x} p(y).$$

   (A) Suppose we are given $X = (X_1, X_2, \ldots, X_n)$. Explain how to compute $q(X)$ in time $O(n)$ (assume that any reasonable operation on real numbers takes constant time).
   (B) Argue that the intervals $[q(x), q(x) + p(x)]$ are disjoint subintervals of $[0,1)$.
   (C) Given (A) and (B), the sequence $X$ can be represented by any point in the interval $I(X) = [q(X), q(X) + p(X)]$. Show that we can choose a codeword in $I(X)$ with $\lceil \log(1/p(X)) \rceil + 1$ binary decimal digits to represent $X$ in such a way that no codeword is the prefix of any other codeword.
   (D) Given a codeword chosen as in (C), explain how to decompress it to determine the corresponding sequence $(X_1, X_2, \ldots, X_n)$.
   (E) Using the Chernoff inequality, argue that $\log(1/p(X))$ is close to $nH(p)$ with high probability. Thus, this approach yields an effective compression scheme.

(a) Let \( S = \sum_{i=1}^{10} 1/i^2 \). Consider a random variable \( X \) such that \( \Pr[X = i] = 1/(Si^2) \), for \( i = 1, \ldots, 10 \). Compute \( \mathbb{H}(X) \).

(b) Let \( S = \sum_{i=1}^{10} 1/i^3 \). Consider a random variable \( X \) such that \( \Pr[X = i] = 1/(Si^3) \), for \( i = 1, \ldots, 10 \). Compute \( \mathbb{H}(X) \).

(c) Let \( S(\alpha) = \sum_{i=1}^{10} 1/i^\alpha \), for \( \alpha > 1 \). Consider a random variable \( X \) such that \( \Pr[X = i] = 1/(S(\alpha)i^\alpha) \), for \( i = 1, \ldots, 10 \). Prove that \( \mathbb{H}(X) \) is either increasing or decreasing as a function of \( \alpha \) (you can assume that \( \alpha \) is an integer).

4. Consider an \( n \)-sided die, where the \( i \)th face comes up with probability \( p_i \). Show that the entropy of a die roll is maximized when each face comes up with equal probability \( 1/n \).

5. The conditional entropy \( \mathbb{H}(Y|X) \) is defined by

\[
\mathbb{H}(Y|X) = \sum_{x,y} \Pr[(X = x) \cap (Y = y)] \log \frac{1}{\Pr[Y = y|X = x]}.
\]

If \( Z = (X, Y) \), prove that

\[
\mathbb{H}(Z) = \mathbb{H}(X) + \mathbb{H}(Y|X).
\]

6. We have shown that we can extract, on average, at least \( \lfloor \log m \rfloor - 1 \) independent, unbiased bits from a number chosen uniformly at random from \( \{0, \ldots, m-1\} \). It follows that if we have \( k \) numbers chosen independently and uniformly at random from \( \{0, \ldots, m-1\} \) then we can extract, on average, at least \( k \lfloor \log m \rfloor - k \) independent, unbiased bits from them. Give a better procedure that extracts, on average, at least \( k \lfloor \log m \rfloor - 1 \) independent, unbiased bits from these numbers.

7. Assume you have a (valid) prefix code with \( n \) codewords, where the \( i \)th codeword is made out of \( \ell_i \) bits. Prove that

\[
\sum_{i=1}^{n} \frac{1}{2^{\ell_i}} \leq 1.
\]