Problem 1.

(a) Since the $X_1, \ldots, X_n$ are i.i.d., so are $p(X_1), p(X_2), \ldots, p(X_n)$, and hence we can apply the law of large numbers to obtain

$$\lim -\frac{1}{n} \log p(X_1, \ldots, X_n) = \lim -\frac{1}{n} \sum \log p(X_i)$$

$$= -E[\log p(X)]$$

$$= -\sum p(x) \log p(x)$$

$$= H(X).$$

(b) Since the $X_1, \ldots, X_n$ are i.i.d., so are $q(X_1), q(X_2), \ldots, q(X_n)$, and hence we can apply the law of large numbers to obtain

$$\lim -\frac{1}{n} \log q(X_1, \ldots, X_n) = \lim -\frac{1}{n} \sum \log q(X_i)$$

$$= -E[\log q(X)]$$

$$= -\sum p(x) \log q(x)$$

$$= \sum p(x) \log \frac{p(x)}{q(x)} - \sum p(x) \log p(x)$$

$$= D(p||q) + H(X).$$

(c) Again, by the law of large numbers,

$$\lim -\frac{1}{n} \log \frac{q(X_1, \ldots, X_n)}{p(X_1, \ldots, X_n)} = \lim -\frac{1}{n} \sum \log \frac{q(X_i)}{p(X_i)}$$

$$= -E \left[ \log \frac{q(X)}{p(X)} \right]$$

$$= -\sum p(x) \log \frac{q(x)}{p(x)}$$

$$= \sum p(x) \log \frac{p(x)}{q(x)}$$

$$= D(p||q).$$

Problem 2.

(a) It is easy to check that $W$ is an i.i.d. process but $Z$ is not. As $W$ is i.i.d. it is also stationary. We want to show that $Z$ is also stationary. To show this, it is sufficient
to prove that the distribution of the process does not change by shift in the time domain.

\[ p_Z(Z_m = a_m, Z_{m+1} = a_{m+1}, \ldots, Z_{m+r} = a_{m+r}) \]
\[ = \frac{1}{2}p_X(X_m = a_m, X_{m+1} = a_{m+1}, \ldots, X_{m+r} = a_{m+r}) \]
\[ + \frac{1}{2}p_Y(Y_m = a_m, Y_{m+1} = a_{m+1}, \ldots, Y_{m+r} = a_{m+r}) \]
\[ = \frac{1}{2}p_X(X_{m+s} = a_m, X_{m+s+1} = a_{m+1}, \ldots, X_{m+s+r} = a_{m+r}) \]
\[ + \frac{1}{2}p_Y(Y_{m+s} = a_m, Y_{m+s+1} = a_{m+1}, \ldots, Y_{m+s+r} = a_{m+r}) \]
\[ = p_Z(Z_{m+s} = a_m, Z_{m+s+1} = a_{m+1}, \ldots, Z_{m+s+r} = a_{m+r}), \]

where we used the stationarity of the \( X \) and \( Y \) processes. This shows the invariance of the distribution with respect to the arbitrary shift \( s \) in time which implies stationarity.

(b) For the \( Z \) process we have

\[ H(Z) = \lim_{n \to \infty} \frac{1}{n} H(Z_1, \ldots, Z_n) \]
\[ = \lim_{n \to \infty} \frac{1}{n} H(Z_1, \ldots, Z_n | \Theta) \]
\[ = \frac{1}{2} H(X_0) + \frac{1}{2} H(Y_0) = 1. \]

\( W \) process is an i.i.d process with the distribution \( p_W(a) = \frac{1}{2} p_X(a) + \frac{1}{2} p_Y(a) \). From concavity of the entropy, it is easy to see that \( H(W) = H(W_0) \geq \frac{1}{2} H(X_0) + \frac{1}{2} H(Y_0) = 1. \) Hence, the entropy rate of \( W \) is greater than the entropy rate of \( Z \) and the equality holds if and only if \( X_0 \) and \( Y_0 \) have the same probability distribution function.

**Problem 3.** Upon noticing \( 0.9^6 > 0.1 \), we obtain \( \{1, 01, 001, 0001, 00001, 000001, 0000001, 0000000\} \) as the dictionary entries.

**Problem 4.** Since the words of a valid and prefix condition dictionary reside in the leaves of a full tree, the Kraft inequality must be satisfied with equality: Consider climbing up the tree starting from the root, choosing one of the \( D \) branches that climb up from a node with equal probability. The probability of reaching a leaf at depth \( l_i \) is then \( D^{-l_i} \). Since the climbing process will certainly end in a leaf, we have

\[ 1 = \Pr(\text{ending in a leaf}) = \sum_i D^{-l_i}. \]

If the dictionary is valid but not prefix-free, by removing all words that already have a prefix in the dictionary we would obtain a valid prefix-free dictionary. Since this reduced dictionary would satisfy the Kraft inequality with equality, the extra words would cause the inequality to be violated.

**Problem 5.**

(a) Let \( I \) be the set of intermediate nodes (including the root), let \( N \) be the set of nodes except the root and let \( L \) be the set of all leaves. For each \( n \in L \) define \( A(n) = \{m \in N : m \text{ is an ancestor of } n\} \) and for each \( m \in N \) define \( D(m) = \{n \in L : m \)
$L : n$ is a descendant of $m \}$. We assume each leaf is an ancestor and a descendant of itself. Then

$$E[\text{distance to a leaf}] = \sum_{n \in L} P(n) \sum_{m \in A(n)} d(m)$$

$$= \sum_{m \in N} d(m) \sum_{n \in D(m)} P(n) = \sum_{m \in N} P(m)d(m).$$

(b) Let $d(n) = -\log Q(n)$. We see that $-\log P(n_j)$ is the distance associated with a leaf. From part (a),

$$H(\text{leaves}) = E[\text{distance to a leaf}]$$

$$= \sum_{n \in N} P(n)d(n)$$

$$= -\sum_{n \in N} P(n) \log Q(n)$$

$$= -\sum_{n \in N} P(\text{parent of } n) Q(n) \log Q(n)$$

$$= -\sum_{m \in I} \sum_{n \text{ : } n \text{ is a child of } m} Q(n) \log Q(n)$$

$$= \sum_{m \in I} P(m) H_{m'}.$$

(c) Since all the intermediate nodes of a valid and prefix condition dictionary have the same number of children with the same set of $Q_n$, each $H_n = H$. Thus $H(\text{leaves}) = H \sum_{n \in I} P(n) = HE[L]$.

PROBLEM 6.

(a) We have

$$E[-\log_2 q(X)] = -\sum_x p(x) \log_2 q(x)$$

$$= \sum_x p(x) \log_2 \frac{p(x)}{p(x)q(x)}$$

$$= \sum_x p(x) \log_2 \frac{1}{p(x)} + \sum_x p(x) \log_2 \frac{p(x)}{q(x)}$$

$$= H(p) + D(p\|q).$$

(b) When $q(x)$ is an integer power of $\frac{1}{2}$, the code which minimizes $\sum_x q(x)[\text{length}[C(x)]]$ will choose $\text{length}[C(x)] = -\log_2 q(x)$.

(c) From part (a) and (b) we see that

$$E[\text{length}[C(x)]] - H(p) = H(p) + D(p\|q) - H(p) = D(p\|q).$$