Problem 1.

(a) Recall that $C$ is uniquely decodable means that $C^*$ is injective, i.e., for any $u^n \neq v^m$ we have $C^n(u^n) \neq C^m(v^m)$. In particular, whenever $u^n \neq v^n$ we have $C^n(u^n) \neq C^n(v^n)$. The last statement is the definition of $C^n$ being injective.

(b) Since we are supposed to show that $u_1 \neq v_1$, we may assume that $|U| \geq 2$.

If $C$ is not uniquely decodable, then there are $u^n \neq v^m$ such that $C^n(u^n) = C^m(v^m)$. Among all such $(u^n, v^m)$ choose one for which $n + m$ is smallest, and assume (without loss of generality) that $m \leq n$. If $m \geq 1$ we are done, since in this case we must have $u_1 \neq v_1$ (because, if not, we can replace $u^n$ by $\tilde{u}^{n-1} = u_2 \ldots u_n$ and $v^m$ by $\tilde{v}^{m-1} = v_2 \ldots v_m$, contradicting $m + n$ being smallest).

Otherwise, $m = 0$ and $v^m = \lambda$ (the null string) with $C(v^m) = \lambda$. Since $u^n \neq v^m = \lambda$ and $C(u^n) = \lambda$, we have a letter $a = u_1 \in U$ such that $C(a) = \lambda$. Take now any letter $b \in U$ with $b \neq a$, and note that $C^2(ab) = C^1(b)$, i.e., there are two source sequences that differ in their first letter and have the same representation.

(c) $C$ is not uniquely decodable means that there is $u^n \neq v^m$ such that $C^n(u^n) = C^m(v^m)$. If $n = m$ then we are done: this would by definition mean that be $C^n$ is not injective. If $n \neq m$, we could attempt the following reasoning: observe $C^*(u^nv^m) = C^*(v^mu^n)$ and conclude that $C^{m+n}$ is not injective. However this reasoning fails because we can’t be sure that $u^mv^n \neq v^mu^n$ just because $u^n \neq v^m$. (E.g., suppose $u^n = a$ and $v^m = aa$). This is the reason the problem has “part (b)”:

As $C$ is not uniquely decodable, we can find $u^n$ and $v^m$ as in part (b). Now observe that (i) $u^nv^m \neq v^mu^n$ (as they differ in their first letter), (ii) $u^n v^m$ and $v^m u^n$ have the same length $k = n + m$, and $C^k(u^n v^m) = C^k(v^m u^n)$, i.e., $C^k$ is not singular.

Morale of the problem: it is clear that the statement “$C^*$ is injective” is a stronger statement than “for every $n$, $C^n$ is injective” — since the first ensures that $u^n \neq v^m$ are assigned different codewords not only when $n = m$ but also for $n \neq m$ — so part (a) is unsurprising. The statement “$C^n$ is injective for each $n$” only means that different source sequences of same length get different representations; it is not immediately clear that this will also imply that source sequences of different lengths also get different representations. Part (c) shows this is indeed the case: that injectiveness of $C^n$ for every $n$ implies the injectiveness of $C^*$. 
Problem 2.

(a) We already know that
\[ H(X) + H(Y) \geq H(XY), \]
\[ H(Y) + H(Z) \geq H(YZ), \]
and
\[ H(Z) + H(X) \geq H(ZX). \]
Adding these inequalities together and diving by two gives
\[ H(X) + H(Y) + H(Z) \geq \frac{1}{2} [H(XY) + H(YZ) + H(ZX)]. \]

(b) The difference between the left and right sides, i.e.,
\[ H(XY) + H(YZ) - H(XYZ) - H(Y), \]
equals
\[ H(X|Y) - H(X|YZ) = I(X; Z|Y), \]
which is always positive.

(c) Using (b) with \((YZX)\) and \((ZXY)\) in the role of \((XYZ)\) gives the inequalities
\[ H(YZ) + H(ZX) \geq H(XYZ) + H(Z) \]
and
\[ H(ZX) + H(XY) \geq H(XYZ) + H(X). \]
Adding the inequality in (b) to these two gives
\[ 2[H(XY) + H(YZ) + H(ZX)] \geq 3H(XYZ) + H(X) + H(Y) + H(Z). \]

(d) Since \(H(X) + H(Y) + H(Z) \geq H(XYZ)\), (c) yields
\[ 2[H(XY) + H(YZ) + H(ZX)] \geq 4H(XYZ). \]

(e) Let \(\{(x_i, y_i, z_i) : i = 1, \ldots, n\}\) be the \(xyz\)-coordinates of the \(n\) points. Let \(X, Y\) and \(Z\) be random variables with \(\Pr((X, Y, Z) = (x_i, y_i, z_i)) = 1/n\) for every \(1 \leq i \leq n\). Then, \(H(XYZ) = \log_2 n\). Furthermore, the random pair \((XY)\) takes values in the projection of the \(n\) points to the \(xy\) plane and similarly for \((YZ)\) and \((ZX)\). Thus \(H(XY) \leq \log_2 n_{xy}, H(YZ) \leq \log_2 n_{yz},\) and \(H(ZX) \leq \log_2 n_{zx}\). Part (d) now yields
\[ \log_2 [n_{xy}n_{yz}n_{zx}] \geq H(XY) + H(YZ) + H(ZX) \geq 2H(XYZ) = 2 \log_2 n, \]
which implies that \(n_{xy}n_{yz}n_{zx} \geq n^2\).

The relationship between \(H(XYZ)\) and \(H(XY), H(YZ)\) and \(H(ZX)\) is a special case of Han’s inequality, which, for a collection of \(n\) random variables relates the sum of the \(\binom{n}{k}\) joint entropies of \(k\) out of \(n\) random variables to the sum of the \(\binom{n}{k+1}\) entropies of \(k + 1\) out of \(n\) random variables.

The combinatorial fact about the projections of points in 3D is known as Shearer’s lemma.
Problem 3.

\[ H(X) = -\sum_{k=1}^{M} P_X(a_k) \log P_X(a_k) \]

\[ = - \sum_{k=1}^{M-1} (1-\alpha)P_Y(a_k) \log[(1-\alpha)P_Y(a_k)] - \alpha \log \alpha \]

\[ = (1-\alpha)H(Y) - (1-\alpha)\log(1-\alpha) - \alpha \log \alpha \]

Since \( Y \) is a random variable that takes \( M - 1 \) values \( H(Y) \leq \log(M-1) \) with equality if and only if \( Y \) takes each of its possible values with equal probability.

Problem 4.

(a) Using the chain rule for mutual information,

\[ I(X, Y; Z) = I(X; Z) + I(Y; Z | X) \geq I(X; Z), \]

with equality iff \( I(Y; Z | X) = 0 \), that is, when \( Y \) and \( Z \) are conditionally independent given \( X \).

(b) Using the chain rule for conditional entropy,

\[ H(X, Y | Z) = H(X | Z) + H(Y | X, Z) \geq H(X | Z) \]

with equality iff \( H(Y | X, Z) = 0 \), that is, when \( Y \) is a function of \( X \) and/or \( Z \).

(c) Using first the chain rule for entropy and then the definition of conditional mutual information,

\[ H(X, Y, Z) - H(X, Y) = H(Z | X, Y) = H(Z | X) - I(Y; Z | X) \]

\[ \leq H(Z | X) = H(X, Z) - H(X), \]

with equality iff \( I(Y; Z | X) = 0 \), that is, when \( Y \) and \( Z \) are conditionally independent given \( X \).

(d) Using the chain rule for mutual information,

\[ I(X; Z | Y) + I(Z; Y) = I(X, Y; Z) = I(Z; Y | X) + I(X; Z), \]

and therefore

\[ I(X; Z | Y) = I(Z; Y | X) - I(Z; Y) + I(X; Z). \]

We see that this inequality is actually an equality in all cases.

Problem 5. Let \( X' \) denote \( X_1, \ldots, X_i \).

(a) By stationarity we have for all \( 1 \leq i \leq n \),

\[ H(X_n | X^{n-1}) \leq H(X_n | X_{n-i+1}, X_{n-i+2}, \ldots, X_{n-1}) = H(X_i | X^{i-1}) \]

which implies that,

\[ H(X_n | X^{n-1}) = \frac{\sum_{i=1}^{n} H(X_n | X^{n-1})}{n} \]

(1)

\[ \leq \frac{\sum_{i=1}^{n} H(X_i | X^{i-1})}{n} \]

(2)

\[ = \frac{H(X_1, X_2, \ldots, X_n)}{n} \]

(3)
(b) By the chain rule for entropy,

\[
\frac{H(X_1, X_2, \ldots, X_n)}{n} = \sum_{i=1}^{n} \frac{H(X_i|X^{i-1})}{n} \\
= \frac{H(X_n|X^{n-1}) + \sum_{i=1}^{n-1} H(X_i|X^{i-1})}{n} \\
= \frac{H(X_n|X^{n-1}) + H(X_1, X_2, \ldots, X_{n-1})}{n}. \tag{4}
\]

From stationarity it follows that for all \(1 \leq i \leq n\),

\[
H(X_n|X^{n-1}) \leq H(X_i|X^{i-1}),
\]

which further implies, by summing both sides over \(i = 1, \ldots, n-1\) and dividing by \(n-1\), that,

\[
H(X_n|X^{n-1}) \leq \frac{\sum_{i=1}^{n-1} H(X_i|X^{i-1})}{n-1} = \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1}. \tag{7}
\]

Combining (6) and (8) yields,

\[
\frac{H(X_1, X_2, \ldots, X_n)}{n} \leq \frac{1}{n} \left[ \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1} + H(X_1, X_2, \ldots, X_{n-1}) \right] \\
= \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1}. \tag{9}
\]

**Problem 6.** By the chain rule for entropy,

\[
H(X_0|X_{-1}, \ldots, X_{-n}) = H(X_0, X_{-1}, \ldots, X_{-n}) - H(X_{-1}, \ldots, X_{-n}) \tag{11}
\]

\[
= H(X_0, X_1, \ldots, X_n) - H(X_1, \ldots, X_n) \tag{12}
\]

\[
= H(X_0|X_1, \ldots, X_n), \tag{13}
\]

where (12) follows from stationarity.

**Problem 7.** \(X \rightarrow Y \rightarrow (Z, W)\) implies that \(I(X; Z, W|Y) = 0\). Then,

\[
I(X; Y) + I(Z; W) = I(X; Y) + I(X; Z, W|Y) + I(Z; W) = I(X; Y, Z, W) + I(Z; W)
\]

Notice that \(I(X; Y) + I(X; Z, W|Y) = I(X; Y, Z, W)\) follows from chain rule. Using the chain rule for a couple of times, we obtain the following steps.

\[
I(X; Y, Z, W) + I(Z; W) = I(X; Z) + I(X; Y, W|Z) + I(Z; W) \tag{14}
\]

\[
= I(X; Z) + I(X; Y|W, Z) + I(X; W|Z) + I(Z; W) \tag{15}
\]

\[
= I(X; Z) + I(X; Y|W, Z) + I(X, Z; W) \tag{16}
\]

\[
\geq I(X; Z) + I(X; W) \tag{17}
\]

as \(I(X, Z; W) \geq I(X; W)\)