PROBLEM 1.

(a) Let $U$ be a random variable taking values in the alphabet $\mathcal{U}$, and let $f$ be a mapping from $\mathcal{U}$ to $\mathcal{V}$. Show that $H(f(U)) \leq H(U)$.

(b) Let $U$ and $V$ be two random variables taking values in the alphabets $\mathcal{U}$ and $\mathcal{V}$ respectively, and let $f$ be a mapping from $\mathcal{V}$ to $\mathcal{W}$. Show that $H(U|V) \leq H(U|f(V))$.

PROBLEM 2.

(a) Let $U$ and $\hat{U}$ be two random variables taking values in the same alphabet $\mathcal{U}$, and let $p_e = P[U \neq \hat{U}]$. Show that $H(U|\hat{U}) \leq h(p_e) + p_e \log(|\mathcal{U}| - 1)$, where $h(p) = p \log \frac{1}{p} + (1 - p) \log \frac{1}{1 - p}$.

Hint: use the random variable $W \in \{0, 1\}$ defined by

$$W = \begin{cases} 1 & \text{if } U \neq \hat{U}, \\ 0 & \text{otherwise.} \end{cases}$$

(b) Let $U$ and $V$ be two random variables taking values in the alphabets $\mathcal{U}$ and $\mathcal{V}$ respectively, and let $f$ be a mapping from $\mathcal{V}$ to $\mathcal{U}$. Define $p_e = P[U \neq f(V)]$. Show that $H(U|V) \leq h(p_e) + p_e \log(|\mathcal{U}| - 1)$.

PROBLEM 3. The entropy $H(U)$ of a random variable $U$ is a function of the distribution $p_U$ of the random variable. Denote by $h(p)$ the entropy of a random variable with distribution $p$, i.e., $h(p) = \sum_{u \in \mathcal{U}} \frac{p(u) \log \frac{1}{p(u)}}{p(u)}$. Let $p$ and $q$ be two probability distributions on the same alphabet $\mathcal{U}$, and, for $\theta \in [0, 1]$ let $r$ be the probability distribution on $\mathcal{U}$ defined by

$$r(u) = \theta p(u) + (1 - \theta) q(u)$$

for every $u \in \mathcal{U}$. We are going to show that

$$H(r) \geq \theta H(p) + (1 - \theta) H(q).$$

(a) Let $U_1$ and $U_2$ be random variables with distributions $p$ and $q$ respectively. Let $Z \in \{1, 2\}$ be a binary random variable with $P(Z = 1) = \theta$. Finally define the random variable $U$ as

$$U = \begin{cases} U_1 & \text{if } Z = 1, \\ U_2 & \text{if } Z = 2. \end{cases}$$

What is the distribution of $U$?

(b) Compute $H(U)$ and $H(U|Z)$. What can you conclude?

PROBLEM 4. Consider a source $U$ with alphabet $\mathcal{U}$ and suppose that we know that the true distribution of $U$ is either $P_1$ or $P_2$. Define $S = \sum_{u \in \mathcal{U}} \max\{P_1(u), P_2(u)\}$. 
Recall that we have already seen the non-negativity of this quantity in the class.

Now assume that the true distribution of $P$ is not be independent. For every $i$, the sequence and the size of the type class is therefore defined by the number of 1’s in $x_i$ sequence.

Problem 5. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be $n$ pairs of random variables which may or may not be independent. For every $i \geq 1$ and $j \leq n$, define $X_i^j$ to be the sequence $X_i, \ldots, X_j$ if $i \leq j$, and to be $\emptyset$ if $i > j$. Define $Y_i^j$ similarly. Therefore, since $X_n = Y_1 = \emptyset$ we have $I(X_n; Y_n) = I(Y_1; X_1) = 0$ and $I(Y_1^n; X_n|X_{n+1}^n) = I(Y_1^{n-1}; X_n)$.

(a) Show that $I(Y_1^{n-1}; X_n) = \sum_{i=1}^{n-1} I(X_i; Y_i|Y_1^{i-1})$.

(b) Show that $\sum_{i=1}^{n} I(X_i^{n+1}; Y_i|Y_1^{i-1}) = \sum_{i=1}^{n} I(Y_i^{i-1}; X_i|X_{i+1}^n)$.

Problem 6. Define the type $P_x$ (or empirical probability distribution) of a sequence $x = x_1, \ldots, x_n$ be the relative proportion of occurrences of each symbol of $\mathcal{X}$; i.e., $P_x(a) = \frac{N(a|x)}{n}$ for all $a \in \mathcal{X}$, where $N(a|x)$ is the number of times the symbol $a$ occurs in the sequence $x \in \mathcal{X}^n$.

(a) Show that if $X_1, \ldots, X_n$ are drawn i.i.d. according to $Q(x)$, the probability of $x$ depends only on its type and is given by

$$Q^n(x) = 2^{-(H(P_x)+D(P_x||Q))}.$$ 

Define the type class $T(P)$ as the set of sequences of length $n$ and type $P$: 

$$T(P) = \{x \in \mathcal{X}^n : P_x = P\}.$$ 

For example, if we consider binary alphabet, the type is defined by the number of 1’s in the sequence and the size of the type class is therefore $\binom{n}{k}$.

(b) Show for a binary alphabet that

$$|T(P)| = 2^{nH(P)}.$$ 

We say that $a_n \xrightarrow{p} b_n$, if $\lim_{n \to \infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$.

(c) Use (a) and (b) to show that

$$Q^n(T(P)) = 2^{-nD(P||Q)}.$$ 

Note: $D(P||Q)$ is the informational divergence (or Kullback-Leibler divergence) between two probability distributions $P$ and $Q$ on a common alphabet $\mathcal{X}$ and is defined as

$$D(P||Q) = \sum_{a \in \mathcal{X}} P(a) \log \frac{P(a)}{Q(a)}.$$ 

Recall that we have already seen the non-negativity of this quantity in the class.