Sample problems for the 5th lecture (Mar. 29)

1. Let $X_1, X_2, \ldots, X_n, \ldots$ be a stationary discrete random process, where each $X_i \in \mathcal{X}$.

   (a) Let $\{\mathcal{F}_n\}_{n=1}^{\infty}$ be a sequence of sets with each $\mathcal{F}_n \in \mathcal{X}^n$. If
   $$\lim_{n\to\infty} \Pr[X^n \in \mathcal{F}_n] = 1,$$
   can we use this sequence of sets as the *typical sets* to prove Shannon's source coding theorem? Justify your answer.

   (b) Suppose this sequence of sets satisfies the following two properties:
      i. $P_{X^n}(\mathcal{F}_n) > 1 - \delta$
      ii. $|\mathcal{F}_n| \le 2^{n(H(\mathcal{X})+\delta)}$

   Design a simple typical-set encoding as follows:
   $$\begin{cases} x^n \to \text{binary-index } x^n \text{ by } k_n \text{ bits}, & \text{when } x^n \in \mathcal{F}_n \\ x^n \to \text{all-zero binary codeword of length } k_n, & \text{when } x^n \notin \mathcal{F}_n \end{cases}$$
   where k_n is the number of bits used to index all $x^n \in \mathcal{X}^n$, and is restricted to be only a function of n. Argue that this typical-set encoding can achieve
   $$\limsup_{n\to\infty} \frac{k_n}{n} \le H(\mathcal{X}) + \delta \quad \text{and} \quad P_e < \delta,$$
   where P_e is the probability of decoding error.

   (c) Further suppose that other than the two properties in (c), the typical sets satisfies
   $$(\forall\ x^n \in \mathcal{F}_n)\ P_{X^n}(x^n) \le 2^{-n(H(\mathcal{X})-\delta)}.$$
   Now for an alternative encoder that wishes to use only k_n' bits to encode the binary source stream x^n for each n, where
   $$\frac{k_n'}{n} \le H(\mathcal{X}) - 2\delta,$$
   show that its probability of correct decoding P_c' is upper bounded by $\delta + 2^{-n\delta}$.

   Hint: Let $\mathcal{S}_n$ be the set of source streams x^n that can be correctly decoded; then, its size must be upper bounded by $2^{k_n'} \le 2^{n(H(\mathcal{X})-2\delta)}$.

Solution.

(a) The answer is "not necessarily." For example, if we let $\mathcal{F}_n = \mathcal{X}^n$, then $\Pr[X^n \in \mathcal{F}_n] = 1$ for every n; apparently, such choice cannot be used to prove Shannon's source coding theorem.

In fact, we also additionally need that $|\mathcal{F}_n|$ is close to $2^{nH(\mathcal{X})}$, for which such sequence of sets should exist according to Shannon's source coding theorem, where $H(\mathcal{X})$ is the entropy rate of the source.

(b) The number of bits required for this encoder must be upper-bounded by

$$k_n \le \left\lceil \log_2\left(2^{n(H(\mathcal{X})+\delta)} + 1\right)\right\rceil,$$

which implies

$$\limsup_{n\to\infty} \frac{k_n}{n} \le \limsup_{n\to\infty} \frac{1}{n}\left\lceil \log_2\left(2^{n(H(\mathcal{X})+\delta)} + 1\right)\right\rceil = H(\mathcal{X}) + \delta.$$

Since only the all-zero codeword cannot be recovered back to the original source symbols, the error rate satisfies

$$P_e \le P_{X^n}(\mathcal{F}_n^c) < \delta.$$

(c) The probability of correct block decoding satisfies

$$\begin{aligned}
1 - P_e' &= \sum_{x^n \in \mathcal{S}_n} P_{X^n}(x^n) \\
&= \sum_{x^n \in \mathcal{S}_n \cap \mathcal{F}_n^c} P_{X^n}(x^n) + \sum_{x^n \in \mathcal{S}_n \cap \mathcal{F}_n} P_{X^n}(x^n) \\
&\le P_{X^n}(\mathcal{F}_n^c) + |\mathcal{S}_n \cap \mathcal{F}_n| \cdot \max_{x^n \in \mathcal{F}_n} P_{X^n}(x^n) \\
&< \delta + |\mathcal{S}_n| \cdot \max_{x^n \in \mathcal{F}_n} P_{X^n}(x^n) \\
&< \delta + 2^{n(H(X)-2\delta)} \cdot 2^{-n(H(X)-\delta)} \\
&= \delta + 2^{-n\delta}.
\end{aligned}$$

2. For a stationary source $\{X_n\}_{n=1}^{\infty}$, show that for any integer $n > 1$, the following inequalities are true. Also provide the condition under which the equality holds for each inequality.

(a) $\frac{1}{n}H(X^n) \le \frac{1}{n-1}H(X^{n-1})$

Note: This can be interpreted as for a stationary source, the amount of per-source-symbol information content, according to the three axioms proposed by Shannon, is nonincreasing as the processing block size increases.

(b) $\frac{1}{n}H(X^n) \geq H(X_n|X^{n-1})$

Hint: Use the chain rule for entropy and the fact that

$$H(X_i|X_{i-1},\ldots,X_1) = H(X_n|X_{n-1},\ldots,X_{n-i+1})$$

for every i.

Solution:

(a)

$$\begin{aligned}
&\frac{1}{n-1}H(X^{n-1}) - \frac{1}{n}H(X^n) \\
&= \frac{1}{n-1}(H(X_1) + H(X_2|X_1) + \cdots + H(X_{n-1}|X^{n-2})) \\
&\quad -\frac{1}{n}(H(X_1) + H(X_2|X_1) + \cdots + H(X_n|X^{n-1}) \\
&= \frac{1}{n(n-1)}(H(X_1) + H(X_2|X_1) + H(X_3|X_1,X_2) \\
&\quad + \cdots + H(X_{n-1}|X^{n-2})) - \frac{1}{n}H(X_n|X^{n-1}) \\
&= \frac{1}{n(n-1)}[(H(X_1) - H(X_n|X^{n-1})) + (H(X_2|X_1) - H(X_n|X^{n-1})) \\
&\quad + \cdots + (H(X_{n-1}|X^{n-2}) - H(X^n|X^{n-1}))] \\
&= \frac{1}{n(n-1)}[(H(X_n) - H(X_n|X^{n-1})) + (H(X_n|X_{n-1}) - H(X_n|X^{n-1})) \\
&\quad + \cdots + (H(X_n|X_{n-1},\ldots,X_2) - H(X^n|X^{n-1}))] \\
&\quad \text{because of stationarity;} \\
&\geq 0, \text{ because conditioning never increases entropy,}
\end{aligned}$$

with equality holding iff X_n is independent of $\{X_i\}_{i=1}^{n-1}$.

(b)

$$
\begin{aligned}
&\frac{1}{n}H(X^n) - H(X_n|X^{n-1}) \\
&= \frac{1}{n}(H(X_1) + H(X_2|X_1) + \cdot\cdot + H(X_n|X^{n-1})) - H(X_n|X^{n-1}) \\
&= \frac{1}{n}[(H(X_1) - H(X_n|X^{n-1}) + (H(X_2|X_1) - H(X_n|X^{n-1})) \\
&\quad + \cdots + (H(X_n|X^{n-1}) - H(X_n|X^{n-1})] \\
&= \frac{1}{n}[(H(X_n) - H(X_n|X^{n-1}) + (H(X_n|X_{n-1}) - H(X_n|X^{n-1})) \\
&\quad + \cdots + (H(X_n|X^{n-1}) - H(X_n|X^{n-1})] \\
&\geq 0,
\end{aligned}
$$

with equality holding iff X_n is independent of $\{X_i\}_{i=1}^{n-1}$.

3. For each of the following codes, either prove unique decodability (UD) or give an ambiguous concatenated sequence of codewords:

   (a) $\{1, 0, 00\}$.

   (b) $\{1, 01, 00\}$.

   (c) $\{1, 10, 00\}$.

   (d) $\{1, 10, 01\}$.

   (e) $\{0, 01\}$.

   (f) $\{00, 01, 10, 11\}$.

   Solution.

   (a) The code {1,0,00} is *not* UD since, for example, the sequence 000 can be decoded into the following different ways:

   - 0,0,0
   - 0,00
   - 00,0

   Furthermore, Kraft's inequality is violated.

   (b) The code {1,01,00} is UD since it is a prefix code.

   (c) The code {1,10,00} is also UD since it is a suffix code (by reversing the order of the codewords, the reversed codewords satisfy the prefix condition, hence we can uniquely decode them).

(d) The code {1,10,01} is not UD. For example, the sequence 010 has 2 valid parsings.

(e) The code {0,01} is a suffix code, hence it is UD.

(f) The code {00,01,10,11} is a prefix code, hence it is UD.

4. (a) Give a binary prefix code, of which the longest codeword is 3 and which equates Kraft's inequality.

   (b) Determine the largest code size among all binary prefix codes that satisfy the requirements in (a).

Solution.

(a) Equality of Kraft's inequality requires a saturated binary tree, i.e., $n_1 \cdot 2^2 + n_2 \cdot 2 + n_3 = 2^3$, where $0 \leq n_1 < 2$, $0 \leq n_2 < 4$ and $2 \leq n_3 \leq 8$. Note that n_3 must be an even number. The below table then lists all possible values of (n_1, n_2, n_3) that can fulfill the requirements.

n_1	1	1	0	0	0	0
n_2	1	0	3	2	1	0
n_3	2	4	2	4	6	8
code size	4	5	5	6	7	8

Thus, a quick example is $\{000, 001, 010, 011, 100, 101, 110, 111\}$.

(b) The code size is equal to $n_1 + n_2 + n_3$. Thus the table in the solution of (a) indicates that the largest code size is $2^3 = 8$.

5. Theorems 3.21 and 3.22 and their proofs are important in this course. Please check them carefully.