Solutions to Homework Set #1

1. *Monotonicity of entropy per element:* For a stationary stochastic process $X_1, X_2, \ldots,$ show that

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \leq \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1}.$$

Note that this proves the existence of entropy rate for stationary processes.

Solution: By the chain rule for entropy,

$$\begin{aligned} \frac{H(X_1, X_2, \ldots, X_n)}{n} &= \frac{\sum_{i=1}^{n} H(X_i|X^{i-1})}{n} &(1)\\ &= \frac{H(X_n|X^{n-1}) + \sum_{i=1}^{n-1} H(X_i|X^{i-1})}{n} &(2)\\ &= \frac{H(X_n|X^{n-1}) + H(X_1, X_2, \ldots, X_{n-1})}{n}. &(3) \end{aligned}$$

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 averaging both sides, that,

$$\begin{aligned} H(X_n|X^{n-1}) &\leq \frac{\sum_{i=1}^{n-1} H(X_i|X^{i-1})}{n-1} &(4)\\ &= \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1}. &(5) \end{aligned}$$

Combining (3) and (5) yields,

$$\begin{aligned} \frac{H(X_1, X_2, \ldots, X_n)}{n} &\leq \frac{1}{n}\left[\frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1} + H(X_1, X_2, \ldots, X_{n-1})\right] &(6)\\ &= \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1}. &(7) \end{aligned}$$

2. *Entropy is essentially a lower bound on expected length of any lossless code:*

   (a) Give an example of a lossless code that does not satisfy the Kraft inequality.

   (b) Generalize and show that, in fact, the "Kraft sum" associated with a lossless code may be arbitrarily large (for sufficiently large source alphabet).

Our proof that the entropy lower bounds the expected length of a uniquely decodable code heavily relied on the Kraft inequality. Indeed:

(c) Give an example of a lossless code for a random variable X whose expected code-length is less than $H(X)$.

In the remainder of this problem we will prove that, nevertheless, the entropy lower bounds the expected length of a general lossless code up to a small term (dependent on the size of the alphabet of X). Assume with no loss of generality that X takes values in $\{1, \ldots, N\}$ and that $p_1 \geq p_2 \geq \cdots \geq p_N > 0$, where $p_i = \Pr\{X = i\}$. Let L^{opt} denote the minimum expected code length among all lossless codes.

(d) Letting l_i denote the length of the codeword for i, show that there exists a lossless source code attaining L^{opt} for which $l_i = \lceil \log(i+2)/2 \rceil$. Thus

$$L^{\text{opt}} = \sum_{i=1}^{N} p_i \lceil \log(i+2)/2 \rceil. \tag{8}$$

(e) Prove that

$$H(X) - \sum_{i=1}^{N} p_i \log(i+2)/2 \leq \log\left(\sum_{i=1}^{n} 2/(i+2) \right), \tag{9}$$

with equality if and only if $p_i = 2/(i+2)$ for $1 \leq i \leq N$.

(f) Use previous items to show that

$$L^{\text{opt}} \geq H(X) - \log\left(2 \ln(N+2)\right). \tag{10}$$

Solution:

(a) Consider the alphabet $\mathcal{A} = \{0, 1, 2\}$ and a lossless code $C(0) = 0, C(1) = 1, C(2) = 00$. It's easy to see that $\sum_{x \in \mathcal{A}} 2^{-l(x)} = 5/4 > 1$, violating the Kraft's inequality.

(b) In general, the lossless code, which assigns $(0, 1, 00, 01, 10, 11, 000, \ldots)$, clearly has 2^l codewords for each codeword length l (possibly except the longest codeword length). Assume that the alphabet size $|\mathcal{A}| = 2(2^L - 1)$ for some integer L to make the longest codeword length L has exactly 2^L codewords. Then the Kraft sum is

$$\sum_{x \in \mathcal{A}} 2^{-l(x)} = \sum_{l} \sum_{x : l(x) = l} 2^{-l(x)} = \sum_{l} 2^l \cdot 2^{-l} = L,$$

which can be made arbitrarily large by choosing large enough alphabet.

(c) Consider the example in part (a) with the source with $p(0) = p(1) = p(2) = 1/3$. It is easy to check that $4/3 = El(X) < H(X) = \log(3) = 1.585$.

(d)-(f) Refer to

J. G. Dunham, "Optimal noiseless coding of random variables (Corresp.)," IEEE Trans. Info. Theory, vol. IT-26, p. 345, May 1980.

In fact, a tighter bound can be found in

J. J. Rissanen, "Tight lower bounds for optimum code length (Corresp.)," IEEE Trans. Info. Theory, vol. IT-28, pp. 348 - 349, March 1982.