Sample problems for the 6th lecture (April 12)

1. (a) Prove that Huffman code has the minimum average codeword length among all uniquely decodable codes.

   Hint: Prove by contradiction that if the code $\mathcal{C}'$ for the reduced source $\mathcal{Y}$ in Lemma 33 is optimal, then the code $\mathcal{C}$ for its immediate extended source $\mathcal{X}$ is optimal.

   (b) Find a Huffman code, a Fano code and a Shannon-Fano-Elias code for the source with probabilities $\{0.25, 0.25, 0.25, 0.1, 0.1, 0.25\}$ and compute their average codeword lengths.

   (c) Why doesn't the average codeword length of the optimal Huffman code in (b) equal the source entropy?

Solution.

(a) Suppose $\mathcal{C}'$ in Lemma 3.33 is optimal for the reduced source $\mathcal{Y}$ but the code $\mathcal{C}$ for the extended source $\mathcal{X}$ that satisfies

$$\mathrm{ACL}(\mathcal{C}) = \mathrm{ACL}(\mathcal{C}') + p_{M-1} + p_M$$

is not optimal, where ACL is a shorthand for average codeword length. Then, there must exist an optimal code $\mathcal{D}$ with $\mathrm{ACL}(\mathcal{D}) < \mathrm{ACL}(\mathcal{C})$ and with a_{M-1} and a_M as siblings (according to Lemma 3.32). We then combine a_{M-1} and a_M into one symbol $a_{M-1,M}$ and assign its probability as $p_{M-1}+p_M$, which immediately gives a (prefix) code $\mathcal{D}'$ for the reduced source $\mathcal{Y}$, whose average codeword length satisfies

$$\mathrm{ACL}(\mathcal{D}) = \mathrm{ACL}(\mathcal{D}') + p_{M-1} + p_M.$$

The two equations above indicate $\mathrm{ACL}(\mathcal{D}') < \mathrm{ACL}(\mathcal{C}')$; a contradiction to the optimality of $\mathcal{C}'$ for the reduced source $\mathcal{Y}$ is obtained.

(b) Example 3.34 gives a Huffman code as $00, 01, 10, 110, 1110, 1111$, of which the ACL is 2.4.

A Fano code for this source is $00, 01, 10, 110, 1110, 1111$ (which is the same as the Huffman code).

A Shannon-Fano-Elias code for this source is 010, 011, 101, 10110, 11100, 111110, of which the ACL is 3.55.

(c) The minimum per-source-symbol average codeword length guarantees to achieve the source entropy (rate) when block size n for code design goes to infinity. As the design in (b) is optimal only for block size $n = 1$, there is no guarantee that its ACL is equal to the source entropy.

2. Show that the tree representation of a binary uniquely decodable prefix code with the minimum average codeword length must be "saturated".

   Solution. If any codeword of an optimal prefix code has no sibling, then we can shorten the codeword by removing the last bit. A contradiction to the optimality of the code is obtained.

3. Consider two random variables X and Y with values in finite sets $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let $\bar{l}_X$, $\bar{l}_Y$ and $\bar{l}_{XY}$ denote the average codeword lengths of the optimal (first-order) prefix codes

$$f\colon \mathcal{X} \to \{0,1\}^*,$$
$$g\colon \mathcal{Y} \to \{0,1\}^*$$

   and

$$h\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}^*,$$

   respectively; i.e., $\bar{l}_X = E[l(f(X))]$, $\bar{l}_Y = E[l(g(Y))]$ and $\bar{l}_{XY} = E[l(h(X,Y))]$. Prove that:

   (a) $\bar{l}_X + \bar{l}_Y - \bar{l}_{XY} < I(X;Y) + 2$.
   Hint: An optimal prefix code for a source should satisfies

$$H(X) \le \bar{l}_X < H(X) + 1.$$

   (b) $\bar{l}_{XY} \le \bar{l}_X + \bar{l}_Y$.

   Solution.

   (a) Since the three prefix codes are optimal, the following bounds hold:

$$\begin{aligned} H(X) &\le \bar{l}_X < H(X) + 1 \\ H(Y) &\le \bar{l}_Y < H(Y) + 1 \\ H(X,Y) &\le \bar{l}_{XY} < H(X,Y) + 1. \end{aligned}$$

   Thus

$$\begin{aligned} \bar{l}_X &< H(X) + 1 \\ \bar{l}_Y &< H(Y) + 1 \\ -\bar{l}_{XY} &\le -H(X,Y). \end{aligned}$$

   Hence

$$\begin{aligned} \bar{l}_X + \bar{l}_Y - \bar{l}_{XY} &< H(X) + H(Y) - H(X,Y) + 2 \\ &= I(X;Y) + 2. \end{aligned}$$

(b) Let $h' \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}^*$ be a *concatenated* code obtained as follows: to encode $(x,y) \in \mathcal{X} \times \mathcal{Y}$, simply concatenate $f(x)$ and $g(y)$:

$$h'((x,y)) = (f(x), g(y)).$$

Clearly, h' is a prefix code and

$$l(h(x,y)) = l(f(x)) + l(g(y)).$$

Now since h is the *optimal* prefix code on $\mathcal{X} \times \mathcal{Y}$, then

$$\bar{l}_{XY} \leq E[l(h'(X,Y))].$$

Hence

$$\begin{aligned} \bar{l}_{XY} &= E[l(h(X,Y)] \\ &\leq E[l(h'(X,Y))] \\ &= E[l(f(X)) + l(g(Y))] \\ &= E[l(f(X))] + E[l(g(Y))] \\ &= \bar{l}_X + \bar{l}_Y. \end{aligned}$$

4. Prove Observation 3.38.

Solution. This can be proved by induction. Suppose a Huffman code for a reduced source satisfies the sibling property. Then, the Huffman code for the immediate extended source of this reduced source shall also satisfy the sibling property because the newly added two nodes should have the least probable probabilities among all nodes corresponding to the immediate extended source, and hence will be the last two in the sorted list.