Information & Communication Exercise Sheet #3

University of Amsterdam, Bachelor of Computer Science, January 2015
Lecturer: Christian Schaffner

Out: Tuesday, 12 January 2015
(due: Monday, 19 January 2015, 13:00)

To be solved in Class

1. For the Markov chain $X \leftrightarrow Y \leftrightarrow \hat{X}$, show that $H(X|\hat{X}) \geq H(X|Y)$.
2. [Cover-Thomas 2.32]. We are given the following joint distribution of $X \in \{1,2,3\}$ and $Y \in \{a,b,c\}$:

   $$P_{XY}(1,a) = P_{XY}(2,b) = P_{XY}(3,c) = 1/6$$
   $$P_{XY}(1,b) = P_{XY}(1,c) = P_{XY}(2,a) = P_{XY}(2,c) = P_{XY}(3,a) = P_{XY}(3,b) = 1/12.$$

   Let $\hat{X}(Y)$ be an estimator for X (based on Y) and let $p_e = P(\hat{X} \neq X)$.

   (a) Find an estimator $\hat{X}(Y)$ for which the probability of error p_e is as small as possible.

   (b) Evaluate Fano's inequality for this problem and compare.
3. [Cover-Thomas 5.18] Consider the code $C = \{0, 01\}$. Is it prefix-free? Is it uniquely decodable?

Homework

1. *Bottleneck.* Suppose a Markov chain starts in one of n states, necks down to $k < n$ states, and then fans back to $m > k$ states. Thus $X_1 \to X_2 \to X_3$, i.e.,

   $$P_{X_1X_2X_3}(x_1, x_2, x_3) = P_{X_1}(x_1) \cdot P_{X_2|X_1}(x_2|x_1) \cdot P_{X_3|X_2}(x_3|x_2)$$

   for all $x_1 \in \{1,2,...,n\}$, $x_2 \in \{1,2,...,k\}$, $x_3 \in \{1,2,...,m\}$.

   (a) [4 points] Show that the (unconditional) dependence of X_1 and X_3 is limited by the bottleneck by proving that $I(X_1; X_3) \leq \log k$.

   (b) [1 point] Evaluate $I(X_1; X_3)$ for $k = 1$, and explain why no dependence can survive such a bottleneck.
2. Let A, B, C be random variables such that

   $$I(A; B) = 0\,, \tag{1}$$
   $$I(A; C|B) = I(A; B|C)\,, \tag{2}$$
   $$H(A|BC) = 0\,. \tag{3}$$

   [3 points] Which of the three relations $\leq$, $\geq$, $=$ holds between the quantities $H(A)$ and $H(C)$? Prove your answer.

3. *Kraft's Inequality:* Below, six binary codes are shown for the source symbols $x_1, \ldots, x_4$.

	Code A	Code B	Code C	Code D	Code E	Code F
x_1	00	0	0	0	1	1
x_2	01	10	11	100	01	10
x_3	10	11	100	110	001	100
x_4	11	110	110	111	0001	1000

(a) [2 points] Which codes fulfill the Kraft inequality?

(b) [2 points] Is a code that satisfies this inequality always uniquely decodable?

(c) [2 points] Which codes are prefix-free codes?

(d) [2 points] Which codes are uniquely decodable?

4. [2 points] Consider a random variable X that takes on four values with probabilities $\frac{1}{3}, \frac{1}{3}, \frac{1}{4}, \frac{1}{12}$. Show that there exist two different sets of optimal length for the (binary) Huffman codewords.

5. *Huffman Coding:* Jane, a student, regularly sends a message to her parents via a binary channel. The binary channel is lossless (i.e. error-free), but the per-bit costs are quite high, so she wants to send as few bits as possible. Each time, she selects one message out of a finite set of possible messages and sends it over the channel. There are 7 possible messages:

(a) "Everything is fine"
(b) "I am short on money; please send me some"
(c) "I'll come home this weekend"
(d) "I am ill, please come and pick me up"
(e) "My study is going well, I passed an exam (. . . and send me more money)"
(f) "I have a new boyfriend"
(g) "I have bought new shoes"

Based on counting the types of 100 of her past messages, the empirical probabilities of the different messages are:

m	a	b	c	d	e	f	g
$P_M(m)$	19/100	40/100	12/100	2/100	16/100	4/100	7/100

Jane wants to minimize the average number of bits needed to communicate to her parents (with respect to the empirical probability model above).

(a) [2 points] Design a Huffman code for Jane and draw the binary tree that belongs to it.

(b) [4 points] For a binary source X with $P_X(0) = \frac{1}{8}$ and $P_X(1) = \frac{7}{8}$, design a Huffman code for blocks of $N = 1, 2$ and 3 bits. For each of the three codes, compute the average codeword length and divide it by N, in order to compare it to the optimal length, i.e. the entropy of the source. What do you observe?

(c) [1 point] If you were asked at (b) to design a Huffman code for a block of $N = 100$ bits, what problem would you run into?