Information and Communication Theory: First Mini-Test

October 5, 2022

Name: ______________________________ Number: ________

Duration: 45 minutes. Part I scores: correct answer = 3/2 point; wrong answer = –3/4 points.

Useful facts: $\log_a b = (\log_c b)/(\log_c a)$; $\log_2 3 \simeq 1.585$. Unless indicated otherwise, all logarithms are base-2.

Part I

1. Let $X, Y, Z \in \{1, 2, ..., 6\}$ be three random variables representing the outcome of tossing three independent fair dice; then,

   a) $H(X,Y,Z) < 3 + 3\ \log_2(3)$ bits/symbol; ☐

   b) $H(X,Y,Z) = 3 + 3\ \log_2(3)$ bits/symbol; ■

   c) $H(X,Y,Z) > 3 + 3\ \log_2(3)$ bits/symbol. ☐

   Explanation: X, Y, Z are independent, and each of them has uniform distribution, thus

   $$H(X,Y,Z) = H(X) + H(Y) + H(Z) = 3\log_2 6 = 3(\log_2 3 + \log_2 2) = 3 + 3\ \log_2 3 \text{ bits/symbol}$$

2. Let $X, Y, Z \in \{1, 2, ..., 6\}$ be the three random variables defined in question 1 and $A = X + Y + Z$; then,

   a) $H(A) < H(X,Y,Z)$ bits/symbol; ■

   b) $H(A) = H(X,Y,Z)$ bits/symbol; ☐

   c) $H(A) > H(X,Y,Z)$ bits/symbol. ☐

   Explanation: A is given by a non-injective function of (X,Y,Z) (that is, knowing A is not sufficient to know (X,Y,Z)), thus its entropy is strictly smaller than that of X,Y,Z.

3. Let $X, Y, Z \in \{1, 2, ..., 6\}$ be the three random variables defined in question 1 and A the one defined in question 2. Also, let $B = X + Y$ and $C = X - Y$. Then

   a) $H(A,B,C) < H(X,Y,Z)$ bits/symbol; ☐

   b) $H(A,B,C) = H(X,Y,Z)$ bits/symbol; ■

   c) $H(A,B,C) > H(X,Y,Z)$ bits/symbol. ☐

   Explanation: (A,B,C) is given by an injective function of (X,Y,Z); notice that if we know (A,B,C), we can solve for (X,Y,Z). Consequently, $H(A,B,C) = H(X,Y,Z)$.

4. Let B and C be as defined in question 3. Then

   a) $I(B,C) = 0$ bits/symbol; ☐

   b) $I(B,C) = H(B)$ bits/symbol; ☐

   c) none of the previous answers. ■

   Explanation: B and C are not independent; for example, if $B = 12$, we known that $B = C = 6$, thus $C = 0$. Consequently, $I(B;C) \neq 0$. Also, B is not a deterministic function of C; for example, if $C = 4$, we can have $B = 8$ (if $X = 6$ and $Y = 2$) or $B = 6$ (if $X = 5$ and $Y = 1$). Consequently, $I(B;C) \neq H(B)$.

5. Consider a stationary time-invariant Markov source $X = (X_1, ..., X_t, ...)$, where $X_t \in \{1, 2, 3\}$, with transition matrix

   $$\mathbf{P} = \begin{bmatrix} 0.2 & 0.3 & 0.5 \\ 0.2 & 0.3+\alpha & 0.5-\alpha \\ 0.2 & 0.5-\alpha & 0.3+\alpha \end{bmatrix},$$

   where $\alpha \in [0, 0.5]$ is a parameter. Then,

a) this source is not memoryless for any value of α; ■

b) this source is memoryless for any value of α; □

a) none of the previous answers. □

Explanation: in a memoryless source, all rows of the transition matrix are equal. To have the second elements of rows 1 and 2 equal, we need $\alpha = 0$, but in that case the second element of the third row is 0.5.

6. Consider a stationary time-invariant Markov source $Y = (Y_1, ..., Y_t, ...)$, where $Y_t \in \{1, 2, 3\}$, with transition matrix

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3+\beta & 0.5 & 0.2-\beta \\ 0.2-\beta & 0.5 & 0.3+\beta \end{bmatrix},$$

where $\beta \in [0, 0.2]$ is a parameter. Let $C = H(0.2, 0.3, 0.5)$ be entropy of the probability distribution $(0.2, 0.3, 0.5)$. Then, the conditional entropy rate of this source satisfies

a) $H'(Y) \leq C$, for any value of β; ■

b) $H'(Y) = C$, for any value of β; □

c) $H'(Y) \geq C$, for any value of β; □

Explanation: $H'(Y)$ is the weighted average of the entropies of the three rows of the matrix. Since α is positive, $H(0.2-\beta, 0.3+\beta, 0.5) \leq H(0.2, 0.3, 0.5)$, because increasing β leads to a more concentrated distribution.

7. Consider the source $X \in \{a, b, c, d\}$, with probabilities satisfying $\mathbb{P}[X = a] > \mathbb{P}[X = b] > \mathbb{P}[X = c] > \mathbb{P}[X = d]$. The binary code $\{C(a) = 0,\ C(b) = 10,\ C(c) = 110,\ C(d) = 111\}$

a) is optimal for this source; □

b) may or not be optimal for this source, depending on the values of the probabilities; ■

c) is not optimal for this source. □

Explanation: for example, for $\mathbb{P}[X = a] = 0.27, \mathbb{P}[X = b] = 0.26, \mathbb{P}[X = c] = 0.24$, and $\mathbb{P}[X = d] = 0.23$, a code with all words having two bits is better. But for $\mathbb{P}[X = a] = 0.5, \mathbb{P}[X = b] = 0.25, \mathbb{P}[X = c] = 0.125, \mathbb{P}[X = d] = 0.125$, the given code is optimal, even ideal.

8. Consider a source $Z \in \{1, ..., N\}$, with uniform distribution (all symbols have the same probability). Then,

a) in an optimal binary code for this source, all codewords have the same length; □

b) an optimal binary code for Z has expected length equal to the entropy: $H(Z) = \log N$ bits/symbol; □

c) none of the previous questions. ■

Explanation: for example, for $N = 3$, the optimal binary code has two 2-bit words and one 1-bit word. Also for $N = 1$, the optimal code has $L(C) = (1 + 2 + 2)/3 = 5/3$ bits/symbol, while $H(X) = \log_2 3$ bits/symbol.

Part II

1. Let $A \in \{1,2,3,4,5,6\}$ be the outcome of a fair die and $B \in \{0,1\}$ a binary random variable corresponding to a fair coin toss. Finally, let $X = A + B$. Compute $H(X)$, $H(X|A)$, $H(X|B)$, $H(A|X)$, $H(B|X)$, $I(X;A))$, and $I(X;B)$.

   Solution. Probability distribution of X:

x	1	2	3	4	5	6	7
$f_X(x)$	$f_A(1)f_B(0) = \frac{1}{12}$	$f_A(2)f_B(0) + f_A(1)f_B(1) = \frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$f_A(6)f_B(1) = \frac{1}{12}$

   Entropies:

$$H(X) = -\frac{2}{12}\log_2\frac{1}{12} - \frac{5}{6}\log_2\frac{1}{6} = \log_2 12 - \frac{5}{6} = \frac{7}{6} + \log_2 3 \quad \text{bits/symbol}$$
$$H(X|A) = H(B|A) = H(B) = 1 \text{ bit/symbol}$$
$$H(X|B) = H(A|B) = H(A) = \log_2 6 = 1 + \log_2 3 \text{ bit/symbol}$$
$$H(A|X) = H(X|A) + H(A) - H(X) = 1 + 1 + \log_2 3 - \frac{7}{6} - \log_2 3 = \frac{5}{6} \text{ bit/symbol}$$
$$H(B|X) = H(X|B) + H(B) - H(X) = 1 + \log_2 3 + 1 - \frac{7}{6} - \log_2 3 = \frac{5}{6} \text{ bit/symbol}$$
$$I(X;A) = H(X) - H(X|A) = \frac{7}{6} + \log_2 3 - 1 = \frac{1}{6} + \log_2 3 \text{ bit/symbol}$$
$$I(X;B) = H(X) - H(X|B) = \frac{7}{6} + \log_2 3 - 1 - \log_2 3 = \frac{1}{6} \text{ bit/symbol}$$

2. Consider the source A defined above; obtain the Shannon-Fano code for this source and compute its expected length. Is it optimal? Why?

   Solution. Since the distribution of A is uniform, $f_A(a) = 1/6$, for all $a \in \{1, ..., 6\}$. Since

$$\left\lceil -\log_2\frac{1}{6}\right\rceil = \lceil 1 + \log_2 3 \rceil = \lceil 2.585... \rceil = 3,$$

   the Shannon-Fano code uses 6 words of 3 bits, such that no word is a prefix of another word. A possible code is

x	1	2	3	4	5	6
$C(x)$	000	001	010	011	100	101

   The expected length is obviously $L(C) = 3$ bits/symbol. The code is not optimal, because there are better ones; for example,

x	1	2	3	4	5	6
$C'(x)$	00	01	100	101	110	111

   which has expected length $L(C') = (2+2+3+3+3+3)/6 = 8/3 \simeq 2.67$ bits/symbol.

3. Let $A, B \in \{0,1\}$ be two independent binary random variable corresponding to two tosses of a biased coin with $\mathbb{P}[A = 1] = \mathbb{P}[B = 1] = \alpha$, and $X = A + B$. Consider the following binary code for X: $C(0) = 1$, $C(1) = 00$, $C(2) = 01$. For what values of α is this code optimal for X?

   Solution. An optimal binary code for any source with alphabet of three symbols has two words with two bits and one word with one bit. The code is optimal if the 1-bit word corresponds to the most probable symbol. In this case, the distribution is: $f_X(0) = (1-\alpha)^2$, $f_X(1) = 2\alpha(1-\alpha)$, and $f_X(2) = \alpha^2$. For the given code to be optimal, we need:

$$(1-\alpha)^2 \geq 2\alpha(1-\alpha) \Leftrightarrow (\alpha \leq \frac{1}{3}) \text{ or } (\alpha \geq 1)$$

   and

$$(1-\alpha)^2 \geq \alpha^2 \Leftrightarrow \alpha \leq \frac{1}{2}.$$

   The intersection of the two conditions leads to $\alpha \leq \frac{1}{3}$.