Indian Institute of Technology Kharagpur
Course: MA41024/MA60020 Information and Coding Theory
Spring Semester 2021-22
Time : 45 minutes
Class Test - I

Declaration:

- **Each question carries 2 marks.**
- **NO query will be entertained during the examination.**
- **Once a problem is passed, it will not appear in your screen again and hence if a problem appears in your screen then identify the correct option and then go for the next problem.**

---

1. Let Z_1, Z_2, Z_3 be i.i.d. random variables with sample space $S = \{1, 2\}$ and pmf $f(1) = 1/2 = f(2)$. Consider the random variables $X_1 = Z_1, X_2 = Z_1 + Z_2$ and $X_3 = Z_1 + Z_2 + Z_3$. Then $I(X_1; X_2, X_3) =$ —— bits (up to two decimal digits)

   Ans. 0.5, tolerance: 0.01

   Note that $I(X_3; X_1|X_2) = 0$ since Z_1, Z_2, Z_3 are i.i.d. Then by Chain rule,

   $$I(X_1; X_2, X_3) = I(X_2; X_2) = I(Z_1 + Z_2; Z_1) = H(X_2) - H(Z_2).$$

   Now the pmf of X_2 is $f(1) = 1/2$, $f(0) = f(2) = 1/4$. Then $H(X_2) = 3/2$ and $H(Z_2) = 1$. Thus $I(X_1; X_2, X_3) = 1/2 = 0.5$

2. Consider the alphabet (source letters) $\mathcal{A} = \{a, b, c, d, e, f\}$ with the pmf $p(a) = 0.30$, $p(b) = 0.10$, $p(c) = 0.02$, $p(d) = 0.15$, $p(e) = 0.40$, and $p(f) = 0.03$. Then the average binary codeword length of the Shannon-Fano coding of the source letters is —— bits (up to two decimal digits)

   Ans. 2.1, tolerance: 0.05

   The codewords are given by $a \sim 10$, $b \sim 1110$, $c \sim 11111$, $d \sim 110$, $e \sim 0$, $f \sim 11110$. Consequently the avg codeword length is

   $$(0.4 \times 1) + (0.3 \times 2) + (0.15 \times 3) + (0.1 \times 4) + (0.05 \times 5) = 2.1\,\text{bits}$$

3. Let X, Y be random variables associated with the sample space $\{0, 1\}$, and the joint pmf of (X, Y) is given by

   $$p(0, 1) = p(1, 0) = p(0, 0) = 1/3, p(1, 1) = 0.$$

   Then
   $$H(X|Y = 0) - H(X|Y = 1) = -------\text{bits}$$

   (up to two decimal digits)

   Ans. 1, tolerance: 0.01

Note that

$$\begin{aligned} H(X|Y=0) &= -p_{X|y}(0,0)\log p_{X|y}(0,0) - p_{X|y}(1,0)\log p_{X|y}(1,0) \\ &= -\frac{p(0,0)}{p_Y(0)}\log\frac{p(0,0)}{p_Y(0)} - \frac{p(1,0)}{p_Y(0)}\log\frac{p(1,0)}{p_Y(0)} \\ &= -0.5\log 0.5 - 0.5\log 0.5 = 1. \end{aligned}$$

Similarly $H(X|Y=1)=0$.

4. Let a_1, a_2, a_3, a_4, a_5 be the source letters of a source random variable X with pmf $p(a_k) = 1/5$, $1 \leq k \leq 5$. Suppose we are interested to design a binary prefix code C for the source letters such that the average codeword length of C equals $H(X)$. If l_k denotes the length of codeword of a_k then which of the following is true?

   (a) $l_1 = 2, l_2 = 2, l_3 = 2, l_4 = 3, l_5 = 4$

   (b) $l_1 = 1, l_2 = 2, l_3 = 2, l_4 = 3, l_5 = 4$

   (c) Not possible for both (a) and (b)

   Ans. (c)

   It is possible if $2^{-l_k} = p_k$ (Lec - 11). However, p_ks are same for all k and there exist k_1, k_2 such that $l_{k_1} \neq l_{k_2}$.

5. Suppose a_1, a_2, a_3, a_4, a_5 are source letters of a source random variable X with two pmfs p_1, p_2 given by $p_1(a_1) = 1/2$, $p_1(a_2) = 1/4$, $p_1(a_3) = 1/8$, $p_1(a_4) = 1/16$, $p_1(a_5) = 1/16$; and $p_2(a_1) = 1/2$, $p_2(a_2) = 1/8$, $p_2(a_3) = 1/8$, $p_2(a_4) = 1/8$, $p_2(a_5) = 1/8$. Let C_1 and C_2 be two codes for the source letters corresponding to p_1 and p_2 :

$$\begin{aligned} C_1 &: a_1 = 0, a_2 = 10, a_3 = 110, a_4 = 1110, a_5 = 1111 \\ C_2 &: a_1 = 0, a_2 = 100, a_3 = 101, a_4 = 110, a_5 = 111. \end{aligned}$$

   Then which of the following are true?

   (a) C_1 is optimal and C_2 is not optimal

   (b) C_1 is not optimal and C_2 is optimal

   (c) C_1, C_2 both are not optimal

   (d) None of the above

   Ans. (d)

6. Consider two channels in the following figure. Suppose a source letter x is transmitted through both the channels simultaneously. Let X, Y, Z denote random variables corresponding to source letter x, outputs y and z respectively. Suppose the pmf of X is given by $p(0) = 1/4$, $p(1) = 1/4$ and $p(2) = 1/2$. Then $H(Y,Z) = ------$ bits (up to 1 decimal digits)

   Ans. 2, tolerance: 0.01

   Note that $H(Z) = 1$. If $z = 1$ then $x = 2$, and $y = 0, 1$ with equal probability; if $z = 0$ then $x = 0, 1$ with equal probability and as a consequence, $y = 0, 1$ with equal probability. Thus

$$H(Y|Z) = P_z(0)H(Y|z=0) + P_z(1)H(Y|z=1) = 1$$

   and hence

$$H(Y,Z) = H(Z) + H(Y|Z) = 1 + 1 = 2.$$

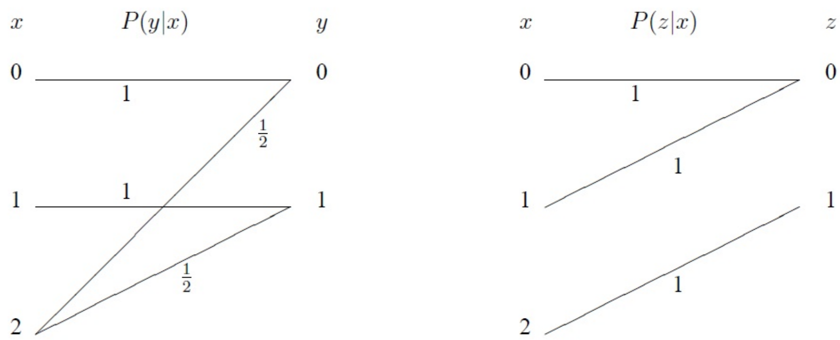

Figure 1: Channels

7. Consider the set of all binary strings of length 4 such that each bit 0 or 1 is drawn independently with equal probability for all the strings. Then the number of all such binary strings whose probability lie in the interval $[0.015625, 0.25]$ can be at most $-------$

   Ans. 64, tolerance: 0.001

   Note that the desired set corresponds to the typical set $A_\epsilon^{(L)}$ where $L = 4$ and $\epsilon = 0.5$, when X is with sample space $\{0, 1\}$ and $p(0) = 1/2 = p(1)$. Hence the desired number can be computed using the bound (Lec 8) of $|A_\epsilon^{(L)}|$.

8. Let $a_k, 1 \le k \le 5$ be source letters corresponding to a source random variable X with pmf $p(a_1) = 0.3$, $p(a_2) = 0.3$, $p(a_3) = 0.2$, $p(a_4) = 0.1$ and $p(a_5) = 0.1$ Then the average codeword length of the binary Huffman code for X is $-----$ bits/symbol. (up to two decimal digits)

   Ans. 2.2, tolerance:0.01

   Note that $a_1 \sim 10$, $a_2 \sim 11$, $a_3 \sim 00$, $a_4 \sim 101$, $a_5 \sim 011$.

9. Suppose X is a source random variable with source letters a_1, a_2. Suppose the source generates 100 source letters at one time and we want to define binary codewords for every such bunch of letters that contain less than or equal to three a_1s. Then the minimum value of the codeword length if we define codewords of constant length, is $-------$ bits

   Ans. 18, tolerance: 0.1

   Note that $\frac{N}{L} \ge \frac{\log K}{\log D}$. Here $L = 1, D = 2$. The value of K is given by

   $$K = \binom{100}{3} + \binom{100}{2} + \binom{100}{1} + \binom{100}{0} = 166751.$$

   Then, $N \ge \lceil \log 166751 \rceil = \lceil \log 17.347 \rceil = 18$

10. Let X be a source random variable with n source letters and pmf is given by $p_1, p_2 \ldots, p_n$. Suppose $l_1, l_2, \ldots, l_n$ are the lengths of codewords of an encoding scheme. Let

    $$L = \sum_{i=1}^{n} p_i l_i.$$

    Suppose $L_1 = \min L$ when the minimum is taken over all prefix codes, and $L_2 = \min L$ when the minimum is taken over all uniquely decodable codes. Then which one of the following is true?

(a) $L_1 < L_2$

(b) $L_2 < L_1$

(c) $L_1 = L_2$

(d) None of the above

Ans. (c)

Note that all prefix codes are uniquely decodable, hence $L_2 \leq L_1$. If a code achieves L_2 should satisfy Kraft inequality and hence we can construct a prefix code with the same codeword lengths, and hence the same L. Thus $L_1 \leq L_2$.