SC222: Tutorial Sheet 11

Problems based on Information Theory.

Pb 1) *Entropy of a disjoint mixture. Let* X_1 *and* X_2 *be discrete random variables drawn according to probability mass functions* $p_1(\cdot)$ *and* $p_2(\cdot)$ *over the respective alphabets* $\chi_1 = \{1, 2, \ldots, m\}$ *and* $\chi_2 = \{m+1, \ldots, n\}$. *Let*

$$X = \begin{cases} X_1 & \textit{with probability } \alpha, \\ X_2 & \textit{with probability } 1-\alpha. \end{cases}$$

a) *Find* $H(X)$ *in terms of* $H(X_1)$, $H(X_2)$, *and* α.

b) *Maximize over* α *to show that* $2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}$ *and interpret using the notion that* $2^{H(X)}$ *is the effective alphabet size.*

Pb 2) Let $p(x, y)$ *be given by*

X \ Y	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

Find:

a) $H(X)$, $H(Y)$.

b) $H(X|Y)$, $H(Y|X)$.

c) $H(X,Y)$.

d) $H(Y) - H(Y|X)$.

e) $I(X; Y)$.

Pb 3) World Series. The World Series is a seven-game series that terminates as soon as either team wins four games. Let X be the random variable that represents the outcome of a World Series between teams A and B; possible values of X are AAAA, BABABAB, and BBBAAAA. Let Y be the number of games played, which ranges from 4 to 7. Assuming that A and B are equally matched and that the games are independent, calculate $H(X)$, $H(Y)$, $H(Y|X)$, *and* $H(X|Y)$.

Pb 4) Run-length coding. Let $X_1, X_2, \ldots, X_n$ *be (possibly dependent) binary random variables. Suppose that one calculates the run lengths* $\mathbf{R} = (R_1, R_2, \ldots)$ *of this sequence (in order as they occur). For example, the sequence* $\mathbf{X}$ = *0001100100 yields run lengths* $\mathbf{R}$ = *(3, 2, 2, 1, 2). Compare* $H(X_1, X_2, \ldots, X_n)$, $H(\mathbf{R})$, *and* $H(X_n, \mathbf{R})$.

Pb 5) Grouping rule for entropy. Let $\mathbf{p} = (p_1, p_2, \ldots, p_m)$ *be a probability distribution on* m *elements (i.e.,* $p_i \geq 0$ *and* $\sum_{i=1}^{m} p_i = 1$*).*

Define a new distribution $\mathbf{q}$ *on m - 1 elements as* $q_1 = p_1$, $q_2 = p_2$, $\ldots$, $q_{m-2} = p_{m-2}$, *and* $q_{m-1} = p_{m-1} + p_m$ *[i.e., the distribution* $\mathbf{q}$ *is the same as* $\mathbf{p}$ *on {1, 2, ..., m - 2}, and the probability of the last element in* $\mathbf{q}$ *is the sum of the last two probabilities of* $\mathbf{p}$*]. Show that*

$$H(\mathbf{p}) = H(\mathbf{q}) + (p_{m-1} + p_m) H\Big(\frac{p_{m-1}}{p_{m-1} + p_m}, \frac{p_m}{p_{m-1} + p_m}\Big)$$

Pb 6) Discrete entropies. Let X and Y be two independent integer valued random variables. Let X be uniformly distributed over {1, 2, ..., 8}, and let Pr{Y = k} $= 2^{-k}$, *k = 1, 2, 3,....*

a) Find H(X).

b) Find H(Y).

Pb 7) Mutual information of heads and tails

a) Consider a fair coin flip. What is the mutual information between the top and bottom sides of the coin?

b) A six-sided fair die is rolled. What is the mutual information between the top side and the front face (the side most facing you)?

Pb 8) Pure randomness. We wish to use a three-sided coin to generate a fair coin toss. Let the coin X have probability mass function

$$X = \begin{cases} A & p_A \\ B & p_B \\ C & p_C \end{cases}$$

where p_A, p_B, p_C *are unknown.*

a) How would you use two independent flips X_1, X_2 *to generate (if possible) a Bernoulli(*$\frac{1}{2}$*) random variable Z?*

b) What is the resulting maximum expected number of fair bits generated?

Pb 9) Codes. Which of the following codes are

(a) Uniquely decodable?

(b) Instantaneous?

$$C_1 = \{00, 01, 0\}$$
$$C_2 = \{00, 01, 100, 101, 11\}$$
$$C_3 = \{0, 10, 110, 1110, \ldots\}$$
$$C_4 = \{0, 00, 000, 0000\}$$