CS 268 Homework Assignment 1 – Solutions

Information Theory

For the some of the problems in this section, you might need the following:
Definition: Let X and Y be discrete random variables, then X and Y are *independent* iff

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

Boole's Law: Let X and Y be discrete random variables, then the *conditional probability of X given Y*, denoted $P(X|Y)$ is:

$$P(X|Y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

1. 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) $I(X; Y)$

   (5 points)

   (a) $H(X) = \sum_x p(x) \log \frac{1}{p(x)} = \frac{1}{3} \log 3 + \frac{2}{3} log \frac{3}{2} = 0.918$ bits. By symmetry, we see $H(Y) = H(X) = 0.918$ bits.

   (b) $H(Y|X) = \sum_x p(x) \sum_y p(y|x) \log \frac{1}{p(y|x)} = \frac{2}{3}(\frac{1}{2} \log 2 + \frac{1}{2} \log 2) + \frac{1}{3}(1 \log 1) = 0.667$ bits. Once again, by symmetry, $H(X|Y) = H(Y|X) = 0.667$ bits.

   (c) $H(X, Y) = \sum_{x,y} p(x, y) \log \frac{1}{p(x,y)} = \frac{1}{3} \log 3 + \frac{1}{3} \log 3 + \frac{1}{3} \log 3 + 0 \log 0 = 1.585$ bits. Alternatively, $H(X, Y) = H(Y) + H(Y|X) = 0.918 + 0.667 = 1.585$ bits.

   (d) $H(Y) - H(Y|X) = 0.918 - 0.667 = 0.251$ bits.

   (e) $I(X; Y) \stackrel{\text{def}}{=} H(Y) - H(Y|X) = 0.251$ bits.

2. 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 the World Series between teams A and B. Some values of X are AAAA, BABABAB, BBBAAAA. Let Y be the number of games played, which ranges from 4 to 7. Assuming that A and B are equally likely to win each game and that the games are independent, calculate $H(X)$, $H(Y)$, $H(Y|X)$, and $H(X|Y)$.

(8 points)
Y takes values in {4,5,6,7} with respective probabilities 1/8, 1/4, 5/16, and 5/16. Thus $H(Y) = 1.92365$ bits.
Conditional on $Y = 4$, X takes 2 values equiprobably. Conditional on $Y = 5$, X takes 8 values equiprobably. Conditional on $Y = 6$, X takes 20 values equiprobably. Conditional on $Y = 7$, X takes 40 values equiprobably. Thus $H(X|Y) = \sum_y H(X|Y = y)p(y) = 3.88885$ bits.
Since Y can be determined from X, we have $H(Y|X) = 0$, and also $H(X) = H(X, Y) = H(Y) + H(X|Y) = 5.8125$ bits.

3. Let X and Y be random variables that take on values $x_1, x_2, \ldots, x_r$ and $y_1, y_2, \ldots, y_s$, respectively. Let $Z = X + Y$. Show that $H(Z|X) = H(Y|X)$. Argue that if X, Y are independent then $H(Y) \leq H(Z)$ and $H(X) \leq H(Z)$. Thus, the addition of *independent* random variables adds uncertainty.

(10 points)
$Z = X + Y$. Hence, $P(Z = z|X = x) = P(Y = z - x|X = x)$.

$$\begin{aligned} H(Z|X) &= \sum p(x)H(Z|X = x) \\ &= -\sum_x p(x) \sum_z P(Z = z|X = x) \log P(Z = z|X = x) \\ &= \sum_x p(x) \sum_y P(Y = z - x|X = x) \log P(Y = z - x|X = x) \\ &= \sum p(x)H(Y|X = x) \\ &= H(Y|X) \end{aligned}$$

If X and Y are independent, then $H(Y|X) = H(Y)$. Since $I(X; Z) \geq 0$, we have $H(X) \geq H(Z|X) = H(Y|X) = H(Y)$. Similarly, we can show that $H(Z) \geq H(X)$.

4. Let $X = (X_1, X_2, \ldots, X_n)$ and $Y = (Y_1, Y_2, \ldots, Y_m)$ be discrete random vectors. Let f and g be real-valued functions of n and m variables respectively. Show that:

$$I(f(X); g(Y)) \leq I(X; Y)$$

(10 points)

$$\begin{aligned} I(X; Y) &= I(X; Y, g(Y)) \\ &= I(X; g(Y)) + I(X; Y|g(Y)) \\ &\geq I(X; g(Y)) \end{aligned}$$

Writing $I(X; g(Y)) = I(g(Y); X)$ and using this again, gives the desired.

5. Consider the random variable

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.49 & 0.26 & 0.12 & 0.04 & 0.04 & 0.03 & 0.02 \end{pmatrix}$$

(a) Find a binary Huffman code for X.

(b) Find the expected codelength for this encoding.

(c) Find a ternary (i.e., $D = \{0, 1, 2\}$) Huffman code for X.

(6 points)

(a) (0, 10, 110, 11100, 11101, 11110, 11111) is a binary Huffman code on the given distribution.

(b) $L(X) = 0.49 + 0.26 \cdot 2 + 0.12 \cdot 3 + 0.04 \cdot 5 + 0.04 \cdot 5 + 0.03 \cdot 5 \cdot 0.02 \cdot 5 = 2.02$ bits.

(c) (0, 1, 20, 21, 220, 221, 222) is a ternary Huffman code on the given distribution.

6. Consider a random variable X which takes on four values with probabilities ($\frac{1}{3}$,$\frac{1}{3}$,$\frac{1}{4}$,$\frac{1}{12}$).

   (a) Construct a Huffman code for this random variable

   (b) Show that here exist two different sets of optimal lengths for the codewords, namely, show that the codeword length assignments (1,2,3,3) and (2,2,2,2) are both optimal.

   (c) Conclude that there are optimal codes with codeword lengths for some symbols that exceed $\lceil \log \frac{1}{p(x)} \rceil$ (called the *Shannon code length).*

(6 points)

(a) Both $C_1 = (0, 10, 110, 111)$ and $C_2 = (00, 01, 10, 11)$ are binary Huffman codes for this distribution.

(b) See (a). By optimality of Huffman codes, both codes above are optimal.

(c) $p(x_3) = \frac{1}{4}$. In $C_1, l(x_3) = 3$. Thus, $\lceil \log \frac{1}{p(x_3)} \rceil = \lceil \log 4 \rceil = 2 < l(x_3)$.

7. Which of these codes cannot be a *Huffman* code for any probability assignment? Why not? If the code can be a Huffman code, give a probability distribution on the input symbol set that would yield the Huffman code.

   (a) (0, 10, 11)

   (b) (00, 01, 10, 110)

   (c) (01, 10)

(6 points)

(a) This is can be a Huffman code on distribution (0.5, 0.25, 0.25).

(b) This cannot be a Huffman code since it violates the property that the two longest codewords differ only in one bit.

(c) Once again, this code cannot be a Huffman code for the same reason as (b). Alternatively, since the uniform distribution gives the largest entropy. The largest entropy $H(X)$ on any probability distribution on two symbols will be one bit. Since a Huffman code satisfies the property $H(X) \leq L(X) < H(X) + 1$, (where L(X) is the average codeword length) we see that this code cannot be a Huffman code since $L(X) = 2 \not< H(X) + 1$.

Slotted ALOHA

8. Given a value for S of $\frac{1}{2e}$, find two possible values for G.

(3 points) We are trying to find solutions for the equation:

$$\frac{1}{2e} = Ge^{-}G$$

We can solve this numerically (or by careful graphical means) to obtain the values for $G \approx 0.24$ and $G \approx 2.68$.

9. Assume that each backlogged node retransmits with some (fixed) probability q_r in each slot until a successful transmission occurs. Thus, the number of slots passed between a collision and a retransmission (i) is described by a random variable. Give a formula for its probability distribution.

(3 Points) This is a geometric random variable having value $i \geq 1$ with probability:

$$q_r(1-q_r)^{i-1}$$

10. Let n be the number of backlogged nodes at a given slot and let q_a be the probability that an unbacklogged node transmits a packet in a given slot. Each of the n nodes will transmit with probability q_r independently of eachother. Let $Q_a(i,n)$ be the probability that i unbacklogged nodes transmit packets in a given slot, and let $Q_r(i,n)$ be the probability that i backlogged nodes transmit. Give formulas for $Q_a(i,n)$ and $Q_r(i,n)$.

(6 Points) We have n as the number of backlogged nodes at the beginning of each slot. Each of these transmits independently of each other with probability q_r. Each of the $m-n$ other nodes will transmit a packet in the given slot if one or more packets arrived during the previous slot. As arrivals are Poisson with mean λ/m, the probability of no arrivals is $e^{-\lambda/m}$, so the probability that an unbacklogged node transmits a packet in the given slot is

$$q_a = 1 - e^{-\lambda/m}$$

Thus, we have:

$$Q_a(i,n) = \left(\begin{array}{c} m-n \\ i \end{array} \right) (1-q_a)^{m-n-i} q_a^i$$

$$Q_r(i,n) = \left(\begin{array}{c} n \\ i \end{array} \right) (1-q_r)^{n-i} q_r^i$$

11. If we consider the state of the system to be the number of backlogged packets, it will increase by the number of new packets introduced by unbacklogged nodes, less one if a packet is transmitted successfully. Successful transmission occurs only in the following two circumstances: (1) one new packet is introduced with no others backlogged or (2) one backlogged packet is introduced with no new packet. Give the transition probabilities $P_{n,n+i}$ of going from state n to state $n+i$. Draw the Markov chain.

(10 Points)

(a) A packet is transmitted successfully only if one one new arrival occurs when there are no backlogged packets or no new arrivals occur and one backlogged packet is transmitted. So, the state transition probability of going from state n to state $n+i$ is given by:

$$P_{n,n+i} = \begin{cases} Q_a(i,n), & 2 \leq i \leq (m-n) \\ Q_a(1,n)[1-Q_r(0,n)], & i = 1 \\ Q_a(1,n)Q_r(0,n) + Q_a(0,n)[1-Q_r(1,n)], & i = 0 \\ Q_a(0,n)Q_r(1,n), & i = -1 \end{cases}$$

(b) The Markov chain is given in Figure 1.

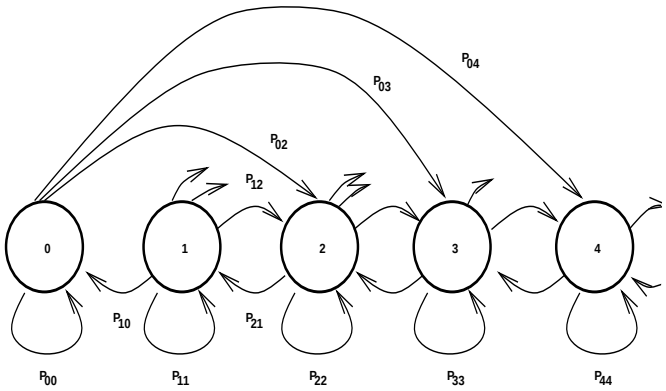

Figure 1: The backlog amount (represented by the states as number of backlogged packets) can decrease by at most one per slot, but may increase by an arbitrary amount.

12. These questions are based on the Markov chain we've just created.

    (a) Verify that the steady state probabilities p_n for the Markov chain are given by the solution to the equations:

$$p_n = \sum_{i=0}^{n+1} p_i P_{in} \qquad\qquad \sum_{n=0}^{m} p_n = 1$$

    (b) For $n < m$, use the last result to express p_{n+1} in terms of $p_0, p_1, \ldots, p_n$.

    (c) Express p_1 in terms of p_0 and then p_2 in terms of p_0.

    (d) For $m = 2$, solve for p_0 in terms of the transition probabilities.

(8 Points)

(a) We are to show the equations properly relate to the Markov chain. The balance equations give us:

$$p_n \sum_{i=n-1}^{m} P_{n,i} = \sum_{i=0}^{n+1} p_i P_{i,n}$$

But, note that:

$$\sum_{i=n-1}^{m} P_{n,i} = Q_a(0,n)Q_r(1,n) +$$

$$Q_a(1,n)Q_r(0,n) + Q_a(0,n)[1 - Q_r(1,n)] + Q_a(1,n)[1 - Q_r(0,n)]$$
$$+Q_a(2,n) + Q_a(3,n) + \ldots$$
$$= \sum_{j=0}^{n} Q_a(j,n) = 1$$

Thus,

$$p_n = \sum_{i=0}^{n+1} p_i P_{i,n}$$

(b)

$$p_n = \sum_{i=0}^{n} p_i P_{i,n} + p_{n+1} P_{n+1,n}$$

$$p_{n+1} = \frac{p_n - \sum_{i=0}^{n} p_i P_{i,n}}{P_{n+1,n}}$$

(c)

$$p_1 = \frac{p_0 - p_0 P_{0,0}}{P_{1,0}} = \frac{p_0(1 - P_{0,0})}{P_{1,0}}$$

$$p_2 = \frac{p_1(1 - P_{1,1}) - p_0 P_{0,1}}{P_{2,1}} = \frac{\frac{p_0(1-P_{0,0})}{P_{1,0}}(1 - P_{1,1}) - p_0 P_{0,1}}{P_{2,1}}$$
$$= \frac{p_0[(1 - P_{0,0})(1 - P_{1,1}) - P_{0,1}P_{1,0}]}{P_{1,0}P_{2,1}}$$

(d)

$$p_0 + p_1 + p_2 = 1$$

$$p_0[1 + \frac{1 - P_{0,0}}{P_{1,0}} + \frac{[(1 - P_{0,0})(1 - P_{1,1}) - P_{0,1}P_{1,0}]}{P_{1,0}P_{2,1}}] = 1$$

$$p_0 = \left(1 + \frac{1 - P_{0,0}}{P_{1,0}} + \frac{[(1 - P_{0,0})(1 - P_{1,1}) - P_{0,1}P_{1,0}]}{P_{1,0}P_{2,1}}\right)^{-1}$$

13. The following questions relate to the instability of Slotted ALOHA.

(a) Define the drift D_n while in state n as the expected change in backlog over one slot time (starting in state n). $D_n = (m-n)q_a - P_{success}$ where:

$$P_{success} = Q_a(1,n)Q_r(0,n) + Q_a(0,n)Q_r(1,n)$$

$G(n) = (m-n)q_a + nq_r$. Assuming q_a and q_r are small, use the approximation $(1-x)^y \approx e^{-xy}$ for small x in the expressions for Q_a and Q_r to show:

$$P_{success} \approx G(n)e^{-G(n)}$$

(b) Plot the departure rate Ge^{-G} and arrival rate $(m-n)q_a$ on the same graph versus G (actually versus n and G, since they are linearly related) for $q_r > q_a$.

(c) The drift D_n represents the difference between the departure and arrival rates while in state n. You should observe three interesting points where the arrival and departure rates are in equilibrium. Label these equilibrium points E_d, E_n, and E_u from left to right, respectively.

(10 Points)

(a)

$$Q_a(0,n) = \begin{pmatrix} m-n \\ 0 \end{pmatrix}(1-q_a)^{m-n} = (1-q_a)^{m-n}$$

$$Q_a(1,n) = \begin{pmatrix} m-n \\ 1 \end{pmatrix}(1-q_a)^{m-n-1}q_a = q_a(m-n)(1-q_a)^{m-n-1}$$

$$Q_r(0,n) = \begin{pmatrix} n \\ 0 \end{pmatrix}(1-q_r)^{n} = (1-q_r)^{n}$$

$$Q_r(1,n) = \begin{pmatrix} n \\ 1 \end{pmatrix}(1-q_r)^{n-1}q_r = nq_r(1-q_r)^{n-1}$$

$$\begin{aligned} P_{success} &\approx q_a(m-n)(1-q_a)^{m-n}(1-q_r)^n + nq_r(1-q_r)^n(1-q_a)^{m-n} \\ &\approx q_a(m-n)e^{-q_a(m-n)}e^{-q_r n} + nq_r e^{-q_r n}e^{-q_a(m-n)} \\ &= (q_a(m-n)+nq_r)(e^{-q_a(m-n)}e^{-q_r n}) = G(n)e^{-G(n)} \end{aligned}$$

(b) The instability of slotted ALOHA is illustrated by Figure 2.

(c) (see Figure 2).

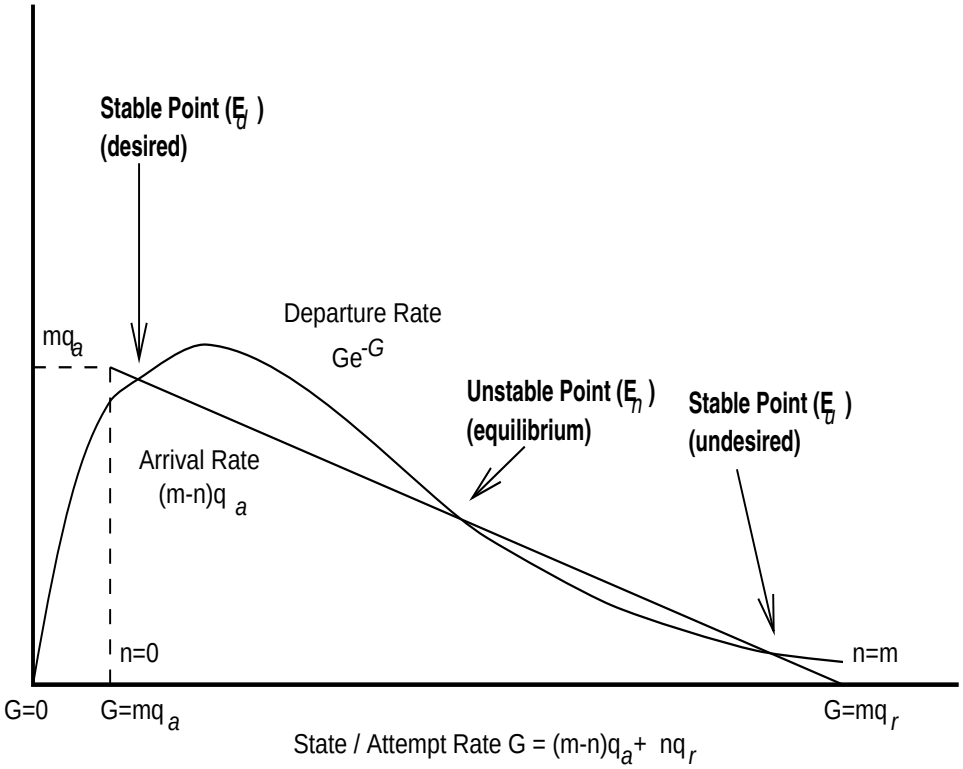

Figure 2: Note that the drift D is negative to the left of the unstable point and positive to the right of the unstable point. Thus, we wish to be operating in the negative drift region to cause the system to tend toward the first (desired) stable point.