16:332:542
Information Theory and Coding
Problem Set 3 Solutions

Chapter 4: 8, 13
Chapter 5: 2, 4, 6, 7

Chapter 4:

8. *Pairwise independence.* Let $X_1, X_2, \ldots, X_{n-1}$ be i.i.d. random variables taking values in $\{0,1\}$, with $\Pr\{X_i = 1\} = \frac{1}{2}$. Let $X_n = 1$ if $\sum_{i=1}^{n-1} X_i$ is odd and $X_n = 0$ otherwise. Let $n \geq 3$.

   (a) Show that X_i and X_j are independent, for $i \neq j, i, j \in \{1, 2, \ldots, n\}$.

   (b) Find $H(X_i, X_j)$, for $i \neq j$.

   (c) Find $H(X_1, X_2, \ldots, X_n)$. Is this equal to $nH(X_1)$?

Solution: (*Pairwise Independence*) $X_1, X_2, \ldots, X_{n-1}$ are i.i.d. Bernoulli(1/2) random variables. We will first prove that for any $k \leq n-1$, the probability that $\sum_{i=1}^{k} X_i$ is odd is $1/2$. We will prove this by induction. Clearly this is true for $k = 1$. Assume that it is true for $k-1$. Let $S_k = \sum_{i=1}^{k} X_i$. Then

$$\begin{aligned} P(S_k \text{ odd}) &= P(S_{k-1} \text{ odd})P(X_k = 0) + P(S_{k-1} \text{ even})P(X_k = 1) && (4.51) \\ &= \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} && (4.52) \\ &= \frac{1}{2}. && (4.53) \end{aligned}$$

Hence for all $k \leq n-1$, the probability that S_k is odd is equal to the probability that it is even. Hence,

$$P(X_n = 1) = P(X_n = 0) = \frac{1}{2}. \qquad (4.54)$$

(a) It is clear that when i and j are both less than n, X_i and X_j are independent. The only possible problem is when $j = n$. Taking $i = 1$ without loss of generality,

$$\begin{aligned} P(X_1 = 1, X_n = 1) &= P(X_1 = 1, \sum_{i=2}^{n-1} X_i \text{ even}) && (4.55) \\ &= P(X_1 = 1)P(\sum_{i=2}^{n-1} X_i \text{ even}) && (4.56) \\ &= \frac{1}{2}\frac{1}{2} && (4.57) \\ &= P(X_1 = 1)P(X_n = 1) && (4.58) \end{aligned}$$

and similarly for other possible values of the pair X_1, X_n. Hence X_1 and X_n are independent.

(b) Since X_i and X_j are independent and uniformly distributed on $\{0,1\}$,

$$H(X_i, X_j) = H(X_i) + H(X_j) = 1 + 1 = 2 \text{ bits}. \qquad (4.59)$$

(c) By the chain rule and the independence of $X_1, X_2, \ldots, X_{n_1}$, we have

$$\begin{aligned} H(X_1, X_2, \ldots, X_n) &= H(X_1, X_2, \ldots, X_{n-1}) + H(X_n | X_{n-1}, \ldots, X_1) && (4.60) \\ &= \sum_{i=1}^{n-1} H(X_i) + 0 && (4.61) \\ &= n - 1, && (4.62) \end{aligned}$$

since X_n is a function of the previous X_i's. The total entropy is not n, which is what would be obtained if the X_i's were all independent. This example illustrates that pairwise independence does not imply complete independence.

since, for $i > 1$, the next position depends only on the previous two (i.e., the dog's walk is 2nd order Markov, if the dog's position is the state). Since $X_0 = 0$ deterministically, $H(X_0) = 0$ and since the first step is equally likely to be positive or negative, $H(X_1|X_0) = 1$. Furthermore for $i > 1$,

$$H(X_i|X_{i-1}, X_{i-2}) \quad = \quad H(.1, .9).$$

Therefore,

$$H(X_0, X_1, \ldots, X_n) = 1 + (n-1)H(.1, .9).$$

(b) From a),

$$\begin{aligned} \frac{H(X_0, X_1, \ldots X_n)}{n+1} &= \frac{1 + (n-1)H(.1, .9)}{n+1} \\ &\rightarrow H(.1, .9). \end{aligned}$$

(c) The dog must take at least one step to establish the direction of travel from whic it ultimately reverses. Letting S be the number of steps taken between reversals we have

$$\begin{aligned} E(S) &= \sum_{s=1}^{\infty} s(.9)^{s-1}(.1) \\ &= 10. \end{aligned}$$

Starting at time 0, the expected number of steps to the first reversal is 11.

13. *Entropy rate of constrained sequences.* In magnetic recording, the mechanism of recording and reading the bits imposes constraints on the sequences of bits that can be recorded. For example, to ensure proper sychronization, it is often necessary to limit the length of runs of 0's between two 1's. Also to reduce intersymbol interference, it may be necessary to require at least one 0 between any two 1's. We will consider a simple example of such a constraint.

Suppose that we are required to have at least one 0 and at most two 0's between any pair of 1's in a sequences. Thus, sequences like 101001 and 0101001 are valid sequences, but 0110010 and 0000101 are not. We wish to calculate the number of valid sequences of length n.

(a) Show that the set of constrained sequences is the same as the set of allowed paths on the following state diagram:

(b) Let $X_i(n)$ be the number of valid paths of length n ending at state i. Argue that

Solution:

Entropy rate of constrained sequences.

(a) The sequences are constrained to have at least one 0 and at most two 0's between two 1's. Let the state of the system be the number of 0's that has been seen since the last 1. Then a sequence that ends in a 1 is in state 1, a sequence that ends in 10 in is state 2, and a sequence that ends in 100 is in state 3. From state 1, it is only possible to go to state 2, since there has to be at least one 0 before the next 1. From state 2, we can go to either state 1 or state 3. From state 3, we have to go to state 1, since there cannot be more than two 0's in a row. Thus we can the state diagram in the problem.

(b) Any valid sequence of length n that ends in a 1 must be formed by taking a valid sequence of length $n-1$ that ends in a 0 and adding a 1 at the end. The number of valid sequences of length $n-1$ that end in a 0 is equal to $X_2(n-1)+X_3(n-1)$ and therefore,

$$X_1(n) = X_2(n-1) + X_3(n-1). \tag{4.79}$$

By similar arguments, we get the other two equations, and we have

$$\begin{bmatrix} X_1(n) \\ X_2(n) \\ X_3(n) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1(n-1) \\ X_2(n-1) \\ X_3(n-1) \end{bmatrix}. \tag{4.80}$$

The initial conditions are obvious, since both sequences of length 1 are valid and therefore $\mathbf{X}(1) = [1\ 1\ 0]^T$.

(c) The induction step is obvious. Now using the eigenvalue decomposition of $A = U^{-1}\Lambda U$, it follows that $A^2 = U^{-1}\Lambda U U^{-1}\Lambda U = U^{-1}\Lambda^2 U$, etc. and therefore

$$\mathbf{X}(n) = A^{n-1}\mathbf{X}(1) = U^{-1}\Lambda^{n-1}U\mathbf{X}(1) \tag{4.81}$$

$$= U^{-1}\begin{bmatrix} \lambda_1^{n-1} & 0 & 0 \\ 0 & \lambda_2^{n-1} & 0 \\ 0 & 0 & \lambda_3^{n-1} \end{bmatrix} U \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \tag{4.82}$$

$$= \lambda_1^{n-1}U^{-1}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} U \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \lambda_2^{n-1}U^{-1}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} U \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$+ \lambda_3^{n-1}U^{-1}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} U \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \tag{4.83}$$

$$= \lambda_1^{n-1}\mathbf{Y}_1 + \lambda_2^{n-1}\mathbf{Y}_2 + \lambda_3^{n-1}\mathbf{Y}_3, \tag{4.84}$$

where $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3$ do not depend on n. Without loss of generality, we can assume that $\lambda_1 > \lambda_2 > \lambda_3$. Thus

$$X_1(n) = \lambda_1^{n-1}\mathbf{Y}_{11} + \lambda_2^{n-1}\mathbf{Y}_{21} + \lambda_3^{n-1}\mathbf{Y}_{31} \tag{4.85}$$

$$X_2(n) = \lambda_1^{n-1}\mathbf{Y}_{12} + \lambda_2^{n-1}\mathbf{Y}_{22} + \lambda_3^{n-1}\mathbf{Y}_{32} \tag{4.86}$$

$$X_3(n) = \lambda_1^{n-1}\mathbf{Y}_{13} + \lambda_2^{n-1}\mathbf{Y}_{23} + \lambda_3^{n-1}\mathbf{Y}_{33} \tag{4.87}$$

For large n, this sum is dominated by the largest term. Thus if $\mathbf{Y}_{1i} > 0$, we have

$$\frac{1}{n}\log X_i(n) \to \log \lambda_1. \tag{4.88}$$

To be rigorous, we must also show that $\mathbf{Y}_{1i} > 0$ for $i = 1,2,3$. It is not difficult to prove that if one of the $\mathbf{Y}_{1i}$ is positive, then the other two terms must also be positive, and therefore either

$$\frac{1}{n}\log X_i(n) \to \log \lambda_1. \tag{4.89}$$

for all $i = 1,2,3$ or they all tend to some other value.

The general argument is difficult since it is possible that the initial conditions of the recursion do not have a component along the eigenvector that corresponds to the maximum eigenvalue and thus $Y_{1i} = 0$ and the above argument will fail. In our example, we can simply compute the various quantities, and thus

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = U^{-1}\Lambda U, \tag{4.90}$$

where

$$\Lambda = \begin{bmatrix} 1.3247 & 0 & 0 \\ 0 & -0.6624 + 0.5623i & 0 \\ 0 & 0 & -0.6624 - 0.5623i \end{bmatrix}, \tag{4.91}$$

and

$$U = \begin{bmatrix} -0.5664 & -0.7503 & -0.4276 \\ 0.6508 - 0.0867i & -0.3823 + 0.4234i & -0.6536 - 0.4087i \\ 0.6508 + 0.0867i & -0.3823i0.4234i & -0.6536 + 0.4087i \end{bmatrix}, \tag{4.92}$$

and therefore

$$\mathbf{Y}_1 = \begin{bmatrix} 0.9566 \\ 0.7221 \\ 0.5451 \end{bmatrix}, \tag{4.93}$$

which has all positive components. Therefore,

$$\frac{1}{n}\log X_i(n) \to \log \lambda_i = \log 1.3247 = 0.4057 \text{ bits.} \tag{4.94}$$

(d) To verify the that

$$\mu = \left[\frac{1}{3-\alpha}, \frac{1}{3-\alpha}, \frac{1-\alpha}{3-\alpha}\right]^T. \tag{4.95}$$

is the stationary distribution, we have to verify that $P\mu = \mu$. But this is straightforward.

rate for a probability transition matrix with the same state diagram. We don't know a reference for a formal proof of this result.

Chapter 5

(b) The last bit in the Huffman code distinguishes between the least likely source symbols. (By the conditions of the problem, all the probabilities are different, and thus the two least likely sequences are uniquely defined.) In this case, the two least likely sequences are $000\ldots00$ and $000\ldots01$, which have probabilities $(1-p_1)(1-p_2)\ldots(1-p_n)$ and $(1-p_1)(1-p_2)\ldots(1-p_{n-1})p_n$ respectively. Thus the last question will ask "Is $X_n = 1$", i.e., "Is the last item defective?".

(c) By the same arguments as in Part (a), an upper bound on the minimum average number of questions is an upper bound on the average length of a Huffman code, namely $H(X_1, X_2, \ldots, X_n) + 1 = \sum H(p_i) + 1$.

---

2. *How many fingers has a Martian?* Let

$$S = \begin{pmatrix} S_1, \ldots, S_m \\ p_1, \ldots, p_m \end{pmatrix}.$$

The S_i's are encoded into strings from a D-symbol output alphabet in a uniquely decodable manner. If $m = 6$ and the codeword lengths are $(l_1, l_2, \ldots, l_6) = (1, 1, 2, 3, 2, 3)$, find a good lower bound on D. You may wish to explain the title of the problem.

Solution: *How many fingers has a Martian?*

Uniquely decodable codes satisfy Kraft's inequality. Therefore

$$f(D) = D^{-1} + D^{-1} + D^{-2} + D^{-3} + D^{-2} + D^{-3} \le 1. \tag{5.4}$$

We have $f(2) = 7/4 > 1$, hence $D > 2$. We have $f(3) = 26/27 < 1$. So a possible value of D is 3. Our counting system is base 10, probably because we have 10 fingers. Perhaps the Martians were using a base 3 representation because they have 3 fingers. (Maybe they are like Maine lobsters ?)

---

4. *Huffman coding.* 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 Huffman code for X.

Solution: *Examples of Huffman codes.*

(a) The Huffman tree for this distribution is

Codeword								
1	x_1	0.49	0.49	0.49	0.49	0.49	0.51	1
00	x_2	0.26	0.26	0.26	0.26	0.26	0.49	
011	x_3	0.12	0.12	0.12	0.13	0.25		
01000	x_4	0.04	0.05	0.08	0.12			
01001	x_5	0.04	0.04	0.05				
01010	x_6	0.03	0.04					
01011	x_7	0.02						

(b) The expected length of the codewords for the binary Huffman code is 2.02 bits. ($H(X) = 2.01$ bits)

(c) The ternary Huffman tree is

Codeword					
0	x_1	0.49	0.49	0.49	1.0
1	x_2	0.26	0.26	0.26	
20	x_3	0.12	0.12	0.25	
22	x_4	0.04	0.09		
210	x_5	0.04	0.04		
211	x_6	0.03			
212	x_7	0.02			

This code has an expected length 1.34 ternary symbols. ($H_3(X) = 1.27$ ternary symbols).

6. *Bad codes.* Which of these codes cannot be Huffman codes for any probability assignment?

   (a) $\{0, 10, 11\}$.

   (b) $\{00, 01, 10, 110\}$.

   (c) $\{01, 10\}$.

Solution: *Bad codes*

   (a) { 0,10,11 } is a Huffman code for the distribution (1/2,1/4,1/4).

   (b) The code { 00,01,10, 110 } can be shortened to { 00,01,10, 11 } without losing its instantaneous property, and therefore is not optimal, so it cannot be a Huffman code. Alternatively, it is not a Huffman code because there is a unique longest codeword.

   (c) The code { 01,10 } can be shortened to { 0,1 } without losing its instantaneous property, and therefore is not optimal and not a Huffman code.

7. *Huffman 20 Questions.* Consider a set of n objects. Let $X_i = 1$ or 0 accordingly as the i-th object is good or defective. Let $X_1, X_2, \ldots, X_n$ be independent with $\Pr\{X_i = 1\} = p_i$; and $p_1 > p_2 > \ldots > p_n > 1/2$. We are asked to determine the set of all defective objects. Any yes-no question you can think of is admissible.

   (a) Give a good lower bound on the minimum average number of questions required.

   (b) If the longest sequence of questions is required by nature's answers to our questions, what (in words) is the last question we should ask? And what two sets are we distinguishing with this question? Assume a compact (minimum average length) sequence of questions.

   (c) Give an upper bound (within 1 question) on the minimum average number of questions required.

Solution: *Huffman 20 Questions.*

   (a) We will be using the questions to determine the sequence $X_1, X_2, \ldots, X_n$, where X_i is 1 or 0 according to whether the i-th object is good or defective. Thus the most likely sequence is all 1's, with a probability of $\prod_{i=1}^n p_i$, and the least likely sequence is the all 0's sequence with probability $\prod_{i=1}^n (1-p_i)$. Since the optimal set of questions corresponds to a Huffman code for the source, a good lower bound on the average number of questions is the entropy of the sequence $X_1, X_2, \ldots, X_n$. But since the X_i's are independent Bernoulli random variables, we have

$$EQ \geq H(X_1, X_2, \ldots, X_n) = \sum H(X_i) = \sum H(p_i). \tag{5.8}$$