

**Date Assigned:** 17 November 2004.

**Date Due:** 1 December 2004 in class.

**Instructions:** I expect you to work on the problems by yourself. You can refer to any textbook (or the technical literature in general), but not confer with any person.

1. *Short Questions*

- (a) How many bits are needed to specify a selection of  $k$  objects from  $n$  objects? ( $n$  and  $k$  are assumed to be known and the selection of  $k$  objects is unordered.)
- (b) Either prove that the following code is uniquely decodable or give an ambiguous concatenated sequence of codewords:

$$\begin{aligned} c_0 &= 101 \\ c_1 &= 0011 \\ c_2 &= 1001 \\ c_3 &= 1110 \\ c_4 &= 00001 \\ c_5 &= 11001 \\ c_6 &= 11100 \\ c_7 &= 010100 \end{aligned}$$

- (c) Consider the memoryless AWGN channel

$$y = x + z$$

where  $z$  is zero mean Gaussian random variable with variance  $\sigma^2$ . The transmit signal  $x$  has an average power constraint of  $P$ . With no other constraints on the input, the capacity of the channel is

$$C = \frac{1}{2} \log_2 \left( 1 + \frac{P}{\sigma^2} \right).$$

Now suppose  $x$  is restricted to the binary alphabet  $\{-\sqrt{P}, +\sqrt{P}\}$ .

- i. Find an expression for the capacity of this restrictive channel and denote it by  $\hat{C}$  (you may not be able to find a closed form expression for  $\hat{C}$  but you should be able to identify the optimal input distribution exactly).
- ii. When is  $\hat{C}$  close to  $C$ ? For small signal-to-noise ratios (defined as the ratio  $P/\sigma^2$ ) or large ones? *Commentary:* This justifies the usual engineering practice of using simple *binary modulation* on the AWGN channel in a certain SNR regime.

2. The frequency  $p_n$  of the  $n$ th most frequent word in English is roughly approximated by

$$p_n \approx \begin{cases} \frac{0.1}{n} & \text{for } 1 \leq n \leq 12367 \\ 0 & n > 12367. \end{cases}$$

(This remarkable  $1/n$  law is known as Zipf's law, and applies to the word frequencies of many languages [4].) If we assume that English is generated by picking words at random according to this distribution, what is the entropy of English (per word)? You might need a computer to help you arrive at the answer.

3. The *Mathematical Games* column of the *Scientific American* featured the following puzzle in 1975.

**The poisoned glass.** ‘Mathematicians are curious birds’, the police commissioner said to his wife. ‘You see, we had all those partly filled glasses lined up in rows on a table in the hotel kitchen. Only one contained poison, and we wanted to know which one before searching that glass for fingerprints. Our lab could test the liquid in each glass, but the tests take time and money, so we wanted to make as few of them as possible by simultaneously testing mixtures of small samples from groups of glasses. The university sent over a mathematics professor to help us. He counted the glasses, smiled and said: ‘ ‘Pick any glass you want, Commissioner. We’ll test it first.’ ‘ ‘But won’t that waste a test?’ I asked. ‘ ‘No,’ he said, ‘it’s part of the best procedure. We can test one glass first. It doesn’t matter which one.’ ‘ ‘How many glasses were there to start with?’ the commissioner’s wife asked. ‘I don’t remember. Somewhere between 100 and 200.’ What was the exact number of glasses?’

- (a) Solve this puzzle.
- (b) Now, explain why the professor was in fact wrong and the commissioner was right. What is the optimal procedure for identifying the one poisoned glass? You will have to make precise the notion of optimality you are considering.
- (c) What is the expected waste relative to this optimum if one followed the professor’s strategy?

*Hint:* How is this problem related to Huffman coding?

#### 4. Graph Entropy

While considering a special kind of communication problem involving data compression with an indistinguishability criterion, Janos Körner introduced a fundamental quantity called *graph entropy*.

A *probabilistic graph*  $(G, \mathbb{P})$  is a graph  $G = (V, E)$  with a probability distribution  $\mathbb{P}$  on its vertices. Let  $\mathcal{A}$  denote the collection of *maximal independent sets* of the graph  $G$ . (Recall that an *independent set* of a graph is a subset of its vertices no pair of which is connected by an edge; a maximal independent set is one that cannot be increased in cardinality by the addition of another vertex. Note that maximal independent sets can be of different cardinalities.)

The graph entropy  $H_G(\mathbb{P})$  of the probabilistic graph  $(G, \mathbb{P})$  is defined as follows: it is the minimum of  $I(X; Y)$  over all pairs of random variables  $(X, Y)$  such that  $X$  takes values in  $V$ , with distribution  $\mathbb{P}$ , and  $Y$  takes values in  $\mathcal{A}$ , and  $X \in Y$  (yes, this is written correctly- it means that conditioned on  $Y = a$ ,  $X$  can only take values among the vertices in the maximally independent set  $a$ ). Equivalently,

$$H_G(\mathbb{P}) = \min_{(X,Y): X \text{ is } V\text{-valued}, Y \text{ is } \mathcal{A}\text{-valued}, X \sim \mathbb{P}, X \in Y} I(X; Y).$$

- (a) Show that, if  $G$  is the complete graph on  $V$ , then  $H_G(\mathbb{P}) = H(\mathbb{P})$ , the usual entropy of the probability distribution  $\mathbb{P}$ .
- (b) What is the graph entropy of the uniform distribution on the vertices of a pentagon?
- (c) Let  $G_1 = (V, E_1)$  and  $G_2 = (V, E_2)$  be two graphs on the same vertex set  $V$ . Let  $E = E_1 \cup E_2$ , let  $G = (V, E)$ , and let  $\mathbb{P}$  be a probability distribution on  $V$ . Show that

$$H_G(\mathbb{P}) \leq H_{G_1}(\mathbb{P}) + H_{G_2}(\mathbb{P}).$$

5. *Source coding with variable-length symbols*

In our discussion of source coding, we considered encoding into a binary alphabet  $\{0, 1\}$  in which both symbols should be used with equal frequency. In this question, we explore how the encoding alphabet should be used if the symbols take different times to transmit. (This is different than putting a cost for certain symbols in the source as studied in the last question of Homework 4 – here we are putting costs on the encoding alphabet.)

A penury-stricken student communicates for free with a friend using a telephone by selecting an integer  $n \in \{1, 2, 3, \dots\}$ , making the friend's phone ring  $n$  times, then hanging up in the middle of the  $n$ th ring. This process is repeated so that a string of symbols  $n_1, n_2, n_3 \dots$  is received. What is the optimal way to communicate? If large integers  $n$  are selected then the message takes longer to communicate. If only small integers  $n$  are used then the information content per symbol is small. We aim to maximize the rate of information transfer, per unit time. Suppose the time taken to transmit a number of rings  $n$  and to redial is  $l_n$  seconds.

- (a) Consider a probability distribution over  $n$ ,  $\mathbf{p} \stackrel{\text{def}}{=} \{p_n\}$ . Define the average duration *per symbol* to be

$$L(\mathbf{p}) = \sum_n p_n l_n.$$

The entropy *per symbol* is defined to be

$$H(\mathbf{p}) = \sum_n p_n \log_2 \frac{1}{p_n}.$$

Show that the largest rate of information transfer in bits per unit time is equal to

$$\sup_{\mathbf{p}} \frac{H(\mathbf{p})}{L(\mathbf{p})}. \quad (1)$$

- (b) Suppose  $l_n = n$ . Solve the optimization problem in (1) explicitly. What is the largest rate of information transfer in this case? Not what one would call mind-numbing speed, but hey, you cannot beat the price.

*Commentary:* Here the information is contained only in the sequence of number of phone rings. But, if we think about it, we can also pack information in the *timing* between the successive phone calls. This means, we choose not to redial instantly but delay it with the purpose of sending information in the duration *between* the rings. But there is typically some (random) lag between when you dial a number and the time the first ring begins at the destination. This is a noisy *timing channel*. The capacity of such a channel is derived in [1]; this work won the *Information Theory Society Best Paper Award* in 1998.

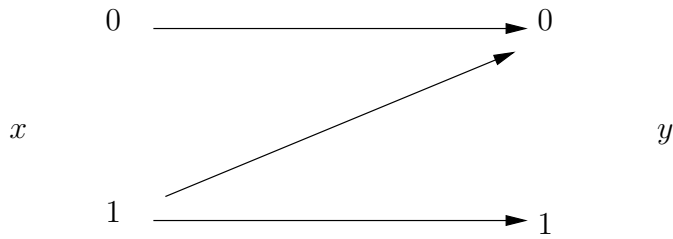


Figure 1: The  $Z$  channel.

### 6. $Z$ Channel

Consider a discrete memoryless channel with binary input and output alphabets, i.e.,  $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ . The channel is described by the following conditional probabilities:

$$\begin{aligned} \mathbb{P}(y = 0|x = 0) &= 1; & \mathbb{P}(y = 0|x = 1) &= q; \\ \mathbb{P}(y = 1|x = 0) &= 0; & \mathbb{P}(y = 1|x = 1) &= 1 - q. \end{aligned}$$

- Show that the capacity of the  $Z$  channel is at least  $0.5(1 - q)$ . *Hint:* The capacity of the *erasure channel* with erasure probability  $q$  is  $1 - q$ . Can you convert two uses of the  $Z$  channel into one over an erasure channel?
- Show that the optimal input distribution  $(p_0^*, p_1^*)$  is given by

$$p_1^* = \frac{1}{(1 - q) \left( 1 + 2^{\frac{H(q)}{1-q}} \right)}, \quad (2)$$

where  $H(q) = -q \log_2 q - (1 - q) \log_2 (1 - q)$  and  $p_0^* = 1 - p_1^*$ .

- What happens to  $p_1^*$  if the noise level  $q$  is very close to 1?
- Argue that  $p_1^*$  is less than 0.5 for all values of  $q$ .
- Why do you think that  $p_1^*$  is always less than 0.5? One could argue that it is good to favor the 0 input, since it is transmitted without error – and also argue that it is good to favor the 1 input, since it often gives rise to the highly prized 1 output, which allows certain identification of the input! Try to make a convincing argument.

## 7. A Channel with Side Information

The channels we have considered in the course are all described stochastically. In many situations of interest, the channel is modeled parametrically and the transmitter or the receiver (or both) may be aware of the parameters. Communication over such channels can now make use of this *side information*. In this problem, we study a very simple model that brings to sharp focus the benefits of side information.

- (a) Consider a 4-card deck with each of the cards individually distinguishable (say numbers 1 through 4 are written on them or they are a subset of the standard 52-card deck). Alice is dealt a 3-card subset from this 4-card deck. Alice can sequentially give 2 (two) of these 3 cards to Bob. How much information can she convey? Let us denote the maximum information she can convey by  $C$  and measure it in bits.

- i. Two cards can be arranged in one of two possible ways, so argue that  $C \geq 1$  bit.
- ii. If the three cards are also shown to Bob at the same time as Alice is dealt these cards, then Alice can pick a 2-card subset in 3 possible ways (and Bob knows which way she picked that subset) and arrange this 2-card subset in two possible ways giving rise to a total of 6 possible ways. Argue that this means that  $C \leq \log_2(6)$  bits.
- iii. Can you find a scheme by which Alice could still actually communicate information at rate  $\log_2(6)$  bits even when Bob has no idea which 3-card subset she is dealt with? *Hint*: Think of the Slepian-Wolf binning scheme.

*Commentary*: This is an example of a channel with side information. Suppose the cards are numbered 1 through 4. The side information  $S$  is a 3-card subset. The channel input  $X$  is forced to be a 2-card subset of the side information  $S$ . The channel output  $Y$  is simply equal to  $X$ . We have just shown the stunning result that the capacity of the channel is the same whether the side information is known to the transmitter alone (Alice) or to both the transmitter and the receiver (Bob).

- (b) Let us now generalize the previous question. Consider an  $n$ -card deck from which Alice is dealt an  $\ell$ -card subset. Alice can then sequentially send Bob a  $k$ -card subset of the  $\ell$  cards she is dealt. Denote the capacity of this channel by  $C$ , measured in bits.

- i. Show that  $C \geq \log_2 k$ .
- ii. Show that  $C \leq \log_2 \ell$ .
- iii. *Bonus*:<sup>1</sup> Calculate  $C$  exactly.

*Commentary*: Side information problems show up in the least expected of situations. The classical problem which motivated this area of study is the engineering problem of “computer read-only memory with faults”. Here, a computer memory (of size  $n$  bits) has some bit locations that are “struck”, i.e., they are fixed to

---

<sup>1</sup>Solutions to the Bonus question will be used to differentiate between the letter grades  $A$  and  $A+$ .

be either 0 or 1 and thus information cannot be written on them. Not only do these faults waste memory (if there are a total of  $k$  struck-at faults, then the useful memory size is only  $n - k$  bits), they cause trouble when trying to read the information: the device reading the bits from the memory has no idea how to differentiate the useful memory locations from the faulty ones. *However*, the struck-at locations are known ahead of time to the device writing the information. It is possible that one can be smart about writing the information bits in the available  $n - k$  locations. A remarkable result in [3] shows that one can be *very smart* in writing (and reading) the information bits: the capacity of the memory is  $n - k$  bits even though the memory reading device has no idea of the  $k$  faulty bit positions.

- (c) A continuous alphabet version of this problem with AWGN also admits a similarly striking result [2]: Consider the memoryless channel

$$y[m] = x[m] + s[m] + z[m]$$

where the additive noise  $z[m]$  is zero mean Gaussian random variable with variance  $\sigma^2$ . The additive *interference*  $s[m]$  is known non-causally to the transmitter but unknown to the receiver (the receiver only knows that statistically  $s[m]$  is generated by an i.i.d. zero mean Gaussian random sequence with variance  $\sigma_s^2$ ). As usual, there is a transmit power constraint of  $P$ . Let us denote the capacity of this channel by  $C$  measured in bits per channel-use.

- i. Show that  $C \geq \frac{1}{2} \log_2 \left( 1 + \frac{P}{\sigma_s^2 + \sigma^2} \right)$ .
- ii. Show that  $C \leq \frac{1}{2} \log_2 \left( 1 + \frac{P}{\sigma^2} \right)$ .
- iii. *Reading exercise:*<sup>2</sup> The capacity of the channel is actually equal to the upper bound above; i.e., the capacity of the channel is unchanged even if the receiver is also made aware of the entire interference sequence.

## References

- [1] V. Anantharam and S. Verdú, “Bits through Queues”, *IEEE Transactions on Information Theory*, Vol. 42, No. 1, pp. 4-18, January 1996.
- [2] M. H. M. Costa, “Writing on dirty paper”, *IEEE Transactions on Information Theory*, Vol.29, pp. 439-441, May 1983.
- [3] A. El Gamal and C. Heegard, “On the Capacity of Computer Memory with Defects,” *IEEE Transactions on Information Theory*, Vol. 29, No. 5, pp. 731-739, September 1983.
- [4] G. K. Zipf, *Human Behavior and the Principle of Least Effort*, Addison-Wesley, 1949.

---

<sup>2</sup>No need to turn this in.