15

Further Exercises on Information Theory

The most exciting exercises, which will introduce you to further ideas in information theory, are towards the end of this chapter.

Refresher exercises on source coding and noisy channels

▷ Exercise 15.1.[2] Let X be an ensemble with $\mathcal{A}_X = \{0, 1\}$ and $\mathcal{P}_X = \{0.995, 0.005\}$. Consider source coding using the block coding of X^{100} where every $\mathbf{x} \in X^{100}$ containing 3 or fewer 1s is assigned a distinct codeword, while the other $\mathbf{x}$s are ignored.

(a) If the assigned codewords are all of the same length, find the minimum length required to provide the above set with distinct codewords.

(b) Calculate the probability of getting an $\mathbf{x}$ that will be ignored.

▷ Exercise 15.2.[2] Let X be an ensemble with $\mathcal{P}_X = \{0.1, 0.2, 0.3, 0.4\}$. The ensemble is encoded using the symbol code $\mathcal{C} = \{0001, 001, 01, 1\}$. Consider the codeword corresponding to $\mathbf{x} \in X^N$, where N is large.

(a) Compute the entropy of the fourth bit of transmission.

(b) Compute the conditional entropy of the fourth bit given the third bit.

(c) Estimate the entropy of the hundredth bit.

(d) Estimate the conditional entropy of the hundredth bit given the ninety-ninth bit.

Exercise 15.3.[2] Two fair dice are rolled by Alice and the sum is recorded. Bob's task is to ask a sequence of questions with yes/no answers to find out this number. Devise in detail a strategy that achieves the minimum possible average number of questions.

▷ Exercise 15.4.[2] How can you use a coin to draw straws among 3 people?

▷ Exercise 15.5.[2] In a magic trick, there are three participants: the magician, an assistant, and a volunteer. The assistant, who claims to have paranormal abilities, is in a soundproof room. The magician gives the volunteer six blank cards, five white and one blue. The volunteer writes a different integer from 1 to 100 on each card, as the magician is watching. The volunteer keeps the blue card. The magician arranges the five white cards in some order and passes them to the assistant. The assistant then announces the number on the blue card.

How does the trick work?

▷ Exercise 15.6.[3] How does *this* trick work?

> 'Here's an ordinary pack of cards, shuffled into random order. Please choose five cards from the pack, any that you wish. Don't let me see their faces. No, don't give them to me: pass them to my assistant Esmerelda. She can look at them.
>
> 'Now, Esmerelda, show me four of the cards. Hmm... nine of spades, six of clubs, four of hearts, ten of diamonds. The hidden card, then, must be the queen of spades!'

The trick can be performed as described above for a pack of 52 cards. Use information theory to give an upper bound on the number of cards for which the trick can be performed.

▷ Exercise 15.7.[2] Find a probability sequence $\mathbf{p} = (p_1, p_2, \ldots)$ such that $H(\mathbf{p}) = \infty$.

▷ Exercise 15.8.[2] Consider a discrete memoryless source with $\mathcal{A}_X = \{a, b, c, d\}$ and $\mathcal{P}_X = \{1/2, 1/4, 1/8, 1/8\}$. There are $4^8 = 65\,536$ eight-letter words that can be formed from the four letters. Find the total number of such words that are in the typical set $T_{N\beta}$ (equation 4.29) where $N = 8$ and $\beta = 0.1$.

▷ Exercise 15.9.[2] Consider the source $\mathcal{A}_S = \{a, b, c, d, e\}$, $\mathcal{P}_S = \{1/3, 1/3, 1/9, 1/9, 1/9\}$ and the channel whose transition probability matrix is

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 2/3 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1/3 & 0 \end{bmatrix}. \tag{15.1}$$

Note that the source alphabet has five symbols, but the channel alphabet $\mathcal{A}_X = \mathcal{A}_Y = \{0, 1, 2, 3\}$ has only four. Assume that the source produces symbols at exactly 3/4 the rate that the channel accepts channel symbols. For a given (tiny) $\epsilon > 0$, explain how you would design a system for communicating the source's output over the channel with an average error probability per source symbol less than ϵ. Be as explicit as possible. In particular, *do not* invoke Shannon's noisy-channel coding theorem.

▷ Exercise 15.10.[2] Consider a binary symmetric channel and a code $C = \{0000, 0011, 1100, 1111\}$; assume that the four codewords are used with probabilities $\{1/2, 1/8, 1/8, 1/4\}$.

What is the decoding rule that minimizes the probability of decoding error? [The optimal decoding rule depends on the noise level f of the binary symmetric channel. Give the decoding rule for each range of values of f, for f between 0 and 1/2.]

Exercise 15.11.[2] Find the capacity and optimal input distribution for the three-input, three-output channel whose transition probabilities are:

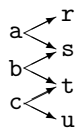
$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{bmatrix}. \tag{15.2}$$

Exercise 15.12.[3, p.239] The input to a channel Q is a word of 8 bits. The output is also a word of 8 bits. Each time it is used, the channel flips *exactly one* of the transmitted bits, but the receiver does not know which one. The other seven bits are received without error. All 8 bits are equally likely to be the one that is flipped. Derive the capacity of this channel.

Show, by describing an *explicit* encoder and decoder that it is possible *reliably* (that is, with *zero* error probability) to communicate 5 bits per cycle over this channel.

▷ Exercise 15.13.[2] A channel with input $x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ and output $y \in \{\mathtt{r}, \mathtt{s}, \mathtt{t}, \mathtt{u}\}$ has conditional probability matrix:

$$\mathbf{Q} = \begin{bmatrix} 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 \end{bmatrix}.$$



What is its capacity?

▷ Exercise 15.14.[3] The ten-digit number on the cover of a book known as the ISBN incorporates an error-detecting code. The number consists of nine source digits $x_1, x_2, \ldots, x_9$, satisfying $x_n \in \{0, 1, \ldots, 9\}$, and a tenth check digit whose value is given by

$$x_{10} = \left(\sum_{n=1}^{9} n x_n \right) \bmod 11.$$

Here $x_{10} \in \{0, 1, \ldots, 9, 10\}$. If $x_{10} = 10$ then the tenth digit is shown using the roman numeral X.

Show that a valid ISBN satisfies:

$$\left(\sum_{n=1}^{10} n x_n \right) \bmod 11 = 0.$$

Imagine that an ISBN is communicated over an unreliable human channel which sometimes *modifies* digits and sometimes *reorders* digits.

Show that this code can be used to detect (but not correct) all errors in which any one of the ten digits is modified (for example, 1-010-00000-4 → 1-010-00080-4).

Show that this code can be used to detect all errors in which any two adjacent digits are transposed (for example, 1-010-00000-4 → 1-100-00000-4).

What other transpositions of pairs of *non-adjacent* digits can be detected?

If the tenth digit were defined to be

$$x_{10} = \left(\sum_{n=1}^{9} n x_n \right) \bmod 10,$$

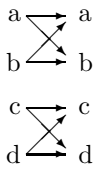
why would the code not work so well? (Discuss the detection of both modifications of single digits and transpositions of digits.)

0-521-64298-1
1-010-00000-4

Table 15.1. Some valid ISBNs. [The hyphens are included for legibility.]

Exercise 15.15.[3] A channel with input x and output y has transition probability matrix:

$$Q = \begin{bmatrix} 1-f & f & 0 & 0 \\ f & 1-f & 0 & 0 \\ 0 & 0 & 1-g & g \\ 0 & 0 & g & 1-g \end{bmatrix}.$$



Assuming an input distribution of the form

$$\mathcal{P}_X = \left\{ \frac{p}{2}, \frac{p}{2}, \frac{1-p}{2}, \frac{1-p}{2} \right\},$$

write down the entropy of the output, $H(Y)$, and the conditional entropy of the output given the input, $H(Y|X)$.

Show that the optimal input distribution is given by

$$p = \frac{1}{1 + 2^{-H_2(g)+H_2(f)}},$$

where $H_2(f) = f \log_2 \frac{1}{f} + (1-f) \log_2 \frac{1}{(1-f)}$.

Remember $\frac{\mathrm{d}}{\mathrm{d}p} H_2(p) = \log_2 \frac{1-p}{p}$.

Write down the optimal input distribution and the capacity of the channel in the case $f = 1/2$, $g = 0$, and comment on your answer.

▷ Exercise 15.16.[2] What are the differences in the redundancies needed in an error-detecting code (which can reliably detect that a block of data has been corrupted) and an error-correcting code (which can detect and correct errors)?

Further tales from information theory

The following exercises give you the chance to discover for yourself the answers to some more surprising results of information theory.

Exercise 15.17.[3] Communication of information from correlated sources. Imagine that we want to communicate data from two data sources $X^{(A)}$ and $X^{(B)}$ to a central location C via noise-free one-way communication channels (figure 15.2a). The signals $x^{(A)}$ and $x^{(B)}$ are strongly dependent, so their joint information content is only a little greater than the marginal information content of either of them. For example, C is a weather collator who wishes to receive a string of reports saying whether it is raining in Allerton ($x^{(A)}$) and whether it is raining in Bognor ($x^{(B)}$). The joint probability of $x^{(A)}$ and $x^{(B)}$ might be

$P(x^{(A)}, x^{(B)})$:

		$x^{(A)}=0$	$x^{(A)}=1$
$x^{(B)}$	0	0.49	0.01
	1	0.01	0.49

(15.3)

The weather collator would like to know N successive values of $x^{(A)}$ and $x^{(B)}$ exactly, but, since he has to pay for every bit of information he receives, he is interested in the possibility of avoiding buying N bits from source A *and* N bits from source B. Assuming that variables $x^{(A)}$ and $x^{(B)}$ are generated repeatedly from this distribution, can they be encoded at rates R_A and R_B in such a way that C can reconstruct all the variables, with the sum of information transmission rates on the two lines being less than two bits per cycle?

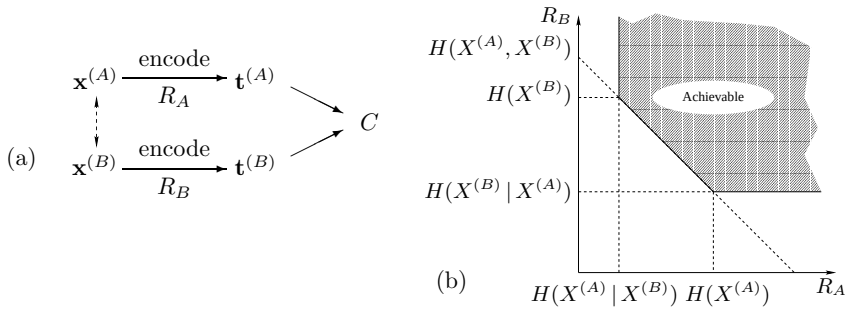


Figure 15.2. Communication of information from dependent sources. (a) $x^{(A)}$ and $x^{(B)}$ are dependent sources (the dependence is represented by the dotted arrow). Strings of values of each variable are encoded using codes of rate R_A and R_B into transmissions $\mathbf{t}^{(A)}$ and $\mathbf{t}^{(B)}$, which are communicated over noise-free channels to a receiver C. (b) The achievable rate region. Both strings can be conveyed without error even though $R_A < H(X^{(A)})$ and $R_B < H(X^{(B)})$.

The answer, which you should demonstrate, is indicated in figure 15.2. In the general case of two dependent sources $X^{(A)}$ and $X^{(B)}$, there exist codes for the two transmitters that can achieve reliable communication of both $X^{(A)}$ and $X^{(B)}$ to C, as long as: the information rate from $X^{(A)}$, R_A, exceeds $H(X^{(A)} \mid X^{(B)})$; the information rate from $X^{(B)}$, R_B, exceeds $H(X^{(B)} \mid X^{(A)})$; and the total information rate $R_A + R_B$ exceeds the joint entropy $H(X^{(A)}, X^{(B)})$ (Slepian and Wolf, 1973).

So in the case of $x^{(A)}$ and $x^{(B)}$ above, each transmitter must transmit at a rate greater than $H_2(0.02) = 0.14$ bits, and the total rate $R_A + R_B$ must be greater than 1.14 bits, for example $R_A = 0.6$, $R_B = 0.6$. There exist codes that can achieve these rates. Your task is to figure out why this is so.

Try to find an explicit solution in which one of the sources is sent as plain text, $\mathbf{t}^{(B)} = \mathbf{x}^{(B)}$, and the other is encoded.

Exercise 15.18.[3] Multiple access channels. Consider a channel with two sets of inputs and one output – for example, a shared telephone line (figure 15.3a). A simple model system has two binary inputs $x^{(A)}$ and $x^{(B)}$ and a ternary output y equal to the arithmetic sum of the two inputs, that's 0, 1 or 2. There is no noise. Users A and B cannot communicate with each other, and they cannot hear the output of the channel. If the output is a 0, the receiver can be certain that both inputs were set to 0; and if the output is a 2, the receiver can be certain that both inputs were set to 1. But if the output is 1, then it could be that the input state was $(0, 1)$ or $(1, 0)$. How should users A and B use this channel so that their messages can be deduced from the received signals? How fast can A and B communicate?

Clearly the total information rate from A and B to the receiver cannot be two bits. On the other hand it is easy to achieve a total information rate $R_A + R_B$ of one bit. Can reliable communication be achieved at rates (R_A, R_B) such that $R_A + R_B > 1$?

The answer is indicated in figure 15.3.

Some practical codes for multi-user channels are presented in Ratzer and MacKay (2003).

Exercise 15.19.[3] Broadcast channels. A broadcast channel consists of a single transmitter and two or more receivers. The properties of the channel are defined by a conditional distribution $Q(y^{(A)}, y^{(B)} \mid x)$. (We'll assume the channel is memoryless.) The task is to add an encoder and two decoders to enable reliable communication of a common message at rate R_0 to both receivers, an individual message at rate R_A to receiver A, and an individual message at rate R_B to receiver B. The *capacity* region of the broadcast channel is the convex hull of the set of achievable rate triplets (R_0, R_A, R_B).

A simple benchmark for such a channel is given by time-sharing (time-division signaling). If the capacities of the two channels, considered separately,

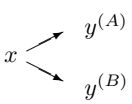


Figure 15.4. The broadcast channel. x is the channel input; $y^{(A)}$ and $y^{(B)}$ are the outputs.

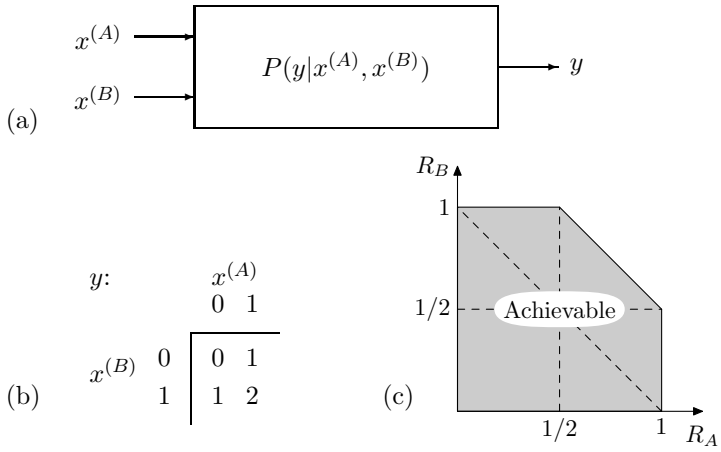


Figure 15.3. Multiple access channels. (a) A general multiple access channel with two transmitters and one receiver. (b) A binary multiple access channel with output equal to the sum of two inputs. (c) The achievable region.

are $C^{(A)}$ and $C^{(B)}$, then by devoting a fraction ϕ_A of the transmission time to channel A and $\phi_B = 1 - \phi_A$ to channel B, we can achieve $(R_0, R_A, R_B) = (0, \phi_A C^{(A)}, \phi_B C^{(B)})$.

We can do better than this, however. As an analogy, imagine speaking simultaneously to an American and a Belarusian; you are fluent in American and in Belarusian, but neither of your two receivers understands the other's language. If each receiver can distinguish whether a word is in their own language or not, then an extra binary file can be conveyed to both recipients by using its bits to decide whether the next transmitted word should be from the American source text or from the Belarusian source text. Each recipient can concatenate the words that they understand in order to receive their personal message, and can also recover the binary string.

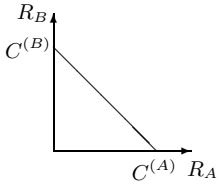


Figure 15.5. Rates achievable by simple timesharing.

An example of a broadcast channel consists of two binary symmetric channels with a common input. The two halves of the channel have flip probabilities f_A and f_B. We'll assume that A has the better half-channel, i.e., $f_A < f_B < 1/2$. [A closely related channel is a 'degraded' broadcast channel, in which the conditional probabilities are such that the random variables have the structure of a Markov chain,

$$x \rightarrow y^{(A)} \rightarrow y^{(B)}, \tag{15.4}$$

i.e., $y^{(B)}$ is a further degraded version of $y^{(A)}$.] In this special case, it turns out that whatever information is getting through to receiver B can also be recovered by receiver A. So there is no point distinguishing between R_0 and R_B: the task is to find the capacity region for the rate pair (R_0, R_A), where R_0 is the rate of information reaching both A and B, and R_A is the rate of the extra information reaching A.

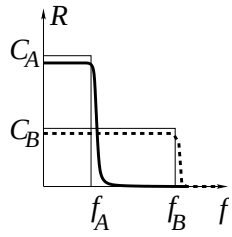


Figure 15.6. Rate of reliable communication R, as a function of noise level f, for Shannonesque codes designed to operate at noise levels f_A (solid line) and f_B (dashed line).

The following exercise is equivalent to this one, and a solution to it is illustrated in figure 15.8.

Exercise 15.20.[3] Variable-rate error-correcting codes for channels with unknown noise level. In real life, channels may sometimes not be well characterized before the encoder is installed. As a model of this situation, imagine that a channel is known to be a binary symmetric channel with noise level either f_A or f_B. Let $f_B > f_A$, and let the two capacities be C_A and C_B.

Those who like to live dangerously might install a system designed for noise level f_A with rate $R_A \simeq C_A$; in the event that the noise level turns out to be f_B, our experience of Shannon's theories would lead us to expect that there

would be a catastrophic failure to communicate information reliably (solid line in figure 15.6).

A conservative approach would design the encoding system for the worst-case scenario, installing a code with rate $R_B \simeq C_B$ (dashed line in figure 15.6). In the event that the lower noise level, f_A, holds true, the managers would have a feeling of regret because of the wasted capacity difference $C_A - R_B$.

Is it possible to create a system that not only transmits reliably at some rate R_0 whatever the noise level, but also communicates some extra, 'lower-priority' bits if the noise level is low, as shown in figure 15.7? This code communicates the high-priority bits reliably at all noise levels between f_A and f_B, and communicates the low-priority bits also if the noise level is f_A or below.

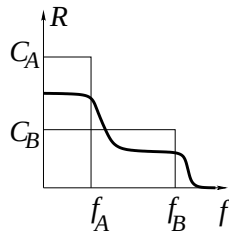


Figure 15.7. Rate of reliable communication R, as a function of noise level f, for a desired *variable-rate* code.

This problem is mathematically equivalent to the previous problem, the degraded broadcast channel. The lower rate of communication was there called R_0, and the rate at which the low-priority bits are communicated if the noise level is low was called R_A.

An illustrative answer is shown in figure 15.8, for the case $f_A = 0.01$ and $f_B = 0.1$. (This figure also shows the achievable region for a broadcast channel whose two half-channels have noise levels $f_A = 0.01$ and $f_B = 0.1$.) I admit I find the gap between the simple time-sharing solution and the cunning solution disappointingly small.

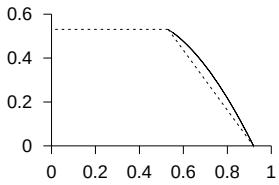


Figure 15.8. An achievable region for the channel with unknown noise level. Assuming the two possible noise levels are $f_A = 0.01$ and $f_B = 0.1$, the dashed lines show the rates R_A, R_B that are achievable using a simple time-sharing approach, and the solid line shows rates achievable using a more cunning approach.

In Chapter 50 we will discuss codes for a special class of broadcast channels, namely erasure channels, where every symbol is either received without error or erased. These codes have the nice property that they are *rateless* – the number of symbols transmitted is determined on the fly such that reliable comunication is achieved, whatever the erasure statistics of the channel.

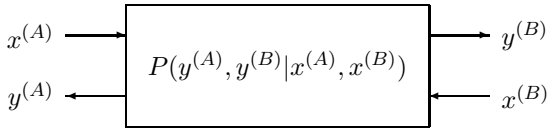
Exercise 15.21.[3] Multiterminal information networks are both important practically and intriguing theoretically. Consider the following example of a two-way binary channel (figure 15.9a,b): two people both wish to talk over the channel, and they both want to hear what the other person is saying; but you can hear the signal transmitted by the other person only if you are transmitting a zero. What simultaneous information rates from A to B and from B to A can be achieved, and how? Everyday examples of such networks include the VHF channels used by ships, and computer ethernet networks (in which *all* the devices are unable to hear *anything* if two or more devices are broadcasting simultaneously).

Obviously, we can achieve rates of $1/2$ in both directions by simple time-sharing. But can the two information rates be made larger? Finding the capacity of a general two-way channel is still an open problem. However, we can obtain interesting results concerning achievable points for the simple binary channel discussed above, as indicated in figure 15.9c. There exist codes that can achieve rates up to the boundary shown. There may exist better codes too.

Solutions

Solution to exercise 15.12 (p.235). $C(Q) = 5$ bits.
Hint for the last part: a solution exists that involves a simple $(8, 5)$ code.

(a)



(b)

$y^{(A)}$:

$x^{(B)}$ \ $x^{(A)}$	0	1
0	0	0
1	1	0

$y^{(B)}$:

$x^{(B)}$ \ $x^{(A)}$	0	1
0	0	1
1	0	0

(c)

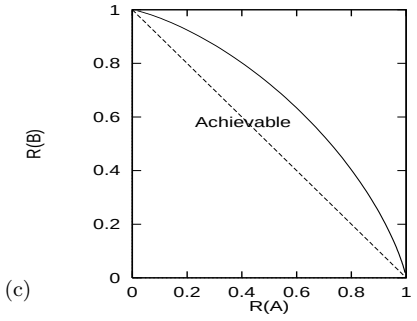


Figure 15.9. (a) A general two-way channel. (b) The rules for a binary two-way channel. The two tables show the outputs $y^{(A)}$ and $y^{(B)}$ that result for each state of the inputs. (c) Achievable region for the two-way binary channel. Rates below the solid line are achievable. The dotted line shows the 'obviously achievable' region which can be attained by simple time-sharing.