

# Handin 2: Information

## Information Theory

Willard Thór Rafnsson

due 23th February 2009

### 1 Mandatory

#### Exercise 1 (Entropy of Sources)

We are given probability distributions for several sources<sup>1</sup>.

$$\begin{array}{ll} P_2 = (0.9, 0.1) & P'_2 = (0.5, 0.5) \\ P_3 = (0.5, 0.3, 0.2) & P'_3 = (0.8, 0.1, 0.1) \\ P_4 = (0.5, 0.2, 0.2, 0.1) & P'_4 = (0.3, 0.3, 0.2, 0.2) \end{array}$$

Let  $\mathcal{S}_k$  and  $\mathcal{S}'_k$  denote a source with probability distribution  $P_k$  and  $P'_k$ , respectively. Denote source symbols in  $\mathcal{S}_k$  by  $s_{k_1}, \dots, s_{k_k}$ , and in  $\mathcal{S}'_k$  by  $s'_{k_1}, \dots, s'_{k_k}$ . Assume  $s_{k_i}$  has probability  $p_{k_i}$  associated with it, where  $p_{k_i}$  is the  $i$ th symbol in  $P_k$ . Likewise for  $s'_{k_i}$ .

- Compute  $\mathbf{l}_2(s_{2_1})$ ,  $\mathbf{l}_2(s_{2_2})$ ,  $\mathbf{H}_2(\mathcal{S}_2)$  and  $\mathbf{H}_2(\mathcal{S}'_2)$ .
- Compute  $\mathbf{l}_2(s_{3_1})$ ,  $\mathbf{l}_3(s_{3_1})$ ,  $\mathbf{H}_2(\mathcal{S}_3)$ ,  $\mathbf{H}_3(\mathcal{S}_3)$ ,  $\mathbf{H}_2(\mathcal{S}'_3)$ , and  $\mathbf{H}_3(\mathcal{S}'_3)$ .
- Compute  $\mathbf{H}_2(\mathcal{S}_4)$ ,  $\mathbf{H}_2(\mathcal{S}'_4)$ ,  $\mathbf{H}_3(\mathcal{S}'_4)$ , and  $\mathbf{H}_4(\mathcal{S}'_4)$ .
- What is the lowest average word length of a *binary* source code for  $\mathcal{S}_4$ ? Using a source code for  $\mathcal{S}_4$  with that average word length, how long is an *encoded* source sequence  $\mathbf{s}$  with  $|\mathbf{s}| = 20$  on average?
- When is the entropy of a source 0? When is it  $\log_r q$ ?

#### Exercise 2 (Channel Matrices and System Entropies)

The symbols emitted by  $\mathcal{S}_4$  from Exercise 1 are to be sent across a *binary symmetric channel* (BSC)  $\Gamma$ . Assume  $\mathcal{S}_4$  emits an infinite sequence of source symbols. To send the sequence over  $\Gamma$ , we encode each symbol in the sequence to binary using the code  $\mathcal{C}_4 = (00, 01, 10, 11)$ . Each symbol in the resulting

---

<sup>1</sup>You can consider  $\mathcal{S}_k$  and  $\mathcal{S}'_k$  for  $k = 2$  to be a biased and unbiased coin tossing experiment, or a streaming of a file in binary. The  $k = 3$  source could be our weather example for different seasons, and  $k = 4$  could be our doctor evaluation example with different weights. Recall, however, that such context is irrelevant in our theory.

bitstream is considered to be a bit emitted by a new *input source*  $\mathcal{A}$ . Let  $a_0 = 0$  and  $a_1 = 1$ . Denote the *output source* by  $\mathcal{B}$ , where  $b_0 = 0$  and  $b_1 = 1$ .

- a) What is the probability distribution of  $\mathcal{A}$ ?
- b) If we were instead sending symbols emitted by  $\mathcal{S}_3$  by the above method and  $\mathcal{C} = (0, 10, 11)$ , what then would the probability distribution of  $\mathcal{A}$  be?
- c) Returning to  $\mathcal{S}_4$ , for correctness probability  $P = 0.6$ , compute  $(P_{ij})$ ,  $(Q_{ij})$ ,  $(R_{ij})$ ,  $H_2(\mathcal{A})$ ,  $H_2(\mathcal{B})$ ,  $H_2(\mathcal{B} \mid a_0)$ ,  $H_2(\mathcal{B} \mid \mathcal{A})$ ,  $H_2(\mathcal{A} \mid b_0)$ ,  $H_2(\mathcal{A} \mid \mathcal{B})$ ,  $\text{Pr}_C$ , and  $\text{Pr}_E$ . Pick a  $\Delta : A \rightarrow B$  by the *ideal observer* rule<sup>2</sup>.
- d) For  $P = 0.8$ , compute  $(P_{ij})$ ,  $(Q_{ij})$ ,  $(R_{ij})$ ,  $\text{Pr}_C$ , and  $\text{Pr}_E$ . Pick a  $\Delta : A \rightarrow B$  by the *ideal observer* rule. Also, what happens to  $\Delta$  and  $\text{Pr}_E$  if we replace  $P$  by its inverse, 0.2?
- e) As d), except assume the probability distribution of  $\mathcal{A}$  is unknown, and apply the *maximum likelihood rule* for the computations and to find  $\Delta$ .

### Exercise 3 (Improving Reliability of $\Gamma$ )

There is a limit to how much information we can transmit through  $\Gamma$ .

- a) Compute  $I(\mathcal{A}, \mathcal{B})$  for  $\mathcal{A}$  and  $\Gamma$  from Exercise 2 with  $P = 0.6$ .
- b) Compute  $R$ , and  $C$ .

We now try to improve on our probability of deciding correctly the symbol emitted by the input source.

- c) Make a *repetition code* for  $\mathcal{A}$  and  $\Gamma$  from Exercise 2 with  $P = 0.6$ , where you repeat 3 times. Compute  $(P_{ij})$ ,  $(R_{ij})$ , and  $\text{Pr}_C$  when using this repetition code. What is the resulting *transmission rate*  $R$ ? Contrast the new  $R$  and the new  $\text{Pr}_C$  to those without repetition; how much better error correction did you obtain by repetition, and at what cost?

### Exercise 4 (Hamming Distance is a Metric)

Let  $A$  be a *field*<sup>3</sup>. For  $\mathbf{u}, \mathbf{v} \in A^n$  where  $\mathbf{u} = (u_1, \dots, u_n)$  and  $\mathbf{v} = (v_1, \dots, v_n)$ , the *Hamming distance*  $d_H(\mathbf{u}, \mathbf{v})$  between  $\mathbf{u}$  and  $\mathbf{v}$  is defined to be the number of subscripts  $i$  such that  $u_i \neq v_i$ . Prove that  $d_H$  is a *metric* (thus making  $(A^n, d_H)$  a *metric space*<sup>4</sup>). That is, for  $\mathbf{u}, \mathbf{v} \in A^n$ ,

- a)  $d_H(\mathbf{u}, \mathbf{v}) \geq 0$ , with equality iff  $\mathbf{u} = \mathbf{v}$ ,
- b)  $d_H(\mathbf{u}, \mathbf{v}) = d_H(\mathbf{v}, \mathbf{u})$ ,
- c)  $d_H(\mathbf{u}, \mathbf{w}) \leq d_H(\mathbf{u}, \mathbf{v}) + d_H(\mathbf{v}, \mathbf{w})$ ,

<sup>2</sup>If you could not solve a), assume the probability distribution  $(\frac{3}{4}, \frac{1}{4})$ .

<sup>3</sup>A set of elements for which  $+$ ,  $-$ ,  $\cdot$ , and  $/$  are well defined, and which is closed under these operations.

<sup>4</sup>Note also that  $(A^n, +, \cdot)$  is a *vector space* (just mentioning this out of interest; you do not need this in the proof).

## 2 Extra

### Exercise 5 (More Channel Matrices and System Entropies)

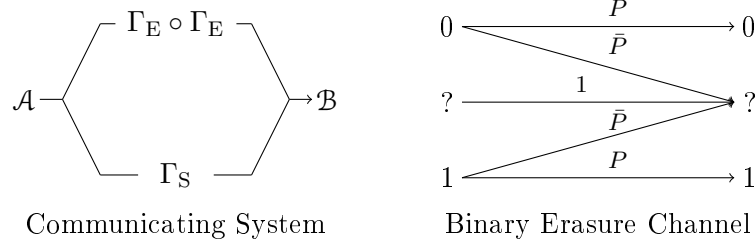
Repeat Exercise 2 for  $S'_4$ .

### Exercise 6 (Better Error Correction)

Devise a scheme which corrects more errors *and* has a better transmission rate than the repetition code in approach Exercise 3 d).

### Exercise 7 (Channel Products)

Let  $\Gamma_S$  denote a *binary symmetric channel*, and let  $\Gamma_E$  denote a *binary erasure channel*<sup>5</sup>. Consider an input-output communication system where we have two paths through which we can send data from  $\mathcal{A}$  to  $\mathcal{B}$ . One path is through a *cascade* of binary erasure channels  $\Gamma_E \circ \Gamma_E$ , the other through an ordinary binary symmetric channel  $\Gamma_S$ . The communicating system, along with the (slightly modified<sup>6</sup>) binary erasure channel, is illustrated below.



This communicating system can be used in three ways: i)  $\Gamma_E \circ \Gamma_E$ , ii)  $\Gamma_S$ , or iii)  $(\Gamma_E \circ \Gamma_E) \times \Gamma_S$ . Let  $P_E = 0.6$ , the correctness probability of  $\Gamma_E$ , and  $P_S = 0.8$ , the correctness probability of  $\Gamma_S$ . Let  $A = (0, ?, 1)$  with distribution  $(0.7, 0, 0.3)$ .

- Verify that approach i) has the *greatest*  $\text{Pr}_C$  (by computing  $(R_{ij})$  for the three approaches).
- Explain why  $\text{Pr}_C$  for  $\Gamma_E$  is so much higher than  $\text{Pr}_C$  for  $\Gamma_S$ , even if the correctness probability for  $\Gamma_S$  is considerably higher.
- Is  $\Gamma_S$  useless? That is, can  $\Gamma_S$  be used to *improve* the *transmission rate* **without** reducing  $\text{Pr}_C$ ? If so, show how. If not, explain why.

### Exercise 8 (Minimum of Conditional Entropy)

Show that if  $H(X | Y) = 0$ , then  $Y$  is a function of  $X$ .

<sup>5</sup>You can consider  $\Gamma_S$  to be a medium which maintains signal strength, but flips digits (like Ethernet on short distances), and  $\Gamma_E$  to be a medium which signal strength fades to intelligibility (like a wireless network over long distance)

<sup>6</sup>We added  $?$  to the input alphabet so that cascading  $\Gamma_E$  would make sense.