

HOMEWORK
STREAM CODES
(MACKEY - CHAPTER 6)

Necessary reading for this assignment:

- *Information Theory, Inference, and Learning Algorithms* (MacKay):
 - Chapter 6.1: *The guessing game*
 - Chapter 6.2: *Arithmetic codes*
 - Chapter 6.4: *Lempel-Ziv coding*
 - Chapter 6.6: *Summary*

Note: The exercises are labeled according to their level of difficulty: [Easy], [Medium] or [Hard]. This labeling, however, is subjective: different people may disagree on the perceived level of difficulty of any given exercise. Don't be discouraged when facing a hard exercise, you may find a solution that is simpler than the one the instructor had in mind!

Exercises.

1. The following exercises regard stream codes.

- (a) (MacKay 6.5) [Medium]
- (b) (MacKay 6.6) [Medium]

2. **(The entropy of a compressed file)** This exercise regards compression algorithms in general.

An information theory student wants to check whether she can beat Shannon's compression limit of $H(X)$ bits per symbol for an optimal code C applied to a source X .

She envisions a lossless compression method in two steps as follows:

Step 1. Apply an optimal lossless code C to the source X , obtaining a compressed file Y .

Step 2. Apply a new optimal lossless code C' to the source Y , obtaining a new compressed file Z .

Recalling Shannon's Source Coding Theorem, the student makes the following claims about her newly proposed compressing method:

Claim 1: Since code C is optimal for the source X , file Y uses approximately $H(X)$ bits to represent each symbol of X .

Claim 2: Since code C' is optimal for the source Y , file Z uses approximately $H(Y)$ bits to represent each symbol of Y .

Claim 3: File Z represents each symbol of X using approximately $H(X)H(Y)$ bits.

- (a) [Easy] Discuss whether or not each of the student's three claims are correct.
- (b) [Medium] What can we say about the size of file Y in comparison to the size of file Z ? Is Z gonna be smaller, larger, or of equal size to Y ? (Hint: Recall that Shannon's Source Coding Theorem must be valid for the compression from X to Z .)

- (c) [Medium] Using your answers to the previous items, what would be an accurate estimation for the value of $H(Y)$?
- (d) [Medium] Using your answers to the previous items, what can the student conclude about the frequency of bits 0 and 1 in any optimally compressed file? How does that relate to the title of this assignment: “*Compression and redundancy*”?