Exercise 1 – Entropy, Compression

Problem 1

We are trying to quantify how much information is provided by an event E. We make following assumptions about the information function $I(E)$:

1. $I(E)$ is a function only of the probability of the event E.
2. Information is a non-negative quantity.
3. Information due to independent events is additive.

a) Find $I(E)$.

b) Suppose we have a distribution X where each event i can happen with probability p_i. Find the average amount of information $H(X)$ that we receive with every event.

Problem 2

Consider the text: AABC
Use arithmetic encoding to encode the text. What is the length and the Shannon entropy (in bits) of the encoded text?

Problem 3

Consider the text: ABADDCCAABABEDAECBDDDAAAABAAAABBCAECEEC

a) Calculate the Shannon entropy (in bits) of the original text.

b) Calculate the length and the Shannon entropy (in bits) of the text using the block encoding

$$\mathsf{A} \mapsto 000, \qquad \mathsf{B} \mapsto 001, \qquad \mathsf{C} \mapsto 010, \qquad \mathsf{D} \mapsto 011, \qquad \mathsf{E} \mapsto 100.$$

Does the length and the entropy depend on the encoding?

c) Calculate the length and the Shannon entropy (in bits) of the text after Shannon-Fano compression.

d) Calculate the length and the Shannon entropy (in bits) of the text after Huffman compression.

Problem 4

Show that there is no compression algorithm that can compress any text to a text that is one bit shorter.