

## Exercise Sheet 4

Submit until Wednesday, November 21 at 4:00pm

### Exercise 1 (5 points)

Prove that the Elias-Gamma and the Elias-Delta code are both prefix-free.

### Exercise 2 (5 points)

Let  $l_i$  be the length of the Elias-Gamma code for the integer  $i > 0$ . Provide an exact formula for  $l_i$  for all  $i > 0$ , and determine the exact value of  $\sum_{i=1}^{\infty} 2^{-l_i}$ .

### Exercise 3 (5 points)

Prove that there is *no* universal encoding scheme (= for encoding all positive integers) which gives code length  $l_1 = 1$  to the number 1 and a code length  $l_i \leq 2 \cdot \log_2 i$  for all integers  $i \geq 2$ .

Why does this not contradict the result from the previous Exercise 2?

### Exercise 4 (5 points)

Assume that a (sorted) inverted list of document ids is randomly created as follows. There are  $N$  document ids, and each document id is included in the list with probability  $p = m/N$ , independent of other document ids. The expected length of the inverted list is then  $m$ .

Prove that there exists a modulus  $M$  such that Golomb codes are within one bit of the entropy-optimal code for encoding the *gaps* of such an inverted list.

*Hint:* First derive a formula for the probability  $\text{Prob}(G = i)$ , for arbitrary  $i$ , where  $G$  is the size of a single fixed gap. When using that formula in the remainder of your argument, you can assume that  $m \ll N$  and hence  $1 - p \approx e^{-p}$ .

Produce a PDF with your solutions (preferably using LaTeX), and commit it to our SVN, in a new sub-directory *exercise-sheet-04*, along with a text file *experiences.txt* with your feedback. As a minimum, say how much time you invested and if you had major problems, and if yes, where.