Logic and Theory of Discrete Systems
RWTH AACHEN UNIVERSITY
Prof. Dr. M. Grohe
E. Fluck, N. Runde

Exercise Sheet 8

Due date: Monday, July 10 until 13:00

- This is the last sheet. To be able to take the exam you need 80 points in total over all 8 sheets.
- Please upload your solutions to RWTH Moodle.
- The due date is on Monday, July 10 until 13:00.
- Hand in your solutions in groups of **two to three students**. If you need to change your group, contact algds@lics.rwth-aachen.de.
- Hand in the solutions of your group as a single PDF file.
- A discussion regarding this exercise sheet will take place on **Friday, July 14 14:30** in room AH II.

Exercise 1 (Frequency Moments and Tug-Of-War) 3+4+1+3+1+2=14 points

Consider the following stream of data elements over the universe $\mathbb{U} = \{1, 2, \ldots, 9\}$:

$$\mathbf{a} = 1, 4, 4, 1, 5, 3, 8, 2, 2, 1, 5.$$

a) Compute $F_0(\mathbf{a})$, $F_1(\mathbf{a})$ and $F_2(\mathbf{a})$.

b) Assume we run the AMS-Estimator (slide 8.36) on $\mathbf{a}$ and evaluate the variables at the end of the while loop (after line 8). Complete the following table with suitable values:

i	1	2	3	4	5	6	7	8	9	10	11
a_i	1			1							
a	1										2
r	1		2				2				2

c) What is the estimated result for $F_2(\mathbf{a})$ returned by the AMS-Estimator in b)?

d) What is the result x^2 returned by the Tug-of-War estimator on $\mathbf{a}$ if the randomly chosen hash function is given by

u	1	2	3	4	5	6	7	8	9
$h(u)$	1	1	-1	1	-1	1	1	1	-1

e) Is there a hash function $h' \colon \mathbb{U} \to \{-1, 1\}$ such that Tug-of-War returns a better (i. e. closer) estimate for F_2 than it does in part d)? If yes, give such a hash function. If no, argue why not.

f) Now think of any stream b with n elements, of which m are distinct. What are the minimum and maximum possible values of $F_2(b)$ (as a function of m and n).

Exercise 2 (Improve the Probability) 6 points

Consider an algorithm $\mathfrak{A}(h)$ that uses a (truly) random hash function $h \in \mathcal{H}$ and gives an estimate $\hat{x} = \mathfrak{A}(h)$ of the true value x of some variable. Suppose that:

$$\Pr_{h \in \mathcal{H}} \left(\frac{x}{4} \leq \hat{x} \leq 4x\right) \geq 0.6.$$

The probability of the estimate is with respect to choice of the hash function. How would you compute an estimate x' that has an improved probability of:

$$\Pr\left(\frac{x}{4} \leq x' \leq 4x\right) \geq 0.8?$$

Hint: Since we do not know the variance, taking the **average** of multiple runs may not help.

Exercise 3 (Minimum Memory for Distinct Elements Approximation) 0 points

Show that any deterministic algorithm that even guarantees to **approximate** the number of distinct elements in a data stream over universe $\mathbb{U} = \{1, ..., m\}$ with error less than $\frac{m}{16}$ must use $\Omega(m)$ bits of memory. This is even the case for data streams of length less than $2m$.

Hint: There is a constant $c > 0$, for which it is possible to create 2^{cm} subsets of $\{1, ..., m\}$, each with $m/2$ elements, such that no two of the subsets have more than $3m/8$ elements in common. You can use this fact without proving it.