TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

Department of Computer Science
Algorithms & Complexity Group

Issue Date: 22.04.2014
Version: 2014-04-29 12:09

Exercise Sheet 1 for Computational Biology (Part 2), SS 14

Hand In: *Until Tuesday, 06.05.2014, 10:00 am, email to* `wild@cs...` *or in lecture.*

Organisational Stuff

- Please hand in your solutions as teams of 2 to 3 students.
- In total, you need at least 40 % of the reachable points to be allowed to take the oral exam (the sum over all sheets, not on every single sheet).
- Please make sure you have enrolled for this course in the OLAT online system.

  The link is on our course website
  `http://wwwagak.cs.uni-kl.de/Vorlesung/bioinf2-14.html`.

Problem 1

4 points

Consider the following *Bernoulli game* with players Alice and Bob: Alice chooses a word $a \in \{0,1\}^k$. We assume Bob knows a and then chooses $b \in \{0,1\}^k$ with $b \neq a$. Afterwards a random 0-1 word $s = s_1 s_2 \cdots$ is generated with $\Pr[s_i = 0] = \Pr[s_i = 1] = \frac{1}{2}$ independently for all i. The winner is the player whose word appears first as a *subword* of s. We say that *Bob probably wins* iff

$$\Pr[b \text{ occurs first}] > \Pr[a \text{ occurs first}].$$

A word s is called *A-win* iff b does not appear in s and a appears exactly once in s, namely as a suffix. An *A-almost-win* is a word s, for which $s \cdot a$ is an A-win. We call the set of all A-almost-wins S_a. B-wins, B-almost-wins and the set S_b are defined similarly.

For two sets of words X and Y, we define their concatenation as usual:

$$X \cdot Y := \{x \cdot y \mid x \in X \wedge y \in Y\}\ .$$

Furthermore for two words a and b the set H_{ab} is defined by

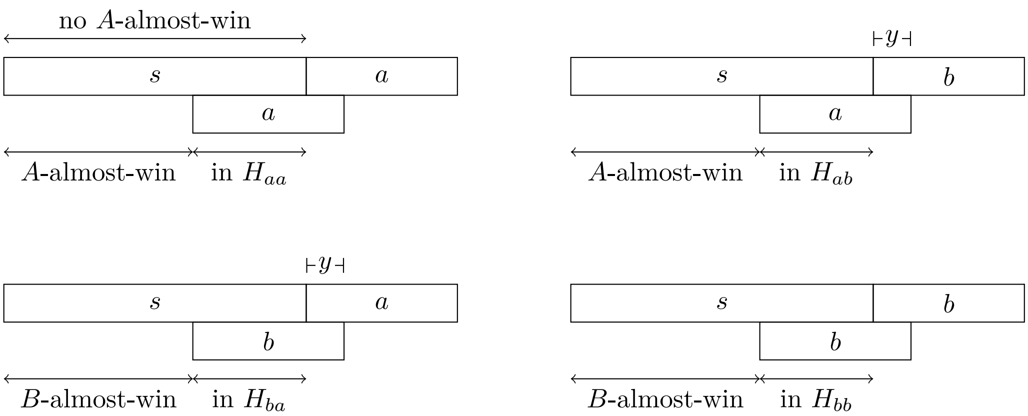
$$H_{ab} := \{x \mid a = x \cdot y \text{ and } y \text{ is a prefix of } b \text{ with } y \neq \varepsilon\}\ .$$

a) Let $S := \{s \in \{0,1\}^\star \mid \text{neither } a \text{ nor } b \text{ is subword of } s\}$. Show that

$$S = (S_a \cdot H_{aa}) \cup (S_b \cdot H_{ba}) \text{ and } S = (S_b \cdot H_{bb}) \cup (S_a \cdot H_{ab})$$

holds.

Hint: Note the connection between H_{ab} and the sets of A/B-wins:



b) *Conway's Inequality* states that Bob probably wins iff

$$P_{aa} - P_{ab} > P_{bb} - P_{ba}\ ,$$

where $P_{ab} := \sum_{u \in H_{ab}} 2^{-|u|}$. Prove this equivalence.

Hint: Use a).

Problem 2

4 points

Consider again the *Bernoulli game* from Problem 1 and prove that for $k \geq 3$, Bob can always win probably, i. e. for each a there is a b, s. t. Bob probably wins.

Hint: Use Conway's inequality.
If $a = a_1 a_2 \cdots a_k$ try using $b = \boxed{?} \cdot a_1 \cdots a_{k-1}$.

Problem 3

3 points

Consider the following unfair coin: With probability p, it shows H (heads) and with probability $q := 1 - p$, it shows T (tails).

What is the expected number of coin tosses until the pattern THTTH appears for the first time? Also determine the corresponding variance as a function of p.

Problem 4

4 points

Progress in lecture was slower than anticipated; therefore Problem 4 is deferred to the next exercise sheet.

We consider the data structure from the lecture for efficiently solving the *lce*-problem. Recall: It is based on a compact suffix tree and uses binary numbers in additional node labels.

Find necessary and sufficient conditions for a node u being a predecessor of node v. The condition may only involve the binary numbers i and j that u respectively v are labelled with.

Hint: The function h may be useful for that, where $h(k)$ is the position (counted from the right end) of the least significant 1 in the binary representation of k.
For example $h(8) = h(1000_2) = 4$ and $h(5) = h(101_2) = 1$.