HW 5 Randomized Algorithms

MPRI 1.24 Tue. Dec. 16, 2014 - **Due on Tue. Jan. 6, 2015**

You are asked to complete the exercise marked with a [★] and to send me your solutions at:
nicolas.schabanel@liafa.fr
(or drop it in my mail box at the 4th floor of Sophie Germain) on **Tue. Jan. 6, 2015**.

■ **Exercise 1 (Traffic monitoring: uniformity detection).** Imagine that we are running a huge website and we want to prevent attacks by keeping track of the origins of the various clients currently connected to the server. Along time, clients connect and then disconnect from the website. And we want to detect if all the clients connected are from the same IP address. But we do not want to slow down the server and wish to dedicate to this task only a *constant* memory, i.e. only a constant number of integers. We model the problem as follows:

We are given an infinite stream of events $e_1, e_2, \ldots, e_n, \ldots$ where each e_i is either **connect**(x) or **disconnect**(x) where x is a positive integer standing for the IP address of the client (dis-)connecting. We assume that the stream is wellformed, i.e. that there are always at least as many events **connect**(x) as **disconnect**(x) from the beginning of the stream to any position for every integer x. We want to detect when all the clients connected have the same IP address x.

► ***Question* 1.1)** *Spot when to set the alarm on in the following sequence where x denotes the event* **connect**(x) *and $\bar{x}$ the event* **disconnect**(x)*:*

$$1, 2, 3, \bar{2}, \bar{3}, 1, 1, \bar{1}, 4, 6, 7, \bar{1}, \bar{6}, \bar{1}, 2, \bar{2}, \bar{4}, 8, 3, \bar{3}, \bar{7}, 9$$

We consider the following algorithm that uses only three integer variables:

- start with $\mathtt{n} := 0$, $\mathtt{a} := 0$ and $\mathtt{b} := 0$ at $t = 0$;
- on event **connect**(x): do $\mathtt{n} := \mathtt{n} + 1$, $\mathtt{a} := \mathtt{a} + x$ and $\mathtt{b} := \mathtt{b} + x^2$;
- on event **disconnect**(x): do $\mathtt{n} := \mathtt{n} - 1$, $\mathtt{a} := \mathtt{a} - x$ and $\mathtt{b} := \mathtt{b} - x^2$;
- set on the alarm every time that $\mathtt{n} > 0$ and $\mathtt{b} = \mathtt{a}^2/\mathtt{n}$.

The right way to the correctness of this deterministic algorithm passes through the analysis of a random variable. Consider a random variable X taking positive integer values. We denote by $\mathsf{supp}(X) = \{\, x : \Pr\{X = x\} > 0 \,\}$ and assume that $|\mathsf{supp}(X)| < \infty$. We denote by $\mathbb{E}[X]$ and $\mathbb{V}\mathsf{ar}[X] = \mathbb{E}[(X - \mathbb{E}(X))^2]$ respectively the expectation and the variance of X.

► ***Question* 1.2)** *Show that $|supp(X)| = 1$ if and only if $\mathbb{V}\mathsf{ar}[X] = 0$.*

► ***Question* 1.3)** *Conclude that the algorithm is correct.*

■ **Exercise 2 (Using fingerprints to check matrix multiplication).** [★]
The best known algorithms for multiplying two $n \times n$-matrices take time $O(n^\alpha)$ for some $\alpha > 2$. However, there is a simple randomized algorithm for verifying matrix products with high probability in $O(n^2)$ time.

Suppose I claim that $AB = C$ where A, B, and C are integer-valued $n \times n$-matrices. You can confirm this by applying both sides of this equation to a random vector v, and checking that $ABv = Cv$.

► ***Question* 2.1)** *Describe how to check if $ABv = Cv$ in $O(n^2)$ time, i.e. without computing the $n \times n$-matrix product AB.*

We desire to improve furthermore the checking procedure by using fingerprints. Assume that $AB \neq C$ and let $\alpha = \max_{i,j}\{|A_{ij}|, |B_{ij}|, |C_{ij}|\}$ bound the maximum coefficient in A, B and C.

► ***Question* 2.2)** *Explain how to choose a prime number $p = O(\log(n\alpha) \log\log(n\alpha))$ such that with probability at least $\frac{9}{10}$, we have $AB - C \neq 0 \mod p$.*
▷ Hint. *Use that the kth prime number p_k verifies $0.91k \ln k < p_k < 1.7k \ln k$, as proved by Felgner in 1990.*

Assume now that we are lucky and $AB - C \neq 0 \mod p$.

► ***Question* 2.3)** *Show that if v is chosen uniformly at random from $\{0, 1, \ldots, p-1\}^n$, then the probability that $ABv = Cv \mod p$ holds is at most $1/p$.*