Universität Siegen
Lehrstuhl Theoretische Informatik
Markus Lohrey

Exercise 6

Task 1. A family $\mathcal{H}$ of hash functions $h\colon A \to \mathbb{F}_p$ has the *uniform difference property* provided that, for all $a, a' \in A$ with $a \neq a'$, the difference $h(a) - h(a') \bmod p$ is uniformly distributed in $\mathbb{F}_p$ when h is drawn uniformly at random from the family $\mathcal{H}$.

(a) Consider $\mathcal{H}_p = \{h_{x,y} \mid x, y \in \mathbb{F}_p\}$, where $h_{x,y}\colon \mathbb{F}_p \to \mathbb{F}_p$ with $h_{x,y}(a) = ax + y \bmod p$, as in the lectures (slide 116). Show that $\mathcal{H}_p$ has the uniform difference property.

Let us now fix a family of hash functions $\mathcal{H}$ with the uniform difference property as above. For any vector of hash functions $H = (h_1, \ldots, h_n) \in \mathcal{H}^n$, we define the hash function

$$h_H\colon A^n \to \mathbb{F}_p \quad \text{by} \quad h_H(a_1, \ldots, a_n) = h_1(a_1) + \ldots + h_n(a_n) \bmod p.$$

(b) Show that the family $\mathcal{H}^{(n)} := \{h_H \mid H \in \mathcal{H}^n\}$ has the uniform difference property and, in particular, that $\mathcal{H}^{(n)}$ is a universal family of hash functions.

Task 2. Prove the following statements; see slides 122 and 123, respectively.

(a) If $X_1, \ldots, X_n$ are independent random variables, then

$$\mathbf{E}[X_1 \cdot \ldots \cdot X_n] = \mathbf{E}[X_1] \cdot \ldots \cdot \mathbf{E}[X_n].$$

(b) If $X_1, \ldots, X_n$ are pairwise independent random variables, then

$$\mathbf{Var}[X_1 + \ldots + X_n] = \mathbf{Var}[X_1] + \ldots + \mathbf{Var}[X_n].$$

Task 3. Consider the stream $S = (101, 011, 010, 111, 011, 101, 000, 001)$. Approximate the number $F_0(S)$ of distinct elements in the stream S with the algorithm of Alon, Matias, and Szegedy (slide 130) using the hash functions $h_{x,y}(a) = ax + y$ over $\mathbb{F}_{2^3}$ with

(a) $x = 101$ and $y = 001$, (b) $x = 100$ and $y = 101$.

You can use that addition in the field $\mathbb{F}_{2^3}$ is just bitwise XOR, whereas multiplicaiton with the relevant values of x is defined according to the following table.

a	000	001	010	011	100	101	110	111
$101 \cdot a$	000	101	001	100	010	111	011	110
$100 \cdot a$	000	100	011	111	110	010	101	001