Algorithmic Data Science - Exercises Series 2

Konstantinos Papadakis
Data Science and Machine Learning 03400149
konstantinospapadakis@mail.ntua.gr

June 23, 2022

Exercise 1

(a)

We have that

$$\begin{aligned} &h_{a,b}(x) = h_{a,b}(y) \\ \iff &ax + b \equiv ay + b \mod m \\ \iff &a(x - y) \equiv 0 \mod m \end{aligned}$$

which in the case of $x = m,\ y = 0$ is true $\forall a, b$ therefore

$$P(h_{a,b}(x) = h_{a,b}(y)) = 1 > \frac{1}{m}$$

meaning that the family is not universal.

(b)

This exercise is *Theorem 11.5* in the book *Introduction to Algorithms by Cormen et al.*.

Let $x, y \in \mathbb{Z}_p : x \neq y$.

Define

$$\begin{aligned} u &:= ax + b \mod p \\ v &:= ay + b \mod p \end{aligned}$$

Note that $u \neq v$ since $u - v \equiv a(x - y) \not\equiv 0 \mod p$ because $a \neq 0$ and $x \not\equiv y \mod p$, the later holding because by hypothesis $x \neq y$ and $x, y < p$ Therefore, there are no collisions when we apply $x \mapsto ax + b \mod p$.

We proceed to show that $(a, b) \mapsto (ax+b \mod p,\ ax+b \mod p)$ is a bijection between the pairs $(a, b) \in \mathbb{Z}_p^* \times \mathbb{Z}_p$ and the pairs $(u, v) \in \mathbb{Z}_p \times \mathbb{Z}_p : u \neq v$.

We can solve for a, b and get a unique solution

$$\begin{aligned} a &= \frac{u - v}{x - y} \mod p \\ b &= r - ak \mod p \end{aligned}$$

Where $\frac{1}{t}$ is the inverse of t in $\mathbb{Z}_p$

Therefore the mapping is one to one. Since we also have that both the domain and the codomain have $p(p - 1)$ elements, the mapping is a bijection. Thus, if (a, b) is uniformely distributed, so is (u, v).

Therefore, the probability that $x, y \in \mathbb{Z}_p : x \neq y$ collide is equal to the probability that $u \equiv v \mod m$ collide when $(u, v) \in \mathbb{Z}_p \times \mathbb{Z}_p : u \neq v$ are chosen uniformely randomly. We proceed to calculate that probability.

Given u, of the $p - 1$ possible remaining values for v we have that at most $\lceil \frac{p}{m} \rceil - 1 \leq \frac{p-1}{m}$ can collide with u.

Therefore the probability of colision is $\leq \frac{1}{m}$, meaning that the hash function family is universal.

(c)

The proof in is still valid, since $x \in U \implies x < p$ is still valid. Therefore the hash function family remains universal.

Exercise 2