Lecture 14: Universal Hash Function Family

Recall: k-wise Independence

Recall the definition of k-wise Independent Function family. Let $\mathcal{D}$ is the domain and $\mathcal{R}$ is the range.

Definition

Let $\mathcal{H}$ be a set of functions $\mathcal{D} \to \mathcal{R}$. For distinct $x_1, x_2, \dots, x_k \mathcal{D}$ and any $y_1, y_2, \dots, y_k \in \mathcal{R}$, the class of hash function $\mathcal{H}$ satisfies the following condition.

$$\mathbb{P}\left[h(x_1) = y_1, \dots, h(x_k) = y_k \colon h \leftarrow \mathcal{H}\right] = \frac{1}{|\mathcal{R}|^k}$$

Intuition: The first k inputs are answered independently and uniformly at random from $\mathcal{R}$.

One construction: For $\mathcal{D} = \mathcal{R} = \mathbb{F}$ a field,

$$\mathcal{H} = \left\{ h_{a_0, a_1, \dots, a_{k-1}} \colon a_0, a_1, \dots, a_{k-1} \in \mathbb{F} \right\}$$

where $h_{a_0, a_1, \dots, a_{k-1}}(X) = a_0 + a_1 X + \cdots + a_{k-1} X^{k-1}$.

2-Independence

- A hash function family $\mathcal{H}$ is 2-Independent if it is k-wise Independent, for $k = 2$
- So, they satisfy the following constraint for all distinct $x_1, x_2 \in \mathcal{D}$ and $y_1, y_2 \in \mathcal{R}$.

$$\mathbb{P}\left[h(x_1) = y_1, h(x_2) = y_2\right] = \frac{1}{|\mathcal{R}|^2}$$

Universal Hash Function Family

Definition (Universal Hash Function Family)

A set $\mathcal{H}$ of functions $\mathcal{D} \to \mathcal{R}$ is a universal hash function family if, for every distinct $x_1, x_2 \in \mathcal{D}$ the hash function family $\mathcal{H}$ satisfies the following constraint.

$$\mathbb{P}\left[h(x_1) = h(x_2) \colon h \overset{\$}{\leftarrow} \mathcal{H}\right] \leqslant \frac{1}{|\mathcal{R}|}$$

Intuition: Given any two distinct inputs x_1 and x_2, a random $h \overset{\$}{\leftarrow} \mathcal{H}$ ensures that the output of $h(x_1)$ and $h(x_2)$ does not collide with high probability

2-wise Independence implies Universality I

- Underlying Intuition: Note that if the first two inputs are answered uniformly and independently at random by a function then they outputs are unlikely to collide
- So, can we prove the following result

Theorem

Let $\mathcal{H}$ be a 2-wise independent hash function family then $\mathcal{H}$ is also a universal hash function family.

2-wise Independence implies Universality II

Proof.

- Since $\mathcal{H}$ is a 2-wise independent hash function family then it satisfies the following condition. For distinct $x_1, x_2 \in \mathcal{D}$ and any $y_1, y_2 \in \mathcal{R}$ we have:

$$\mathbb{P}\left[h(x_1) = y_1, h(x_2) = y_2 \colon h \xleftarrow{\$} \mathcal{H}\right] = \frac{1}{|\mathcal{R}|^2}$$

- Fix $y_2 = y_1$. Now, we have the guarantee

$$\mathbb{P}\left[h(x_1) = h(x_2) = y_1 \colon h \xleftarrow{\$} \mathcal{H}\right] = \frac{1}{|\mathcal{R}|^2}$$

- Summing over all possible $y_1 \in \mathcal{R}$, we have

$$\sum_{y_1 \in \mathcal{R}} \mathbb{P}\left[h(x_1) = h(x_2) = y_1 \colon h \xleftarrow{\$} \mathcal{H}\right] = \sum_{y_1 \in \mathcal{R}} \frac{1}{|\mathcal{R}|^2} = \frac{1}{|\mathcal{R}|}$$

2-wise Independence implies Universality III

- Now, note that

$$\mathbb{P}\left[h(x_1) = h(x_2) \colon h \xleftarrow{\$} \mathcal{H}\right] = \sum_{y \in \mathcal{R}} \mathbb{P}\left[h(x_1) = h(x_2) = y \colon h \xleftarrow{\$} \mathcal{H}\right]$$
$$= \frac{1}{|\mathcal{R}|} \qquad \text{(from above)}$$

- This proves that $\mathcal{H}$ is a universal hash function family

Food for thought

- We saw that if $\mathcal{H}$ is 2-wise independent then $\mathcal{H}$ is universal. Does this work the other way? That is, if $\mathcal{H}$ is universal then $\mathcal{H}$ is also 2-wise independent.
- The definition of universal hash function family states that the collision probability is $\leqslant \frac{1}{|\mathcal{R}|}$. Can the collision probability be $< \frac{1}{|\mathcal{R}|}$?

We will start answering both these questions simultaneously using an example. We shall prove the formal version of this result in the next lecture.

Observations I

Observation

When the range $\mathcal{R}$ is large than the domain $\mathcal{D}$, a universal hash function family need not necessarily be 2-wise independent.

- So, we need to demonstrate one counterexample $\mathcal{H}$ that is universal hash function family but is not 2-wise independent
- Pick any $\mathcal{D}$ with size $\geqslant 2$
- Let h^* be any one-to-one function $\mathcal{D} \to \mathcal{R}$ (since, $\mathcal{R}$ is at least as large as $\mathcal{D}$, such a function exists)
- Let $\mathcal{H} = \{h^*\}$
- Note that $\mathcal{H}$ is a universal hash function family (because the function is one-to-one)

Observations II

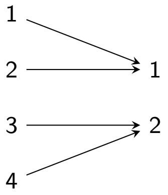
- Note that $\mathcal{H}$ is <u>not</u> a 2-wise independent hash function family. We can choose any two distinct $x_1, x_2 \in \mathcal{D}$ and $y_1 = h^*(x_1)$ and $y_2 = h^*(x_2)$. Now, we have

$$\mathbb{P}\left[h(x_1) = y_1, h(x_2) = y_2 \colon h \stackrel{\$}{\leftarrow} \mathcal{H}\right] = 1 \not\leq \frac{1}{|\mathbb{R}|^2}$$

Observation

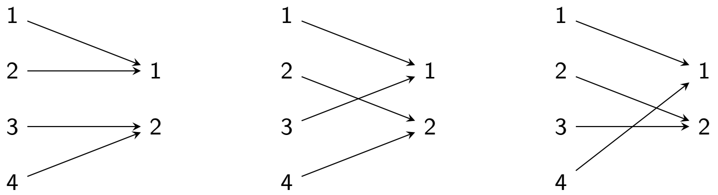
We can design universal hash function families $\mathcal{H}$ such that the collision probability is $< \frac{1}{|\mathcal{R}|}$, where the range $\mathcal{R}$ is smaller is size than the domain $\mathcal{D}$

- For such a construction we shall use $\mathcal{D} = \{1, 2, 3, 4\}$ and $\mathcal{R} = \{1, 2\}$
- We shall use a pictorial representation for functions for brevity. The picture below represents the function $f : \mathcal{D} \to \mathcal{R}$ such that $f(1) = 1$, $f(2) = 1$, $f(3) = 2$, and $f(4) = 2$.



Observations IV

- Consider the three functions h_1, h_2, h_3 defined below



- Define $\mathcal{H} = \{h_1, h_2, h_3\}$
- Collision Probability. Check that the collision probability is $\frac{1}{3} < \frac{1}{2}$. So, this is a universal hash function family with collision probability $< \frac{1}{|\mathcal{R}|}$
- 2-wise Independence. Pick $x_1 = 1$, $x_2 = 4$, $y_1 = 1$, and $y_2 = 2$. Note that

$$\mathbb{P}\left[h(x_1) = y_1, h(x_2) = y_2 \colon h \xleftarrow{\$} \mathcal{H}\right] = \frac{2}{3} \not\leqslant \frac{1}{4} = \frac{1}{|\mathcal{R}|^2}$$

Observations V

- Therefore, we have a construction of hash function family that is universal but not 2-wise independent!

Food for Thought

What is the smallest possible achievable collision probability?

Looking Ahead

In the next lecture, we shall prove the following result. For any class of hash function family $\mathcal{H}$, we shall prove the following bound

Theorem

Let $\mathcal{H}$ is a hash function family from the domain $\mathcal{D}$ to the range $\mathcal{R}$. We shall prove that, there exists distinct $x_1, x_2 \in \mathcal{D}$ such that

$$\mathbb{P}\left[h(x_1) = h(x_2) \colon h \xleftarrow{\$} \mathcal{H}\right] \geqslant \frac{\frac{N}{M} - 1}{N - 1},$$

where $|\mathcal{D}| = N$, $|\mathcal{R}| = M$, and $N/M \geqslant 1$. Further, this bound is achievable when M divides N.

And note that we always have $\frac{\frac{N}{M}-1}{N-1} < \frac{1}{M}$. We can show that the class of hash functions that achieves equality in the above bound is not a 2-wise independent hash function family!