Lecture 12: k-wise Independent Hash Function Family

Intuition

- We want to design a hash function family $\mathcal{H}$ from a field $\mathbb{F}$ to $\mathbb{F}$
- For $h \stackrel{\$}{\leftarrow} \mathcal{H}$, we want the first k inputs to be answered randomly

Definition

Definition (k-wise Independence)

Let $\mathcal{H} = \{h_1, \ldots, h_t\}$ be a family of hash functions such that $h_i \colon \mathcal{D} \to \mathcal{R}$, where $\mathcal{D}$ is the domain and $\mathcal{R}$ is the range. For any distinct $x_1, \ldots, x_k \in \mathcal{D}$ and any $y_1, \ldots, y_k \in \mathcal{R}$ we have the following guarantee.

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

Elaborating the Definition I

- From the definition, we can compute the probability that $h(x_1) = y_1, \ldots, h(x_{k-1}) = y_{k-1}$ (note we dropped the constraint that $h(x_k) = y_k$)
- For any $x_k \in \mathcal{D}$, we can simplify the probability as follows:

$$\mathbb{P}\left[h(x_1) = y_1, \ldots, h(x_{k-1}) = y_{k-1} \colon h \xleftarrow{\$} \mathcal{H}\right]$$

$$= \sum_{y_k \in \mathcal{R}} \mathbb{P}\left[h(x_1) = y_1, \ldots, h(x_{k-1}) = y_{k-1}, f(x_k) = y_k \colon h \xleftarrow{\$} \mathcal{H}\right]$$

$$= \sum_{y_k \in \mathcal{R}} \frac{1}{|\mathcal{R}|^k} = \frac{1}{|\mathcal{R}|^{k-1}}$$

- Remark: This proof-technique should seem similar to the solution to one of your HW01 problems!

Elaborating the Definition II

- Proceeding inductively, we can prove the following statement. For all $i \in \{1, 2, \dots, k\}$, we have

$$\mathbb{P}\left[h(x_1) = y_1, \dots, h(x_i) = y_i \colon h \xleftarrow{\$} \mathcal{H}\right] = \frac{1}{|\mathcal{R}|^i}$$

- Using these guarantees, we can prove the following statements:

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

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

$$\vdots$$

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

Elaborating the Definition III

- The definition of k-wise independent function insists that the probability of a random $h \xleftarrow{\$} \mathcal{H}$ simultaneously mapping $x_1 \mapsto y_1,\ x_2 \mapsto y_2,\ \dots,\ x_k \mapsto y_k$ is $1/|\mathcal{R}|^k$
- The first input is uniformly randomly answered: We have proved above that $h(x_1) = y_1$ is $\frac{1}{|\mathcal{R}|}$.
- The second input is uniformly randomly answered (even conditioned on the first input's answer): We have proved above that $h(x_2) = y_2$ conditioned on $h(x_1) = y_1$) is $\frac{1}{|\mathcal{R}|}$.
- Similarly, for any $i \in \{1, \dots, k\}$, the i-th input is uniformly randomly answered (even conditioned on the previous $(i-1)$ inputs' answers): We have proved above that $h(x_i) = y_i$ conditioned on $h(x_1) = y_1,\ \dots,\ h(x_{i-1}) = y_{i-1}$ is $\frac{1}{|\mathcal{R}|}$
- Summarizing: It says that the first k inputs to the function are answered independently and uniformly at random

Example Construction

- Let $\mathcal{D} = \mathcal{R} = \mathbb{F}$, where $\mathbb{F}$ is a field
- The hash function family is defined as follows:

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

where the hash function $h_{a_0, a_1, \dots, a_{k-1}} : \mathbb{F} \to \mathbb{F}$ is defined as follows

$$h_{a_0, a_1, \dots, a_{k-1}}(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_{k-1} X^{k-1}$$

- Intuitively, $\mathcal{H}$ is the set of all polynomials of degree $< k$ with coefficients in the field $\mathbb{F}$
- You have already proved in HW01 that this hash function family is k-wise independent!

Food for Thought

- Let $\mathcal{D} = \mathbb{F}^m$ and $\mathcal{R} = \mathbb{F}$, where $\mathbb{F}$ is a field
- The hash function family is defined as follows:

$$\mathcal{H} = \left\{ h_{a_1,\ldots,a_m} \colon a_1, \ldots, a_m \in \mathbb{F} \right\}$$

  where the hash function $h_{a_1,\ldots,a_m} \colon \mathbb{F}^m \to \mathbb{F}$ is defined as follows

$$h_{a_1,\ldots,a_m}(x_1, \ldots, x_m) = a_1 x_1 + \cdots + a_m x_m$$

- Is this k-wise independent hash function family?