Advanced Algorithms 2024 21/10, 2024

Exercise 05

Lecturer: Maximilian Probst *Teaching Assistant: Jakob Nogler*

1 Hashing with Chaining via 2-wise Independence

Recall that, when using a truly random hash function with chaining, we can get a hashing algorithm with an expected time of $\mathcal{O}(1 + n/m)$ for all operations. However, in this exercise, you are only allowed to use a uniform, 2-wise independent hash function. Argue that, using chaining, you can still obtain a hashing algorithm with an expected time of $\mathcal{O}(1 + n/m)$ for all operations.

2 Extending Hash Functions to Non-Prime Domains

In the lecture, you have seen an example of uniform, k-wise independent families of hash functions $h : U \to [m]$ for each m that is prime. However, the assumption that m is prime might turn out to be quite inconvenient for some applications. In this exercise, we want to show that we can allow m to not be prime at a small cost of h being only almost uniform.

We let $m \in \mathbb{N}$ be arbitrary (not necessarily prime) and choose any $c \geq 1$. Moreover, we pick a prime number p such that $4c \cdot m \geq p \geq 2c \cdot m$, which exists by Bertrand's postulate. The family of hash functions we want to consider consists of the functions

$$g_k(x) = \left(\sum_{i=1}^{k} a_i x^{i-1} \mod p \right) \mod m,$$

where $a_1, a_2, \ldots, a_k$ are drawn independently and uniformly at random from $[p]$.

Prove that $g_k(x)$ is strongly universal k-wise independent with uniformity $1 + 1/2c$.

3 Linear Probing with 3-wise Independence

In class, we showed that using a 5-independent hash function, we can implement linear probing with expected time per operation $O(1)$.

Show that using a 3-independent hash function, we can still prove that each operation only takes expected time $O(\log n)$.

4 Method of Moments

Let $Y_0, \ldots, Y_{n-1}$ be 0/1 random variables, each taking the value 1 with probability p, and let $X = \sum_{i=0}^{n-1}$. Let us write $\mu = \mathbb{E}[X] = np$. In class, you have seen that if the Y_i's are 4-wise independent, then for any $d > 0$

$$\mathbb{P}[|X - \mu| > d\sqrt{\mu}] < 4/d^4,$$

which is strictly better than what Chebyshev's bound would give us.

In this exercise, we want to show that under stricter independence assumptions, we can obtain even stronger bounds. Specifically, we will show that for any **even** $k \geq 2$, if we assume that the Y_i's are k-wise independent, then for any $d > 0$ we have

$$\mathbb{P}[|X - \mu| > d\sqrt{\mu}] = O(k^k/(2d)^k).$$

Let therefore $k \geq 2$ be even and assume the Y_i's are k-wise independent. Generalize the proof you have seen in class to prove this statement. You can follow the steps below.

1. The key to our proof is to again upper bound the k-th moment which is defined as $\mathbb{E}[(X-\mu)^k] = \sum_{i_1,i_2,\ldots,i_k \in [n]} \mathbb{E}[(Y_{i_1} - p)(Y_{i_2} - p) \cdots (Y_{i_k} - p)]$.

   Consider now any indices $i_1, i_2, \ldots, i_k \in [n]$. We let $j_1 < j_2 < \ldots < j_c$ be the distinct indices among $i_1, i_2, \ldots, i_k$ and let m_h be the multiplicity of j_h, i.e. how often the j_h appears among $i_1, i_2, \ldots, i_k$.

   Argue that if $c > k/2$, then $\mathbb{E}[(Y_{j_1} - p)^{m_1}(Y_{j_2} - p)^{m_2} \cdots (Y_{j_c} - p)^{m_c}] = 0$. Argue that otherwise, we have $\mathbb{E}[(Y_{j_1} - p)^{m_1}(Y_{j_2} - p)^{m_2} \cdots (Y_{j_c} - p)^{m_c}] \leq p^c$.

2. Use the insight from above to argue that $\mathbb{E}[(X-\mu)^k] \leq c^k \cdot (np)^c$. Then, conclude the proof.

   Hint:
   For any set of distinct indices $j_1 < j_2 < \ldots < j_c$, crudely upper bound the number of (non-distinct) indices $i_1, i_2, \ldots, i_k$ that produce these distinct indices. Use that $\binom{k}{c} \leq c^k$.