CS 170 Final Review: Hashing

1 Chernon's Educated Guess

Chernon gets one try to guess a very important value. All Chernon has at hand is an algorithm that yields a ϵ-accurate answer with probability 0.6, and an inaccurate answer with probability 0.4. How can Chernon give a ϵ-accurate answer with a probability of at least 0.999?

Solution: Chernon should run his algorithm k times, and then return the median. The higher k is, the higher chance of getting a median that is within ϵ of the true value. We can use the Chernoff bound to give us a lower bound for k.

Let's attempt to calculate the probability that our median of k answers is *not* ϵ-accurate. Assuming we have an odd k, to output a final ϵ-accurate answer, we just need half or more of our iterations to give ϵ-accurate outputs. This is analagous to flipping a biased, heads-favored coin a total of k flips, and hoping we get more heads than tails.

Let X_i be a Bernoulli random variable that is 1 if the i^{th}output is ϵ-accurate, and 0 if not. We know these probabilities to be 0.6 and 0.4, respectively. We attempt to count the number of times that the answer is ϵ-accurate:

Let $X = X_1 + X_2 + ... + X_n$. Then the exact probability is:

$$\Pr[X > \frac{n}{2}] = \sum_{\lfloor \frac{n}{2} \rfloor + 1}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

Using the Chernoff inequality, this can be bounded:

$$\Pr[X > \frac{n}{2}] \geq 1 - e^{-\frac{n}{2p}(p-\frac{1}{2})^2}$$

$$\Pr[X \leq \frac{n}{2}] = 1 - \Pr[X > \frac{n}{2}] = e^{-\frac{n}{2p}(p-\frac{1}{2})^2}$$

In our case, p is 0.6, so we want this quantity to be smaller than 0.001:

$$\Pr[X \leq \frac{n}{2}] = e^{-\frac{n}{2\times 0.6}(0.6-\frac{1}{2})^2}$$

Then we need to solve the following:

$$e^{-\frac{n}{1.2}0.01} = 0.001$$

$$n = -\ln(0.001)\frac{1.2}{0.01}$$

Solving this yields $n = 828.9$. So if we let $k = 829$, then we will certainly output a ϵ-accurate with probability greater than 0.999.

Note: There are other variants of the Chernoff Bound that you can work with, and they could give you varying results. All such results will still be valid. Most variants derive from the Moment-Generating-Function form, and give some leeway so that the formula looks nice.

2 Universal Hashing

The definition of a universal hash family that we use in this class is that of pairwise independence: for every fixed pair, $\langle x, y \rangle$ of keys where $x \neq y$, and for any h chosen uniformly at random from the hash family, the pair $\langle h(x), h(y) \rangle$ is equally likely to be any of the m^2 pairs of elements from $\{0, 1, \ldots, m-1\}$. (The probability it taken only over the random choice of the hash function.)

A weaker definition of universality (not used in this course) is as follows: a family $\mathcal{H}$ is "weakly-universal" if for every pair of keys $\langle x, y\rangle$ such that $x \neq y$, for any h chosen uniformly at random from $\mathcal{H}$, $\Pr_{h\in\mathcal{H}}[h(x) = h(y)] = \frac{1}{m}$.

(a) Show that, if $\mathcal{H}$ is universal, then it is weakly-universal.

Solution: If $\mathcal{H}$ is universal, then for every pair of distinct keys x and y, and for every $i \in \{0, 1, \ldots, m-1\}$,

$$\Pr_{h\in\mathcal{H}}[\langle h(x), h(y)\rangle = \langle i, i\rangle] = \frac{1}{m^2}$$

There are exactly m possible ways for us to have x and y collide, i.e., $h(x) = h(y) = i$ for $i \in \{0, 1, \ldots, m-1\}$. Thus,

$$\Pr_{h\in\mathcal{H}}[h(x) = h(y)] = \sum_{i=0}^{m-1}\left(\Pr_{h\in\mathcal{H}}[\langle h(x), h(y)\rangle = \langle i, i\rangle]\right) = \frac{m}{m^2} = \frac{1}{m}$$

Therefore, by definition, $\mathcal{H}$ is weakly-universal.

(b) Suppose that an adversary knows the hash family $\mathcal{H}$ and controls the keys we hash, and the adversary wants to force a collision. In this problem part, suppose that $\mathcal{H}$ is weakly-universal. The following scenario takes place: we choose a hash function h randomly from $\mathcal{H}$, keeping it secret from the adversary, and then the adversary chooses a key x and learns the value $h(x)$. Can the adversary now force a collision? In other words, can it find a $y \neq x$ such that $h(x) = h(y)$ with probability greater than $1/m$?

Solution: We can construct a scenario where the adversary can force a collision. On a universe $\mathcal{U} = \{x, y, z\}$, consider the following family $\mathcal{H}$:

	x	y	z
h_1	0	0	1
h_2	1	0	1

$\mathcal{H}$ is a weakly-universal hash family: x and y collide with probability $1/2$, x and z collide with probability $1/2$, and y and z collide with probability $0 < 1/2$.

The adversary can determine whether we have selected h_1 or h_2 by giving us x to hash. If $h(x) = 0$, then we have chosen h_1, and the adversary then gives us y. Otherwise, if $h(x) = 1$, we have chosen h_2 and the adversary gives us z.

(c) Consider the hash family containing $h_{a_1,a_2,a_3}(x_1, x_2) = a_1x_1 + a_2x_2 + a_3x_1 \mod m$ for every $a_1, a_2, a_3 \in \{0, ..., m-1\}$, where m is prime and $x_1, x_2 \in \{0, ..., m-1\}$. Show that this family is weakly-universal.

Solution: Any function in this family can be rewritten as $h_{a_1,a_2,a_3}(x_1, x_2) = (a_1+a_3)x_1+a_2x_2 = a'x_1+a_2x_2 \mod m$, where $a' = a_1+a_3 \mod m$. Every value of a' occurs with the same frequency, so the family can be rewritten as $h_{a',a_2}(x_1, x_2) = a'x_1 + a_2x_2 \mod m$, which we already know to be weakly-universal from discussion.

(d) Show that the family from the previous part is not universal (not pairwise independent).

Solution: Observe the following two inputs: $(0, 0)$ and $(1, 0)$. Then for any hash function in the family, $h_{a_1,a_2,a_3}(0, 0) = a_1 * 0 + a2 * 0 + a_3 * 0 = 0$ and $h_{a_1,a_2,a_3}(1, 0) = a_1 + a_3 \mod m$. The probability $Pr[h(0, 0) = 1, h(1, 0) = 1] = 0 \neq \frac{1}{m^2}$ because it is impossible for $h(0, 0)$ to equal one for any h in the hash family. Thus, the family is not universal.