ESO207A: Data Structures and Algorithms

Practice Set *9: Hash Functions* –

Problem 1. Let $[m]$ denote the set $\{0, 1, 2 \ldots, m-1\}$. For each of the following families of hash functions, say whether or not it is universal, and determine how many random bits are needed to choose a function from the family.

1. $H = \{h_{a_1,a_2} \mid a_1, a_2 \in [m]\}$, where m is a fixed prime and

$$h_{a_1,a_2}(x_1, x_2) = a_1 x_1 + a_2 x_2 \mod m \ .$$

   Here $x_1, x_2 \in \{0, \ldots m-1\}$.

2. H is the same as before except that $m = 2^k$, for some fixed positive integer k.

3. H is the set of all functions $f : [m] \to [m-1]$.

Solution. (1) is universal. The proof is very similar to what has been done in the class (where we used a four tuple). To repeat the argument, we have to calculate the probability that $h(x_1, x_2) = h(y_1, y_2)$, given that $(x_1, x_2) \neq (y_1, y_2)$. Assume that $x_2 \neq y_2$. Then, the above equation is

$$a_1 x_1 + a_2 x_2 = a_1 y_1 + a_2 y_2 \mod m$$

or, that

$$a_1(x_1 - y_1) = a_2(y_2 - y_1) \mod m \ .$$

Since, m is prime, and $y_2 \neq y_1$, $y_2 - y_1$ has an inverse modulo m, which fixes a_2 in the above equation (modulo m). Thus, $\Pr(h(x_1, x_2) = h(y_1, y_2)) = \frac{1}{m}$, since, a_1 can take any value, but that fixes exactly 1 out of m values of a_2. Hence, the probability (if a_1, a_2 are chosen randomly) that $h(x_1, x_2) = h(y_1, y_2)$ is $\frac{1}{m}$.

(2) Now suppose $m = 2^k$. Then, $y_2 - y_1$ if non-zero need not have an inverse, if it is even (for example). So H is not universal.

(3) H is the set of all functions from $[m]$ to $[m-1]$. So, for $x \neq y$, $h(x) = h(y)$ has how many solutions? Since, $h(x) = h(y)$, x, y together get mapped to the same value in $m-1$ ways. Every other value $z \in [m] \setminus \{x, y\}$ gets mapped in $m-1$ ways. Total is $(m-1)^{m-2+1} = (m-1)^{m-1}$. The possible number of hash functions, that is, the size $|H| = (m-1)^m$. Hence, the probability is

$$\Pr(h(x) = h(y)) = \frac{\text{Number of hash functions s.t. } h(x) = h(y)}{\text{Total number of hash functions}} = \frac{(m-1)^{m-1}}{(m-1)^m} = \frac{1}{m-1}$$

and hence is universal.