Tutorial 8

Exercise 1 (Light revision)
Matt is practicing basketball and he wants to practice free throws. As he is a beginner, the probability of succesfully scoring on a free throw is $p \in (0, 1]$ and it is independent on all his previous attempts. Let X be a random variable that denotes the number of attempts made until the first time Matt scores (including the last attempt when he scored). Show that $\mathbb{E}[X] = \frac{1}{p}$.

Exercise 2 (Collision probability)
Show that in a hash-table of size $m = n^2$ with n elements, the probability of a collision is at most $1/2$, if we assume the hashing function to be uniformly random.

Exercise 3 (Fixed points of permutations)
Let us have a uniformly random permutation on n elements. Compute the expected number of fixed points of the permutation.

Exercise 4 (Black box)
You are given a hash function $h : \mathcal{U} \to [m]$. If you do not know anything else about the function, how many evaluations of h do you need to always find a k-tuple of elements that share the same bucket?

Exercise 5 (Independence and universality)
Prove the following:

- if a hashing system is (k, c)-independent, it is also $(k-1, c)$-independent (for $k \geq 2$),
- if a hashing system is $(2, c)$-independent, it is also c-universal.

Exercise 6 (Truly practical systems)
Let us consider the hashing function system $\mathcal{H}_1 = \{\text{id}\}$ that contains just one function, the identity that maps x to x. Is $\mathcal{H}_1$ c-universal for some c? Is $\mathcal{H}_1$ (k, c)-independent for some k and c?
Next, consider the system $\mathcal{H}_2 = \{h_a(x) = a \colon a \in [m]\}$. Prove that this system is (1,1)-independent. Next, show that $\mathcal{H}_2$ is neither $(2, c)$-independent nor c-universal for any c.

Exercise 7 (Modulo of a universal system does not need to be universal)
Show that, if we have a universal system of hash functions $\mathcal{H}$, then the system $\mathcal{H}'$, where each function is computed modulo m, does not have to be universal. Formally: Show that for every c and $m > c$, there exists a universe $\mathcal{U}$ and a system $\mathcal{H}$ from $\mathcal{U}$ to $\mathcal{U}$ such that $\mathcal{H}$ is universal but $\mathcal{H}'$ is not c-universal.

Useful notions

Proposition (Union bound)**.** For elements A_1, A_2, we have $\Pr[A_1 \cup A_2] \leq \Pr[A_1] + \Pr[A_2]$.

Proposition (Linearity of expectation)**.** For random variables X, Y and coefficients $\alpha, \beta \in \mathbb{R}$, we have $\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y]$.

Definition (Indicator, independence of random variables)**.** Let A be an event in a discrete probability space. Then the indicator of A is a random variable I_A defined as: $I_A(\omega) = 0 \Leftrightarrow \omega \notin A$, otherwise $I_A(\omega) = 1$.
Random variables X, Y on a discrete probability space $(\Omega, 2^\Omega, P)$ are independent, of $\forall \alpha, \beta \in \mathbb{R}$, the events $\{\omega \in \Omega : X(\omega) \leq \alpha\}, \{\omega \in \Omega : Y(\omega) \leq \beta\}$ are independent.

Definition (c-universal function system)**.** A system $\mathcal{H}$ of functions $h : \mathcal{U} \to [m]$ is c-universal for $c > 0$, if for all $x \neq y$, it holds that $\Pr_{h \in \mathcal{H}}[h(x) = h(y)] \leq \frac{c}{m}$.
A system $\mathcal{H}$ is universal, if it is c-universal for some $c > 0$.

Definition (k-independent function system)**.** A system $\mathcal{H}$ of functions $h : \mathcal{U} \to [m]$ is (k, c)-independent for some $k \geq 1, c > 0$, if $\Pr_{h \in \mathcal{H}}[h(x_1) = a_1 \wedge \ldots \wedge h(x_k) = a_k] \leq \frac{c}{m^k}$ for any pairwise distinct $x_1, \ldots, x_k$ and any not necessarily distinct $a_1, \ldots, a_k$.
A system $\mathcal{H}$ is k-independent, if it is (k, c)-independent for some fixed constant c.