Tutorial 8

Data Structures 1, 11. 4. 2025 https://iuuk.mff.cuni.cz/~chmel/2425/ds1en/

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 $\mathbb{E}[X] = \frac{1}{p}$.

Solution
Using the memory-less property: $\mathbb{E}[X] = p + (1-p)(1 + \mathbb{E}[X]) = 1 + (1-p)\mathbb{E}[X] \rightsquigarrow p\mathbb{E}[X] = 1 \rightsquigarrow \mathbb{E}[X] = \frac{1}{p}$.

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.

Solution
$P[\text{collision}] = P[\exists i \neq j \in [n] : h(i) = h(j)] = P[\bigcup_{i \neq j \in [n]}(h(i) = h(j))] \leq \sum_{i \neq j \in [n]} P[(h(i) = h(j))] = \binom{n}{2} \cdot \frac{1}{m}$

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.

Solution
We use indicators: if F is a random variable that denotes the number of fixed points of the random permutation, we can write $F = I_1 + I_2 + \ldots + I_n$, where I_ℓ is the indicator of the event that $\pi(\ell) = \ell$. Then $\mathbb{E}[F] = \mathbb{E}[I_1 + I_2 + \ldots + I_n] = \mathbb{E}[I_1] + \mathbb{E}[I_2] + \ldots + \mathbb{E}[I_n]$ by the linearity of expectation, and $\mathbb{E}[I_\ell] = \frac{(n-1)!}{n!} = \frac{1}{n}$, and thus $\mathbb{E}[F] = 1$.

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?

Solution
Pigeonhole principle: we have m holes, in each $k-1$ pigeons, and having one more attempt guarantees, that we really get a k-tuple, thus the number of attempts is $1 + m(k-1)$.

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.

Solution

- We want to show $(k-1, c)$-independence, so we have given $x_1, \ldots, x_{k-1} \in \mathcal{U}, a_1, \ldots, a_{k-1} \in [m]$. Next, we choose $x \neq x_i \forall i \in [k-1]$ (such x exists from k-independence). We then compute $\Pr_h[h(x_1) = a_1 \wedge \ldots \wedge h(x_{k-1}) = a_{k-1}] = \sum\limits_{a \in [m]} \Pr_h[h(x_1) = a_1 \wedge \ldots \wedge h(x_{k-1}) = a_{k-1} \wedge h(x) = a] \leq \sum\limits_{a \in [m]} \frac{c}{m^k} = \frac{c}{m^{k-1}}$.
- Let us have $x \neq y \in \mathcal{U}$. We attempt to bound from above: $\Pr_h[h(x) = h(y)] = \sum\limits_{a \in [m]} \Pr_h[h(x) = a \wedge h(y) = a] \leq \sum\limits_{a \in [m]} \frac{c}{m^2} = \frac{c}{m}$.

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.

Solution
$\mathcal{H}_1$ is ε-universal for every $\varepsilon > 0$. The problem is that $\Pr[h(x) = x] = 1$ and therefore, if $|\mathcal{U}| > 1$, it can never be independent.
For the second system: (1,1)-independence follows from the fact that $\Pr[h_a(x) = b] = \frac{1}{m}$, as we only ever randomly choose a. On the other hand, for $x \neq y$ we have $\Pr[h_a(x) = b \wedge h_a(y) = b] = \frac{1}{m} > \frac{c}{m^2}$ for any constant, and this the system is not 2-independent. It is also clear that $\Pr[h_a(x) = h_a(y)] = 1$ and thus c-universality is not satisfied either.

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.

Solution
Consider $\mathcal{H}_1 = \{\text{id}\}$ from the previous exercise, then $\mathcal{H}_1 \mod m$ cannot be c-universal as for $m < |\mathcal{U}|$, elements 1 and $m + 1$ will always map to the element 1.

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.