CS3230: Assignment for Week 6 Solutions

Due: Sunday, 27th Feb 2022, 11:59 pm SGT.

---

1. As in the naive algorithm, we begin by comparing $P[1\,..\,m]$ and $T[1\,..\,m]$, character by character. If these two strings are the same, then we can proceed to compare $P[1\,..\,m]$ with $T[m{+}1\,..\,2m]$; this is because $P[1\,..\,m]$ cannot be the same as $T[i{+}1\,..\,i{+}m]$ for any $i < m$ due to the assumption that all characters in P are different (in particular, since $T[i+1] = P[i+1]$ and $P[i+1] \neq P[1]$, it must be that $T[i+1] \neq P[1]$). Similarly, if $P[1\,..\,m]$ and $T[1\,..\,m]$ are not the same, then assume that $P[j] \neq T[j]$ is the first mismatch. If $P[1] = T[j]$, we can compare $P[1\,..\,m]$ with $T[j\,..\,j+m-1]$, while if $P[1] \neq T[j]$, we can compare $P[1\,..\,m]$ with $T[j+1\,..\,j+m]$. Since each character in T is compared at most twice, the algorithm runs in time $O(n)$.

2. Create a hash table of size $\Theta(n^2)$. For each pair $a \in A, b \in B$, insert $a+b$ into the hash table, resolving collision by chaining. For each $c \in C$, we then check whether $2022 + c$ is in the hash table; if so, we return Yes. Otherwise, if for every $c \in C$ we do not return Yes, we return No.

   Correctness is obvious: if $a + b - c = 2022$ for some $a \in A, b \in B, c \in C$, then $a + b = 2022 + c$ is in the table and we return Yes. On the other hand, if $a + b - c \neq 2022$ for all a, b, c, then for no c is it the case that $2022 + c$ is in the table, and we return No.

   For the running time, since we insert $O(n^2)$ elements into the hash table, recall from Claim 5.3.2 in the lecture notes that if the table has size $\Theta(n^2)$, then the expected number of collisions for each inserted element is $O(1)$ (the expectation is over the choice of hash function drawn from a universal family). Hence, the algorithm runs in expected time $O(n^2)$.

3. The proof is by induction. For the base case $k = 1$, $X = x_1$ is indeed uniformly distributed over the only number you have seen. For the inductive step, suppose that after you have seen $x_1, \ldots, x_{k-1}$, X is uniformly distributed over the $k-1$ numbers. Upon the arrival of x_k, $X = x_k$ with probability $\frac{1}{k}$, while for each $1 \leq i < k$, $X = x_i$ with probability $\left(1 - \frac{1}{k}\right) \cdot \frac{1}{k-1} = \frac{1}{k}$. Hence, X is uniformly distributed over the k numbers, completing the induction.