CPS815/CP8201 - Assignment 4

1. (50 marks) Let A be an array containing n integers. An element a in A is said to be a majority element if it appears in at least half of the entries of A. For example, 3 is a majority element in the following array.

1	3	3	6	3	8	3	2	3	5	3	3

   The *majority problem* for A is to decide whether there exists a majority element in A. Suppose there is an algorithm T that can solve the majority problem for any given array with probability $p = 1/2 + 1/100$. That means T will output the correct answer only with probability p. Write a randomized algorithm that uses T to solve the majority problem with high probability, say with with probability $\geq 1 - 2^{-20}$. The input to the algorithm is an integer array A of length n. The output is 1 if there exists a majority element in A, otherwise the output is 0. Your algorithm should call T only a **constant** number of times.

   Hint: Run T with input A for a constant number of times, say c times, and then decide according to the majority of T's outputs. Then argue, using the Chernoff bound, that your output is correct with high probability. You should choose c large enough to end up with success probability $\geq 1 - 2^{-20}$.

2. (50 marks) Recall the definition a universal hash family from the lectures. You can also find details in Section 13.6 of the textbook. Denote by $\mathbb{Z}_p$ the set of numbers mod p, that is $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$ with operations done modulo p. An $n \times m$ matrix M over $\mathbb{Z}_p$ is a matrix with n rows and m columns with entries from $\mathbb{Z}_p$. Similarly, a vector v of length m over $\mathbb{Z}_p$ is a vector of m elements from $\mathbb{Z}_p$. Note that the matrix-vector product Mv is a vector of length n over $\mathbb{Z}_p$.

   Assume $m \geq n + 1$. Let $\mathcal{M}_{n,m}$ be the set of $n \times m$ matrices over $\mathbb{Z}_p$, and let $\mathcal{V}_k$ be the set of vectors of length k over $\mathbb{Z}_p$. Then for any matrix M from $\mathcal{M}_{n,m}$ we can construct a hash function $f_M : \mathcal{V}_m \to \mathcal{V}_n$ defined by

$$f_M(v) = Mv.$$

   Therefore, the hash function f_M maps vectors of length m to vectors of length n. Show that the set $\mathcal{H} = \{f_M \mid M \in \mathcal{M}_{n,m}\}$ is a universal family of hash functions.

   Hint: the proof is similar to the one in the book/lecture. In there, we had the hash function f_v where v was a vector.