Algorithms, Winter 2020 at CIS

Homework 2 **Due: 2/2/20 11:59pm China time**

---

1. What is an algorithm with the smallest number of comparisons you can find for outputting both the maximum and the minimum of n numbers?

2. Suppose $\mathcal{H}$ is a 3-universal family of hash functions h, where each function $h \in \mathcal{H}$ is a mapping from $\{1, 2, \ldots, M\}$ to $\{1, 2, \ldots, N\}$. Show that $\mathcal{H}$ is also a 2-universal family of hash functions.

3. Consider the following data stream algorithm for estimating counts of items. We choose a universal hash function $h : \Sigma \to \{0, 1, 2, \ldots, r-1\}$ and a 2-universal hash function $s : \Sigma \to \{-1, 1\}$. We initialize an array A of length r to be all zeros. When we see the stream item a_i, we set $A[h(a_i)] = A[h(a_i)] + s(a_i)$. At the end of the stream, for an element $e \in \Sigma$, suppose we output $\text{est}(e) = s(e) \cdot A[h(e)]$.

   (a) Show that the expected value $\mathbf{E}[\text{est}(e)] = \text{count}(e)$, where $\text{count}(e)$ is the number of occurrences of e in the stream.

   (b) Recall the variance $\mathbf{Var}[X]$ of a random variable X is $\mathbf{E}[(X - \mathbf{E}[X])^2]$, which is also equal to $\mathbf{E}[X^2] - \mathbf{E}^2[X]$. The smaller the variance is, the more likely your random variable is to be close to its expected value. Show $\mathbf{Var}[\text{est}(e)] \leq \frac{1}{r} \sum_{e' \in \Sigma} \text{count}^2(e')$. It might help to use that for independent random variables X and Y, it holds that $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.