COMP5408: Fall 2021 — Assignment 2

Please write up your solutions on paper (word processed in LaTeX would be nice) and email them to me as a *single PDF file*.

1. Suppose we have a probability distribution $p_1,\dots,p_n$ where each p_i is of the form 2^{-x_i} and x_i is an integer. Describe a data structure that allows us to generate random numbers according to this distribution using only coin flips. The expected number of coin flips required to generate a number should be $H(p_1,\dots,p_n)$.

2. Suppose you have access to a random number generator that allows you to generate integers uniformly at random in $\{0,\dots,W-1\}$ in constant time. You are given a probability distribution $p_1,\dots,p_n$, where each p_i is of the form k_i/W, for some integer k_i. Describe a data structure that allows you to generate random numbers according to this distribution. Your data structure should be as fast as possible and use as little space as possible. Ideally, you could generate each sample in constant time using a data structure that uses $O(n)$ space.

3. The w-bit RAM model is the model of computer you're used to, that allows constant time operations on binary integers containing w bits. There exists a data structure of size $O(n)$ for the w-bit RAM model that makes it possible to search an n-element set S of w-bit integers in $O(\log_w n)$ time.

   Using the structure mentioned in the previous paragraph, and anything we've seen in class, show how this makes it possible to have a data structure of size $O(n)$ that stores an n-element set of w-bit integers and has search time $O(\sqrt{\log n})$.

4. Let S be a set of n strings of total length M and let $p_1,\dots,p_n$ be a probability distribution over these strings.

   (a) Show how to build a ternary trie on S so that the time to search for the ith string $s \in S$ is $O(|s| + \log(1/p_i))$.

   (b) Using this data structure, what is the expected cost of searching for a random string drawn according to the distribution $p_1,\dots,p_n$?