Homework Assignment 11 is due Friday April 30

Exercise 1: Given a stream of n positive real numbers $a_1, a_2, \ldots a_n$, upon seeing $a_1, a_2, \ldots a_i$ keep track of the sum $a = a_1 + a_2 + \ldots + a_i$ and a sample $a_j, j \le i$ drawn with probability proportional to its value. On reading a_{i+1}, with probability $\frac{a_{i+1}}{a + a_{i+1}}$ replace the current sample with a_{i+1} and update a. Prove that the algorithm selects an a_i from the stream with the probability of picking an element being proportional to its value.

Exercise 2: Given a stream of symbols $s_1, s_2, \ldots, s_n$, give an algorithm that will select one symbol uniformly at random from the stream. How much memory does your algorithm require?

Exercise 3: How would one pick a random word from a very large book where the probability of picking a word is proportional to the number of occurrences of the word in the book.

Exercise 4: Suppose $a_1, a_2, \ldots, a_m$ are non-negative reals. Show that the minimum of $\sum_{k=1}^{m} \frac{a_k}{x_k}$ subject to the constraints $x_k \ge 0; \sum_k x_k = 1$ is attained when x_k are proportional to $\sqrt{a_k}$.

Exercise 5: Show that for a 2-universal hash family $\Pr\left(h(x) = z\right) = \frac{1}{M+1}$ for all $x \in \{1, 2, \ldots m\}$ and $z \in \{0, 1, 2, \ldots M\}$.

Exercise 6: Generate two 100 by 100 matrices A and B with integer values between 1 and 100. Compute the product AB both directly and by sampling. Plot the difference in L_2 norm between the results as a function of the number of samples.