Homework 1 for CMSC 498U/644

Due 02/14

February 7, 2018

1 Problem 1

Recall in class, we talk about an example of Bonferroni's Principle. The following assumptions are made:

1. There are one billion people who might be evil-doers.
2. Everyone goes to a hotel one day in 100.
3. A hotel holds 100 people, Hence, there are 100,000 hotels - enough to hold the 1% of a billion people who visit a hotel on any given day.
4. We shall examine hotel records for 1000 days

To find evil-doers, we look for people who, on two different days, were both at the same hotel. Using this information, what would be the number of suspected pairs if the following changes were made to the data (and all other numbers remained as they were in the section)?

(a) The number of days of observation was raised to 2000.

(b) The number of people observed was raised to 2 billion (and there were therefore 200,000 hotels).

(c) We only reported a pair as suspect if they were at the same hotel at the same time on three different days.

2 Problem 2

We covered the Bloom Filter in class. Consider the following bloom filter:

1. An array a of n cells, initially all 0's.
2. One hash function h that maps "key" value to n buckets, corresponding to the n cells of the array.
3. A set S of n key values.

On seeing a value x, we compute $h(x)$, and set the corresponding cell $a[h(x)]$ to be one. What is the probability that a certain cell remains zero after we have seen all the n key values?

3 Problem 3

Please use your favourite programming language (including but not limited to C, C++, Java, Python, etc.) to implement the following task:

Read a list of numbers (one number each line), and calculate a hash function h on each of them. Count the number of occurance of each possible hash output. For example, we have a list (34, 58, 32, 11, 57, 28), and the hash function is $h(x) = x \mod 5$, and we output

```
0: 0
1: 1
2: 2
3: 2
4: 1
```

Since $h(34) = 4, h(58) = 3, h(32) = 2, h(11) = 1, h(57) = 2, h(28) = 3$, and we get zero 0s, one 1, two 2s, two 3s, and one 4.

Implement the task for the following cases:

1. $h(x) = ax + b \mod p$, where $a = 3, b = 1, p = 11$
2. $h(x) = ax + b \mod p$, where $a = 3, b = 1, p = 8$

There are two input file: `input_1.txt`, and `input_2.txt`. Please run your program (two possible hash functions) on both inputs, and look at the distribution you output. Think of why you would need a prime. In your solution, you need: your source code and the four output file.