Bloom filter

Introduction

In this project, I implemented Bloom filters in Java and compared their theoretical False Positive rate with the achieved results, for different values of c and n.

Expected FP rate is calculated as (1-exp(-k/c)).

I've defined my Universe as all numbers from 0 to Max_Integer/2. The same was used for all calculations.

FP rate is the number of FPs/ (FP + TN)

Graphs

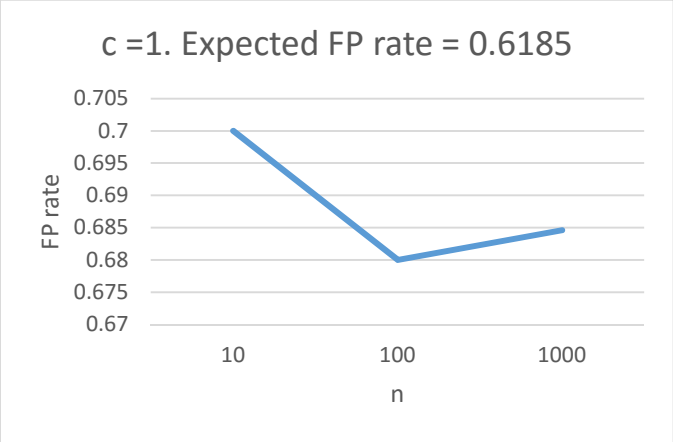

Figure 1: c is set to 1.

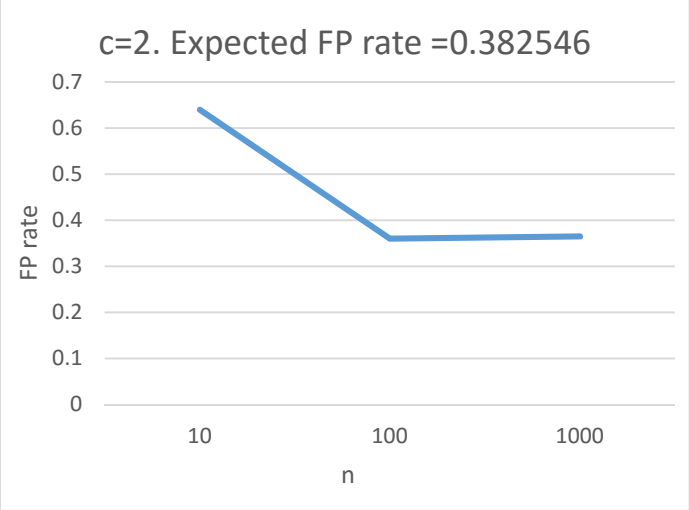

Figure 2: c is set to 2.

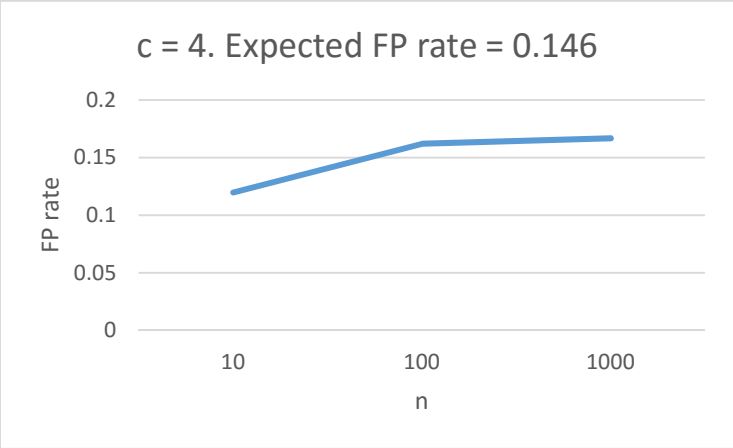

Figure 3: c is set to 4.

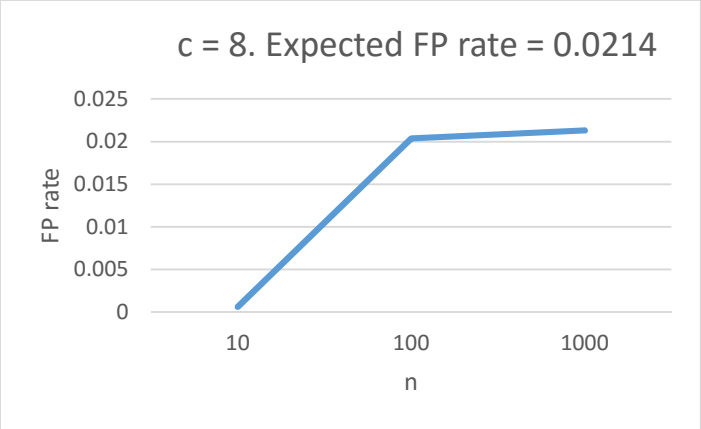

Figure 4: c is set to 8.

Conclusions:

In most cases, as n increased, the fp rate reached the theoretical rates. When c>=10, the false positive rate was close to 0. This implementation of Bloom filter achieved results close to the theoretical claims.

Bloom filter proves to be a very useful data structure when false positives are permissible but false negatives are not.