



UNIVERSITÀ DI PISA

Fifth hands-on: Bloom Filters

Algorithm Design (2021/2022)

Gabriele Pappalardo

Email: g.pappalardo4@studenti.unipi.it

Department of Computer Science

March 2022

1 Introduction

The problem is composed in two parts:

1. Consider the Bloom filters where a single random universal hash random function $h : U \rightarrow [m]$ is employed for a set $S \subseteq U$ of keys, where U is the universe of keys. Consider its binary array B of m bits. Suppose that $m \geq c|S|$, for some constant $c > 1$, and that both c and $|S|$ are unknown to us. Estimate the expected number of 1s in B under a uniform choice at random of $h \in H$. Is this related to $|S|$? Can we use it to estimate $|S|$?

2. Consider B and its rank function: show how to use extra $O(m)$ bits to store a space-efficient data structure that returns, for any given i , the following answer in constant time: $\text{rank}(i) = \#1s \in B[1..i]$

Hint: Easy to solve in extra $O(m \log m)$ bits. To get $O(m)$ bits, use prefix sums on B , and sample them. Use a lookup table for pieces of B between any two consecutive samples.

2 Solution

TODO