## HW3作业：BF实验报告

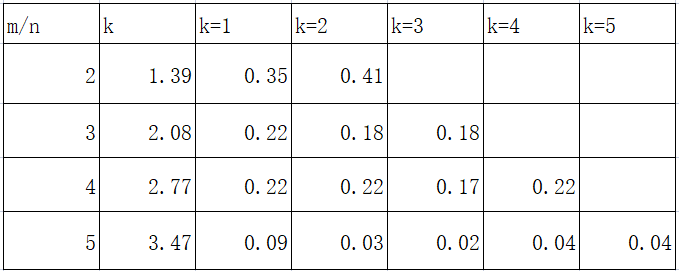
1. 实验思路：

本次实验的思路主要来源于课件PPT，通过构造一个int类型的数组来模拟bit-数组，并通过一个或多个hash函数来模拟BF实际的数据存储与查找的过程。

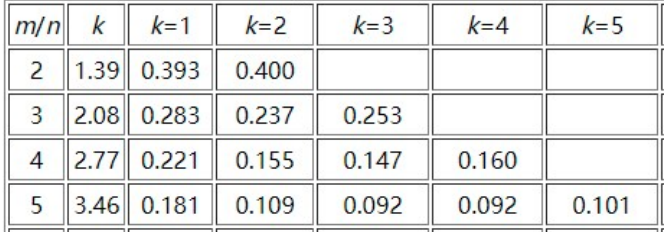
1. 实验过程

本次实验选用的数据集为助教已经给定的输入数据集与测试数据集。具体的实现过程即通过自定义的基于std::hash的k个hash函数在一个长度为m的大数组上进行数据的查找与插入。而在此过程中，遇到的最大的问题即为hash函数的选取。通过很多次的测试，最终选择了

NumofHash \* MAX\_SIZE作为相邻hash函数的偏移量来构造相互关联的NumofHash个hash函数。之后获得的具体结果如下：



而课件给出的理论值如下：



通过上述理论值与实际值的对比可以发现，两者之间的数据变化趋势是基本相同的，但是仍然发现部分数据（如m/n = 3时的数据和m/n=4时的数据）并不完全符合理论数值，并且m/n=5时的数值明显偏小于理论数值。原因可能为：

Std自定义的hash函数大多是以桶长度为依据进行的，而不同的桶长度就可能导致不同的结果索引值，而这就可能导致此hash函数可能会使得测试使用的数据分散的并不是特别均匀，从而使得数据出现一些偏差。