Homework of Week 6

Deadline: 9:00am, December 3 (Thursday), 2020

1. Bloom filters can be used to estimate set differences. Suppose Alice has a set X and Bob has a set Y, both with n elements. For example, the sets might represent their 100 favorite songs. Alice and Bob create Bloom filters of their sets respectively, using the same number of bits m and the same k hash functions. Determine the expected number of bits where our Bloom filters differ as a function of m, n, k and $|X \bigcap Y|$. Explain how this could be used as a tool to find people with the same taste in music more easily than comparing lists of songs directly.

2. Suppose there is a set of size m. Consider two approaches to hashing this set. One is a Bloom filter with n bits and $k = \frac{n}{m} \ln 2$ hash functions. The other is k independent Bloom filters, each having n' bits and 1 hash function. Choose n' such that the probabilities of false positive of the two approaches are equal. Compare n and kn'.

3. Do Bernoulli experiment for 20 trials, using a new 1-Yuan coin. Record the result in a string $s_1 s_2 ... s_i ... s_{20}$, where s_i is 1 if the i^{th} trial gets Head, and otherwise is 0.