Homework of Week 7

Deadline: 9:00am, November 11 (Thursday), 2021

1. (**Optional. Bonus score 5 points**) Let $X_i^{(m)}, 1 \le i \le n$, be the load of bin i in the (m, n)-Bins&Balls model, and $Y_1^{(m)}, \cdots, Y_n^{(m)}$ are independent Poisson random variables each having expectation m/n. Assume that f is a nonnegative n-ary function. Prove that if $E[f(X_1^{(m)}, ...X_n^{(m)})]$ is monotonically decreasing in m, then $E[f(X_1^{(m)}, ...X_n^{(m)})] \le 2E[f(Y_1^{(m)}, ...Y_n^{(m)})]$.

2. Consider the probability that every bin receives exactly one ball when n balls are thrown randomly into n bins.

   - Give an upper bound on this probability using the condition-free Poisson approximation.
   - Determine the exact probability of this event.

3. Bloom filters can be used to estimate set differences. Suppose Alice has a set X and Bob has a set Y, both with m 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 n of bits 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.

4. 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'.

5. 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.