You are allowed to discuss with others but not allowed to use any references except the course notes and the books "Probability and Computing" and "Randomized Algorithms". Please list your collaborators for each question. This will not affect your marks. In any case, you must write your own solutions.

There are totally 60 marks, and the full mark is 50. This homework is counted 10% of the course. The extra marks will not be carried to other parts of the course. Please read the course outline for the late submission policy.

---

1. **Maximum Load**

   (10 marks) We throw $n^{0.99}$ balls into n bins, where each ball is thrown to a uniform random bin independently. Prove the best upper bound on k so that the maximum load of a bin is at most k with probability at least $1/2$. You can assume that n is large enough to establish your bound (i.e. n is larger than some big constant).

2. **Coupon Collectors**

   (10 marks) Now we know that it requires $\Theta(n \log n)$ coupons to collect n different types of coupons, with at least one coupon per type. We wonder whether it would be more efficient if a group of k people cooperate, such that each person buys cn coupons and as a group they have at least k coupons per type (so that everyone gets a complete set of coupons).

   Prove the best bound you can on k to ensure that, with probability at least 0.9, each person only needs to buy at most $10n$ coupons. You will get full marks if k is of the correct order in terms of n.

3. **3-Wise Independent Bits**

   (10 marks) Consider the following modification of the construction of pairwise independent bits in L05. Let $X_1, X_2, \ldots, X_b$ be b independent uniformly random bits. We generate 2^{b-1} bits as follows. For each subset $S \subseteq \{1, 2, \ldots, b\}$ with $|S|$ *odd*, we generate a bit $Y_S = (\sum_{i \in S} X_i) \bmod 2$. Prove that these 2^{b-1} bits $\{Y_S \mid |S| \text{ is an odd number}\}$ are 3-wise independent.

4. **Maximum Load by Universal Family**

   (10 marks) We have shown that the maximum load when n items are hashed into n bins using a hash function chosen from a 2-universal family of hash functions is at most $\sqrt{2n}$ with probability at least $1/2$. Generalize this argument to k-universal hash functions. That is, find a value such that the probability that the maximum load is larger than that value is at most $1/2$. Then, find the smallest value of k such that the maximum load is at most $3 \ln n / \ln \ln n$ with probability at least $1/2$ when choosing a random hash function from a k-universal family.

5. **Streaming Sampling**

   (10 marks) Suppose we have a long sequence of numbers coming one at a time. We would like to maintain a set of numbers of size k with the property that for each number we have seen so far, the probability that the number appears in the set are equal.

   We want to accomplish this without knowing the total number of items in advance or storing all of the items that we have seen.

   Prove that the following simple algorithm works. When the first k numbers come, we put it in the set. After that when the m-th number appears, with probability k/m, we replace a random number in the set with the m-th number.

6. **Distinct Elements**

   (10 marks) In the distinct element algorithm in L06, we showed that with probability at least 2/3, we have $(1-\epsilon)D \le Y \le (1+\epsilon)D$ where Y is the answer that our algorithm returns. Suppose we run k independent copies of the algorithm in parallel and obtain the estimates $Y_1, \ldots, Y_k$ and return the median Y' of $Y_1, \ldots, Y_k$ as our answer. Prove that $(1-\epsilon)D \le Y' \le (1+\epsilon)D$ with probability $1-\delta$ when $k = O(\log \frac{1}{\delta})$.

7. ***k*-Wise Independence Bits** (optional; generalization of Q3)

   (No marks) Suppose that we are given m vectors $v_1, \ldots, v_m \in \{0,1\}^b$ such that any k of the vectors are linearly independent modulo 2. Let $v_i = (v_{i,1}, v_{i,2}, \ldots, v_{i,b})$ for $1 \le i \le m$. Let X be chosen uniformly at random from $\{0,1\}^b$. Let $Y_i = (\sum_{j=1}^{b} v_{i,j} X_j) \mod 2$. Show that the Y_i are uniform, k-wise independent random bits.