CS225: Probability and Computating
Shanghai Jiao Tong University

Homework 1
due Friday 8 July

## Problem 1

(Due to J. von Neumann)

(a) Suppose you are given a coin for which the probability of HEADS, say $\rho$, is *unknown*. How can you use this coin to generate unbiased (i.e. $\Pr[\text{HEADS}] = \Pr[\text{TAILS}] = 1/2$) coin-flips? Give a scheme for which the expected number of flips the biased coin for extracting one unbiased coin-flip is no more than $1/[\rho(1-\rho)]$.

(b) Devise an extension of the scheme that extracts the largest possible number of independent, unbiased coin-flips from a given number of flips of the biased coins.

## Problem 2

(a) Consider a sequence of $n$ flips of an unbiased coins. Let $H_i$ denote the absolute value of the excess of the number of HEADS over the number of TAILS seen in the first $i$ flips. Define $H = \max_i H_i$. Show that $\mathbb{E}[H] = \Theta(\sqrt{n})$.

(b) Suppose $\pi$ is a uniformly random permutation of the ordered set $N = \{1, 2, \ldots, n\}$ over all permutation. Let $L(\pi)$ denote the length of the longest increasing subsequence in permutation $\pi$. For large $n$ and some positive constant $c$, prove that $\mathbb{E}[L(\pi)] \geqslant c\sqrt{n}$. Is this bound tight?

## Problem 3

Consider the problem of deciding whether two integer *multiset* $S_1$ and $S_2$ are identical, we mean that each element occurs the same number of times in both sets. This problem can be solved by sorting the two sets in $O(n \log n)$ time, where $n$ is the cardinality of the multisets. Give a way of representing this as a problem involving a verification of a polynomial identity, and thereby obtain an efficient randomized algorithms. Discuss the relative merits of the two algorithms, keeping in mind issues such as the model of computation and the size of the integers being operated upon.

## Problem 4

Let $Q(x_1, x_2, \ldots, x_n)$ be a multivariate polynomial over a field $\mathbb{Z}_2$ with the degree sequence $(d_1, d_2, \ldots, d_n)$. A degree sequence is defined as follows: let $d_1$ be the maximum exponent of $x_1$ in $Q$, and

$Q_1(x_2, \dots, x_n)$ be the coefficient of $x_1^{d_1}$ in $Q$; then let $d_2$ be the maximum exponent of $x_2$ in $Q_1$, and $Q_2(x_3, \dots, x_n)$ be the coefficient of $x_2^{d_2}$ in $Q_1$; and so on.

Let $S_1, S_2, \dots, S_n \subseteq \mathbb{Z}_2$ be arbitrary subsets. For $r_i \in S_i$ chosen independently and uniformly at random, show that

$$\Pr\left[Q(r_1, r_2, \dots, r_n) = 0 \mid Q \not\equiv 0\right] \leqslant \sum_{i=1}^{n} \frac{d_i}{|S_i|}.$$

## Problem 5

Let $G = (V, E)$ be a simple graph and suppose that each vertex $v$ is associated with a set $S(v)$ of colors of size at least $10d$, where $d \geqslant 1$. Besides, for any $v$ and $c \in S(v)$, there are at most $d$ neighbors $u$ of $v$ such that $c$ lies in $S(u)$. Prove that there is a proper coloring of $G$ assigning to each vertex $v$ a color from its class $S(v)$.

## Problem 6

Let $G = (V, E)$ be a cycle of length $4n$ and let $V = V_1 \cup V_2 \cup \cdots \cup V_n$ be a partition of its $4n$ vertices into $n$ pairwise disjoint subsets, each of cardinality 4. Is it true that there must be an independent set of $G$ containing precisely one vertex from each $V_i$?

## Problem 7

Now you know that it requires $\Theta(n \log n)$ coupons to collect $n$ different types of coupons, with at least one coupon per type. You wonder whether it would be more efficient if a group of $k$ people cooperate, such that each person buys $cn$ coupons and exchange with each other.

Prove that 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.

## Problem 8

In Balls-and-Bins model, we throw $n$ balls independently and uniformly at random into $n$ bins, then the maximum load is $\Theta\left(\frac{\log n}{\log\log n}\right)$ with high probability. The two-choice paradigm is another way to throw $n$ balls into $n$ bins: each ball is thrown into the least loaded of two bins chosen independently and uniformly at random(it could be the case that the two chosen bins are exactly the same, and then the ball will be thrown into that bin), and breaks the tie arbitrarily. The maximum load of two-choice paradigm is $\Theta(\log\log n)$ with high probability, which is exponentially less than the previous one. This phenomenon is called *the power of two choices*.

Here are the questions.

(a) Consider the following paradigm: we still have $n$ balls and $n$ bins. The first $\frac{n}{2}$ balls are thrown into bins independently and uniformly at random. The remaining $\frac{n}{2}$ balls are thrown into bins using two-choice paradigm. What is the maximum load with high probability? You need to give an asymptotically tight bound (i.e. $\Theta(\cdot)$).

(b) Replace the above paradigm to the following: the first $\frac{n}{2}$ balls are thrown into bins using two-choice paradigm while the remaining $\frac{n}{2}$ balls are thrown into bins independently and uniformly at random. What is the maximum load with high probability in this case? You need to give an asymptotically tight bound (i.e. $\Theta(\cdot)$).

(c) Replace the above paradigm to the following: assume all $n$ balls are in a sequence. For every $1 \leqslant i \leqslant n$, if $i$ is odd, we throw $i$th ball into bins independently and uniformly at random, otherwise, we throw it into bins using two-choice paradigm. What is the maximum load with high probability in this case? You need to give an asymptotically tight bound (i.e. $\Theta(\cdot)$).