Homework 2 (due anytime before Christams)

Obs! Full credit for correct solutions to any seven problems

Exercises 1 and 2 can be solved using only a Second Moment Method. The same is true of **Q.3**, but there is an extra twist which makes the analysis considerably more complicated. So don't worry if you can't solve **Q.3**. Exercises 4-9 involve Chernoff bounds and emphasise CS applications.

Q.1. A set A of integers is called a *Sidon set*, if the sums $a_1 + a_2$, for $a_1, a_2 \in A$, are all distinct. Here $a_1 = a_2$ is allowed. So, for example, $A = \{0, 1, 3, 6\}$ is not a Sidon set, since $0 + 6 = 3 + 3$, whereas $A = \{0, 1, 3, 7\}$ is a Sidon set, since the 10 possible sums of two elements of A are all distinct. Indeed, $A + A = \{0, 1, 2, 3, 4, 6, 7, 8, 10, 14\}$.

Now let n denote a prime number[1] and $p \in [0, 1]$ a probability. Let $A = A(n, p)$ denote a random subset of $\mathbb{Z}_n$, the field of integers modulo n, where each number is chosen independently with probability p. This is a natural number theory analogue of the Erdős-Renyi model for random graphs. Determine, with proof, threshold functions $p = p(n)$ for the following events :

(i) "A contains no 3-term arithmetic progressions"

(ii) "A is not a Sidon set".

Q.2 With notation as above, let $X = X(n, p)$ be the random variable which denotes the cardinality of the sumset $A + A$. In this exercise, we suppose that $n^{-1} \ll p = p(n) \ll n^{-1/2}$.

Let $\mu_n = \mu(n, p) := \mathbb{E}[X]$. Show that

$$\mu_n \sim \frac{[n \cdot p(n)]^2}{2}, \quad \text{as } n \to \infty,$$

and that X is strongly concentrated about its mean in the following sense : for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}[(1 - \epsilon)\mu < X < (1 + \epsilon)\mu] = 1.$$

***Q.3** Now suppose that $p = p(n) = c \cdot n^{-1/2}$, for some fixed constant $c > 0$. Show that

$$\mu_n \sim g\left(\frac{c^2}{2}\right) \cdot n,$$

[1] I'd like to use p for primes, but I'm already using p for probability.

where $g : (0, \infty) \to (0, 1)$ is the function

$$g(x) = 1 - e^{-x}.$$

Show that X is also strongly concentrated about its mean, in the same sense as in **Q.2**.

Q.4 Let $X_1, \cdots, X_n$ be independent random variables chosen uniformly from the set $\{0, 1, 2\}$ and let $X := \sum_i X_i$. Derive a Chernoff bound on $\mathbb{P}(X > (1 + \delta)n)$ and $\mathbb{P}(X < (1 - \delta)n)$ for $0 < \delta < 1$.

Q.5 In class we introduced the random variable $W(k, p)$ as the number of trials required to get k successes with a coin of success probability p and derived a concentration result on this *negative Binomial* distribution by relating it to the Binomial distribution $B(n, p)$, for which we derived CH bounds. Write $W(k, p)$ as the sum of k independent random variables with a *geometric* distribution, and employ the method of the CH bounds by computing the exponential moment generating function $e^{\lambda W}$ explicitly. Choose a suitable λ and compare the resulting bound to the one you get via the relation to $B(n, p)$.

Q.6 A function $f : \mathbb{R} \to \mathbb{R}$ is *convex* if for all x, y and $0 \leq \lambda \leq 1$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y),$$

that is, the graph of f between x and y always lies below the line joining $(x, f(x))$ and $(y, f(y))$.

(a) Let Z be a random variable that takes on a finite set of values in $[0, 1]$ and let $p := \mathbb{E}[Z]$. Define the $0/1$ random variable X so that $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$. Show that $\mathbb{E}[f(Z)] \leq \mathbb{E}[f(X)]$ for any convex function f.

(b) Use this to give an alternate proof of the Hoeffding extension of the Chernoff bound.

Q.7 Suppose n balls are thrown independently and uniformly at random into n bins, and let $B_1, \cdots, B_n$ be the number of balls in each of the n bins. Let $L := \max_i B_i$ be the maximm number of balls in any bin.

(a) Show that, with high probability, $L = O(\log n))$, by a direct application of a CH bound.

(b) Show that, with high probability, $L = O(\frac{\log n}{\log \log n})$. You may need to use the following form of the Chernoff bound:

$$\mathbb{P}(X > (1 + \delta)\mathbb{E}(X)] < \left(\frac{e^{\delta}}{(1 + \delta)^{(1+\delta)}} \right)^{\mathbb{E}(X)}.$$

Q.8 In class, in the analysis of the randomized two phase bit fixing protocol for routing, we assumed unbounded queue sizes at each vertex. Give a high probability analysis for the maximum size of the queue at any vertex during the execution of the algorithm in the worst case.

Q.9 Modify the randomized median selection algorithm we discussed in class (Mitzenmacher-Upfal, p. 54) by replacing just the first step i.e. how the random sample is chosen: let R be formed by selecting each element of S independently with probability $p := n^{-1/4}$ (so R is a set not a multiset). Analyse the resulting algorithm using CH bounds. (Start by giving a high probability bound for the size of the set R etc.)