E0 234 : Homework 2

Deadline : 20th February 2023, 2pm

Instructions

- Please write your answers using LaTeX. Handwritten answers will not be accepted.
- You are forbidden from consulting the internet. You are strongly encouraged to work on the problems on your own.
- You may discuss these problems with another student. However, you must write your own solutions and must mention your collaborator's name. Otherwise, it will be considered as plagiarism.
- Academic dishonesty/plagiarism will be dealt with severe punishment.
- Late submissions are accepted only with prior approval (on or before the day of posting of HW) or a medical certificate.

1. *Bound on sum.*
   Show that there is a positive constant $\beta \in (0,1)$ such that the following holds. For n reals $a_1, a_2, \ldots a_n$ satisfying $\sum_i a_i^2 = 1$, if $(\alpha_1, \ldots, \alpha_n)$ is $\{+1, -1\}$-random vector obtained by choosing each α_i independently and uniformly at random to be $+1$ or -1, then:
   $$\mathbb{P}\left[\mid \sum_{i=1}^{n} \alpha_i a_i \mid \leq 1 \right] \geq \beta.$$

2. *Balls and bins.*
   Consider the following balls-and-bins process: We start by throwing n balls into n bins. From that point on, at each iteration we remove every bin that is occupied by the balls thrown at the previous iteration, and then throw n *new* balls into the remaining bins. The process ends when there are no more bins left. Show that the expected number of iterations performed by the process is $O(\log^* n)$.
   Hint. Break the process into several "sub-steps" as in the coupon collector problem.

3. Given an $n \times n$ matrix $A := \{a_{ij} \in \{+1, -1\}\}, 1 \leq i, j \leq n\}$ of lights (A may not be symmetric), and row switches x_i and column switches y_j (for $1 \leq i, j \leq n$), the objective is to find $x_i \in \{\pm 1\}, y_j \in \{\pm 1\}$ so as to maximize $|\sum_{i,j} a_{ij} x_i y_j|$. Show that there exists an A for which this maximum cannot be made larger than $cn^{3/2}$ for a suitable constant $c > 0$.

   (*Hint:* Chernoff bounds.)

4. *Independent Bounded Difference Inequality.*
   Consider the following variant of the bin packing problem. We are given n items and the item sizes are independent random variables $X_1, \ldots, X_n$ with uniform probability distribution on $[0, 1]$. Let B_n be the minimum number of unit capacity bins needed to pack all items, such that the sum of the sizes of the items in any given bin does not exceed its capacity.

   Show that B_n grows linearly in n, i.e., $\mathbb{E}[B_n] = \beta n$ for some positive constant $\beta \geq 1/2$.

   Also show the concentration around the expectation, i.e., $\mathbb{P}[|B_n - \beta n| \geq \epsilon n] \leq 2\exp\{-2\epsilon^2 n\}$. You may use the following inequality (McDiarmid's inequality):

Let $X := f(X_1, X_2, \ldots, X_n)$. If f is such that for all i, all $x_1, x_2, \ldots, x_n$ in the range of $(X_1, X_2, \ldots, X_n)$, and x, x' in the range of X_i,

$$|f(x_1, x_2, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n) - f(x_1, x_2, \ldots, x_{i-1}, x', x_{i+1}, \ldots, x_n)| \leq c_i,$$

then

$$\Pr[|X - \mathrm{E}[f(X_1, \ldots, X_n)]] \geq t] \leq 2\exp(-2t^2/\sum_i c_i^2).$$

5. *Chernoff bounds.*
In this exercise, we will derive the consequences of the following bound Chernoff bound shown in class.

Suppose $X_1, X_2, \ldots, X_n$ are independent 0-1 random variables, where each X_i is 1 with probability p_i. Let $X = \sum_i X_i$ and $p = (\sum_i p_i)/n$. Then, for $t \geq 0$ and $p + t \leq 1$, we have

$$\Pr[X \geq (p+t)n] \leq \exp(-\mathsf{KL}((p+t, 1-p-t), (p, 1-p))n); \tag{1}$$

where

$$\mathsf{KL}((p+t, 1-p-t), (p, 1-p)) = \left[\left(\frac{p}{p+t}\right)^{p+t} \left(\frac{1-p}{1-p-t}\right)^{1-p-t} \right] \tag{2}$$

is the Kullback-Leibler divergence of the two distributions $(p+t, 1-p-t)$ and $(p, 1-p)$. [Note that when $p + t = 1$, we take $((1-p)/(1-p-t))^{1-p-t} = 1$]

Assume $t = \delta p$, for $0 \leq \delta \leq 1$ (and $p + t < 1$). Show that

$$\Pr[X \geq (1+\delta)pn] \leq \exp(-\delta^2 pn/3);$$
$$\Pr[X \geq (1-\delta)pn] \leq \exp(-\delta^2 pn/2).$$

[In class, we derived (1) by letting $f(x) = \mathsf{KL}((p+x, 1-p-x), (p, 1-p))$ and showing that $f''(x) = 1/((p+x)(1-p-x)) \geq 4$. To show the above inequalities, note that $f''(x) \geq \max\{1/(p+x), 1/(1-p-x)\}$.]

Recommended practice problems (not for grading)

1. Let X be a random variable with expectation $\mathbb{E}[X] = 0$ and variance σ^2. Prove that for all $\lambda > 0$, $\mathbb{P}[X \geq \lambda] \leq \frac{\sigma^2}{\sigma^2+\lambda^2}$.
2. *Book:* **Mitzenmacher-Upfal (2nd edition): 3.11, 3.18, 3.24, 3.25, 4.17, 4.19, 4.21, 5.4, 5.11.**