Problem Set 2

Due: Noon, Monday September 26, email the pdf to toolkit2016homework@gmail.com

---

Homework policy: Exactly the same as last time.

Notational conventions: The notation $[n] = \{1, 2, \ldots, n\}$ is very standard in theoretical computer science. Some people like to use **boldface** to denote random variables; you might like to do this too.

---

1. (a) Let $\boldsymbol{X}$ be a random variable which is 1 with probability p and 0 with probability $1-p$. We "empirically estimate the mean of $\boldsymbol{X}$", by defining $\overline{\boldsymbol{X}} = \frac{1}{n}(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)$, where $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are independent copies of $\boldsymbol{X}$. We want to choose $n = n(\varepsilon, \delta)$ sufficiently large so that "$\overline{\boldsymbol{X}}$ is ε-accurate with δ-confidence", meaning $\mathbf{Pr}[|\overline{\boldsymbol{X}} - p| > \varepsilon] \le \delta$. Show that $n = O(\frac{1}{\varepsilon^2}\log(1/\delta))$ is sufficient (as $\varepsilon, \delta \to 0^+$).

   (b) Let $\boldsymbol{Y}$ be a random variable with a continuous probability distribution. We estimate the median of $\boldsymbol{Y}$ by defining $\boldsymbol{m} = \text{median}(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)$, where $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are independent copies of $\boldsymbol{Y}$. We wish to have

   $$\frac{1}{2} - \varepsilon \le \mathbf{Pr}[\boldsymbol{Y} \le \boldsymbol{m}] \le \frac{1}{2} + \varepsilon,$$

   except with probability at most δ. Again, show that $n = O(\frac{1}{\varepsilon^2}\log(1/\delta))$ is sufficient.

2. Prove the following "one-sided" version of Chebyshev's Inequality: If $\boldsymbol{X}$ is a random variable with $\mathbf{E}[\boldsymbol{x}] = \mu$ and $\mathbf{stddev}[\boldsymbol{X}] = \sigma > 0$, then for every $t > 0$,

   $$\mathbf{Pr}[\boldsymbol{X} \ge \mu + t\sigma] \le \frac{1}{t^2+1}.$$

   (Hint: Mimic the proof of Chebyshev's Inequality. "Standardize" $\boldsymbol{X}$, then prove and use the fact that $\frac{(x+1/t)^2}{(t+1/t)^2} \ge 1_{\{x \ge t\}}$.)

3. It is a basic fact of linear algebra that if we have m orthogonal unit-length vectors $\vec{u}_1, \ldots, \vec{u}_m$ in $\mathbb{R}^n$, then $m \le n$. (Recall that

   $$\text{“orthogonal”} \quad \Leftrightarrow \quad \measuredangle(\vec{u}_i, \vec{u}_j) = \pi/2 = 90^\circ \quad \Leftrightarrow \quad \vec{u}_i \cdot \vec{u}_j = 0,$$

   where $\vec{u}_i \cdot \vec{u}_j = \|\vec{u}_i\|\|\vec{u}_j\| \cos(\measuredangle(\vec{u}_i, \vec{u}_j))$ is the dot-product.)

   However, in this problem you will show the rather surprising fact that if we are willing for the unit vectors to only be *almost* orthogonal, we can have *exponentially* many such vectors.

   Suppose we define random vectors $\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \ldots, \vec{\boldsymbol{v}}_m \in \mathbb{R}^n$ by choosing every coordinate of each of the vectors to be ± 1 with probability $1/2$ each. Then we put $\vec{\boldsymbol{w}}_i = \vec{\boldsymbol{v}}_i/\sqrt{n}$ for each $i \in [m]$ so as to get unit vectors (meaning $\|\vec{\boldsymbol{w}}_i\| = 1$ for all $i \in [m]$).

(a) Suppose $i \neq j$. Let $\boldsymbol{\theta}_{ij} = \measuredangle(\vec{\boldsymbol{w}}_i, \vec{\boldsymbol{w}}_j)$. Show that

$$\mathbf{Pr}\big[|\cos \boldsymbol{\theta}_{ij}| \geq \delta\big] \leq \exp\big(-\Omega(\delta^2 n)\big).$$

Deduce

$$\mathbf{Pr}\big[|\pi/2 - \boldsymbol{\theta}_{ij}| \geq \delta\big] \leq \exp\big(-\Omega(\delta^2 n)\big).$$

(b) Show that even for some $m = \exp\big(\Omega(\delta^2 n)\big)$ we will have

$$\mathbf{Pr}\big[\pi/2 - \delta \leq \boldsymbol{\theta}_{ij} \leq \pi/2 + \delta \text{ for } \textit{all} \text{ pairs } i \neq j\big] \geq .99.$$

4. Let $\boldsymbol{X}$ be a random variable that is always nonnegative. Assume also that $\boldsymbol{X}$ only takes on finitely many different values.[1]

   (a) Prove

$$\mathbf{E}[\boldsymbol{X}] = \int_0^\infty \mathbf{Pr}[\boldsymbol{X} \geq t]\, dt.$$

   (b) Prove

$$\mathbf{E}[\boldsymbol{X}^2] = 2\int_0^\infty t\, \mathbf{Pr}[\boldsymbol{X} \geq t]\, dt.$$

5. Let $\boldsymbol{X}$ be a nonnegative random variable.

   (a) Prove that $\mathbf{Pr}[\boldsymbol{X} = 0] \leq \frac{\mathbf{Var}[\boldsymbol{X}]}{\mathbf{E}[\boldsymbol{X}]^2}$.

   (b) Prove that $\mathbf{Pr}[\boldsymbol{X} > 0] \geq \frac{\mathbf{E}[\boldsymbol{X}]^2}{\mathbf{E}[\boldsymbol{X}^2]}$.

   (c) In the Erdős–Rényi random graph model, we start with n vertices, and then each of the $\binom{n}{2}$ potential edges is included independently with probability p (where p may be a function of n). This is denoted $\boldsymbol{G} \sim \mathcal{G}(n, p)$. Suppose that $p = o(n^{-2/3})$. Show that

$$\mathbf{Pr}[\boldsymbol{G} \text{ contains a 4-clique}] = o(1) \qquad (n \to \infty).$$

   (Hint: Let $\boldsymbol{X}$ be the number of 4-cliques in $\boldsymbol{G}$. Compute $\mathbf{E}[\boldsymbol{X}]$ exactly as a function of n and p; then use Markov.)

   (d) On the other hand, show that if $p = \omega(n^{-2/3})$ then

$$\mathbf{Pr}[\boldsymbol{G} \text{ doesn't contain a 4-clique}] = o(1) \qquad (n \to \infty).$$

   (Hint: use part (a) or (b). You'll have to carefully calculate the probability of 4-cliques occurring simultaneously on vertex sets A and B when $|A \cap B| \geq 2$.)

6. In this problem, let $\boldsymbol{Z} \sim N(0, 1)$ denote a standard Gaussian random variable, with probability density function $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Before solving part (a) below, you might try differentiating $\varphi(x)$, just for fun.

   (a) Compute $\int_0^\infty x\varphi(x)\, dx$. Deduce $\mathbf{E}[|\boldsymbol{Z}|] = \sqrt{2/\pi}$.

[1] This isn't really necessary, but it keeps things simple.

(b) Let $a_1, \ldots, a_n$ be real numbers satisfying $\sum_i a_i^2 = 1$ and write $\varepsilon = \max\{|a_i| : i \in [n]\}$. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be i.i.d. random variables, each being ± 1 with equal probability. Let $\boldsymbol{S} = \sum_i a_i \boldsymbol{x}_i$. Show that

$$\left| \mathbf{E}\big[|\boldsymbol{S}|\big] - \sqrt{2/\pi} \right| = o(1) \quad (\text{as } \varepsilon \to 0^+).$$

Here the $o(1)$ function *may not depend on n or the a_i's*; it must be a function of ε only. For full credit, you should achieve a bound of $O(\varepsilon\sqrt{\log(1/\varepsilon)})$.

Hint: for this problem you will need the Berry–Esseen Theorem, which will be covered on Wednesday:

Berry–Esseen Theorem. *There is a universal constant c (e.g., $c = .56$ suffices) such that the following holds: Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent random variables with $\mathbf{E}[\boldsymbol{X}_i] = 0$ for all $i \in [n]$, $\mathbf{Var}[\boldsymbol{X}_i] = \sigma_i^2$, $\sum_{i=1}^n \sigma_i^2 = 1$, and $\sum_{i=1}^n \mathbf{E}[|\boldsymbol{X}_i|^3] = \beta$. Write $\boldsymbol{S} = \boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n$. Then for all $u \in \mathbb{R}$,*

$$\Big| \mathbf{Pr}[\boldsymbol{S} \le u] - \mathbf{Pr}[\boldsymbol{Z} \le u] \Big| \le c\beta.$$