Problem Set 3 CS265, Fall 2020

Due: October 9 (Friday) at 23:59 (Pacific Time)

Please follow the homework policies on the course website.

---

1. **(7 pt.) Sub-Gaussian Properties**

   We say that a random variable X is σ-sub-Gaussian for $\sigma > 0$ if the following holds:

   $$\mathbb{E}\left[e^{t(X-\mathbb{E}[X])}\right] \le e^{\sigma^2 t^2/2} \text{ for all } t \in \mathbb{R},$$

   i.e., the moment-generating function of $X - \mathbb{E}[X]$ is upper bounded by $e^{\sigma^2 t^2/2}$ at every point. Recall from lecture videos that if X follows the Gaussian distribution $N(\mu, \sigma^2)$, the moment-generating function of $X - \mathbb{E}[X]$ is exactly $e^{\sigma^2 t^2/2}$, so the term "sub-Gaussian" makes sense.

   (a) **(2 pt.)** Suppose that independent random variables X and Y are σ_X-sub-Gaussian and σ_Y-sub-Gaussian respectively. Prove that $X + Y$ is σ-sub-Gaussian for $\sigma = \sqrt{\sigma_X^2 + \sigma_Y^2}$.

   (b) **(3 pt.)** Let X be σ-sub-Gaussian. Prove the following concentration inequality for X:

   $$\Pr[|X - \mathbb{E}[X]| \ge \epsilon] \le 2\exp\left(-\frac{\epsilon^2}{2\sigma^2}\right) \text{ for any } \epsilon > 0.$$

   [**HINT:** *Mimic the proof of Chernoff bounds and upper bound the moment-generating function of $X - \mathbb{E}[X]$ using the sub-Gaussian property of X.*]

   (c) **(2 pt.)** *Hoeffding's lemma* states that any random variable X that satisfies $X \in [a, b]$ almost surely is $\frac{b-a}{2}$-sub-Gaussian. Using Hoeffding's lemma without proof, prove Hoeffding's Inequality: Suppose that $X_1, \ldots, X_n$ are independent random variables with $X_i \in [a_i, b_i]$ almost surely for each $i \in [n]$. Then for any $\epsilon > 0$,

   $$\Pr\left[\left|\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}[X_i]\right| \ge \epsilon\right] \le 2\exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

2. **(11 pt.) Concentration without Independence**

   A computer system has n different failure modes, each of which happens with a small probability. Fortunately, the system is designed to be sufficiently robust in the following sense: as long as less than half of the failures occur, things are fine; otherwise, a large-scale crash will happen. We want to make sure that the probability of crashing is small enough.

   To model the above scenario, we define n Bernoulli random variables $X_1, \ldots, X_n$. Each X_i is the indicator of the i-th failure mode, i.e., $X_i = 1$ if failure i occurs and $X_i = 0$ otheriwse. Our goal is to upper bound the probability $\Pr\left[\sum_{i=1}^n X_i \ge n/2\right]$.

   (a) **(2 pt.)** Let's first assume that the n failure events are independent and the probability of each failure is at most $1/3$. Formally, we have:

   Assumption 1. $\Pr[X_i = 1] \le 1/3$ *for every* $i \in [n]$ *and* $X_1, \ldots, X_n$ *are independent.*

Prove that under Assumption 1, for some constant $C > 0$ that does not depend on n,

$$\Pr\left[\sum_{i=1}^{n} X_i \geq n/2\right] \leq e^{-Cn}. \tag{1}$$

Thus, the probability of a crash is exponentially small in n.

[**HINT:** *Feel free to use (without proof) any of the Chernoff bounds in lecture note #5 (including Theorem 2 and Corollaries 5 and 6) and also the inequality* $\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \leq e^{-\delta^2/3}$ *for* $\delta \in [0, 1]$.]

(b) **(1 pt.)** Now we decide that Assumption 1 is too unrealistic, since many of the failure modes are known to be strongly correlated. Show that only assuming $\Pr[X_i = 1] \leq 1/3$ (and not the independence), the probability of crashing can be as large as $\Omega(1)$. In particular, prove that for any $n \geq 1$, there exist random variables $X_1, \ldots, X_n$ that satisfy: (1) $\Pr[X_i = 1] \leq 1/3$ for every $i \in [n]$; (2) $\Pr\left[\sum_{i=1}^{n} X_i \geq n/2\right] \geq 1/3$.

(c) **(2 pt.)** Let's try the following relaxation of Assumption 1, which states that the probability for k different failures to occur simultaneously is exponentially small in k:

Assumption 2. *For any* $S \subseteq [n]$, $\Pr[X_i = 1 \text{ for all } i \in S] \leq (1/3)^{|S|}$.

Show that Assumption 2 is strictly weaker than Assumption 1 by proving: (1) Assumption 1 implies Assumption 2; (2) the implication on the other direction does not hold, i.e., there exist some n and $X_1, \ldots, X_n$ that satisfy Assumption 2 but not Assumption 1.

[**HINT:** *For (2), there exists a counterexample for* $n = 2$.]

(d) **(6 pt.)** Prove that under Assumption 2, inequality (1) holds for some constant $C > 0$. In your proof, you can appeal to the proof of the Chernoff bounds from lecture videos/notes if you need to write it out verbatim at some point. For example, if you manage to upper bound $\Pr\left[\sum_{i=1}^{n} X_i \geq n/2\right]$ by an expression involving the moment-generating function of some random variable Y that is the sum of n independent Bernoulli random variables, you can simply say that "the rest of the proof is exactly the proof of Theorem 2 from Lecture #5".

[**HINT:** *Consider independent Bernoulli random variables* $Y_1, \ldots, Y_n$ *with* $Pr[Y_i = 1] = 1/3$ *for each* $i \in [n]$. *For distinct indices* $i, j, \ell \in [n]$, *does* $\mathbb{E}[X_i X_j X_\ell] \leq \mathbb{E}[Y_i Y_j Y_\ell]$ *hold? Can you extend your proof of the inequality to the case with repeating indices?*]

[**HINT:** *Let* $X = \sum_{i=1}^{n} X_i$ *and* $Y = \sum_{i=1}^{n} Y_i$. *What can we say about* $\mathbb{E}\left[X^k\right]$ *and* $\mathbb{E}\left[Y^k\right]$ *for integer* $k \geq 0$? *Considering the identity* $e^z = \sum_{k=0}^{+\infty} \frac{z^k}{k!}$, *what can we say about* $\mathbb{E}\left[e^{tX}\right]$ *and* $\mathbb{E}\left[e^{tY}\right]$ *for any* $t > 0$?]

(e) **(0 pt.) [Optional: this won't be graded.]** Can you construct counterexamples for Part 2b that satisfy *pairwise independence* but have a crashing probability of $\Omega(1/n)$? Formally, prove that there exists $C > 0$ such that for any $n \geq 2$, there exist $X_1, \ldots, X_n$ that satisfy: (1) $\Pr[X_i = 1] \leq 1/3$; (2) X_i and X_j are independent for distinct $i, j \in [n]$; (3) $\Pr\left[\sum_{i=1}^{n} X_i \geq n/2\right] \geq C/n$.

[**NOTE:** *This shows that unlike Chebyshev's inequality, Chernoff bounds do not hold if we only assume pairwise independence.*]

[**HINT:** *Recall pairwise independent hash functions if you have seen them before. You can use the Bertrand-Chebyshev theorem, which states that for any integer* $n \geq 1$, *there exists a prime number* p *with* $n < p < 2n$.]

3. **(8 pt.) Poisson Approximation**

Suppose that n balls are thrown into n bins independently and uniformly at random. Let random variable X_i denote the number of balls that end up in the i-th bin.

(a) **(1 pt.)** Consider $\Pr[X_1 = \cdots = X_n = 1]$, the probability that each bin receives exactly one ball. Explain why this is equal to $n!/n^n$.

(b) **(3 pt.)** Alternatively, we can approximate the above probability by replacing $X_1, \ldots, X_n$ with independent Poisson random variables $Y_1, \ldots, Y_n \sim \mathrm{Poi}(1)$. Find $\Pr[Y_1 = \cdots = Y_n = 1]$, and prove that it is smaller than the actual probability by a factor of $\Theta(\sqrt{n})$, i.e., for some constants $C_1, C_2 > 0$:

$$C_1\sqrt{n} \le \frac{\Pr[X_1 = \cdots = X_n = 1]}{\Pr[Y_1 = \cdots = Y_n = 1]} \le C_2\sqrt{n}. \tag{2}$$

[**HINT:** *Stirling's approximation* $\sqrt{2\pi}n^{n+1/2}e^{-n} \le n! \le en^{n+1/2}e^{-n}$ *might be handy.*]

(c) **(0 pt.)** While the above approximation is off by a $\Theta(\sqrt{n})$ factor, we will show in the remainder of this problem that the previous example is essentially the worst case: informally, $\mathbb{E}[$ *any* function of $X_1, \ldots, X_n]$ is within a $\Theta(\sqrt{n})$ factor of the corresponding expression for $Y_1, \ldots, Y_n$.

We'll state that formally (and you'll prove it!) in the next part, but we'll need the following two facts about $(X_1)_{i=1}^n$ and $(Y_i)_{i=1}^n$:

- $\Pr[Y_1 + \cdots + Y_n = n] \ge \frac{1}{e\sqrt{n}}$.
- Conditioning on $Y_1 + \cdots + Y_n = n$, the distribution of $(Y_1, \ldots, Y_n)$ is the same as that of $(X_1, \ldots, X_n)$.

You may assume these facts without proof for the next part.

[**NOTE:** *There is no question here, we're just stating the above facts.*]

(d) **(4 pt.)** Let $\mathbb{N}$ denote $\{0, 1, 2, \ldots\}$ and $f : \mathbb{N}^n \to [0, +\infty)$ be an arbitrary function that maps n nonnegative integers to a nonnegative real number. Random variables $(X_i)_{i=1}^n$ and $(Y_i)_{i=1}^n$ are defined as above. Prove the following inequality:

$$\frac{\mathbb{E}[f(X_1, \ldots, X_n)]}{\mathbb{E}[f(Y_1, \ldots, Y_n)]} \le e\sqrt{n}. \tag{3}$$

[**NOTE:** *This states that the Poisson approximation* $\mathbb{E}[f(Y)]$ *underestimates* $\mathbb{E}[f(X)]$ *by a factor of at most* $O(\sqrt{n})$. *When the function* $f(t_1, \ldots, t_n)$ *is chosen to be the indicator of* $t_1 = \cdots = t_n = 1$, $\mathbb{E}[f(X)]$ *and* $\mathbb{E}[f(Y)]$ *are exactly* $\Pr[X_1 = \cdots = X_n = 1]$ *and* $\Pr[Y_1 = \cdots = Y_n = 1]$, *and the bound* (3) *is consistent with inequality* (2).]

(e) **(0 pt.) [Optional: this part won't be graded]** Prove the facts from part (c).

4. **(0 pt.) [Optional: this won't be graded.] Moment vs Chernoff Bounds**

Let X be a non-negative random variable and fix $\epsilon > 0$. So far we have seen two approaches to upper bounding the tail probability $\Pr[X \ge \epsilon]$. One is based on the moments of X: assuming that we know (either exactly or a good upper bound of) $\mathbb{E}[X^1], \mathbb{E}[X^2], \ldots$, for any integer

$k \geq 1$ we have $\Pr[X \geq \epsilon] = \Pr[X^k \geq \epsilon^k] \leq \frac{\mathbb{E}[X^k]}{\epsilon^k}$. Choosing the k that minimizes the right-hand side gives us the best *moment bound*:

$$\inf_{k \in \mathbb{Z}, k \geq 1} \frac{\mathbb{E}\left[X^k\right]}{\epsilon^k}.$$

Another approach is based on the moment-generating function of X: for any $t > 0$, we have $\Pr[X \geq \epsilon] = \Pr[e^{tX} \geq e^{t\epsilon}] \leq \frac{\mathbb{E}[e^{tX}]}{e^{t\epsilon}}$. Similarly, the best *Chernoff bound* is obtained by choosing t optimally:

$$\inf_{t>0} \frac{\mathbb{E}\left[e^{tX}\right]}{e^{t\epsilon}}.$$

Prove that the best *moment bound* is always as good as the best *Chernoff bound*, i.e.,

$$\min\left\{\inf_{k \in \mathbb{Z}, k \geq 1} \frac{\mathbb{E}\left[X^k\right]}{\epsilon^k}, 1\right\} \leq \inf_{t>0} \frac{\mathbb{E}\left[e^{tX}\right]}{e^{t\epsilon}}.$$