COS 521: Homework 3

Due on October 31, 2022

Professor Matt Weinberg

Nameless Author :)
Collaborators: I can't tell you :)

Problem 1

We say a random variable Z is *subgamma* with parameters (σ^2, B) if

$$\mathbb{E}\left[e^{\lambda(Z-\mathbb{E}[Z])}\right] \leq e^{\lambda^2\sigma^2/2},$$

for all $|\lambda| \leq B$.

Part A

Proof. Let $Z = \sum_{i=1}^m Z_i$ be the sum of the independent random variables. Then,

$$\mathbb{E}[Z] = \sum_{i=1}^m \mathbb{E}[Z_i] \implies Z - \mathbb{E}[Z] = \sum_{i=1}^m Z - \mathbb{E}[Z_i]$$

Since the random variables are independent, we have that

$$\mathbb{E}\left[e^{\lambda(Z-\mathbb{E}[Z])}\right] = \mathbb{E}\left[\prod_{i=1}^m e^{\lambda(Z-\mathbb{E}[Z_i])}\right] = \prod_{i=1}^m \mathbb{E}\left[e^{\lambda(Z-\mathbb{E}[Z_i])}\right]$$

Now, for any λ such that $|\lambda| \leq B = min_{i\in[m]} B_i$, all of the subgamma conditions for all the Z_i are satisfied, and we can say that

$$\prod_{i=1}^m \mathbb{E}\left[e^{\lambda(Z-\mathbb{E}[Z_i])}\right] \leq \prod_{i=1}^m e^{\lambda^2\sigma_i^2/2} = e^{\lambda^2\sigma^2/2} \implies \mathbb{E}\left[e^{\lambda(Z-\mathbb{E}[Z])}\right] \leq e^{\lambda^2\sigma^2/2},$$

where $\sigma^2 = \sum_{i=1}^m \sigma_i^2$. So, we see that $\boxed{Z \text{ is subgamma with parameters } \left(\sum_{i=1}^m \sigma_i^2, min_{i\in[m]} B_i\right)}$ ■

Part B

Proof. Suppose that Z is subgamma with parameters (σ^2, B). Define $\Delta := Z - \mathbb{E}[Z]$ for notation, and observe that, since e^x is monotone increasing for positive x, we have that for all $\lambda \in (0, B]$

$$\mathbb{P}[\Delta > t] = \mathbb{P}\left[e^{\lambda\Delta} > e^{\lambda t}\right] \leq \frac{\mathbb{E}\left[e^{\lambda\Delta}\right]}{e^{\lambda t}} \leq \frac{e^{\lambda^2\sigma^2/2}}{e^{\lambda t}} = e^{\frac{\lambda^2\sigma^2}{2} - \lambda t},$$

where the first inequality is just Markov's Inequality and the second inequality comes from the fact that Z is subgamma. Similarly, we can bound the other tail by exponentiating with $e^{\lambda(\cdot)}$ for $\lambda \in [-B, 0)$, which flips the inequality:

$$\mathbb{P}[\Delta < -t] = \mathbb{P}\left[e^{\lambda\Delta} > e^{-\lambda t}\right] \leq \frac{\mathbb{E}\left[e^{\lambda\Delta}\right]}{e^{-\lambda t}} \leq \frac{e^{\lambda^2\sigma^2/2}}{e^{-\lambda t}} = e^{\frac{\lambda^2\sigma^2}{2} + \lambda t},$$

where we also apply Markov's Inequality and the subgamma condition. We can combine these results and show that the tails are both bounded by $e^{\lambda^2\sigma^2/2 - |\lambda| t}$, where we select $\lambda \in (0, B]$ for the upper tail and $\lambda \in [-B, 0)$ for the lower tail. Now, there are two cases:

- $(\frac{t}{\sigma^2} \leq B)$ If this is the case, we can set $\lambda = \frac{t}{\sigma^2}$ for the upper tail and $\lambda = -\frac{t}{2}$ for the lower tail, and the subgamma condition will be satisfied. Plugging this into the bound yields that both tails are at most
$$e^{\frac{\lambda^2\sigma^2}{2} - |\lambda| t} = e^{\frac{t^2\sigma^2}{2\sigma^4} - \frac{t^2}{\sigma^2}} = e^{-t^2/2\sigma^2}$$
- $(\frac{t}{\sigma^2} > B)$ In this case, we can use the fact that $|\lambda| \leq B$ to see that both tails are at most
$$e^{\frac{B^2\sigma^2}{2} - Bt} \leq e^{\frac{Bt\sigma^2}{2\sigma^2} - Bt} = e^{-Bt/2},$$
where the inequality comes from one application of the fact that $B < \frac{t}{\sigma^2}$.

Problem 1 continued on next page...

Note that $e^{-Bt/2} > e^{-t^2/2\sigma^2}$ if and only if $B < \frac{t}{\sigma^2}$, and so we see that the shared bound on both tails takes the value of

$$max\left\{e^{-t^2/2\sigma^2}, e^{-Bt/2}\right\}$$

as desired. ■

Part C

Proof. Let Z be a geometric random variable such that $\mathbb{P}[Z = k] = p \cdot (1-p)^{k-1}$ for all integers $k \geq 1$. Then, we have that the expectation of Z is $\mathbb{E}[Z] = \frac{1}{p}$. We can use the sum definition of expectation to say that

$$\mathbb{E}\left[e^{\lambda(Z-\mathbb{E}[Z])}\right] = \mathbb{E}\left[e^{\lambda(Z-1/p)}\right] = \sum_{k=1}^{\infty} e^{\lambda k-\lambda/p} \cdot \mathbb{P}[Z = k] = \frac{p}{1-p}e^{-\lambda/p} \cdot \sum_{k=1}^{\infty} e^{\lambda k}(1-p)^k$$

Suppose that $|\lambda| \leq \frac{p}{2} \implies e^{\lambda}(1-p) \leq e^{\frac{p}{2}}(1-p) \leq e^{\frac{p}{2}-p} = e^{-\frac{p}{2}} < 1$, which is always less than 1 for $p \in (0, 1]$. So, we can say that this geometric sum converges (since the ratio $e^{\lambda}(1-p) < 1$), yielding

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda(Z-1/p)}\right] &= \frac{p}{1-p}e^{-\lambda/p} \cdot \frac{e^{\lambda}(1-p)}{1-e^{\lambda}(1-p)} = \frac{pe^{\lambda-\lambda/p}}{1-e^{\lambda}(1-p)} \\
&= \frac{pe^{-\lambda/p}}{e^{-\lambda}-(1-p)} \leq \frac{pe^{-\lambda/p}}{1-\lambda-(1-p)} = \frac{pe^{-\lambda/p}}{p-\lambda} \\
&= \frac{e^{-\lambda/p}}{1-\frac{\lambda}{p}},
\end{aligned}$$

where the inequality comes from an application of $e^{-\lambda} \geq 1-\lambda$. Now, for λ with $|\lambda| \leq \frac{p}{2} \implies \left|\frac{\lambda}{p}\right| \leq \frac{1}{2}$, we can use the other inequality in the hint to say that

$$e^{-\lambda/p} \cdot \frac{1}{1-\frac{\lambda}{p}} \leq e^{-\lambda/p} \cdot e^{\lambda/p+\lambda^2/p^2} = e^{\frac{\lambda^2}{p^2}} \implies \boxed{\mathbb{E}\left[e^{\lambda(Z-\mathbb{E}[Z])}\right] \leq e^{\frac{\lambda^2}{p^2}}},$$

where all of the above reasoning holds for all λ with $|\lambda| \leq \frac{p}{2}$. This is precisely the statement that Z is subgamma with parameters $\left(\frac{2}{p^2}, \frac{p}{2}\right)$. ■

Problem 2

Let $b := 1 + \alpha$ for immense notational convenience.

Part A

Proof. Consider calling ***inc()*** to try and increment X from $k-1$ to k. Each of these trials are independent, each with a probability of $p_k = b^{-(k-1)}$ of success (incrementing). Each of these trials are then independent Bernoulli variables with parameter p_k, and so the number of trials necessary before the first success is distributed as a geometric variable with parameter p_k (this is the definition of the geometric distribution). So, $Y_k \sim Geom(p_k)$. From Problem 1(c), we get that Y_k is subgamma with parameters

$$\left(\frac{2}{p_k^2}, \frac{p_k}{2}\right) = \boxed{\left(2b^{2k-2}, \frac{b^{-(k-1)}}{2}\right)}$$

■

Part B

Proof. Let $\epsilon > 0$ be arbitrary. Let k be such that

$$\widetilde{n}(k) = (1-\epsilon)n \implies \frac{b^k - 1}{\alpha} = (1-\epsilon)n \implies k = \log_b((1-\epsilon)\cdot \alpha n + 1)$$

We assume for simplification that this value of k, which is the value the counter must take such that $\widetilde{n}(k) = (1-\epsilon)n$, is an integer (ED post 93 says we can do this, thank you Huacheng!). So, $k \in \mathbb{N}$. Therefore, we are interested in bounding the probability that after n ***inc()*** calls,

$$\widetilde{n}(X) < (1-\epsilon)n = \widetilde{n}(k) \iff X < k$$

Note that we can say this since $b = (1+\alpha) > 1 \implies b^{(\cdot)}$ is monotone increasing, allowing us to take the log base b of both sides of the inequality. The above event occurring is exactly equivalent to the event that after n calls to ***inc()***, the counter has not yet incremented to a value of k. In other words, this event is equivalent to the event that $\sum_{i=1}^k Y_i > n$, in which case it would have taken more than n calls to ***inc()*** to reach a counter value of k (since each Y_i is the number of calls needed to increment from $i-1$ to i). Let $Y^{(k)} := \sum_{i=1}^k Y_i$ be the random variable for the number of calls to ***inc()*** that would have been needed to reach a counter value of k. We arrive at the fact that

$$\mathbb{P}\left[\widetilde{n}(X) < (1-\epsilon)n\right] = \mathbb{P}\left[Y^{(k)} > n\right]$$

Lemma 1. *$Y^{(k)}$ is subgamma with parameters $\left(2\cdot\frac{b^{2k}-1}{b^2-1}, \frac{1}{2b^{k-1}}\right)$ and expectation $\mathbb{E}\left[Y^{(k)}\right] = (1-\epsilon)n$.*

Proof of Lemma 1. Firstly, note that since each of the Y_i's are independent from each other (each increment trial is independent), we can apply Problem 1(a) to see that $Y^{(k)}$ is subgamma with parameters

$$\sigma^2 = \sum_{i=1}^k \sigma_i^2 = \sum_{i=1}^k 2b^{2i-2} = 2\cdot\sum_{i=1}^k (b^2)^{i-1} = 2\cdot\left(\frac{(b^2)^k - 1}{b^2 - 1}\right) = 2\cdot\frac{b^{2k}-1}{b^2-1},$$

$$B = \min_{i\in[k]}\{B_i\} = \min_{i\in[k]}\left\{\frac{b^{-(i-1)}}{2}\right\} = \frac{b^{-(k-1)}}{2} = \frac{1}{2b^{k-1}},$$

where for σ^2 we used the finite geometric series with ratio b^2 and for B we used the fact that $b = 1+\alpha > 1$, which means that the minimum of $b^{-(i-1)}$ happens for the largest i, which is $i = k$. We can also determine that since a geometric random variable with parameter p has expectation $\frac{1}{p}$,

$$\mathbb{E}\left[Y^{(k)}\right] = \mathbb{E}\left[\sum_{i=1}^{k} Y_i\right] = \sum_{i=1}^{k} \mathbb{E}\left[Y_i\right] = \sum_{i=1}^{k} \frac{1}{p_k} = \sum_{i=1}^{k} b^{i-1} = \frac{b^k - 1}{b-1} = \frac{(1+\alpha)^k - 1}{\alpha} = \widetilde{n}(k) = (1-\epsilon)n,$$

where we again used the finite geometric series and plugged in our earlier definition of k. ■

We immediately apply Lemma 1 to say that

$$\mathbb{P}\left[\widetilde{n}(X) < (1-\epsilon)n\right] = \mathbb{P}\left[Y^{(k)} > n\right] = \mathbb{P}\left[Y^{(k)} - (1-\epsilon)n > n - (1-\epsilon)n\right] = \mathbb{P}\left[Y^{(k)} - \mathbb{E}\left[Y^{(k)}\right] > \epsilon n\right]$$

We can apply Problem 1(b) with $t = \epsilon n > 0$ to see that, since $Y^{(k)}$ is subgamma,

$$\mathbb{P}\left[Y^{(k)} - \mathbb{E}\left[Y^{(k)}\right] > \epsilon n\right] \leq max\left\{e^{-\frac{\epsilon^2 n^2}{2\sigma^2}}, e^{-\frac{\epsilon n B}{2}}\right\} = \max\left\{e^{-\frac{\epsilon^2 n^2 (b^2-1)}{4(b^{2k}-1)}}, e^{-\frac{\epsilon n}{4b^{k-1}}}\right\}$$

Lemma 2. *There exists some constant C such that for all n with $\alpha n > C$, we have that* $\max\left\{e^{-\frac{\epsilon^2 n^2 (b^2-1)}{4(b^{2k}-1)}}, e^{-\frac{\epsilon n}{4b^{k-1}}}\right\}$ *is of the order $e^{-\Omega\left(\frac{\epsilon^2}{\alpha}\right)}$.*

Proof of Lemma 2. We want to show that the exponents in both branches are of the order $-\Omega\left(\frac{\epsilon^2}{\alpha}\right)$. We first tackle the left one. We can plug in our expression for k to see that $b^k = (1-\epsilon)\cdot \alpha n + 1$, and so since $b^2 - 1 = \alpha^2 + 2\alpha$,

$$\frac{\epsilon^2 n^2 (b^2-1)}{4(b^{2k}-1)} = \frac{\epsilon^2 n^2 (\alpha^2 + 2\alpha)}{4(((1-\epsilon)\cdot\alpha n + 1)^2 - 1)} = \frac{\epsilon^2 n^2 (\alpha^2 + 2\alpha)}{4((1-\epsilon)^2 \cdot \alpha^2 n^2 + 2(1-\epsilon)\cdot \alpha n)}$$

Note that, as $n \to \infty$, this expression approaches the limit of $\frac{\epsilon^2(\alpha^2+2\alpha)}{4(1-\epsilon)^2\alpha^2}$ from below. So, for all $\delta_1 > 0$, there exists some large constant C_1 such that, for all n with $\alpha n > \alpha C_1$ (this is a way to rigorously define a limit as getting arbitrarily close to the result for large enough n),

$$\frac{\epsilon^2 n^2 (\alpha^2 + 2\alpha)}{4((1-\epsilon)^2 \cdot \alpha^2 n^2 + 2(1-\epsilon)\cdot \alpha n)} + \delta_1 > \frac{\epsilon^2(\alpha^2 + 2\alpha)}{4(1-\epsilon)^2\alpha^2} = \frac{\epsilon^2(\alpha+2)}{4(1-\epsilon)^2\alpha} > \frac{\epsilon^2}{4} + \frac{\epsilon^2}{2\alpha} > \frac{\epsilon^2}{2\alpha},$$

where the first inequality uses a nice and formal way to define the limit, the second inequality uses that $1 - \epsilon < 1$, and the third inequality is because $\frac{\epsilon^2}{4} > 0$. (Note that this method with the δ_1 is equivalent to saying that, for large enough n, we are within a *constant* of the limit; meaning, for large enough n we behave asymptotically the same as the limit in terms of ϵ and α).

$$\implies \frac{\epsilon^2 n^2 (\alpha^2 + 2\alpha)}{4((1-\epsilon)^2 \cdot \alpha^2 n^2 + 2(1-\epsilon)\cdot \alpha n)} = \Omega\left(\frac{\epsilon^2}{\alpha}\right)$$

For the right branch, we can also plug in $b^k = (1-\epsilon)\cdot\alpha n + 1$ to see that

$$\frac{\epsilon n}{4b^{k-1}} = \frac{\epsilon n \cdot b}{4((1-\epsilon)\cdot \alpha n + 1)} = \frac{\epsilon n(1+\alpha)}{4(1-\epsilon)\alpha n + 4}$$

Once again, note that the limit as $n \to \infty$ of this expression approaches $\frac{\epsilon(1+\alpha)}{4(1-\epsilon)\alpha}$ from below. So, for all $\delta_2 > 0$, there exists some large constant C_2 such that, for all n with $n > C_2 \implies \alpha n > \alpha C_2$,

$$\frac{\epsilon n(1+\alpha)}{4(1-\epsilon)\alpha n + 4} + \delta_2 > \frac{\epsilon(1+\alpha)}{4(1-\epsilon)\alpha} > \frac{\epsilon(1+\alpha)}{4\alpha} = \frac{\epsilon}{4\alpha} + \frac{\epsilon}{4} > \frac{\epsilon^2}{4\alpha} = \Omega\left(\frac{\epsilon^2}{\alpha}\right),$$

where for the first inequality we used a formal limit definition, for the second inequality we used that $1-\epsilon < 1$, and for the third inequality we used that $\epsilon < 1 \implies \epsilon > \epsilon^2$ and $\frac{\epsilon}{4} > 0$. So, if we take $C = \max\{\alpha C_1, \alpha C_2\}$, we get that for all n with $\alpha n > C$,

$$\min\left\{\frac{\epsilon^2 n^2(b^2-1)}{4(b^{2k}-1)}, \frac{\epsilon n}{4b^{k-1}}\right\} = \Omega\left(\frac{\epsilon^2}{\alpha}\right) \implies \max\left\{e^{-\frac{\epsilon^2 n^2(b^2-1)}{4(b^{2k}-1)}}, e^{-\frac{\epsilon n}{4b^{k-1}}}\right\} = e^{-\Omega\left(\frac{\epsilon^2}{\alpha}\right)}$$

■

We can now apply Lemma 2 to our earlier result to arrive at the fact that, for some C, for all n s.t. $\alpha n > C$, after n calls to ***inc()*** it holds that

$$\mathbb{P}\left[\widetilde{n}(X) < (1-\epsilon)n\right] = \mathbb{P}\left[Y^{(k)} - \mathbb{E}\left[Y^{(k)}\right] > \epsilon n\right] \le e^{-\Omega\left(\frac{\epsilon^2}{\alpha}\right)}$$

■

Part C

Proof. Given N, ϵ, δ, we wish to set α and T to achieve the desired result. From the result of Problem 2(b), we have that both of the events $\widetilde{n}(X) < (1-\epsilon)n$ and $\widetilde{n}(X) > (1+\epsilon)n$ occur with probability at most $e^{-\Omega\left(\frac{\epsilon^2}{\alpha}\right)}$. We may apply the union bound to say that the probability that either of these two events happening is at most $2e^{-\Omega\left(\frac{\epsilon^2}{\alpha}\right)}$. Therefore, the probability that neither of these events happen is greater than $1 - 2e^{-\Omega\left(\frac{\epsilon^2}{\alpha}\right)}$. Therefore, if we set $\alpha = O\left(\frac{\epsilon^2}{\log 2/\delta}\right) \implies \frac{1}{\alpha} = \Omega\left(\frac{\log 2/\delta}{\epsilon^2}\right)$,

$$\mathbb{P}\left[\widetilde{n}(X) \in (1 \pm \epsilon)n\right] > 1 - 2e^{-\Omega\left(\frac{\epsilon^2}{\alpha}\right)} = 1 - O(\delta)$$

as desired. So, we can say that the Morris Counter approximates n to arbitrary precision with high probability (for large enough n). This means that, with high probability, the worst case (largest counter possible) takes the form

$$\widetilde{n}(X_{max}) \in (1 \pm \epsilon)N \implies \frac{(1+\alpha)^{X_{max}} - 1}{\alpha} = O(N) \implies (1+\alpha)^{X_{max}} = O(N) \implies X_{max} = O(\log N)$$

Since X_{max} is an integer, it takes $\log(X_{max}) = O(\log\log N)$ bits to store the counter; this occurs with high probability $1 - O(\delta)$. We also need to store α, which takes $\log\left(\frac{1}{\alpha}\right)$ bits. With the value of α determined above, we find the result that for large enough n, it occurs with high probability $1 - O(\delta)$ that it takes

$$O(\log\log N) + \log\left(\frac{1}{\alpha}\right) = O(\log\log N) + O\left(\log\frac{\log 2/\delta}{\epsilon^2}\right) = O\left(\log\log N + \log\frac{1}{\epsilon} + \log\log\frac{1}{\delta}\right)$$

bits for the whole Morris Counter, as desired. To handle the cases for tiny enough n that the reasoning from Problem 1(b) breaks down, we can keep an exact counter for all $n \le T$, where $T = \log(N)$; this exact counter also takes $O(\log T) = O(\log\log N)$ bits, but maintains perfect accuracy for the small n that may mess up the Morris Counter. We arrive at the final result: given N, ϵ, δ, we can create a Morris Counter such that after n calls to ***inc()***,

$$\mathbb{P}\left[\widetilde{n}(X) \in (1 \pm \epsilon)n\right] > 1 - O(\delta)$$

using

$$O\left(\log\log N + \log\frac{1}{\epsilon} + \log\log\frac{1}{\delta}\right)$$

bits. ■

Problem 3

Solution

Proof. Consider the Johnson-Lindenstrauss dimensionality reduction method described in lecture: $x \to \Pi x$ where each entry in $\Pi \in \mathbb{R}^{m \times d}$ equals $\Pi_{ij} = c \cdot g_{ij}$ for some fixed scaling factor c and $g_{ij} \sim \mathcal{N}(0,1)$.

Lemma 3. *For any vector $\vec{x} = (x_1, ..., x_d) \in \mathbb{R}^d$, we have that the expectation over values of Π of the L1 norm of $\Pi\vec{x}$ is*

$$\mathbb{E}\left[||\Pi\vec{x}||_1\right] = ||\vec{x}||_2 \cdot cm \cdot \sqrt{\frac{2}{\pi}}$$

Proof of Lemma 3. Let $w_i = \sum_{j=1}^d x_j g_{ij}$. Then, we can write

$$\Pi\vec{x} = (cw_1, ..., cw_m) \implies ||\Pi\vec{x}||_1 = c \cdot \sum_{i=1}^m |w_i| \implies \mathbb{E}\left[||\Pi\vec{x}||_1\right] = c \cdot \sum_{i=1}^m \mathbb{E}\left[|w_i|\right]$$

We can evaluate each expectation above as follows: note that each w_i is a linear combination of independent unit normal random variables g_{ij}, weighted by components x_j. So, by the properties of linear combinations of Gaussians, $w_i = x_1 g_{i1} + x_2 g_{i2} + ... + x_d g_{id} \sim \mathcal{N}(0, x_1^2 + x_2^2 + ... + x_d^2) = \mathcal{N}(0, ||\vec{x}||_2^2)$. Then, we can say that

$$\mathbb{E}\left[|w_i|\right] = \int_{-\infty}^{\infty} |w_i| \cdot \frac{e^{-\frac{w_i^2}{2||\vec{x}||_2^2}}}{\sqrt{2\pi||\vec{x}||_2^2}} dw_i$$

Let $u = \frac{w_i}{\sqrt{2||\vec{x}||_2^2}}$. Then, we get that

$$\mathbb{E}\left[|w_i|\right] = 2 \cdot \int_0^{\infty} w_i \cdot \frac{e^{-\frac{w_i^2}{2||\vec{x}||_2^2}}}{\sqrt{2\pi||\vec{x}||_2^2}} dw_i = 2 \cdot \sqrt{2}||\vec{x}||_2 \int_0^{\infty} u \cdot \frac{e^{-u^2}}{\sqrt{\pi}} du = 2 \cdot \sqrt{2}||\vec{x}||_2 \cdot \frac{1}{2\sqrt{\pi}} = ||\vec{x}||_2 \sqrt{\frac{2}{\pi}},$$

where the integral evaluation is a simple Gaussian integral. Then, we get that

$$\mathbb{E}\left[||\Pi\vec{x}||_1\right] = c \cdot \sum_{i=1}^m \mathbb{E}\left[|w_i|\right] = c \cdot \sum_{i=1}^m ||\vec{x}||_2 \sqrt{\frac{2}{\pi}} = ||\vec{x}||_2 \cdot cm \cdot \sqrt{\frac{2}{\pi}}$$

■

So, consider the set of vectors in $\mathbb{R}^d$ for some d (we will find d later) given by

$$\vec{x_0} = (0, ..., 0), \quad \vec{x_1} = (1, 0, ..., 0), \quad \vec{x_2} = (1, ..., 1),$$

i.e. $\vec{x_0}$ is the zero vector, $\vec{x_1}$ is the first basis vector, and $\vec{x_2}$ is the sum of all the basis vectors (1 for every component). Then, we have the L1 norms

$$||\vec{x_1} - \vec{x_0}||_1 = 1, \quad ||\vec{x_2} - \vec{x_0}||_1 = d,$$

and the L2 norms

$$||\vec{x_1} - \vec{x_0}||_2 = 1, \quad ||\vec{x_2} - \vec{x_0}||_2 = \sqrt{d}$$

Then, we can apply Lemma 3 to see that

$$\mathbb{E}\left[||\Pi(\vec{x_1} - \vec{x_0})||_1\right] = ||\vec{x_1} - \vec{x_0}||_2 \cdot cm \cdot \sqrt{\frac{2}{\pi}} = cm \cdot \sqrt{\frac{2}{\pi}} = cm \cdot \sqrt{\frac{2}{\pi}}||\vec{x_1} - \vec{x_0}||_1$$

and

$$\mathbb{E}\left[||\Pi(\vec{x_2} - \vec{x_0})||_1\right] = ||\vec{x_2} - \vec{x_0}||_2 \cdot cm \cdot \sqrt{\frac{2}{\pi}} = cm \cdot \sqrt{\frac{2d}{\pi}} = cm \cdot \sqrt{\frac{2}{\pi d}}||\vec{x_1} - \vec{x_0}||_1$$

In order for both of these estimations to be correct within a factor of 2, we require that

$$cm \cdot \sqrt{\frac{2}{\pi}}, \quad cm \cdot \sqrt{\frac{2}{\pi d}} \in [0.5, 2.0]$$

Since $d > 1$, the most extreme value of d for which this can be possible happens exactly when

$$cm \cdot \sqrt{\frac{2}{\pi}} = 2, \quad cm \cdot \sqrt{\frac{2}{\pi d}} = 0.5 \implies \sqrt{d} = 4 \implies d = 16$$

(To see this, note that the largest value of d occurs when the ratio between $cm \cdot \sqrt{\frac{2}{\pi}}$ *and* $cm \cdot \sqrt{\frac{2}{\pi d}}$*, which is* $\sqrt{d}$*, is as large as possible; the largest such ratio within this range is* $2/0.5 = 4$*).* Therefore, for any value of $d > 16$, we cannot have that both

$$cm \cdot \sqrt{\frac{2}{\pi}} \in [0.5, 2.0]$$

and

$$cm \cdot \sqrt{\frac{2}{\pi d}} \in [0.5, 2.0]$$

So, this example shows that for $d > 16$, the JL dimensionality reduction method cannot preserve L1 distances within a factor of 2 within the example set $\{\vec{x_0}, \vec{x_1}, \vec{x_2}\} \subset \mathbb{R}^d$. ■

Problem 4

Solution

Proof. We start by reformulating the k-means objective in terms of pairwise L2 distances. Note that all norms in this problem are L2 norms.

Lemma 4. *For vectors $\vec{x_1}, ..., \vec{x_n}$ and clusters $C_1, ..., C_k$ with centroids $\vec{c_1}, ..., \vec{c_k}$, we have that*

$$f_X(C_1, ..., C_K) = \sum_{j=1}^{k} \sum_{i \in C_j} ||\vec{c_j} - \vec{x_i}||^2 = \sum_{j=1}^{k} \frac{1}{2|C_j|} \cdot \sum_{r,s \in C_j} ||\vec{x_r} - \vec{x_s}||^2$$

Proof of Lemma 4. We perform some cute algebra with the following property:

$$||x - y||^2 = ||x||^2 + ||y||^2 - 2\langle x, y \rangle$$

Getting started,

$$\begin{aligned}\sum_{j=1}^{k} \sum_{i \in C_j} ||\vec{c_j} - \vec{x_i}||^2 &= \sum_{j=1}^{k} \sum_{i \in C_j} ||\vec{c_j}||^2 + ||\vec{x_i}||^2 - 2\langle c_j, x_i \rangle \\ &= \sum_{j=1}^{k} |C_j| \cdot ||\vec{c_j}||^2 + \sum_{i \in C_j} ||\vec{x_i}||^2 - 2\langle \vec{c_j}, \vec{x_i} \rangle\end{aligned}$$

Plugging in that $\vec{c_j} = \frac{1}{|C_j|} \cdot \sum_{i \in C_j} \vec{x_i}$, we find that

$$\begin{aligned}&= \sum_{j=1}^{k} |C_j| \cdot \frac{1}{|C_j|} \cdot \sum_{i \in C_j} \langle \vec{c_j}, \vec{x_i} \rangle + \sum_{i \in C_j} ||\vec{x_i}||^2 - 2\langle \vec{c_j}, \vec{x_i} \rangle \\ &= \sum_{j=1}^{k} \sum_{i \in C_j} ||\vec{x_i}||^2 - \langle \vec{c_j}, \vec{x_i} \rangle \\ &= \sum_{j=1}^{k} \frac{1}{2|C_j|} \left(\sum_{r,s \in C_j} ||\vec{x_r}||^2 + ||\vec{x_s}||^2 \right) - \frac{1}{|C_j|} \left(\sum_{r,s \in C_j} \langle \vec{x_r}, \vec{x_s} \rangle \right),\end{aligned}$$

where the first double sum over C_j is just a clever rewriting of the sum of squared norms $||\vec{x_i}||^2$ (dividing by $2|C_j|$ to avoid double counting), and the second double sum comes from plugging in $\vec{c_j}$. This gives

$$\begin{aligned}&= \sum_{j=1}^{k} \frac{1}{2|C_j|} \left(\sum_{r,s \in C_j} ||\vec{x_r}||^2 + ||\vec{x_s}||^2 - 2\langle \vec{x_r}, \vec{x_s} \rangle \right) \\ &= \sum_{j=1}^{k} \frac{1}{2|C_j|} \cdot \sum_{r,s \in C_j} ||\vec{x_r} - \vec{x_s}||^2\end{aligned}$$

as desired. ■

From here on out, we use Lemma 4 to rewrite the k-means objective. We apply the result of the Johnson-Lindenstrauss Theorem: namely, that if Π is a JL map into $s = O\left(\frac{\log n}{(\epsilon/3)^2}\right)$ dimensions, then we can say that for all $\vec{x}, \vec{y} \in X$,

$$\mathbb{P}\left[||\Pi\vec{x} - \Pi\vec{y}||^2 \in \left(1 \pm \frac{\epsilon}{3}\right) ||\vec{x} - \vec{y}||^2\right] > 1 - \frac{1}{n}$$

Problem 4 continued on next page. . .

Suppose that $C_1, ..., C_k$ are the optimal clusters that obtain minimal objective $OPT_X = f_X(C_1, ..., C_k)$ in the metric space X (i.e. over the vectors $\vec{x_1}, ..., \vec{x_n}$) and that $\widetilde{C_1}, ..., \widetilde{C_k}$ are the optimal clusters that minimize $OPT_{\Pi X} = f_{\Pi X}(\widetilde{C_1}, ..., \widetilde{C_k})$ in the dimensionality-reduced metric space (i.e. over the vectors $\Pi\vec{x_1}, ..., \Pi\vec{x_n}$). Note that the JL Theorem grants that, with probability greater than $1 - \frac{1}{n}$, the following two relations (the next half of the page) hold:

$$f_{\Pi X}(C_1, ..., C_k) = \sum_{j=1}^{k} \frac{1}{2|C_j|} \cdot \sum_{r,s \in C_j} ||\Pi\vec{x_r} - \Pi\vec{x_s}||^2 < \sum_{j=1}^{k} \frac{1}{2|C_j|} \cdot \sum_{r,s \in C_j} \left(1 + \frac{\epsilon}{3}\right) ||\vec{x_r} - \vec{x_s}||^2$$
$$= \left(1 + \frac{\epsilon}{3}\right) \sum_{j=1}^{k} \frac{1}{2|C_j|} \cdot \sum_{r,s \in C_j} ||\vec{x_r} - \vec{x_s}||^2 = \left(1 + \frac{\epsilon}{3}\right) OPT_X,$$

where we used the fact that $C_1, ..., C_k$ are the optimal clustering in the X metric space. We can also perform a similar thing to see that, for the optimal clusters $\widetilde{C_1}, ..., \widetilde{C_k}$ in the ΠX metric space,

$$OPT_{\Pi X} = f_{\Pi X}(\widetilde{C_1}, ..., \widetilde{C_k}) = \sum_{j=1}^{k} \frac{1}{2|\widetilde{C_j}|} \cdot \sum_{r,s \in \widetilde{C_j}} ||\Pi\vec{x_r} - \Pi\vec{x_s}||^2$$
$$> \sum_{j=1}^{k} \frac{1}{2|\widetilde{C_j}|} \cdot \sum_{r,s \in \widetilde{C_j}} \left(1 - \frac{\epsilon}{3}\right) ||\vec{x_r} - \vec{x_s}||^2$$
$$= \left(1 - \frac{\epsilon}{3}\right) \sum_{j=1}^{k} \frac{1}{2|\widetilde{C_j}|} \cdot \sum_{r,s \in \widetilde{C_j}} ||\vec{x_r} - \vec{x_s}||^2 = \left(1 - \frac{\epsilon}{3}\right) f_X(\widetilde{C_1}, ..., \widetilde{C_k})$$

Lastly, note that because of the fact that $\widetilde{C_1}, ..., \widetilde{C_k}$ is optimal (minimizes the objective) in the ΠX metric space, we have that

$$OPT_{\Pi X} \leq f_{\Pi X}(C_1, ..., C_k)$$

To recap, we showed that with probability $1 - \frac{1}{n}$, the following three inequalities hold:

$$f_{\Pi X}(C_1, ..., C_k) < \left(1 + \frac{\epsilon}{3}\right) OPT_X,$$
$$OPT_{\Pi X} > \left(1 - \frac{\epsilon}{3}\right) f_X(\widetilde{C_1}, ..., \widetilde{C_k}),$$
$$OPT_{\Pi X} \leq f_{\Pi X}(C_1, ..., C_k)$$

Chaining these three together yields

$$\left(1 - \frac{\epsilon}{3}\right) f_X(\widetilde{C_1}, ..., \widetilde{C_k}) < OPT_{\Pi X} \leq f_{\Pi X}(C_1, ..., C_k) < \left(1 + \frac{\epsilon}{3}\right) OPT_X$$
$$\implies \left(1 - \frac{\epsilon}{3}\right) f_X(\widetilde{C_1}, ..., \widetilde{C_k}) < \left(1 + \frac{\epsilon}{3}\right) OPT_X$$
$$\implies f_X(\widetilde{C_1}, ..., \widetilde{C_k}) < \frac{1 + \frac{\epsilon}{3}}{1 - \frac{\epsilon}{3}} OPT_X$$

The final step is to note that for $\epsilon \in (0, 1)$ (this range is ok because we are interested in behavior for small ϵ, see ED post 99), it is the case that

$$\epsilon > \epsilon^2 \implies 1 \leq 1 + \frac{\epsilon}{3} - \frac{\epsilon^2}{3} \implies 1 + \frac{\epsilon}{3} \leq 1 + \frac{2\epsilon}{3} - \frac{\epsilon^2}{3} = \left(1 - \frac{\epsilon}{3}\right)(1 + \epsilon) \implies \frac{1 + \frac{\epsilon}{3}}{1 - \frac{\epsilon}{3}} \leq 1 + \epsilon$$

This yields the final result that with probability greater than $1 - \frac{1}{n}$ (i.e. with high probability),

$$\boxed{f_X(\widetilde{C_1}, ..., \widetilde{C_k}) < (1 + \epsilon) OPT_X}$$

■

Problem 5

Note: the below proof is done for unweighted, undirected graphs. Also, by ED post 100, we suppose that k is constant, although the final step of the proof does indeed hold for $k = o(\log n)$.

Solution

Proof. We start with a Lemma relating the number of cycles of length $\leq k$ to the number of edges in a $k-1$-spanner.

Lemma 5. *If a graph $G = (V, E)$ has no cycle of length $\leq k$, then any $k-1$ spanner of G must contain exactly $|E|$ edges.*

Proof of Lemma 5. We prove this by showing that for such a graph G, removing any edge from consideration in the formation of a spanner disallows a $k-1$ spanner to be formed (i.e. for all edges $(u, v) \in E$, we want to show that any subgraph of G that doesn't contain (u, v) cannot be a $k-1$ spanner). Note that if G is disconnected or has no cycles, removing an edge from consideration disallows *any* spanners (since there would be unreachable vertices). So, suppose G is connected and has a cycle. Consider any arbitrary edge $(u, v) \in E$. If (u, v) is not a part of some cycle of G, then removing this edge disconnects u and v, once again disallowing *any* spanner from being formed. So, suppose that (u, v) is a part of some cycle of G; by assumption, this cycle must be of length $> k$. This necessarily means that any simple path between u and v along G is either just the edge (u, v) or has a length $> k-1$ (either we traverse (u, v) or go the whole way around the cycle, which has total length k). So, if we were to remove (u, v) from consideration in a spanner formation, any path from u to v in the spanner must have length at $> k-1 \implies$ it cannot be a $k-1$ spanner. Since this line of reasoning holds for all edges in E, we obtain the result that any subgraph of G with less than $|E|$ edges *cannot* be a $k-1$ spanner. Therefore, any $k-1$ spanner must have $|E|$ edges. ■

The rest of the proof goes as follows: we want to show that there exists some graph $G = (V, E)$ with $|V| = n$ vertices and $|E| > O\left(n^{1+\frac{1}{k}}\right)$ edges that has no cycle of length $\leq k$. From here, we could apply Lemma 5 to see that any $k-1$ spanner has $> O\left(n^{1+\frac{1}{k}}\right)$ edges, and there is therefore no $k-1$ spanner with $O\left(n^{1+\frac{1}{k}}\right)$ edges. By ED post 100, this is what we are trying to show (the first bullet point in the post).

Now, consider a graph $G = (V_G, E_G)$ with $|V_G| = n$ vertices and each possible edge $(u, v) \in V_G \times V_G$ existing with probability $p = n^{-\left(1-\frac{1}{k}-\frac{1}{k^2}\right)}$ independently. We can show that, in expectation, such a graph G has an expected number of edges

$$\mathbb{E}\left[|E_G|\right] = \binom{n}{2} \cdot \frac{1}{n^{1-\frac{1}{k}-\frac{1}{k^2}}} = \frac{n-1}{2} \cdot \frac{1}{n^{-\frac{1}{k}-\frac{1}{k^2}}} = \frac{1}{2} \cdot (n-1) \cdot n^{\frac{1}{k}+\frac{1}{k^2}}$$

Lemma 6. *The expected number of cycles of length $\leq k$ is at most $n^{1+\frac{1}{k}}$.*

Proof of Lemma 6. Define N_l to be the random variable of the number of cycles of length l in G. Note that from any subset of l vertices, in order for there to be a cycle of length l within this subset there must be precisely l edges made. So, within each set of vertices of size l (say, $V_l \subset V_G$), the probability that it contains a cycle of length l is

$$\frac{l!}{2l} \cdot \left(\frac{1}{n^{1-1/k-1/k^2}}\right)^l,$$

where we need the factor of $\frac{l!}{2l}$ to account for the different possible permutations of the cycle ($l!$), as well as the symmetry of a cycle to starting position ($\frac{1}{l}$) and direction ($\frac{1}{2}$). So, since there are $\binom{n}{l}$ possible subsets of size l, each forming an l-cycle with the above probability, we find that the expected number of cycles of length l is

$$\mathbb{E}\left[N_l\right] = \binom{n}{l} \cdot \frac{l!}{2l} \cdot \left(\frac{1}{n^{1-1/k-1/k^2}}\right)^l = \frac{n!}{2l!} \cdot \left(\frac{1}{n^{1-1/k-1/k^2}}\right)^l$$

Therefore, the expected number of cycles with length $\leq k$ is given by

$$\sum_{l=3}^{k} \mathbb{E}\left[N_l\right] \leq \sum_{l=0}^{k} \frac{n!}{2l!} \cdot \left(\frac{1}{n^{1-1/k-1/k^2}}\right)^l \leq \frac{1}{2} \sum_{l=0}^{k} \left(\frac{n}{n^{1-1/k-1/k^2}}\right)^l,$$

where the last inequality comes from the fact that $\frac{n!}{l!} \leq n^l$. So, we can use a geometric sum to see that this is equal to

$$= \frac{1}{2} \cdot \left(\frac{n^{1+2/k+1/k^2} - 1}{n^{1/k+1/k^2} - 1}\right) \leq n^{1+1/k},$$

where this last inequality holds because the left term inside the parenthesis converges to the right hand side, and is always much closer than a factor of $\frac{1}{2}$ (you can use Desmos to see this). So, the expected number of cycles of length $\leq k$ is as desired. ■

So, we have seen that the graph G, constructed as above, in expectation has the properties that it has

$$\frac{1}{2} \cdot (n-1) \cdot n^{\frac{1}{k}+\frac{1}{k^2}}$$

edges and no more than $n^{1+1/k}$ cycles. So, we can say that such a graph G' certainly exists, since the random graph G in expectation is G'. From here, we can proceed deterministically on G', confident that it exists. If we take G' and remove one edge from each of its cycles of length $\leq k$ (thus breaking each such cycle), we result in a new graph, say $\widetilde{G'}$, with no cycles of length $\leq k$ and with at least

$$\frac{1}{2} \cdot (n-1) \cdot n^{\frac{1}{k}+\frac{1}{k^2}} - n^{1+1/k}$$

edges. Note that the term $n \cdot n^{1/k+1/k^2} = n^{1+1/k+1/k^2}$ dominates the entire expression, and so we find that the number of edges in $\widetilde{G'}$ is of the order

$$O\left(n^{1+1/k+1/k^2}\right) > O\left(n^{1+1/k}\right)$$

So, we have shown the existence of a graph $\widetilde{G'}$ with more than $O\left(n^{1+1/k}\right)$ edges that has no cycles of length $\leq k$. By Lemma 5, we see that every $k-1$ spanner of $\widetilde{G'}$ must have $O\left(n^{1+1/k+1/k^2}\right)$ edges, and so there is no $k-1$ spanner with $O\left(n^{1+1/k}\right)$ edges. ■