NYU CS-GY 6763: Algorithmic Machine Learning and Data Science	Spring 2025

Homework Solution key

Name: Solution key

Problem 1

1. $\mathbb{E}[X] = 1/2$, so by Markov's $\Pr[X \geq 7/8] \leq \frac{1/2}{7/8} = .571$.

2. $\mathbb{E}[(X - \mathbb{E}\, X)^2] = \int_{0^1} (x - .5)^2 dx = .08333\ldots$ via Wolfram Alpha. So by Chebyshev's $\Pr[|X - .5| \geq (7/8 - .5)] \leq \frac{.08333}{(7/8-.5)^2} = .593$.

3. $\mathbb{E}[X^2] = \int_0^1 x^2 dx = \frac{1}{3}$, so by Markov's $\Pr[X^2 \geq (7/8)^2] \leq \frac{1/3}{(7/8)^2} = .435$. So the uncentered moment gives a better bound.

4. The general equation for an upper bound is:

$$\frac{1/(q+1)}{(7/8)^q}.$$

For $q = 3, 4, \ldots, 10$ we get values:

$$\begin{bmatrix} .373 & .341 & .325 & .318 & .318 & .323 & .333 & .346 \end{bmatrix}$$

So using higher moments at first improves our bound, but then eventually starts giving a weaker bound. The tighted bound is obtained at $q = 6$ and $q = 6$.

5. Let g be the step function which is 0 for $X < 7/8$ and 1 for $X \geq 7/8$. Then we have $\Pr[X \geq 7/8] = \Pr[g(X) \geq 1]$. We have $\mathbb{E}[g(X)] = \frac{1}{8}$, so $\Pr[g(X) \geq 1] \leq \frac{1}{8}$ by Markov's, which gives the tight bound.

Problem 2

I will write this later. This is the standard "median trick". The difficult observation is that, for the median to give error $> \epsilon$, it must be that more than half of the runs of the algorithm give error $> \epsilon$. Why? Suppose the median gave an answer more than ϵ above the correct value. Then all values above the median (which is 1/2 of the runs) have to also be $> \epsilon$ above the correct value.

Once you make this observation, they just need to prove that if you flip a coin that comes up heads with probability 2/3, you won't have $< 50\%$ heads in a sample of size $O(\log(1/\delta))$ with probability $1 - \delta$. This was actually done as an example in Lecture 3 (not with exactly those numbers). Best way is to use Chernoff bound.

Problem 3

1. Once there are $n+1$ servers in this setup, the expected number of items on the $(n+1)^{\text{st}}$ server is $\frac{m}{n+1}$, by symmetry. All of these items (and only these items) must have been relocated when the $(n+1)^{\text{st}}$ server was added. So the expected number of items that move is $\frac{m}{n+1}$.

2. For a server S to own more than a $c \log n/n$ fraction of the interval , it would need to be that *no other server* falls within distance $c \log n/n$ to the left of the server. We can choose the random location of server S first. Then the probability of any one server landing within distance $c \log n/n$ from S's left is $c \log n/n$. So the probability *no servers* land that close is:

$$(1 - c \log n/n)^{n-1} \leq \frac{1}{10n},$$

as long as we choose c to be a large enough constant (same analysis as homework 1). By a *union bound*, we thus have that no server owns more than an $O(\log n/n)$ fraction of the interval with probability $\geq 1 - n\frac{1}{10n} = \frac{9}{10}$ which proves the claim.

3. From Part 2, we could have equivalently proven that no server owns more than a $c \log n/n$ fraction of the interval with probability 19/20 (by choosing c larger). For the rest of the problem, assume that this event happening.

For servers $S_1, \ldots, S_n$ let $Y_i^{(j)}$ be the indicator random variable that item j lands within distance $c \log n/n$ to S_i's left. Let X_i equal $X_i = \sum_{j=1}^{m} Y_i^{(j)}$. Since we assumed that no server owns more than a $c \log n/n$ fraction of the interval, X_i is an *upper bound* on the number of items assigned to server i. So it suffices to show that X_i is not too large for all i.

To do so, note that, for a fixed i, $Y_i^{(1)}, Y_i^{(2)}, \ldots, Y_i^{(m)}$ are an independent $\{0, 1\}$ random variables, where each is 1 with probability exactly $c \log n/n$. So they are just biased coin flips!

Let $c > 2$ be a sufficiently large constant. Using the Chernoff bound from class with $\epsilon = c$, we get that:

$$\Pr[X_i \geq 2c \cdot \frac{m \log n}{n}] \leq e^{\frac{-c^2 m \log n/n}{2+c}} \leq e^{\frac{-c \log n}{2}} \leq \frac{1}{20n},$$

for large enough c. The last inequality uses that $m > n$ (as specified in the problem).

We conclude via a union bound that no server is assigned more than $O(m \log n/n)$ items with probability $\frac{19}{20}$.

There's one last step – we needed two events to hold for our proof to go through: 1) no server owns more than a $c \log n/n$ fraction of the interval and 2) no server was assigned two many items. Since each holds with probability 19/20, by another union bound, both hold with probability 9/10.

Problem 4 (a)

1. **Expectation Calculation.** As in class, we have that $\mathbb{E}[\|\Pi x\|_2^2] = \mathbb{E}[\langle \pi, x\rangle^2]$, where π is a single unscaled row from the matrix Π. I.e. π has length n and contains random ± 1 entries. We have:

$$\mathbb{E}[\langle \pi, x\rangle^2] = \mathbb{E}\left[\left(\sum_{j=1}^{n} \pi_j x_j\right)^2\right] = \mathbb{E}\left[\sum_{j=1}^{n} \pi_j^2 x_j^2\right] + \mathbb{E}\left[\sum_{i\neq j}^{n} \pi_i\pi_j x_j x_i\right]$$
$$= \sum_{j=1}^{n} \mathbb{E}\left[\pi_j^2\right] x_j^2 + \sum_{i\neq j}^{n} \mathbb{E}\left[\pi_i\pi_j\right] x_j x_i.$$

The last equality follows from linearity of expectation. Since π_i is independent of π_j, we have that for $j \neq i$, $\mathbb{E}\left[\pi_i\pi_j\right] = \mathbb{E}[\pi_i]\,\mathbb{E}[\pi_j] = 0$. On the other hand $\pi_j^2 = 1$ deterministically, so we have $\mathbb{E}\left[\pi_j^2\right] = 1$. Plugging in above, we find that

$$\mathbb{E}[\langle \pi, x\rangle^2] = \sum_{j=1}^{n} x_j^2 + \sum_{i\neq j}^{n} 0 \cdot x_j x_i = \sum_{j=1}^{n} x_j^2 = \|x\|_2^2,$$

as desired.

Variance Calculation. Since $\|\Pi x\|_2^2 = \frac{1}{k}\sum_{i=1}^{k}\langle \pi^i, x\rangle^2$, where $\pi^1, \ldots, \pi^k$ are the unscaled rows of Π, we first observe that $\operatorname{Var}[\|\Pi x\|_2^2] = \frac{1}{k}\operatorname{Var}[\langle \pi, x\rangle^2]$ for a single random ± 1 vector π. So we just need to bound $\operatorname{Var}[\langle \pi_i, x\rangle^2]$. This gets a bit tricky! There are many ways to do it, but I think the easiest way is to take advantage of linearity of variance by writing:

$$\langle \pi, x\rangle^2 = \sum_{j=1}^{n} \pi_j^2 x_j^2 + 2\sum_{i>j} \pi_i\pi_j x_i x_j.$$

The terms in the first part of the sum are actually deterministic, since $\pi_j = 1$. The terms in the second part of the sum are random, but they are *pairwise independent* since $\pi_i\pi_j$ is random ± 1 and independent from any $\pi_i\pi_k$, $\pi_k\pi_j$, or $\pi_k\pi_\ell$. They are not mutually independent, but we only need pairwise independence to apply linearity of variance. Note that to make this claim it's important that I used the form $2\sum_{i>j}$ instead of $\sum_{i\neq j}$. If I did the later, there would be repeated random variables in the sum ($\pi_i\pi_j x_i x_j$ and $\pi_j\pi_i x_j x_i$). Writing the other way removes duplicates.

$$\operatorname{Var}[\langle \pi, x\rangle^2] = \sum_{j=1}^{n} \operatorname{Var}[\pi_j^2 x_j^2] + 4\sum_{i>j} \operatorname{Var}[\pi_i\pi_j x_j x_i] = 0 + 4\sum_{i>j} x_j^2 x_i^2.$$

Then finally we observe that:

$$\|x\|_2^4 = \|x\|_2^2 \cdot \|x\|_2^2 = (x_1^2 + \ldots + x_n^2)\cdot(x_1^2 + \ldots + x_n^2) \geq 2\sum_{i>j} x_j^2 x_i^2.$$

Putting this together we have that $\operatorname{Var}[\langle \pi, x\rangle^2] \leq 2\|x\|_2^4$ and the result follows since $\operatorname{Var}[\|\Pi x\|_2^2] = \frac{1}{k}\operatorname{Var}[\langle \pi, x\rangle^2]$ as claimed above.

2. This just follows directly from Chebyshev's.

3. It's almost the same analysis as in part 1. The first thing to observe is that:

$$\langle \Pi x, \Pi y \rangle = \frac{1}{k} \sum_{i=1}^{k} \langle \pi^i, x \rangle \langle \pi^i, y \rangle.$$

So we have that $\mathbb{E}[\langle \Pi x, \Pi y \rangle] = \mathbb{E}[\langle \pi, x \rangle \langle \pi, y \rangle]$ and $\mathrm{Var}[\langle \Pi x, \Pi y \rangle] = \frac{1}{k} \mathrm{Var}[\langle \pi, x \rangle \langle \pi, y \rangle]$, where π is a single random ± 1 vector. We also have that

$$\langle \pi, x \rangle \langle \pi, y \rangle = \left(\sum_{j=1}^{n} \pi_j x_j \right) \cdot \left(\sum_{j=1}^{n} \pi_j y_j \right) = \sum_{i=1}^{n} \pi_i^2 x_i y_i + \sum_{j \neq i} \pi_i \pi_j x_i y_j.$$

From this it's clear that

$$\mathbb{E}[\langle \Pi x, \Pi y \rangle] = \mathbb{E}[\langle \pi, x \rangle \langle \pi, y \rangle] = \sum_{i=1}^{n} x_i y_i = \langle x, y \rangle,$$

as desired.

The variance calculation is also a bit tricky since we need to make sure our sums involve pairwise independent random variables. We have that:

$$\langle \pi, x \rangle \langle \pi, y \rangle = \sum_{i=1}^{n} \pi_i^2 x_i y_i + \sum_{j > i} \pi_i \pi_j (x_i y_j + x_j y_i).$$

Applying linearity of variance, we find that

$$\begin{aligned} \mathrm{Var}[\langle \pi, x \rangle \langle \pi, y \rangle] = \sum_{j > i} (x_i y_j + x_j y_i)^2 &= \sum_{j > i} x_i^2 y_j^2 + x_j^2 y_i^2 + 2 x_i x_j y_i y_j \\ &\leq 2 \sum_{j > i} x_i^2 y_j^2 + x_j^2 y_i^2 \\ &\leq 2(x_1^2 + \ldots + x_n^2)(y_1^2 + \ldots + y_n^2) \\ &= 2 \|x\|_2^2 \|y\|_2^2. \end{aligned}$$

In second to last inequality we have used that for any a, b, $2ab \leq a^2 + b^2$, which follows from the fact that $(a - b)^2 \geq 0$ for all a, b (this is technically called the AM-GM inequality).

Overall, we get a variance bound of:

$$\mathrm{Var}[\langle \Pi x, \Pi y \rangle] \leq \frac{2}{k} \|x\|_2^2 \|y\|_2^2.$$

Once they get the mean and variance, the bound just follows from applying Chebyshev inequality again. .

Problem 4 (b)

1. Construct 2 length U binary vectors x and y where $x_i = 1$ if $i \in X$ and 0 otherwise, and $y_i = 1$ if $i \in Y$ and 0 otherwise. Note that $|X \cap Y|$ is exactly equal to $\langle x, y \rangle$, so we can estimate the quantity using sketches Πx and Πy. If we set $k = O(1/\epsilon^2)$, then with 9/10 probability we will have:

$$|\langle x, y \rangle - \langle \Pi x, \Pi y \rangle| \leq \epsilon \|x\|_2 \|y\|_2$$

Note that $\|x\|_2^2 = |X|$ and $\|y\|_2^2 = |Y|$, which yields the bound.