36-755, Fall 2017
Homework 2 Solution

Due Wed Oct 4 by 5:00 pm in Jisu's mailbox

Points: 100 + 2 pts total for the assignment.

1. Let $(X_1, \ldots, X_n)$ be zero-mean $SG(\sigma^2)$ random variables (not assumed independent). Give bounds on
$$\mathbb{E}\left[\max_i |X_i|\right]$$
and
$$\mathbb{P}\left(\max_i |X_i| \geq t\right), \quad t \geq 0.$$
You can use the corresponding result, proved in class, for $\max_i X_i$.

Points: 10 + 2 pts.

Solution.
Let $Y_1, \cdots, Y_{2n}$ be random variables defined as for $1 \leq i \leq n$, $Y_i := X_i$ and $Y_{n+i} := -X_i$. Then for all $1 \leq i \leq n$, $\max\{Y_i, Y_{n+i}\} = \max\{X_i, -X_i\} = |X_i|$, and hence
$$\max_{1 \leq i \leq 2n} Y_i = \max_{1 \leq i \leq n} |X_i|.$$
And since $X_i \in SG(\sigma^2)$ implies $-X_i \in SG(\sigma^2)$ as well, so for all $1 \leq i \leq 2n$, $Y_i \in SG(2\sigma^2)$. Hence from Lecture 4, the expectation $\mathbb{E}\left[\max_{1 \leq i \leq n} |X_i|\right]$ can be bounded as
$$\begin{aligned}\mathbb{E}\left[\max_{1 \leq i \leq n} |X_i|\right] &= \mathbb{E}\left[\max_{1 \leq i \leq 2n} Y_i\right] \\ &\leq \sqrt{2\sigma^2 \log(2n)}.\end{aligned}$$
Also from Hoeffding's inequality and union bound, the probability $\mathbb{P}\left(\max_{1 \leq i \leq n} |X_i| \geq t\right)$ for $t \geq 0$ can be bounded as
$$\mathbb{P}\left(\max_{1 \leq i \leq n} |X_i| \geq t\right) \leq n \sum_{i=1}^{n} \mathbb{P}\left(|X_i| \geq t\right) \leq 2n \exp\left(-\frac{t^2}{2\sigma^2}\right).$$
There is 2 bonus points for those who get the upper bound of $\mathbb{E}\left[\max_i |X_i|\right]$ as of order $\sigma\sqrt{\log n}$.

2. Let $(X_1, \ldots, X_n)$ be independent random variables with mean zero and let $(a_1, \ldots, a_n) \in \mathbb{R}^n$. Compute bounds for
$$\mathbb{P}\left(|\sum_{i=1}^{n} a_i X_i| \geq t\right)$$
under the assumption that the X_i's are in $SG(\sigma^2)$ and in $SE(\nu^2, \alpha)$. Compare the bounds. When does one dominate the other?

Points: 10 pts.

Solution.

Note that $\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i\right] = 0$.

First, consider the case when $X_i \in SE(\nu^2, \alpha)$. Note that $X_i \in SE(\nu^2, \alpha)$ implies that

$$\mathbb{E}\left[\exp\left(\lambda a_i X_i\right)\right] \leq \exp\left(\frac{\lambda^2 a_i^2 \nu^2}{2}\right)$$

holds for $|\lambda| \leq \frac{1}{a_i \alpha}$. And since $X_1, \ldots, X_n$ are independent,

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{n} a_i X_i\right)\right] &= \prod_{i=1}^{n} \mathbb{E}\left[\exp\left(\lambda a_i X_i\right)\right] \\
&\leq \prod_{i=1}^{n} \exp\left(\frac{\lambda^2 a_i^2 \nu^2}{2}\right) \\
&= \exp\left(\frac{\lambda^2 \nu^2 \|a\|_2^2}{2}\right)
\end{aligned}$$

holds for $|\lambda| \leq \min_{1 \leq i \leq n} \frac{1}{|a_i| \alpha}$. Hence $\sum_{i=1}^{n} a_i X_i \in SE\left(\|a\|_2^2 \nu^2, \max_{1 \leq i \leq n} \{|a_i|\} \alpha\right)$. Then from two-sided sub-exponential tail bound,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i\right| \geq t\right) \leq \begin{cases} 2\exp\left(-\frac{t^2}{2\|a\|_2^2 \nu^2}\right) & \text{for } 0 \leq t \leq \frac{\|a\|_2^2 \nu^2}{\max_{1 \leq i \leq n} \{|a_i|\} \alpha} \\ 2\exp\left(-\frac{t}{2\max_{1 \leq i \leq n} \{|a_i|\} \alpha}\right) & \text{for } t > \frac{\|a\|_2^2 \nu^2}{\max_{1 \leq i \leq n} \{|a_i|\} \alpha} \end{cases}$$

Now consider the case when $X_i \in SG(\sigma^2)$. Since $X_i \in SE(\sigma^2, 0)$, $\sum_{i=1}^{n} a_i X_i \in SE\left(\|a\|_2^2 \sigma^2, 0\right) = SG(\|a\|_2^2 \sigma^2)$ as well. Then from two-sided Hoeffding's inequality,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\|a\|_2^2 \sigma^2}\right).$$

For comparison, assume that $\nu^2 = \sigma^2$. Note that for $t > \frac{\|a\|_2^2 \sigma^2}{\max_{1 \leq i \leq n} \{|a_i|\} \alpha}$,

$$2\exp\left(-\frac{t}{2\max_{1 \leq i \leq n} \{|a_i|\} \alpha}\right) \leq 2\exp\left(-\frac{t^2}{2\|a\|_2^2 \sigma^2}\right)$$

holds, so bounds for $SG(\sigma^2)$ is always better than bounds for $SE(\nu^2, \alpha)$.

Now, let $n \to \infty$, $\max_{1 \leq i \leq n} \{|a_i|\}$ is remaining constant, and consider the case when $||a||_2 \approx c_1 \times \max_{1 \leq i \leq n} \{|a_i|\}$ and $||a||_2 \approx c_2 n \times \max_{1 \leq i \leq n} \{|a_i|\}$ for some constant c_1, c_2 that does not depend on n, i.e. when a is sparse and when a is evenly distributed, and consider the sub-exponential bound. For the first case,

$$\frac{\|a\|_2^2 \nu^2}{\max_{1 \leq i \leq n} \{|a_i|\} \alpha} \approx \frac{c_1^2 \max_{1 \leq i \leq n} \{|a_i|\} \nu^2}{\alpha},$$

so tail bound remains to have exponential tail behavior. For the second case,

$$\frac{\|a\|_2^2\,\nu^2}{\max\limits_{1\le i\le n}\{|a_i|\}\,\alpha} \approx \frac{c_2^2 n^2 \max\limits_{1\le i\le n}\{|a_i|\}\,\nu^2}{\alpha} \to \infty,$$

so tail bound asymptotically behaves like Gaussian. This is related to the Lindeberg condition in the sense that when the variances of independent variables $a_i X_i$ are evenly distributed, then Central Limit Theorem can be applied to their sum $\sum\limits_{i=1}^{n} a_i X_i$ so that the sum asymptotically behaves like Gaussian.

3. Orlicz norms. We have defined sub-gaussian and sub-exponential variables in terms of bounds on the moment generating functions. There exists an equivalent and more general way of expressing these properties using *Orlicz Norms* of random variables, which is more abstract but, at the same time, leads to simpler calculation. You will explore these concepts in this exercise. First, do the following problems in the book:

   (a) Exercise 2.18: (Orlicz norms).
   Let $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be a strictly increasing convex function that satisfies $\psi(0) = 0$. The ψ-*Orlicz* norm of a random variable X is defined as

   $$\|X\|_\psi := \inf\left\{t > 0 | \mathbb{E}\left[\psi(t^{-1}|X|) \le 1\right]\right\},$$

   where $\|X\|_\psi$ is infinite if there is no finite t for which the expectation $\mathbb{E}\left[\psi(t^{-1}|X|)\right]$ exists. For the functions $u \mapsto u^q$ for some $q \in [1, \infty]$, then the Orlicz norm is simply the usual ℓ_q-norm $\|X\|_q = (\mathbb{E}[|X|^q])^{1/q}$. In this exercise, we consider the Orlicz norms $\|\cdot\|_{\psi_q}$ defined by the convex functions $\psi_q(u) = \exp(u^q) - 1$, for $q \ge 1$.

      i. If $\|X\|_{\psi_q} < +\infty$, show that there exist positive constants c_1, c_2 such that

      $$\mathbb{P}[|X| > t] \le c_1 \exp(-c_2 t^q) \quad \text{for all } t > 0. \tag{1}$$

      (In particular, you should be able to show that this bound holds with $c_1 = 2$ and $c_2 = \|X\|_{\psi_q}^{-q}$.)

      ii. Suppose that a random variable Z satisfies the tail bound (1). Show that $\|X\|_{\psi_q}$ is finite.

   (b) Exercise 2.19. (Maxima of Orlicz variables).
   Recall the definition of Orlicz norm from Exercise 2.18. Let $\{X_i\}_{i=1}^n$ be an i.i.d. sequence of zero-mean random variables with finite Orlicz norm $\sigma = \|X_i\|_\psi$. Show that

   $$\mathbb{E}\left[\max_{i=1,\ldots,n} |X_i|\right] \le \sigma \psi^{-1}(n).$$

   In this context, a random variable is said to be sub-gaussian if there exists a $K > 0$ such that

   $$\mathbb{E}\left[e^{X^2/K^2}\right] \le 2 \tag{2}$$

   and sub-exponential if there exists a constant $K' > 0$ such that

   $$\mathbb{E}\left[e^{|X|/K'}\right] \le 2. \tag{3}$$

   If X is sub-gaussian, its *sub-gaussian norm* is the smallest K satisfying (2), which correspond to $\|X\|_{\psi_2}$. Similarly, if X is sub-exponential, its *sub-exponential norm* is $\|X\|_{\psi_1}$, the smallest K' satisfying (3).

(c) Prove that X is sub-gaussian if and only if X^2 is sub-exponential and

$$\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2$$

(d) If X and Y are sub-gaussians, then XY is sub-exponential with

$$\|XY\|_{\psi_1} \leq \|X\|_{\psi_2}\|Y\|_{\psi_2}.$$

The last two properties would have made problems 2 and 8 in Homework 1 easier...

Remarks. (Please read) it is possible to show that the above definitions are equivalent to the ones given in class: see the Appendix of Chapter 2 of the textbook. In particular, if X is sub-exponential then

$$\mathbb{E}\left[e^{\lambda X}\right] \leq \exp \lambda^2 \|X\|_{\psi_1}^2, \qquad \forall |\lambda| \leq \frac{1}{\|X\|_{\psi_1}}.$$

From this, it is possible to derive the following, equivalent, versions of Hoeffding and Bernstein inequalities which you will also find in the literature.

- **Hoeffding inequality**. Let $X_1, \ldots, X_n$ be independent, mean-zero sub-gaussian variables. Then, there exists a universal constant $c > 0$ such that, for any $t \geq 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \geq t\right) \leq 2 \exp\left(-\frac{ct^2}{\sum_{i=1}^n \|X_i\|_{\psi_2}^2}\right)$$

- **Bernstein inequality**. Let $X_1, \ldots, X_n$ be independent, mean-zero sub-exponential variables. Then, there exists a universal constant $c > 0$ such that, for any $t \geq 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \geq t\right) \leq 2 \exp\left(-c \min\left\{-\frac{t^2}{\sum_{i=1}^n \|X_i\|_{\psi_1}^2}, \frac{t}{\sum_{i=1}^n \|X_i\|_{\psi_1}}\right\}\right)$$

In other words, mapping to the notation used in class, $\sigma = \|X\|_{\psi_2}$ and $\nu = \alpha = \|X\|_{\psi_1}$.

Points: 21 pts = 6 + 6 + 4 + 5.

Solution.

(a) i.

Note that $\|X\|_{\psi_q} < +\infty$ implies that for all $t > \|X\|_{\psi_q}$, $\mathbb{E}\left[\psi_q\left(t^{-1}|X|\right)\right] \leq 1$. And since $\mathbb{E}\left[\psi_q\left(t^{-1}|X|\right)\right] = \mathbb{E}\left[\exp\left(t^{-q}|X|^q\right) - 1\right]$ is a continuous function of t, hence $\mathbb{E}\left[\exp\left(\frac{|X|^q}{\|X\|_{\psi_q}^q}\right) - 1\right] \leq 1$ holds, i.e.

$$\mathbb{E}\left[\exp\left(\frac{|X|^q}{\|X\|_{\psi_q}^q}\right)\right] \leq 2.$$

Then by using Markov's inequality, $\mathbb{P}\left[|X| > t\right]$ can be bounded as

$$\begin{aligned}
\mathbb{P}\left(|X| > t\right) = \mathbb{P}&\left(\exp\left(\frac{|X|^q}{\|X\|_{\psi_q}^q}\right) > \exp\left(\frac{t^q}{\|X\|_{\psi_q}^q}\right)\right) \\
&\leq \exp\left(-\frac{t^q}{\|X\|_{\psi_q}^q}\right) \mathbb{E}\left[\exp\left(\frac{|X|^q}{\|X\|_{\psi_q}^q}\right)\right] \qquad \text{(using Markov inequality)} \\
&\leq 2 \exp\left(-\|X\|_{\psi_q}^{-q} t^q\right).
\end{aligned}$$

(a) ii.

Suppose X satisfies $\mathbb{P}\left[|X| > t\right] \le c_1 \exp(-c_2 t^q)$. Then $\exp\left(t^{-q}|X|^q\right) \ge 1$ implies that

$$\int_0^1 \mathbb{P}\left(\exp\left(t^{-q}|X|^q\right) \ge u\right) du = \int_0^1 1 du = 1,$$

hence

$$\begin{aligned}
\mathbb{E}\left[\psi_q\left(t^{-1}|X|\right)\right] &= \mathbb{E}\left[\exp\left(t^{-q}|X|^q\right)\right] - 1 \\
&= \int_1^\infty \mathbb{P}\left(\exp\left(t^{-q}|X|^q\right) \ge u\right) du \\
&= \int_1^\infty \mathbb{P}\left(|X| \ge t\left(\log u\right)^{\frac{1}{q}}\right) du \\
&\le \int_1^\infty c_1 \exp\left(-c_2 t^q \log u\right) du \\
&= c_1 \int_1^\infty u^{-c_2 t^q} du \\
&= \frac{c_1}{c_2 t^q - 1},
\end{aligned}$$

when $c_2 t^q > 1$ is satisfied. Then $t \ge \left(\frac{c_1+1}{c_2}\right)^{\frac{1}{q}}$ implies both $c_2 t^q > 1$ and $\mathbb{E}\left[\psi_q\left(t^{-1}|X|\right)\right] \le \frac{c_1}{c_2 t^q - 1} \le 1$. Hence

$$\|X\|_{\psi_q} \le \left(\frac{c_1+1}{c_2}\right)^{\frac{1}{q}} < +\infty.$$

(b)

Since ψ is a convex function, Jensen's inequality can be applied as

$$\begin{aligned}
\psi\left(\mathbb{E}\left[\sigma^{-1}\max_{i=1,\dots,n}|X_i|\right]\right) &\le \mathbb{E}\left[\psi\left(\sigma^{-1}\max_{i=1,\dots,n}|X_i|\right)\right] \quad \text{(using Jensen's inequality)} \\
&= \mathbb{E}\left[\max_{i=1,\dots,n}\psi\left(\sigma^{-1}|X_i|\right)\right] \\
&\le \sum_{i=1}^n \mathbb{E}\left[\psi\left(\sigma^{-1}|X_i|\right)\right] \le n.
\end{aligned}$$

And since ψ is strictly increasing function with $\psi(0) = 0$, $\psi^{-1} : \mathbb{R}_+ \to \mathbb{R}_+$ is well defined and preserves inequality, hence

$$\mathbb{E}\left[\max_{i=1,\dots,n}|X_i|\right] \le \sigma\psi^{-1}(n).$$

(c)

Note that X being sub-gaussian implies $\mathbb{E}\left[\exp\left(\frac{X^2}{\|X\|_{\psi_2}^2}\right)\right] \le 2$. Then $\mathbb{E}\left[\exp\left(\frac{|X^2|}{\|X\|_{\psi_2}^2}\right)\right] \le 2$ as well, so X^2 is sub-exponential with $\|X^2\|_{\psi_1} \le \|X\|_{\psi_2}^2$. Conversely, X^2 being sub-exponential implies $\mathbb{E}\left[\exp\left(\frac{|X^2|}{\|X\|_{\psi_1}}\right)\right] \le 2$. Then $\mathbb{E}\left[\exp\left(\frac{X^2}{\left(\sqrt{\|X^2\|_{\psi_1}}\right)^2}\right)\right] \le 2$ as well, so X is sub-gaussian with $\|X\|_{\psi_2} \le \sqrt{\|X^2\|_{\psi_1}}$. Hence X is sub-gaussian if and only if X^2 is sub-exponential, and

$$\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2.$$

(d)

If X and Y are sub-gaussian, then $\mathbb{E}\left[\exp\left(\frac{X^2}{\|X\|_{\psi_2}^2}\right)\right] \leq 2$ and $\mathbb{E}\left[\exp\left(\frac{Y^2}{\|Y\|_{\psi_2}^2}\right)\right] \leq 2$. Now, from $\exp(\cdot)$ being increasing function, AM-GM inequality, and Cauchy-Schwarz inequality,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\frac{|XY|}{\|X\|_{\psi_2}\|Y\|_{\psi_2}}\right)\right] \\
&\leq \mathbb{E}\left[\exp\left(\frac{1}{2}\frac{X^2}{\|X\|_{\psi_2}^2} + \frac{1}{2}\frac{Y^2}{\|Y\|_{\psi_2}^2}\right)\right] \text{ (exp increasing, AM-GM inequality)} \\
&\leq \sqrt{\mathbb{E}\left[\exp\left(\frac{X^2}{\|X\|_{\psi_2}^2}\right)\right]\mathbb{E}\left[\exp\left(\frac{Y^2}{\|Y\|_{\psi_2}^2}\right)\right]} \text{ (Cauchy-Schwarz)} \\
&\leq 2.
\end{aligned}$$

Hence XY is sub-exponential with

$$\|XY\|_{\psi_1} \leq \|X\|_{\psi_2}\|Y\|_{\psi_2}.$$

4. (Reading exercise. **Not to be graded for correctness, but only for effort**)
Suppose that $X_1, \ldots, X_n$ are zero-mean, independent random variables belonging to the class $SG(\sigma^2)$ and $A = (A_{i,j})$ a $n \times n$ matrix. Let

$$\|A\|_{\text{op}} = \max_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|}{\|x\|}$$

and

$$\|A\|_{\text{HS}} = \sqrt{\sum_{i=1}^n \sum_{j=1}^n A_{i,j}^2}$$

be the operator and the Hilbert-Schmidt (or Frobenius) norm of A. Notice that $\|A\|_{\text{op}}$ is also the largest absolute eigenvalue of A. The goal of this exercise is to derive an exponential inequality for the probability

$$\mathbb{P}\left(\left|X^\top AX - \mathbb{E}\left[X^\top AX\right]\right| \geq t\right), \forall t \geq 0.$$

Do so by reproducing the proof of Theorem 1.1 from the following reference, using the definition of sub-Gaussian and sub-Exponential variables given in class.

- Rudelson, M., and Vershynin, R. (2013). Hanson-Wright inequality and sub-gaussian concentration. Electron. Commun. Probab., 18(82), 1- 9.

Notice that the definitions of sub-gaussian and sub-exponential variables in this paper is different than the ones given in class and correspond to the ones in problem 3. Make sure to keep track of the constants that depend on σ^2.

Points: 10 pts.

Solution.

The difference $X^\top AX - \mathbb{E}\left[X^\top AX\right]$ can be factorized as

$$
\begin{aligned}
X^\top AX - \mathbb{E}\left[X^\top AX\right] &= \sum_i A_{ii}\left(X_i^2 - \mathbb{E}\left[X_i^2\right]\right) &(4)\\
&\quad + \sum_{i\neq j} A_{ij}X_iX_j. &(5)
\end{aligned}
$$

Step 1: diagonal sum.

We consider (4) first. Note that from HW1 Problem 8, $X_i^2 \in SE\left(256\sigma^4, 16\sigma^2\right)$ holds, and hence $A_{ii}X_i^2 \in SE\left(256\sigma^4 A_{ii}^2, 16\sigma^2|A_{ii}|\right)$ holds. Since X_i^2, $i = 1,\dots,n$ are independent, from HW1 Problem 6 Details,

$$
\sum_i A_{ii}X_i^2 \in SE\left(256\sigma^4 \sum_i A_{ii}^2, 16\sigma^2 \max_i |A_{ii}|\right).
$$

Then $\sum_i A_{ii}^2 \le \sum_{i,j} A_{ij}^2 = \|A\|_{HS}^2$ and $\max_i |A_{ii}| \le \|A\|$, hence

$$
\sum_i A_{ii}X_i^2 \in SE\left(256\sigma^4 \|A\|_{HS}^2, 16\sigma^2 \|A\|\right).
$$

Hence by applying sub-exponential tail bound,

$$
\begin{aligned}
\mathbb{P}\left(\sum_i A_{ii}\left(X_i^2 - \mathbb{E}\left[X_i^2\right]\right) > t\right) &\le \begin{cases} \exp\left(-\frac{t^2}{512\sigma^4\|A\|_{HS}^2}\right) & 0 \le t \le 16\sigma^2\frac{\|A\|_{HS}^2}{\|A\|}, \\ \exp\left(-\frac{t}{32\sigma^2\|A\|}\right) & t \ge 16\sigma^2\frac{\|A\|_{HS}^2}{\|A\|}, \end{cases}\\
&= \exp\left(-\min\left\{\frac{t^2}{512\sigma^4\|A\|_{HS}^2}, \frac{t}{32\sigma^2\|A\|}\right\}\right). \qquad (6)
\end{aligned}
$$

Step 2: decoupling.

Now consider (5), and let $S := \sum_{i\neq j} A_{ij}X_iX_j$. Consider independent Bernoulli random variables $\delta_i \sim Bernoulli\left(\frac{1}{2}\right)$ that are independent of X_i's. Since

$$
\mathbb{E}\left[\delta_i(1-\delta_j)\right] = \begin{cases} 0, & i = j \\ \frac{1}{4}, & i \neq j \end{cases}
$$

we have

$$
S = 4\mathbb{E}\left[S_\delta | X\right], \qquad \text{where} \qquad S_\delta = \sum_{i,j} \delta_i(1-\delta_j)A_{ij}X_iX_j
$$

Then for any $\lambda \in \mathbb{R}$, by applying Jensen's inequality on convex function $x \mapsto \exp(\lambda x)$,

$$
\begin{aligned}
\mathbb{E}\left[\exp\left(\lambda S\right)\right] &= \mathbb{E}\left[\exp\left(4\lambda\mathbb{E}\left[S_\delta|X\right]\right)\right]\\
&\le \mathbb{E}\left[\mathbb{E}\left[\exp\left(4\lambda S_\delta\right)|X\right]\right]\\
&= \mathbb{E}\left[\exp\left(4\lambda S_\delta\right)\right].
\end{aligned}
$$

Now for fixed δ, consider the set of indices $\Lambda_\delta = \{i \in [n] : \delta_i = 1\}$ and express

$$
S_\delta = \sum_{i\in\Lambda_\delta, j\in\Lambda_\delta^c} A_{ij}X_iX_j = \sum_{j\in\Lambda_\delta^c} X_j\left(\sum_{i\in\Lambda_\delta} A_{ij}X_i\right)
$$

Now we condition on δ and $X_{\Lambda_\delta} := (X_i)_{i\in\Lambda_\delta}$. Then S_δ is a linear combination of subgaussian random variables $X_j \in SG(\sigma^2)$, $j \in \Lambda_\delta^c$. Hence from HW1 Problem 6,

$$S_\delta|\delta, X_{\Lambda_\delta} \in SG\left(\sigma_\delta^2\right), \qquad \text{where} \qquad \sigma_\delta^2 = \sigma^2 \sum_{j\in\Lambda_\delta^c}\left(\sum_{i\in\Lambda_\delta} A_{ij}X_i\right)^2$$

Hence

$$\begin{aligned}\mathbb{E}\left[\exp\left(4\lambda S_\delta\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\exp\left(4\lambda S_\delta\right)|\delta, X_{\Lambda_\delta}\right]\right] \\ &\le \mathbb{E}\left[\exp\left(\frac{(4\lambda)^2}{2}\sigma_\delta^2\right)\right]\end{aligned}$$

Now for fixed δ, let $E_\delta := \mathbb{E}\left[\exp\left(8\lambda^2\sigma_\delta^2\right)|\delta\right]$. It remains to estimate E_δ.

Step 3: reduction to normal random variables.

Consider $g = (g_1, \ldots, g_n)$ where g_i are independent $N(0,1)$ random variables. For each fixed δ and X, we have

$$Z := \sigma \sum_{j\in\Lambda_\delta^c} g_j\left(\sum_{i\in\Lambda_\delta} A_{ij}X_i\right)|\delta, X_{\Lambda_\delta} \sim N\left(0, \sigma_\delta^2\right).$$

Then since the moment generating function of $N(\mu, \sigma^2)$ is $s \mapsto \exp\left(\mu s + \frac{1}{2}\sigma^2 s^2\right)$, hence

$$\mathbb{E}\left[\exp(sZ)|\delta, X_{\Lambda_\delta}\right] = \exp\left(\frac{s^2}{2}\sigma_\delta^2\right),$$

and hence

$$E_\delta = \mathbb{E}\left[\exp\left(4\lambda Z\right)|\delta\right].$$

Now by rearranging the terms, we can write $Z = \sum_{i\in\Lambda_\delta} X_i\left(\sum_{j\in\Lambda_\delta^c} A_{ij}g_j\right)$. Now we condition Z on δ and g. Then Z is a linear combination of subgaussian random variables $X_i \in SG(\sigma^2)$, $i \in \Lambda_\delta$. Hence from HW1 Problem 6,

$$Z|\delta, g \in SG\left(\sigma^4 \sum_{i\in\Lambda_\delta}\left(\sum_{j\in\Lambda_\delta^c} A_{ij}g_j\right)^2\right).$$

Let P_δ denote the coordinate projection of $\mathbb{R}^n$ onto $\mathbb{R}^{\Lambda_\delta}$, and define the matrix $A_\delta = P_\delta A(I - P_\delta)$. Hence we get

$$Z|\delta, g \in SG\left(\sigma^4\left\|A_\delta g\right\|_2^2\right),$$

and we get an upper bound of E_δ as

$$\begin{aligned}E_\delta &= \mathbb{E}\left[\mathbb{E}\left[\exp\left(4\lambda Z\right)|\delta, g\right]|\delta\right] \\ &\le \mathbb{E}\left[\exp\left(\frac{(4\lambda)^2}{2}\sigma^4\left\|A_\delta g\right\|_2^2\right)|\delta\right].\end{aligned}$$

where the original random variable X_i's are removed in the upper bound.

Step 4: calculation for normal random variables.

Now for fixed δ, let $A_\delta = UDV^\top$ be the singular value decomposition of A_δ, and let $s_1, \cdots, s_n$ be the diagonal values of D, i.e. singular values of A_δ. Then

$$\|A_\delta g\|_2^2 = g^\top V D^\top U^\top U D V^\top g = g^\top V D^\top D V^\top g = \left\|DV^\top g\right\|_2^2.$$

And since $Var(V^\top g) = V^\top V = I = Var(g)$, so $V^\top g \stackrel{d}{=} g$, and hence

$$\|A_\delta g\|_2^2 \stackrel{d}{=} \|Dg\|_2^2 = \sum_i s_i^2 g_i^2.$$

Hence by independence of g_i's conditioned on δ, we have

$$\begin{aligned} E_\delta &\le \mathbb{E}\left[\exp\left(8\lambda^2\sigma^4 \|A_\delta g\|_2^2\right) |\delta\right] \\ &= \mathbb{E}\left[\exp\left(8\lambda^2\sigma^4 \sum_i s_i^2 g_i^2\right) |\delta\right] \\ &= \prod_i \mathbb{E}\left[\exp\left(8\lambda^2\sigma^4 s_i^2 g_i^2\right) |\delta\right]. \end{aligned}$$

Then each $g_i^2|\delta$ is the chi-square distribution with one degree of freedom, hence its moment generating function is $\mathbb{E}\left[\exp\left(tg_i^2\right)\right] = (1-2t)^{-\frac{1}{2}}$ for $|t| < \frac{1}{2}$. Therefore

$$E_\delta \le \prod_i \left(1 - 2(8\lambda^2\sigma^4 s_i^2)\right)^{-\frac{1}{2}} \qquad \text{provided} \qquad |\lambda| < \frac{1}{4\sigma^2 \max_i |s_i|}.$$

Using the inequality $(1-z)^{-\frac{1}{2}} \le e^z$ which is valid for all $0 \le z \le \frac{1}{2}$, we can further simplify as

$$E_\delta \le \prod_i \exp\left(16\lambda^2\sigma^4 s_i^2\right) = \exp\left(16\lambda^2\sigma^4 \sum_i s_i^2\right) \qquad \text{provided} \qquad |\lambda| < \frac{1}{4\sigma^2 \max_i |s_i|}.$$

Then since $\max_i |s_i| = \|A_\delta\| \le \|A\|$ and $\sum_i s_i^2 = \|A_\delta\|_{HS}^2 \le \|A\|_{HS}^2$, we can further bound as

$$E_\delta \le \exp\left(16\lambda^2\sigma^4 \|A\|_{HS}^2\right) \qquad \text{for} \qquad |\lambda| < \frac{1}{4\sigma^2 \|A\|}.$$

Since this is a uniform bound for all δ, we have

$$\begin{aligned} \mathbb{E}\left[\exp\left(\lambda S\right)\right] &\le \mathbb{E}\left[E_\delta\right] \\ &\le \exp\left(16\lambda^2\sigma^4 \|A\|_{HS}^2\right) \qquad \text{for} \qquad |\lambda| < \frac{1}{4\sigma^2 \|A\|}, \end{aligned}$$

i.e. $S \in SE\left(32\sigma^4 \|A\|_{HS}^2, 4\sigma^2 \|A\|\right)$. Hence by applying sub-exponential tail bound,

$$\begin{aligned} \mathbb{P}\left(S \ge t\right) &\le \begin{cases} \exp\left(-\frac{t^2}{64\sigma^4 \|A\|_{HS}^2}\right) & 0 \le t \le 8\sigma^2 \frac{\|A\|_{HS}^2}{\|A\|}, \\ \exp\left(-\frac{t}{8\sigma^2 \|A\|}\right) & t \ge 8\sigma^2 \frac{\|A\|_{HS}^2}{\|A\|}, \end{cases} \\ &= \exp\left(-\min\left\{\frac{t^2}{64\sigma^4 \|A\|_{HS}^2}, \frac{t}{8\sigma^2 \|A\|}\right\}\right). \end{aligned} \tag{7}$$

Step 5: conclusion.

By putting (6) and (7) together, we have

$$
\begin{aligned}
&\mathbb{P}\left(X^\top AX - \mathbb{E}\left[X^\top AX\right] \geq t\right) \\
&= \mathbb{P}\left(\sum_i A_{ii}\left(X_i^2 - \mathbb{E}\left[X_i^2\right]\right) \geq \frac{2}{3}t\right) + \mathbb{P}\left(\sum_{i\neq j} A_{ij}X_iX_j \geq \frac{1}{3}t\right) \\
&\leq \exp\left(-\min\left\{\frac{t^2}{1152\sigma^4\left\|A\right\|_{HS}^2}, \frac{t}{48\sigma^2\left\|A\right\|}\right\}\right) + \exp\left(-\min\left\{\frac{t^2}{576\sigma^4\left\|A\right\|_{HS}^2}, \frac{t}{24\sigma^2\left\|A\right\|}\right\}\right) \\
&\leq 2\exp\left(-\min\left\{\frac{t^2}{1152\sigma^4\left\|A\right\|_{HS}^2}, \frac{t}{48\sigma^2\left\|A\right\|}\right\}\right).
\end{aligned}
$$

By plugging in $-A$ in place of A, we have the other tail inequality as well. Hence, we get the two-sided upper bound as

$$
\begin{aligned}
&\mathbb{P}\left(\left|X^\top AX - \mathbb{E}\left[X^\top AX\right]\right| \geq t\right) \\
&\leq 4\exp\left(-\min\left\{\frac{t^2}{1152\sigma^4\left\|A\right\|_{HS}^2}, \frac{t}{48\sigma^2\left\|A\right\|}\right\}\right).
\end{aligned}
$$

5. **Robust statistics and the median-of-mean estimator**. Suppose we observe n i.i.d. random variables with distribution P and would like to construct a $1-\alpha$ confidence set for the expected value of P, where $\alpha \in (0,1)$.

   (a) If the common distribution P is in the class $SG(\sigma^2)$ provide such a confidence interval.

   (b) Now let's drop the assumption that P is an $SG(\sigma^2)$ distribution and in particular allow for very thick tails.

   How can we proceed?

   Here is a simple method. Assume that $\mathrm{Var}[X] = \sigma^2 < \infty$. For a fixed $\alpha \in [e^{1-n/2}, 1)$, set $b = \lceil \ln(1/\alpha) \rceil$ and note that $b \leq n/2$. Next, partition $[n] = \{1, \ldots, n\}$ into b blocks $B_1, \ldots, B_b$ each of size $|B_i| \geq \lfloor n/b \rfloor \geq 2$ and compute the sample mean in each block:

$$\overline{X}_i = \frac{1}{|B_i|}\sum_{j\in B_i} X_j, \quad i = 1, \ldots, b.$$

   Finally define **the median-of-means** estimator as

$$\hat{\mu} = \hat{\mu}(\alpha) = \text{median}\left\{\overline{X}_1, \ldots, \overline{X}_b\right\},$$

   where, for any b-tuple of numbers $(x_1, \ldots, x_b)$,

$$\text{median}\left\{x_1, \ldots, x_b\right\} = x_{j^*},$$

   with

$$|\{k \in [b] \colon x_k \leq x_{j^*}\}| \geq b/2 \quad \text{and} \quad |\{k \in [b] \colon x_k \geq x_{j^*}\}| \geq b/2,$$

   (if more than one such x_{j^*} satisfies the above inequalities, pick one of them at random).

Show that the median-of-means estimator yields, up to constants, the same type of sub-Gaussian confidence interval obtained in the first part, but without requiring the assumption of sub-Gaussianity. That is, show that

$$\mathbb{P}\left(|\hat{\mu} - \mu| \geq C\sqrt{\frac{\sigma^2 \log(1/\alpha)}{n}}\right) \leq \alpha,$$

for some constant C, where $\sigma^2 = \mathrm{Var}[X]$. You may want to consult these paper:

- M. Lerasle and R. I. Oliveira (2011). Robust empirical mean estimators. https://arxiv.org/pdf/1112.3914v1.pdf
- Luc Devroye, Matthieu Lerasle, Gabor Lugosi and Roberto I. Oliveira (2016). Sub-Gaussian mean estimators. https://arxiv.org/pdf/1509.05845v1.pdf

(c) The median-of-means estimator has an obvious drawback. What is it? *Hint: think of the situation when you want to use this estimator to compute confidence intervals at different levels α and α'...*

Points: 14 pts = 5 + 7 + 2.

Solution.

(a)

Let $X_1, \dots, X_n$ be the i.i.d. random variables with $\mathbb{E}[X_i] = \mu$.
If $X_i \in SG(\sigma^2)$, then from two-sided Hoeffding's inequality,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}(X_i - \mu)\right| > nt\right) \leq 2\exp\left(-\frac{(nt)^2}{2n\sigma^2}\right).$$

So by letting $\bar{X} := \frac{1}{n}\sum\limits_{i=1}^{n} X_i$, then $\left|\sum\limits_{i=1}^{n}(X_i - \mu)\right| \leq nt \iff \mu \notin \left[\bar{X} - t, \bar{X} + t\right]$, hence

$$\mathbb{P}\left(\mu \in \left[\bar{X} - t, \bar{X} + t\right]\right) \geq 1 - 2\exp\left(-\frac{nt^2}{2\sigma^2}\right).$$

And $1 - \alpha = 1 - 2\exp\left(-\frac{nt^2}{2\sigma^2}\right) \iff t = \sqrt{\frac{2\sigma^2\log(2/\alpha)}{n}}$. Hence the confidence interval can be computed as

$$\mathbb{P}\left(\mu \in \left[\bar{X} - \sqrt{\frac{2\sigma^2\log(2/\alpha)}{n}}, \bar{X} + \sqrt{\frac{2\sigma^2\log(2/\alpha)}{n}}\right]\right) \geq 1 - \alpha,$$

i.e. $\left[\bar{X} - \sqrt{\frac{2\sigma^2}{n}\log\left(\frac{2}{\alpha}\right)}, \bar{X} + \sqrt{\frac{2\sigma^2}{n}\log\left(\frac{2}{\alpha}\right)}\right]$ is $1-\alpha$ confidence interval for i.i.d. $SG(\sigma^2)$ random variables.

(b)

Fix $x > 0$, and for $i = 1, \dots, b$, let $Y_i := I\left(\left|\bar{X}_i - \mu\right| \geq x\right)$. Note that $|\hat{\mu} - \mu| \geq x$ implies that $\left|\left\{i \in [b] \colon \left|\bar{X}_i - \mu\right| \geq x\right\}\right| \geq \frac{b}{2}$, which is equivalent to $\sum\limits_{i=1}^{b} I\left(\left|\bar{X}_i - \mu\right| \geq x\right) \geq \frac{b}{2}$. Hence

$$\mathbb{P}\left(|\hat{\mu} - \mu| \geq x\right) \leq \mathbb{P}\left(\sum_{i=1}^{b} Y_i \geq \frac{b}{2}\right).$$

Let $p_i := \mathbb{P}\left(\left|\bar{X}_i - \mu\right| \geq x\right)$ and $\bar{p} := \frac{1}{b}\sum_{i=1}^{b} p_i$. Note that since $Var[X_j] = \sigma^2$ and each X_j's are independent, $\left\lfloor \frac{n}{\lceil b \rceil} \right\rfloor \geq \frac{n}{b}$

$$Var[\bar{X}_i] = Var\left[\frac{1}{|B_i|}\sum_{j\in B_i} X_j\right] = \frac{1}{|B_i|}Var[X_j] \leq \frac{\sigma^2}{\lfloor n/b \rfloor}$$

Hence by using Chebyshev inequality,

$$p_i = \mathbb{P}\left(\left|\bar{X}_i - \mu\right| \geq x\right) \leq \frac{Var[\bar{X}_i]}{x^2} \leq \frac{\sigma^2}{x^2\lfloor n/b \rfloor}.$$

Let $x > \sqrt{\frac{\sigma^2}{4\lfloor n/b \rfloor}}$. Then there exists $W \sim Binomial\left(b, \frac{\sigma^2}{x^2\lfloor n/b \rfloor}\right)$ such that $\sum_{i=1}^{b} Y_i \leq W$ a.s.. Hence

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{b} Y_i \geq \frac{b}{2}\right) &\leq \mathbb{P}\left(W \geq \frac{b}{2}\right) \\
&= \sum_{j=\lceil \frac{b}{2} \rceil}^{b} \binom{b}{j} \left(\frac{\sigma^2}{x^2\lfloor n/b \rfloor}\right)^j \left(1 - \frac{\sigma^2}{x^2\lfloor n/b \rfloor}\right)^{b-j} \\
&\leq \left(\frac{\sigma^2}{x^2\lfloor n/b \rfloor}\right)^{\lceil \frac{b}{2} \rceil} \sum_{j=0}^{b} \binom{b}{j} \\
&\leq \left(\frac{4\sigma^2}{x^2\lfloor n/b \rfloor}\right)^{\frac{b}{2}},
\end{aligned}$$

and hence

$$\mathbb{P}\left(|\hat{\mu} - \mu| \geq x\right) \leq \left(\frac{4\sigma^2}{x^2\lfloor n/b \rfloor}\right)^{\frac{b}{2}}.$$

Then $\left(\frac{4\sigma^2}{x^2\lfloor n/b \rfloor}\right)^{\frac{b}{2}} \leq \alpha$ if and only if $x \geq \sqrt{\frac{4\sigma^2}{\lfloor n/b \rfloor \alpha^{\frac{2}{b}}}}$. Now $\frac{1}{\alpha^{\frac{2}{b}}} \leq e^2$, so

$$\mathbb{P}\left(|\hat{\mu} - \mu| \geq \sqrt{\frac{4e^2\sigma^2}{\lfloor n/b \rfloor}}\right) \leq \alpha.$$

(c)
In (a), the estimator $\bar{X}$ does not depend on α, and once the estimator is computed, confidence interval $\left[\bar{X} - \sqrt{\frac{2\sigma^2\log(2/\alpha)}{n}}, \bar{X} + \sqrt{\frac{2\sigma^2\log(2/\alpha)}{n}}\right]$ for different levels α and α' can be computed easily. In (c), the $b = \lceil \ln(1/\alpha) \rceil$ blocks $B_1, \ldots, B_b$ depends on α, and hence the estimator $\hat{\mu}(\alpha) = \text{median}\left\{\overline{X}_1, \ldots, \overline{X}_b\right\}$ also depends on α. Hence the estimator $\hat{\mu}(\alpha)$ has to be computed for different levels α and α' for the confidence interval $\left[\hat{\mu} - \sqrt{\frac{4e^2\sigma^2}{\lfloor n/b \rfloor}}, \hat{\mu} + \sqrt{\frac{4e^2\sigma^2}{\lfloor n/b \rfloor}}\right]$, and hence it is computationally heavier. This also makes difference in streamed data case: for (a), once $\bar{X}$ is computed, the data set $\{X_1, \cdots, X_n\}$ need not be maintained anymore to produce confidence intervals for different levels α and α', while in (b), the data set $\{X_1, \cdots, X_n\}$ should be maintained to compute confidence intervals for different levels α and α'.

6. Let $(\mathcal{X}, d)$ be a metric space and, for $\delta > 0$, let $N(\delta)$ and $M(\delta)$ denote the δ-covering and δ-packing number, respectively. Show that

$$M(2\delta) \leq N(\delta) \leq M(\delta).$$

Points: 12 pts.

Solution.
$(M(2\delta) \leq N(\delta))$
Let $\mathcal{X}$ be covered by $N(\delta)$ balls $\left\{B(x_1, \delta), \cdots, B(x_{N(\delta)}, \delta)\right\}$, and let $\{y_1, \cdots, y_k\}$ be a 2δ-packing of $\mathcal{F}$. For any $1 \leq i \neq j \leq k$, since $\|y_i - y_j\| \geq 2\delta$, y_i and y_j cannot be in the same ball $\mathcal{B}(x_l, \delta)$. Hence there is a one-to-one function from $\{y_1, \cdots, y_k\}$ to $\left\{B(x_1, \delta), \cdots, B(x_{N(\delta)}, \delta)\right\}$, so

$$k = |\{y_1, \cdots, y_k\}| \leq \left|\left\{B(x_1, \delta), \cdots, B(x_{N(\delta)}, \delta)\right\}\right| = N(\delta).$$

Since this holds for any 2δ-packing, $M(2\delta) \leq N(\delta)$ holds.
$(N(\delta) \leq M(\delta))$
Let $\left\{y_1, \cdots, y_{M(\delta)}\right\}$ be a δ-packing of $\mathcal{X}$. Suppose $\left\{B(y_1, \delta), \cdots, B(y_{M(\delta)}, \delta)\right\}$ don't cover $\mathcal{X}$, then there exists $x \in \mathcal{X}$ with $x \notin \bigcup_{i=1}^{M(\delta)} B(y_i, \delta)$. This implies $\|x - y_i\| > \delta$ for all $i = 1, \cdots, M(\delta)$. (note that $B(y_i, \delta)$ is a closed ball in Lecture note 6(Sep 18)). Then $\left\{x, y_1, \cdots, y_{M(\delta)}\right\}$ is also a δ-packing of $\mathcal{X}$ with size $M(\delta) + 1 > M(\delta)$, which is a contradiction to the definition of $M(\delta)$. Hence $\left\{y_1, \cdots, y_{M(\delta)}\right\}$ is also a δ-covering of $\mathcal{X}$. Since $N(\delta)$ is the size of the smallest δ-covering,

$$N(\delta) \leq M(\delta).$$

7. **Efron-Stein inequality** In this exercise you will derive a nice result, known as the Efron-Stein inequality, that yields useful bounds for the variance of functions of independent variables.

   Let $X_1, \ldots, X_n$ be independent random variables and $Z = f(X_1, \ldots, X_n)$. We make no assumptions on the function $f : \mathbb{R}^n \to \mathbb{R}$ other than $\mathbb{E}[Z^2] < \infty$. For any $i = 1, \ldots, n$, let

$$\mathbb{E}_i[Z] = \mathbb{E}[Z | X_j, j \neq i].$$

   (a) Prove the Efron-Stein Inequality:

$$\mathbb{V}[Z] \leq \sum_{i=1}^{n} \mathbb{E}\left[(Z - \mathbb{E}_i[Z])^2\right]$$

   Hints: by Doob's representation, $Z - \mathbb{E}[Z] = \sum_{i=1}^{n} Y_i$, *for a martingale difference sequence* $(Y_1, \ldots, Y_n)$. *Then show that* $\mathbb{V}[Z] = \sum_{i=1}^{n} \mathbb{E}\left[Y_i^2\right]$. *Finally, show that, for all* i,

$$Y_i^2 \leq \mathbb{E}\left[(Z - \mathbb{E}_i[Z])^2 | X_1, \ldots, X_i\right].$$

   (b) For any $i = 1, \ldots, n$, let $Z_i = f_i(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, for an arbitrary function $g_i : \mathbb{R}^{n-1} \to \mathbb{R}$ of the $(n-1)$ variables X_j, $j \neq i$. Use the Efron-Stein inequality to show that

$$\mathbb{V}[Z] \leq \sum_{i=1}^{n} \mathbb{E}\left[(Z - Z_i)^2\right].$$

   Hint: use conditionally the fact that, for any random variable X, $\mathbb{V}[X] \leq \mathbb{E}\left[(X - c)^2\right]$, *for any* $c \in \mathbb{R}$.

(c) Use the previous result to show that, if f satisfies the bounded difference property with constants $(c_1, \ldots, c_n)$, then

$$\mathbb{V}[Z] \leq \frac{1}{4} \sum_{i=1}^{n} c_i^2.$$

(d) **Application to Kernel density estimation.** Let p be a Lebesgue density over the real line. Let $X_1, \ldots, X_n$ be an i.i.d. sample from the distribution P with density p. Let K be a nonnegative function with $\int_{\mathbb{R}} K(t)dt = 1$ (a kernel). For a $h > 0$ (the bandwidth parameter) define the random function $\hat{p}_h$ given by

$$x \in \mathbb{R} \mapsto \hat{p}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

This is the kernel density estimator of p with bandwidth h. Let

$$Z = \int_{\mathbb{R}} |\hat{p}_h(x) - p(x)| dx,$$

be the L_1 distance between p and $\hat{p}_h$. Show that

$$\mathbb{V}[Z] \leq \frac{1}{n}.$$

Points: 23 pts = 7 + 4 + 6 + 6.

Solution.

(a)

For $i = 1, \ldots, n$, define Y_i as $Y_i = \mathbb{E}[Z|X_1, \ldots, X_i] - \mathbb{E}[Z|X_1, \ldots, X_{i-1}]$. Then $Z - \mathbb{E}[Z]$ can be factorized as

$$Z - \mathbb{E}[Z] = \sum_{i=1}^{n} Y_i.$$

Also, for $i < j$, using tower property and that Y_i is a function of $X_1, \ldots, X_i$,

$$\begin{aligned}
\mathbb{E}\left[Y_i Y_j | X_1, \ldots, X_i\right] &= Y_i \mathbb{E}\left[\mathbb{E}[Z|X_1, \ldots, X_j] - \mathbb{E}[Z|X_1, \ldots, X_{j-1}] | X_1, \ldots, X_i\right] \\
&= Y_i \left(\mathbb{E}[Z|X_1, \ldots, X_i] - \mathbb{E}[Z|X_1, \ldots, X_i]\right) = 0.
\end{aligned}$$

Hence $\mathbb{V}[Z]$ can be factorized as

$$\begin{aligned}
\mathbb{V}[Z] = \mathbb{E}\left[(Z - \mathbb{E}[Z])^2\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[Y_i^2\right] + 2 \sum_{i<j} \mathbb{E}\left[Y_i Y_j\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[Y_i^2\right] + 2 \sum_{i<j} \mathbb{E}\left[Y_i Y_j | X_1, \ldots, X_i\right] = \sum_{i=1}^{n} \mathbb{E}\left[Y_i^2\right].
\end{aligned}$$

Now, note that since $X_1, \ldots, X_n$ are independent,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}_i[Z] | X_1, \ldots, X_i\right] &= \mathbb{E}\left[\mathbb{E}\left[Z | X_j, j \neq i\right] | X_1, \ldots, X_i\right] = \mathbb{E}\left[\mathbb{E}\left[Z | X_j, j \neq i\right] | X_1, \ldots, X_{i-1}\right] \\
&= \mathbb{E}\left[Z | X_1, \ldots, X_{i-1}\right].
\end{aligned}$$

And hence by applying Jensen's inequality,

$$\begin{aligned} Y_i^2 &= (\mathbb{E}[Z|X_1,\dots,X_i] - \mathbb{E}[Z|X_1,\dots,X_{i-1}])^2 = \left(\mathbb{E}\left[Z - \mathbb{E}_i[Z]\right| X_1,\dots,X_i]\right)^2 \\ &\leq \mathbb{E}\left[(Z-\mathbb{E}_i[Z])^2 \left| X_1,\dots,X_i\right.\right]. \end{aligned}$$

Hence applying this gives

$$\begin{aligned} \mathbb{V}[Z] &= \sum_{i=1}^n \mathbb{E}\left[Y_i^2\right] \leq \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left[(Z-\mathbb{E}_i[Z])^2 \left| X_1,\dots,X_i\right.\right]\right] \\ &= \sum_{i=1}^n \mathbb{E}\left[(Z-\mathbb{E}_i[Z])^2\right]. \end{aligned}$$

(b)
For each i, note that Z_i is a function of the $(n-1)$ variables X_j, $j \neq i$, and hence

$$\mathbb{E}_i\left[(Z-\mathbb{E}_i[Z])^2\right] \leq \mathbb{E}_i\left[(Z-Z_i)^2\right].$$

Hence applying this to (a) yields

$$\begin{aligned} \mathbb{V}[Z] &\leq \sum_{i=1}^n \mathbb{E}\left[(Z-\mathbb{E}_i[Z])^2\right] = \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}_i\left[(Z-\mathbb{E}_i[Z])^2\right]\right] \\ &\leq \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}_i\left[(Z-Z_i)^2\right]\right] = \sum_{i=1}^n \mathbb{E}\left[(Z-Z_i)^2\right]. \end{aligned}$$

(c)
For each i, note that conditioned on X_j, $j \neq i$, Z is a bounded random variable with satisfying $a \leq Z \leq b$ with $b - a \leq c_i$. Hence from Popoviciu's inequality below,

$$\mathbb{E}_i\left[(Z-\mathbb{E}_i[Z])^2\right] \leq \frac{1}{4}c_i^2.$$

Hence applying this to (a) yields

$$\begin{aligned} \mathbb{V}[Z] &\leq \sum_{i=1}^n \mathbb{E}\left[(Z-\mathbb{E}_i[Z])^2\right] = \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}_i\left[(Z-\mathbb{E}_i[Z])^2\right]\right] \\ &\leq \frac{1}{4}\sum_{i=1}^n c_i^2. \end{aligned}$$

Lemma. (Popoviciu's inequality) If a random variable X satisfies $a \leq X \leq b$ for some $a, b \in \mathbb{R}$, then

$$\mathbb{V}[X] \leq \frac{1}{4}(b-a)^2.$$

Proof of Lemma.
Note that $\mathbb{V}[X] \leq \mathbb{E}\left[\left(X - \frac{a+b}{2}\right)^2\right]$. Then from $a \leq X \leq b$, $\left|X - \frac{a+b}{2}\right| \leq \frac{b-a}{2}$ holds, and hence

$$\mathbb{V}[X] \leq \mathbb{E}\left[\left(X - \frac{a+b}{2}\right)^2\right] \leq \frac{1}{4}(b-a)^2.$$

(d)
Note first that $Z \leq \int_{\mathbb{R}} \hat{p}_h(x)dx + \int_{\mathbb{R}} p(x)dx = 2$, and hence $\mathbb{E}[Z^2] < \infty$ is finite. Let $f : \mathbb{R}^n \to \mathbb{R}$ be

$$f(x_1, \ldots, x_n) = \int_{\mathbb{R}} \left| \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) - p(x) \right| dx,$$

then $Z = f(X_1, \ldots, X_n)$ holds. Then, note that for any $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n \in \mathbb{R}$ and $y, y' \in \mathbb{R}$, from $\left| \int g(x)dx - \int h(x)dx \right| \leq \int |g(x) - h(x)|dx$,

$$\begin{aligned}
&\left| f(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, y', x_{i+1}, \ldots, x_n) \right| \\
&\leq \int_{\mathbb{R}} \left| \frac{1}{nh} K\left(\frac{x - y}{h}\right) - \frac{1}{nh} K\left(\frac{x - y'}{h}\right) \right| dx \\
&\leq \int_{\mathbb{R}} \frac{1}{nh} K\left(\frac{x - y}{h}\right) dx + \int_{\mathbb{R}} \frac{1}{nh} K\left(\frac{x - y'}{h}\right) dx \\
&= \frac{2}{n}.
\end{aligned}$$

Hence applying (c) gives

$$\mathbb{V}[Z] \leq \frac{1}{4} \sum_{i=1}^{n} \left(\frac{2}{n}\right)^2 = \frac{1}{n}.$$