Points for problems: 1, 1, 0.5, 1, 1
Total: 4.5

Vo Ngoc Bich Uyen

Home assignment 2

High-dimensional probability & statistics

Problem 1: Let $X \in \mathbb{R}^n$ be a random vector with i.i.d. Gaussian components $X_i \sim N(0,1)$ for all $i \in \{1, \dots, n\}$. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric positive semidefinite matrix and denote $Z = X^T A X$. Prove the Hanson-Wright inequality:

$$\mathbb{P}(Z \geq \mathrm{Tr}(A) + t) \leq \exp\left\{ -c \cdot \min\left\{ \frac{t}{\|A\|}, \frac{t^2}{\|A\|_F^2} \right\} \right\} \text{ where } c > 0 \text{ - an absolute constant.}$$

→ Proof: + $\mathrm{Tr}(A) = \mathbb{E} X^T A X = \sum_i a_{ii} \mathbb{E} X_i^2 \quad \left(= \sum_{i=1}^n a_{ii}\right)$

+ $Z - \mathrm{Tr}(A) = X^T A X - \mathbb{E} X^T A X = \sum_{i,j}^n a_{ij} X_i X_j - \sum_i^n a_{ii} \mathbb{E} X_i^2$

$$= \sum_i^n a_{ii}(X_i^2 - \mathbb{E} X_i^2) + \sum_{\substack{i \neq j \\ i,j}}^n a_{ij} X_i X_j$$

$$\Rightarrow \mathbb{P}(Z \geq \mathrm{Tr}(A) + t) \leq \mathbb{P}\left(\sum_i^n a_{ii}(X_i^2 - \mathbb{E} X_i^2) \geq \frac{t}{2} \right) + \mathbb{P}\left(\sum_{\substack{i,j \\ i \neq j}}^n a_{ij} X_i X_j \geq \frac{t}{2} \right) = P_1 + P_2 \quad (1)$$

+ We have X_i are independent, sub-Gaussian random variables, so $X_i^2 - \mathbb{E}X_i^2$ are independent, mean zero, sub-exponential random variables.

+ Applying Bernstein's inequality for $X_i^2 - \mathbb{E}X_i^2$, ~~$a = (a_1, \dots, a_n) \in \mathbb{R}^n$~~ we receive:

$$P_1 = \mathbb{P}\left(\sum_i^n a_{ii}(X_i^2 - \mathbb{E}X_i^2) \geq \frac{t}{2} \right) \leq \exp\left\{ -c_1 \min\left(\frac{t^2}{K^2 \|a\|_2^2}, \frac{t}{K \|a\|_\infty} \right) \right\}$$

where $\cancel{K = \max_i \|X_i\|_{\psi_1}}$ $K = \max_i \|X_i^2 - \mathbb{E}X_i^2\|_{\psi_1}$

+ As we know, sub-exponential is sub-gaussian squared; i.e. $\|X_i^2\|_{\psi_1} = \|X_i\|_{\psi_2}^2$

Following the centering inequality, we have $\|X_i^2 - \mathbb{E}X_i^2\|_{\psi_1} \leq \|X_i^2\|_{\psi_1}$

$$\rightarrow \|X_i^2 - \mathbb{E}X_i^2\|_{\psi_1} \leq \|X_i^2\|_{\psi_1} \leq \|X_i\|_{\psi_2}^2 \leq 1 \quad ; \text{ i.e. } K = 1$$

$$\Rightarrow P_1 \leq \exp\left\{ -c_1 \min\left\{ \frac{t}{\max_i |a_{ii}|}, \frac{t^2}{\sum_i^n a_{ii}^2} \right\} \right\} \leq \exp\left\{ -c_1 \min\left(\frac{t}{\|A\|}, \frac{t^2}{\|A\|_F^2} \right) \right. \quad (2)$$

+ Let $M = \sum_{\substack{i,j \\ i \neq j}}^n a_{ij} X_i X_j$

Applying Markov's inequality, we have:

$$P_2 = \mathbb{P}\left(M \geq \frac{t}{2}\right) \underset{(\lambda > 0)}{=} \mathbb{P}\left(\lambda M \geq \lambda \frac{t}{2}\right) \leq \exp\left(-\frac{\lambda t}{2}\right) \mathbb{E} e^{\lambda M}$$

+ Using decoupling inequality for function $F = \exp(x)$, we have $\mathbb{E}\exp\left(\sum_{\substack{i,j \\ i \neq j}}^n a_{ij} X_i X_j \right) \leq \mathbb{E}\exp\left(4 \sum_{i,j}^n a_{ij} X_i X_j' \right)$

where X' is independent copy of X.

$$\Rightarrow \mathbb{E}e^{\lambda M} \le \mathbb{E}e^{4\lambda X^T A X'} \le \left[\begin{array}{l}\text{Comparison lemma} \\ \mathbb{E}e^{\lambda X^T A X'} \le \mathbb{E}e^{C\kappa^2 \lambda g^T A g'} \\ \text{where } g, g' \overset{i.i.d}{\sim} N(0, I_n)\end{array}\right] \le \mathbb{E}e^{C_0 \lambda g^T A g'} \le \left[\begin{array}{l}\text{MGF of Gaussian chaos} \\ \mathbb{E}e^{\lambda X^T A X'} \le e^{c\lambda^2 \|A\|_F^2}\end{array}\right]$$

$$\le e^{C_1 \lambda^2 \|A\|_F^2} \quad \text{provided that } |\lambda| \le \frac{C_1}{\|A\|}$$

$$\Rightarrow p_2 \le e^{-\frac{\lambda t}{2}} e^{C_1 \lambda^2 \|A\|_F^2} = \exp\left(-\frac{\lambda t}{2} + C_1 \lambda^2 \|A\|_F^2\right)$$

+) Optimizing over $0 \le \lambda \le \frac{C_1}{\|A\|}$, we have $p_2 \le \exp\left\{-C_1 \min\left(\frac{t}{\|A\|}, \frac{t^2}{\|A\|_F^2}\right)\right\}$ (3)

From (1), (2) and (3)

$$\Rightarrow P\left(Z \ge Tr(A) + t\right) \le 2\exp\left\{-C_1 \cdot \min\left(\frac{t}{\|A\|}, \frac{t^2}{\|A\|_F^2}\right)\right\} = \exp\left\{-C \cdot \min\left(\frac{t}{\|A\|}, \frac{t^2}{\|A\|_F^2}\right)\right\} \square$$

Problem 2: Let $p \ge 1$ and let $X_1, \dots, X_n$ be r.v. with finite ψ_p-norms. Show that there exists an absolute constant $C > 0$, such that $\|\max_{1\le i\le n} X_i\|_{\psi_p} \le C \cdot \max_{1\le i\le n} \|X_i\|_{\psi_p} (\log n)^{1/p}$

$\longrightarrow$ Proof: +) We have $\psi_p(x) = e^{x^p} - 1 \Rightarrow \psi_p^{-1}(n) = [\log(n+1)]^{1/p}$

+) For some constant c, we have $\displaystyle\lim_{a,b\to\infty} \frac{\psi_p(a)\psi_p(b)}{\psi_p(cab)} < \infty$

or $\psi_p(a) \cdot \psi_p(b) \le \psi_p(cab) \ \forall a, b \ge 1$; i.e. $\psi_p\left(\frac{a}{b}\right) \le \frac{\psi_p(ca)}{\psi_p(b)} \ \forall, a \ge b \ge 1$ (1)

+) Applying (1) for $a = \frac{|X_i|}{C_1}$; $b = y \ge 1$; $C_1 > 0$, we get

$$\max \psi_p\left(\frac{|X_i|}{C_1 y}\right) \le \max\left[\frac{\psi_p\left(\frac{c|X_i|}{C_1}\right)}{\psi_p(y)} 1\left\{\frac{|X_i|}{C_1 y} \ge 1\right\} + \psi_p\left(\frac{|X_i|}{C_1 y}\right) \cdot 1\left\{\frac{|X_i|}{C_1 y} < 1\right\}\right]$$

$$\le \max\left(\frac{\psi_p\left(\frac{c|X_i|}{C_1}\right)}{\psi_p(y)}\right) + \psi_p(1) \le \sum_{i=1}^n \frac{\psi_p\left(\frac{c|X_i|}{C_1}\right)}{\psi_p(y)} + \psi_p(1)$$

Set $C_1 = c \cdot \max \|X_i\|_{\psi_p}$

$$\mathbb{E}\psi_p\left(\frac{\max|X_i|}{C_1 y}\right) \le \mathbb{E}\max\psi_p\left(\frac{|X_i|}{C_1 y}\right) \le \sum_{i=1}^n \mathbb{E}\left[\frac{\psi_p\left(\frac{|X_i|}{\max\|X_i\|_{\psi_p}}\right)}{\psi_p(y)}\right] + \psi_p(1) \le \frac{n}{\psi_p(y)} + \psi_p(1) \overset{(2)}{\le} 1$$

$$(2) \Leftrightarrow \frac{n}{\psi_p(y)} \le 1 - \psi_p(1) \Leftrightarrow \frac{n}{1 - \psi_p(1)} \le \psi_p(y) \Leftrightarrow 2n \le \frac{n}{1 - \psi_p(1)} \le \psi_p(y)$$

$$[\psi_p(1) \ge \tfrac{1}{2}]$$

$$\Leftrightarrow y \ge \psi_p^{-1}(2n) \Leftrightarrow C_1 y \ge c \cdot \max_{1\le i\le n} \|X_i\|_{\psi_p} \cdot \psi_p^{-1}(2n) \quad (\text{because } C_1 = c \cdot \max\|X_i\|_{\psi_p})$$

$$\Rightarrow \|\max_{1\le i\le n} X_i\|_{\psi_p} \le 2c \cdot \psi_p^{-1}(n) \cdot \max_{1\le i\le n} \|X_i\|_{\psi_p}$$

Hence $\|\max_{1\le i\le n} X_i\|_{\psi_p} \le C \cdot \max_{1\le i\le n} \|X_i\|_{\psi_p} \cdot (\log n)^{1/p}$ $\square$

Problem 3: Let $f: \mathbb{R}^2 \to [-b, b]$ be a symmetric function of its arguments. Let $\varepsilon_1, \dots, \varepsilon_n$ be independent random variables. Denotes $U = \frac{2}{n(n-1)} \sum_{i<j} f(\varepsilon_i, \varepsilon_j)$.

Prove that $P(U - EU \geq t) \leq e^{-\frac{nt^2}{32b^2}} \quad \forall t \geq 0$

→ Proof: + Let $X_{ij} = f(\varepsilon_i, \varepsilon_j)$ for $i<j$

+ Denote $Y_{ij} = X_{ij} - E[X_{ij}]$; $Y = \sum_{i<j} Y_{ij}$

+ We have $U = \frac{2}{n(n-1)} \sum_{i<j} X_{ij} = \frac{2}{n(n-1)} \left(\sum_{i<j} Y_{ij} + \sum E[X_{ij}] \right) = \frac{2}{n(n-1)} Y + EU$

$$\Rightarrow U - EU = \frac{2}{n(n-1)} \cdot Y$$

+ Applying the Hoeffding inequality to each Y_{ij}, we have:

$$P(Y_{ij} \geq t) \leq e^{-\frac{2t^2}{(b-(-b))^2}} = e^{-\frac{t^2}{2b^2}}$$

+ Hence, $\forall t > 0$

$$P(U - EU \geq t) = P\left(\frac{2}{n(n-1)} Y \geq t \right) = P\left(Y \geq \frac{n(n-1)}{2} \cdot t \right) \leq e^{-\frac{n^2(n-1)^2 t^2}{8b^2}}$$

+ To obtain the desired ~~er~~ inequality $P(U - EU \geq t) \leq e^{-\frac{nt^2}{32b^2}}$, we have to prove that $-\frac{n^2(n-1)^2 t^2}{8b^2} \leq -\frac{nt^2}{32b^2} \quad \forall t > 0, n > 1$ (for U to be defined)

$(\Leftrightarrow) \quad 4n^2(n-1)^2 \geq n$

~~$\Rightarrow 4n(n-1)^2 \geq 1$ is always true for all $n \geq 1$~~

$\Rightarrow \quad 4n(n-1)^2 \geq 1 \qquad (*)$

Y_{ij}'s are not independent, you cannot apply Hoeffding's inequality in this case

+ We can use induction to prove $(*)$

- $n = 2$: $\quad 4n(n-1)^2 = 8 \geq 1$
- Assume that $n = k$ is true, we have $4k(k-1)^2 \geq 1 \quad \forall k > 2$
- We have to prove that $n = k+1$ is true; i.e. $4(k+1) \cdot k^2 \geq 1$

Indeed, $\begin{cases} 4(k+1)k^2 = 4k(k-1)^2 + 12k(k-1) + 8k \\ 4k(k-1)^2 \geq 1 \\ 12k(k-1) + 8k \geq 0 \quad \forall k > 2 \end{cases}$

Therefore, we can conclude that $4n(n-1)^2 \geq 1 \quad \forall n > 1$

and

$$P(U - EU \geq t) \leq e^{-\frac{n^2(n-1)^2 t^2}{8b^2}} \leq e^{-\frac{nt^2}{32b^2}} \qquad \square$$

Problem 4: Prove the following variational representation:

$$\mathrm{Ent}(e^{\lambda f(X)}) = \sup_{g : Ee^g \leq 1} E[g(X) \cdot e^{\lambda f(X)}]$$

where the supremum is taken with respect to all measurable functions g.

→ Proof: +) Case 1: Choose $g: \mathbb{E}e^g = 1$

· We have $Ent(e^{\lambda f(x)}) - \mathbb{E}[g.e^{\lambda f(x)}] = \mathbb{E}[e^{\lambda f(x)}.\log(e^{\lambda f(x)})] - \mathbb{E}[e^{\lambda f(x)}.\log e^{g(x)}] -$

$- \mathbb{E}[e^{\lambda f(x)}].\log \mathbb{E}[e^{\lambda f(x)}] = [dQ = e^g dP] = \mathbb{E}_Q[e^{-g(x)}.e^{\lambda f(x)}.\log(e^{-g(x)}.e^{\lambda f(x)})] -$

$- \mathbb{E}_Q[e^{-g(x)}.e^{\lambda f(x)}].\log \mathbb{E}_Q[e^{-g(x)}.e^{\lambda f(x)}]$

As $x \to x\log x$ is convex, from Jensen's inequality, we have ~~Ent(e^λf(x)) ≥ E[~~

$Ent(e^{\lambda f(x)}) - \mathbb{E}[g.e^{\lambda f(x)}] \geq 0 \quad \forall g: \mathbb{E}e^g = 1$

"=" $\Leftrightarrow \quad g(x) = \log\left(\frac{e^{\lambda f(x)}}{\mathbb{E}e^{\lambda f(x)}}\right)$

+) Case 2: Choose ~~g: Ee^g ≤ 1 and~~ g such that $\mathbb{E}[g.e^{\lambda f(x)}] \leq Ent(e^{\lambda f(x)})$

· If $\mathbb{E}e^g = 0$: there's nothing to prove (because $\mathbb{E}e^g = 0 < 1$)

· If $\mathbb{E}e^g \neq 0$:

Given a positive integer n, larger enough to ensure that $x_n = \mathbb{E}e^{\min(g,n)} > 0$

$\Rightarrow \mathbb{E}\left[g.\frac{e^{\min(g,n)}}{x_n}\right] \leq Ent\left(\frac{e^{\min(g,n)}}{x_n}\right)$

$\Rightarrow \frac{1}{x_n}.\mathbb{E}[g.e^{\min(g,n)}] \leq \frac{1}{x_n}\left[\mathbb{E}(\min(g,n)e^{\min(g,n)}) - \log x_n\right]$

$\Rightarrow \quad \log x_n \leq 0$

$\Rightarrow \quad \min(g,n) \leq 0$

$\Rightarrow \quad g(x) \leq 0$ (because n is a positive integer)

$\Rightarrow \quad \mathbb{E}e^{g(x)} \leq e^0 = 1$

We can conclude that $Ent(e^{\lambda f(x)}) = \sup_{g: \mathbb{E}e^g \leq 1} \mathbb{E}[g(x).e^{\lambda f(x)}]$.

Problem 5: Let X be a standard n-dimensional Gaussian random vector $N(0, I_n)$

Let $f: \mathbb{R}^n \to \mathbb{R}$ be a L-Lipschitz function w.r.t. the Euclidean norm, that's

$$|f(x) - f(y)| \leq L\|x - y\| \quad \text{for all } x, y \in \mathbb{R}^n$$

Introduce a function $H(\lambda) = \frac{1}{\lambda} \log \mathbb{E} e^{\lambda f(X)}$, $\lambda > 0$, $H(0) = \mathbb{E} f(X)$

Using the relation $H'(\lambda) = \frac{\text{Ent}(e^{\lambda f})}{\lambda^2 \mathbb{E} e^{\lambda f}}$ and the logarithmic Sobolev inequality

$$\text{Ent}(g^2) = \mathbb{E} g^2(X) \log g^2(X) - \mathbb{E} g^2(X) \log \mathbb{E} g^2(X) \leq 2\mathbb{E}\|\nabla g(X)\|^2$$

which holds for all almost everywhere differentiable g, prove that

$$H(\lambda) \leq \mathbb{E} f(X) + \frac{L^2 \lambda}{2} \quad \text{for all } \lambda > 0$$

and hence $\log \mathbb{E} e^{\lambda(f(X) - \mathbb{E} f(X))} \leq \frac{L^2 \lambda^2}{2}$

$\to$ Proof: +) Assume that f is differentiable with gradient uniformly bounded by L; ~~$\mathbb{E}f(X)=0$~~.

+ Applying the logarithmic Sobolev inequality for $g = e^{\frac{\lambda f(X)}{2}}$, we have:

$$\begin{cases} \text{Ent}(e^{\lambda f}) \leq 2\,\mathbb{E}\|\nabla e^{\lambda f/2}\|^2 = \frac{\lambda^2}{2} \mathbb{E}\left[e^{\lambda f(X)} \|\nabla f(X)\|^2\right] \leq \frac{\lambda^2 . L^2}{2} . \mathbb{E} e^{\lambda f(X)} \quad (1) \\ \text{Ent}(e^{\lambda f(X)}) = \mathbb{E}\left[\lambda f(X) . e^{\lambda f(X)}\right] - \mathbb{E} e^{\lambda f(X)} . \log(\mathbb{E} e^{\lambda f(X)}) \end{cases}$$

$$(1) \Rightarrow \frac{\text{Ent}(e^{\lambda f(X)})}{\lambda^2 \mathbb{E} e^{\lambda f(X)}} \leq \frac{L^2}{2} \quad (\Leftrightarrow) \quad H'(\lambda) \leq \frac{L^2}{2}$$

+) Integrating the inequality, we have $\int_0^\lambda H'(\lambda) d\lambda \leq \int_0^\lambda \frac{L^2}{2} d\lambda$

$$(\Leftrightarrow) \quad H(\lambda) \leq \frac{L^2 \lambda}{2} + H(0)$$

$$\Rightarrow H(\lambda) \leq \mathbb{E} f(X) + \frac{L^2 \lambda}{2} \quad (\text{because } H(0) = \mathbb{E} f(X))$$

$$\Rightarrow \lambda(H(\lambda) - \mathbb{E} f(X)) \leq \frac{L^2 \lambda^2}{2}$$

$\left[H(\lambda) = \frac{1}{\lambda} \log \mathbb{E} e^{\lambda f(X)} = f(X)\right]$

$$(\Leftrightarrow) \log \mathbb{E} e^{\lambda(f(X) - \mathbb{E} f(X))} \leq \frac{L^2 \lambda^2}{2} \quad \square$$

Được quét bằng CamScanner