Fundamentals of Statistical Learning | Prof. Dr. Celisse Alain | Homework 1

Anton Zaitsev | 0230981826 | anton.zaitsev.001@student.uni.lu | University of Luxembourg

March 30, 2024

Exercise 1

Let $A \subset \mathbb{R}^n$ denote a subset (not necessarily a vector space) of vectors $a = (a_1, \ldots, a_n)^T$, and define the Rademacher complexity of the set A as

$$\mathcal{R}(A) = \mathbb{E}_\epsilon \left[\sup_{a \in A} \langle a, \epsilon \rangle \right],$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ the ϵ_is are independent Rademacher random variables.

---

1. Basic facts

(a)

We aim to justify that

$$\mathcal{R}(A) = \mathcal{R}(-A)$$

Proof

Based on the inner product definition, we can rewrite Rademacher complexity as follows:

$$\begin{aligned}
\mathcal{R}(A) &= \mathbb{E}_\epsilon \left[\sup_{a \in A} \langle a, \epsilon \rangle \right] \\
&= \mathbb{E}_\epsilon \left[\sup_{a \in A} \sum_{i=1}^n a_i \epsilon_i \right]
\end{aligned}$$

Now, consider set $-A$ of vectors $-a = (-a_1, \ldots, -a_n)^T$. For any $-A$, we have:

$$\begin{aligned}
\mathcal{R}(-A) &= \mathbb{E}_\epsilon \left[\sup_{a \in -A} \sum_{i=1}^n a_i \epsilon_i \right] \\
&= \mathbb{E}_\epsilon \left[\sup_{a \in A} \sum_{i=1}^n (-a_i) \epsilon_i \right] \\
&= \mathbb{E}_\epsilon \left[\sup_{a \in A} \sum_{i=1}^n a_i (-\epsilon_i) \right]
\end{aligned}$$

Since ϵ_i and $-\epsilon_i$ follow the same distribution, we have:

$$\begin{aligned}
\mathcal{R}(-A) &= \mathbb{E}_\epsilon \left[\sup_{a \in A} \sum_{i=1}^n a_i (-\epsilon_i) \right] \\
&= \mathbb{E}_\epsilon \left[\sup_{a \in A} \sum_{i=1}^n a_i \epsilon_i \right] \\
&= \mathcal{R}(A)
\end{aligned}$$

(b)

We aim to show that

$$\mathcal{R}(A) \geq 0$$

Proof

Notice that $\sup_{a\in A}\sum_{i=1}^{n} a_i\epsilon_i$ is a convex function, since it is the supremum of a linear function $\sum_{i=1}^{n} a_i\epsilon_i$. Then, by Jensen's inequality for convex functions and the fact that $\forall i \in \{1,\dots,n\}, \mathbb{E}[\epsilon_i]=0$:

$$\begin{aligned}\mathcal{R}(A) &= \mathbb{E}_\epsilon\left[\sup_{a\in A}\sum_{i=1}^{n} a_i\epsilon_i\right]\\ &\geq \sup_{a\in A}\mathbb{E}_\epsilon\left[\sum_{i=1}^{n} a_i\epsilon_i\right]\\ &\geq \sup_{a\in A}\sum_{i=1}^{n}\mathbb{E}[\epsilon_i]\,a_i\\ &= 0\end{aligned}$$

(c)

We aim to prove that

$$\mathcal{R}(A\cup -A) = \mathbb{E}_\epsilon\left[\sup_{a\in A}|\langle a,\epsilon\rangle|\right],$$

and we will give an example to justify that in general

$$\mathcal{R}(A\cup -A) \neq \mathcal{R}(A)$$

Proof

By the definition of Rademacher complexity:

$$\mathcal{R}(A\cup -A) = \mathbb{E}_\epsilon\left[\sup_{a\in A\cup -A}\langle a,\epsilon\rangle\right]$$

Notice that:

$$\sup_{a\in A\cup -A}\langle a,\epsilon\rangle = \sup\left\{\sup_{a\in A}\langle a,\epsilon\rangle,\ \sup_{a\in -A}\langle a,\epsilon\rangle\right\}$$

Since for any $a\in -A, -a\in A$ we have $\langle -a,\epsilon\rangle = -\langle a,\epsilon\rangle$, we get that $\sup_{a\in -A}\langle a,\epsilon\rangle = \sup_{a\in A}\langle -a,\epsilon\rangle = \sup_{a\in A} -\langle a,\epsilon\rangle$ and thus:

$$\begin{aligned}\sup_{a\in A\cup -A}\langle a,\epsilon\rangle &= \sup\left\{\sup_{a\in A}\langle a,\epsilon\rangle,\ \sup_{a\in -A}\langle a,\epsilon\rangle\right\}\\ &= \sup\left\{\sup_{a\in A}\langle a,\epsilon\rangle, \sup_{a\in A} -\langle a,\epsilon\rangle\right\}\\ &= \sup_{a\in A}|\langle a,\epsilon\rangle|\end{aligned}$$

Thus, we get:

$$\begin{aligned}\mathcal{R}(A\cup -A) &= \mathbb{E}_\epsilon\left[\sup_{a\in A\cup -A}\langle a,\epsilon\rangle\right]\\ &= \mathbb{E}_\epsilon\left[\sup_{a\in A}|\langle a,\epsilon\rangle|\right]\end{aligned}$$

To give an example to show that $\mathcal{R}(A \cup -A) \neq \mathcal{R}(A)$, let us define $A = \{(1,1)\} \subset \mathbb{R}^2$ containing a single vector. Then:

$$\begin{aligned}\mathcal{R}(A) &= \frac{1}{2}\left(\frac{1}{4}(1+1) + \frac{1}{4}(1-1) + \frac{1}{4}(-1+1) + \frac{1}{4}(-1-1)\right) \\ &= 0\end{aligned}$$

Now, $A \cup -A = \{(1,1), (-1,1)\}$ and $\mathcal{R}(A \cup -A)$ is calculated as follows:

$$\begin{aligned}\mathcal{R}(A \cup -A) &= \frac{1}{2}\left(\frac{1}{4}\max(1+1, -1-1) + \frac{1}{4}\max(1-1, -1+1) + \frac{1}{4}\max(-1+1, 1-1) + \frac{1}{4}\max(-1-1, 1+1)\right) \\ &= \frac{1}{2} \\ &\neq \mathcal{R}(A)\end{aligned}$$

---

2. Show the next properties

(a)

We aim to prove that

$$A \subset B => \mathcal{R}(A) \leq \mathcal{R}(B)$$

Proof

Since $A \subset B$ and the fact that the supremum is non-decreasing function, then for every ϵ:

$$\sup_{a \in A} \langle a, \epsilon \rangle \leq \sup_{b \in B} \langle b, \epsilon \rangle$$

Thus, since we take expectation over the same ϵ:

$$\mathbb{E}_{\epsilon}\left[\sup_{a \in A} \langle a, \epsilon \rangle\right] \leq \mathbb{E}_{\epsilon}\left[\sup_{b \in B} \langle b, \epsilon \rangle\right]$$

$$\iff$$

$$\mathcal{R}(A) \leq \mathcal{R}(B)$$

(b)

We aim to prove that

$$\mathcal{R}(cA + \{b\}) = |c|\mathcal{R}(A)$$

Proof

First, let us see what happens when we multiply A by a constant c. We have 2 cases: $c \geq 0$ and $c < 0$. Let us first discuss $c \geq 0$.

If $c \geq 0$, then $\sup_{a \in A} \sum_{i=1}^{n} c \cdot a_i \epsilon_i = c \cdot \sup_{a \in A} \sum_{i=1}^{n} a_i \epsilon_i = |c| \sup_{a \in A} \sum_{i=1}^{n} a_i \epsilon_i$.

If $c < 0$, then $\sup_{a \in A} \sum_{i=1}^{n} c \cdot a_i \epsilon_i = \sup_{a \in A} \sum_{i=1}^{n} -|c| a_i \epsilon_i = |c| \sup_{a \in A} \sum_{i=1}^{n} -a_i \epsilon_i = |c| \sup_{a \in A} \sum_{i=1}^{n} a_i(-\epsilon_i)$.

Again, using the fact that ϵ_i and $-\epsilon_i$ follow the same distribution, we achieve:

$$\begin{aligned}\mathcal{R}(A) &= \mathbb{E}_\epsilon\left[\sup_{a\in A}\sum_{i=1}^n c\cdot a_i\epsilon_i\right]\\ &= \mathbb{E}_\epsilon\left[|c|\sup_{a\in A}\sum_{i=1}^n a_i\epsilon_i\right]\end{aligned}$$

Now, let us add $b \in \mathbb{R}^n$:

$$\begin{aligned}\mathcal{R}(cA+\{b\}) &= \mathbb{E}_\epsilon\left[\sup_{a\in A}\sum_{i=1}^n (ca_i+b_i)\epsilon_i\right]\\ &= \mathbb{E}_\epsilon\left[\sup_{a\in A}\sum_{i=1}^n c\cdot a_i\epsilon_i + \sum_{i=1}^n b_i\epsilon_i\right]\\ &= \mathbb{E}_\epsilon\left[\sup_{a\in A}\sum_{i=1}^n c\cdot a_i\epsilon_i\right] + \mathbb{E}\left[\sum_{i=1}^n b_i\epsilon_i\right]\\ &= \mathbb{E}_\epsilon\left[\sup_{a\in A}\sum_{i=1}^n c\cdot a_i\epsilon_i\right] + \sum_{i}^n b_i\mathbb{E}\left[\epsilon_i\right] (= 0 \text{ since } \forall i\in\{1,\dots,n\}, \mathbb{E}\left[\epsilon_i\right]=0)\\ &= \mathbb{E}_\epsilon\left[|c|\sup_{a\in A}\sum_{i=1}^n a_i\epsilon_i\right]\\ &= |c|\mathbb{E}_\epsilon\left[\sup_{a\in A}\sum_{i=1}^n a_i\epsilon_i\right]\\ &= |c|\mathcal{R}(A)\end{aligned}$$

(c)

We aim to prove that

$$\mathcal{R}(A+B) = \mathcal{R}(A) + \mathcal{R}(B)$$

Proof

Let us rewrite $\mathcal{R}(A+B)$ as follows:

$$\begin{aligned}\mathcal{R}(A+B) &= \mathcal{R}(a+b|a\in A, b\in B)\\ &= \mathbb{E}_\epsilon\left[\sup_{a\in A, b\in B}\sum_{i=1}^n (a_i+b_i)\epsilon_i\right]\\ &= \mathbb{E}_\epsilon\left[\sup_{a\in A}\sum_{i=1}^n a_i\epsilon_i + \sup_{b\in B}\sum_{i=1}^n b_i\epsilon_i\right]\\ &= \mathbb{E}_\epsilon\left[\sup_{a\in A}\sum_{i=1}^n a_i\epsilon_i\right] + \mathbb{E}_\epsilon\left[\sup_{b\in B}\sum_{i=1}^n b_i\epsilon_i\right]\\ &= \mathcal{R}(A)+\mathcal{R}(B)\end{aligned}$$

- **(d)** $\mathcal{R}(Conv(A)) = \mathcal{R}(A)$, with $Conv(A) = \{\sum_{i=1}^n \theta_i a_i | (\theta_i)_{1\le i\le n} \in (\mathbb{R}_+)^n, \sum_{i=1}^n \theta_i = 1\}$

$$
\begin{aligned}
\mathcal{R}(Conv(A)) &= \mathcal{R}\left(\sum_{i=1}^{n}\theta_i a_i\right) \\
&= \mathbb{E}_\epsilon\left[\sup_{a\in A,(\theta_i)_{1\le i\le n}\in(\mathbb{R}_+)^n,\sum_{i=1}^n\theta_i=1}\sum_{i=1}^{n}\left(\sum_{i=1}^{n}\theta_i a_i\right)\epsilon_i\right]
\end{aligned}
$$

Since θ_i are non-negative scalars, we can use the fact that for f real-valued function, $\sup_{i\in\{1,\dots,n\}}\sum_i^n\theta_i f(a_i) = \sum_i^n\theta_i\sup_{i\in\{1,\dots,n\}} f(a_i)$. In our case, $f(a_i) = a_i$ and we get:

$$
\begin{aligned}
\mathcal{R}(Conv(A)) &= \mathbb{E}_\epsilon\left[\sup_{a\in A,(\theta_i)_{1\le i\le n}\in(\mathbb{R}_+)^n,\sum_{i=1}^n\theta_i=1}\sum_{i=1}^{n}\left(\sum_{i=1}^{n}\theta_i a_i\right)\epsilon_i\right] \\
&= \mathbb{E}_\epsilon\left[\sum_{i=1,(\theta_i)_{1\le i\le n}\in(\mathbb{R}_+)^n,\sum_{i=1}^n\theta_i=1}^{n}\theta_i\sup_{a\in A}\sum_{i=1}^{n}a_i\epsilon_i\right] \\
&= \sum_{i=1,(\theta_i)_{1\le i\le n}\in(\mathbb{R}_+)^n,\sum_{i=1}^n\theta_i=1}^{n}\theta_i\mathbb{E}_\epsilon\left[\sup_{a\in A}\sum_{i=1}^{n}a_i\epsilon_i\right] \\
&= \sum_{i=1,(\theta_i)_{1\le i\le n}\in(\mathbb{R}_+)^n,\sum_{i=1}^n\theta_i=1}^{n}\theta_i\mathcal{R}(A) \\
&= \mathcal{R}(A)(\text{since }\sum_{i=1}^{n}\theta_i = 1)
\end{aligned}
$$

---

3. Bounded-difference inequality

(a)

We aim to prove that

$$
|\varphi(D_n) - \varphi(D'_n(i))| \le \frac{2M}{n}
$$

Proof

Let us rewrite $|\varphi(D_n) - \varphi(D'_n(i))|$ according to the definition of the $\varphi(x)$ function:

$$
\begin{aligned}
|\varphi(D_n) - \varphi(D'_n(i))| &= \left|\sup_{f\in F}\left\{\frac{1}{n}\sum_{j=1}^{n}(\mathbb{E}[f(X_j)] - f(X_j))\right\} - \sup_{f\in F}\left\{\frac{1}{n}\sum_{j=1}^{n}\left(\mathbb{E}[f(X'_j)] - f(X'_j)\right)\right\}\right| \\
&= \left|\sup_{f\in F}\left\{\frac{1}{n}\sum_{j=1}^{n}(\mathbb{E}[f(X_j)] - f(X_j))\right\} - \sup_{f\in F}\left\{\frac{1}{n}\left(\sum_{j=1,j\ne i}^{n}(\mathbb{E}[f(X_j)] - f(X_j)) + \mathbb{E}[f(X'_i)] - f(X'_i)\right)\right\}\right|
\end{aligned}
$$

Now, we can use the fact that for $f, g \in \mathbb{R}^I, \forall x \in I, \sup(f-g)(I) \ge \sup f(I) - \sup g(I)$, ("triangle inequality for the supremum") and get:

$$
\begin{aligned}
|\varphi(D_n) - \varphi(D'_n(i))| &= \left| \sup_{f\in F} \left\{ \frac{1}{n}\sum_{j=1}^{n} (\mathbb{E}[f(X_j)] - f(X_j)) \right\} - \sup_{f\in F} \left\{ \frac{1}{n} \left(\sum_{j=1, j\neq i}^{n} (\mathbb{E}[f(X_j)] - f(X_j)) + \mathbb{E}[f(X'_i)] - f(X'_i) \right) \right\} \right| \\
&\leq \left| \sup_{f\in F} \left\{ \frac{1}{n}\sum_{j=1}^{n} (\mathbb{E}[f(X_j)] - f(X_j)) - \frac{1}{n} \left(\sum_{j=1, j\neq i}^{n} (\mathbb{E}[f(X_j)] - f(X_j)) + \mathbb{E}[f(X'_i)] - f(X'_i) \right) \right\} \right| \\
&= \left| \sup_{f\in F} \left\{ \frac{1}{n} \left(\sum_{j=1}^{n} (\mathbb{E}[f(X_j)] - f(X_j)) - \sum_{j=1, j\neq i}^{n} (\mathbb{E}[f(X_j)] - f(X_j)) + \mathbb{E}[f(X'_i)] - f(X'_i) \right) \right\} \right| \\
&= \left| \sup_{f\in F} \left\{ \frac{1}{n} (\mathbb{E}[f(X_i)] - f(X_i) - \mathbb{E}[f(X'_i)] + f(X'_i)) \right\} \right|
\end{aligned}
$$

Since X_i and X'_i have the same distribution, $\mathbb{E}[f(X_i)] = \mathbb{E}[f(X_i)']$. Thus:

$$
\begin{aligned}
|\varphi(D_n) - \varphi(D'_n(i))| &\leq \left| \sup_{f\in F} \left\{ \frac{1}{n} (\mathbb{E}[f(X_i)] - f(X_i) - \mathbb{E}[f(X'_i)] + f(X'_i)) \right\} \right| \\
&= \left| \sup_{f\in F} \left\{ \frac{1}{n} (f(X'_i) - f(X_i)) \right\} \right| \\
&= \frac{1}{n} \left| \sup_{f\in F} \{ f(X'_i) - f(X_i) \} \right|
\end{aligned}
$$

We know that $\sup_{f\in F} \|f\|_\infty \leq M < +\infty$, and thus by triangle inequality: $|f(X'_i) - f(X_i)| \leq |f(X'_i)|_{\leq M} + |f(X_i)|_{\leq M} \leq 2M$

$$
\begin{aligned}
|\varphi(D_n) - \varphi(D'_n(i))| &\leq \frac{1}{n} \left| \sup_{f\in F} \{ f(X'_i) - f(X_i) \} \right| \\
&= \frac{2M}{n}
\end{aligned}
$$

(b)

We aim to prove that for every $t > 0$,

$$
\mathbb{P}\left[\sup_f (P - P_n) f - \mathbb{E}\left[\sup_f (P - P_n) f \right] > t \right] \leq e^{-\frac{2nt^2}{(2M)^2}}
$$

$$
\mathbb{P}\left[\sup_f (P_n - P) f - \mathbb{E}\left[\sup_f (P_n - P) f \right] > t \right] \leq e^{-\frac{2nt^2}{(2M)^2}}
$$

Proof

From **(a)** we know that $\varphi : \mathcal{X}^n \to \mathbb{R}$ is a function such that:

$$
\begin{aligned}
|\varphi(D_n) - \varphi(D'_n(i))| &= \sup_{X_1,\ldots,X_n,X'_i \in R} |\varphi(X_1,\ldots,X_i,\ldots,X_n) - \varphi(X_1,\ldots,X'_i,\ldots,X_n)| \\
&\leq \frac{2M}{n}
\end{aligned}
$$

Thus, we can apply the **BDI** theorem to this function and get:

$$
\begin{aligned}
\mathbb{P}\left[\sup_f (P-P_n)f - \mathbb{E}\left[\sup_f (P-P_n)f\right] > t\right] &\le e^{-\frac{2t^2}{\sum_{i=1}^n (2M/n)^2}} \\
&= e^{-\frac{2nt^2}{(2M)^2}}
\end{aligned}
$$

Same applies for the following:

$$
\begin{aligned}
\mathbb{P}\left[\sup_f (P-P_n)f - \mathbb{E}\left[\sup_f (P-P_n)f\right] < -t\right] &\le e^{-\frac{2t^2}{\sum_{i=1}^n (2M/n)^2}} \\
&= e^{-\frac{2nt^2}{(2M)^2}}
\end{aligned}
$$

---

4. Symmetrization

(a)

We aim to justify that

$$
\mathbb{E}\left[\sup_{f\in F} (P-P_n)f\right] \le \mathbb{E}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n (f(X_i)-f(X_i'))\right\}\right]
$$

Proof

Let us rewrite $\mathbb{E}\left[\sup_{f\in F}(P-P_n)f\right]$ as follows (definition):

$$
\mathbb{E}\left[\sup_{f\in F} (P-P_n)f\right] = \mathbb{E}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n (\mathbb{E}[f(X_i)]-f(X_i))\right\}\right]
$$

Since X_i' is a copy of X_i, then $\mathbb{E}[X_i]=\mathbb{E}[X_i']$ and $\mathbb{E}_{X_i'}[X_i]=X_i$. Thus, we can rewrite the above inequality as follows:

$$
\begin{aligned}
\mathbb{E}\left[\sup_{f\in F} (P-P_n)f\right] &= \mathbb{E}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n (\mathbb{E}[f(X_i)]-f(X_i))\right\}\right] \\
&= \mathbb{E}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n \mathbb{E}_{X_i'}\left[f(X_i')-f(X_i)\right]\right\}\right]
\end{aligned}
$$

Now, due to the fact that the supremum of an expectation is **at most** an expectation of a supremum (by Jensen's inequality), we can bound above as follows:

$$
\begin{aligned}
\mathbb{E}\left[\sup_{f\in F} (P-P_n)f\right] &= \mathbb{E}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n \mathbb{E}_{X_i'}\left[f(X_i')-f(X_i)\right]\right\}\right] \\
&\le \mathbb{E}\left[\mathbb{E}_{X_i'}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n (f(X_i')-f(X_i))\right\}\right]\right] \\
&\le \mathbb{E}_{X_i,X_i'}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n (f(X_i')-f(X_i))\right\}\right]
\end{aligned}
$$

Lastly, since we take the expectation of $f(X_i)-f(X_i')$, we know that this would be equal to the expecation of $f(X_i')-f(X_i)$, since they follow the same probability distribution. Thus, we finally get:

$$
\begin{aligned}
\mathbb{E}\left[\sup_{f\in F}(P-P_n)f\right] &\leq \mathbb{E}_{X_i,X_i'}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(f(X_i')-f(X_i)\right)\right\}\right] \\
&= \mathbb{E}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(f(X_i)-f(X_i')\right)\right\}\right]
\end{aligned}
$$

Notice that here to denote $\mathbb{E}_{X_i,X_i'}[\dots]$ is the same as to denote $\mathbb{E}[\dots]$.

(b)

We aim to show that

$$
f(X_i)-f(X_i') \stackrel{\mathcal{L}}{=} \epsilon_i\left[f(X_i)-f(X_i')\right]
$$

and deduce that the join distribution of $(f(X_i)-f(X_i'))_{i\leq i\leq n}$ is equal to the one of $(\epsilon_i\left[f(X_i)-f(X_i')\right])_{i\leq i\leq n}$

Proof

Let us study the distribution of $\epsilon_i(f(X_i)-f(X_i'))$:

$$
\begin{aligned}
\mathbb{P}\left[\epsilon_i(f(X_i)-f(X_i'))\leq t\right] &\stackrel{\text{Law of Total Prob.}}{=} \mathbb{P}\left[\epsilon_i(f(X_i)-f(X_i'))\leq t|\epsilon_i=1\right]\mathbb{P}[\epsilon_i=1] \\
&\quad + \mathbb{P}\left[\epsilon_i(f(X_i)-f(X_i'))\leq t|\epsilon_i=-1\right]\mathbb{P}[\epsilon_i=-1] \\
&= \frac{1}{2}\mathbb{P}\left[f(X_i)-f(X_i')\leq t\right]+\frac{1}{2}\mathbb{P}\left[f(X_i')-f(X_i)\leq t\right]
\end{aligned}
$$

Since X_i and X_i' are independent copies, we get that $\mathbb{P}\left[f(X_i)-f(X_i')\leq t\right]=\mathbb{P}\left[f(X_i')-f(X_i)\leq t\right]$ and thus:

$$
\mathbb{P}\left[\epsilon_i(f(X_i)-f(X_i'))\leq t\right]=\mathbb{P}\left[f(X_i)-f(X_i')\leq t\right] \quad ,
$$

thus proving the fact that $f(X_i)-f(X_i') \stackrel{\mathcal{L}}{=} \epsilon_i\left[f(X_i)-f(X_i')\right]$.

For the joint distribution of $(\epsilon_i(f(X_i)-f(X_i')))_{1\leq i\leq n}$ we have that:

$$
\mathbb{P}\left[\epsilon_1(f(X_1)-f(X_1'))\leq t_1,\dots,\epsilon_n(f(X_n)-f(X_n'))\leq t_n\right]=\prod_{i=1}^{n}\mathbb{P}\left[\epsilon_i(f(X_i)-f(X_i'))\leq t_i\right] \quad ,
$$

which is true since $\epsilon_i(f(X_i)-f(X_i'))\perp\epsilon_j(f(X_j)-f(X_j')) \quad \forall 1\leq i,j\leq n, i\neq j$.

We also know that $\mathbb{P}\left[\epsilon_i(f(X_i)-f(X_i'))\leq t\right]=\mathbb{P}\left[f(X_i)-f(X_i')\leq t\right]$ and thus we conclude:

$$
\begin{aligned}
\mathbb{P}\left[\epsilon_1(f(X_1)-f(X_1'))\leq t_1,\dots,\epsilon_n(f(X_n)-f(X_n'))\leq t_n\right] &= \prod_{i=1}^{n}\mathbb{P}\left[\epsilon_i(f(X_i)-f(X_i'))\leq t_i\right] \\
&= \prod_{i=1}^{n}\mathbb{P}\left[f(X_i)-f(X_i')\leq t_i\right] \\
&\stackrel{\perp}{=} \mathbb{P}\left[(f(X_1)-f(X_1'))\leq t_1,\dots,(f(X_n)-f(X_n'))\leq t_n\right]
\end{aligned}
$$

So we deduce that the join distribution of $(f(X_i)-f(X_i'))_{i\leq i\leq n}$ is equal to the one of $(\epsilon_i\left[f(X_i)-f(X_i')\right])_{i\leq i\leq n}$

(c)

We aim to deduce that

$$\mathbb{E}\left[\sup_f (P-P_n)f\right] \leq 2\mathbb{E}_{D,\epsilon}\left[\sup_f \left\{\frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i)\right\}\right]$$

Proof

From part **(a)** we know that:

$$\mathbb{E}\left[\sup_{f\in F}(P-P_n)f\right] \leq \mathbb{E}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n (f(X_i')-f(X_i))\right\}\right]$$

Using **(b)**, i.e. the fact that the join distribution of $(f(X_i)-f(X_i'))_{i\leq i\leq n}$ is equal to the one of $(\epsilon_i\left[f(X_i)-f(X_i')\right])_{i\leq i\leq n}$, the right side term in the above inequality is equal to the following:

$$\mathbb{E}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n (f(X_i')-f(X_i))\right\}\right] = \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n \epsilon_i\left(f(X_i')-f(X_i)\right)\right\}\right]$$

Thus:

$$\mathbb{E}\left[\sup_{f\in F}(P-P_n)f\right] \leq \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n \epsilon_i\left(f(X_i')-f(X_i)\right)\right\}\right]$$

Let us study the term on the right side of the inequality:

$$\begin{aligned}
\mathbb{E}\left[\sup_{f\in F}(P-P_n)f\right] &\leq \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n \epsilon_i\left(f(X_i')-f(X_i)\right)\right\}\right]\\
&\leq \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i') - \frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i)\right\}\right]\\
&\leq \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i') + \frac{1}{n}\sum_{i=1}^n (-\epsilon_i) f(X_i)\right\}\right]
\end{aligned}$$

Since supremum of the sum is less than the sum of the supremums:

$$\begin{aligned}
\mathbb{E}\left[\sup_{f\in F}(P-P_n)f\right] &\leq \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i') + \frac{1}{n}\sum_{i=1}^n (-\epsilon_i) f(X_i)\right\}\right]\\
&\leq \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i')\right\}\right] + \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n (-\epsilon_i) f(X_i)\right\}\right]
\end{aligned}$$

Notice that because of the symmetry of the Rademacher random variables, the expectations over positive and negative ϵ_i are the same. Hence, this expression simplifies to twice the expectation over positive ϵ_i:

$$\begin{aligned}
\mathbb{E}\left[\sup_{f\in F}(P-P_n)f\right] &\leq \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i')\right\}\right] + \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n (-\epsilon_i) f(X_i)\right\}\right]\\
&= \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i')\right\}\right] + \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i)\right\}\right]
\end{aligned}$$

Since X_i' is independent copy of X_i:

$$\begin{aligned}
\mathbb{E}\left[\sup_{f\in F}(P-P_n)f\right] &\le \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i')\right\}\right] + \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i)\right\}\right] \\
&= \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i)\right\}\right] + \mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i)\right\}\right] \\
&= 2\mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i)\right\}\right]
\end{aligned}$$

(d)

Let $\mathcal{F}(D_n) = \{(f(X_1), \dots, f(X_n)) \in \mathbb{R}^n \mid f \in \mathcal{F}\}$. We aim to conclude that

$$\mathbb{E}\left[\sup_{f\in F}(P-P_n)f\right] \le 2\mathbb{E}_D\left[\mathcal{R}(\mathcal{F}(D_n)/n)\right]$$

Proof

By denoting

$$\mathcal{R}(\mathcal{F}(D_n)/n) = \frac{1}{n}\mathbb{E}_\epsilon\left[\sup_{f\in F}\left\{\sum_{i=1}^{n}\epsilon_i f(X_i)\right\}\right],$$

we can conclude from **(c)** that

$$\begin{aligned}
\mathbb{E}\left[\sup_{f\in F}(P-P_n)f\right] &\le 2\mathbb{E}_{D,\epsilon}\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i)\right\}\right] \\
&\le 2\mathbb{E}_D\left[\frac{1}{n}\mathbb{E}_\epsilon\left[\sup_{f\in F}\left\{\sum_{i=1}^{n}\epsilon_i f(X_i)\right\}\right]\right] \\
&\le 2\mathbb{E}_D\left[\mathcal{R}(\mathcal{F}(D_n)/n)\right]
\end{aligned}$$

Exercise 2

1. ψ_2-Orlicz norm

We define ψ_2-Orlicz norm as:

$$||X||_{\psi_2} := \inf \left\{ t > 0 \mid \mathbb{E}\left[\exp\left(\frac{X^2}{t^2}\right)\right] \leq 2 \right\}$$

(a)

We aim to prove that

$$||X||_{\psi_2} = 0 => \mathbb{E}[X^2] \leq t^2 \ln(2), \forall t \geq 0$$

Then, we will deduce that $||X||_{\psi_2} = 0$ implies $X = 0$ a.s.

Proof

Let $f(x) = \exp(x)$ a convex function. Jensen's inequality states that for any random variable X and convex function f:

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$$

According to the Jensen's inequality and the definition of the ψ_2-Orlicz norm:

$$\exp\left(\mathbb{E}\left[\frac{X^2}{t^2}\right]\right)^2 \leq \mathbb{E}\left[\exp\left(\frac{X^2}{t^2}\right)\right] \leq 2$$

Thus, we get:

$$\exp\left(\mathbb{E}\left[\frac{X^2}{t^2}\right]\right) \leq 2$$

$$\iff$$

$$\mathbb{E}\left[\frac{X^2}{t^2}\right] \leq \ln(2)$$

Due to the linearity of expectation, we can rewrite the left term as:

$$\mathbb{E}\left[\frac{X^2}{t^2}\right] = \frac{1}{t^2}\mathbb{E}\left[X^2\right]$$

Thus, we get:

$$\frac{1}{t^2}\mathbb{E}\left[X^2\right] \leq \ln(2)$$

$$\iff$$

$$\mathbb{E}\left[X^2\right] \leq t^2 \ln(2)$$

According to the definition of ψ_2-norm, if $||X||_{\psi_2} = 0$, then $t = 0$, meaning that $\mathbb{E}\left[X^2\right] \leq 0$. Notice that for any real-valued random variable X, $\mathbb{E}\left[X^2\right] \geq 0$ with equality if and only if $X = 0$ almost surely. Thus, we get that if $||X||_{\psi_2} = 0$, then $X = 0$ almost surely.

(b)

For all $\lambda \in \mathbb{R}$, we aim to prove that

$$||\lambda X||_{\psi_2} = |\lambda| \, ||X||_{\psi_2}$$

Proof

$$||\lambda X||_{\psi_2} = \inf\left\{t > 0 \mid \mathbb{E}\left[\exp\left(\frac{(\lambda X)^2}{t^2}\right)\right] \leq 2\right\}$$

Let $t = |\lambda| \, k$. Then:

$$\begin{aligned} ||\lambda X||_{\psi_2} &= \inf\left\{|\lambda| \, k > 0 \mid \mathbb{E}\left[\exp\left(\frac{(\lambda X)^2}{(|\lambda| \, k)^2}\right)\right] \leq 2\right\} \\ &= \inf\left\{|\lambda| \, k > 0 \mid \mathbb{E}\left[\exp\left(\frac{X^2}{k^2}\right)\right] \leq 2\right\} \end{aligned}$$

We can take $|\lambda|$ out of the infimum since k does not depend on it. Thus:

$$\begin{aligned} ||\lambda X||_{\psi_2} &= |\lambda| \inf\left\{k > 0 \mid \mathbb{E}\left[\exp\left(\frac{X^2}{k^2}\right)\right] \leq 2\right\} \\ &= |\lambda| \, ||X||_{\psi_2} \end{aligned}$$

(c)

We aim to prove that

$$||X + Y||_{\psi_2} \leq ||X||_{\psi_2} + ||Y||_{\psi_2},$$

for X, Y with finite ψ_2-norm (triangle inequality).

Proof

Let $f(u) = \exp\left(u^2\right)$ a convex and increasing function. We can write:

$$f\left(\frac{|X+Y|}{a+b}\right) \leq f\left(\frac{|X| + |Y|}{a+b}\right)$$

The inequality holds because of the triangle inequality for the absolute value function and the fact that f is convex and increasing. Then, by Jensen's inequality for real convex functions, we have:

$$\begin{aligned} f\left(\frac{|X+Y|}{a+b}\right) \leq f\left(\frac{|X| + |Y|}{a+b}\right) &= f\left(\frac{|X|}{a}\frac{a}{a+b} + \frac{|Y|}{b}\frac{b}{a+b}\right) \\ &\leq \frac{a}{a+b} f\left(\frac{|X|}{a}\right) + \frac{b}{a+b} f\left(\frac{|Y|}{b}\right) \end{aligned}$$

Let $a = ||X||_{\psi_2}, b = ||Y||_{\psi_2}$. Then:

$$f\left(\frac{|X+Y|}{||X||_{\psi_2} + ||Y||_{\psi_2}}\right) \leq \frac{||X||_{\psi_2}}{||X||_{\psi_2} + ||Y||_{\psi_2}} f\left(\frac{|X|}{||X||_{\psi_2}}\right) + \frac{||Y||_{\psi_2}}{||X||_{\psi_2} + ||Y||_{\psi_2}} f\left(\frac{|Y|}{||Y||_{\psi_2}}\right)$$

Let us take *expectation* on both sides:

$$\mathbb{E}\left[f\left(\frac{|X+Y|}{||X||_{\psi_2}+||Y||_{\psi_2}}\right)\right] \leq \frac{||X||_{\psi_2}}{||X||_{\psi_2}+||Y||_{\psi_2}}\mathbb{E}\left[f\left(\frac{|X|}{||X||_{\psi_2}}\right)\right] + \frac{||Y||_{\psi_2}}{||X||_{\psi_2}+||Y||_{\psi_2}}\mathbb{E}\left[f\left(\frac{|Y|}{||Y||_{\psi_2}}\right)\right]$$

Notice that

$$\mathbb{E}\left[f\left(\frac{|X+Y|}{||X||_{\psi_2}+||Y||_{\psi_2}}\right)\right] = \mathbb{E}\left[\exp\left(\frac{|X+Y|}{||X||_{\psi_2}+||Y||_{\psi_2}}\right)^2\right] \leq 2,$$
$$\mathbb{E}\left[f\left(\frac{|X|}{||X||_{\psi_2}}\right)\right] = \mathbb{E}\left[\exp\left(\frac{X}{||X||_{\psi_2}}\right)^2\right] \leq 2,$$
$$\mathbb{E}\left[f\left(\frac{|Y|}{||Y||_{\psi_2}}\right)\right] = \mathbb{E}\left[\exp\left(\frac{|Y|}{||Y||_{\psi_2}}\right)^2\right] \leq 2,$$

by the definition of ψ_2-norm. Thus:

$$\begin{aligned}\mathbb{E}\left[\exp\left(\frac{X+Y}{||X||_{\psi_2}+||Y||_{\psi_2}}\right)^2\right] &\leq \frac{||X||_{\psi_2}}{||X||_{\psi_2}+||Y||_{\psi_2}}2 + \frac{||Y||_{\psi_2}}{||X||_{\psi_2}+||Y||_{\psi_2}}2 \\ &= \frac{2||X||_{\psi_2}+2||Y||_{\psi_2}}{||X||_{\psi_2}+||Y||_{\psi_2}} \\ &= 2\end{aligned}$$

We can notice that $||X||_{\psi_2}+||Y||_{\psi_2}$ belongs to the set $S = \left\{t>0|\mathbb{E}\left[\exp\left(\frac{Z}{t}\right)^2\right] \leq 2\right\}$. Now, $||X+Y||_{\psi_2}$ is also within the set S, since it is the **smallest** value of t that satisfies the inequality for the random variable $X+Y$. Given that $||X||_{\psi_2}+||Y||_{\psi_2}$ and $||X+Y||_{\psi_2}$ are both within the set S and the latter is the smallest, we conclude that:

$$||X+Y||_{\psi_2} \leq ||X||_{\psi_2}+||Y||_{\psi_2}$$

---

2. Centering Lemma

Given: X is a real-valued random variable with finite ψ_2-norm and

$$||X||_k \leq \sqrt{\frac{e}{2}}||X||_{\psi_2}k, \quad \forall k \in \mathbb{N}^*$$

(a)

We aim to prove that for every $t>0$,

$$\mathbb{P}\left[|X|>t\right] \leq 2\exp\left(-\frac{t^2}{||X||_{\psi_2}}\right)$$

Proof

Let $f(t) = \exp(\frac{t^2}{||X||_{\psi_2}})$, where X is a real-valued random variable and $||X||_{\psi_2}$ is finite. Then, by Markov's inequality:

$$
\begin{aligned}
\mathbb{P}\left[|X|>t\right] &= \mathbb{P}\left[|f(X)|>f(t)\right] \\
&= \mathbb{P}\left[\exp\left(\frac{X^2}{||X||_{\psi_2}}\right) > \exp\left(\frac{t^2}{||X||_{\psi_2}}\right)\right] \\
&\leq \frac{\mathbb{E}[\exp\left(\frac{X^2}{||X||_{\psi_2}}\right)] \leq 2}{\exp\left(\frac{t^2}{||X||_{\psi_2}}\right)} \\
&\leq 2\exp\left(-\frac{t^2}{||X||_{\psi_2}}\right)
\end{aligned}
$$

(b)

We aim to show that

$$
||X||_k \leq \frac{k^{1/k}}{\sqrt{2}}||X||_{\psi_2}\sqrt{k} \leq \frac{\sqrt{e}}{\sqrt{2}}||X||_{\psi_2}\sqrt{k}, \quad \forall k \in \mathbb{N}^*
$$

Proof

$\forall k \in \mathbb{N}^*$ we have that $||X||_k = \mathbb{E}[|X|^k]^{\frac{1}{k}}$. Thus we can rewrite it as follows:

$$
\begin{aligned}
||X||_k^k &= \mathbb{E}[|X|^k] \\
&= \mathbb{E}[\int_0^{|X|^k} du] \\
&= \mathbb{E}\left[\int_0^{+\infty} 1_{\{|X|^k>u\}} du\right] \\
&\overset{\text{Fubini-Tonelli}}{=} \int_0^{+\infty} \mathbb{E}\left[1_{\{|X|^k>u\}}\right] du \\
&= \int_0^{+\infty} \mathbb{P}\left[|X|^k > u\right] du \\
&= \int_0^{+\infty} \mathbb{P}\left[|X| > u^{\frac{1}{k}}\right] du \quad |u^{\frac{1}{k}} = t, du = kt^{k-1}dt \\
&= \int_0^{+\infty} kt^{k-1}\left(\mathbb{P}\left[|X|>t\right]\right) dt \\
&\overset{\text{from (a)}}{\leq} \int_0^{+\infty} kt^{k-1}\left(2\exp\left(-\frac{t^2}{||X||_{\psi_2}}\right)\right) dt \quad |v = \frac{t^2}{||X||_{\psi_2}}, t \quad = \sqrt{v}\sqrt{||X||_{\psi_2}}, dt = \frac{\sqrt{||X||_{\psi_2}}}{2}\frac{1}{\sqrt{v}}dv \\
&= \int_0^{+\infty} k\left(v||X||_{\psi_2}\right)^{\frac{k-1}{2}} 2\exp\left(-v\right)\frac{\sqrt{||X||_{\psi_2}}}{2}\frac{1}{\sqrt{v}}dv \\
&= k||X||_{\psi_2}^{\frac{k}{2}} \int_0^{+\infty} v^{\frac{k-2}{2}} \exp\left(-v\right) dv \\
&= k||X||_{\psi_2}^{\frac{k}{2}} \int_0^{+\infty} v^{\frac{k-2}{2}} \exp\left(-v\right) dv \\
&= k||X||_{\psi_2}^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right) \\
&\leq k||X||_{\psi_2}^{\frac{k}{2}}\left(\frac{k}{2}\right)^{\frac{k}{2}}
\end{aligned}
$$

Thus, we have:

$$||X||_k^k \leq k||X||_{\psi_2}^{\frac{k}{2}} \left(\frac{k}{2}\right)^{\frac{k}{2}}$$

$$\iff$$

$$\begin{aligned} ||X||_k &\leq \frac{k^{1/k}}{\sqrt{2}} \sqrt{||X||_{\psi_2}} \sqrt{k} \\ &\leq \frac{k^{1/k}}{\sqrt{2}} ||X||_{\psi_2} \sqrt{k} \quad \text{for } ||X||_{\psi_2} \geq 1 \end{aligned}$$

Since $k^{\frac{1}{k}} = \exp\left(\frac{\ln(k)}{k}\right)$ and $\frac{\ln(k)}{k} \leq \frac{1}{2}$, we can bound $\exp\left(\frac{\ln(k)}{k}\right) \leq \exp\left(\frac{1}{2}\right)$ and rewrite the above inequality as:

$$||X||_k \leq \frac{\exp\left(\frac{1}{2}\right)}{\sqrt{2}} ||X||_{\psi_2} \sqrt{k} \quad \text{for } ||X||_{\psi_2} \geq 1$$

(c)

We aim to prove that

$$||\mathbb{E}[X]||_{\psi_2} \leq \frac{\mathbb{E}[X]}{\sqrt{\log(2)}}$$

Proof

From the definition of the ψ_2-norm:

$$||\mathbb{E}[X]||_{\psi_2} := \inf \left\{ t > 0 \mid \mathbb{E}\left[\exp\left(\frac{\mathbb{E}[X]}{t}\right)^2\right] \leq 2 \right\}$$

From the Jensen's inequality and the definition of the ψ_2-norm we know that:

$$\exp\left(\mathbb{E}\left[\frac{\mathbb{E}[X]}{t}\right]^2\right) \leq \mathbb{E}\left[\exp\left(\frac{\mathbb{E}[X]}{t}\right)^2\right] \leq 2$$

$$\iff$$

$$\exp\left(\mathbb{E}\left[\frac{\mathbb{E}[X]}{t}\right]^2\right) \leq 2$$

$$\iff$$

$$\mathbb{E}\left[\frac{\mathbb{E}[X]}{t}\right]^2 \leq \log(2)$$

$$\iff$$

$$\frac{1}{t}\mathbb{E}\left[\mathbb{E}[X]\right] \leq \sqrt{\log(2)}$$

$$\iff$$

$$t \geq \frac{\mathbb{E}\left[\mathbb{E}[X]\right]}{\sqrt{\log(2)}}$$

$$\iff$$

$$t \geq \frac{\mathbb{E}\left[X\right]}{\sqrt{\log(2)}}$$

$$\iff$$

$$||\mathbb{E}[X]||_{\psi_2} \leq \frac{\mathbb{E}\left[X\right]}{\sqrt{\log(2)}}$$

(d)

We aim to deduce that

$$\exists c_2 > 0 \text{ s.t. } ||\mathbb{E}[X]||_{\psi_2} \leq c_2 ||X||_{\psi_2}$$

Proof

From **(c)** we know that

$$\begin{aligned} ||\mathbb{E}[X]||_{\psi_2} &\leq \frac{\mathbb{E}\left[X\right]}{\sqrt{\log(2)}} \\ &\leq \frac{|\mathbb{E}\left[X\right]|}{\sqrt{\log(2)}} \end{aligned}$$

By using the Jensen's inequality we get:

$$\begin{aligned} ||\mathbb{E}[X]||_{\psi_2} &\leq \frac{|\mathbb{E}\left[X\right]|}{\sqrt{\log(2)}} \\ &\leq \frac{\mathbb{E}\left[|X|\right]}{\sqrt{\log(2)}} \end{aligned}$$

Notice that $\mathbb{E}\left[|X|\right] = ||X||_1$. From **(b)** we know that:

$$||X||_1 \leq \frac{||X||_{\psi_2}}{\sqrt{2}} \iff \mathbb{E}\left[|X|\right] \leq \frac{||X||_{\psi_2}}{\sqrt{2}}$$

Thus, we conclude:

$$
\begin{aligned}
||\mathbb{E}[X]||_{\psi_2} &\leq \frac{\mathbb{E}\left[|X|\right]}{\sqrt{\log(2)}} \\
&\leq \frac{||X||_{\psi_2}}{\sqrt{2\log(2)}} \\
&\leq c_2||X||_{\psi_2}, \text{ with } c_2 = \frac{1}{\sqrt{(2\log(2))}}
\end{aligned}
$$

(e)

We aim to show that

$$||X - \mathbb{E}[X]||_{\psi_2} \leq (1 + c_2)||X||_{\psi_2}$$

Proof

By using the triangle inequality that we proved in **Exercise 2, 1, (c)**, we can bound $||X - \mathbb{E}[X]||_{\psi_2}$ as:

$$||X - \mathbb{E}[X]||_{\psi_2} \leq ||X||_{\psi_2} + ||\mathbb{E}[X]||_{\psi_2}$$

From **Exercise 2, 2, (d)**, we know that

$$||\mathbb{E}[X]||_{\psi_2} \leq c_2||X||_{\psi_2}$$

Thus, substituting it into the triangle inequality, we get:

$$
\begin{aligned}
||X - \mathbb{E}[X]||_{\psi_2} &\leq ||X||_{\psi_2} + ||\mathbb{E}[X]||_{\psi_2} \\
&\leq ||X||_{\psi_2} + c_2||X||_{\psi_2} \\
&\leq ||X||_{\psi_2}(1 + c_2)
\end{aligned}
$$