

Uniform Laws of Large Numbers

Jiuzhou Miao

March 19, 2025

Proof of Theorem 4. Denote $\check{f}(x) = f(x) - \mathbb{E}\{f(X)\}$ and

$$G(\mathbf{x}) = G(x_1, \dots, x_n) = \sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n \check{f}(x_i) \right|.$$

Concentration around mean: Let $\mathbf{y} = (y_1, \dots, y_n)^\top$ and $\mathbf{x}^{\sim k} = (x_1^{\sim k}, \dots, x_n^{\sim k})^\top$ with $x_i^{\sim k} = x_i$ for $i \neq k$ and $x_k^{\sim k} = y_k$ for $1 \leq i, k \leq n$. Note that

$$\begin{aligned} \left| n^{-1} \sum_{i=1}^n \check{f}(x_i) \right| - G(\mathbf{x}^{\sim k}) &= \left| n^{-1} \sum_{i=1}^n \check{f}(x_i) \right| - \sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n \check{f}(x_i^{\sim k}) \right| \\ &\leq \left| n^{-1} \sum_{i=1}^n \check{f}(x_i) \right| - \left| n^{-1} \sum_{i=1}^n \check{f}(x_i^{\sim k}) \right| \\ &\leq \left| n^{-1} \sum_{i=1}^n \{ \check{f}(x_i) - \check{f}(x_i^{\sim k}) \} \right| \\ &= n^{-1} |f(x_k) - f(y_k)| \\ &\leq 2bn^{-1}, \end{aligned}$$

by taking supremum over $f \in \mathcal{F}$ one has that $G(\mathbf{x}) \leq G(\mathbf{x}^{\sim k}) + 2bn^{-1}$. Since the same argument may be applied with the roles of x and y reversed, one can conclude that $|G(\mathbf{x}) - G(\mathbf{x}^{\sim k})| \leq 2bn^{-1}$, which means that the function $G(\cdot)$ satisfies the bounded difference property with parameters $(2bn^{-1}, \dots, 2bn^{-1})$ for each k . By using bounded differences inequality, one has that for any $t > 0$,

$$\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \leq \mathbb{E}(\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}}) + t$$

with \mathbb{P} -probability at least $1 - e^{-\frac{nt^2}{2b^2}}$.

Upper bound on mean: Let $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ be a copy of $\mathbf{X} = (X_1, \dots, X_n)^\top$. Then

one has that

$$\begin{aligned}
\mathbb{E}(\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}}) &= \mathbb{E}_{\mathbf{X}} \left[\sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n f(X_i) - n^{-1} \sum_{i=1}^n \mathbb{E}_{Y_i} \{f(Y_i)\} \right| \right] \\
&= \mathbb{E}_{\mathbf{X}} \left[\sup_{f \in \mathcal{F}} \left| \mathbb{E}_{\mathbf{Y}} \left[n^{-1} \sum_{i=1}^n \{f(X_i) - f(Y_i)\} \right] \right| \right] \\
&\leq \mathbb{E}_{\mathbf{X}, \mathbf{Y}} \left[\sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n \{f(X_i) - f(Y_i)\} \right| \right] \\
&= \mathbb{E}_{\mathbf{X}, \mathbf{Y}, \varepsilon} \left[\sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n \varepsilon_i \{f(X_i) - f(Y_i)\} \right| \right] \\
&\leq 2\mathbb{E}_{\mathbf{X}, \varepsilon} \left\{ \sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n \varepsilon_i f(X_i) \right| \right\} = 2\mathcal{R}_n(\mathcal{F}).
\end{aligned}$$

Then the argument can be obtained by using

$$\left\{ \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \leq \mathbb{E}(\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}}) + t \right\} \subseteq \left\{ \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \leq 2\mathcal{R}_n(\mathcal{F}) + t \right\}.$$

□

Proof of Proposition 5. By Jensen's inequality, one has that

$$\begin{aligned}
\mathbb{E}_{\mathbf{X}, \varepsilon} \left\{ \phi(\|S_n\|_{\check{\mathcal{F}}}/2) \right\} &= \mathbb{E}_{\mathbf{X}, \varepsilon} \left\{ \phi \left(2^{-1} \sup_{\check{f} \in \check{\mathcal{F}}} \left| n^{-1} \sum_{i=1}^n \varepsilon_i \check{f}(X_i) \right| \right) \right\} \\
&= \mathbb{E}_{\mathbf{X}, \varepsilon} \left\{ \phi \left(2^{-1} \sup_{f \in \mathcal{F}} \left| \mathbb{E}_{\mathbf{Y}} \left[n^{-1} \sum_{i=1}^n \varepsilon_i \{f(X_i) - f(Y_i)\} \right] \right| \right) \right\} \\
&\leq \mathbb{E}_{\mathbf{X}, \mathbf{Y}, \varepsilon} \left\{ \phi \left(2^{-1} \sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n \varepsilon_i \{f(X_i) - f(Y_i)\} \right| \right) \right\} \\
&= \mathbb{E}_{\mathbf{X}, \mathbf{Y}} \left\{ \underbrace{\phi \left(2^{-1} \sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n \{f(X_i) - f(Y_i)\} \right| \right)}_{T_1} \right\}.
\end{aligned}$$

Note that

$$T_1 \leq 2^{-1} \sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n f(X_i) - \mathbb{E}\{f(X)\} \right| + 2^{-1} \sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n f(Y_i) - \mathbb{E}\{f(Y)\} \right|,$$

by the convexity of ϕ one has that

$$\begin{aligned}
\phi(T_1) &\leq \phi \left(2^{-1} \sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n f(X_i) - \mathbb{E}\{f(X)\} \right| + 2^{-1} \sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n f(Y_i) - \mathbb{E}\{f(Y)\} \right| \right) \\
&\leq 2^{-1} \phi \left(\sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n f(X_i) - \mathbb{E}\{f(X)\} \right| \right) + 2^{-1} \phi \left(\sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n f(Y_i) - \mathbb{E}\{f(Y)\} \right| \right).
\end{aligned}$$

Then by taking expectations, one has that

$$\begin{aligned}\mathbb{E}_{\mathbf{X},\epsilon}\left\{\phi(\|S_n\|_{\tilde{\mathcal{F}}}/2)\right\} &\leq 2^{-1}\mathbb{E}_{\mathbf{X}}\left\{\phi\left(\sup_{f\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^n f(X_i) - \mathbb{E}\{f(X)\}\right|\right)\right\} \\ &\quad + 2^{-1}\mathbb{E}_{\mathbf{Y}}\left\{\phi\left(\sup_{f\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^n f(Y_i) - \mathbb{E}\{f(Y)\}\right|\right)\right\} \\ &= \mathbb{E}_{\mathbf{X}}\left\{\phi(\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}})\right\}.\end{aligned}$$

The first inequality holds.

Note that

$$\begin{aligned}\mathbb{E}_{\mathbf{X}}\left\{\phi(\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}})\right\} &= \mathbb{E}_{\mathbf{X}}\left\{\phi\left(\sup_{f\in\mathcal{F}}\left|\mathbb{E}_{\mathbf{Y}}\left[n^{-1}\sum_{i=1}^n \{f(X_i) - f(Y_i)\}\right]\right|\right)\right\} \\ &\leq \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left\{\phi\left(\sup_{f\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^n \{f(X_i) - f(Y_i)\}\right|\right)\right\} \\ &= \mathbb{E}_{\mathbf{X},\mathbf{Y},\epsilon}\left\{\phi\left(\underbrace{\sup_{f\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^n \varepsilon_i \{f(X_i) - f(Y_i)\}\right|}_{T_2}\right)\right\},\end{aligned}$$

by the convexity of ϕ one has that

$$\phi(T_2) = \phi(2T_2/2) \leq 2^{-1}\phi\left(2\sup_{f\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^n \varepsilon_i f(X_i)\right|\right) + 2^{-1}\phi\left(2\sup_{f\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^n \varepsilon_i f(Y_i)\right|\right).$$

Then by taking expectations, one has that

$$\begin{aligned}\mathbb{E}_{\mathbf{X}}\left\{\phi(\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}})\right\} &\leq 2^{-1}\mathbb{E}_{\mathbf{X},\epsilon}\left\{\phi\left(2\sup_{f\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^n \varepsilon_i f(X_i)\right|\right)\right\} \\ &\quad + 2^{-1}\mathbb{E}_{\mathbf{Y},\epsilon}\left\{\phi\left(2\sup_{f\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^n \varepsilon_i f(Y_i)\right|\right)\right\} \\ &= \mathbb{E}_{\mathbf{X},\epsilon}\left\{\phi(2\|S_n\|_{\mathcal{F}})\right\}.\end{aligned}$$

The second inequality holds. □

Proof of Proposition 8. Let $\tilde{\mathcal{F}} = \mathcal{F} \cup (-\mathcal{F})$ where $-\mathcal{F} = \{-f : f \in \mathcal{F}\}$, then one has that

$$\sup_{f\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^n \varepsilon_i f(x_i)\right| = \sup_{\tilde{f}\in\tilde{\mathcal{F}}}\left\{n^{-1}\sum_{i=1}^n \varepsilon_i \tilde{f}(x_i)\right\}$$

and

$$\text{Card}(\tilde{\mathcal{F}}(x_1^n)) \leq \text{Card}(\mathcal{F}(x_1^n)) + \text{Card}(-\mathcal{F}(x_1^n)) \leq 2(n+1)^\nu.$$

Let

$$\mathcal{A} = \{(a_1, \dots, a_n)^\top : a_i = n^{-1}\tilde{f}(x_i), \tilde{f} \in \tilde{\mathcal{F}}\},$$

then one has

$$\sup_{\tilde{f} \in \tilde{\mathcal{F}}} \left\{ n^{-1} \sum_{i=1}^n \varepsilon_i \tilde{f}(x_i) \right\} = \sup_{\mathbf{a} \in \mathcal{A}} \sum_{i=1}^n a_i \varepsilon_i = \sup_{\mathbf{a} \in \mathcal{A}} \langle \mathbf{a}, \boldsymbol{\varepsilon} \rangle.$$

For each $\mathbf{a} \in \mathcal{A}$, $a_i \varepsilon_i$ is a bounded random variable on $[-|a_i|, |a_i|]$ which means that $a_i \varepsilon_i$ is sub-Gaussian with parameter $|a_i|$, one has that for any $\lambda > 0$,

$$\mathbb{E}(e^{\lambda \langle \mathbf{a}, \boldsymbol{\varepsilon} \rangle}) = \prod_{i=1}^n \mathbb{E}(e^{\lambda a_i \varepsilon_i}) \leq e^{\lambda^2 \|\mathbf{a}\|_2^2} \leq e^{\lambda^2 D^2(x_1^n)/n}.$$

Note that $\text{Card}(\mathcal{A}) = \text{Card}(\tilde{\mathcal{F}}(x_1^n)) \leq 2(n+1)^\nu$, by using Jensen's inequality, one has that

$$\begin{aligned} \exp \left\{ \lambda \mathbb{E} \left(\sup_{\mathbf{a} \in \mathcal{A}} \langle \mathbf{a}, \boldsymbol{\varepsilon} \rangle \right) \right\} &\leq \mathbb{E}(e^{\lambda \sup_{\mathbf{a} \in \mathcal{A}} \langle \mathbf{a}, \boldsymbol{\varepsilon} \rangle}) \\ &= \mathbb{E} \left(\sup_{\mathbf{a} \in \mathcal{A}} e^{\lambda \langle \mathbf{a}, \boldsymbol{\varepsilon} \rangle} \right) \\ &\leq \text{Card}(\mathcal{A}) e^{\lambda^2 D^2(x_1^n)/n} \\ &\leq 2(n+1)^\nu e^{\lambda^2 D^2(x_1^n)/n}. \end{aligned}$$

By taking logarithm, one has that

$$\begin{aligned} \mathbb{E} \left(\sup_{\mathbf{a} \in \mathcal{A}} \langle \mathbf{a}, \boldsymbol{\varepsilon} \rangle \right) &\leq \{ \log 2 + \nu \log(n+1) \} \lambda^{-1} + \lambda n^{-1} D^2(x_1^n) \\ &\leq 4\nu \log(n+1) \lambda^{-1} + \lambda n^{-1} D^2(x_1^n). \end{aligned}$$

By taking $\lambda = 2\sqrt{n\nu \log(n+1)}/D(x_1^n)$, one has that

$$\mathbb{E} \left(\sup_{\mathbf{a} \in \mathcal{A}} \langle \mathbf{a}, \boldsymbol{\varepsilon} \rangle \right) \leq 4D(x_1^n) \sqrt{\nu \log(n+1)/n}.$$

□

Proof of Theorem 9. Denote

$$\mathcal{F} = \{ \mathbb{I}_{(-\infty, x]}(\cdot) : x \in \mathbb{R} \}.$$

We only need to show that $\text{Card}(\mathcal{F}(x_1^n)) \leq n+1$. By ordering all points in x_1^n by $x_{(1)} \leq \dots \leq x_{(n)}$, then \mathbb{R} is divided by $n+1$ intervals: $(-\infty, x_{(1)})$, $\{[x_{(i)}, x_{(i+1)})\}_{i=1}^{n-1}$ and $[x_{(n)}, \infty)$. For given $x \in \mathbb{R}$, $\mathbb{I}_{(-\infty, x]}(\cdot)$ takes the value 1 for all $x_{(i)} \leq x$ and the value 0 for all other samples. One has that $\text{Card}(\mathcal{F}(x_1^n)) \leq n+1$. □