EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Mathematisch-Naturwissenschaftliche Fakultät

Fachbereich Mathematik

Prof. Dr. Andreas Prohl
Dr. Abhishek Chaudhary

Statistical Learning 2

Summer-Semester 2023 **Tübingen, 14.06.2023**

Homework 5

Let

$$\mathbf{Z} \text{ and } \mathbf{Z}_1, \dots, \mathbf{Z}_n \quad \text{be i.i.d. } \mathbb{R}^d\text{-valued random variables, and}$$
$$\mathcal{G} \quad \text{be a set of functions } g : \mathbb{R}^d \to \mathbb{R}, \text{ with } \mathbb{E}\big[g(\mathbf{Z})\big] < \infty.$$

This exercise sheet is on the property of $\mathcal{G}$ to satisfy the *uniform law of large numbers (ULLN)*, *i.e.*,

$$\sup_{g \in \mathcal{G}} \Big| \frac{1}{n} \sum_{i=1}^n g(\mathbf{Z}_i) - \mathbb{E}\big[g(\mathbf{Z})\big] \Big| \overset{n\uparrow\infty}{\longrightarrow} 0 \qquad \mathbb{P}\text{-a.s.}$$

The following theorem elaborates sufficient criteria for $\mathcal{G}$ to satisfy ULLN. It uses the envelope $G(\mathbf{x}) = \sup_{g\in\mathcal{G}} \big|g(\mathbf{x})\big|$, where $\mathbf{x} \in \mathbb{R}^d$.

Theorem 1. Let $\mathcal{G}$ be given, and G be its envelope. Assume

$$\mathbb{E}\big[G(\mathbf{Z})\big] < \infty \qquad \text{and} \qquad V_{\mathcal{G}^+} < \infty\,.$$

Then ULLN holds for $\mathcal{G}$.

Problem 1. Prove Theorem 1.

<u>Hints</u>: 1) Fix $L > 0$, and consider the set of truncated functions $\mathcal{G}_L := \big\{g \cdot \mathbb{1}_{\{G \leq L\}};\, g \in \mathcal{G}\big\}$. Now estimate the relevant term as follows:

$$\begin{aligned}
\Big|\frac{1}{n}\sum_{i=1}^n g(\mathbf{Z}_i) - \mathbb{E}\big[g(\mathbf{Z})\big]\Big| \quad \leq \quad & \overbrace{\Big|\frac{1}{n}\sum_{i=1}^n g(\mathbf{Z}_i) - \frac{1}{n}\sum_{i=1}^n g(\mathbf{Z}_i)\mathbb{1}_{\{G(\mathbf{Z}_i)\leq L\}}\Big|}^{\leq \frac{1}{n}\sum_{i=1}^n G(\mathbf{Z}_i)\mathbb{1}_{\{G(\mathbf{Z}_i)>L\}} =: \mathtt{I}_n} \\
& + \underbrace{\Big|\frac{1}{n}\sum_{i=1}^n g(\mathbf{Z}_i)\mathbb{1}_{\{G(\mathbf{Z}_i)\leq L\}} - \mathbb{E}\big[g(\mathbf{Z})\mathbb{1}_{\{G(\mathbf{Z})\leq L\}}\big]\Big|}_{:=\mathtt{II}_n} + \underbrace{\Big|\mathbb{E}\big[g(\mathbf{Z})\mathbb{1}_{\{G(\mathbf{Z})\leq L\}}\big] - \mathbb{E}\big[g(\mathbf{Z})\big]\Big|}_{\leq \mathbb{E}[G(\mathbf{Z})\mathbb{1}_{\{G(\mathbf{Z})>L\}}] =: \mathtt{III}_L}\,.
\end{aligned}$$

Now show that all terms tend to zero independently, for $n, L \uparrow \infty$.

2) a) For $\{\mathtt{I}_n\}_n$ we may argue with the law of large numbers, since only the envelope G is involved.
b) For $\{\mathtt{III}_L\}_L$ we may use the monotone convergence theorem (why)?

c) For the last term, reason why it suffices to verify that

$$\mathtt{II}_n \leq \sup_{g \in \mathcal{G}_L} \Big| \frac{1}{n} \sum_{i=1}^{n} g(\mathbf{Z}_i) - \mathbb{E}\big[g(\mathbf{Z})\big] \Big| \overset{n \uparrow 0}{\longrightarrow} 0 \qquad \mathbb{P}\text{-a.s.}$$

Therefore, start with Pollard's *uniform exponential inequality*, and use the results from the lecture to bring it in a form where $V_{\mathcal{G}_L^+}$ enters. Now **assume** the following estimate:

$$V_{\mathcal{G}_L^+} \leq V_{\mathcal{G}^+} \,.$$

Date of Submission: 12.00 on 21.06.2023.