Radamacher complexity and Uniform Laws of Large Numbers

Advisor: Ambrus Tamás

Béla Szlovicsák

December 17, 2023

During this semester I have read the first few chapters of [1] and [2] to gain a deeper understanding of concentration inequalities and their different uses. I familiarised myself with many different techniques, such as entropy and martingale methods. To illustrate these methods, I present here another broad way of utilising concentration inequalities and their utility in non-asymptotic results. These results are widely used in statistical learning as the problems of regression or classification can be formulated with the goal of minimising the empirical risk, in which case radamacher complexity can be used to upper bound the population risk.

1 Generalisation of Glivenko-Cantelli

Definition 1.1. *The **empirical CDF**, belonging to a collection of samples $X_1...X_n$ drawn i.i.d. from the same distribution is given by*

$$\hat{F}_n(t) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,t]}(X_i),$$

where $\mathbb{1}_{(-\infty,t]}(x)$ is the indicator function.

Theorem 1.1 (Glivenko-Cantelli)**.** *For any distribution, the empirical CDF $\hat{F}_n(t)$ is a strongly consistent estimator of the population CDF in the uniform norm, that is*

$$||\hat{F}_n - F||_\infty \xrightarrow{a.s.} 0.$$

This result is quite useful for many settings where one wishes to estimate some functional of the CDF, such as the Quantile or the Expected Value functional. However one might wish to see if widening the scope of this theorem might provide us with more applicable tools. This turns out to be the case. Consider the following Definition:

Definition 1.2. *Let $\mathcal{F}$ be a class of integrable real-valued functions over the domain $\mathcal{X}$, and let $X_1...X_n$ be a collection of i.i.d. random variables taking values in $\mathcal{X}$, drawn according to the same distribution $\mathbb{P}$. We then use the notation:*

$$||\mathbb{P}_n - \mathbb{P}||_{\mathcal{F}} := \sup_{f\in\mathcal{F}} \Big|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X_i)]\Big|.$$

*We say that $\mathcal{F}$ is a **Glivenko-Cantelli** class for $\mathbb{P}$ if $||\mathbb{P}_n - \mathbb{P}||_{\mathcal{F}}$ converges to zero in probability as $n \longrightarrow \infty$.*

Here the naming convention makes sense, as for any series of i.i.d. random variables the function class $\mathcal{F} = \{\mathbb{1}_{(-\infty,t]} \mid t \in \mathbb{R}\}$ gives us back the Glivenko-Cantelli Theorem. One ubiquitous class of examples where this notion proves to be quite useful is estimation problems, where we want to check the goodnes of estimation in terms of some loss function. Using this loss functions one can define the empirical and the population risk for the given estimand. We can minimize the empirical risk, but we are unsure whether it approaches the population risk. Luckily their difference can be upper bounded by the random variable $||\mathbb{P}_n - \mathbb{P}||_{\mathcal{F}}$ where the function class of the loss functions is used. Therefore it becomes evident that for examining estimators using empirical risk minimalization one needs to check wheter the class of loss functions used is a Glivenko-Cantelli class for any given $\mathbb{P}$. That is one needs to establish a uniform law of large numbers for this function class.

2 Results using Radamacher Complexity

Definition 2.1. *Given $\mathcal{F}$, a class of real-valued functions, a collection of real numbers $\mathbf{x} := (x_1, ..., x_n)$ and a collection of i.i.d. Radamacher random variables $(\epsilon_1...\epsilon_n)$, that is $\mathbb{P}[\epsilon_i = -1] = \mathbb{P}[\epsilon_i = 1] = \frac{1}{2}$. Then the **empirical Radamacher complexity** is given by applying $\mathcal{F}$ to $\mathbf{x}$:*

$$\mathcal{R}(\mathcal{F}(\mathbf{x})/n) := \mathbb{E}_\epsilon[\sup_{f\in\mathcal{F}} \Big|\frac{1}{n}\sum_{i=1}^{n} \epsilon_i f(x_i)\Big|]$$

Definition 2.2. *Given a class of real-valued functions $\mathcal{F}$, as before and $\mathbf{X} := (X_1, ..., X_n)$ a collection of i.i.d. samples, the empirical Radamacher complexity $\mathcal{R}(\mathcal{F}(\mathbf{X})/n)$ is a random variable. The expectation of this random variable is the* ***Radamacher complexity*** *of the function class $\mathcal{F}$, given by*

$$\mathcal{R}_n(\mathcal{F}) := \mathbb{E}_{\mathbf{X},\epsilon}[\sup_{f\in\mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i f(X_i) \right|].$$

This deterministic quantity measures, in a sense, the correlation between $(f(X_1), ..., f(X_n))$ and the "noise vector" of the Radamacher random variables. Meaning that a function class will be too large to be useful, when we can always find a function that gives a high correlation with the Radamacher variables. In the case however when this quantity tends to 0 as n tends to ∞, the function class does not contain functions that correlate highly with random noise. The following theorem formalises this notion and gives a sufficient condition for a function class to be Glivenko-Cantelli. It apllies to classes of b-uniformly bounded functions, meaning $||f||_\infty \leq b$.

Theorem 2.1. *For any b-uniformly bounded class of functions $\mathcal{F}$, any positive integer $n \geq 1$ and any real $\delta \geq 0$, we have*

$$||\mathbb{P}_n - \mathbb{P}||_\mathcal{F} \leq 2\mathcal{R}_n(\mathcal{F}) + \delta,$$

with probability at least $1 - \exp(-\frac{n\delta^2}{2b^2})$. Consequently, if $\mathcal{R}_n(\mathcal{F}) = o(1)$, we have $||\mathbb{P}_n - \mathbb{P}||_\mathcal{F} \xrightarrow{a.s.} 0$.

This theorem serves as a uniform law of large numbers characterized by the Radamacher complexity of the given function class. This result follows from a concentration inequality around the expected value, known as McDiarmid's inequality and a small lemma. We show these necessary results here and refer the reader to [2] for a full proof of the theorems without the proof of Lemma 2.1.

Theorem 2.2 (McDiarmid's inequality)**.** *Suppose that for the function $f : \mathbb{R}^n \to \mathbb{R}$ we have for all $k \in \{1...n\}$*

$$\sup_{x_i, x_j \in \mathbb{R}} |f(x_1, ..., x_{k-1}, x_i, x_{k+1}, ..., x_n) - f(x_1, ..., x_{k-1}, x_j, x_{k+1}, ..., x_n)| \leq L_k,$$

and that the random vector $\mathbf{X} := (X_1, ..., X_n)$ has independent components. Than we have

$$\mathbb{P}[|f(X) - \mathbb{E}[f(X)]| \ \geq \ t] \leq 2\exp(-\frac{2t^2}{\sum_{k=1}^{n} L_k^2}).$$

The property defined for the function is known as the bounded difference property, and the theorem is also known as the Bounded Differences inequality. For a proof of this theorem see [2].

Lemma 2.1. *For a given function class $\mathcal{G}$ and a random variable X we have*

$$\sup_{g\in\mathcal{G}} \mathbb{E}[g(X)] \leq \mathbb{E}[\sup_{g\in\mathcal{G}} |g(X)|]$$

Definition 2.3. *A class of functions $\mathcal{F}$ over a domain $\mathcal{X}$ has* ***polynomial discrimination*** *of order $\nu \geq 1$ if for each positive integer n and collection $\mathbf{x} = \{x_1, ..., x_n\}$ the cardinality of the set $\mathcal{F}(\mathbf{x})$ is upper bounded as*

$$card(\mathcal{F}(\mathbf{x})) \leq (n+1)^\nu$$

Lemma 2.2. *Suppose $\mathcal{F}$ has polynomial discrimination of order ν and that it is b uniformly bounded. Then for all positive integers n we have*

$$\mathcal{R}_n(\mathcal{F}) \leq 4b\sqrt{\frac{\nu \log(n+1)}{n}}$$

For indicator functions one can easily see that both $b = 1$ and $\nu = 1$. Therefore this lemma combined with Theorem 2.1. implies the classical Glivenko-Cantelli theorem as a corollary.

Corollary 2.1 (Glivenko-Cantelli)**.** *Using our earlier notation we have for all $\delta > 0$*

$$\mathbb{P}[||\hat{F}_n - F||_\infty \geq 8\sqrt{\frac{\log(n+1)}{n}} + \delta] \leq e^{\frac{-n\delta^2}{2}}$$

from which $||\hat{F}_n - F||_\infty \xrightarrow{a.s.} 0$ follows.

References

[1] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities. A nonasymptotic theory of independence*. Oxford: Clarendon Press, 2012.

[2] Martin J. Wainwright. *High-Dimensional Statistics. A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 2019.