

## Lecture 3: Uniform concentration inequality

Lecturer: Ben Dai

*“There is Nothing More Practical Than A Good Theory.”*

— Kurt Lewin

## 1 Introduction

As indicated in Lecture 2, we will focus on the asymptotics of the empirical process of the estimation error, that is, try to find a  $\delta_n \rightarrow 0$  for any small  $\varepsilon > 0$ ,

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)| \geq \varepsilon\right) \leq \delta_n$$

Motivated by **concentration**, how a random variable deviates from its expectation, rewrite the probability as:

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)| - \mathbb{E} \sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)| \geq \varepsilon - \mathbb{E} \sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)|\right) \leq \delta_n.$$

To investigate this bound, we itemize two aims:

- **A1.** The asymptotics of

$$\mathbb{E} \sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)|.$$

- **A2.** The concentration inequality of

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)| - \mathbb{E} \sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)| \geq \varepsilon\right).$$

For **A1**, the minimum requirement is that

$$\mathbb{E} \sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)| = o(1),$$

to ensure asymptotically vanishing of (the upper bound of) the estimation error. For **A2**, we can regard it as a uniform version of concentration inequalities.

## 2 From pointwise to uniform

### 2.1 From Hoeffding's inequality to McDiarmid's inequality

**Theorem 2.1** (Hoeffding's Inequality). *Suppose  $Z_1, \dots, Z_n$  are independent random variables such that  $a_i \leq Z_i \leq b_i$  almost surely, then for any  $\varepsilon > 0$ :*

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n Z_i - \frac{1}{n} \sum_{i=1}^n \mathbb{E}(Z_i) \geq \varepsilon\right) \leq \exp\left(\frac{-2n^2 \varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Note that we can not use the Hoeffding's inequality of bound **A2**, since there is a supremum on the average. McDiarmid's inequality is a general form of Hoeffding's inequality, which enables us to directly bound the probabilistic bound in **A2**.

**Theorem 2.2** (McDiarmid's inequality). *Suppose  $Z_1, \dots, Z_n$  are independent random variables, and there is a function  $g: \mathbb{R}^n \rightarrow \mathbb{R}$  such that the variation on  $i$ -th coordinate is upper bounded, that is, for all  $i = 1, \dots, n$  and all  $(z_1, \dots, z_i, z'_i, \dots, z_n)$ ,*

$$|g(z_1, \dots, z_i, \dots, z_n) - g(z_1, \dots, z'_i, \dots, z_n)| \leq c_i.$$

Then,

$$\mathbb{P}(g(Z_1, \dots, Z_n) - \mathbb{E}g(Z_1, \dots, Z_n) \geq \varepsilon) \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n c_i^2}\right).$$

The idea of McDiarmid's inequality is quite similar to Hoeffding's inequality, yet it uses the boundness of the overall function  $g(Z_1, \dots, Z_n)$ . We demonstrate the McDiarmid's inequality for our Aim **A2**.

Let  $Z_i = l(\mathbf{Y}_i, f(\mathbf{X}_i))$ , and

$$g(Z_1, \dots, Z_n) = \sup_{f \in \mathcal{F}} (\widehat{R}_n(f) - R(f)) = \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n (Z_i - \mathbb{E}l(\mathbf{Y}, f(\mathbf{X}))).$$

Assume that  $0 \leq l(\mathbf{Y}_i, f(\mathbf{X}_i)) \leq U$ , we have

$$|g(z_1, \dots, z_i, \dots, z_n) - g(z_1, \dots, z'_i, \dots, z_n)| \leq \left| \sup_{f \in \mathcal{F}} \frac{1}{n} (z_i - z'_i) \right| = U/n.$$

Then, McDiarmid's inequality yields that

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)| - \mathbb{E} \sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)| \geq \varepsilon\right) \leq \exp\left(-\frac{2n\varepsilon^2}{U^2}\right).$$

We summarize the result as the following corollary.

**Corollary 2.3.** *For a loss function  $l(\cdot, \cdot)$  uniformly bounded by a constant  $U$ , then for any  $\varepsilon > 0$ ,*

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)| - \mathbb{E} \sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)| \geq \varepsilon\right) \leq \exp\left(-\frac{2n\varepsilon^2}{U^2}\right).$$

*Remark 2.4.* The information used in Hoeffding's inequality and McDiarmid's inequality: *boundness of the loss function.*

## 2.2 From Bernstein's inequality to Talagrand's inequality

Hoeffding's inequality does not use any information about the randomness of random variables. Bernstein's inequality is a sharper inequality to consider the *variance of the random variable*.

**Theorem 2.5** (Bernstein's inequality). *Let  $Z_1, \dots, Z_n$  be independent random variables with  $|Z_i| \leq U$  almost surely, for all  $i = 1, \dots, n$ . Then, for all  $\varepsilon > 0$ ,*

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n Z_i - \frac{1}{n} \sum_{i=1}^n \mathbb{E}(Z_i) \geq \varepsilon\right) \leq \exp\left(-\frac{n\varepsilon^2}{2\sigma^2 + 2U\varepsilon/3}\right),$$

where  $\sigma^2 = \frac{1}{n} \sum_{i=1}^n \text{Var}(Z_i)$ .

The uniform Bernstein's inequality is a much harder problem which was solved by *Talagrand* [Talagrand, 1996b, Talagrand, 1996a].

**Theorem 2.6** (Talagrand's inequality). *For a loss function  $l(\cdot, \cdot)$  uniformly bounded by a constant  $U$ , the for  $\varepsilon > 0$ ,*

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)| - \mathbb{E} \sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)| \geq \varepsilon\right) \leq K \exp\left(-\frac{1}{K} \frac{n\varepsilon}{U} \log\left(1 + \frac{\varepsilon U}{V}\right)\right),$$

where  $K > 0$  is a universal constant and  $V$  is any constant satisfying

$$V \geq \mathbb{E} \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \left(l(\mathbf{Y}_i, f(\mathbf{X}_i)) - \mathbb{E}l(\mathbf{Y}_i, f(\mathbf{X}_i))\right)^2.$$

The constant  $V$  is analog to the variance in Bernstein's inequality. However, find a tight constant  $V$  to bound the "variance" of the functional space is not easy. Now, given the results of Talagrand's inequality, we slight modify our aims:

- **A1**. The asymptotics of

$$\mathbb{E} \sup_{f \in \mathcal{F}} |\widehat{R}_n(f) - R(f)|.$$

- **A2'**. Find a tight constant  $V$  such that

$$V \geq \mathbb{E} \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \left(l(\mathbf{Y}_i, f(\mathbf{X}_i)) - \mathbb{E}l(\mathbf{Y}_i, f(\mathbf{X}_i))\right)^2.$$

In the sequel, we will show that **A1** and **A2'** are crossed in the same direction.

## References

[Talagrand, 1996a] Talagrand, M. (1996a). New concentration inequalities in product spaces. *Inventiones mathematicae*, 126(3):505–563.

[Talagrand, 1996b] Talagrand, M. (1996b). A new look at independence. *The Annals of probability*, pages 1–34.