MATH888: High-dimensional probability and statistics Fall 2021

Lecture 17 — October 15, 2021

Sebastien Roch, UW-Madison *Scribe: Jenny Wei, Sebastien Roch*

1 Overview

In the last lecture we have (a) derived the concentration bound of the least square estimator of linear regression model, and (b) investigated a more general case: non-linear models.

In this lecture we will explore the behavior of the suprema of a random processes defined on some index set T. To see why this is useful in solving high dimensional problems, we can see the following example. When studying the behavior of the 2-norm of some random matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$, we can view the 2-norm as the suprema of the following random processes:

$$X_{\boldsymbol{x}} = \|\mathbf{A}\boldsymbol{x}\|, \ \boldsymbol{x} \in \mathbb{S}^{m-1}$$

That is:

$$\|\mathbf{A}\|_2 = \max_{\boldsymbol{x}\in\mathbb{S}^{m-1}} X_{\boldsymbol{x}} = \max_{\boldsymbol{x}\in\mathbb{S}^{m-1}} \|\mathbf{A}\boldsymbol{x}\|.$$

Remark 1. *This lecture aligns closely with section 5.1 and 5.2 in [2].*

2 Suprema of Random Process with Finite Index Set

2.1 Extreme Value Theory with i.i.d Random variables

We begin with finding the maximum of a random process $(X_t)_{t\in T}$ whose index set T is finite.

When $X_1, \ldots, X_n$ are i.i.d random variables with cumulative distribution function (CDF) $F(x) = \mathbb{P}(X_i \le x)$ for $x \in \mathbb{R}$, the problem becomes simple because we can compute the CDF of the finite maxima $M_n = \max_{1\le i\le n} X_i$ directly:

$$\mathbb{P}(M_n \le x) = (F(x))^n$$

In particular, for specific examples with polynomial and exponential tails, the following hold as $n \to \infty$.

Theorem 2 (See e.g. Exercise 3.2.2 in [1])**.** *1. If $F(x) = 1 - x^{-\alpha}$, for $x \ge 1$ and $\alpha > 0$, then*

$$\mathbb{P}\left(\frac{M_n}{n^{1/\alpha}} \le y\right) \to \exp\left(-y^{-x}\right).$$

2. If $F(x) = 1 - e^{-x}$ for $x \ge 0$, then

$$\mathbb{P}(M_n - \log n \le y) \to \exp\left(-e^{-y}\right).$$

2.2 Extreme Value Theory with Finite Step Sub-Gaussian Random Process

Let $(X_t)_{t\in T}$ be a random process where T is an arbitrary index set, and $(X_t)_{t\in T}$ need not to be i.i.d random variables.

Example 1. *Let* $\mathbf{A}$ *be a random matrix in* $\mathbb{R}^{p\times p}$ *and* $T = \mathbb{S}^{p-1}$*, the* p*-dimensional unit ball. In this case, we can consider* $\|\mathbf{A}\|_2$ *as the suprema of a random processes* $X_{\boldsymbol{u}} = \|\mathbf{A}\boldsymbol{u}\|_2$ *for all* $\boldsymbol{u} \in \mathbb{S}^{p-1}$*.*

2.2.1 Naive Bound

Asuming we have a finite index set T. How can one bound the maximum of a finite set of random variables? The most naive approach is to bound the supremum by a sum:

$$\sup_{t\in T} X_t \le \sum_{t\in T} |X_t|$$

Using this inequality, we could get the following bound:

$$\begin{aligned}
\mathbb{E}\left(\sup_{t\in T} X_t\right) &\le \mathbb{E}\left(\sup_{t\in T} |X_t|\right) \\
&\le \mathbb{E}\left(\sum_{t\in T} |X_t|\right) \\
&\le \sum_{t\in T} \mathbb{E}\,|X_t| \\
&\le |T| \sup_{t\in T} \mathbb{E}\,|X_t|
\end{aligned}$$

This indicates that if we could control the magnitude of every individual random variable X_t, we can get a bound that grows linearly in the cardinality $|T|$. Extending this bound a little bit, by the Jensen's Inequality, we get that for $p \ge 1$,

$$\begin{aligned}
\mathbb{E}\left(\sup_{t\in T} X_t\right) &\le \mathbb{E}\left(\sup_{t\in T} |X_t|^p\right)^{1/p} \\
&\le |T|^{1/p} \sup_{t\in T} \left(\mathbb{E}\,|X_t|^p\right)^{1/p}
\end{aligned}$$

Thus if the random variables X_t have bounded p-th moment, the dependence on $|T|$ for this naive bound can be improved to $|T|^{1/p}$.

2.2.2 Maximal Inequality for Sub-Gaussian Processes

In the naive bound, we did not make any assumptions for each random variable X_t. In this class, we are mostly interested in sub-Gaussian random variables. So, for sub-Gaussian random process $(X_t)_{t\in T}$, the following theorem provides us a way to bound its suprema. One could look for a general version of this following theorem in Lemma 5.1 of [2].

Theorem 3 (Maximal Inequality)**.** *Let T be a finite index set. $(X_t)_{t\in T}$ is a random process where for any $t \in T$, X_t has zero mean and $\|X_t\|_{\psi_2}^2 \le \sigma^2$. Then,*

$$\mathbb{E}\left(\sup_{t\in T} X_t\right) \le \sqrt{2C\sigma^2 \log |T|}$$

Proof. By Jensen's Inequality, for any $\lambda > 0$, we have

$$\begin{aligned}
\mathbb{E}\left(\sup_{t\in T} X_t\right) &\le \frac{1}{\lambda}\log \mathbb{E}\left(e^{\lambda \sup_{t\in T} X_t}\right) \\
&\le \frac{1}{\lambda}\log \sum_{t\in T} \mathbb{E}\left(e^{\lambda X_t}\right) \\
&\le \frac{1}{\lambda}\log \sum_{t\in T} \mathbb{E}\left(e^{\frac{\lambda^2 C\sigma^2}{2}}\right) \\
&= \frac{1}{\lambda}\log \left(|T| e^{\frac{\lambda^2 C\sigma^2}{2}}\right) \\
&= \frac{\log |T|}{\lambda} + \frac{\lambda^2 C\sigma^2}{2}
\end{aligned}$$

Now optimize over λ, we get the desired bound. □

Exercise 4 (Maximal Tail Inequality, Lemma 5.2 of [2])**.** *Show that*

$$\mathbb{P}\left(\sup_{t\in T} X_t \ge \sqrt{2C\sigma^2 \log |T|} + x\right) \le e^{-x^2/2\sigma^2} \quad \textit{for all } x \ge 0.$$

Hint. Use Markov's inequality and proceed as above.

3 Towards understanding Suprema of Random Process with Infinite Index Set

Before we consider random processes defined on an infinite index set, we need to introduce a couple of tools first.

Definition 5 (ϵ-net, Definition 5.5 in [2])**.** *Let (T, d) be a metric space, $\epsilon > 0$ and a set $K \subseteq T$. A subset $N \subseteq K$ is an ϵ-net of K if for $\forall \boldsymbol{x} \in K$, $\exists \boldsymbol{x}_0 \in N$ such that $d(x, x_0) \le \epsilon$. Equivalently, N is an ϵ-net of K if and only if K can be covered by balls with centers in N and radius ϵ. See Figure 1 (Figure 4.1 in [3]).*

Definition 6 (Lipschitz process, Definition 5.4 in [2])**.** *The random process $(X_t)_{t\in T}$ is Lipschitz for a metric d on T if there exists a random variable L such that for all $\boldsymbol{t}, \boldsymbol{s} \in T$,*

$$|X_{\boldsymbol{t}} - X_{\boldsymbol{s}}| \le L\, d(\boldsymbol{t}, \boldsymbol{s}) \quad a.s.$$

Example 2 (Example of a Lipschitz process)**.** *Consider again the first example, where $X_{\boldsymbol{u}} = \|\mathbf{A}\boldsymbol{u}\|_2$, for all $u \in \mathbb{S}^{p-1}$. Let the metric d be the Euclidean distance $d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\|_2$. So following the definition of Lipschitz process, we will find a random variable L so that*

$$|X_{\boldsymbol{u}} - X_{\boldsymbol{v}}| = |\|\mathbf{A}\boldsymbol{u}\|_2 - \|\mathbf{A}\boldsymbol{v}\|_2| \le L\|\boldsymbol{u} - \boldsymbol{v}\|_2.$$

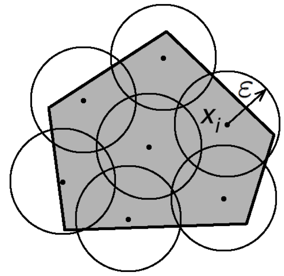

Figure 1: This covering of a pentagon K by 7 ϵ-balls shows that $|N| = 7$

Note that

$$\begin{aligned}\|\mathbf{A}\boldsymbol{u}\|_2 &= \|\mathbf{A}\left(\boldsymbol{u}-\boldsymbol{v}+\boldsymbol{v}\right)\|_2\\ &\leq \|\mathbf{A}\boldsymbol{v}\|_2 + \|\mathbf{A}\left(\boldsymbol{u}-\boldsymbol{v}\right)\|_2\\ &\leq \|\mathbf{A}\boldsymbol{v}\|_2 + \|\mathbf{A}\|_2\,\|\boldsymbol{u}-\boldsymbol{v}\|_2.\end{aligned}$$

Hence, we obtain that $L = \|\mathbf{A}\|_2$.

References

[1] Rick Durrett, *Probability theory and examples (fifth edition)*, Cambridge University Press, 2019.

[2] Ramon van Handel, *APC 550: Probability in High Dimension*, Lecture Notes, 2016. `https://web.math.princeton.edu/~rvan/APC550.pdf`

[3] Roman Vershynin, *High-dimensional probability: An introduction with applications in data science*, Cambridge University Press, 2018.