IDS.160 – Mathematical Statistics: A Non-Asymptotic Approach

Lecturer: P Rigollet
Scribe: Hussein Mozannar and Arnab Sarker

Lecture 5
Feb. 20, 2020

Goals: Develop tools to characterize the behavior of the maximum of a set of random variables.

In the last two lectures we covered Bernstein's and Hoeffding's inequalities, which provide concentration inequalities on the average of independent random variables $\bar{X}_n$, and can be generally extended to linear combinations of random variables. However, we may often be concerned with the maximum of a set of random variables, for example in the study of empirical risk minimization. In this lecture, we will turn our focus on maximal inequalities to upper bound the maximum of a collection of random variables which may not necessarily be independent.

1. MAXIMUM OVER A FINITE SET

Our first problem will be to consider the maximum over a finite collection of random variables $X_1, \cdots, X_N$ which may *not necessarily be independent*.

As a first attempt, we may introduce the following inequalities:

$$\max_{1\le 1\le N} X_i \le \max_{1\le i\le N} |X_i| \le \sum_{i=1}^{N} |X_i| \,.$$

Taking expectation on both sides of the above inequality we obtain

$$\mathbb{E}\left[\max_{1\le i\le N} X_i\right] \le N \max_{1\le i\le N} \mathbb{E}\left[|X_i|\right] \,.$$

The bound above has a linear dependence on N, the size of our collection, but our analysis can be refined by considering the p-norm of the random variables.

$$\begin{aligned}
\mathbb{E}\left[\max_{1\le i\le N} X_i\right] &\le \mathbb{E}\left[\left(\max_{1\le i\le N} |X_i|^p\right)^{\frac{1}{p}}\right] && (\forall\ p \ge 1)\\
&\le \left(\mathbb{E}\left[\max_{1\le i\le N} |X_i|^p\right]\right)^{\frac{1}{p}} && \text{(Jensen's Inequality)}\\
&\le \left(\sum_{i=1}^{N} \mathbb{E}\left[|X_i|^p\right]\right)^{\frac{1}{p}}\\
&\le N^{\frac{1}{p}} \max_{1\le i\le N} \|X_i\|_p
\end{aligned}$$

We now have a polynomial dependence on N; however, we might suffer through the p-norm of X_i. As a specific example when the random variables have finite p-norms for all $p \ge 1$,

consider the case where $X_i \sim \mathsf{subG}(\sigma^2)$ for all i. Then, setting $p = \log(N)$ we obtain

$$\begin{aligned}\mathbb{E}\left[\max_{1\le i\le N} X_i\right] &\le N^{\frac{1}{\log(N)}} \max_{1\le i\le N} \|X_i\|_{\log(N)} \\ &\le ec\sigma\sqrt{\log(N)}\,, \qquad \left(N^{\frac{1}{\log(N)}} = e^{\frac{\log N}{\log N}} = e\right)\end{aligned}$$

For some constant c as discussed in Lecture 2. The following theorem makes the above result more precise, and refines the constant factor.

Theorem: Let $X_1, \dots, X_N$ be N random variables (not necessarily independent) such that $X_i \sim \mathsf{subG}(\sigma^2)$.

Then

$$\mathbb{E}\left[\max_{1\le i\le N} X_i\right] \le \sigma\sqrt{2\log(N)}\,, \quad \text{and} \quad \mathbb{E}\left[\max_{1\le i\le N} |X_i|\right] \le \sigma\sqrt{2\log(2N)}\,.$$

Moreover, for any $t > 0$,

$$\mathbb{P}\left(\max_{1\le i\le N} X_i > t\right) \le Ne^{-\frac{t^2}{2\sigma^2}}\,, \quad \text{and} \quad \mathbb{P}\left(\max_{1\le i\le N} |X_i| > t\right) \le 2Ne^{-\frac{t^2}{2\sigma^2}}\,.$$

Proof. For any $s > 0$,

$$\begin{aligned}\mathbb{E}\left[\max_{1\le i\le N} X_i\right] &\le \frac{1}{s}\mathbb{E}\left[\log e^{s\max_{1\le i\le N} X_i}\right] \\ &\le \frac{1}{s}\log\mathbb{E}\left[e^{s\max_{1\le i\le N} X_i}\right] && \text{(By Jensen's Inequality)} \\ &= \frac{1}{s}\log\mathbb{E}\left[\max_{1\le i\le N} e^{sX_i}\right] \\ &\le \frac{1}{s}\log\sum_{i=1}^{N}\mathbb{E}\left[e^{sX_i}\right] \\ &\le \frac{1}{s}\log\sum_{i=1}^{N} e^{\frac{\sigma^2 s^2}{2}} && (X_i \sim \mathrm{subG}(\sigma^2)) \\ &= \frac{\log(N)}{s} + \frac{\sigma^2 s}{2}\,.\end{aligned}$$

To obtain the tightest bound we minimize over $s > 0$ the RHS of the above bound, since it is convex we can set the derivative with respect to s to 0 and get $s = \sqrt{2\log(N)/\sigma^2}$.

For the high probability bound,

$$\begin{aligned}\mathbb{P}\left(\max_{1\le i\le N} X_i > t\right) &= \mathbb{P}\left(\bigcup_{i=1}^{N}\{X_i > t\}\right) \\ &\le \sum_{i=1}^{N}\mathbb{P}(X_i > t) && \text{(Union Bound)} \\ &\le Ne^{-\frac{t^2}{2\sigma^2}}\,.\end{aligned}$$

To prove the inequalities involving the absolute values of $|X_i|$, define $\tilde{X}_i = X_i$ and $\tilde{X}_{N+i} = -X_i$ for $i = 1, \cdots, N$, then note that:

$$\max_{1 \le i \le 2N} \tilde{X}_i = \max_{1 \le i \le N} |X_i|$$

This new collection is of course *not independent* but since we did not require independence in our proof then we can apply the results we just proved above to a collection of $2N$ random variables which amounts to replacing N by $2N$ in all the bounds.

□

One might be interested in studying the maximum or supremum over an infinite collection of random variables; however, the following simple example shows that such results may not always generalize to an infinite collection of random variables.

Let $X_1, \cdots, X_N \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma^2)$. Then, for any $N \ge 1$, and any $t > 0$,

$$\begin{aligned} \mathbb{P}\left(\max_{1 \le i \le N} X_i > t \right) &= 1 - \mathbb{P}\left(\max_{1 \le i \le N} X_i \le t \right) \\ &= 1 - \mathbb{P}(X_1 \le t)^N \to 1 \quad (N \to \infty)\,. \end{aligned}$$

Therefore, in this setting, the maximum of an infinite set of random variables is unbounded almost surely. On the other hand, if $X_1 = X_2 = \cdots = X_N$, i.e., all the random variables are equal to the same random variable X_1, then we have for any $t > 0$,

$$\mathbb{P}\left(\max_{1 \le i \le N} X_i > t \right) = \mathbb{P}(X_1 > t) < 1\,, \quad \forall N \ge 1\,.$$

This simple example illustrates that if an infinite collection of random variables has a certain structure, then their maximum may in fact be finite. The following sections review other examples where we can exploit the structure of the set of random variables to analyze the behavior of the maximum of the set.

2. MAXIMUM OVER A CONVEX POLYTOPE

We now turn our attention to an uncountably infinite set of random variables. Specifically, given a random vector X taking values in $\mathbb{R}^d$, we seek to understand the set of random variables $\{\theta^\top X \mid \theta \in P\}$ where $P \subset \mathbb{R}^d$. We first consider the case in which P is a *convex polytope*, and provide a general result regarding maximization over a convex polytope.

Definition (Convex Polytope): A convex polytope P is a compact set with a finite number of vertices, $\mathcal{V}(P)$ (also called extreme points), such that $P = \mathsf{conv}\left(\mathcal{V}(P)\right)$

In the definition above, conv refers to the *convex hull* of a set of points. Formally,

$$\mathsf{conv}\left(\{v_1, \ldots, v_k\}\right) = \left\{ \sum_{i=1}^k \lambda_i v_i \;\middle|\; \lambda_i \ge 0 \;\forall i,\; \sum_{i=1}^k \lambda_i = 1 \right\}$$

We next show the following lemma regarding maximization of a linear function over a convex polytope.

Lemma: For any $c \in \mathbb{R}^d$, and any convex polytope P with extreme points $\mathcal{V}(P)$, it holds

$$\max_{x \in P} c^\top x = \max_{x \in \mathcal{V}(P)} c^\top x$$

Proof. First, note that since $\mathcal{V}(P) \subseteq P$, we must have:

$$\max_{x \in P} c^\top x \geq \max_{x \in \mathcal{V}(P)} c^\top x \,.$$

Next, fix some $x \in P$, and let $\mathcal{V}(P) = \{v_1, \ldots, v_N\}$. By definition of a convex polytope, we may write $x = \sum_{i=1}^N \lambda_i v_i$ for some non-negative values λ_i such that $\sum_{i=1}^N \lambda_i = 1$. Therefore, we may write

$$c^\top x = \sum_{i=1}^N \lambda_i c^\top v_i \leq \left(\max_{1 \leq i \leq N} c^\top v_i \right) \sum_{i=1}^N \lambda_i = \max_{1 \leq i \leq N} c^\top v_i = \max_{v \in \mathcal{V}(P)} c^\top v$$

Since the above holds for any $x \in P$, we can take the maximum over x in the left-hand side to get

$$\max_{x \in P} c^\top x \leq \max_{x \in \mathcal{V}(P)} c^\top x \,.$$

This completes the proof of the lemma. □

The above lemma leads to the following maximal inequality, as we may consider maximization over a convex polytope to have similar properties as maximization over a finite set.

Corollary: Consider a convex polytope P with N vertices $\mathcal{V}(P) = \{v_1, \ldots, v_N\}$. If for each v_i, we have $v_i^\top X \sim \mathsf{subG}(\sigma^2)$, then

$$\mathbb{E}\left[\max_{\theta \in P} \theta^\top X\right] \leq \sigma\sqrt{2\log(N)}\,, \quad \text{and} \quad \mathbb{E}\left[\max_{\theta \in P} |\theta^\top X|\right] \leq \sigma\sqrt{2\log(2N)}\,.$$

Further, for any $t > 0$,

$$\mathbb{P}\left[\max_{\theta \in P} \theta^\top X > t\right] \leq N e^{-\frac{t^2}{2\sigma^2}}\,, \quad \text{and} \quad \mathbb{P}\left[\max_{\theta \in P} |\theta^\top X| > t\right] \leq 2N e^{-\frac{t^2}{2\sigma^2}}\,.$$

Note that the condition $v^\top X \sim \mathsf{subG}(\sigma^2)$ for all $v \in \mathcal{V}(P)$ is equivalent to $v^\top X \sim \mathsf{subG}(\sigma^2)$ for all $v \in P$.

Proof. From the lemma above, we see that the following two random variables are equivalent:

$$\max_{\theta \in P} \theta^\top X = \max_{\theta \in \mathcal{V}(P)} \theta^\top X$$

Hence, we may apply the theorem from section 2 to the N random variables $v_1^\top X, \ldots, v_N^\top X$. □

The theorem above is most useful when considering polytopes with a small number of vertices. One particular convex polytope of interest is the ℓ_1 ball,

$$\mathcal{B}_1 = \left\{ x = (x^{(1)}, \ldots, x^{(d)})^\top \in \mathbb{R}^d,\ |x|_1 := \sum_{i=1}^d |x^{(i)}| \leq 1 \right\}.$$

Here, $x^{(i)}$ denotes the ith element of the vector x. The ℓ_1 ball has vertices at each of the standard basis vectors e_i (defined as the vector with a 1 in the ith position and 0's elsewhere) and their negations $-e_i$, for a total of $2d$ vertices.

3. MAXIMUM OVER THE EUCLIDEAN BALL

We now consider the case in which θ belongs to the set of vectors in the Euclidean ball,

$$\mathcal{B}_2 = \left\{ x \in \mathbb{R}^d,\ |x|_2^2 := \sum_{i=1}^d |x^{(i)}|^2 \leq 1 \right\}.$$

The Euclidean ball $\mathcal{B}_2$ is not a polytope, as has an infinite number of extreme points. However, we have that[1] $\mathcal{B}_2 \subset \sqrt{d}\mathcal{B}_1$.

Next, observe that for any $x \in \mathcal{B}_2$,

$$\begin{aligned}
\sum_{i=1}^d |x_i| &\leq \sqrt{\left(\sum_{i=1}^d |x^{(i)}|^2\right)\left(\sum_{i=1}^d 1^2\right)} && \text{(Cauchy-Schwarz inequality)} \\
&= |x|_2\sqrt{d} \\
&\leq \sqrt{d} && (x \in \mathcal{B}_2)
\end{aligned}$$

Therefore $x \in \sqrt{d}\mathcal{B}_1$ and this completes the proof that $\mathcal{B}_2 \subset \sqrt{d}\mathcal{B}_1$. Hence,

$$\max_{\theta \in \mathcal{B}_2} \theta^\top X \leq \sqrt{d} \max_{\theta \in \mathcal{B}_1} \theta^\top X = \sqrt{d} \max_{1 \leq i \leq d} |X^{(i)}|.$$

Therefore, if $X^{(i)} \sim \mathsf{subG}(\sigma^2)$ for all i, then

$$\mathbb{E}\left[\max_{\theta \in \mathcal{B}_2} \theta^\top X\right] \leq \sqrt{d}\mathbb{E}\left[\max_{1 \leq i \leq d} |X^{(i)}|\right] \leq \sigma\sqrt{2d\log(2d)}.$$

However, we can refine our analysis and remove the dependence on $\log d$. This will require the notion of a ε-net (covering).

Definition: Fix $K \subset \mathbb{R}^d$ and $\varepsilon > 0$. A set $\mathcal{N}$ is called an ε-net of K with respect to a distance $d(\cdot, \cdot)$ on $\mathbb{R}^d$, if $\mathcal{N} \subset K$ and for any $z \in K$, there exists $x \in \mathcal{N}$ such that $d(x, z) \leq \varepsilon$

If $\mathcal{N}$ is an ε-net of K with respect to a distance $d(\cdot, \cdot)$, then every point of K is at distance at most ε from a point in $\mathcal{N}$. Note that K is trivially an ε-net of K with respect to any distance. Moreover, every compact set admits a finite ε-net by definition.

We will be interested in ε-nets of small size and the following lemma gives an upper bound on the size of the smallest ε-net of $\mathcal{B}_2$.

[1] we use the notation that if $A \subset R^d$, then for $a \in \mathbb{R}$, and $b \in \mathbb{R}^d$, $aA + b$ is defined as the set $\{ax + b \mid x \in A\}$.

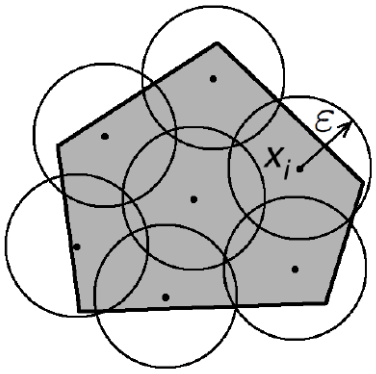

Figure 1: An example of an ε-net of size 7 with respect to the Euclidean norm for a pentagon. Figure from [Ver18]

Lemma: Fix $\varepsilon \in (0,1)$. The unit Euclidean ball $\mathcal{B}_2$ admits an ε-net $\mathcal{N}$ with respect to the Euclidean distance such that:

$$|\mathcal{N}| \leq \left(1 + \frac{2}{\varepsilon}\right)^d \leq \left(\frac{3}{\varepsilon}\right)^d$$

Proof. The following proof is an example of a "Volume argument" where we consider the ratio of the volume of the covering to the size of the set we are trying to cover. We show the proof by constructing an ε-net of $\mathcal{B}_2$ with bounded size.
Consider the following iterative construction of the ε-net:

1. Set $x_1 = 0$
2. For $i \geq 2$, take x_i to be any $x \in \mathcal{B}_2 \setminus \bigcup_{j=1}^{i-1}\{\varepsilon\mathcal{B}_2 + x_j\}$.
   If no such x_i exists, then output $\mathcal{N} = \{x_1, \cdots, x_{i-1}\}$ as the ε-net.

Clearly this procedure will create an ε-net. We now compute its size.

Observe that for any $x, y \in \mathcal{N}$, $|x - y|_2 > \varepsilon$. Hence, the set of Euclidean balls centered at x_j with radius $\varepsilon/2$ for $j = 1, \cdots, |\mathcal{N}|$ are disjoint. Their total volume is

$$\text{vol}\left(\bigcup_{j=1}^{|\mathcal{N}|}\left\{\frac{\varepsilon}{2}\mathcal{B}_2 + x_j\right\}\right) = \sum_{j=1}^{|\mathcal{N}|}\text{vol}\left(\left\{\frac{\varepsilon}{2}\mathcal{B}_2\right\}\right) = |\mathcal{N}|\left(\frac{\varepsilon}{2}\right)^d \text{vol}(\mathcal{B}_2)\,.$$

Moreover,

$$\bigcup_{j=1}^{|\mathcal{N}|}\{\frac{\varepsilon}{2}\mathcal{B}_2 + x_j\} \subset (1 + \frac{\varepsilon}{2})\mathcal{B}_2\,.$$

This is justified as the farthest point $x \in \mathcal{N}$ could lie from the origin is on the surface of $\mathcal{B}_2$, and hence the collection of balls with center $x \in \mathcal{N}$ with radius $\epsilon/2$ lie inside the enlarged

$(1+\frac{\varepsilon}{2})\mathcal{B}_2$. Translating the above set relation in terms of volumes we obtain:

$$|\mathcal{N}|\left(\frac{\varepsilon}{2}\right)^d \mathrm{vol}(\mathcal{B}_2) \leq \left(1+\frac{\varepsilon}{2}\right)^d \mathrm{vol}(\mathcal{B}_2)$$
$$\iff |\mathcal{N}| \leq \left(1+\frac{2}{\varepsilon}\right)^d .$$

Then, since $\varepsilon \in (0,1)$, $\left(1+\frac{2}{\varepsilon}\right)^d \leq \left(\frac{3}{\varepsilon}\right)^d$, completing the proof. □

Theorem: Assume that for any $u \in \mathbb{R}^d$, we have $u^\top X \sim \mathsf{subG}(\sigma^2|u|_2^2)$. Then,

$$\mathbb{E}\left[\max_{\theta\in\mathcal{B}_2} \theta^\top X\right] = \mathbb{E}\left[\max_{\theta\in\mathcal{B}_2} |\theta^\top X|\right] \leq 4\sigma\sqrt{d}$$

Moreover, for any $\delta > 0$, with probability $1-\delta$, it holds

$$\max_{\theta\in\mathcal{B}_2} \theta^\top X = \max_{\theta\in\mathcal{B}_2} \left|\theta^\top X\right| \leq 4\sigma\sqrt{d} + 2\sigma\sqrt{2\log(1/\delta)}$$

Proof. Let $\mathcal{N}$ be a 1/2-net of $\mathcal{B}_2$ with respect to the Euclidean norm obtained from the construction in the previous Lemma. We have $|\mathcal{N}| \leq 5^d$. For every $\theta \in \mathcal{B}_2$, there exists $z \in \mathcal{N}$ and x such that $|x|_2 \leq 1/2$ and $\theta = z + x$. Therefore,

$$\max_{\theta\in\mathcal{B}_2} \theta^\top X \leq \max_{z\in\mathcal{N}} z^\top X + \max_{x\in\frac{1}{2}B_2} x^\top X$$

the rightmost term on the RHS is nothing but

$$\max_{x\in\frac{1}{2}B_2} x^\top X = \frac{1}{2}\max_{\theta\in B_2} \theta^\top X$$

Combining the inequalities above and referring to the theorem in section 2 on the maximum of a finite collection of random variables, we obtain:

$$\mathbb{E}\left[\max_{\theta\in\mathcal{B}_2} \theta^\top X\right] \leq 2\mathbb{E}\left[\max_{z\in\mathcal{N}} z^\top X\right] \leq 2\sigma\sqrt{2\log(|\mathcal{N}|)} \leq 2\sigma\sqrt{2\log(5)d} \leq 4\sigma\sqrt{d}\,.$$

For the high probability bound,

$$\mathbb{P}\left(\max_{\theta\in\mathcal{B}_2} \theta^\top X > t\right) \leq \mathbb{P}\left(2\max_{z\in\mathcal{N}} z^\top X > t\right) \leq |\mathcal{N}| e^{-\frac{t^2}{8\sigma^2}} \leq 5^d e^{-\frac{t^2}{8\sigma^2}}\,.$$

Setting the RHS equal to δ yields

$$t \geq 2\sqrt{2\log(5)}\sigma\sqrt{d} + 2\sigma\sqrt{2\log(1/\delta)}$$

which is sufficient thus completing the proof. □

3.1 Application: Operator Norm of a Random Matrix

Consider a random matrix $A = (A_{ij})_{\substack{1 \le i \le m \\ 1 \le j \le n}}$, where each entry $A_{ij} \overset{\text{i.i.d}}{\sim} \mathsf{subG}(\sigma^2)$. We wish to analyze the *operator norm* of the random matrix,

$$\|A\|_{\text{op}} = \max_{x \in \mathcal{B}_2(\mathbb{R}^n)} |Ax|_2 = \max_{x \in \mathcal{B}_2(\mathbb{R}^n)} \max_{y \in \mathcal{B}_2(\mathbb{R}^m)} y^\top A x \,.$$

Proposition: For the random matrix A defined above,

$$\mathbb{E}\left[\|A\|_{\text{op}}\right] \le 2\sigma\sqrt{2(n+m)\log 9}\,.$$

Proof. Let $\mathcal{N}_n$ and $\mathcal{N}_m$ be $\frac{1}{4}$-nets of $\mathcal{B}_2(\mathbb{R}^n)$ and $\mathcal{B}_2(\mathbb{R}^m)$, respectively. We may select $\mathcal{N}_n$ and $\mathcal{N}_m$ such that $|\mathcal{N}_n| \le 9^n$, and $|\mathcal{N}_m| \le 9^m$, from the arguments above. Then, since any $x \in \mathcal{B}_2(\mathbb{R}^n)$ may be written as $z + \delta$ for some $z \in \mathcal{N}_n$ and $\delta \in \frac{1}{4}\mathcal{B}_2(\mathbb{R}^n)$,

$$\begin{aligned}
\|A\|_{\text{op}} &= \max_{x \in \mathcal{B}_2(\mathbb{R}^n)} |Ax|_2 \\
&\le \max_{z \in \mathcal{N}_n} |Az|_2 + \max_{\delta \in \frac{1}{4}\mathcal{B}_2(\mathbb{R}^n)} |A\delta|_2 \\
&= \max_{z \in \mathcal{N}_n} |Az|_2 + \frac{1}{4}\|A\|_{\text{op}} \\
&= \max_{z \in \mathcal{N}_n} \max_{y \in \mathcal{B}_2(\mathbb{R}^m)} y^\top A z + \frac{1}{4}\|A\|_{\text{op}} \,. \qquad (3.1)
\end{aligned}$$

Next, for fixed $z \in \mathcal{N}_n$, we similarly note

$$\begin{aligned}
\max_{y \in \mathcal{B}_2(\mathbb{R}^m)} y^\top A z &\le \max_{w \in \mathcal{N}_m} w^\top A z + \max_{\delta \in \frac{1}{4}\mathcal{B}_2(\mathbb{R}^m)} \delta^\top A z \\
&\le \max_{w \in \mathcal{N}_m} w^\top A z + \frac{1}{4}\|A\|_{\text{op}} \,.
\end{aligned}$$

Combining the above inequality with (3.1) and rearranging, we get

$$\|A\|_{\text{op}} \le 2 \max_{z \in \mathcal{N}_n} \max_{w \in \mathcal{N}_m} w^\top A z \,.$$

Further, we note that each random variable $w^\top A z$ can be written as $\sum_{i,j} w^{(i)} A_{ij} z^{(j)}$, which means that each random variable $w^\top A z \sim \mathsf{subG}(\sigma^2 |w|_2^2 |z|_2^2)$ as each A_{ij} is independent. Since w and z lie within the Euclidean ball, their norms are at most one, and we see that $\|A\|_{\text{op}}$ is bounded by the maximum of 9^{n+m} random variables which are sub-Gaussian with variance proxy σ^2. Therefore,

$$\begin{aligned}
\mathbb{E}\left[\|A\|_{\text{op}}\right] &\le 2\mathbb{E}\left[\max_{z \in \mathcal{N}_n} \max_{w \in \mathcal{N}_m} w^\top A z\right] \\
&\le 2\sigma\sqrt{2\log(9^{n+m})} \\
&= 2\sigma\sqrt{2(n+m)\log 9}\,,
\end{aligned}$$

completing the proof. □

Exercise. Provide a high probability bound on the size of $\|A\|_{\text{op}}$.

Summary: In this lecture, we developed tools for characterizing the behavior of the maximum of a set of variables. We began by considering the case in which the set of random variables is finite, and showed that the expectation of the maximum of N random variables is bounded by a term on the order of $\sqrt{\log(N)}$.

We then considered the case in which the set of random variables is characterized by a convex polytope, and found a similar bound– when the convex polytope has N vertices, the expected maximum of associated random variables again is bounded by a term on the order of $\sqrt{\log(N)}$.

Finally, we considered the case when the set of random variables is characterized by the Euclidean ball in $\mathbb{R}^d$, and found the expected maximum of associated random variables to be bounded by a term on the order of $\sqrt{d}$ – we then extended the reasoning to characterize the operator norm of a random matrix with independent sub-Gaussian entries.

References

[Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.