Lecture 4: Maximal Inequalities

Isabella Zhu

13 February 2025

§1 TLDR

We will look at bounds involving $\max_i X_i$.

§2 Maximum Over Finite Set

Let $X_1, X_2, \dots X_n$ be r.v., not necessarily independent.

§2.0.1 Attempt 1

$$\mathbb{E}\left[\max X_i\right] \leq \mathbb{E}\left[\sum |X_i|\right] \leq \sigma n.$$

This bound blows up very fast.

§2.0.2 Attempt 2

For every $p \geq 1$, we have

$$\mathbb{E}[|X_i|^p]^{1/p} \leq \sigma\sqrt{p}$$

By Jensen, we have

$$\mathbb{E}[\max|X_i|] \leq \left(\mathbb{E}[\max_i |X_i|^p]\right)^{1/p} \leq \left(\mathbb{E}[\sum_i |X_i|^p]\right)^{1/p} \leq \sigma n^{1/p}\sqrt{p}.$$

Then we pick p to make this bound as small as possible. If p is on the order of $\log n$, then we get a much tighter bound of

$$\mathbb{E}[\max|X_i|] \leq e\sigma\sqrt{\log n}.$$

Proposition 2.1

Let $X_1, X_2, \ldots X_n$ be subgaussian with proxy σ^2, *not necessarily i.i.d.* Then, we have

$$\mathbb{E}[\max X_i] \leq \sigma\sqrt{2\log n}, \ \ \mathbb{E}[\max X_i] \leq \sigma\sqrt{2\log 2n}$$

Furthermore, we have tail bounds

$$\mathbb{P}(\max X_i \geq t) \leq ne^{-t^2/2\sigma^2}, \ \ \mathbb{P}(\max X_i \geq t) \leq 2ne^{-t^2/2\sigma^2}.$$

We can rewrite this as

$$\max X_i \leq \sigma\sqrt{2\log(N/\delta)} \text{ with probability at least } 1-\delta.$$

Proof. We have

$$\begin{aligned}
\mathbb{E}[\max X_i] &= \frac{1}{s}\log\exp(s\mathbb{E}[\max X_i]) \\
&\leq \frac{1}{s}\log\mathbb{E}[\exp(s\max X_i)] \quad \textit{(Jensen)} \\
&\leq \frac{1}{s}\log\sum_{i=1}^{n}\mathbb{E}[e^{sX_i}] \\
&\leq \frac{1}{s}\log ne^{s^2\sigma^2/2} \\
&= \frac{\log n}{s} + \frac{s\sigma^2}{2}
\end{aligned}$$

and then we take the optimal s that produces a minimal bound of $\sigma\sqrt{2\log n}$.

For the tail bound, we have

$$\mathbb{P}(\max X_i > t) \leq \sum_{i=1}^{N}\mathbb{P}(X_i > t) \leq Ne^{-t^2/2\sigma^2}.$$

These bounds are tight. To see this, just take $X_1, X_2, \ldots X_n \sim N(0, \sigma^2)$ and compute.

Remark 2.2. The more dependence structure there is between the X_is, the stricter bounds we can impose on the maximum of X_i.

§3 Maximum Over Convex Polytope

Definition 3.1. A **convex polytope** is defined as

$$P = \{v \in \mathbb{R}^d, \sum_{i=1}^{N}\lambda_i v_i \mid \lambda_i. \geq 0, \sum\lambda_i = 1\}$$

$V(P) = \{v_1, v_2, \ldots v_n\}$ are the vertices of P. Convex hull of finite number of points.

We care about $\max_{v\in P} v^T X$, where $X \in \mathbb{R}^d$ is a random vector. Clearly, this set has infinite size, so we can't directly apply the results from the previous section.

Remark 3.2. One of the key ideas in this section is to reduce infinite sets to finite sets.

Lemma 3.3

For every $x \in \mathbb{R}^d$, convex polytope P, then

$$\max_{v \in P} v^T x = \max_{v \in V(P)} v^T x$$

Proof. For any $v \in P$, we decompose v as weighted average of v_is.

$$v^T x = \sum_{i=1}^{N} \lambda_i v_i^T x \leq \max_{v \in V(P)} v^T x.$$

Theorem 3.4

If P is a convex polytope with vertices $v_1, v_2, \ldots v_n$. If $v_j^T x$ are subgaussian with variance proxy σ^2, then

$$\mathbb{E}[\max_{v \in P} v^T x] = \mathbb{E}[\max v_j^T x] \leq \sigma \sqrt{2 \log N}.$$

Proof. Follows directly from lemma and results in section 2.

§3.1 Examples

Example 3.5

Some common polytopes:

1. $B_\infty = [0,1]^d$. We have $N = 2^d$, so bound is $\sigma\sqrt{2d}$.
2. $B_1 = \{X \mid \sum_{i=1}^n |X_i| \leq 1\}$. We have $N = 2d$, so bound is $\sigma\sqrt{2 \log 2d}$.

§4 Maximum Over Euclidean Ball

Definition 4.1. B_2 is the **Euclidean ball** such that

$$\{X \mid |X|_2^2 = \sum_{i=1}^{d} X_i^2 \leq 1\}$$

We have containment

$$B_2 \subset \sqrt{d} B_1$$

by Cauchy Schwarz.

We care about finding $\mathbb{E}[\max_{v \in B_2} v^T X]$. We will reduce to finite max by discretization.

Definition 4.2. Fix $K \subset \mathbb{R}^d$ and $\epsilon > 0$. A set N is an ϵ-net for K w.r.t $d(.,.)$ if

1. $N \subset K$.
2. For every point $z \in K$, there exists $x \in N$ such that $d(z, x) < \epsilon$.

We want to build the smallest ϵ-net.

Lemma 4.3

Fix $\epsilon \in (0, 1)$. Then B_2 admits N, ϵ-net of size at most

$$|N| \le (1 + \frac{2}{\epsilon})^d \le (\frac{3}{\epsilon})^d.$$

Proof. We will construct N sequentially. Just keep picking points outside balls to center a new ball at until all points are covered.

Let U_j be the ball $x_j + \frac{\epsilon}{2}B_2$. Note that none of the U_j can intersect. Thus, we have

$$Vol\left(\bigcup U_j\right) = \sum Vol(U_j) = (\frac{\epsilon}{2})^d |N| Vol(B_2).$$

Consider the enlargement ball $(1 + \frac{\epsilon}{2})B_2$. All small balls are fully contained in this big ball. So,

$$(1 + \frac{\epsilon}{2})^d Vol(B_2) \ge Vol\left(\bigcup U_j\right) = \sum Vol(U_j) = (\frac{\epsilon}{2})^d |N| Vol(B_2).$$

and we get that

$$|N| \le \left(1 + \frac{2}{\epsilon}\right)^d \le \left(\frac{3}{\epsilon}\right)^d.$$

Theorem 4.4

If for every $v \in B_2$, $v^T X$ is subgaussian with variance proxy σ^2, then

$$\mathbb{E}[\max_{v \in B_2} v^T X] \le 4\sigma\sqrt{d}$$

and

$$\max_{v \in B_2} v^T X \le 4\sigma\sqrt{d} + 2\sigma\sqrt{2\log(1/\delta)} \text{ with probability at least } 1 - \delta.$$

Proof. Let N be a $\frac{1}{2}$-net for B_2. Then, $|N| \le 5^d$ by above lemma. For any $v \in B_2$, there exists $z \in N$ such that $v = z + \delta$ and $|\delta| \le 1/2$.

Then, we have

$$\max_{v \in B_2} v^T X \le \max_{z \in N} z^T X + \max_{|\delta| \le 1/2} \delta^T X$$

However, note that

$$\max_{|\delta| \le 1/2} \delta^T X = \max_{v \in B_2} v^T X.$$

So, we get

$$\max_{v \in B_2} v^T X \leq 2 \max_{z \in N} z^T X \leq 2\sigma \sqrt{2 \log |N|} \leq 2\sigma \sqrt{2d \log 5} \leq 4\sigma \sqrt{d}.$$