The Method of Chaining

Pan Jing Bin

MA5249 Project Presentation Part 1

Outline

1. Introduction
2. Finite Maxima
3. Covering and Packing
4. Wasserstein Law of Large Numbers

Introduction to the Task

Motivation: Given a random process $\{X_t\}_{t \in T}$, we would like to control its **regularity**.

Introduction to the Task

Motivation: Given a random process $\{X_t\}_{t\in T}$, we would like to control its **regularity**.

Main task: Control the regularity of the supremum of a random process $\{X_t\}_{t\in T}$ by establishing **lower bounds** and **upper bounds** on

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \quad \text{and} \quad P\left(\sup_{t\in T} X_t \geq x\right).$$

Introduction to the Task

Motivation: Given a random process $\{X_t\}_{t\in T}$, we would like to control its **regularity**.

Main task: Control the regularity of the supremum of a random process $\{X_t\}_{t\in T}$ by establishing **lower bounds** and **upper bounds** on

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \quad \text{and} \quad P\left(\sup_{t\in T} X_t \geq x\right).$$

Example: Operator norm of random matrices

$$\|M\|_{\mathrm{op}} := \sup_{\|v\|_2=1} \|Mv\|_2$$

Introduction to the Task

Motivation: Given a random process $\{X_t\}_{t\in T}$, we would like to control its **regularity**.

Main task: Control the regularity of the supremum of a random process $\{X_t\}_{t\in T}$ by establishing **lower bounds** and **upper bounds** on

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \quad \text{and} \quad P\left(\sup_{t\in T} X_t \geq x\right).$$

Example: Operator norm of random matrices

$$\|M\|_{\text{op}} := \sup_{\|v\|_2=1} \|Mv\|_2 = \sup_{v\in\mathbb{S}^n} M_v.$$

Introduction to the Task

Motivation: Given a random process $\{X_t\}_{t \in T}$, we would like to control its **regularity**.

Main task: Control the regularity of the supremum of a random process $\{X_t\}_{t \in T}$ by establishing **lower bounds** and **upper bounds** on

$$\mathbb{E}\left[\sup_{t \in T} X_t\right] \quad \text{and} \quad P\left(\sup_{t \in T} X_t \geq x\right).$$

Example: Operator norm of random matrices

$$\|M\|_{\text{op}} := \sup_{\|v\|_2 = 1} \|Mv\|_2 = \sup_{v \in \mathbb{S}^n} M_v.$$

Definition 1.1.

Let $\sigma \in \mathbb{R}_{\geq 0}$. A random variable X is $\boldsymbol{\sigma^2}$**-subgaussian** if for all $\lambda \in \mathbb{R}$, we have

$$\mathbb{E}[e^{\lambda X}] \leq e^{\frac{\lambda^2 \sigma^2}{2}}.$$

Finite Maxima

First step: Control the maximum of finitely many variables $\{X_1, X_2, \cdots, X_k\}$.

Finite Maxima

First step: Control the maximum of finitely many variables $\{X_1, X_2, \cdots, X_k\}$.

Naive approach: Bound the supremum by the sum

$$\max_{t \in T} X_t \leq \sum_{t \in T} |X_t|.$$

Finite Maxima

First step: Control the maximum of finitely many variables $\{X_1, X_2, \cdots, X_k\}$.

Naive approach: Bound the supremum by the sum

$$\max_{t \in T} X_t \leq \sum_{t \in T} |X_t|.$$

Key idea: Use **concavity** to improve the dependence on $|T|$. For example,

$$\mathbb{E}\left[\max_{t \in T} X_t\right] \leq \mathbb{E}\left[\max_{t \in T} |X_t|^2\right]^{1/2} \leq |T|^{1/2} \max_{t \in T} \left\{\mathbb{E}\left[\left|X_t\right|^2\right]^{1/2}\right\}.$$

Finite Maximal Inequality

Lemma 2.1. (Maximal Inequality)

Let $\{X_t\}_{t\in T}$ be a random process. Supoose that X_t is σ^2-subgaussian for each $t \in T$. Then we have

$$\mathbb{E}\left[\max_{t\in T} X_t\right] \leq \sqrt{2\sigma^2 \log|T|}.$$

Finite Maximal Inequality

Lemma 2.1. (Maximal Inequality)

Let $\{X_t\}_{t\in T}$ be a random process. Supoose that X_t is σ^2-subgaussian for each $t \in T$. Then we have

$$\mathbb{E}\left[\max_{t\in T} X_t\right] \leq \sqrt{2\sigma^2 \log|T|}.$$

Proof.

For any $\lambda \in \mathbb{R}_{>0}$, Jensen's inequality gives

$$\mathbb{E}\left[\max_{t\in T} X_t\right] \leq \frac{1}{\lambda}\log\left(\mathbb{E}\left[e^{\lambda \max_{t\in T} X_t}\right]\right) \leq \frac{1}{\lambda}\log\left(\sum_{t\in T}\mathbb{E}\left[e^{\lambda X_t}\right]\right) \leq \frac{\log|T|}{\lambda} + \frac{\lambda\sigma^2}{2}.$$

The conclusion follows by optimizing in λ and choosing $\lambda = \dfrac{\sqrt{2\log|T|}}{\sigma}$. □

Finite Maximal Inequality

Lemma 2.1. (Maximal Inequality)

Let $\{X_t\}_{t\in T}$ be a random process. Supoose that X_t is σ^2-subgaussian for each $t \in T$. Then we have

$$\mathbb{E}\left[\max_{t\in T} X_t\right] \leq \sqrt{2\sigma^2 \log |T|}.$$

Proof.

For any $\lambda \in \mathbb{R}_{>0}$, Jensen's inequality gives

$$\mathbb{E}\left[\max_{t\in T} X_t\right] \leq \frac{1}{\lambda}\log\left(\mathbb{E}\left[e^{\lambda \max_{t\in T} X_t}\right]\right) \leq \frac{1}{\lambda}\log\left(\sum_{t\in T}\mathbb{E}\left[e^{\lambda X_t}\right]\right) \leq \frac{\log|T|}{\lambda} + \frac{\lambda\sigma^2}{2}.$$

The conclusion follows by optimizing in λ and choosing $\lambda = \dfrac{\sqrt{2\log|T|}}{\sigma}$. □

Obtain **elementary** but **workable** bounds.

Generalising to the Infinite Case

To have a meaningful theory when $|T| = +\infty$, we need to exploit the relationship between the variables in $\{X_t\}_{t \in T}$.

Generalising to the Infinite Case

To have a meaningful theory when $|T| = +\infty$, we need to exploit the relationship between the variables in $\{X_t\}_{t \in T}$.

Key idea: Leverage on the **rich structure** of the index set T.

Generalising to the Infinite Case

To have a meaningful theory when $|T| = +\infty$, we need to exploit the relationship between the variables in $\{X_t\}_{t\in T}$.

Key idea: Leverage on the **rich structure** of the index set T.

Example: Homogeneous discrete Markov chain

$$\begin{aligned} &P(X_r = k \mid X_i = x_i, \quad \text{for } i = 0, \cdots, r-1, r+1, \cdots, n) \\ =&P(X_r = k \mid X_{r-1} = x_{r-1}, X_{r+1} = x_{r+1}). \end{aligned}$$

Generalising to the Infinite Case

To have a meaningful theory when $|T| = +\infty$, we need to exploit the relationship between the variables in $\{X_t\}_{t \in T}$.

Key idea: Leverage on the **rich structure** of the index set T.

Example: Homogeneous discrete Markov chain

$$\begin{aligned} &P(X_r = k \mid X_i = x_i, \quad \text{for } i = 0, \cdots, r-1, r+1, \cdots, n) \\ =&P(X_r = k \mid X_{r-1} = x_{r-1}, X_{r+1} = x_{r+1}). \end{aligned}$$

The random variable X_r shares the **strongest dependence** with its **adjacent** neighbours.

Lipschitz Processes and ϵ-nets

To generalise this concept to arbitrary processes, we introduce the notion of a **Lipschitz process**.

Definition 3.1. (Lipschitz process)

A random process $\{X_t\}_{t\in T}$ is **Lipschitz** for a metric d on T if there exists a random variable C such that

$$|X_s - X_t| \leq Cd(s,t) \qquad \text{for all } s,t \in T.$$

Lipschitz Processes and ϵ-nets

To generalise this concept to arbitrary processes, we introduce the notion of a **Lipschitz process**.

Definition 3.1. (Lipschitz process)

A random process $\{X_t\}_{t \in T}$ is **Lipschitz** for a metric d on T if there exists a random variable C such that

$$|X_s - X_t| \leq Cd(s,t) \qquad \text{for all } s, t \in T.$$

Approach: Approximate a Lipschitz process by a **finite** set N, and then estimate N using the inequalities defined in the previous section.

Lipschitz Processes and ϵ-nets

To generalise this concept to arbitrary processes, we introduce the notion of a **Lipschitz process**.

Definition 3.1. (Lipschitz process)

A random process $\{X_t\}_{t \in T}$ is **Lipschitz** for a metric d on T if there exists a random variable C such that

$$|X_s - X_t| \leq C d(s,t) \qquad \text{for all } s,t \in T.$$

Approach: Approximate a Lipschitz process by a **finite** set N, and then estimate N using the inequalities defined in the previous section.

Definition 3.2. (ϵ-net and covering number)

A set N is called an **ϵ-net** for (T,d) if for every $t \in T$, there exists $\pi(t) \in N$ such that $d(t, \pi(t)) \leq \epsilon$. The smallest cardinality of an ϵ-net for (T,d) is called the **covering number**

$$N(T,d,\epsilon) := \inf \left\{ |N| : N \text{ is an } \epsilon\text{-net for } (T,d) \right\}.$$

Lipschitz Maximal Inequality

Lemma 3.3. (Lipschitz maximal inequality)

Let $\{X_t\}_{t \in T}$ be a Lipschitz random process. Suppose that X_t is σ^2-subgaussian for each $t \in T$. Then

$$\mathbb{E}\left[\sup_{t \in T} X_t\right] \le \inf_{\epsilon > 0}\left\{\epsilon \mathbb{E}[C] + \sqrt{2\sigma^2 \log N(T, d, \epsilon)}\right\}.$$

Lipschitz Maximal Inequality

Lemma 3.3. (Lipschitz maximal inequality)

Let $\{X_t\}_{t\in T}$ be a Lipschitz random process. Suppose that X_t is σ^2-subgaussian for each $t \in T$. Then

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \leq \inf_{\epsilon>0}\left\{\epsilon\mathbb{E}[C] + \sqrt{2\sigma^2 \log N(T,d,\epsilon)}\right\}.$$

Proof.

Fix $\epsilon \in \mathbb{R}_{>0}$. Choose an ϵ-net N satisfying $|N| = N(T,d,\epsilon)$ and perform the following decomposition:

$$\sup_{t\in T} X_t \leq \sup_{t\in T}\left\{X_t - X_{\pi(t)}\right\} + \sup_{t\in T} X_{\pi(t)} \leq C\epsilon + \max_{t\in N} X_t.$$

Lipschitz Maximal Inequality

Lemma 3.3. (Lipschitz maximal inequality)

Let $\{X_t\}_{t\in T}$ be a Lipschitz random process. Suppose that X_t is σ^2-subgaussian for each $t \in T$. Then

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \leq \inf_{\epsilon>0}\left\{\epsilon\mathbb{E}[C] + \sqrt{2\sigma^2 \log N(T,d,\epsilon)}\right\}.$$

Proof.

Fix $\epsilon \in \mathbb{R}_{>0}$. Choose an ϵ-net N satisfying $|N| = N(T,d,\epsilon)$ and perform the following decomposition:

$$\sup_{t\in T} X_t \leq \sup_{t\in T}\left\{X_t - X_{\pi(t)}\right\} + \sup_{t\in T} X_{\pi(t)} \leq C\epsilon + \max_{t\in N} X_t.$$

Taking expectation and applying the maximal inequality gives

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \leq \inf_{\epsilon>0}\left\{\epsilon\mathbb{E}[C] + \sqrt{2\sigma^2 \log N(T,d,\epsilon)}\right\}.$$

Wasserstein Law of Large Numbers

Let $X_1, X_2, \cdots$ be i.i.d random variables taking values in the interval $[0,1]$ and let $f : [0,1] \to \mathbb{R}$ be a bounded function. By the law of large numbers,

$$\mathbb{E}\left[\left|\sum_{i=1}^{n} \frac{f(X_i)}{n} - \mu_f\right|\right] \lesssim n^{-1/2}$$

where $\mu_f := \mathbb{E}\big[f(X_1)\big]$.

Wasserstein Law of Large Numbers

Let $X_1, X_2, \cdots$ be i.i.d random variables taking values in the interval $[0,1]$ and let $f : [0,1] \to \mathbb{R}$ be a bounded function. By the law of large numbers,

$$\mathbb{E}\left[\left|\sum_{i=1}^{n} \frac{f(X_i)}{n} - \mu_f\right|\right] \lesssim n^{-1/2}$$

where $\mu_f := \mathbb{E}\big[f(X_1)\big]$.

Question: What is the optimal bound that also is **uniform in f?**

Wasserstein Law of Large Numbers

Let $X_1, X_2, \cdots$ be i.i.d random variables taking values in the interval $[0,1]$ and let $f : [0,1] \to \mathbb{R}$ be a bounded function. By the law of large numbers,

$$\mathbb{E}\left[\left|\sum_{i=1}^{n} \frac{f(X_i)}{n} - \mu_f\right|\right] \lesssim n^{-1/2}$$

where $\mu_f := \mathbb{E}\big[f(X_1)\big]$.

Question: What is the optimal bound that also is **uniform in f?**

Problem setting:

$$X_f := \sum_{i=1}^{n} \frac{f(X_i)}{n} - \mu_f \qquad \text{and} \qquad \mathcal{F} := \big\{f \in \mathrm{Lip}([0,1]) : 0 \le f \le 1\big\}.$$

Goal: Establish an upper bound for the quantity

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} X_f\right].$$

Wasserstein Law of Large Numbers

Two preliminary observations:

1. The process $\{X_f\}_{f \in \mathcal{F}}$ is Lipschitz (with Lipschitz constant 2) with respect to the supremum norm on $\mathcal{F}$.
2. Each X_f is $\dfrac{1}{n}$-subgaussian.

Two preliminary observations:

1. The process $\{X_f\}_{f \in \mathcal{F}}$ is Lipschitz (with Lipschitz constant 2) with respect to the supremum norm on $\mathcal{F}$.
2. Each X_f is $\frac{1}{n}$-subgaussian.

By the Lipschitz maximal inequality,

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} X_f\right] \leq \inf_{\epsilon > 0}\left\{2\epsilon + \sqrt{\frac{2}{n}\log N(\mathcal{F}, \|\cdot\|_\infty, \epsilon)}\right\}.$$

Two preliminary observations:

1. The process $\{X_f\}_{f\in\mathcal{F}}$ is Lipschitz (with Lipschitz constant 2) with respect to the supremum norm on $\mathcal{F}$.
2. Each X_f is $\frac{1}{n}$-subgaussian.

By the Lipschitz maximal inequality,

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} X_f\right] \le \inf_{\epsilon>0}\left\{2\epsilon + \sqrt{\frac{2}{n}\log N(\mathcal{F},\|\cdot\|_\infty,\epsilon)}\right\}.$$

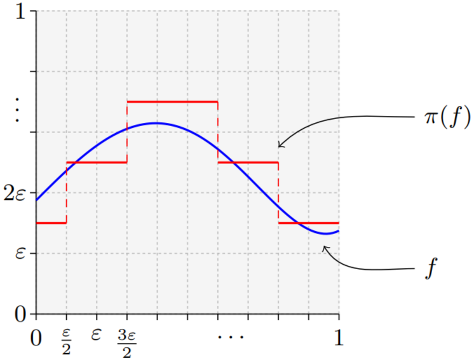
It remains to find a good bound on the **covering number**.

Wasserstein Law of Large Numbers

Lemma 3.8.

There exists a constant $c \in \mathbb{R}$ such that

$$N(\mathcal{F}, \|\cdot\|_\infty, \epsilon) \le e^{c/\epsilon} \text{ for } \epsilon < \frac{1}{2}, \qquad N(\mathcal{F}, \|\cdot\|_\infty, \epsilon) = 1 \text{ for } \epsilon \ge \frac{1}{2}.$$



Source: R. van Handel (2016, p.127)

Wasserstein Law of Large Numbers

Lemma 3.8.

There exists a constant $c \in \mathbb{R}$ such that

$$N(\mathcal{F}, \|\cdot\|_\infty, \epsilon) \leq e^{c/\epsilon} \text{ for } \epsilon < \frac{1}{2}, \qquad N(\mathcal{F}, \|\cdot\|_\infty, \epsilon) = 1 \text{ for } \epsilon \geq \frac{1}{2}.$$

The equation

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} X_f\right] \leq \inf_{\epsilon > 0} \left\{2\epsilon + \sqrt{\frac{2}{n} \log N(\mathcal{F}, \|\cdot\|_\infty, \epsilon)}\right\}.$$

reduces to

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} X_f\right] \leq \inf_{\epsilon > 0} \left\{2\epsilon + \sqrt{\frac{2c}{\epsilon n}}\right\} \lesssim n^{-1/3}$$

Wasserstein Law of Large Numbers

Lemma 3.8.

There exists a constant $c \in \mathbb{R}$ such that

$$N(\mathcal{F}, \|\cdot\|_\infty, \epsilon) \leq e^{c/\epsilon} \text{ for } \epsilon < \frac{1}{2}, \qquad N(\mathcal{F}, \|\cdot\|_\infty, \epsilon) = 1 \text{ for } \epsilon \geq \frac{1}{2}.$$

The equation

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} X_f\right] \leq \inf_{\epsilon > 0} \left\{2\epsilon + \sqrt{\frac{2}{n} \log N(\mathcal{F}, \|\cdot\|_\infty, \epsilon)}\right\}.$$

reduces to

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} X_f\right] \leq \inf_{\epsilon > 0} \left\{2\epsilon + \sqrt{\frac{2c}{\epsilon n}}\right\} \lesssim n^{-1/3}$$

which is unfortunately not **sharp**. In the second part of the presentation, we will see how sharper bounds can be obtained once we have improved our tool further.