The Method of Chaining

Pan Jing Bin

MA5249 Project Presentation Part 2

Outline

Drawback of Our Approach

The key idea is to **decompose** the supremum

$$\sup_{t \in T} X_t \le \sup_{t \in T} \left\{ X_t - X_{\pi(t)} \right\} + \sup_{t \in T} X_{\pi(t)}$$

and use the **Lipschitz property** $|X_s - X_t| \lesssim d(s, t)$ to control the **remainder**.

Drawback of Our Approach

The key idea is to **decompose** the supremum

$$\sup_{t\in T} X_t \leq \sup_{t\in T} \left\{ X_t - X_{\pi(t)} \right\} + \sup_{t\in T} X_{\pi(t)}$$

and use the **Lipschitz property** $|X_s - X_t| \lesssim d(s,t)$ to control the **remainder**.

Problem: The Lipschitz property must hold **almost surely** and is far too restrictive!

Drawback of Our Approach

The key idea is to **decompose** the supremum

$$\sup_{t\in T} X_t \leq \sup_{t\in T} \left\{ X_t - X_{\pi(t)} \right\} + \sup_{t\in T} X_{\pi(t)}$$

and use the **Lipschitz property** $|X_s - X_t| \lesssim d(s,t)$ to control the **remainder**.

Problem: The Lipschitz property must hold **almost surely** and is far too restrictive! What if we only require the Lipschitz property to hold in probability instead?

$$\mathbb{E}\left[X_t - X_{\pi(t)}\right] \lesssim \epsilon.$$

Drawback of Our Approach

The key idea is to **decompose** the supremum

$$\sup_{t\in T} X_t \le \sup_{t\in T} \left\{ X_t - X_{\pi(t)} \right\} + \sup_{t\in T} X_{\pi(t)}$$

and use the **Lipschitz property** $|X_s - X_t| \lesssim d(s,t)$ to control the **remainder**.

Problem: The Lipschitz property must hold **almost surely** and is far too restrictive! What if we only require the Lipschitz property to hold in probability instead?

$$\mathbb{E}\left[X_t - X_{\pi(t)}\right] \lesssim \epsilon.$$

Another problem: Cannot control the remainder directly. For example, the maximum of M i.i.d standard normal variables is asymptotically $\gtrsim \log M$.

Drawback of Our Approach

The key idea is to **decompose** the supremum

$$\sup_{t\in T} X_t \le \sup_{t\in T}\left\{X_t - X_{\pi(t)}\right\} + \sup_{t\in T} X_{\pi(t)}$$

and use the **Lipschitz property** $|X_s - X_t| \lesssim d(s,t)$ to control the **remainder**.

Problem: The Lipschitz property must hold **almost surely** and is far too restrictive! What if we only require the Lipschitz property to hold in probability instead?

$$\mathbb{E}\left[X_t - X_{\pi(t)}\right] \lesssim \epsilon.$$

Another problem: Cannot control the remainder directly. For example, the maximum of M i.i.d standard normal variables is asymptotically $\gtrsim \log M$.

We are seemingly back at where we started...

The Method of Chaining

All hope is not lost: The magnitude of the remainder term $X_t - X_{\pi(t)}$ is **smaller** than the original process.

The Method of Chaining

All hope is not lost: The magnitude of the remainder term $X_t - X_{\pi(t)}$ is **smaller** than the original process.

Key idea: We repeat the same process but with a **finer** ϵ-net. If N' is an $\epsilon/2$-net, then

$$\sup_{t \in T} \left\{ X_t - X_{\pi(t)} \right\} \leq \sup_{t \in T} \left\{ X_t - X_{\pi'(t)} \right\} + \sup_{t \in T} \left\{ X_{\pi'(t)} - X_{\pi(t)} \right\}.$$

The Method of Chaining

All hope is not lost: The magnitude of the remainder term $X_t - X_{\pi(t)}$ is **smaller** than the original process.

Key idea: We repeat the same process but with a **finer** ϵ-net. If N' is an $\epsilon/2$-net, then

$$\sup_{t\in T}\left\{X_t - X_{\pi(t)}\right\} \le \sup_{t\in T}\left\{X_t - X_{\pi'(t)}\right\} + \sup_{t\in T}\left\{X_{\pi'(t)} - X_{\pi(t)}\right\}.$$

We can repeat this process **any number of times:**

$$\sup_{t\in T} X_t \le \sup_{t\in T}\overbrace{\left\{X_t - X_{\pi_n(t)}\right\}}^{\sim 2^{-n}} + \sum_{k=1}^{n}\sup_{t\in T}\overbrace{\left\{X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\right\}}^{\sim 2^{-k}} + \sup_{t\in T} X_{\pi_0(t)}.$$

The Method of Chaining

All hope is not lost: The magnitude of the remainder term $X_t - X_{\pi(t)}$ is **smaller** than the original process.

Key idea: We repeat the same process but with a **finer** ϵ-net. If N' is an $\epsilon/2$-net, then

$$\sup_{t \in T} \left\{ X_t - X_{\pi(t)} \right\} \leq \sup_{t \in T} \left\{ X_t - X_{\pi'(t)} \right\} + \sup_{t \in T} \left\{ X_{\pi'(t)} - X_{\pi(t)} \right\}.$$

We can repeat this process **any number of times:**

$$\sup_{t \in T} X_t \leq \sup_{t \in T} \overbrace{\left\{ X_t - X_{\pi_n(t)} \right\}}^{\sim 2^{-n}} + \sum_{k=1}^{n} \sup_{t \in T} \overbrace{\left\{ X_{\pi_k(t)} - X_{\pi_{k-1}(t)} \right\}}^{\sim 2^{-k}} + \sup_{t \in T} X_{\pi_0(t)}.$$

If we can control the **telescoping series** and ensure that the remainder term **vanishes**, then a good estimate can still be obtained!

Subgaussian Separable Processes

Definition 5.1. (Subgaussian process)

A random process $\{X_t\}_{t \in T}$ on the metric space (T, d) is **subgaussian** if $\mathbb{E}[X_t] = 0$ and

$$\mathbb{E}\left[e^{\lambda(X_s - X_t)}\right] \leq e^{\frac{\lambda^2 d(s,t)^2}{2}} \qquad \text{for all } s, t \in T \text{ and } \lambda \geq 0.$$

Subgaussian Separable Processes

Definition 5.1. (Subgaussian process)

A random process $\{X_t\}_{t\in T}$ on the metric space (T, d) is **subgaussian** if $\mathbb{E}[X_t] = 0$ and

$$\mathbb{E}\left[e^{\lambda(X_s - X_t)}\right] \le e^{\frac{\lambda^2 d(s,t)^2}{2}} \qquad \text{for all } s, t \in T \text{ and } \lambda \ge 0.$$

Definition 5.3. (Separable process)

A random process $\{X_t\}_{t\in T}$ is **separable** if there exists a countable subset $T_0 \subseteq T$ and an event E of probability 1 such that for all $\omega \in E$ and $t \in T$, there exists a sequence $(t_k)_{k=1}^{\infty}$ in T_0 satisfying

$$\lim_{k\to\infty} X_{t_k}(\omega) = X_t(\omega).$$

Dudley Chaining Argument

$$\sup_{t\in T} X_t \leq \sup_{t\in T} \overbrace{\left\{X_t - X_{\pi_n(t)}\right\}}^{\sim 2^{-n}} + \sum_{k=1}^{n} \sup_{t\in T} \overbrace{\left\{X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\right\}}^{\sim 2^{-k}} + \sup_{t\in T} X_{\pi_0(t)}.$$

With the **separability** and **subgaussian** assumptions, we have

Dudley Chaining Argument

$$\sup_{t\in T} X_t \le \sup_{t\in T} \overbrace{\left\{X_t - X_{\pi_n(t)}\right\}}^{\sim 2^{-n}} + \sum_{k=1}^{n} \sup_{t\in T} \overbrace{\left\{X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\right\}}^{\sim 2^{-k}} + \sup_{t\in T} X_{\pi_0(t)}.$$

With the **separability** and **subgaussian** assumptions, we have

Theorem 5.5. (Dudley)

Let $\{X_t\}_{t\in T}$ be a separable subgaussian process on the metric space (T, d). Then we have the following estimate:

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \le 6 \sum_{k\in\mathbb{Z}} 2^{-k} \sqrt{\log N(T, d, 2^{-k})}.$$

Dudley Chaining Argument

$$\sup_{t\in T} X_t \leq \sup_{t\in T} \overbrace{\left\{X_t - X_{\pi_n(t)}\right\}}^{\sim 2^{-n}} + \sum_{k=1}^{n} \sup_{t\in T} \overbrace{\left\{X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\right\}}^{\sim 2^{-k}} + \sup_{t\in T} X_{\pi_0(t)}.$$

With the **separability** and **subgaussian** assumptions, we have

Theorem 5.5. (Dudley)

Let $\{X_t\}_{t\in T}$ be a separable subgaussian process on the metric space (T, d). Then we have the following estimate:

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \leq 6 \sum_{k\in\mathbb{Z}} 2^{-k} \sqrt{\log N(T, d, 2^{-k})}.$$

Corollary 5.5.1. (Entropy Integral)

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \leq 12 \int_0^\infty \sqrt{\log N(T, d, \epsilon)}\, d\epsilon.$$

Wasserstein Law of Large Numbers Revisited

The process $\{X_f\}_{f \in \mathcal{F}}$ is subgaussian with respect to the rescaled metric $d(f, g) = n^{-1/2} \|f - g\|_\infty$.

Wasserstein Law of Large Numbers Revisited

The process $\{X_f\}_{f\in\mathcal{F}}$ is subgaussian with respect to the rescaled metric $d(f,g) = n^{-1/2}\,\|f-g\|_\infty$.

Lemma 3.8.

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

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

Using the entropy integral inequality and Lemma 3.8, we get

Wasserstein Law of Large Numbers Revisited

The process $\{X_f\}_{f \in \mathcal{F}}$ is subgaussian with respect to the rescaled metric $d(f, g) = n^{-1/2} \|f - g\|_\infty$.

Lemma 3.8.

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

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

Using the entropy integral inequality and Lemma 3.8, we get

$$\begin{aligned}\mathbb{E}\left[\sup_{f \in \mathcal{F}} X_f\right] &= 12 \int_0^\infty \sqrt{\log N\left(\mathcal{F}, n^{-1/2}\|\cdot\|_\infty, \epsilon\right)}\, d\epsilon \\ &= \frac{12}{\sqrt{n}} \int_0^\infty \sqrt{\log N\left(\mathcal{F}, \|\cdot\|_\infty, \epsilon\right)}\, d\epsilon \le \frac{12}{\sqrt{n}} \int_0^{1/2} \sqrt{\frac{c}{\epsilon}}\ d\epsilon.\end{aligned}$$

Wasserstein Law of Large Numbers Revisited

The process $\{X_f\}_{f\in\mathcal{F}}$ is subgaussian with respect to the rescaled metric $d(f,g) = n^{-1/2}\|f-g\|_\infty$.

Lemma 3.8.

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

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

Using the entropy integral inequality and Lemma 3.8, we get

$$\begin{aligned}\mathbb{E}\left[\sup_{f\in\mathcal{F}} X_f\right] &= 12\int_0^\infty \sqrt{\log N\big(\mathcal{F}, n^{-1/2}\|\cdot\|_\infty, \epsilon\big)}\, d\epsilon \\ &= \frac{12}{\sqrt{n}}\int_0^\infty \sqrt{\log N\big(\mathcal{F}, \|\cdot\|_\infty, \epsilon\big)}\, d\epsilon \le \frac{12}{\sqrt{n}}\int_0^{1/2}\sqrt{\frac{c}{\epsilon}}\, d\epsilon.\end{aligned}$$

Since the integral converges, we obtain the improved estimate

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} X_f\right] \lesssim n^{-1/2}$$

which is asymptotically **optimal!**

Lower Bounds for Gaussian Proceees

Lower Bounds for Gaussian Proceees

Question: In general, when are the upper bounds established by our method **sharp?**

Lower Bounds for Gaussian Proceees

Question: In general, when are the upper bounds established by our method **sharp?**

One possible approach: Supplement our upper bounds with corresponding **lower bounds.**

Lower Bounds for Gaussian Proceees

Question: In general, when are the upper bounds established by our method **sharp?**

One possible approach: Supplement our upper bounds with corresponding **lower bounds.**

Definition 7.1. (Gaussian process)

The random process $\{X_t\}_{t\in T}$ is called a **(centered) Gaussian process** if for all $n \in \mathbb{Z}_{\geq 1}$ and indices $t_1, \cdots, t_n$, the random variables $\{X_{t_1}, \cdots, X_{t_n}\}$ are centered (i.e. $\mathbb{E}[X_{t_j}] = 0$ for each j) and jointly Gaussian.

Definition 7.3. (Natural distance)

A Gaussian process $\{X_t\}_{t\in T}$ is subgaussian on (T, d) under the natural distance $d(s,t) := \mathbb{E}\left[|X_s - X_t|^2\right]^{1/2}$.

Finite Minima

Guiding philosophy: First study the **finite** case, and then generalise to the **infinite** case.

Finite Minima

Guiding philosophy: First study the **finite** case, and then generalise to the **infinite** case.

Lemma 7.4.

If $X_1, \cdots, X_n$ are i.i.d $\mathcal{N}(0, \sigma^2)$ random variables, then

$$c\sqrt{\sigma^2 \log n} \leq \mathbb{E}\left[\max_{i \leq n} X_i\right] \leq \sqrt{2\sigma^2 \log n}$$

for some universal constant $c \in \mathbb{R}_{\geq 0}$.

Generalisation to the infinite case

Problem: How to reduce a Gaussian random process $\{X_t\}_{t\in T}$ to the case of finitely many **independent** random variables?

Generalisation to the infinite case

Problem: How to reduce a Gaussian random process $\{X_t\}_{t\in T}$ to the case of finitely many **independent** random variables?

Theorem 7.5. (Slepian-Fernique)

Let $X \sim \mathcal{N}(0, \Sigma_X)$ and $Y \sim \mathcal{N}(0, \Sigma_Y)$ be n-dimensional Gaussian vectors. Suppose that we have

$$\mathbb{E}\left[|X_i - X_j|^2\right] \geq \mathbb{E}\left[|Y_i - Y_j|^2\right] \quad \text{for all } i, j \in \{1, \cdots, n\}.$$

Then

$$\mathbb{E}\left[\max_{1\leq k\leq n} X_k\right] \geq \mathbb{E}\left[\max_{1\leq k\leq n} Y_k\right].$$

Generalisation to the infinite case

Problem: How to reduce a Gaussian random process $\{X_t\}_{t \in T}$ to the case of finitely many **independent** random variables?

Theorem 7.5. (Slepian-Fernique)

Let $X \sim \mathcal{N}(0, \Sigma_X)$ and $Y \sim \mathcal{N}(0, \Sigma_Y)$ be n-dimensional Gaussian vectors. Suppose that we have

$$\mathbb{E}\left[|X_i - X_j|^2\right] \geq \mathbb{E}\left[|Y_i - Y_j|^2\right] \quad \text{for all } i, j \in \{1, \cdots, n\}.$$

Then

$$\mathbb{E}\left[\max_{1 \leq k \leq n} X_k\right] \geq \mathbb{E}\left[\max_{1 \leq k \leq n} Y_k\right].$$

Approach: Find well-separated points $\{X_{t_1}, X_{t_2}, \cdots, X_{t_k}\}$ (i.e. $\mathbb{E}\left[|X_i - X_j|^2\right] \geq \delta$) reduce to an **independent** process by choosing the parameters accordingly.

Generalisation to the infinite case

Problem: How to reduce a Gaussian random process $\{X_t\}_{t\in T}$ to the case of finitely many **independent** random variables?

Theorem 7.5. (Slepian-Fernique)

Let $X \sim \mathcal{N}(0, \Sigma_X)$ and $Y \sim \mathcal{N}(0, \Sigma_Y)$ be n-dimensional Gaussian vectors. Suppose that we have

$$\mathbb{E}\left[|X_i - X_j|^2\right] \geq \mathbb{E}\left[|Y_i - Y_j|^2\right] \quad \text{for all } i, j \in \{1, \cdots, n\}.$$

Then

$$\mathbb{E}\left[\max_{1\leq k\leq n} X_k\right] \geq \mathbb{E}\left[\max_{1\leq k\leq n} Y_k\right].$$

Approach: Find well-separated points $\{X_{t_1}, X_{t_2}, \cdots, X_{t_k}\}$ (i.e. $\mathbb{E}\left[|X_i - X_j|^2\right] \geq \delta$) reduce to an **independent** process by choosing the parameters accordingly.

Also agrees with our principle that **nearby points** are **highly dependent** and **points further away** are **nearly independent**.

Generalisation to the infinite case

Theorem 7.6. (Sudakov)

For a Gaussian process $\{X_t\}_{t \in T}$, we have the lower bound

$$\mathbb{E}\left[\sup_{t \in T} X_t\right] \geq \widetilde{c} \sup_{\epsilon > 0} \epsilon \sqrt{\log N(T, d, \epsilon)}$$

for a universal constant $\widetilde{c} \in \mathbb{R}_{\geq 0}$.

Generalisation to the infinite case

Theorem 7.6. (Sudakov)

For a Gaussian process $\{X_t\}_{t\in T}$, we have the lower bound

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \geq \widetilde{c} \sup_{\epsilon>0} \epsilon\sqrt{\log N(T,d,\epsilon)}$$

for a universal constant $\widetilde{c} \in \mathbb{R}_{\geq 0}$.

In conclusion:

$$\sup_{\epsilon>0} \epsilon\sqrt{\log N(T,d,\epsilon)} \lesssim \mathbb{E}\left[\sup_{t\in T} X_t\right] \lesssim \int_0^\infty \sqrt{\log N(T,d,\epsilon)}\, d\epsilon.$$

The End

Thank you for your attention.