Empirical Process

(Chapter 4) Chaining and uniform entropy

Qiangqiang Zhu

April 29, 2022

School of Mathematical Sciences
University of Southampton

Email: qz6u21@soton.ac.uk

Table of Content

Theorem 4.1 (Dudley's entropy bound for finite T)

Suppose that $\{X_t : t \in T\}$ is a mean zero stochastic process such that for every $s, t \in T$ and $u \geq 0$,

$$\mathbb{P}\{|X_t - X_s| \geq u\} \leq 2\exp\left(-\frac{u^2}{2d^2(s,t)}\right) \tag{1}$$

Also, assume that (T, d) is a finite metric space. Then, we have

$$\mathbb{E}\sup_{t\in T} X_t \leq C\int_0^\infty \sqrt{\log N(\epsilon, T, d)}\,d\epsilon \tag{2}$$

where $C > 0$ is a constant.

Proposition 4.2

Let T be a finite set and let $X_t, t \in T$ be a stochastic process. Suppose that for every $t \in T$ and $u \geq 0$, the inequality

$$\mathbb{P}\left(|X_t| \geq u\right) \leq 2\exp\left(-\frac{u^2}{2\sigma^2}\right) \tag{3}$$

holds. Then, for a universal positive constant C, we have

$$\mathbb{E}\max_{t \in T}|X_t| \leq C\sigma\sqrt{\log(2|T|)} \tag{4}$$

Proof of Proposition 4.2

First,

$$\mathbb{E} \max_{t \in T} |X_t| = \int_0^\infty \mathbb{P}\left(\max_{t \in T} |X_t| \geq u \right) du$$

we can write

$$\mathbb{P}\left(\max_{t \in T} |X_t| \geq u \right) = \mathbb{P}\left(\cup_{t \in T} \{ |X_t| \geq u \} \right) \leq \sum_{t \in T} \mathbb{P}\left(|X_t| \geq u \right) \leq 2|T| \exp\left(-\frac{u^2}{2\sigma^2} \right)$$

This bound is good for $u \geq u_0$ for some u_0 to be specified later. This gives

$$\begin{aligned} \mathbb{E} \max_{t \in T} |X_t| &= \int_0^{u_0} \mathbb{P}\left(\max_{t \in T} |X_t| \geq u \right) du + \int_{u_0}^\infty \mathbb{P}\left(\max_{t \in T} |X_t| \geq u \right) du \\ &\leq u_0 + \int_{u_0}^\infty 2|T| \exp\left(-\frac{u^2}{2\sigma^2} \right) du \\ &\leq u_0 + \int_{u_0}^\infty 2|T| \frac{u}{u_0} \exp\left(-\frac{u^2}{2\sigma^2} \right) du = u_0 + \frac{2|T|}{u_0} \sigma^2 \exp\left(-\frac{u_0^2}{2\sigma^2} \right) \end{aligned}$$

Proof of Proposition 4.2 (continued)

Here, we can set

$$u_0 = \sqrt{2}\sigma\sqrt{\log(2|T|)}$$

that is,

$$\exp\left(\frac{u_0^2}{2\sigma^2}\right) = 2|T|$$

This gives

$$\mathbb{E}\max_{t \in T}|X_t| \leq \sqrt{2}\sigma\sqrt{\log(2|T|)} + \frac{\sigma^2}{\sqrt{2\sigma^2\log(2|T|)}} \leq C\sigma\sqrt{\log(2|T|)}$$

which proves the result.

Theorem 4.3

Suppose (T, d) is a finite metric space and $\{X_t, t \in T\}$ is a stochastic process such that (1) hold. Then, for a universal positive constant C, the following inequality holds for every $t_0 \in T$:

$$\mathbb{E} \max_{t \in T} |X_t - X_{t_0}| \leq C \int_0^\infty \sqrt{\log D(\epsilon, T, d)} d\epsilon \lesssim \int_0^\infty \sqrt{\log N(\epsilon, T, d)} d\epsilon \qquad (5)$$

Here $D(\epsilon, T, d)$ denotes the ϵ-packing number of the space (T, d).

Remark 4.1

Let $\tilde{D}$ denote the diameter of the metric space T. Then $D(\epsilon, T, d)$ clearly equals 1 for $\epsilon \geq \tilde{D}$. Therefore,

$$\int_0^\infty \sqrt{\log D(\epsilon, T, d)} d\epsilon = \int_0^{\tilde{D}} \sqrt{\log D(\epsilon, T, d)} d\epsilon$$

Moreover,

$$\begin{aligned}\int_0^{\tilde{D}} \sqrt{\log D(\epsilon, T, d)} d\epsilon &= \int_0^{\tilde{D}/2} \sqrt{\log D(\epsilon, T, d)} d\epsilon + \int_{\tilde{D}/2}^{\tilde{D}} \sqrt{\log D(\epsilon, T, d)} d\epsilon \\ &= \int_0^{\tilde{D}/2} \sqrt{\log D(\epsilon, T, d)} d\epsilon + \int_0^{\tilde{D}/2} \sqrt{\log D(\epsilon + (\tilde{D}/2), T, d)} d\epsilon \\ &\leq 2 \int_0^{\tilde{D}/2} \sqrt{\log D(\epsilon, T, d)} d\epsilon\end{aligned}$$

because $D(\epsilon + (\tilde{D}/2), T, d) \leq D(\epsilon, T, d)$ for every ϵ.

Remark 4.1 (continued)

We can thus state Dudley's bound as

$$\mathbb{E}\max_{t\in T}|X_t - X_{t_0}| \leq C\int_0^{\tilde{D}/2}\sqrt{\log D(\epsilon, T, d)}d\epsilon$$

Similarly, again by splitting the above integral in two parts, we can also state Dudley's bound as

$$\mathbb{E}\max_{t\in T}|X_t - X_{t_0}| \leq C\int_0^{\tilde{D}/4}\sqrt{\log D(\epsilon, T, d)}d\epsilon$$

Proof of Theorem 4.3

For $n \geq 1$, let T_n be a maximal $\tilde{D}2^{-n}$-separated subset of T and T_n be a maximal cardinality subject to the separation restriction. The cardinality of T_n is given by the packing number $D\left(\tilde{D}2^{-n}, T, d\right)$. Because of the maximality,

$$\max_{t \in T} \min_{s \in T_n} d(s,t) \leq \tilde{D}2^{-n} \tag{6}$$

Because T is finite and $d(s,t) > 0$ for all $s \neq t$, the set T_n will equal T when n is large. Let

$$N := \min\{n \geq 1 : T_n = T\}$$

For each $n \geq 1$, let $\pi_n : T \to T_n$ denote the function which maps each point $t \in T$ to the point in T_n that is closest to T. In other words, $\pi_n(t)$ is chosen so that

$$d\left(t, \pi_n(t)\right) = \min_{s \in T_n} d(t,s)$$

Proof of Theorem 4.3 (continued)

From (6), we have

$$d\left(t, \pi_n(t)\right) \leq \tilde{D} 2^{-n} \qquad \text{for all } t \in T \text{ and } n \geq 1 \tag{7}$$

Note that $\pi_0(t) = t_0$ and $\pi_N(t) = t$ for all $t \in T$. Now

$$X_t - X_{t_0} = \sum_{n=1}^{N} \left(X_{\pi_n(t)} - X_{\pi_{n-1}(t)} \right) \qquad \text{for every } t \in T \tag{8}$$

By (8), we obtain

$$\max_{t \in T} |X_t - X_{t_0}| \leq \max_{t \in T} \sum_{n=1}^{N} \left| X_{\pi_n(t)} - X_{\pi_{n-1}(t)} \right| \leq \sum_{n=1}^{N} \max_{t \in T} \left| X_{\pi_n(t)} - X_{\pi_{n-1}(t)} \right|$$

so that

$$\mathbb{E} \max_{t \in T} |X_t - X_{t_0}| \leq \sum_{n=1}^{N} \mathbb{E} \max_{t \in T} \left| X_{\pi_n(t)} - X_{\pi_{n-1}(t)} \right| \tag{9}$$

Proof of Theorem 4.3 (continued)

For the elementary bound given by Proposition 4.2, note first that by (1), we have

$$\mathbb{P}\left\{\left|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\right| \geq u\right\} \leq 2\exp\left(\frac{-u^2}{2d^2\left(\pi_n(t), \pi_{n-1}(t)\right)}\right)$$

Now

$$d\left(\pi_n(t), \pi_{n-1}(t)\right) \leq d\left(\pi_n(t), t\right) + d\left(\pi_{n-1}(t), t\right) \leq \tilde{D}2^{-n} + \tilde{D}2^{-(n-1)} = 3\tilde{D}2^{-n}$$

Thus Proposition 4.2 can be applied with $\sigma := 3\tilde{D}2^{-n}$ so that we obtain

$$\begin{aligned} \mathbb{E}\max_{t\in T}\left|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\right| &\leq C\frac{3\tilde{D}}{2^n}\sqrt{\log\left(2\left|T_n\right|\left|T_{n-1}\right|\right)} \\ &\leq C\tilde{D}2^{-n}\sqrt{\log(2\left|T_n\right|^2)} \\ &\leq C\tilde{D}2^{-n}\sqrt{\log\left(2D\left(\tilde{D}2^{-n}, T, d\right)\right)} \end{aligned}$$

Proof of Theorem 4.3 (continued)

Plugging the above bound into (9), we deduce

$$\begin{aligned}
\mathbb{E}\max_{t\in T}|X_t - X_{t_0}| &\leq C\sum_{n=1}^{N}\frac{\tilde{D}}{2^n}\sqrt{\log\left(2D\left(\tilde{D}2^{-n}, T, d\right)\right)} \\
&\leq 2C\sum_{n=1}^{N}\int_{\tilde{D}/2^{n+1}}^{\tilde{D}/2^n}\sqrt{\log(2D(\epsilon, T, d))}d\epsilon \\
&\leq 2C\int_0^{\tilde{D}/4}\sqrt{\log(2D(\epsilon, T, d))}d\epsilon
\end{aligned}$$

Note that for $\epsilon \leq \tilde{D}/4$, the packing number $D(\epsilon, T, d) \geq 2$ so that

$$\log(2D(\epsilon, T, d)) = \log 2 + \log D(\epsilon, T, d) \leq 2\log D(\epsilon, T, d)$$

We have thus proved that

$$\mathbb{E}\max_{t\in T}|X_t - X_{t_0}| \leq 2\sqrt{2}C\int_0^{\tilde{D}/4}\sqrt{\log D(\epsilon, T, d)}d\epsilon$$

Definition 4.4 (Separable stochastic process)

Let (T, d) be a metric space. The stochastic process $\{X_t, t \in T\}$ indexed by T is said to be separable if there exists a null set N and a countable subset $\tilde{T}$ of T such that for all $\omega \notin N$ and $t \in N$, there exists a sequence $\{t_n\}$ in $\tilde{T}$ with $\lim_{n\to\infty} d(t_n, t) = 0$ and $\lim_{n\to\infty} X_{t_n}(\omega) = X_t(\omega)$.

If $\{X_t, t \in T\}$ is a separable stochastic process, then

$$\sup_{t\in T} |X_t - X_{t_0}| = \sup_{t\in \tilde{T}} |X_t - X_{t_0}| \qquad \text{almost surely} \tag{10}$$

for every $t_0 \in T$. Here $\tilde{T}$ is a countable subset of T which appears in the definition of separability of $X_t, t \in T$.

Theorem 4.5

Let (T, d) be a separable metric space and let $(X_t, t \in T)$ be a separable stochastic process. Suppose that for every $s, t \in T$ and $u \geq 0$, we have

$$\mathbb{P}\left\{|X_s - X_t| \geq u\right\} \leq 2\exp\left(-\frac{u^2}{2d^2(s,t)}\right)$$

Then for every $t_0 \in T$, we have

$$\mathbb{E}\sup_{t\in T}|X_t - X_{t_0}| \leq C\int_0^{\tilde{D}/4}\sqrt{\log D(\epsilon, T, d)}d\epsilon \tag{11}$$

where $\tilde{D}$ is the diameter of the metric space (T, d).

Proof of Theorem 4.5

Let $\tilde{T}$ be a countable subset of T. We may assume that $\tilde{T}$ contains t_0 (otherwise simply add t_0 to $\tilde{T}$). For each $k \geq 1$, let $\tilde{T}_k$ be the finite set obtained by taking the first k elements of $\tilde{T}$.

Applying Theorem 4.3 to $\{X_t, t \in \tilde{T}_k\}$, we obtain

$$\mathbb{E} \max_{t \in \tilde{T}_k} |X_t - X_{t_0}| \leq C \int_0^{\mathrm{diam}(\tilde{T}_k)/4} \sqrt{\log D(\epsilon, \tilde{T}_k, d)} d\epsilon \leq C \int_0^{\tilde{D}/4} \sqrt{\log D(\epsilon, T, d)} d\epsilon$$

Letting $k \to \infty$, we obtain

$$\mathbb{E} \sup_{t \in \tilde{T}} |X_t - X_{t_0}| \leq C \int_0^{\tilde{D}/4} \sqrt{\log D(\epsilon, T, d)} d\epsilon$$

Definition 4.6 (Uniform entropy bound)

A class $\mathcal{F}$ of measurable functions with measurable envelope F satifies the uniform entropy bound if and only if $J(1, \mathcal{F}, F) < \infty$ where

$$J(\delta, \mathcal{F}, F) := \int_0^\delta \sup_Q \sqrt{\log N\left(\epsilon \|F\|_{Q,2}, \mathcal{F} \cup \{0\}, L_2(Q)\right)} d\epsilon, \quad \delta > 0 \tag{12}$$

Fitness of the integral will be referred to as the uniform entropy condition.

Theorem 4.7

If $\mathcal{F}$ is a class of measuable functions with measurable envelop function F, then

$$\mathbb{E}\left[\|\mathbb{G}_n\|_{\mathcal{F}}\right] \lesssim \mathbb{E}\left[J\left(\theta_n, \mathcal{F}, F\right) \|F\|_n\right] \lesssim J(1, \mathcal{F}, F) \|F\|_{P,2} \tag{13}$$

where $\theta_n := \sup_{f \in \mathcal{F}} \|f\|_n / \|F\|_n$ and $\mathbb{G}_n(f) = \sqrt{n}\left(\mathbb{P}_n - P\right)(f)$.

Proof of Theorem 4.7

Recall that

Lemma 3.12 (Hoeffding's inequality for Rademacher variables)

Let $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ be a vector of constants and $\varepsilon_1, \ldots, \varepsilon_n$ be rademacher random variables. Then

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i \varepsilon_i\right| \geq x\right) \leq 2e^{-x^2/\left(2\|a\|^2\right)}$$

where $\|a\|$ denotes the Euclidean norm of a.

Theorem 3.17 (Symmetrisation)

Fro any class of measurable function $\mathcal{F}$,

$$\mathbb{E}\left\|\mathbb{P}_n - P\right\|_{\mathcal{F}} \leq 2\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i f(X_i)\right\|_{\mathcal{F}}$$

Proof of Theorem 4.7 (continued)

It suffices to bound $\mathbb{E}\left\|\mathbb{G}_n^o\right\|_{\mathcal{F}}$; recall that $\mathbb{G}_n^o(f)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i f\left(X_i\right)$ where ε_i's are i.i.d. Rademacher (by Theorem 3.17). Given $X_1,\ldots,X_n$, the process $\mathbb{G}_n^o$ is sub-Gaussian for the $L_2(\mathbb{P}_n)$-seminorm $\|\cdot\|_n$ (by Lemma 3.12), i.e.,

$$\mathbb{P}\left(\left|\sum_{i=1}^n \varepsilon_i \frac{f(X_i)}{\sqrt{n}}-\sum_{i=1}^n \varepsilon_i \frac{g(X_i)}{\sqrt{n}}\right| \geq u \mid X_1,\ldots,X_n\right) \leq 2 e^{-u^2/(2\|f-g\|_n^2)}$$

$$\forall f, g \in \mathcal{F}, \forall u \geq 0$$

The value $\sigma_{n,2}^2:=\sup_{f\in\mathcal{F}} \mathbb{P}_n f^2=\sup_{f\in\mathcal{F}}\|f\|_n^2$ is an upper bound for the squared radius of $\mathcal{F}\cup\{0\}$ with respect to this norm. We add the function $f\equiv 0$ to $\mathcal{F}$, so that the symmetrised process is zero at some parameter.

Theorem 4.3 (with $X_{t_0}=0$) gives

$$\mathbb{E}_{\varepsilon}\left\|\mathbb{G}_n^o\right\|_{\mathcal{F}} \lesssim \int_0^{\sigma_{n,2}} \sqrt{\log N\left(\epsilon, \mathcal{F}\cup\{0\}, L_2\left(\mathbb{P}_n\right)\right)} d\epsilon$$

where $\mathbb{E}_{\varepsilon}$ is the expectation with respect to the Rademacher variables.

Proof of Theorem 4.7 (continued)

The right side can be bounded by

$$\int_0^{\sigma_{n,2}/\|F\|_n} \sqrt{\log N\left(\epsilon\|F\|_n, \mathcal{F}\cup\{0\}, L_2\left(\mathbb{P}_n\right)\right)}d\epsilon\|F\|_n \leq J\left(\theta_n, \mathcal{F}, F\right)\|F\|_n$$

Since $\theta_n \leq 1$, we have that $J\left(\theta_n, \mathcal{F}, F\right) \leq J(1, \mathcal{F}, F)$. Furthermore, by Jensen's inequality applied to the root function,

$$\mathbb{E}\|F\|_n \leq \sqrt{\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n F^2\left(X_i\right)\right]} = \|F\|_{P,2}$$

This gives the inequality on the right side of the theorem.

Thank you!