High Dimensional Probability and Statistics **2nd Semester, 2023-2024**

Lecture 5: Expectation of Suprema: Chaining

Instructor: Ke Wei *Scribe: Ke Wei (Updated: 2024/04/15)*

Recap and motivation: Recall from the last lecture that we have established the following bound based on (one-step) finite approximation:

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \le \mathbb{E}\left[\sup_{t\in T}\left(X_t - X_{\pi(t)}\right)\right] + \mathbb{E}\left[\sup_{t\in T} X_{\pi(t)}\right], \tag{5.1}$$

where $\pi(t)$ is the projection of t onto a covering set of T, denoted N. Assume T has infinite number of points. Then the first term on the right side of (5.1) still has infinite number of terms. Though for some problems (i.e., computing the expectation of the spectral norm of a random matrix), it can be shown that $\mathbb{E}\left[\sup_{t\in T}\left(X_t - X_{\pi(t)}\right)\right]$ is a small proportion of $\mathbb{E}\left[\sup_{t\in T} X_t\right]$, in general it is still not very convenient to process the first term. In order to mitigate this, we may continue to approximate T by a finer covering set, denoted N', and obtain

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \le \mathbb{E}\left[\sup_{t\in T}\left(X_t - X_{\pi'(t)}\right)\right] + \mathbb{E}\left[\sup_{t\in T}\left(X_{\pi'(t)} - X_{\pi(t)}\right)\right] + \mathbb{E}\left[\sup_{t\in T} X_{\pi(t)}\right],$$

where $\pi'(t)$ denotes the projection of t onto N'. Of course, we can iterate this process and the term with infinite number of points is expected to diminish to zero as the approximation becomes finer and finer. The chaining argument applies this mechanism from the coarsest covering (with only one point), which yields a more tractable bound for $\mathbb{E}\left[\sup_{t\in T} X_t\right]$.

Agenda:

- The chaining method
- Examples
- Generic chaining

5.1 The Chaining Method

Definition 5.1 (Sub-Gaussian process) *A random process $\{X_t\}_{t\in T}$ defined on a metric space (T,d) is called sub-Gaussian if $\mathbb{E}[X_t] = 0$ and $X_t - X_s$ is $d(t,s)^2$-sub-Gaussian for all $t,s\in T$.*

Since the variations within the random process is determined by the metric space (T,d), it is expected to exploit the structure of (T,d) to bound $\mathbb{E}[\sup_t X_t]$. The chaining method will be our focus of this lecture. It provides one way to exploit the structure of (T,d). (A more refined way (see [2] for example, equivalent to generic chaining which is presented in the last section but not required, may improve the chaining bound in some situations.)

Example 5.2 *Consider $X_t = \langle g, t\rangle$, where $g \in \mathcal{N}(0, I_d)$ and $t \in T \subset \mathbb{R}^d$. Since $X_t - X_s = \langle g, t-s\rangle$ is a $\|t-s\|_2^2$-Gaussian, $\{X_t\}_{t\in T}$ is certainly a sub-Gaussian process.*

Theorem 5.3 (Discrete Dudley inequality) *Let $\{X_t\}_{t\in T}$ be a separable sub-Gaussian process on the metric space (T,d). Then*[1]*,*

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

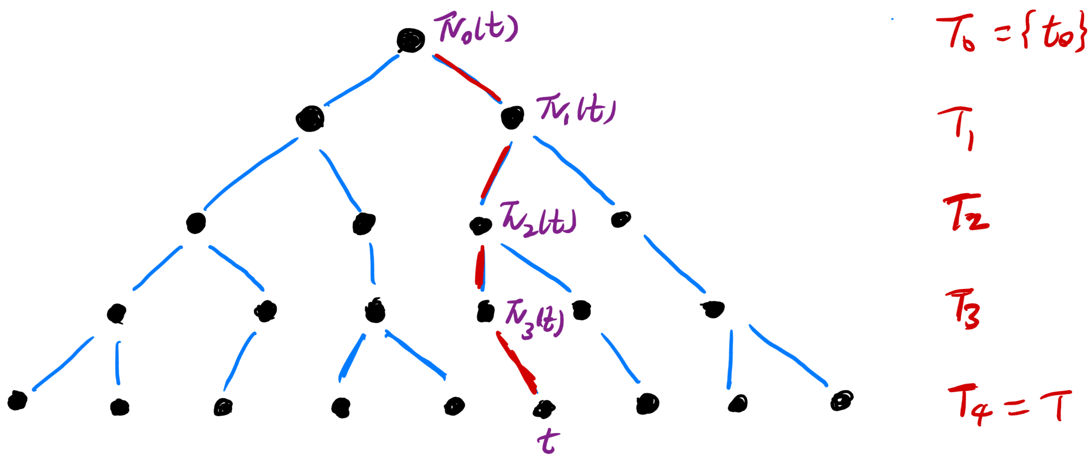

Figure 5.1: Illustration for chaining.

Proof: Without loss of generality we may assume $|T|<\infty$ since the separability of $\{X_t\}_{t\in T}$ implies that $\mathbb{E}\left[\sup_{t\in T} X_t\right] = \lim_{n\to\infty} \mathbb{E}\left[\sup_{t\in T_n} X_t\right]$ where T_n in an increasing finite subset of T.

Let $k_0\in\mathbb{Z}$ such that $2^{-k_0} > \mathrm{diam}(T)$. Then any singleton $T_0=\{t_0\}$ is an 2^{-k_0}-net of T. For $k>k_0$, let T_k be the 2^{-k}-net of T with covering number $N(T,d,2^{-k})$. Moreover, since $|T|<\infty$, there exists a sufficiently large K such that $T_K = T$, see Figure 5.1. Thus, we have

$$X_t = X_{t_0} + \sum_{k=k_0+1}^{K}\left(X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\right),$$

where $\pi_k(t)$ maps t to the nearest point in T_k. It follows that

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \le \sum_{k=k_0+1}^{K} \mathbb{E}\left[\sup_{t\in T}\left(X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\right)\right]. \tag{5.2}$$

First note that there are at most

$$|T_k||T_{k-1}| \le |T_k|^2 = N(T,d,2^{-k})^2$$

terms in $\sup_{t\in T}\left(X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\right)$. Moreover, since

$$d(\pi_k(t),\pi_{k-1}(t)) \le d(\pi_k(t),t) + d(t,\pi_{k-1}(t)) \le 3\times 2^{-k},$$

[1]The negative k in the sum denotes the approximation in the coarse (or large) scale with a small metric entropy. If $\mathrm{diam}(T)<\infty$, there exists a sufficiently small k_0 such that for all $k\le k_0$, $N(T,d,2^k)=1$ and thus $\log N(T,d,2^k)=0$.

and $X_{\pi_k(t)} - X_{\pi_{k-1}(t)}$ is $d(\pi_k(t), \pi_{k-1}(t))^2$-sub-Gaussian by the assumption, we have

$$\mathbb{E}\left[\sup_{t\in T}\left(X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\right)\right] \lesssim 2^{-k}\sqrt{\log N(T, d, 2^{-k})}.$$

Inserting this into (5.2) completes the proof. ■

Remark 5.4 *A careful reader may find out that what we have actually established in Theorem 5.3 is that*

$$\mathbb{E}\left[\sup_{t\in T}|X_t - X_{t_0}|\right] \lesssim \sum_{k\in\mathbb{Z}} 2^{-k}\sqrt{\log N(T, d, 2^{-k})}.$$

This observation will be useful in one of the examples in the sequel.

Discrete Dudley inequality bounds $\mathbb{E}\left[\sup_{t\in T} X_t\right]$ by a sum of (geometric structured) covering scales times the corresponding square root of metric entropies. The result can be written in an integral form since the sum can be viewed as a Riemann sum approximation to a certain integral.

Theorem 5.5 (Dudley integral) *Let $\{X_t\}_{t\in T}$ be a separable sub-Gaussian process on the metric space (T, d). Then*

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

Proof: The claim follows from

$$\begin{aligned}
\sum_{k\in\mathbb{Z}} 2^{-k}\sqrt{\log N(T, d, 2^{-k})} &= 2\sum_{k\in\mathbb{Z}}\int_{2^{-(k+1)}}^{2^{-k}} \sqrt{\log N(T, d, 2^{-k})}d\varepsilon \\
&\le 2\sum_{k\in\mathbb{Z}}\int_{2^{-(k+1)}}^{2^{-k}} \sqrt{\log N(T, d, \varepsilon)}d\varepsilon \\
&= 2\int_0^\infty \sqrt{\log N(T, d, \varepsilon)}d\varepsilon,
\end{aligned}$$

which completes the proof. ■

Remark 5.6 *In the proof we have shown that*

$$\sum_{k\in\mathbb{Z}} 2^{-k}\sqrt{\log N(T, d, 2^{-k})} \le 2\int_0^\infty \sqrt{\log N(T, d, \varepsilon)}d\varepsilon.$$

Actually, we can also establish that

$$\sum_{k\in\mathbb{Z}} 2^{-k}\sqrt{\log N(T, d, 2^{-k})} = \sum_{k\in\mathbb{Z}}\int_{2^{-k}}^{2^{-(k-1)}} \sqrt{\log N(T, d, 2^{-k})}d\varepsilon \ge \int_0^\infty \sqrt{\log N(T, d, \varepsilon)}d\varepsilon.$$

Thus nothing is lost in expressing the chaining bound as an integral rather than a sum, up to a constant factor.

Remark 5.7 . *It is worthing that we always have $N(T, d, \varepsilon) = 1$ when $\varepsilon \geq \operatorname{diam}(T)$. Thus, it is sufficient to take integral up to $\varepsilon = \operatorname{diam}(T)$.*

Remark 5.8 *It is not always the case that the bound by the Dudley integral is better than the one step discretization bound. Note that $N(T, d, \varepsilon)$ may approaches ∞ as ε approaches 0. Then the Dudley integral in an indefinite integral at the point 0. If $\sqrt{\log N(T, d, \varepsilon)}$ diverges very fast, the Dudley integral can be infinite. In this case the one step-discretization would still give a nontrivial bound even when the covering number is not integrable. Thus, sometimes, it is useful to combine the chaining method and the one step-discretization method to obtain a bound which mixes the Dudley integral (from a point strictly larger than zero) and the uniform one step-discretization bound, see for example Problem 5.11 in [2].*

5.2 Examples

5.2.1 Gaussian Complexity of $\mathbb{B}_2^d$

Recall that the Gaussian complexity of $\mathbb{B}_2^d$ is given by

$$\mathcal{G}(\mathbb{B}_2^d) = \mathbb{E}\left[\sup_{t \in \mathbb{B}_2^d} \langle g, t \rangle\right], \quad g \sim \mathcal{N}(0, I_d).$$

Letting $X_t = \langle g, t \rangle$, we known that $X_t - X_s$ is $\|t - s\|_2^2$-sub-Gaussian. Moreover, the covering number of $(\mathbb{B}_2^d, \|\cdot\|_2)$ at a scale $0 < \varepsilon < 1$ can be bounded by $(3/\varepsilon)^d$ (see Lecture 4). Thus, by the Dudley integral, we have

$$\begin{aligned}
\mathcal{G}(\mathbb{B}_2^d) &\lesssim \int_0^1 \sqrt{\log N(\mathbb{B}_2^d, \|\cdot\|_2, \varepsilon)} d\varepsilon \\
&= \sqrt{d} \int_0^1 \sqrt{\log \frac{3}{\varepsilon}} d\varepsilon \lesssim \sqrt{d},
\end{aligned}$$

which captures the correct order of $\mathcal{G}(\mathbb{B}_2^d)$, see Lecture 4.

5.2.2 A Failure Example

Let $T = \left\{ \frac{e_k}{\sqrt{1+\log k}} : \ k = 1, \cdots, n \right\}$, where e_k is the k-th canonical vector. Consider the Gaussian complexity of T,

$$\mathcal{G}(T) = \mathbb{E}\left[\sup_{t \in T} \langle g, t \rangle\right], \quad g \sim \mathcal{N}(0, I_d).$$

Note that $\mathcal{G}(T)$ can be explicitly written as

$$\mathcal{G}(T) = \mathbb{E}\left[\sup_{k=1,\cdots,n} \frac{g_k}{\sqrt{1 + \log k}}\right].$$

Thus, it can be shown that there exists a universal constant $C > 0$ such that for all n,

$$\mathcal{G}(T) \leq \mathbb{E}\left[\sup_{k=1,\cdots,n} \frac{|g_k|}{\sqrt{1 + \log k}}\right]$$

$$
\begin{aligned}
&= \int_0^\infty \mathbb{P}\left[\sup_{k=1,\cdots,n} \frac{|g_k|}{\sqrt{1+\log k}} \geq t\right] dt \\
&= \int_0^a \mathbb{P}\left[\sup_{k=1,\cdots,n} \frac{|g_k|}{\sqrt{1+\log k}} \geq t\right] dt + \int_a^\infty \mathbb{P}\left[\sup_{k=1,\cdots,n} \frac{|g_k|}{\sqrt{1+\log k}} \geq t\right] dt \\
&\leq a + \sum_{k=1}^n \int_a^\infty \mathbb{P}\left[\frac{|g_k|}{\sqrt{1+\log k}} \geq t\right] dt \\
&\leq C \qquad (\textbf{complete this step by choosing } a \textbf{ properly!})\ .
\end{aligned}
$$

However, we will show that the bound from the Dudley integral diverges as $n \to \infty$. Here, we consider the case $n = 2^{2^L}$. First note that the first m vectors in T is $1/\sqrt{\log m}$ separated. Thus, the packing number satisfies

$$
P(T, \|\cdot\|_2, 1/\sqrt{\log m}) \geq m.
$$

It follows that

$$
\begin{aligned}
&\int_0^\infty \sqrt{\log N(T, \|\cdot\|_2, \varepsilon)} d\varepsilon \\
&\geq \int_0^{\frac{1}{2\sqrt{\log(n)}}} \sqrt{\log N(T, \|\cdot\|_2, \varepsilon)} d\varepsilon \\
&+ \int_{\frac{1}{2\sqrt{\log(n)}}}^{\frac{1}{2\sqrt{\log\left(n^{1/2}\right)}}} \sqrt{\log N(T, \|\cdot\|_2, \varepsilon)} d\varepsilon \\
&+ \cdots\cdots \\
&+ \int_{\frac{1}{2\sqrt{\log\left(n^{1/2^{L-1}}\right)}}}^{\frac{1}{2\sqrt{\log\left(n^{1/2^L}\right)}}} \sqrt{\log N(T, \|\cdot\|_2, \varepsilon)} d\varepsilon \\
&\geq \int_0^{\frac{1}{2\sqrt{\log(n)}}} \sqrt{\log N\left(T, \|\cdot\|_2, \frac{1}{2\sqrt{\log n}}\right)} d\varepsilon \\
&+ \int_{\frac{1}{2\sqrt{\log(n)}}}^{\frac{1}{2\sqrt{\log\left(n^{1/2}\right)}}} \sqrt{\log N\left(T, \|\cdot\|_2, \frac{1}{2\sqrt{\log\left(n^{1/2}\right)}}\right)} d\varepsilon \\
&+ \cdots\cdots \\
&+ \int_{\frac{1}{2\sqrt{\log\left(n^{1/2^{L-1}}\right)}}}^{\frac{1}{2\sqrt{\log\left(n^{1/2^L}\right)}}} \sqrt{\log N\left(T, \|\cdot\|_2, \frac{1}{2\sqrt{\log\left(n^{1/2^L}\right)}}\right)} d\varepsilon \\
&\geq \int_0^{\frac{1}{2\sqrt{\log(n)}}} \sqrt{\log P\left(T, \|\cdot\|_2, \frac{1}{\sqrt{\log n}}\right)} d\varepsilon
\end{aligned}
$$

$$
\begin{aligned}
&+ \int_{\frac{1}{2\sqrt{\log(n)}}}^{\frac{1}{2\sqrt{\log\left(n^{1/2}\right)}}} \sqrt{\log P\left(T, \|\cdot\|_2, \frac{1}{\sqrt{\log\left(n^{1/2}\right)}}\right)} d\varepsilon \\
&+ \cdots\cdots \\
&+ \int_{\frac{1}{2\sqrt{\log\left(n^{1/2^{L-1}}\right)}}}^{\frac{1}{2\sqrt{\log\left(n^{1/2^{L}}\right)}}} \sqrt{\log P\left(T, \|\cdot\|_2, \frac{1}{\sqrt{\log\left(n^{1/2^{L}}\right)}}\right)} d\varepsilon \\
&\geq \frac{1}{2} + \frac{1}{2}\left(1 - \frac{1}{\sqrt{2}}\right) L \to \infty \text{ as } L \to \infty,
\end{aligned}
$$

where the third inequality follows from the relationship between covering number and packing number, and the last one uses the note that the first m vectors in T is $1/\sqrt{\log m}$ separated.

Thus, the Dudley inequality/integral is not able to capture the right bound for the Gaussian complexity of T in this example. Next we will present a method that works well for this example.

5.3 Generic Chaining

Before introducing generic chaining, we first reformulate the Dudley inequality into an equivalent form. To this end, we need to give the definition of *admissible sequence.* Let $\{T_k\}_{k=1}^{\infty}$ be a sequence of subsets of T. If

$$
|T_0| = 1, \quad |T_k| \leq 2^{2^k}, \; k = 1, 2, \cdots \tag{5.3}
$$

$\{T_k\}_{k=1}^{\infty}$ is called an *admissible sequence.*

Lemma 5.9 *We have*

$$
\int_0^{\infty} \sqrt{\log N(T, d, \varepsilon)} d\varepsilon \asymp \inf_{\{T_k\}_{k=0}^{\infty}} \sum_{k=0}^{\infty} 2^{k/2} \sup_{t \in T} d(t, T_k), \tag{5.4}
$$

where the infimum is taken over all the admissible sequence satisfying (5.3)*.*

Proof: First note that the righthand side (5.4) is equivalent to

$$
\sum_{k=0}^{\infty} 2^{k/2} \inf_{T_k} \sup_{t \in T} d(t, T_k).
$$

Then letting $e_k(T) = \inf_{T_k} \sup_{t \in T} d(t, T_k)$, one can easily see that

$$
e_0(T) = \inf\{\varepsilon : \; N(T, d, \varepsilon) = 1\}, \quad e_k(T) = \inf\{\varepsilon : \; N(T, d, \varepsilon) \leq 2^{2^k}\} \text{ for } k \geq 1.
$$

Therefore, for $\varepsilon < e_k(T)$, $N(T, d, \varepsilon) \geq 2^{2^k} + 1$. It follows that

$$
\int_{e_{k+1}(T)}^{e_k(T)} \sqrt{\log N(T, d, \varepsilon)} d\varepsilon \gtrsim 2^{k/2}(e_k(T) - e_{k+1}(T)).
$$

Consequently,

$$\begin{aligned}\int_0^{e_0(T)} \sqrt{\log N(T,d,\varepsilon)}d\varepsilon &\gtrsim \sum_{k=0}^{\infty} 2^{k/2}(e_k(T) - e_{k+1}(T)) \\ &= \sum_{k=0}^{\infty} 2^{k/2} e_k(T) - \sum_{k=1}^{\infty} 2^{(k-1)/2} e_k(T) \\ &\geq \left(1 - \frac{1}{\sqrt{2}}\right) \sum_{k=0}^{\infty} 2^{k/2} e_k(T),\end{aligned}$$

which completes the proof of one direction.

For the other direction, we have

$$\begin{aligned}\int_0^{\infty} \sqrt{\log N(T,d,\varepsilon)}d\varepsilon &= \int_0^{e_0(T)} \sqrt{\log N(T,d,\varepsilon)}d\varepsilon \\ &= \sum_{k=0}^{\infty} \int_{e_{k+1}(T)}^{e_k(T)} \sqrt{\log N(T,d,\varepsilon)}d\varepsilon \\ &\lesssim \sum_{k=0}^{\infty} 2^{(k+1)/2}(e_k(T) - e_{k+1}(T)) \\ &\lesssim \sum_{k=0}^{\infty} 2^k e_k(T).\end{aligned}$$

Now the proof is complete. ■

Remark 5.10 *The above lemma means that if we choose the sequence of covering numbers properly, fixing the sequence of the covering numbers and computing the related covering errors is equivalent to fixing the sequence of covering errors and computing the covering numbers in Dudley inequality. The derivation above also reveals why it requires* $|T_k| \leq 2^{2^k}$ *in the admissible sequence. Basically, we would like to have a matching lower and upper bound for Dudley integral in the form of the righthand of* (5.4)*. Alternatively, if we want* $\sqrt{\log N(T,d,\varepsilon)}$ *to be integrable towards* 0*,* $\sqrt{\log N(T,d,\varepsilon)}$ *is at most of order* $1/\sqrt{\varepsilon}$ *(other smaller than* 1 *power also be fine). Then,* $N(T,d,\varepsilon) \approx e^{1/\varepsilon} = e^{2^k}$ *when* $\varepsilon = 2^{-k}$.

The generic chaining will allow us to pull the supremum outside the sum and thus leads to a potentially smaller bound.

Theorem 5.11 (Generic chaining) *Let* $\{X_t\}_{t\in T}$ *be a separable sub-Gaussian process on the metric space* (T,d)*. Then*

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \lesssim \gamma(T,d) := \inf_{\{T_k\}_{k=0}^{\infty}} \sup_{t\in T} \sum_{k=0}^{\infty} 2^{k/2} d(t, T_k), \tag{5.5}$$

where the infimum is taken over all the admissible sequences.

Proof: As before, we can still assume $|T| < \infty$. Thus, it holds that

$$X_t - X_{t_0} = \sum_{k=1}^{K} \left(X_{\pi_k(t)} - X_{\pi_{k-1}}(t) \right),$$

where $\pi_k(t)$ maps t to the closest point in T_k. The overall goal is to show that

$$\mathbb{P}\left[\sup_{t\in T} |X_t - X_{t_0}| \gtrsim u\gamma(T,d) \right] \lesssim \exp\left(-u^2/2\right) \quad \text{for } u \geq c,$$

where $c > 0$ is an absolute constant. The claim will then follow immediately.

To this end, we will consider each term in the chaining sum and then take a uniform bound. Because $|X_{\pi_k(t)} - X_{\pi_{k-1}(t)}|$ is $d(\pi_k(t), \pi_{k-1}(t))^2$-sub-Gaussian, we have

$$\mathbb{P}\left[|X_{\pi_k(t)} - X_{\pi_{k-1}(t)}| \geq Cu2^{k/2} d(\pi_k(t), \pi_{k-1}(t)) \right] \leq 2\exp\left(-u^2 2^k\right), \tag{5.6}$$

where $C > 0$ is an absolute and fixed constant. Let Ω_u be the event such that

$$|X_{\pi_k(t)} - X_{\pi_{k-1}(t)}| \leq Cu2^{k/2} d(\pi_k(t), \pi_{k-1}(t)) \quad \text{for all } t \in T \text{ and } k.$$

Since there are at most $|T_k||T_{k-1}|$ terms in $|X_{\pi(t)} - X_{\pi_{k-1}(t)}|$, we have

$$\mathbb{P}\left[\Omega_u^c\right] \leq 2 \sum_{k\geq 1} 2^{2^{k+1}} \exp\left(-u^2 2^k\right).$$

Note that whenever Ω_u occurs, we have[2]

$$\sup_{t\in T} |X_t - X_{t_0}| \leq Cu \sup_{t\in T} \sum_{k=1}^{\infty} 2^{k/2} d(\pi_k(t), \pi_{k-1}(t)).$$

Consequently,

$$\mathbb{P}\left[\sup_{t\in T} |X_t - X_{t_0}| \geq Cu \sup_{t\in T} \sum_{k=1}^{\infty} 2^{k/2} d(\pi_k(t), \pi_{k-1}(t)) \right] \leq 2 \sum_{k\geq 1} 2^{2^{k+1}} \exp\left(-u^2 2^k\right)$$

Noting that $d(\pi_k(t), \pi_{k-1}(t)) \leq d(t, T_k) + d(t, T_{k-1})$, we have

$$\mathbb{P}\left[\sup_{t\in T} |X_t - X_{t_0}| \gtrsim u\gamma(T,d) \right] \leq 2 \sum_{k\geq 1} 2^{2^{k+1}} \exp\left(-u^2 2^k\right).$$

Thus, it only remains to bound $\sum_{k\geq 1} 2^{2^{k+1}} \exp\left(-u^2 2^k\right)$. Noting that

$$u^2 2^k \geq u^2/2 + u^2 2^{k-1} \geq u^2/2 + 2^{k+1}$$

for $u \geq 2$, one can easily obtain that $\sum_{k\geq 1} 2^{2^{k+1}} \exp\left(-u^2 2^k\right) \lesssim \exp\left(-u^2/2\right)$. ■

It is worth noting that the tail bound version of the Dudley integral and the generic chaining can also be established, see for example [2] or [3]. The generic chaining bound is not as convenient to use as the Dudley integral since constructing a good admissible sequence is not always easy. However, the difference between the generic chaining bound and the Dudley integral can look minor, but sometimes it is real. To see this, we revisit the failure example for the Dudley integral in Section 5.2.2.

[2] Basically, the argument here tensorize well without first triggering the sup in the first place.

Example 5.12 (Revisit of example in Section 5.2.2) *Consider the case* $n = 2^{2^L}$. *For notational convenience, we let* $t_k = e_k/\sqrt{1+\log k}$. *For any admissible sequence* $\{T_k\}$ *satisfying* $|T_k| \le 2^{2^k}$, *it is easy to show that*

$$\sup_{t\in T} d(t, T_k) \gtrsim 1/\sqrt{1+\log 2^{2^k}} \asymp 2^{-k/2}.$$

Thus, the bound obtained from Dudley integral is about

$$\sum_{k=1}^{L} O(1) = O(L) \to \infty.$$

In contrast, to apply generic chaining, we can construct an admissible sequence as follows:

$$T_0 = \{t_n\}, \quad T_k = \{t_2, \cdots, t_{2^{2^k}}, t_n\}, \ k = 1, \cdots, L-1.$$

Then give any $t \in T$, *there exists a* K *such that the index of* t *satisfies* $2^{2^K} < i(t) \le 2^{2^{K+1}}$. *It follows that*

$$\sum_{k=0}^{\infty} 2^{k/2} d(t, T_k) = \sum_{k=0}^{K} 2^{k/2} d(t, T_k) \lesssim \sum_{k=0}^{K} 2^{(k-K)/2} = O(1).$$

Here t_n *is included in* T_k *in order for* $d(t, T_k) \asymp 2^{-K/2}$, $k \le K$ *(independent of* k*). Because* t *is arbitrary, we can conclude that the generic chaining can capture the right magnitude in this special example.*

Reading Materials

[1] Martin Wainwright, *High Dimensional Statistics – A non-asymptotic viewpoint*, Chapter 5.3.

[2] Ramon van Handel, *Probability in High Dimension*, Chapter 5.3.

[3] Roman Vershynin, *High-Dimensional Probability: An introduction with applications in data science*, Chapter 8.