Chapter 6: Maximal Inequalities and Chaining

Empirical Processes (MATH-522)

Myrto Limnios

April, 2025

This chapter focuses on providing an upperbound of $\mathbb{E}\left[\sup_{t\in T}|X_t|\right]$ and shows why it measures the *size* of a generic process $\{X_t\}_{t\in T}$, when the index set is considered to be infinite. We will present a general method, namely *chaining method*, for obtaining sharp bounds of the quantity $\mathbb{E}\left[\sup_{t\in T}|X_t|\right]$ called *maximal inequalities*. We see that if the *size* of the index set T can be analyzed w.r.t. a distance based on the process X as in the Chapter 4, then we can control the *worst* deviation of the process uniformly on T.

Notations. We will use the notations of Chapter 5 without further notice. Recall that we considered an i.i.d. samples $X_1, \dots X_n$ defined on the p.s. $(\Omega, \mathcal{A}, \mathbb{P})$, valued in a measurable space $\mathcal{X} \subseteq \mathbb{R}^d$, $d \in \mathbb{N}^*$, of probability distribution P and of empirical p.d. P_n.

1 Introduction to the Chaining Method

1.1 Finite Index Set

Suppose the set T to be finite. We want to upperbound the maximum of a finite number of r.v.s. Notice that

$$\mathbb{E}[\sup_{t\in T} X_t] \leq \mathbb{E}[\sum_{t\in T}|X_t|] \leq |T| \sup_{t\in T}\mathbb{E}|X_t| \ .$$

Remark 1.1. *1. Controlling the magnitude of each of the r.v.s X-s seems unsatisfactory, and we want to take advantage of possibly some tail assumption of the r.v.s.*

2. The bound grows linearly with the size of T that is, again, unsatisfactory. It seems that we cannot get any good conclusion from this.

Suppose now that the r.v.s X have bounded p-moment, then Jensen's inequality helps us understand a refined control

$$\mathbb{E}[\sup_{t\in T} X_t] \leq \mathbb{E}[\sup_{t\in T}|X_t|^p]^{1/p} \leq |T|^{1/p} \sup_{t\in T}\mathbb{E}[|X_t|^p]^{1/p} \ .$$

Remark 1.2. *This bound is more interesting: if p is large, then the tails of the r.v.s are vanishing implying a smaller value for the expectation. In addition, the larger p and the slower growth in terms of $|T|$.*

We prove a first *maximal inequality* to see how to use more general functionals related to Cramér-Chernoff method resulting in sharper bounds. Before stating the first result, recall an important definition.

Definition 1.1. A process $(X_t)_{t\in T}$ defined on a metric space (T,d) is centered $(\nu>0)$-sub-Gaussian if $PX_t = 0$, if, for all $\lambda > 0$, for all $t\in T$,

$$\psi_X(\lambda) = \log \mathbb{E}[e^{\lambda X_t}] \le \frac{\lambda^2 \nu}{2} \ .$$

Lemma 1.2. *Consider a finite collection of elements $|T| < \infty$, and the process X_t be centered ν-sub-Gaussian. Suppose that we observe an i.i.d. sequence of X_t of size n of empirical measure P_n. Then it holds true that*

$$\mathbb{E}\left[\max_{t\in T} |P_n t|\right] \le \sqrt{\frac{\nu \log(2|T|)}{n}} \ . \tag{1}$$

Remark 1.3.

1. *The result is interesting as soon as we choose $n \ge \sqrt{\nu \log(2|T|)}$.*
2. *We study the max because the class of functions is finite, thus the supnorm is attainable.*
3. *Notice that the size of the class enters into play similarly as the square-root of the entropy.*
4. *This is the strongest result we can have, and it is sharp: If we know the best constant v upperbounding the variance of the process, then we cannot obtain better upperbound.*
5. *Notice that it can be related to Massart's inequality (Massart (2000), Lemma 5.2). Notice that we consider the absolute valued here, yielding a 2 in the log.*

Proof. By Jensen's inequality:

$$\begin{aligned}
\mathbb{E}\left[\max_{t\in T} |P_n t|\right] &= \frac{1}{\lambda}\mathbb{E}\left[\log \exp \lambda \max_{t\in T} |P_n t|\right] \\
&\le \frac{1}{\lambda}\log \mathbb{E}\left[\exp \lambda \max_{t\in T} |P_n t|\right] \\
&= \frac{1}{\lambda}\log \mathbb{E}\left[\max_{t\in T} \exp \lambda |P_n t|\right] \\
&\le \frac{1}{\lambda}\log \sum_{t\in T} \mathbb{E}\left[\exp \lambda |P_n t|\right] \ .
\end{aligned}$$

Notice that $e^{|x|} \le e^x + e^{-x}$, so that for $\lambda \ge 0$, and using the sub-Gaussianity assumption yields

$$\mathbb{E}[\exp\{\lambda |P_n t|\}] \le \mathbb{E}[\exp\{\lambda P_n t\}] + \mathbb{E}[\exp\{-\lambda P_n t\}] \le 2e^{\lambda^2\nu/2n} \ .$$

Thus

$$\mathbb{E}\left[\max_{t\in T} |P_n t|\right] \le \frac{\log(2|T|)}{\lambda} + \frac{\lambda\nu}{2n} \ .$$

Minimizing w.r.t. $\lambda^* = \sqrt{2n \log 2|T|/\nu}$ yields the result.

□

We can see this result similarly to the Chernoff bound that we studied in Chapter 2, i.e., if $\log \mathbb{E}[e^{\lambda X_t}] \le \psi(\lambda)$, then $\mathbb{P}(X_t \ge u) \le e^{-\psi^*(u)}$, for all $u \ge 0$ and $t \in T$. We thus formulate the maximal tail inequality.

Lemma 1.3 (Maximal tail inequality)**.** *Consider the process X_t as defined in Lemma 1.2, then for all $\delta \in (0,1)$, with probability at least $1-\delta$*

$$\max_{t\in T} |P_n t| \le \frac{\sqrt{\nu \log(2N)}}{n} + \frac{\sqrt{\nu \log(1/\delta)}}{n} \ .$$

Remark 1.4. *This result is again sharp, as soon as all the X-s are independent. Notice that if there is dependence, and for example all r.v.s are equal for all t, then essentially* $\max_{i \le N} |P_n t_j| = |P_n t_1|$ *and thus the bound is far from being optimal.*

The next section provides a generic method for obtaining similar rate of convergence, when the index set is uncountable.

1.2 One-step Discretization Chain under an Entropic Condition

From that simple derivation when considering the set T to be finite/countable, we saw that we can replace the supremum by the maximum, to then invoke the union bound. The idea is to approximate the supremum over T by a maximum of increments over an increasing sequence of covering sets with accuracy ε, plus an approximation error depending on ε and converging to 0. We will see that the sup can be upperbounded by the maximum of increments that depend on their size and number. Consider now an index class that can be approximated with an ε-cover w.r.t. the stochastic distance based d_X based on the process X_t. A key tool for the following results is to be able to control the size of the increments of the process $X_s - X_t$ in terms of the distance between the two points s and t, formulated below.

Definition 1.4. A process $(X_t)_{t\in T}$ defined on a metric space (T, d) is said to satisfy the **increment condition** if, for all $u > 0$, for all $s, t \in T$,

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

We say that the diameter of (T, d) is defined by $D(T) = \sup_{s,t\in T} d(s,t)$.

Reminder 1.5. *A centered process* $(X_t)_{t\in T}$ *defined on a metric space* (T, d) *is said to be sub-Gaussian iff*

$$\mathbb{E}[e^{\lambda(X_s - X_t)}] \le e^{\lambda^2 d_X(s,t)^2/2} ,$$

for all $\lambda \in \mathbb{R}$*, and for all* $s, t \in T$*.*

Remark 1.6. *Sub-Gaussian processes satisfy the increment condition w.r.t. a stochatstic metric* d_X *that it can be a pseudometric. It is typically* $d_X(s,t) = \mathbb{E}[|X_s - X_t|^2]^{1/2}$*. Gaussian or Rademacher processes indexed on* $[0,1]$*, we consider the Euclidean metric* $d(s,t) = \|s - t\|_2$*.*

Lemma 1.5. *Suppose* $(X_t)_{t\in T}$ *to be a sub-Gaussian centered process w.r.t* d_X*. Then for any* $\varepsilon \in [0, D(T)]$*, such that* $N(\varepsilon, T, d_X) \ge c$*, with* $c > 0$ *universal constant, it holds true that*

$$\mathbb{E}\left[\sup_{t,t'\in T}(X_t - X_{t'})\right] \le 2\mathbb{E}\left[\sup_{t,t'\in T,\ d_X(t,t')\le\varepsilon}(X_t - X_{t'})\right] + 4D(T)\sqrt{\log N(\varepsilon, T, d_X)} .$$

Proof. The idea of the proof is to use a cover of T to approximate the increment $(X_t - X_{t'})$ by the increments based on the centers of the covering sequence, with an additional approximation error.

Let $\varepsilon > 0$. Define by $t^1, \dots, t^N$ the centers of the ε-cover of T. Then, for any $t \in T$, there exists an index $i \le N$, such that $d_X(t, t^i) \le \varepsilon$. Hence

$$\begin{aligned} X_t - X_{t^1} &= \underbrace{X_t - X_{t^i}}_{\text{increment in } X \text{ between } t \text{ and its closest center } t^i} \\ &+ \underbrace{X_{t^i} - X_{t^1}}_{\text{increment in } X \text{ between the } best \text{ approximation of } t \text{ and any center of the cover}} \end{aligned}$$

Notice that

$$X_t - X_{t^i} \leq \sup_{t,t' \in T,\, d_X(t,t') \leq \varepsilon} (X_t - X_{t'})$$

because both t and t^i are in the ith element of the cover, thus $d_X(t, t^i) \leq \varepsilon$ by construction. And

$$X_{t^i} - X_{t^1} \leq \max_{i \leq N} |X_{t^i} - X_{t^1}|$$

because both t^i and t^1 are centers of the cover of T. Thus

$$X_t - X_{t^1} \leq \sup_{t,t' \in T,\, d_X(t,t') \leq \varepsilon} (X_t - X_{t'}) + \max_{i \leq N} |X_{t^i} - X_{t^1}| \ . \tag{2}$$

It holds true for any point $t' \in T$ as well (the cover is independent of t), so that we can add both bounds to obtain

$$\sup_{t,t' \in T} (X_t - X_{t'}) \leq 2 \sup_{t,t' \in T,\, d_X(t,t') \leq \varepsilon} (X_t - X_{t'}) + 2 \max_{i \leq N} |X_{t^i} - X_{t^1}| \ .$$

Notice that because the r.v.s are sub-Gaussian, then each increament is centered sub-Gaussian as well, with at most $d_X(t^i, t^1) \leq D(T)$ for the index set of the max. Lemma 1.2 applies, to the maximum on the right

$$\mathbb{E}\left[\max_{i \leq N} |X_{t^i} - X_{t^1}|\right] \leq 2D(T)\sqrt{\log N} \ .$$

We can select the optimal size of the cover to be $N = N(\varepsilon, T, d_X)$, and the result is proved. □

Remark 1.7. *It is not clear why we consider increments until we point out the following fact. For any fixed $t' \in T$, then*

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

and it is evident now that

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

We consider some examples.

Example 1.8 (Lipschitz processes)**.** *Consider the assumptions from Lemma 1.5, and that there exists a r.v. L such that X_t is L-Lipschitz, then*

$$\mathbb{E}\left[\sup_{t,t' \in T} (X_t - X_{t'})\right] \leq 2 \inf_{\varepsilon \in [0, D(T)]} \{\varepsilon \mathbb{E}[L] + 4D(T)\sqrt{\log N(\varepsilon, T, d_X)}\} \ .$$

Example 1.9 (Localized Canonical Gaussian Complexity)**.** *Recall that $\mathcal{G}(T) = \mathbb{E}[\sup_{t \in T} G_t] = \mathbb{E}[\sup_{t \in T} \langle \eta, t \rangle]$, with $T \subset \mathbb{R}^d$. Suppose $0 \in T$, and consider the ℓ_2-ball of radius ε by $T(\varepsilon) = \{t - t' \in \mathbb{R}^d, \ \|t - t'\|_2 \leq \varepsilon\}$. The natural metric d_X is the Euclidean $\| \cdot \|_2$. Thus Lemma 1.5 shows that we can control the Gaussian complexity of T by its localized complexity based on the ball $T(\varepsilon)$*

$$\mathcal{G}(T) \leq \inf_{\varepsilon \in [0, D(T)]} \{\mathcal{G}(T(\varepsilon)) + 2D(T)\sqrt{\log N_2(\varepsilon, T)}\} \ ,$$

with $N_2(\varepsilon, T)$ being the ε-covering number of T w.r.t. the Euclidean norm. Now, using Example 3.2 of Chapter 4, we have the explicit bound $\mathcal{G}(T(\varepsilon)) \leq \varepsilon\sqrt{d}$. It remains to compute an explicit upperbound of $N_2(\varepsilon, T)$ (left as exercise). Notice that we can get rid of the constant 2 in that case.

Example 1.10 (VC-classes of functions)**.** *Suppose $\mathcal{H}$ to be a* VC*-class of functions with finite* VC*-dimension. By Lemma 4.6, we have that $\sqrt{\log N(\varepsilon, T, d_X)} \leq C\sqrt{\mathcal{V} \log(1/\varepsilon)}$.*

1.3 Generic chaining based on covering sets - Dudley's entropy integral

Important intuition of the chaining method. Before, the supremum was approximated by a finite maximum over an ε-cover with an additional approximation error. We will now write the supremum as a finite sum of maxima indexed by successively refined sets.

Definition 1.6. Let (T, d) be a pseudometric space, and consider an ε-cover of finite covering number $N(\varepsilon, T, d)$ such that *Dudley's entropy integral* is well-defined by

$$\mathcal{J}(\varepsilon, D(T)) = \int_{\varepsilon}^{D(T)} \sqrt{\log N(u, T, d)} \mathrm{d}u \ ,$$

where $D(T) = \sup_{s,t \in T} d(s, t)$ is the diameter of T.

Theorem 1.7 (Wainwright , Theorem 5.22)**.** *Consider X_t a sub-Gaussian centered process w.r.t. the induced pseudometric d_X on T. Then, for any $\varepsilon \in [0, D(T)]$,*

$$\mathbb{E}\left[\sup_{t,t' \in T} (X_t - X_{t'})\right] \leq 2\mathbb{E}\left[\sup_{t,t' \in T,\ d_X(t,t') \leq \varepsilon} (X_t - X_{t'})\right] + 32 \int_{\varepsilon/4}^{D(T)} \sqrt{\log N(u, T, d_X)} \mathrm{d}u \ .$$

Corollary 1.8 (Dudley's Entropy Integral)**.** *Consider X_t a sub-Gaussian centered process w.r.t. the induced pseudometric d_X on T.*

$$\mathbb{E}\left[\sup_{t \in T} X_t\right] \leq 32 \int_{0}^{D(T)} \sqrt{\log N(\varepsilon, T, d_X)} \mathrm{d}\varepsilon \ .$$

Proof. Board.

□

Remark 1.11. *Why is this method called chaining? We want to refine the control of the second term in Lemma 1.5.*

Recall that we start with the ε-cover of T having $N(\varepsilon) = N(\varepsilon, t, d)$ elements, as in Lemma 1.5, defined by $T_0 = \{t^1, \dots, t^N\}$. For any $m = 1 \dots, M$ we consider refined covers T_m of T_0 composed of $N_m \leq N(\varepsilon_m)$ elements, with $\varepsilon_m = 2^{-m}D$, where M is such that $T_M = T_0$ (it exists since T_0 is finite), thus $\varepsilon = 2^{-(M-1)}(D/2) = 2^{-M}D$. By defining π_m the to be the best approximation of $t \in T_0$ by an element of the subset T_m - where optimality is in the sense of minimizing d - the proof relies on the decomposition relating X_t to X_{γ_1}, where $\gamma_1 \in T_1$ via a telescopic sum, where $\gamma_M = t$, and $\gamma_m = \pi_m(t)$. The chain is

$$T_0 \ni t = \gamma_M \to \gamma_{M-1} := \pi_{M-1}(t) \to \dots \to \gamma_1 \in T_1 \ . \tag{3}$$

Then one can control the increment

$$|X_t - X_{\gamma_1}| = \sum_{m=2}^{M} |X_{\gamma_m} - X_{\gamma_{m-1}}| \leq \sum_{m=2}^{M} \max_{u \in T_m} |X_u - X_{\pi_{m-1}(u)}|$$

A similar chain γ' can be obtained for any other element $t' \in T_0$, and thus we can upperbound

$$\max_{t,t' \in T_0} |X_t - X_{t'}| \leq \max_{t,t' \in T_1} |X_t - X_{t'}| + 2 \sum_{m=2}^{M} \max_{u \in T_m} |X_u - X_{\pi_{m-1}(u)}|$$

that refines the second term of Lemma 1.5.

Maximal inequality for VC-classes of functions. We first need an important result relating VC-dimension for VC-classes of functions, and the size of a covering set.

Lemma 1.9. *Let $\mathcal{H}$ be a VC-class of functions with measurable envelope function $H(t)$, and let $r \in \mathbb{N}^*$. Then, for any $\varepsilon > 0$, there exists a universal constant $C > 0$, such that*

$$N(\varepsilon\|H\|_r, \mathcal{H}, L_r(Q)) \leq C\mathcal{V}(16e)^{\mathcal{V}} \left(\frac{1}{\varepsilon}\right)^{2\mathcal{V}} ,$$

for any p.m. Q s.t. $\|H\|_r = (\int |H(x)|^r \mathrm{d}Q(x))^{1/r} < \infty$.

Corollary 1.10. *Let $\mathcal{H}$ be a VC-class of functions uniformly bounded by a finite constant $K > 0$, and of finite VC-dimension $\mathcal{V} > 0$. Consider an i.i.d. random sample $X_1, \ldots, X_n$ drawn from P. There exists a universal constant $C > 0$, such that*

$$\mathbb{E}\left[\|P_n - P\|_{\mathcal{H}}\right] \leq C\sqrt{\frac{\mathcal{V}}{n}} .$$

Remark 1.12. *Compare this rate to the ones obtained in Exercise session 7: what is the gain from Theorem 1.7 compared to the one-step discretization (Lemma 1.5)?*

Proof. By symmetrization we write

$$\mathbb{E}\left[\|P_n - P\|_{\mathcal{H}}\right] \leq 2\mathbb{E}\left[\|P_n^0\|_{\mathcal{H}}\right]$$

and notice that $P_n^0 = (1/n)\sum_{i=1}^n \varepsilon_i h(X_i)$ is the empirical Rademacher complexity based on the random sample $X_1, \ldots, X_n$. We want to apply Dudley's entropy integral with $\varepsilon = 0$ to the process $\{\sqrt{n}P_n^0 h,\ h \in \mathcal{H}\}$ conditionally on the X_i's. The process is centered, it has sub-Gaussian increments w.r.t. the empirical distance conditionally on the X_i's

$$d_X(h, g)^2 = \frac{1}{n}\sum_{i=1}^n (h(X_i) - g(X_i))^2$$

and that $\sup_{h,g\in\mathcal{H}} d_X(h, g) \leq 2K$. Sequentially using Remark 1.7, and Dudley's maximal inequality

$$\mathbb{E}_\varepsilon\left[\|P_n^0\|_{\mathcal{H}}\right] \leq \frac{1}{\sqrt{n}}\mathbb{E}_\varepsilon\left[\sup_{h,g\in\mathcal{H}} \frac{1}{\sqrt{n}}|\sum_{i=1}^n \varepsilon_i h(X_i) - g(X_i)|\right] \leq \frac{1}{\sqrt{n}}32\int_0^{2K} \sqrt{\log N(\varepsilon, T, d_X)}\mathrm{d}\varepsilon .$$

Now using Lemma 1.9, where the envelope function here is supposed to be uniformly bounded by K, it yields,

$$N(\varepsilon, \mathcal{H}, d_X) \leq C\mathcal{V}(16e)^{\mathcal{V}} \left(\frac{K}{\varepsilon}\right)^{2\mathcal{V}} ,$$

thus expressing the log and separating the terms depending on $\mathcal{V}$, ε, the result is obtained by lastly integrating w.r.t. P. □

2 Lower Bounds: Optimal rates

This last section proves a complementary result to that of Lemma 1.5 (one-step discretization) and Theorem 1.7 (Dudley's entropy integral). We will prove a lower bound for the uniform deviation of Gaussian processes in expectation. First, we provide general results for comparing two processes indexed by the same set. Then, we define a new metric measure for a set, namely ıpacking numbers, to prove the main theorem for Gaussian processes, known as Sudakov's Theorem.

2.1 Gaussian Comparison Inequalities

We want to compare two processes X_t, Y_t, indexed by the same set T, in the following sense. Given a real-valued function Φ, we would like to compare the real value $\mathbb{E}[\Phi(X)]$ to $\mathbb{E}[\Phi(Y)]$, and in particular for $\Phi(X_t) = \sup_{t \in T} X_t$. We focus on Gaussian processes.

Reminder 2.1. *We say that $X_t, t \in T$ is a centered Gaussian process if the random variables $\{X_{t_1}, \ldots, X_{t_M}\}$, are centered and jointly Gaussian, i.e., any linear combination is Gaussian, for all $M \geq 1$, and $\{t^1, \ldots, t^M\} \subset T$.*
In addition, for centered Gaussian r.v.s, its sub-Gaussian parameters equals its variance, and the canonical metric defined on T is given by

$$d_X(t, t') = \mathbb{E}[(X_t - X_{t'})^2] \ .$$

The first two results are formulated in terms of correlations and variances.

Theorem 2.1. *Let $\{X_1, \ldots, X_M\}$ and $\{Y_1, \ldots, Y_M\}$ be two centered Gaussian vectors. Suppose that there exist two disjoint subsets A, B of $\{1, \ldots, M\}^2$ such that*

$$\begin{aligned} \mathbb{E}[X_i X_j] &\leq \mathbb{E}[Y_i Y_j], \quad \forall i, j \in A \\ \mathbb{E}[X_i X_j] &\geq \mathbb{E}[Y_i Y_j], \quad \forall i, j \in B \\ \mathbb{E}[X_i X_j] &= \mathbb{E}[Y_i Y_j], \quad \forall i, j \notin A \cup B \ . \end{aligned}$$

Let $\Phi : \mathbb{R}^N \to \mathbb{R}$ be twice-differentiable, such that

$$\begin{aligned} \frac{\partial^2 \Phi(u)}{\partial u_i \partial u_j} &\geq 0, \quad \forall i, j \in A \\ \frac{\partial^2 \Phi(u)}{\partial u_i \partial u_j} &\leq 0, \quad \forall i, j \in B \ . \end{aligned}$$

Then, it holds true that

$$\mathbb{E}[\Phi(X)] \leq \mathbb{E}[\Phi(Y)] \tag{4}$$

Proof. Exercise. □

An important Corollary is the following.

Corollary 2.2 (Slepian's inequality)**.** *Let $\{X_1, \ldots, X_M\}$ and $\{Y_1, \ldots, Y_M\}$ be two centered Gaussian vectors. If*

$$\begin{aligned} \mathbb{E}[X_i X_j] &\geq \mathbb{E}[Y_i Y_j], \quad \forall i \neq j \\ \mathbb{E}[X_i^2] &= \mathbb{E}[Y_i^2], \quad \forall i \in \{1, \ldots, M\} \ , \end{aligned}$$

then it holds true that

$$\mathbb{E}\left[\max_{i=1\ldots M} X_i\right] \leq \mathbb{E}\left[\max_{i=1\ldots M} Y_i\right] \ . \tag{5}$$

Proof. Board. □

We now end with a result using the pseudometrics induced by the processes, namely $d_X = \mathbb{E}[(X_i - X_j)^2]$ and $d_Y = \mathbb{E}[(Y_i - Y_j)^2]$.

Theorem 2.3 (Sudakov-Fernique's Theorem)**.** *Let $\{X_1, \ldots, X_M\}$ and $\{Y_1, \ldots, Y_M\}$ be two centered Gaussian vectors. If*

$$\mathbb{E}[(X_i - X_j)^2] \leq \mathbb{E}[(Y_i - Y_j)^2], \quad \forall i, j \ ,$$

then

$$\mathbb{E}\left[\max_{i=1\ldots M} X_i\right] \leq \mathbb{E}\left[\max_{i=1\ldots M} Y_i\right] \ . \tag{6}$$

2.2 Lower Bound for Gaussian Processes

We are now ready to state the main result: lower bound for Gaussian Processes based on packing numbers. We first formulate a lower bound for finite maxima of Gaussian processes.

Lemma 2.4. *Let $\{X_1, \ldots, X_M\}$ be a set of centered Gaussian i.i.d. r.v.s of variance $\sigma^2 > 0$. Then there exists a small constant $c > 0$ such that*

$$\mathbb{E}\left[\max_{i \le M} X_i\right] \ge c\sigma\sqrt{\log M} \ .$$

Proof. Exercise. □

Packing Numbers.

Definition 2.5. An ε-packing of a set T w.r.t. a metric d, is a set $\{t^1, \ldots, t^M\} \subset T$ s.t. $d(t^i, t^j) > \varepsilon$, for all $i \neq j$. The ε-packing number $M(\varepsilon, T, d)$ is defined as the cardinality of the largest ε-packing of T.

Remark 2.2. *An ε-packing can be a collection of balls of at most $\varepsilon/2$ radii (the centers should be strictly varepsilon-appart), centered on elements of the index set T, and such that the intersections are empty.*

Lemma 2.6. *Let $\varepsilon > 0$. The packing and covering numbers are comparable*

$$M(2\varepsilon, T, d) \le N(\varepsilon, T, d) \le M(\varepsilon, T, d) \ .$$

Thus, when $\varepsilon \to 0$, packing and covering numbers have similar behavior.

Example 2.3 (Unit cubes)**.**

- *$d = 1$. Let $T = [-1, 1]$, equipped with $d(t, t') = |t - t'|$. Let $\varepsilon > 0$. Then Lemma 2.6 yields*

$$\lfloor 1/\varepsilon \rfloor \le N(\varepsilon, T, |\cdot|) \le 1/\varepsilon + 1$$

  and for ε very small, $N(\varepsilon, T, |\cdot|) = O(1/\varepsilon)$

- **Exercise.** *Prove that for $d \in \mathbb{N}^*$, it holds true that $N(\varepsilon, [-1, 1]^d, \|\cdot\|_\infty) = O((1/\varepsilon)^d)$, where $\|t\|_\infty = \sup_{i \le d} |t_i|$.*

Lower bound. We state the main Theorem of this section.

Theorem 2.7 (Sudakov minoration)**.** *Let $\{X_t,\ t \in T\}$ be a centered Gaussian process indexed by $T \neq \emptyset$. Then, there exists a small constant $c > 0$, such that*

$$\mathbb{E}\left[\sup_{t \in T} X_t\right] \ge \sup_{\varepsilon > 0} c\varepsilon\sqrt{\log M(\varepsilon, T, d_X)} \ ,$$

where $d_X(t, t') = \mathbb{E}[(X_t - X_{t'})^2]$.

Proof. Board. □

Example 2.4 (Lower bounds for Gaussian complexity)**.**