ORIE 7790 High Dimensional Probability and Statistics	Lecture 10 - 02/20/2020

Lecture 10: Random Processes and Metric Entropy

Lecturer: Yudong Chen *Scribe: Lucy Huo*

References:

- R. Vershynin, High-dimensional Probability, Sections 4.2, 7.1, 7.4.
- M. J. Wainwright, High-dimensional Statistics: A Non-asymptotic Viewpoint, Sections 5.1, 5.2, 5.5.

1 Random Processes

Definition 1 (Random Process)**.** *A random process $(Z_\theta)_{\theta \in T}$ refers to a collection of random variables in the same probability space indexed by $\theta \in T$.*

Remark Stochastic processes is one example of random processes, where the index θ refers to time. For random processes, the index set T can be more general, .e.g., multi-dimensional.

Examples
Here, we give some examples of random processes with $T \subset \mathbb{R}^d$. The first three examples involve $T \subset \mathbb{R}^d$.

1. Rademacher Process: $Z_\theta = \langle \varepsilon, \theta \rangle = \sum_{i=1}^d \varepsilon_i \theta_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} \text{unif}\{\pm 1\}$.
2. Gaussian Process: $\forall T_0 \subset T$ with $|T_0| < \infty$, $(Z_\theta)_{\theta \in T_0}$ is jointly Gaussian.
3. Canonical Gaussian Process: $Z_\theta = \langle g, \theta \rangle = \sum_{i=1}^d g_i \theta_i, \quad g_i \overset{\text{iid}}{\sim} N(0, 1)$.

   In the next example, $T = \mathcal{F}$ is a class of functions $\mathcal{X} \to \mathbb{R}$.
4. Empirical Process: $Z_f = \frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}(f(X_1)), \quad \{X_i\}$ iid random variables.

2 Sub-Gaussian Increments

We are interested in developing an upper/lower bound of $\mathbb{E}(\sup_{\theta \in T} Z_\theta)$ using the covering number/metric entropy of T. To obtain a good bound, we need some assumptions on the structure of random processes. Therefore, we introduce the concept of sub-Gaussian increments.

Definition 2 (Sub-Gaussian Increments)**.** *$(Z_\theta)_{\theta \in T}$ has sub-Gaussian increments w.r.t. metric ρ on T, if*

$$\mathbb{E}\left[e^{\lambda(Z_\theta - Z_{\theta'})}\right] \leq e^{\lambda^2 \rho(\theta, \theta')^2/2}, \quad \forall \theta, \theta' \in T, \lambda \in \mathbb{R}$$

i.e., $Z_\theta - Z_{\theta'}$ is sub-Gaussian with parameter $\rho(\theta, \theta')^2$.

Examples
Following are some examples of random processes with sub-Gaussian increments:

1. Rademacher Process: $Z_\theta - Z_{\theta'} = \langle \varepsilon, \theta - \theta' \rangle$ is $\|\theta - \theta'\|_2^2$-sub-Gaussian.
   $\Rightarrow (Z_\theta)$ has sub-Gaussian increments w.r.t. $\rho(\theta, \theta') = \|\theta - \theta'\|_2$.
2. Gaussian Process: $Z_\theta - Z_{\theta'} \sim N(0, \mathbb{E}(Z_\theta - Z_{\theta'})^2)$.
   $\Rightarrow (Z_\theta)$ has sub-Gaussian increments w.r.t. $\rho(\theta, \theta') \triangleq \sqrt{\mathbb{E}(Z_\theta - Z_{\theta'})^2}$.
3. Canonical Gaussian Process: $Z_\theta - Z_{\theta'} = \langle g, \theta - \theta' \rangle \sim N(0, \|\theta - \theta'\|_2^2)$
   $\Rightarrow (Z_\theta)$ has sub-Gaussian increments w.r.t. $\rho(\theta, \theta') = \|\theta - \theta'\|_2$.

3 Sudakov's Lower Bound

Recall:

- $N(\varepsilon, T, \rho)$ is the covering number of T w.r.t. ρ.
- $\log(N(\varepsilon, T, \rho))$ is the metric entropy of T w.r.t. ρ.

We will introduce Sudakov's minorization inequality shortly. To prove Sudakov's, we will need one definition and two lemmas. Therefore, we present the definition and two lemmas first.

Definition 3 (Packing Number)**.** *$T_\varepsilon \subset T$ is called an ε-packing of T if $\rho(\theta, \theta') > \varepsilon, \forall \theta, \theta' \in T_\varepsilon$. The largest cardinality of ε-packing is called the packing number of T, denote as $M(\varepsilon, T, \rho)$.*

Lemma 1. *$\forall \varepsilon > 0$, $M(2\varepsilon, T, \rho) \leq N(\varepsilon, T, \rho) \leq M(\varepsilon, T, \rho)$.*

Proof The proof is left as an exercise. (See R. Vershynin, Exercise 7.4.2). □

Lemma 2. *Let $X_i \overset{iid}{\sim} N(0, \sigma^2), i = 1, 2, \ldots, N$. Then $\mathbb{E}\left[\max_{i=1,\ldots,N} X_i\right] \gtrsim \sigma\sqrt{\log N}$.*

Remark Note that the expectation is taken over a finite collection of Gaussian variables.

Proof The proof of this lemma is again left as an exercise.
Hint: utilize Markov Inequality and then directly calculate $\mathbb{P}(\max_i X_i \geq \sigma\sqrt{\log N})$. You could refer to this Gautam Kamath's writing for a similar calculation.[1] □

Theorem 1 (Sudakov's Minorization Inequality)**.** *Let $(Z_\theta)_{\theta \in T}$ be a zero-mean Gaussian process. Then*

$$\mathbb{E}\left[\sup_{\theta \in T} Z_\theta\right] \geq \frac{\varepsilon}{2}\sqrt{\log N(\varepsilon, T, \rho)}, \quad \forall \varepsilon \geq 0$$

where the metric is $\rho(\theta, \theta') \triangleq \sqrt{\mathbb{E}(Z_\theta - Z_{\theta'})^2}$

Remark Minorization means finding the lower bound, while majorization means upper.

Proof Let T_ε be an maximal ε-packing of T, with $|T_\varepsilon| = M(\varepsilon, T, \rho) \geq N(\varepsilon, T, \rho)$ by Lemma 1. Then,

$$\mathbb{E}\left[\sup_{\theta \in T} Z_\theta\right] \geq \mathbb{E}\left[\sup_{\theta \in T_\varepsilon} Z_\theta\right].$$

Next, we compare $(Z_\theta)_{\theta \in T_\varepsilon}$ with another process $(Y_\theta)_{\theta \in T_\varepsilon}$, where $Y_\theta \overset{\text{iid}}{\sim} N(0, \frac{\varepsilon^2}{2}), \theta \in T_\varepsilon$. Check that,

$$\forall \theta, \theta' \in T_\varepsilon, \quad \mathbb{E}(Z_\theta - Z_{\theta'})^2 = \rho(\theta, \theta')^2 > \varepsilon^2 = \mathbb{E}(Y_\theta - Y_{\theta'})^2.$$

The first equality in the above equation holds by the definition of metric ρ, the second inequality holds by the property of packing, and the third equality holds by the definition of (Y_θ).

By Sudakov-Fernique Comparison Theorem (Recall: Lecture7-Random Matrix I), we arrive at

$$\mathbb{E}\left[\sup_{\theta \in T_\varepsilon} Z_\theta\right] \geq \mathbb{E}\left[\sup_{\theta \in T\varepsilon} Y_\theta\right] \geq \frac{\varepsilon}{\sqrt{2}}\sqrt{\log |T_\varepsilon|} \gtrsim \varepsilon\sqrt{\log N(\varepsilon, T, \rho)}$$

The second inequality holds by Lemma 2. As such, we have proven Sudakov's Minorization Inequality. □

[1] `http://http://www.gautamkamath.com/writings/gaussian_max.pdf`

4 Applications of Sudakov's Minorization Inequality

Application 1. Gaussian Complexity of Unit ℓ_2 Ball $\mathbb{B}^d$

Here, we would like to bound $\mathbb{E}[\sup_{\theta\in\mathbb{B}^d}\langle\theta,g\rangle]$, with $g_i \stackrel{\text{iid}}{\sim} N(0,1)$. It is easy to obtain an upper bound:

$$\mathbb{E}\left[\sup_{\theta\in\mathbb{B}^d}\langle\theta,g\rangle\right] \le \mathbb{E}\|g\|_2 \le \sqrt{\mathbb{E}\|g\|_2^2} = \sqrt{d}$$

The first inequality holds due to Cauchy-Schwarz, and the second inequality holds by Jensen's Inequality.

By Sudakov's minorization, we could obtain the following lower bound,

$$\mathbb{E}\left[\sup_{\theta\in\mathbb{B}^d}\langle\theta,g\rangle\right] \gtrsim \varepsilon\sqrt{\log N(\varepsilon,\mathbb{B}^d,\|\cdot\|_2)} \ge \varepsilon\sqrt{\log\left(\frac{1}{\varepsilon}\right)^d} \gtrsim \sqrt{d}.$$

The first inequality holds by Sudakov's minorization, while the second holds by the property of covering number (recall Lecture 8 Lemma 1 Remark). The third inequality holds when we take $\varepsilon = \frac{1}{e}$.

Therefore, we can conclude that the upper bound is tight up to a constant.

Application 2. Lower Bound on Max Singular Value

For $X\in\mathbb{R}^{n\times n}$, with $X_{ij}\stackrel{\text{iid}}{\sim} N(0,1)$, we have $\forall\varepsilon>0$:

$$\mathbb{E}\|X\|_{op} = \mathbb{E}\left[\sup_{u,v\in\mathbb{S}^{n-1}}\langle X,uv^T\rangle\right] = \mathbb{E}\left[\sup_{u,v\in\mathbb{B}^n}\langle X,uv^T\rangle\right] \gtrsim \varepsilon\sqrt{\log N(\varepsilon,\mathbb{B}^n\times\mathbb{B}^n,\|uv^T-\tilde{u}\tilde{v}^T\|_F)}.$$

To lower bound the last right hand side, we use the inequality from Lemma 1:

$$N(\varepsilon,\mathbb{B}^n\times\mathbb{B}^n,\|uv^T-\tilde{u}\tilde{v}^T\|_F) \ge M(2\varepsilon,\mathbb{B}^n\times\mathbb{B}^n,\|uv^T-\tilde{u}\tilde{v}^T\|_F),$$

and then find a lower bound for the packing number. Consider a maximal packing set, $\{u_1,u_2,\ldots,u_m\}$, for $\mathbb{B}^n$, i.e., $M(2\varepsilon,\mathbb{B}^n,\|\cdot\|_2)=m$. By the definition of a packing set, we know that $\|u_i-u_j\|_2 > 2\varepsilon$ for all $i\ne j\in[m]$. Next, fix an arbitrary point $v\in\mathbb{S}^{n-1}\subset\mathbb{B}^n$. It is clear that $S:=\{(u_i,v)\}_{i=1,\ldots,m}\subseteq\mathbb{B}^n\times\mathbb{B}^n$. Furthermore, we claim that the set S is a 2ε-packing for $\mathbb{B}^n\times\mathbb{B}^n$ under the metric $\|uv^T-\tilde{u}\tilde{v}^T\|_F$. Indeed, for each pair $i\ne j\in[m]$, we have

$$\begin{aligned}\|u_iv^\top-u_jv^\top\|_F &= \|(u_i-u_j)v^\top\|_F\\ &= \|u_i-u_j\|_2\cdot\|v\|_2\\ &= \|u_i-u_j\|_2 > 2\varepsilon.\end{aligned}$$

It follows that

$$M(2\varepsilon,\mathbb{B}^n\times\mathbb{B}^n,\|uv^T-\tilde{u}\tilde{v}^T\|_F) \ge |S| = m = M(2\varepsilon,\mathbb{B}^n,\|\cdot\|_2).$$

Finally, by Lemma 1 we have

$$M(2\varepsilon,\mathbb{B}^n,\|\cdot\|_2) \ge N(2\varepsilon,\mathbb{B}^n,\|\cdot\|) \gtrsim \left(\frac{1}{2\varepsilon}\right)^n,$$

where the last inequality was used in Application 1 above.

Combining pieces, we obtain that for any $\varepsilon>0$:

$$\mathbb{E}\|X\|_{op} \gtrsim \varepsilon\sqrt{\log\frac{1}{(2\varepsilon)^n}} = \varepsilon\sqrt{n\log\frac{1}{2\varepsilon}} \gtrsim \sqrt{n},$$

where the last step holds by choosing a suitable value for ε, e.g., $\varepsilon=\frac{1}{2e}$.

Compare the above lower bound with the upper bound we derived in Lecture 7:

$$\mathbb{E}\|X\|_{op} \leq 2\sqrt{n}.$$

We see that the two bounds match up to a constant.

Sudakov's minorization can also be used in reverse to upper bound the covering number and metric entropy, which will be covered in the next lecture.