Lecturer: Prof. Homrighausen Scribe: Zach Weller

In this lecture we conclude our coverage of the "concentration of measure" section by finishing up the uniform risk example. We discussed bracketing numbers for a function class $\mathcal{F}$ with the special case of Lipschitz in a parameter. We finish by discussing sub-Gaussianity.

1 Review: motivation & goal

We want to get uniform coverage of a stochastic process (i.e. $\{sup_{t\in T} X_t\}$). This type of result is fundamental for establishing performance guarantees of many algorithms (Lafferty et al. 2008). We achieve this by combining two structures:

1. Concentration inequalities: show that a random quantity is close to its mean with high probability
2. Uniform bounds: guarantee that a set of random quantities are all simultaneously close to their means with high probability

A nice chapter on concentration of measure from Lafferty, Liu, and Wasserman can be found online here.

2 Uniform Risk, continued

1. Recall, we wanted to show that the empirical risk looks like the true risk for a set of estimators. To do this, it is easier to generate the set of functions $\mathcal{L}_n = \{\ell_f : f \in \mathcal{F}_n\}$ and look at

$$\sup_{\ell\in\mathcal{L}_m} |\hat{\mathbb{P}}\ell - \mathbb{P}\ell|.$$

2. Since the function class $\mathcal{F}_n$ is (usually) infinite, we need to bound/quantify its complexity. This achieved through covering numbers, VC dimension, or bracketing numbers.
3. **Bracketing Numbers**
   (a) Punchline: bracketing numbers can be easier to bound/compute than covering numbers or VC-dimension
   (b) Idea: make a bracketing of $\mathcal{F}$ with a collections of brackets, $[L_j, U_j] = \{f \in \mathcal{F} : L_j(x) \le f(x) \le U_j(x), \forall x\}$. We are then interested in the bracketing number, $N_{[]}(\epsilon, \mathcal{F}, ||\cdot||_{L_q(\mathbb{P})})$, the smallest $\epsilon - L_q(\mathbb{P})$-bracketing. (aside: this is analogous to the ϵ-covers when using packing numbers).
   (c) We noted that bracketing numbers are a bit larger than covering numbers, $N(\epsilon, \mathcal{F}, ||\cdot||_{L_q(\mathbb{P})}) \le N_{[]}(2\epsilon, \mathcal{F}, ||\cdot||_{L_q(\mathbb{P})})$. However, they provide a stronger control over complexity, e.g. "Fundamental GC lemma"

   Theorem 2.1. *If $N_{[]}(\epsilon, \mathcal{F}, ||\cdot||_{L_1(\mathbb{P})}) < \infty \ \forall \epsilon > 0$, then $\mathcal{F}$ is Glivenko-Cantelli.*

   (d) An important case/application of bracketing numbers comes when the loss class, $\mathcal{L}$,is Lipschitz in a parameter, say $\beta \in \mathcal{B}$. If we have a Lipschitz type condition, we can hope to translate the complexity of $\mathcal{L}$ to the complexity of $\mathcal{B}$.

Definition 2.2 (Lipschitz)**.** *A function f is Lipschitz if $|f(x) - f(y)| \leq C|x - y|$ $\forall x, y$ where C is a constant independent of x and y.*

(e) Idea: a bracket on $\mathcal{B}$ can make a bracket on $\mathcal{L}$, provided $\ell_\beta - \ell_{\beta'}$ can't change too much relative to $\beta - \beta'$.

Example 2.3. *Let $(\mathcal{B}, ||\cdot||)$ be a normed subset of $\mathbb{R}^{p+1}$. If there is a function m where*

$$|\ell_\beta(z) - \ell_{\beta'}(z)| \leq m(z)||\beta - \beta'||$$

then

$$N_{[\,]}(\epsilon, \mathcal{F}, ||\cdot||_{L_q(\mathbb{P})}) \leq \left(\frac{4\sqrt{p+1}\ diam(\mathcal{B})||m||_{L_q(\mathbb{P})}^q}{\epsilon} \right)^{p+1}$$

(f) To get a more careful bound, we need to use the contraction theorem. We use the idea that we can map from the range space to the domain space with some level of control. We can use this to bound the complexity of the loss class, conditioned on the data. We can use the contraction theorem if

i. Z is bounded a.s., then the loss is bounded;

ii. with high probability, f is close to f_*, then use local Lipschitz property.

Theorem 2.4 (contraction theorem)**.** *If the loss ℓ is Lipschitz, then for any function $f_* \in \mathcal{F}$ and nonrandom z_i, we have*

$$\mathbb{E}\left(\sup_{f\in\mathcal{F}} \left|\sum_{i=1}^n \epsilon_i \left(\ell_f(z_i) - \ell_{f_*(z_i)}\right)\right|\right) \leq 2\mathbb{E}\left(\sup_{f\in\mathcal{F}} \left|\sum_{i=1}^n \epsilon_i \left(f(z_i) - f_*(z_i)\right)\right|\right).$$

3 Sub-Gaussian Bounds

1. Recall, a sub-Gaussian random variable is one that has tail decay at least as fast as a Gaussian. A sub-Gaussian process is a $\{X_t\}_{t\in T}$ where

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

* Important to remember that a process is sub-Gaussian w.r.t. the pseudo-metric on the index set (e.g. $t \in T \leftrightarrow f \in \mathcal{F}$).

2. **Main result:** we want to bound the expected value of the maximum of an infinite set of random variables. We do so via Dudley's inequality:

Lemma 3.1. *Suppose we have a sub-Gaussian process $(X_t)_{t\in T}$ such that $||X_s - X_t||_{\psi_2} \leq Cd(s,t)$. Also, let the diameter of T be $D = \sup_{s,t\in T} d(s,t)$. Then[1] if X_t is zero mean*

$$\mathbb{E}\sup_{t\in T} X_t \leq K\int_0^D \sqrt{\log(N(\epsilon, T, d))}d\epsilon$$

See Chapter 1.2 of Talagrand's Generic Chaining. For the symmetrizing statement for lower bounds, see Lemma 1.2.8.

[1] There are some other technical conditions, notably separability,to this result (the sup of a measurable process isn't necessarily measurable)

3. Lafferty et. all (2008) gives an example of applying this inequality:

Example 3.2 (7.102 from Lafferty et al.)**.** *Let $Y_1, \ldots, Y_n$ be a sample from a continuous CDF F on $[0,1]$ with bounded density. Let $X_s = \sqrt{n}(F_n(s) - F(s))$ where F_n is the empirical CDF. The collection $\{X_s : s \in [0,1]\}$ can be shown to be sub-Gaussian and sample continuous[2] w.r.t. the Euclidean metric on $[0,1]$. The covering number is $N([0,1], r) = 1/r$. Hence,*

$$\mathbb{E}\left(\sup_{0\leq s\leq 1} X_s\right) = \mathbb{E}\left(\sup_{0\leq s\leq 1} \sqrt{n}(F_n(s) - F(s))\right) \leq 12\int_0^{1/2} \sqrt{log(1/\epsilon}\, d\epsilon \leq C$$

, for some $C > 0$. Thus,

$$\mathbb{E}\left(\sup_{0\leq s\leq 1} \sqrt{n}(F_n(s) - F(s))\right) \leq \frac{C}{\sqrt{n}}$$

[2]See original document for this definition