Math 547: Mathematical Foundations of Statistical Learning Theory Fall 2017

Lecture 14 — September 22

Instructor: Stas Minsker *Scribe: Mose Wintner*

14.1 Generic chaining and metric entropy

(T,d) a finite metric space, $\{x(t),\, t\in T\}$ a stochastic process with sub-Gaussian increments with respect to d, meaning
$$x(t)-x(s)\in SG(d^2(t,s)).$$
A sequence $T_n\subset T$ not necessarily nested such that $card(T_n)\le 2^{2^n}$ with $|T_0|=1$ is an admissible sequence of subsets. The generic chaining complexity is

$$\gamma_2(T,d)=\inf_{T_n}\sup_{t\in T}\sum_{j\ge 0}2^{j/2}d(t,T_n).$$

For given T_n it's the maximum weighted sum of distances to T_n among all $t\in T$.

Theorem 1. Under the assumptions on $\{x(t),\, t\in T\}$, for all $u>0$,

$$Pr(\sup_{t\in T}|x(t)-x(t_0)|\ge 2u\gamma_2(T,d))\le 10e^{-u^2/4}.$$

In other words, $\sup_{t\in T}|x(t)-x(t_0)|-\mathbb{E}\sup_{t\in T}|x(t)-x(t_0)|\in SG(c_0\gamma_2(T,d))$, so that

$$\sup_{t\in T}(x(t)-x(t_0))-\mathbb{E}\sup_{t\in T}(x(t)-x(t_0))\in SG(c_1\gamma_2(T,d)).$$

Problem 1. What is $\mathbb{E}\sup_{t\in T}(x(t)-x(t_0))$?
Assume that T is arbitrary (i.e. not necessarily finite). Then

$$\mathbb{E}\sup_{t\in T}(x(t)-x(t_0)):=\sup_{\substack{S\subseteq T\\ S\text{ finite}}}\mathbb{E}\sup_{t\in S}x(t)-x(t_0).$$

Remark 1.
$$\sup_{t\in T}(x(t)-x(t_0))\ge 0.$$

For any nonnegative random variable Y,

$$\mathbb{E}Y=\int_0^\infty Pr(Y\ge t)\,dt.$$

Corollary 1.
$$\mathbb{E}\sup_{t\in T}(x(t)-x(t_0))\le c_2\gamma_2(T,d)$$

for c_2 a numerical constant.

14.1.1 Metric entropy and Dudley's theorem

Suppose $x(t)$ is a Gaussian process, i.e. for all $k \geq 1$, $t_1, \dots, t_k \in T$, $(x(t_1), \dots, x(t_k))$ is multivariate normal. Assume $\mathbb{E}x(t) = 0$ for all t. Then $x(t)$ has sub-Gaussian increments with respect to $d(s,t)$. Then

$$d^2(s,t) = \mathrm{Var}(x(s) - x(t))$$

is a natural metric associated to the Gaussian process.

Definition 1. Let (T, d) be a relatively compact metric space. The ε covering number of (T, d), written $N(T, d, \varepsilon)$ is the minimum n such that there exist $t_1, \dots, t_n \in T$ with $T \subseteq \cup_{j=1}^n B(t_j, \varepsilon)$.

Definition 2. The metric entropy of (T, d) is defined as

$$H(\varepsilon) = \log N(T, d, \varepsilon).$$

Example 1. $T = [0,1]^d$
a) $d(t,s) = ||t - s||_\infty$, $N(\varepsilon) = \varepsilon^{-d}$, $H(\varepsilon) = -d \log \varepsilon$.
b) $d(t,s) = ||t - s||_2$.

$$\frac{1}{vol(B_2(\epsilon))} = const(d)\varepsilon^{-d} \leq N(\epsilon) \leq \left(\frac{\sqrt{d}}{\varepsilon}\right)^d$$

c) T is a set of smooth functions $f : [0,1]^d \to \mathbb{R}$, $f \in C^\alpha$. $d(f,g) = \sup_x |f(x) - g(x)|$, then $H(\varepsilon) \propto \varepsilon^{-d/k}$, $N(T, d, \varepsilon) \propto e^{\varepsilon^{-d/k}}$.

Theorem 2. Let $x(t)$ be a process with sub-Gaussian increments with respect to d, then for all $t_0 \in T$,

$$\mathbb{E} \sup_{t \in T} |x(t) - x(t_0)| \leq c \int_0^\infty \sqrt{H(\varepsilon)}\, d\varepsilon.$$

Note that $H(\varepsilon) = 0$ for ε greater than the diameter.