Lecture 13 — September 20

Instructor: Stas Minsker *Scribe: Mose Wintner*

13.1 Generic chaining and Dudley's entropy integral

Let $\{x(t), t \in T\}$ be a stochastic process, i.e. a family of random variables indexed by T.

Example 1. Brownian motion.

Example 2. $\mathcal{F} = \{f : S \to \{\pm 1\}\}$ a collection of binary classifiers. $(X_1, Y_1), \ldots, (X_n, Y_n) \in S \times \{\pm 1\}$ iid from P.

Define a function of f by

$$Z_n(f) = \frac{1}{n} \sum_{j=1}^{n} I\{Y_j \neq f(X_j)\} - Pr(Y \neq f(X)).$$

For fixed f the terms are small by LLN, and Hœffding's inequality bounds the total difference.

Problem 1. What is the size of $\sup_{t \in T} |x(t) - \mathbb{E}x(t)|$?

To ensure that this quantity is measurable, we will *define*

$$\sup_{t \in T} x(t) := \sup_{S \subset T \text{ finite}} \max_{t \in S} x(t).$$

Assume now that (T, d) is a metric space and that T is finite (we lose no generality here by our definition of sup above). For $S \subset T$, $t \in T$, define

$$d(t, S) = \inf_{s \in S} d(t, s).$$

We also now make the assumption that the process $\{x(t)\, t \in T\}$ has sub-Gaussian increments with respect to d, meaning for all t_1, $t_2 \in T$,

$$x(t_1) - x(t_2) = SG(d(t_1, t_2)^2).$$

Example 3. Brownian motion: $\{w(t), t \in (0, 1)\}$, $w(t) - w(s) \sim \mathcal{N}(0, |t - s|)$, $d(s, t) = \sqrt{|t - s|}$.

Let $T_0 \subset T_1 \subset T_2 \subset \cdots \subset T_n \subset \cdots = T$, and assume that the cardinality of T_k is $2^{2^{k-1}}$ (note that this chain terminates because T is finite).

Theorem 1. For all $u \geq 0$ and any approximating sequence T_n,

$$Pr\left(\sup_{t\in T}|x(t)-x(t_0)| \geq 2u \sup_{t\in T}\sum_{n=0} 2^{n/2}d(t,T_n)\right) \leq ce^{-u^2/4}$$

for some constant c.
Proof:

$$\gamma_2(T) := \inf_{T_n}\sup_{t}\sum_{n\geq 0} 2^{n/2}d(t,T_n)$$

is called the *generic chaining complexity* and depends only on the metric space (T,d).
First let $u > 0$ be given. Define $\pi_n : T \to T_n$ to be the nearest-neighbor function which sends an element of T to its nearest neighbor in T_n, so that $d(t,\pi_n t) = d(t,T_n)$. Then

$$x(t) - x(t_0) = \sum_{j=0}^{\infty}\left(x(\pi_{j+1}t) - x(\pi_j t)\right).$$

Define the event

$$E = \{|x(s_1)-x(s_2)| \leq u2^{j/2}d(s_1,s_2)\,\forall\, s_1 \in T_j,\ s_2 \in T_{j+1},\ j \geq 0\}.$$

We want to control $Pr(E^c)$; we have

$$\begin{aligned}
Pr(E^c) &= Pr\left(\bigcup_{\substack{j\geq 0\\ s_1\in T_j\\ s_2\in T_{j+1}}}\{|x(s_1)-x(s_2)| \geq u2^{j/2}d(s_1,s_2)\}\right)\\
&\leq \sum_{j=0}\ \sum_{(s_1,s_2)\in T_j\times T_{j+1}} Pr(|x(s_1)-x(s_2)| > u2^{j/2}d(s_1,s_2))\\
&\leq 2\mathrm{exp}\left(-\frac{u^2 2^j d(s_1,s_2)^2}{2d(s_1,s_2)^2}\right) = 2e^{-u^2 2^{j-1}}\\
&\leq \sum_{j\geq 0}\operatorname{card}(T_j)\operatorname{card}(T_{j+1})2e^{u^2 2^{j-1}}\\
&= 2\sum_{j\geq 0} 2^{2^j+2^{j+1}}e^{-u^2 2^{j-1}}\\
&= 2\sum_{j\geq 0} e^{-u^2 2^{j-1}+2^{j-1}(2+4)\log 2}\\
&\leq 2\sum_{j\geq 0}\exp\left(-u^2 2^{j-1} + 6\cdot 2^{j-1}\log 2\right).
\end{aligned}$$

If $u^2 > 24\log 2$, then the sum will decay quickly, and this sum is $\leq 2e^{-u^2 2^{j-3}} \leq ce^{-u^2/8}$ for some c. On E,

$$|x(t)-x(t_0)| \leq \sum_{j\geq 0}|x(\pi_{j+1}t)-x(\pi_j t)| \leq u\sum_{j\geq 0}2^{j/2}d(\pi_j t,\pi_{j+1}t) \leq u\sum_{j\geq 0}2^{j/2}d(t,T_j)+\frac{1}{\sqrt{2}}\sum_{j\geq 0}2^{(j+1)/2}d(t,T_{j+1})$$

which proves the result.