Math 547: Mathematical Foundations of Statistical Learning Theory	**Fall 2017**

Lecture 27 — October 23

Instructor: Stas Minsker *Scribe: Mose Wintner*

27.1 Talagrand's Contraction Inequality

First, from last time: $P(|Z-\mathbb{E}Z|\geq t)\leq 2e^{-2t^2/\sum c_j^2}$ by a proof a bit more careful than ours.

Let ϕ be a convex functions which is nondecreasing on $[0,\infty)$ and let φ be a contraction, meaning it's Lipschitz with constant 1. Then $\varphi(0)=0$.

Theorem 1. Let $T\subset\mathbb{R}^n$. Then

$$\phi\left(\frac{1}{2}\mathbb{E}\sup_{t\in T}\left|\sum\varepsilon_j\varphi(t_j)\right|\right)\leq\mathbb{E}\phi\left(\sup_{t\in T}\left|\sum\varepsilon_jt_j\right|\right).$$

Corollary 1. Let $\mathcal{F}=\{f:S\to\mathbb{R}\}$ such that $f\in\mathcal{F}\Rightarrow|f|\leq U$ a.s. Also let $X_1,\dots,X_n\in S$, $T=\{(f(X_1),\dots,f(X_n)):\ f\in\mathcal{F}\}$. If $X_1,\dots,X_n$ are iid from P, then

$$\mathbb{E}_\varepsilon||R_n||_{\varphi\circ\mathcal{F}}\leq 2\mathbb{E}_\varepsilon||R_n||_{\mathcal{F}}.$$

For example, if $\tilde{\varphi}(x)=x^2$ on the interval $[-U,U]$, $\varphi(x)=x^2/2U$ is a contraction. Hence

$$\mathbb{E}\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum\varepsilon_jf^2(X_j)\right|\leq 4U\mathbb{E}\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum\varepsilon_jf(X_j)\right|.$$

Proof of theorem:
First, we show that $\forall A:T\to\mathbb{R}$ measurable,

$$\mathbb{E}\phi\left(\sup_{t\in T}[A(t)+\sum_{j=1}^n\varepsilon_j\varphi(t_j)]\right)\leq\mathbb{E}\phi\left(\sup_{t\in T}[A(t)+\sum_{j=1}^n\varepsilon_jt_j]\right).$$

Turns out that $n=1$ is sufficient to prove the above. Indeed, if

$$\mathbb{E}\phi\left(\sup_{t\in T}[A(t)+\varepsilon\varphi(t)]\right)\leq\phi\left(\sup_{t\in T}[A(t)+\varepsilon t]\right),$$

then

$$\begin{aligned}\mathbb{E}\phi\left(\sup_{t\in T}[A(t)+\sum_{j=1}^{n-1}\varepsilon_j\varphi(t_j)+\varepsilon_n\varphi(t_n)]\right)&=\mathbb{E}\phi\left(\sup_{t\in T}[\widehat{A(t)}+\varepsilon_n\varphi(t_n)]\right)\\&\leq\mathbb{E}_{\varepsilon,\varepsilon_{n-1}}\phi\left(\sup_{t\in T}[A(t)+\sum_1^{n-1}\varepsilon_j\varphi_j(t_j)+\varepsilon_nt_n]\right)\\&\cdots\leq\mathbb{E}\phi\left(\sup_{t\in T}[A(t)+\sum_1^n\varepsilon_jt_j]\right).\end{aligned}$$

Want:

$$\mathbb{E}\phi\left(\sup_{t\in T}[A(t)+\varepsilon\varphi(t)]\right)\le\mathbb{E}\phi\left(\sup_{t\in T}[A(t)+\varepsilon t]\right)\Leftrightarrow\phi\left(\sup_{t}[A(t)-\varphi(t)]\right)+\varphi\left(\sup_{t}[A(t)+\varphi(t)]\right)$$
$$\le\phi\left(\sup_{t}[A(t)-t]\right)+\phi\left(\sup_{t}[A(t)+t]\right).$$

Let $A(t)=t_1$, $t=t_2$; we've reduced to $t\in\mathbb{R}$. Then want

$$\phi\left(\sup_{t_1,t_2\in\tilde{T}}t_1-\varphi(t_2))\right)+\phi\left(\sup_{t_1,t_2}(t_1+\varphi(t_2))\right)\le\phi\left(\sup_{t_1,t_2}(t_1-t_2)\right)+\phi\left(\sup_{t_1,t_2}(t_1+t_2)\right).$$

We can assume WLOG that supremum is attained (if not, just do argument for supremum minus ϵ for small $\epsilon>0$. Let $\overline{t_1},\overline{t_2};\overline{s_1},\overline{s_2}$ be the argmaxes of the above LHS expressions.

Case (a): $\overline{t_2}\ge 0$, $\overline{s_2}\ge 0$, $\overline{t_2}\ge\overline{s_2}$. This case is similar to the same with $\overline{t_2}\le\overline{s_2}$ after renaming the $\overline{t}$ and $\overline{s}$ and replacing φ with $-\varphi$. We will show

$$\phi(\overline{t_1}+\varphi(\overline{t_2}))+\phi(\overline{s_1}-\varphi(\overline{s_2}))\le\phi(\overline{t_1}+\overline{t_2})+\phi(\overline{s_1}-\overline{s_2}).$$

Now some cumbersome arithmetic...define

$$a=\overline{t_1}+\varphi(\overline{t_2})\qquad b=\overline{t_1}+\overline{t_2}\qquad c=\overline{s_1}-\overline{s_2}\qquad d=\overline{s_1}-\varphi(\overline{s_2}).$$

Then $a\le b$ since φ is a contraction, $c\le d$. Also $b-a=\overline{t_2}-\varphi(\overline{t_2})\ge\overline{s_2}-\varphi(\overline{s_2})=d-c$, so that

$$\varphi(\overline{t_2})-\varphi(\overline{s_2})\le\overline{t_2}-\overline{s_2}.$$

Then $c\le a\le d\le b$; it remains to show $c\le a$.

$$a=\overline{t_1}+\varphi(\overline{t_2})\ge\overline{s_1}+\varphi(\overline{s_2})\ge\overline{s_1}-\overline{s_2}=c,$$

the first due to nature of $\overline{t_1}$ as argmax, and the second because φ is a contraction. Then by convexity

$$\phi(b)-\phi(a)\ge\phi(d)-\phi(c),$$

which evidently completes our proof.

Case (b): $\overline{s_2}\ge\overline{t_2}$ reduces to previous case.
Case (c): $\overline{t_2}\le 0$, $\overline{s_2}\le 0$; then rename $(\overline{t_1},\overline{t_2})\mapsto(\overline{t_1},-\overline{t_2})$, same for s, $\tilde{\varphi(x)}=\varphi(-x)$. Reduces to previous case.
Case (d): $\overline{t_2}\ge 0$, $\overline{s_2}\le 0$. Then $\varphi(\overline{t_2})\le\overline{t_2})$, $\varphi(\overline{s_2})\le-\overline{s_2}$. Since ϕ is monotone,

$$\phi(\overline{t_1}+\varphi(\overline{t_2}))+\phi(\overline{s_1}-\varphi(\overline{s_2}))\le\phi(\overline{t_1}+\overline{t_2})+\phi(\overline{s_1}-\overline{s_2}).$$