Maximal Inequalities

BIO 251,
Lab 6

March 24, 2014

Orlicz norm

Definition.

Orlicz norm of a non-decreasing, convex function ψ with $\psi(0) = 0$ for a random variable X: $||X||_\psi$ is defined as:

$$||X||_\psi = \inf\{C > 0 : \mathrm{E}\,\psi\left(\frac{|X|}{C}\right) \leq 1\}$$

Examples

Example

- $\psi(x) = x^p$, gives us L_p
- $\psi_p(x) = exp(x^p) - 1$

Some Inequalities

- $||X||_p \leq ||X||_{\psi_p}$
- $||X||_{\psi_p} \leq (\log 2)^{p/q} ||X||_{\psi_q}$, $p \leq q$
- $||X||_p \leq p! ||X||_{\psi_1}$

Tail Bounds with Orlicz norms

Tail Bounds

- $P(|X| > x) \leq \dfrac{1}{\psi(x/||X||_\psi)}$, by Markov's Inequality
- Thus for ψ_p we get a bound of the sort $\approx \exp(-Cx^p)$

Lemma.

Let X be a random variable with $P(|X| > x) \leq K \exp(-Cx^p)$, for every x for some constants K and C, and for $p \geq 1$. Then its Orlicz norm satisfies $||X||_{\psi_p} \leq ((1+K)/C)^{1/p}$.

An inequality for L_p norms

For L_p norms

$$\left|\left|\max_{1\leq i\leq m} X_i\right|\right|_p = \left(\mathrm{E}\max_{1\leq i\leq m}|X_i|^p\right)^{1/p} \leq m^{1/p}\max_{1\leq i\leq m}||X_i||_p$$

In general for ψ norms we have:

Lemma

Let ψ be a convex, nondecreasing, nonzero function with $\psi(0) = 0$, and $\limsup_{x,y\to\infty} \frac{\psi(x)\psi(y)}{\psi(cxy)} < \infty$ for some constant c. Then for any random variables $X_1, X_2, \ldots, X_m$ we have:

$$\left|\left|\max_{1\leq i\leq m} X_i\right|\right|_\psi \leq K\psi^{-1}(m)\max_i ||X_i||_\psi$$

where the constant K depends on ψ only.

What happens when $\psi = \psi_p$

$\psi = \psi_p$

For $\psi(x) = \psi_p(x) = e^{x^p} - 1$ we have $\psi_p^{-1}(m) = (\log(1 + m))^{1/p}$.

Packing and Covering numbers

Let (T, d) be a an arbitrary semimetric space.

Definition (Covering numbers).

The *covering number* $N(\varepsilon, d)$ is the minimal number of balls of radius ε needed to cover T.

Definition (Packing numbers).

Call a collection of points ε-separated if the minimum distance between any two points is is strictly larger than ε. The *packing number* $D(\varepsilon, d)$ is the maximal number of ε-separated points in T.

Packing and Covering numbers

We have the following relationship

$$N(\varepsilon, d) \leq D(\varepsilon, d) \leq N(\frac{1}{2}\varepsilon, d)$$

Important Theorem

Theorem

Let ψ satisfy the previous requirements. Let $\{X_t : t \in T\}$, be a separable stochastic process, with:

$$||X_s - X_t||_\psi \leq Cd(s,t), \text{ for every } s, t$$

for some semimetric d on T and a constant C. Then, for any $\eta, \delta > 0$,

$$\left|\left| \sup_{d(s,t)\leq\delta} |X_s - X_t| \right|\right|_\psi \leq K \left[\int_0^\eta \psi^{-1}(D(\varepsilon, d))d\varepsilon + \delta\psi^{-1}(D^2(\eta, d)) \right]$$

for a constant K depending on ψ and C only.

Corollary

Corollary

The constant K can be chosen such that:

$$\left\| \sup_{s,t} |X_s - X_t| \right\|_\psi \leq K \int_0^{\operatorname{diam} T} \psi^{-1}(D(\varepsilon, d)) d\varepsilon$$

Hoeffding's Inequality (Special Case)

Lemma (Hoeffding's Inequality).

Let $a_1, \ldots, a_n$ are constants and $\varepsilon_1, \ldots, \varepsilon_n$ are iid *Rademacher* random variables: i.e. $P(\varepsilon = 1) = P(\varepsilon = -1) = 1/2$. Then we have:

$$P(|\sum \varepsilon_i a_i| > x) \leq 2e^{-\frac{1}{2}x^2/||a||^2}$$

for the Euclidean norm $||a||$. Consequently, $||\sum \varepsilon_i a_i||_{\psi_2} \leq \sqrt{6}||a||$.

Sub-Gaussian Processes

Definition

A stochastic process is called *sub-Gaussian* with respect to the semimetric d on its index set if:

$$P(|X_s - X_t| > x) \leq 2\exp^{-\frac{1}{2}x^2/d^2(s,t)}, \text{ for every } s,t \in T, x > 0$$

Note

Any sub-Gaussian process satisfies: $||X_s - X_t||_{\psi_2} \leq \sqrt{6}d(s,t)$

Examples

Gaussian Processes

Any Gaussian process is sub-Gaussian with respect to $d(s,t) = \sigma(X_s - X_t)$.

Rademacher Processes

$$X_a = \sum_{i=1}^{n} a_i \varepsilon_i, a \in \mathbb{R}^n$$

for Rademacher variables $\varepsilon_i, \ldots, \varepsilon_n$. By Hoeffding's inequality, this is sub-Gaussian wrt to the Euclidean distance $d(a,b) = ||a - b||$.

Corollary for sub-Gaussian Processes

Corollary

Let $\{X_t : t \in T\}$ be a separable sub-Gaussian process. Then for any $\delta > 0$:

$$\mathrm{E} \sup_{d(s,t)<\delta} |X_s - X_t| \leq K \int_0^\delta \sqrt{\log(D(\varepsilon, d))} d\varepsilon$$

for a universal constant K. In particular, for any t_0:

$$\mathrm{E} \sup_t |X_t| \leq \mathrm{E} |X_{t_0}| + K \int_0^\infty \sqrt{\log(D(\varepsilon, d))} d\varepsilon$$