

Symmetrization and GC Classes

BIO 251,
Lab 7

March 30, 2014

Symmetrization

Lemma (Symmetrization).

For every convex, nondecreasing $\Phi : \mathbb{R} \mapsto \mathbb{R}$ and a class of measurable functions \mathcal{F} ,

$$E^* \Phi(\|\mathbb{P}_n - P\|_{\mathcal{F}}) \leq E^* \Phi(2\|\mathbb{P}_n^0\|_{\mathcal{F}})$$

Measurable Class

Definition (Measurable class).

A class \mathcal{F} of measurable functions $f : \mathcal{X} \mapsto \mathbb{R}$ on a probability space $(\mathcal{X}, \mathcal{A}, P)$ is called a P -measurable class if the map:

$$(X_1, X_2, \dots, X_n) \mapsto \left\| \sum_{i=1}^n e_i f(X_i) \right\|_{\mathcal{F}}$$

is measurable for all $\{e_1, \dots, e_n\} \in \mathbb{R}^n$ on the completion of the space $(\mathcal{X}^n, \mathcal{A}^n, P^n)$.

Example

The use of the above definition is that we can claim that

$$E_X E_\varepsilon = E^*$$

Bracketing numbers

Definition (Bracketing numbers).

Given two functions l and u , the *bracket* $[l, u]$ is the set of all functions f with $l \leq f \leq u$ pointwise. An ε -*bracket* is a bracket $[l, u]$ with $\|u - l\| < \varepsilon$. The *bracketing number* $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ ($\|\cdot\|$ here is the norm on \mathcal{F}) is the minimum number of ε -brackets needed to cover \mathcal{F} . Note that u and l need not belong to \mathcal{F} , but are assumed to have finite norms.

GC Class Sufficient Condition 1

Theorem

Let \mathcal{F} be a class of measurable functions such that $N_{[]}(\varepsilon, \mathcal{F}, L_1(P)) < \infty$ for every $\varepsilon > 0$. Then \mathcal{F} is Glivenko-Cantelli.

Example

Example

Take \mathcal{F} to be the set of all indicator functions $\mathbb{1}_{(-\infty, c]}$. This class possesses a finite bracketing number, for any underlying distribution and for any $\varepsilon > 0$ and thus is GC.

GC Class Sufficient Condition 2

Theorem

Let \mathcal{F} be a P -measurable class of measurable functions with envelope F such that $P^*F < \infty$. Let \mathcal{F}_M be the class of functions $f \mathbb{1}_{F \leq M}$ when f ranges over \mathcal{F} . If $\log N(\varepsilon, \mathcal{F}_M, L_1(\mathbb{P}_n)) = o_P^*(n)$ for every ε and $M > 0$, then $\|\mathbb{P}_n - P\|_{\mathcal{F}}^* \rightarrow 0$ both almost surely and in mean. In particular \mathcal{F} is a GC class.