STAT6018 Research Frontiers in Data Science

Topic II: Introduction to empirical process theory

Yu Gu, PhD
Assistant Professor

Department of Statistics & Actuarial Science
The University of Hong Kong

Table of Contents

Table of Contents

Glivenko-Cantelli (GC) class

Definition 1 (GC class)

A function class $\mathcal{F}$ is called P-GC if

$$\|\mathbb{P}_n - P\|_{\mathcal{F}} \xrightarrow{a.s.} 0$$

under the probability measure P.

- $\|Q\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |Qf|$
- uniform almost sure convergence across $\mathcal{F}$

GC theorem with bracketing

Bracket number $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$:

- minimum number of brackets $[\ell, u]$ with $\|\ell - u\| < \epsilon$ needed to cover $\mathcal{F}$
- entropy with bracketing: $\log N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$

Theorem 2 (GC with bracketing)

Let $\mathcal{F}$ be a class of P-measurable functions such that

$$N_{[]}(\epsilon, \mathcal{F}, L_1(P)) < \infty, \qquad \text{for every } \epsilon > 0.$$

Then $\mathcal{F}$ is P-GC.

GC theorem with bracketing (cont.)

Proof.

For every $f \in [\ell_i, u_i]$, we have

$$\begin{cases} (\mathbb{P}_n - P)f \leq \mathbb{P}_n u_i - P\ell_i \leq (\mathbb{P}_n - P)u_i + \|u_i - \ell_i\|_{L_1(P)} \\ (\mathbb{P}_n - P)f \geq \mathbb{P}_n \ell_i - Pu_i \geq (\mathbb{P}_n - P)\ell_i - \|u_i - \ell_i\|_{L_1(P)} \end{cases}$$

Thus,

$$\begin{cases} \sup_{f\in\mathcal{F}}(\mathbb{P}_n - P)f \leq \max_i(\mathbb{P}_n - P)u_i + \epsilon \overset{a.s.}{\to} \epsilon \\ \inf_{f\in\mathcal{F}}(\mathbb{P}_n - P)f \geq \min_i(\mathbb{P}_n - P)\ell_i - \epsilon \overset{a.s.}{\to} -\epsilon \end{cases} \quad \text{(by SLLN)}$$

$$\Rightarrow \limsup_n \|\mathbb{P}_n - P\|_{\mathcal{F}} \leq \epsilon \text{ almost surely.}$$

Letting $\epsilon \downarrow 0$ yields the desired result. $\square$

GC theorem without bracketing

Covering number $N(\epsilon, \mathcal{F}, \|\cdot\|)$:

- minimum number of balls $B(f;\epsilon) := \{g : \|g - f\| \le \epsilon\}$ needed to cover $\mathcal{F}$
- entropy without bracketing: $\log N(\epsilon, \mathcal{F}, \|\cdot\|)$

Envelope function F: $|f(x)| \le F(x)$ for every $x \in \mathcal{X}$ and $f \in \mathcal{F}$

Theorem 3 (GC without bracketing)

Let $\mathcal{F}$ be a class of P-measurable functions with envelope F such that $PF < \infty$. Let $\mathcal{F}_M$ be the class of functions $f\mathbb{1}\{F \le M\}$ when f ranges over $\mathcal{F}$. Then $\mathcal{F}$ is P-GC if and only if

$$n^{-1}\log N(\epsilon, \mathcal{F}_M, L_1(\mathbb{P}_n)) \xrightarrow{p} 0, \quad \forall \epsilon, M > 0.$$

GC theorem without bracketing (cont.)

Symmetrization (Theorem 1.26):

$$E\left\|\mathbb{P}_n - P\right\|_{\mathcal{F}} \leq 2E\left\|\mathbb{P}_n^o\right\|_{\mathcal{F}}$$

Proof of sufficiency.

$$\begin{aligned} E\left\|\mathbb{P}_n - P\right\|_{\mathcal{F}} &\leq 2E_X E_\varepsilon \left\| \frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}} && \text{(symmetrization)} \\ &\leq 2E_X E_\varepsilon \left\| \frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}_M} + 2P[F\mathbb{1}\{F > M\}] && \text{(triangle inequality)} \end{aligned}$$

For sufficiently large M, $P[F\mathbb{1}\{F > M\}]$ is arbitrarily small.

GC theorem without bracketing (cont.)

Maximal inequality for Rademacher linear combinations (Corollary 1.25):

$$E \max_{1 \le i \le N} |\xi_i| \le C\sqrt{\log N} \max_{1 \le i \le N} \|a^{(i)}\|$$

Proof of sufficiency (cont.)

Let $\mathcal{G}$ denote the ϵ-cover associated with $N(\epsilon, \mathcal{F}_M, L_1(\mathbb{P}_n))$. For any $f \in \mathcal{F}_M$,

$$\begin{aligned}
\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i f\left(X_i\right)\right| &\le \left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i g\left(X_i\right)\right| + \left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i \left[f\left(X_i\right) - g\left(X_i\right)\right]\right| \\
&\le \left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i g\left(X_i\right)\right\|_{\mathcal{G}} + \epsilon \\
&\le C\sqrt{\frac{\log N\left(\epsilon, \mathcal{F}_M, L_1\left(\mathbb{P}_n\right)\right)}{n}} \max_{g \in \mathcal{G}} \sqrt{\mathbb{P}_n g^2} + \epsilon \quad \text{(maximal inequality)} \\
&\xrightarrow{p} \epsilon
\end{aligned}$$

GC theorem without bracketing (cont.)

Proof of sufficiency (cont.)

Letting $\epsilon \downarrow 0$ yields $\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i f(X_i)\|_{\mathcal{F}} \xrightarrow{p} 0$. Since $|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i f(X_i)| \leq M$, it follows by the dominated convergence theorem that $E_X E_\varepsilon \left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{F}_M} \to 0$.
Thus, we conclude that $E\|\mathbb{P}_n - P\|_{\mathcal{F}} \to 0$.
By Lemma 2.4.5 of VW, $\|\mathbb{P}_n - P\|_{\mathcal{F}}$ is a reverse sub-martingale, thus converges almost surely to a constant, which must be 0 by the convergence in mean. $\square$

GC theorem with uniform covering

Corollary 4

Let $\mathcal{F}$ be a class of P-measurable functions with envelope F such that $PF < \infty$. Then $\mathcal{F}$ is P-GC if

$$\sup_Q N(\epsilon \|F\|_{L_1(Q)}, \mathcal{F}, L_1(Q)) < \infty, \quad \forall \epsilon > 0,$$

where the supremum is over all probability measures Q with $0 < QF < \infty$.

Proof.

Assume that $PF > 0$ (otherwise the result is trivial). There exists an $\eta \in (0, \infty)$ such that $1/\eta < \mathbb{P}_n F < \eta$ for all n large enough. For any $\epsilon > 0$, there exists a K_ϵ such that with probability 1,

$$\log N(\epsilon\eta, \mathcal{F}, L_1(\mathbb{P}_n)) \leq \log N(\epsilon \mathbb{P}_n F, \mathcal{F}, L_1(\mathbb{P}_n)) \leq K_\epsilon$$

for all n large enough. Thus, for any $\epsilon, M > 0$,

$$\log N(\epsilon, \mathcal{F}_M, L_1(\mathbb{P}_n)) \leq \log N(\epsilon, \mathcal{F}, L_1(\mathbb{P}_n)) = O_p(1).$$

The desired result follows by Theorem 3. □

Table of Contents