1. Application of GC theorem
   - Introdunction of least square regression
   - Theorem 3.20 and relative proof
2. Bounded differences inequality - a simple concentration inequality
   - Bounded differences inequality and relative proof
   - Application of Theorem 3.24
3. Supremum of the emperical process for a bounded class of functions
   - Use bounded differences inequality into emperical process
   - Classical Glivenko-Cantelli poblem

Application of GC theorem

Consistency of least square regression

$$Y_i = g_0(z_i) + W_i \qquad \text{for i = 1,2,...,n}$$

- $Y_i \in \mathbb{R}$ is the observed response variable.
- $z_i \in \mathcal{Z}$ is a covariate and W_i is the unobserved error.
- W_i is assumed to be independent random variables with $\mathbb{E}W_i = 0$ and $Var(W_i) \leq \sigma_0^2 < \infty$.
- The covariates $z_1, ..., z_n$ are fixed.

- The function $g_0 : \mathcal{Z} \to \mathbb{R}$ is unknown, but we assume that $g_0 \in \mathcal{G}$, where $\mathcal{G}$ is a given class of regression functions.

- The unknown regression function can be estimated by the least squares estimator (LSE) $\hat{g}_n$, which is defined by

$$\hat{g}_n = \arg\min_{g \in \mathcal{G}} \sum_{i=1}^{n} \left(Y_i - g\left(z_i\right)\right)^2$$

When can we say that $\left\|\hat{g}_n - g_0\right\|_n \xrightarrow{\mathbb{P}} 0$?

- $Q_n := \frac{1}{n}\sum_{i=1}^{n}\delta_{z_i}$ denote the empirical measure of the design points.

- We shall need to control the entropy of subclasses $\mathcal{G}_n(R)$, which are defined as

$$\mathcal{G}_n(R) = \{g \in \mathcal{G} : \|g - g_0\|_n \leq R\}$$

$\log N(\delta, G_n(R), L(Q$

- For $g : \mathcal{Z} \rightarrow \mathbb{R}$, we write $\|g\|_n^2 := \frac{1}{n}\sum_{i=1}^{n} g^2(z_i)$,

$$\|Y - g\|_n^2 := \frac{1}{n}\sum_{i=1}^{n}(Y_i - g(z_i))^2, \langle W, g\rangle_n := \frac{1}{n}\sum_{i=1}^{n} W_i g(z_i).$$

- $\|Y - \hat{g}_n\|_n^2 \leq \|Y - g_0\|_n^2 \quad \Rightarrow \quad \|\hat{g}_n - g_0\|_n^2 \leq 2\langle W, \hat{g}_n - g_0\rangle_n \qquad (1)$

$\|Y - \hat{g}_n - g_0 + g_0\|_n^2$

Theorem 3.20 Suppose that

$$\lim_{K \to \infty} \lim\sup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left(W_i^2 1_{\{|W_i| > K\}}\right) = 0$$

and

$$\frac{\log N\left(\delta, \mathcal{G}_n(R), L_1\left(Q_n\right)\right)}{n} \to 0, \qquad \text{for all } \delta > 0, R > 0$$

Then, $\left\|\hat{g}_n - g_0\right\|_n \xrightarrow{p} 0$.

Proof Theorem 3.20 :

Let $\eta, \delta > 0$ be given. We will show that $\mathbb{P}\left(\|\hat{g}_n - g_0\|_n > \delta\right)$ can be made arbitrarily small, for all n sufficiently large.

Note that for any $R > \delta$, we have

$$\mathbb{P}\left(\|\hat{g}_n - g_0\|_n > \delta\right) \leq \mathbb{P}\left(\delta < \|\hat{g}_n - g_0\|_n < R\right) + \mathbb{P}\left(\|\hat{g}_n - g_0\|_n > R\right)$$

We will first prove the second term. From (1), using Cauchy-Schwarz inequality $\|\hat{g}_n - g_0\|^2 \leq 2\langle W, \hat{g}_n - g_0\rangle_n \leq 2\|W\|_n \cdot \|\hat{g}_n - g_0\|_n$

Hence, it follows that

$|\langle u, v\rangle| \leq \|u\| \cdot \|v\|$

$$\|\hat{g}_n - g_0\|_n \leq 2\left(\frac{1}{n}\sum_{i=1}^{n} W_i^2\right)^{1/2}$$

Thus, using Markov's inequality,

$$\begin{aligned}\mathbb{P}\left(\|\hat{g}_n - g_0\|_n > R\right) &\leq \mathbb{P}\left(2\left(\frac{1}{n}\sum_{i=1}^{n} W_i^2\right)^{1/2} > R\right)\\ &\leq \frac{4}{R^2}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}W_i^2 \leq \frac{4\sigma_0^2}{R^2} = \eta\end{aligned}$$

where $R^2 := 4\sigma_0^2/\eta$.

Then we will prove the first term. Now, using (1) again,

$$\mathbb{P}\left(\delta < \|\hat{g}_n - g_0\|_n < R\right) \leq \mathbb{P}\left(\sup_{g \in \mathcal{G}_n(R)} 2\left\langle W, g - g_0\right\rangle_n \geq \delta^2\right)$$

$$\leq \mathbb{P}\left(\sup_{g \in \mathcal{G}_n(R)} \left\langle W 1_{\{|W| \leq K\}}, g - g_0\right\rangle_n \geq \frac{\delta^2}{4}\right) + \mathbb{P}\left(\sup_{g \in \mathcal{G}_n(R)} \left\langle W 1_{\{|W| > K\}}, g - g_0\right\rangle_n \geq \frac{\delta^2}{4}\right)$$

In this part we will prove $\mathbb{P}\left(\sup_{g\in\mathcal{G}_n(R)}\left\langle W1_{\{|W|>K\}}, g-g_0\right\rangle_n \geq \frac{\delta^2}{4}\right) \leq \eta$

Using cauchy-Schwarz inequality

$$\sup_{g\in\mathcal{G}_n(R)}\left\langle W1_{\{|W|>K\}}, g-g_0\right\rangle_n \leq \sup_{g\in\mathcal{G}_n(R)}\|W1_{\{|W|>K\}}\|_n \cdot \|g-g_0\|_n$$
$$= \left(\frac{1}{n}\sum_{i=1}^{n} W_i^2 1_{\{|W_i|>K\}}\right)^{1/2} \cdot R$$

Using Markov's inequality:

$$\mathbb{P}\left(\sup_{g\in\mathcal{G}_n(R)}\left\langle W1_{\{|W|>K\}}, g-g_0\right\rangle_n \geq \frac{\delta^2}{4}\right) \leq \mathbb{P}\left(\left(\frac{1}{n}\sum_{i=1}^{n} W_i^2 1_{\{|W_i|>K\}}\right)^{1/2} \geq \frac{\delta^2}{4R}\right)$$
$$\leq \left(\frac{4R}{\delta^2}\right)^2 \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} W_i^2 1_{\{|W_i|>K\}}\right) \leq \eta$$

by choosing $K = K(\delta,\eta)$ sufficiently large and using

$$\lim_{K\to\infty}\lim\sup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(W_i^2 1_{\{|W_i|>K\}}\right) = 0$$

This part we will prove $\mathbb{P}\left(\sup_{g\in\mathcal{G}_n(R)} \left\langle W1_{\{|W|\leq K\}}, g-g_0\right\rangle_n \geq \frac{\delta^2}{4}\right) \leq \frac{4\eta}{\delta^2}$

Using Markov's inequality

$$\mathbb{P}\left(\sup_{g\in\mathcal{G}_n(R)} \left\langle W1_{\{|W|\leq K\}}, g-g_0\right\rangle_n \geq \frac{\delta^2}{4}\right) \leq \frac{4}{\delta^2}\mathbb{E}\| \left\langle W1_{\{|W|\leq K\}}, g-g_0\right\rangle_n \|_{\mathcal{G}_n(R)}$$

Next proof will mimic to proof of **Theoroem 3.5**, and get

$$\frac{4}{\delta^2}\mathbb{E}\| \left\langle W1_{\{|W|\leq K\}}, g-g_0\right\rangle_n \|_{\mathcal{G}_n(R)} \leq \eta$$

Bounded differences inequality

We are interested in bounding the random fluctuations of functions of many independent random variables.

Let $X_1, \ldots, X_n$ be independent random variables taking values in $\mathcal{X}$.

Let $f \colon \mathcal{X}^n \to \mathbb{R}$, and let $Z = f(X_1, \ldots, X_n)$ be the random variable of interest.

We seek upper bounds for

$$\mathbb{P}(Z > \mathbb{E}Z + t) \quad \text{and} \quad \mathbb{P}(Z < \mathbb{E}Z - t) \quad \text{for } t > 0$$

Recall :

Lemma 3.9 (Hoeffding's inequality). Let $X_1, \dots, X_n$ be independent bounded random variables such that $X_i \in [a_i, b_i]$ with probability 1. $Z := S_n = \sum_{i=1}^n X_i$. Then, we obtain,

$$\mathbb{P}\left(S_n - \mathbb{E}S_n \geq t\right) \leq e^{-2t^2/\sum_{i=1}^n (b_i - a_i)^2}$$

and

$$\mathbb{P}\left(S_n - \mathbb{E}S_n \leq -t\right) \leq e^{-2t^2/\sum_{i=1}^n (b_i - a_i)^2}$$

Theorem 3.24 (Bounded differences inequality or McDiarmid's inequality). Suppose that $Z = f(X_1, \dots, X_n)$ and f is a function with bounded differences, then

$$\mathbb{P}(|Z - \mathbb{E}(Z)| > t) \leq 2e^{-2t^2 / \sum_{i=1}^n c_i^2}$$

Definition 3.23 (Functions with bounded differences). We say that a function $f \colon \mathcal{X}^n \to \mathbb{R}$ has the bounded difference property if for some nonnegative constants $c_1, \dots, c_n$,

$$\sup_{x_1, \dots, x_n, x_i' \in \mathcal{X}} |f(x_1, \dots, x_n) - f(x_1, \dots, x_{i-1}, x_i', x_{i+1}, \dots, x_n)| \leq c_i, \quad 1 \leq i \leq n$$

Proof Theorem 3.24 :

Here we try to express $Z - \mathbb{E}(Z)$ as a sum of variables.

Let $X_1, \ldots, X_n$ be independent random variables taking values in $\mathcal{X}$. Let $f : \mathcal{X}^n \to \mathbb{R}$ and

$$Z = f(X_1, \ldots, X_n)$$

be the random variable of interest.

Martingale

Given a sequence $\{Y_k\}_{k=1}^{\infty}$ of random variables adapted to a filtration $\{\mathcal{F}_k\}_{k=1}^{\infty}$ (e.g., $\mathcal{F}_k = \sigma(X_1, \ldots, X_k)$), the pair $\{Y_k, \mathcal{F}_k\}_{k=1}^{\infty}$ is a martingale if, for all $k \geq 1$,

$$\mathbb{E}[|Y_k|] < \infty, \quad \text{and} \quad \mathbb{E}[Y_{k+1} \mid \mathcal{F}_k] = Y_k.$$

Note that if we define

$$Y_k := \mathbb{E}\left[Z \mid X_1, \dots, X_k\right], \qquad \text{for } k = 1, \dots, n$$

then $\{Y_k\}_{k=0}^n$ is a martingale adapted to a filtration generated by $\{X_k\}_{k=1}^n$.

Denote by $\mathbb{E}_i[\cdot] := \mathbb{E}\left[\cdot \mid X_1, \dots, X_i\right]$. Thus, $\mathbb{E}_0(Z) = \mathbb{E}(Z)$, $\mathbb{E}_k(Z) = Y_k$ and $\mathbb{E}_n(Z) = Z$, for $k = 1, \dots, n$. Writing

$$\Delta_i := \mathbb{E}_i[Z] - \mathbb{E}_{i-1}[Z]$$

we have

$$Z - \mathbb{E}Z = \sum_{i=1}^{n} \Delta_i$$

Lemma 3.23 (Azuma-Hoeffding inequality) Let $\{Y_0, Y_1, \cdots\}$ be a martingale with respect to filtration $\{\mathcal{F}_0, \mathcal{F}_1, \cdots\}$.

Assume there are predictable processes $\{A_0, A_1, \cdots\}$ and $\{B_0, B_1, \ldots\}$ with respect to $\{\mathcal{F}_0, \mathcal{F}_1, \cdots\}$, i.e. for all i, A_i, B_i are $\mathcal{F}_{i-1}$-measurable, and constants $0 < c_1, c_2, \cdots < \infty$.

Such that $A_i \le Y_i - Y_{i-1} \le B_i$ and $B_i - A_i \le c_i$ almost surely. Then for all $\epsilon > 0$,

$$\mathrm{P}\left(Y_n - Y_0 \ge \epsilon\right) \le \exp\left(-\frac{2\epsilon^2}{\sum_{t=1}^{n} c_i^2}\right)$$

We use Lemma 3.23 to prove Theorem 3.24

We define

$$A_i = \inf_x \mathbb{E}\left[Z \mid X_1, \ldots, X_{i-1}, x\right] - \mathbb{E}\left[Z \mid X_1, \ldots, X_{i-1}\right]$$
$$= \inf_x \int f(X_1, \ldots, X_{i-1}, x, x_{i+1}, \ldots, x_n)\, dP(x_{i+1}) \cdots dP(x_n) - \mathbb{E}_{i-1}[\cdot]$$
$$B_i = \sup_x \mathbb{E}\left[Z \mid X_1, \ldots, X_{i-1}, x\right] - \mathbb{E}\left[Z \mid X_1, \ldots, X_{i-1}\right]$$
$$= \sup_x \int f(X_1, \ldots, X_{i-1}, x, x_{i+1}, \ldots, x_n)\, dP(x_{i+1}) \cdots dP(x_n) - \mathbb{E}_{i-1}[\cdot]$$

then we have

$$A_i \leq \Delta_i \leq B_i \quad \text{a.s. } \forall i = 1, \ldots, n$$

We need to bound the quantity $B_i - A_i$. By independence of the X_i and the bounded difference assumption

$$B_i - A_i = \sup_x \mathbb{E}\left[Z \mid X_1, \ldots, X_{i-1}, x\right] - \inf_x \mathbb{E}\left[Z \mid X_1, \ldots, X_{i-1}, x\right]$$
$$= \sup_{x,x'} \int \Big(f(X_1, \ldots, X_{i-1}, x, x_{i+1}, \ldots, x_n)$$
$$- f(X_1, \ldots, X_{i-1}, x', x_{i+1}, \ldots, x_n) \Big) dP(x_{i+1}) \cdots dP(x_n)$$
$$\leq c_i$$

Application of Theorem 3.24

Kernel density estimation

Let $X_1, \ldots, X_n$ are i.i.d from a distribution P on $\mathbb{R}$ with density ϕ.

We want to estimate ϕ nonparametrically using the kernel density estimator (KDE) $\hat{\phi}_n : \mathbb{R} \to [0, \infty)$ defined as

$$\hat{\phi}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right), \qquad \text{for } x \in \mathbb{R}$$

- $h_n > 0$ is the smoothing bandwidth.
- K is a nonnegative kernel (i.e., $K \geq 0$ and $\int K(x)dx = 1$).

The L_1-error of the estimator $\hat{\phi}_n$ is

$$Z \equiv f(X_1, \ldots, X_n) := \int \left| \hat{\phi}_n(x) - \phi(x) \right| dx$$

- The random variable Z provides a measure of the difference between $\hat{\phi}_n$ and ϕ.
- Z also captures the difference between P_n and P in the total variation distance. $(Z = 2 \sup_A |P_n(A) - P(A)|)$

We now use Theorem 3.24 to get exponential tail bounds for Z.

For $x_1, \ldots, x_n, x_i' \in \mathcal{X}$

$$
\begin{aligned}
&|f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \\
&= \left| \int |\hat{\phi}_{n1}(x) - \phi(x)| dx - \int |\hat{\phi}_{n2}(x) - \phi(x)| dx \right| \\
&\leq \left| \int |\hat{\phi}_{n1}(x) - \hat{\phi}_{n2}(x)| dx \right| \\
&\leq \frac{1}{nh_n} \int \left| K\left(\frac{x - x_i}{h_n}\right) - K\left(\frac{x - x_i'}{h_n}\right) \right| dx \leq \frac{2}{n}
\end{aligned}
$$

$||a| - |b|| \leq |a-b|$

$\int k(x)\,dx = 1$

Thus, using Theorem 3.24 with $c_i = 2/n$, for all $i = 1, \ldots, n$.

$$\mathbb{P}(|Z - \mathbb{E}(Z)| > t) \leq 2e^{-nt^2/2} \quad \Rightarrow \quad \mathbb{P}(\sqrt{n}|Z - \mathbb{E}(Z)| > t) \leq 2e^{-t^2/2}$$

Z concentrates around its expectation $\mathbb{E}[Z]$ at the rate $n^{-1/2}$.

Supremum of the emperical process for a bounded class of functions

$$Z := \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E}\left[f(X_1)\right] \right|$$

- $X_1, \ldots, X_n$ are i.i.d. random objects taking values in $\mathcal{X}$
- $\mathcal{F}$ is a collection of real-valued functions on $\mathcal{X}$.
- $\mathcal{F}$ is assumed that all functions in $\mathcal{F}$ are bounded by a positive constant B, i.e.,

$$\sup_{x \in \mathcal{X}} |f(x)| \leq B \quad \text{for all } f \in \mathcal{F}$$

Let

$$g(x_1, \ldots, x_n) := \left| \frac{1}{n} \sum_{i=1}^{n} f(x_i) - \mathbb{E}[f(X_1)] \right|$$

Next, find the bound of effect of i_{th} variable on function g.

$$\begin{aligned} g(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n) &= \left| \frac{1}{n} \sum_{j \neq i} f(x_j) + \frac{f(x_i')}{n} - \mathbb{E}[f(X_1)] \right| \\ &= \left| \frac{1}{n} \sum_{j=1}^{n} f(x_j) - \mathbb{E}[f(X_1)] + \frac{f(x_i')}{n} - \frac{f(x_i)}{n} \right| \\ &\leq \left| \frac{1}{n} \sum_{j=1}^{n} f(x_j) - \mathbb{E}[f(X_1)] \right| + \frac{2B}{n} \\ &\leq g(x_1, \ldots, x_n) + \frac{2B}{n} \end{aligned}$$

Then, use **Theorem 3.24** with $c_i = 2B/n$ for $i = 1, \dots, n$

$$\mathbb{P}(|Z - \mathbb{E}Z| > t) \leq 2\exp\left(-\frac{nt^2}{2B^2}\right), \quad \text{for every } t \geq 0$$

Setting $\delta := \exp\left(-\frac{nt^2}{2B^2}\right)$, we can deduce that

$$|Z - \mathbb{E}[Z]| \leq B\sqrt{\frac{2}{n}\log\frac{1}{\delta}}$$

holds with probability at least $1 - 2\delta$ for every $\delta > 0$. This inequality implies that $\mathbb{E}[Z]$ is usually the dominating term for understanding the behavior of Z.

Theorem 3.26 Suppose that $X_1, \ldots, X_n$ are *i.i*.d. random variables on $\mathbb{R}$ with distribution P and c.d.f. F. Let $\mathbb{F}_n$ be the empirical d.f. of the data. Then,

$$\mathbb{P}\left[\|\mathbb{F}_n - F\|_\infty \geq 8\sqrt{\frac{\log(n+1)}{n}} + t\right] \leq e^{-nt^2/2}, \qquad \text{forall } t > 0.$$

Hence, $\|\mathbb{F}_n - F\|_\infty \overset{\text{a.s.}}{\to} 0$.

Proof :

- The function class is $\mathcal{F} := \left\{ 1_{(-\infty,t]}(\cdot) : t \in \mathbb{R} \right\}$.
- $Z := \|\mathbb{P}_n - P\|_{\mathcal{F}} = \|\mathbb{F}_n - F\|_\infty$ ($\mathbb{F}_n = \frac{1}{n}\sum_{i=1}^n 1_{(-\infty,x]}(X_i)$).
- We have to bound upper bound $\mathbb{E}[Z]$ via symmetrization, i.e., $\mathbb{E}[Z] \le 2\mathbb{E}_X \left[\mathbb{E}_\varepsilon \left[\sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i) \right| \right] \right]$, where $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. Rademachers independent of the X_i 's.
  (Rademachers random varibale ε take values ± 1 with equal probability $1/2$)
- For a fixed $(x_1, \ldots, x_n) \in \mathbb{R}^n$, define

$$\Delta_n(\mathcal{F}; x_1, \ldots, x_n) := \{(f(x_1), \ldots, f(x_n)) : f \in \mathcal{F}\}$$

Observe that although $\mathcal{F}$ has uncountable many functions, for every $(x_1, \ldots, x_n) \in \mathbb{R}^n$, $\Delta_n(\mathcal{F}; x_1, \ldots, x_n)$ can take at most $n+1$ distinct values.

Thus, $\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(x_i) \right|$ is at most the supremum of $n+1$ such variables, and we can apply Lemma 3.16 to show that

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(X_i) \right| \right] \leq 4\sqrt{\frac{\log(n+1)}{n}}$$

This can show

$P[Z - EZ \geq t]$

$$\mathbb{P}\left[\|\mathbb{F}_n - F\|_\infty \geq 8\sqrt{\frac{\log(n+1)}{n}} + t \right] \leq e^{-nt^2/2}, \quad \text{forall } t > 0.$$

This implies $\|\mathbb{F}_n - F\|_\infty \overset{\text{a.s.}}{\to} 0$.