SDS 387
Linear Models

Fall 2024

Lecture 10 - Thu, Sep 26, 2024

Instructor: Prof. Ale Rinaldo

- Fixed some typos on HW 1. I will upload the new version after class.

- Last time: Berry - Esseen bounds.

- More general ways to think about CLT and more generally Gaussian approximations.

- First, let's talk about some probability distances.

  Let $\mathcal{F}$ be a class of test functions. Then we can consider integral probability metrics of the form

$$\underset{\text{distance}}{d_{\mathcal{F}}}(P_X, P_Y) = \sup_{f \in \mathcal{F}} \left| \mathbb{E}_{X \sim P_X}[f(X)] - \mathbb{E}_{Y \sim P_Y}[f(Y)] \right|$$

- The choice of $\mathcal{F}$ dictates the property of metric.

- A natural choice in $\mathbb{R}$ is $\mathcal{F} = \{ \mathbb{1}_{(-\infty, x]}, x \in \mathbb{R} \}$

This gives the Kolmogorov-Smirnov distance.

$$KS(P_X, P_Y) = \sup_{z \in \mathbb{R}} | \underbrace{F_X(z)}_{\text{cdf of } X} - \underbrace{F_Y(z)}_{\text{cdf } Y} |$$

- Let $\mathcal{F} = \{ f : \mathbb{R} \xrightarrow{\text{measurable}} [0,1] \}$ Then we obtain

$$\underbrace{d_{TV}}_{\substack{\text{Total variation} \\ \text{distance}}} (P_X, P_Y) = \sup_{\substack{B \subseteq \mathbb{R} \\ \text{Borel-measurable}}} | \mathbb{P}(X \in B) - \mathbb{P}(Y \in B) |$$

$$= \frac{1}{2} \int_{\mathbb{R}} | \underbrace{f_X(z) - f_Y(z)}_{\text{densities of } X \text{ and } Y} | dz$$

HW

= sum of type I and type II error for testing the null hypothesis that say $X \sim P_X$ vs $X \sim P_Y$

It is also possible to show that

$$d_{TV}(P_X, P_Y) = \inf_{\substack{\text{all coupling} \\ \text{of } X, Y}} \mathbb{P}(X \neq Y)$$

→ joint distributions for (X, Y) s.t. $X \sim P_X$, $Y \sim P_Y$

- Wasserstein's distance in $\mathbb{R}$ is obtained by considering

$$\mathcal{F} = \{ f : \mathbb{R} \to \mathbb{R} \quad 1\text{- Lipschitz} \}$$

$$\Downarrow$$

$$|f(x) - f(y)| \le |x - y|$$

$\forall x, y$ in the domain of f.

Note: Wasserstein distance metrizes convergence in distribution

$Wass(P_{X_n}, P_X) \to 0$ as $n \to \infty$ is equivalent to

i) $X_n \xrightarrow{d} X$ and ii) $\mathbb{E}|X_n| \to \mathbb{E}|X|$ as $n \to \infty$

- Relationship btw KS and Wass. It is possible to show that

HW $\leftarrow$ $$KS(P_X, P_Y) \le 2\sqrt{C \, Wass(P_X, P_Y)}$$ where

$$C = \|f_Y\|_\infty$$

↳ density of Y

$\Downarrow$

One can establish a Berry - Esseen bounds using convergence in Wasserstein distance but the rate is sub-optimal. But dealing with Wasserstein distance is more convenient because $\mathcal{F}$ is "better behaved".

- A Generalization of the Lindeberg Principle

  by S. Chatterjee, Annals of Probability, 2006

- Let X and Y be random vectors in $\mathbb{R}^n$ with Y having independent components. For $i = 1, \dots, n$ let

$A_i = B_i = 0$ if x_i's are indep and $\mathbb{E}[X_i] = \mathbb{E}[Y_i]$ and $\mathbb{E}[X_i^2] = \mathbb{E}[Y_i^2]$

$$\left\{ \begin{aligned} A_i &= \mathbb{E}\Big[\Big| \mathbb{E}[X_i \mid X_1, \dots, X_{i-1}] - \mathbb{E}[Y_i] \Big|\Big] \\ B_i &= \mathbb{E}\Big[\Big| \mathbb{E}[X_i^2 \mid X_1, \dots, X_{i-1}] - \mathbb{E}[Y_i^2] \Big|\Big] \end{aligned} \right.$$

Let $M_3 = \max_{i=1,\dots,n} \left\{ \mathbb{E}[|X_i|^3] + \mathbb{E}[|Y_i|^3] \right\}$. Then

$$\Big| \mathbb{E}[f(X)] - \mathbb{E}[f(Y)] \Big| \leq \sum_{i=1}^n \left(A_i L_1(f) + \frac{1}{2} B_i L_2(f) \right) + \frac{1}{6} n M_3 L_3(f)$$

where for $j = 1, 2, 3$

$$L_j(f) = \sup_x \max_{i=1,\dots,n} \left| \frac{\partial^j f}{\partial x_i^j}(x) \right| < \infty$$

wher $f: \mathbb{R}^n \to \mathbb{R}$ that is 3 times continuously diff.

- Remark if $A_i = B_i = 0$ all i the bound becomes $\frac{n}{6} M_3 L_3(f)$ which become

$$\frac{M_3 L_3(g)}{6\sqrt{n}} \quad \text{if} \quad f(x) = g\left(\tfrac{1}{\sqrt{n}} \sum x_i\right)$$

$\hookrightarrow$ smooth

Pf/ The proof is based on the leave one out method or Lindeberg swapping or smoothing that can be used in, e.g., Banach spaces.

<u>Notation</u> let $\partial_i f$ be $\frac{\partial f}{\partial x_i}$

$\partial_{ij} f$ be $\frac{\partial^2 f}{\partial x_i \partial x_j}$

Let $Z_i = (X_1, \ldots, X_i, Y_{i+1}, \ldots, Y_n)$

$Z_i^\circ = (X_1, \ldots, X_{i-1}, 0, Y_{i+1}, \ldots, Y_n)$

Then $Z_0 = Y$ and $Z_n = X$ and by telescoping

$$\mathbb{E}[f(X)] - \mathbb{E}[f(Y)] = \sum_{i=1}^{n} \Big(\underbrace{\mathbb{E}[f(Z_i)] - \mathbb{E}[f(Z_{i-1})]}_{\text{we need to handle each of these terms separately}} \Big)$$

We do a 3rd order Taylor series expansion of each term in the sum.

$$\left| f(Z_i) - f(Z_i^\circ) - X_i \partial_i f(Z_i^\circ) - \frac{X_i^2}{2} \partial_{ii} f(Z_i^\circ) \right| \le \frac{|X_i|^3 L_3(f)}{6}$$

similarly

$$\left| f(Z_{i-1}) - f(Z_i^\circ) - Y_i \, \partial_i f(Z_i^\circ) - \frac{Y_i^2}{2} \partial_{ii} f(Z_i^\circ) \right|$$

$$\leq \frac{|Y_i|^3 L_3(f)}{6}$$

By independence of the Y_i's:

$$\mathbb{E}\left[(X_i - Y_i)\, \partial_i f(Z_i^\circ)\right] = \mathbb{E}\left[\mathbb{E}\left[(X_i - Y_i)\, \partial_i f(Z_i^\circ) \,\middle|\, \begin{matrix} X_1, \dots, X_{i-1} \\ Y_{i+1}, \dots, Y_n \end{matrix}\right]\right]$$

$$= \mathbb{E}\left[\left(\mathbb{E}[X_i \mid X_1, \dots, X_{i-1}] - \mathbb{E}[Y_i]\right) \partial_i f(Z_i^\circ)\right]$$

similarly

$$\mathbb{E}\left[(X_i^2 - Y_i^2)\, \partial_{ii} f(Z_i^\circ)\right] = \mathbb{E}\left[\left(\mathbb{E}[X_i^2 \mid X_1, \dots, X_{i-1}] - \mathbb{E}[Y_i^2]\right) \partial_{ii} f(Z_i^\circ)\right]$$

So, for all i.

$$\left| \mathbb{E} f(Z_i) - \mathbb{E}[f(Z_{i-1})] \right| \leq A_i L_1(f) + B_i L_2 f + \frac{L_3(f)}{6}\left[|X_i|^3 + |Y_i|^3\right]$$

The result follows by summing over i and by triangle inequality ∎

Q: Can you get a CLT (an asymptotic statement) from this result?

i.e. can you prove that

$$\left| \mathbb{P}\left(\frac{\sum X_i}{B_n} \leq x \right) - \mathbb{P}(Z \leq x) \right| \to 0$$

as $n \to \infty$?

Yes HW!