SDS 387
Linear Models

Fall 2025

Lecture 12 - Thu, Oct 9, 2025

Instructor: Prof. Ale Rinaldo

- Comment about example of using Berry-Esseen bound for sum independent Bernoulli. The bound is

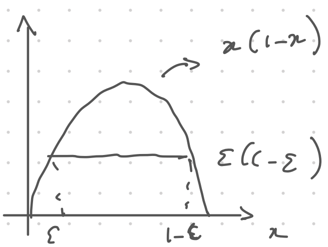
$$C \frac{\sum_{i=1}^n p_i(1-p_i)}{\left(\sum_{i=1}^n p_i(1-p_i)\right)^{3/2}} \leq C \cancel{\frac{n\,\varepsilon(1-\varepsilon)}{(n\,\varepsilon(1-\varepsilon))^{3/2}}}$$

wrong

$\varepsilon < p_i < 1-\varepsilon$ all $i$, $\varepsilon \in (0, 1/2)$

$$= C \frac{1}{\sqrt{\sum_{i=1}^n p_i(1-p_i)}}$$

$$\leq C \frac{1}{\sqrt{n \min_i p_i(1-p_i)}}$$




$$\leq C \frac{1}{\sqrt{n\,\varepsilon(1-\varepsilon)}}$$

because $p_i(1-p_i) \geq \varepsilon(1-\varepsilon)$ all $i$ due to concavity of the function $x \in [\varepsilon, 1-\varepsilon] \mapsto x(1-x)$

- Weakening of this bound. Assuming only $2+\delta$ moments where $\delta \in (0,1]$ (i.e. $\mathbb{E}\left[|X_i - \mathbb{E}[X_i]|^{2+\delta}\right] < \infty$ all $i$.)

Then

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left( \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{B_n} \le x \right) - \underbrace{\Phi(x)}_{\mathbb{P}(Z \le x),\ \sim N(0,1)} \right| \le C \frac{\sum_{i=1}^n \mathbb{E}|X_i - \mathbb{E}[X_i]|^{2+\delta}}{B_n^{2+\delta}}$$

$$B_n^2 = \sum_{i=1}^n \mathrm{Var}[X_i]$$

- It is useful to rewrite the Berry-Esseen bound as

centered

$$\sup_x \left| \mathbb{P}\left( \frac{\sum_i X_i}{B_n} \le x \right) - \mathbb{P}\left( \underbrace{\frac{\sum_i Z_i}{B_n}}_{\sim N(0,1)} \le x \right) \right| \le C \frac{\sum_i \mathbb{E}|X_i|^3}{B_n^3}$$

where $Z_1, \ldots, Z_n$ indep. Gaussians with

$$\mathbb{E}[Z_i] = \mathbb{E}[X_i]$$
$$\mathrm{Var}[Z_i] = \mathrm{Var}[X_i]$$

or

$$\sup_{A \in \mathcal{A}} \left| \mathbb{P}\left( \frac{\sum_i X_i}{B_n} \in A \right) - \mathbb{P}\left( \frac{\sum_i Z_i}{B_n} \in A \right) \right| \le \cdots$$

$$\mathcal{A} = \{ (-\infty, x],\ x \in \mathbb{R} \}$$

- Now suppose we want to get an analogous result in $\mathbb{R}^d$, assuming now that

$$X_i \sim (0, \Sigma_i) \quad \text{independent}$$

$$Z_i \sim N(0, \Sigma_i) \quad \text{independent}$$

Now let $\mathcal{A}$ be a class of subsets in $\mathbb{R}^d$.

There are many choices for $\mathcal{A}$:

$\mathcal{A}$: convex sets

$\mathcal{A}$: balls or ellipsoids

$\mathcal{A}$: hyper-rectangles

- Bentkus (2005) + Raič (2019). Let

$$\Sigma_n = \frac{1}{n} \sum_{i=1}^{n} \Sigma_i$$

Assume some regularity conditions on $\mathcal{A}$: in particular if $A \in \mathcal{A}$ then $A^\varepsilon = \{x \in \mathbb{R}^d, d(x, A) \leq \varepsilon\} \in \mathcal{A}$

$\inf_{y \in A} \|x - y\|$

↳ Note $\mathcal{A}$ cannot be the class of hyper-rectangles

$$\sup_{A \in \mathcal{A}} \left| \mathbb{P}\left( \frac{\sum_i X_i}{\sqrt{n}} \in A \right) - \mathbb{P}\left( \frac{\sum_i Z_i}{\sqrt{n}} \in A \right) \right| \leq$$

$$\leq C(\mathcal{A}, d) \frac{1}{\sqrt{n}} \left( \sum_{i=1}^{n} \frac{\mathbb{E} \| \Sigma_n^{-1/2} X_i \|^3}{n} \right)$$

where $C(\mathcal{A}, d)$ is the Gaussian isoperimetric constant of $\mathcal{A}$:

$\forall A \in \mathcal{A}$

$$\left.\begin{array}{l} \mathbb{P}(\overset{\to N(0, I_d)}{Z} \in A^{\varepsilon} \setminus A) \\ \mathbb{P}(Z \in A \setminus A^{-\varepsilon}) \end{array}\right\} \leq C(\mathcal{A}, d)\, \varepsilon$$

If $\mathcal{A}$ is sufficiently regular then

$$\int_{\partial A} \underbrace{\phi(z)}_{\text{pdf of } N(0, I_d)} \, d \underbrace{\mathcal{H}^{d-1}}_{\text{Hausdorff measure}} \leq C(\mathcal{A}, d)$$

- If $\mathcal{A}$ is

| | |
|---|---|
| class of balls in $\mathbb{R}^d$ | $C(\mathcal{A}, d) = 1$ |
| convex sets | $d^{1/4}$ |
| hyper-rectangles | $\log d$ |

- Assume $\mathcal{A}$ class of convex sets. Then the Berry-Esseen bound is of order

$$\frac{d^{1/4}}{\sqrt{n}} \sum_{i=1}^{n} \frac{\mathbb{E} \| \Sigma_n^{-1/2} X_i \|^{3 \cdot 2}_{2}}{n}$$

To lower bound it notice that

by Jensen

$$\geq \frac{d^{1/4}}{\sqrt{n}} \left( \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \| \Sigma_n^{-1/2} X_i \|^2 \right)^{3/2}$$

$\|x\|^2 = x^T x$
$\Leftrightarrow$
$\|A^{-1/2} x\|^2 = x^T A^{-1} x$

$$= \frac{d^{1/4}}{\sqrt{n}} \left( \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left( X_i^T \Sigma_n^{-1} X_i \right) \right)^{3/2}$$

$x^T A^{-1} x = \mathrm{tr}(x^T A^{-1} x)$
$= \mathrm{tr}(A^{-T} x x^T)$

cyclicity of $\mathrm{tr}(\cdot)$

$$= \frac{d^{1/4}}{\sqrt{n}} \left( \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ \mathrm{tr} \left( \Sigma_n^{-1} X_i X_i^T \right) \right] \right)^{3/2}$$

$$= \frac{d^{1/4}}{\sqrt{n}} \left( \frac{1}{n} \sum_{i=1}^{n} \left[ \mathrm{tr} \left( \Sigma_n^{-1} \underbrace{\mathbb{E}[X_i X_i^T]}_{\Sigma_i} \right) \right] \right)^{3/2}$$

$$= \frac{d^{1/4}}{\sqrt{n}} \left( \mathrm{tr} \left( \Sigma_n^{-1} \underbrace{\frac{1}{n} \sum_{i=1}^{n} \Sigma_i}_{\Sigma_n} \right) \right)^{3/2}$$

$$= \frac{d^{1/4}}{\sqrt{n}} \left( \mathrm{tr} \left( \underbrace{\Sigma_n^{-1} \Sigma_n}_{I_d} \right) \right)^{3/2}$$

$$= \frac{d^{1/4}}{\sqrt{n}} d^{3/2} = \frac{d^{7/4}}{\sqrt{n}} = \frac{d}{n^{2/4}} \to 0$$

as $n, d \to \infty$

if $d = o(n^{2/7})$

- Remarkable result: high-dim Berry-Esseen bound for $\mathcal{A}$ = hyper-rectangles

$\|x\|_\infty = \max_i |x_i|$

of order

$$\frac{1}{\sqrt{n}} \cdot \frac{\log^2 d}{\sigma_{min}(\Sigma_n)} \times \text{moment conditions}$$

$$\Downarrow$$

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \|X_i\|_\infty^b$$

$b = 3, 4$

- Recall that we view Berry-Esseen bound as a probabilistic bound on the difference between

$$\frac{\sum X_i}{\sqrt{n}} \quad \text{and} \quad \frac{\sum Z_i}{\sqrt{n}}$$

$X_i$ and $Z_i$ have matching 2 moments and the $Z_i$'s are Gaussians

Chatterjee's result says that

$$\left| \mathbb{E}\left[f(X_1, \dots, X_n)\right] - \mathbb{E}\left(f(Z_1, \dots, Z_n)\right) \right| \le \frac{1}{6} \underbrace{n}_{= \text{number } \#X_i} L_3(f) M_3$$

where $M_3 = \max_i \left\{ \mathbb{E}|X_i|^3 + \mathbb{E}|Z_i|^3 \right\}$

$$L_3(f) = \sup_{x \in \mathbb{R}^n} \max_i \left| \frac{\partial^3}{\partial x_i^3} f(x) \right|$$

When $f(x) = g\left(\frac{1}{\sqrt{n}} \sum x_i\right)$ the bound becomes $\frac{L_3(g) M_3}{\sqrt{n}}$

A Generalization of the Lindeberg Principle

Annals of Prob. 2006