1

LECTURE 2

Quick review of weak, vague, dist. convergence:

Def: (Klenke 13.12) For $\mu_1, \mu_2, \ldots, \mu_\infty$ finite measures on $\mathbb{R}$,

- $\mu_n \longrightarrow \mu_\infty$ <u>weakly</u> if $\int f \, d\mu_n \to \int f \, d\mu_\infty$ $\forall f$ cts, bdd.
- $\mu_n \longrightarrow \mu_\infty$ <u>vaguely</u> if " " , compact support.

Rk: Sensible for $\mathbb{R} \rightsquigarrow$ general metric space.

Ex: $\mu_n = \delta_n$, $\xrightarrow{\text{vague}}$ zero measure $= \mu_\infty$, but not weakly:

$\int 1 \, d\mu_n = \mu_n(\mathbb{R}) = 1 \not\to \mu_\infty$.

only obstruction

Thm: (K13.16) TFAE:

- $\mu_n \longrightarrow \mu_\infty$ weakly
- $\mu_n \longrightarrow \mu_\infty$ vaguely and $\int 1 \, d\mu_n = \mu_n(\mathbb{R}) \to \mu_\infty(\mathbb{R}) = \int 1 \, d\mu_\infty$

<u>Always true</u> if all μ_i prob. measures. If so, TFAE:

- $\int f \, d\mu_n \longrightarrow \int f \, d\mu_\infty$ $\forall f$ cts, bdd
- " " " cts, compact support
- " " " smooth, compact support
- " " " $f(x) = \begin{cases} p(x) & \text{for } |x| \le K \\ 0 & \text{otherwise} \end{cases}$ for <u>polynomial</u> p, (Weierstrass approx.)

(↪ many other conditions, like these.)

- $\mu_n(A) \longrightarrow \mu_\infty(A)$ $\forall$ measurable A with $\mu_\infty(\partial A) = 0$.
- $\mu_n((-\infty, x]) \longrightarrow \mu_\infty((-\infty, x])$ whenever $\mu_\infty(\{x\}) = 0$.

Def: X_n, X_∞ random variables, then $X_n \longrightarrow X_\infty$ if $\underset{(\mu_n)}{\mathrm{Law}(X_n)} \xrightarrow{\text{(weak)}} \underset{(\mu_\infty)}{\mathrm{Law}(X_\infty)}$

Translate above conditions:

$\int f \, d\mu_n \longrightarrow \int f \, d\mu_\infty \quad \longleftrightarrow \quad \mathbb{E} f(X_n) \longrightarrow \mathbb{E} f(X_\infty)$

$\mu_n(A) \longrightarrow \mu_\infty(A) \quad \longleftrightarrow \quad \mathbb{P}[X_n \in A] \longrightarrow \mathbb{P}[X_\infty \in A]$.

Rk: Continuity thm $\Rightarrow$ another sufficient set of "test functions" f is $f(x) = \exp(itx)$ for $t \in \mathbb{R}$.

Lindeberg approach to CLT: will prove a "weak" CLT,

Thm: If $(X_i)_{i \geq 1}$ iid w/ $\mathbb{E}X_i = \mu$, $\operatorname{Var} X_i = \sigma^2$, $\mathbb{E}|X_i|^3 < \infty$, $\hat{S}_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \Longrightarrow \mathcal{N}(0, \sigma^2)$.

Pf: WLOG $\mu = 0$, $\sigma^2 = 1$. Suffices to show, for f smooth, cpt. supp.:

$$\mathbb{E} f(\hat{S}_n) \longrightarrow \mathbb{E} f(N)$$

$N \sim \mathcal{N}(0,1)$ $\qquad$ $Y_i \overset{iid}{\sim} \mathcal{N}(0,1)$

$$\mathbb{E} f(\hat{S}_n) = \mathbb{E} f\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right) \qquad \mathbb{E} f(N) = \mathbb{E} f\left(\frac{Y_1 + \cdots + Y_n}{\sqrt{n}}\right)$$

[**Prop:** • $X \sim \mathcal{N}(\mu, \sigma^2)$ then $\operatorname{Law}(aX + b) = \mathcal{N}(a\mu + b, a^2\sigma^2)$
• $X \sim \mathcal{N}(\mu, \sigma^2)$, $X' \sim \mathcal{N}(\mu', \sigma'^2)$ ind. then $\operatorname{Law}(X + X') = \mathcal{N}(\mu + \mu', \sigma^2 + \sigma'^2)$.]

Idea: swap X_i for Y_i one by one.

$$\mathbb{E} f\left(\underbrace{\frac{X_1 + \cdots + X_{n-1}}{\sqrt{n}}}_{=: Z_1} + \frac{X_n}{\sqrt{n}}\right) - \mathbb{E} f\left(\frac{X_1 + \cdots + X_{n-1} + Y_n}{\sqrt{n}}\right)$$

$$= \mathbb{E} f\left(Z_1 + \frac{X_n}{\sqrt{n}}\right) - \mathbb{E} f\left(Z_1 + \frac{Y_n}{\sqrt{n}}\right)$$

Taylor: $f\left(Z_1 + \frac{X_n}{\sqrt{n}}\right) = f(Z_1) + f'(Z_1)\frac{X_n}{\sqrt{n}} + \frac{1}{2} f''(Z_1)\frac{X_n^2}{n} + \Delta$

3

<u>Lagrange remainder</u>: $\Delta = \Delta(Z_1, X_n/\sqrt{n}) = \frac{1}{6} f'''(\tilde{Z}_1) X_n^3/n^{3/2}$
for some $\tilde{Z}_1 = \tilde{Z}_1(Z_1, X_n/\sqrt{n})$ b/w Z_1 and $Z_1 + X_n/\sqrt{n}$.

$$\Longrightarrow |\mathbb{E}\Delta| \le \frac{1}{6n^{3/2}} \mathbb{E}|X_n|^3 \underbrace{\sup_{x\in\mathbb{R}} |f'''(x)|}_{=: \|f'''\|_{L^\infty}}$$

Z_1 and X_n <u>independent</u> $\rightsquigarrow$

$$\mathbb{E} f\left(Z_1 + \frac{X_n}{\sqrt{n}}\right) = \underbrace{\mathbb{E} f(Z_1) + \mathbb{E} f'(Z_1)\frac{\mathbb{E}X_n}{\sqrt{n}} + \mathbb{E} f''(Z_1)\frac{\mathbb{E}X_n^2}{n}} + \mathbb{E}\Delta$$

$\mathbb{E}X_n = \mathbb{E}Y_n$, $\mathbb{E}X_n^2 = \mathbb{E}Y_n^2 \longrightarrow$ first 3 terms match!

$$\Longrightarrow \left|\mathbb{E} f\left(Z_1 + \frac{X_n}{\sqrt{n}}\right) - \mathbb{E} f\left(Z_1 + \frac{Y_n}{\sqrt{n}}\right)\right| \le \frac{1}{6n^{3/2}}\left(\mathbb{E}|X_n|^3 + \overbrace{\mathbb{E}|Y_n|^3}^{=2\sqrt{\frac{2}{\pi}}}\right) \|f'''\|_{L^\infty}$$

Repeat n times $\Longrightarrow$ $\boxed{\left|\mathbb{E} f(\hat{S}_n) - \mathbb{E} f(N)\right| = O\left(\frac{1}{n^{1/2}}\right)}$.

<u>Remarks</u>: 0 Same argument with any Y_i iid $\longrightarrow$ universality w/o knowing Gaussian limit.

1 Can remove $\mathbb{E}|X_i|^3 < \infty$ assumption by <u>truncation</u> trick, like in LLN, expanding $X_i = X_i \mathbb{1}\{|X_i| \le \varepsilon\sqrt{n}\} + X_i \mathbb{1}\{|X_i| > \varepsilon\sqrt{n}\}$.

2 <u>Quantitative rates</u>: $\mathbb{E} f(\hat{S}_n) = \mathbb{E} f(N) + O\left(\frac{1}{\sqrt{n}}\left(1 + \mathbb{E}|X_1|^3\right)\|f'''\|_{L^\infty}\right)$
Shows "flatter" f converge faster.
In similar spirit [redacted] (pf of weaker version by smoothing ("mollifying") $f(x) = \mathbb{1}\{x \le t\}$.

Thm: (Berry - Esséen) $\mathbb{P}[\hat{S}_n \le t] = \mathbb{P}[N \le t] + O\left(\frac{1}{\sqrt{n}} \mathbb{E}|X_1|^3\right)$.

If $\mathbb{E}X_i^k = \mathbb{E}N^k \ \forall\, 1 \le k \le \ell$, $\mathbb{E}|X_i|^{\ell+1} < \infty \rightsquigarrow$ both rates $O\left(\frac{1}{n^{\frac{\ell-1}{2}}}\right)$.

③ Non-identical distributions: by identical proof,

Thm: Suppose $(X_{i,n})_{1 \le i \le n}$ triangular array of ind. r.v. with $\mathbb{E} X_{i,n} = 0$, $\mathbb{E} X_{i,n}^2 = 1$, $\mathbb{E}|X_{i,n}|^3 \le K$ absolute constant.
$\hat{S}_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n X_{i,n} \longrightarrow$ same conclusions.

Slightly harder / more delicate:

Thm: (Lindeberg CLT) $\mathbb{E} X_{i,n} = 0$, $\mathbb{E} X_{i,n}^2 < \infty$,
$V_n^2 := \sum_{i=1}^n \mathbb{E} X_{i,n}^2 = \operatorname{Var} S_n$
Suppose Lindeberg condition holds: $\forall \varepsilon > 0$,
$$\frac{1}{V_n^2} \sum_{i=1}^n \mathbb{E} X_{i,n}^2 \mathbb{1}\{|X_{i,n}| > \varepsilon V_n\} \longrightarrow 0$$
Then, for $\hat{S}_n := \frac{1}{V_n} \sum_{i=1}^n X_{i,n}$, $\hat{S}_n \Longrightarrow N(0,1)$.

④ Non-linear universality: defining $s(X_1, \dots, X_n) := \frac{1}{\sqrt{n}} \sum X_i$, we found $\mathbb{E} f(s(X_1, \dots, X_n)) \approx \mathbb{E} f(s(Y_1, \dots, Y_n))$.
Did not use much of structure of s!

Thm: (informal) Same holds for "nice" multilinear low-degree polynomials,
$$s(X_1, \dots, X_n) = \sum_{\substack{T \subseteq [n] \\ |T| \le d}} \alpha_T \prod_{i \in T} X_i$$
$|T| \le d \leftarrow$ degree
, + Berry-Esséen variant.

Ex: $X_i \sim \mathrm{Unif}(\{\pm 1\}) \longrightarrow Y_i \sim N(0,1)$ (Rk: $(Y_1, \dots, Y_n)$ rotationally invariant)

Analysis of Boolean functions, e.g. voting rules. | Gaussian measure + associated geometry.