STAT 7032 Probability
CLT part

Wlodek Bryc

Created: Friday, Jan 2, 2014
Printed: April 15, 2020 `File:`
`Grad-Prob-2020-slides.TEX`

Facts to use

$\varphi(t) = E\exp(itX)$

- For standard normal distribution $\varphi(t) = e^{-t^2/2}$
- The following are equivalent:
  - $X_n \xrightarrow{\mathcal{D}} X$
  - $\varphi_n(t) \to \varphi(t)$ for all $t \in \mathbb{R}$.
- If X is square integrable with mean zero and variance σ^2 then

$$\left|\varphi(t) - (1 - \tfrac{\sigma^2 t^2}{2})\right| \leq E(\min\{\frac{1}{6}|tX|^3, (tX)^2\}) \qquad (1)$$

Proof: $\varphi(t) = Ee^{-itX}$.

This relies on two integral identities applied to $x = tX(\omega)$ under the integral: $\left|e^{ix} - (1 + ix - \frac{x^2}{2})\right| = \left|\frac{i}{2}\int_0^x (x-s)^2 e^{is} ds\right| \leq \frac{|x^3|}{6}$

$$\left|e^{ix} - (1 + ix - \tfrac{x^2}{2})\right| = \left|\int_0^x (x-s)(e^{is}-1)ds\right| \leq x^2$$

□

Last time we used inequality $|z_1^n - z_2^n| \le n|z_1 - z_2|$ complex numbers of modulus at most 1 which we now generalize.

Lemma

If $z_1, \dots, z_m$ and $w_1, \dots, w_m$ are complex numbers of modulus at most 1 then

$$|z_1 \dots z_m - w_1 \dots w_m| \le \sum_{k=1}^{m} |z_k - w_k| \tag{2}$$

Proof.

Write the left hand side of (2) as a telescoping sum:

$$\begin{aligned} z_1 \dots z_m - w_1 \dots w_m &= z_1 \dots z_m - w_1 z_2 \dots z_m + w_1 z_2 \dots z_m - w_1 w_2 \dots z_m \\ &\quad \dots + w_1 w_2 \dots w_{m-1} z_m - w_1 w_2 \dots w_m \\ &= \sum_{k=1}^{m} w_1 \dots w_{k-1} (z_k - w_k) z_{k+1} \dots z_m \end{aligned}$$

□

Lindeberg's theorem

For each n we have a triangular array of random variables that are independent in each row

$$\begin{array}{ccc} X_{1,1}, X_{1,2}, & \dots & , X_{1,r_1} \\ X_{2,1}, X_{2,2}, & \dots & , X_{2,r_2} \\ & \vdots & \\ X_{n,1}, X_{n,2}, & \dots & , X_{n,r_n} \\ & \vdots & \end{array}$$

and we set $S_n = X_{n,1} + \cdots + X_{n,r_n}$. We assume that random variables are square-integrable with mean zero, and we use the notation

$$E(X_{n,k}) = 0, \; \sigma_{nk}^2 = E(X_{n,k}^2), \; s_n^2 = \sum_{k=1}^{r_n} \sigma_{nk}^2 \tag{3}$$

Definition (The Lindeberg condition)

We say that the *Lindeberg condition* holds if

$$\forall_{\varepsilon > 0} \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{r_n} \int_{|X_{nk}| > \varepsilon s_n} X_{nk}^2 dP = 0 \tag{4}$$

Remark (Important Observation)

Under the Lindeberg condition, we have

$$\lim_{n\to\infty} \max_{k\le r_n} \frac{\sigma_{nk}^2}{s_n^2} = 0 \tag{5}$$

Proof.

$$\sigma_{nk}^2 = \int_{|X_{nk}|\le\varepsilon s_n} X_{nk}^2 dP + \int_{|X_{nk}|>\varepsilon s_n} X_{nk}^2 dP \le \varepsilon s_n^2 + \int_{|X_{nk}|>\varepsilon s_n} X_{nk}^2 dP$$

So

$$\max_{k\le r_n} \frac{\sigma_{nk}^2}{s_n^2} \le \varepsilon + \frac{1}{s_n^2} \max_{k\le r_n} \int_{|X_{nk}|>\varepsilon s_n} X_{nk}^2 dP$$

$$\le \varepsilon + \frac{1}{s_n^2} \sum_{k=1}^{r_n} \int_{|X_{nk}|>\varepsilon s_n} X_{nk}^2 dP$$

□

Theorem (Lindeberg CLT)

Suppose that for each n the sequence $X_{n1} \dots X_{n,r_n}$ is independent with mean zero. If the Lindeberg condition holds for all $\varepsilon > 0$ then $S_n/s_n \xrightarrow{\mathcal{D}} Z$.

Example (Suppose $X_1, X_2, \dots,$ are iid mean m variance $\sigma^2 > 0$. Then $S_n = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_k - m) \xrightarrow{\mathcal{D}} Z$.)

- Triangular array: $X_{n,k} = \frac{X_k - m}{\sqrt{n}\sigma}$ and $s_n = 1$.
- The Lindeberg condition is

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \int_{|X_k - m| > \varepsilon\sigma\sqrt{n}} \frac{(X_k - m)^2}{\sigma^2} dP$$
$$= \lim_{n\to\infty} \frac{1}{\sigma^2} \int_{|X_1 - m| > \varepsilon\sigma\sqrt{n}} (X_1 - m)^2 dP = 0$$

by Lebesgue dominated convergence theorem.

Proof of Lindeberg CLT I

Without loss of generality we may assume that $s_n^2 = 1$ so that $\sum_{k=1}^{r_n} \sigma_{nk}^2 = 1$.

- Denote $\varphi_{nk} = E(e^{itX_{nk}})$. By (1) we have

$$\left|\varphi_{nk}(t) - (1 - \frac{1}{2}t^2\sigma_{nk}^2)\right| \leq E\left(\min\{|tX_{nk}|^2, |tX_{nk}|^3\}\right)$$

$$\leq \int_{|X_{nk}|\leq\varepsilon} |tX_{nk}|^3 dP + \int_{|X_{nk}|>\varepsilon} |tX_{nk}|^2 dP$$

$$\leq t^3\varepsilon \int_{|X_{nk}|dP\leq\varepsilon} X_{nk}^2 dP + t^2 \int_{|X_{nk}|>\varepsilon} X_{nk}^2 dP \leq t^3\varepsilon\sigma_{nk}^2 + t^2 \int_{|X_{nk}|>\varepsilon} X_{nk}^2$$

- Using (2), $\boxed{|z_1 \dots z_m - w_1 \dots w_m| \leq \sum_{k=1}^m |z_k - w_k|}$ we see that for n large enough so that $\frac{1}{2}t^2\sigma_{nk}^2 < 1$

$$\left|\varphi_{S_n}(t) - \prod_{k=1}^{r_n}(1 - \tfrac{1}{2}t^2\sigma_{nk}^2)\right|$$

Proof of Lindeberg CLT II

Since $\varepsilon > 0$ is arbitrary and $t \in \mathbb{R}$ is fixed, this shows that

$$\lim_{n\to\infty} \left| \varphi_{S_n}(t) - \prod_{k=1}^{r_n} (1 - \tfrac{1}{2} t^2 \sigma_{nk}^2) \right| = 0$$

It remains to verify that $\lim_{n\to\infty} \left| e^{-t^2/2} - \prod_{k=1}^{r_n} (1 - \frac{1}{2} t^2 \sigma_{nk}^2) \right| = 0$. To do so, we apply the previous proof to the triangular array $Z_{n,k} = \sigma_{n,k} Z_k$ of independent normal random variables. Note that

$$\varphi_{\sum_{k=1}^{r_n} Z_{nk}}(t) = \prod_{k=1}^{r_n} e^{-t^2 \sigma_{nk}^2/2} = e^{-t^2/2}$$

We only need to verify the Lindeberg condition for $\{Z_{nk}\}$.

Proof of Lindeberg CLT III

$$\int_{|Z_{nk}|>\varepsilon} Z_{nk}^2 dP = \sigma_{nk}^2 \int_{|x|>\varepsilon/\sigma_{nk}} x^2 f(x) dx$$

So for $\varepsilon > 0$ we estimate (recall that $\sum_k \sigma_{nk}^2 = 1$)

$$\begin{aligned}
\sum_{k=1}^{r_n} \int_{|Z_{nk}|>\varepsilon} Z_{nk}^2 dP &\le \sum_{k=1}^{r_n} \sigma_{nk}^2 \int_{|x|>\varepsilon/\sigma_{nk}} x^2 f(x) dx \\
&\le \max_{1\le k\le r_n} \int_{|x|>\varepsilon/\sigma_{nk}} x^2 f(x) dx \\
&= \int_{|x|>\varepsilon/\max_k \sigma_{nk}} x^2 f(x) dx
\end{aligned}$$

The right hand side goes to zero as $n \to \infty$, because by $\max_{1\le k\le r_n} \sigma_{nk} \to 0$ by (5). QED

Lyapunov's theorem

Theorem

Suppose that for each n the sequence $X_{n1} \dots X_{n,r_n}$ is independent with mean zero. If the Lyapunov's condition

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{r_n} E|X_{nk}|^{2+\delta} = 0 \tag{7}$$

holds for some $\delta > 0$, then $S_n/s_n \xrightarrow{\mathcal{D}} Z$

Proof.

We use the following bound to verify Lindeberg's condition:

$$\frac{1}{s_n^2} \sum_{k=1}^{r_n} \int_{|X_{nk}|>\varepsilon s_n} X_{nk}^2 dP \leq \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{k=1}^{r_n} \int_{|X_{nk}|>\varepsilon s_n} |X_{nk}|^{2+\delta} dP$$

$$\leq \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{k=1}^{r_n} E|X_{nk}|^{2+\delta}$$

□

Corollary

Suppose X_k are independent with mean zero, variance σ^2 and that $\sup_k E|X_k|^{2+\delta} < \infty$. Then $S_n/\sqrt{n} \xrightarrow{\mathcal{D}} \sigma Z$.

Proof.

Let $C = \sup_k E|X_k|^{2+\delta}$. WLOG $\sigma > 0$. Then $s_n = \sigma\sqrt{n}$ and $\frac{1}{s_n^{2+\delta}} \sum_{k=1}^n E(|X_k|^{2+\delta}) \le \frac{Cn}{\sigma^{2+\delta} n^{1+\delta/2}} = \frac{C}{\sigma^{2+\delta} n^{\delta/2}} \to 0$, so Lyapunov's condition is satisfied. □

Corollary

Suppose X_k are independent, uniformly bounded, and have mean zero. If $\sum_n Var(X_n) = \infty$, then $S_n/\sqrt{Var(S_n)} \xrightarrow{\mathcal{D}} N(0,1)$.

Proof.

Suppose $|X_n| \le C$ for a constant C. Then

$$\frac{1}{s_n^3} \sum_{k=1}^{n} E|X_n|^3 \le C\frac{s_n^2}{s_n^3} = \frac{C}{s_n} \to 0$$

□

The end

Lets stop here

- Homework 11, due Monday - two exercises from Ch 11 of the notes.
- There is also a sizeable list of exercises from past prelims
- Things to do on Friday:
  - CLT without Lindeberg condition, when normalization is not by variance
  - Multivariate characteristic functions and multivariate normal distribution.

Thank you