

Lecture 12 : Setup for the Central Limit Theorem

STAT205 *Lecturer: Jim Pitman*

Scribe: *Jing Lei* <jinglei@statberkeley.edu>

This set of notes is a revision of the work of David S. Rosenberg and Nate Coehlo.

Abstract

See Durrett's book section 2.4b for an equivalent formulation and a proof using characteristic functions. That proof leans on the continuity theorem for characteristic functions, (3.4) on page 99, which in turn relies on Helly's selection theorem (2.5) on page 88. The present approach, due to Lindeberg, is more elementary in that it does not require these tools, but note that the basic idea in both arguments is to estimate the expected value of a sum of independent variables using a Taylor expansion with error bound.

12.1 Triangular Arrays

Roughly speaking, a sum of many small independent random variables will be approximately normally distributed. To formulate such a limit theorem, we must consider a sequence of sums of more and more, smaller and smaller random variables. Therefore, throughout this section we shall study the sequence of sums

$$S_i = \sum_j X_{ij}$$

obtained by summing the rows of a *triangular array* of random variables

$$\begin{array}{cccc} X_{11}, X_{12}, \dots, & X_{1n_1} \\ X_{21}, X_{22}, \dots, & X_{2n_2} \\ X_{31}, X_{32}, \dots, & X_{3n_3} \\ \vdots & \vdots & \vdots & \vdots \end{array}$$

It will be assumed throughout that the triangular arrays we consider satisfy 3 *Triangular Array Conditions*¹ (here i ranges over $\{1, 2, \dots\}$, and j ranges over $\{1, 2, \dots, n_i\}$):

¹This is not standard terminology, but is used here as a simple referent for these conditions.

1. For each i , the n_i random variables $X_{i1}, X_{i2}, \dots, X_{in_i}$ in the i th row are mutually independent.
2. $\mathbb{E}(X_{ij}) = 0$ for all i, j , and
3. $\sum_j \mathbb{E}X_{ij}^2 = 1$ for all i .

We have some remarks for these conditions:

- It is *not* assumed that random variables in each row are identically distributed.
- It is *not* assumed that different rows are independent. In fact, a common application of triangular arrays is sums $X_1 + X_2 + \dots + X_n$ obtained from a sequence of independent random variables $X_1, X_2, \dots$
- It will usually be the case that $n_i \rightarrow \infty$ as $i \rightarrow \infty$. And according to the nature of our problem, we should have the variables in each row tend to be smaller and smaller as i increases. Both of these two conditions are implied by the Lindeberg Condition which we will discuss below.

12.2 The Lindeberg Condition and Some Consequences

Theorem 12.1 (Lindeberg's Theorem) *Suppose that in addition to the Triangular Array Conditions, the triangular array satisfies Lindeberg's condition:*

$$\forall \epsilon > 0, \lim_{i \rightarrow \infty} \sum_{j=1}^{n_i} \mathbb{E}[X_{ij}^2 \mathbf{1}(|X_{ij}| > \epsilon)] = 0 \quad (12.1)$$

Then $S_i \xrightarrow{d} \mathcal{N}(0, 1)$.

The Lindeberg condition makes precise the sense in which the random variables must be smaller and smaller. It says that for arbitrarily small $\epsilon > 0$, the contribution to the total row variance from the terms with absolute value greater than ϵ becomes negligible as you go down the rows. We see this as follows:

$$\begin{aligned} X_{ij}^2 &\leq \epsilon^2 + X_{ij}^2 \mathbf{1}(|X_{ij}| > \epsilon) \\ \mathbb{E}X_{ij}^2 &\leq \epsilon^2 + \mathbb{E}X_{ij}^2 \mathbf{1}(|X_{ij}| > \epsilon) \\ \mathbb{E}X_{ij}^2 &\leq \epsilon^2 + \sum_j \mathbb{E}X_{ij}^2 \mathbf{1}(|X_{ij}| > \epsilon) \end{aligned}$$

This last inequality is true for all j , so we have:

$$\max_j \mathbb{E}X_{ij}^2 \leq \epsilon^2 + \sum_j \mathbb{E}X_{ij}^2 \mathbf{1}(|X_{ij}| > \epsilon) \quad (12.2)$$

The Lindeberg condition says that, as $i \rightarrow \infty$, the summation on the RHS of (12.2) tends to zero. Since (12.2) holds for all $\epsilon > 0$, we get

$$\lim_{i \rightarrow \infty} \max_j \mathbb{E}X_{ij}^2 = 0, \quad (12.3)$$

which implies $n_i \rightarrow \infty$ as $i \rightarrow \infty$, since we assume in Triangular Array Condition that $\sum_j \mathbb{E}X_{ij}^2 = 1$ for all i . Another consequence follows from (12.3) and Chebyshev's inequality: since we have

$$\mathbb{P}(|X_{ij}| > \epsilon) \leq \frac{\mathbb{E}(X_{ij}^2)}{\epsilon^2} \quad \text{for all } \epsilon > 0,$$

taking the maximum over j and $i \rightarrow \infty$, we get that $X_{ij} \xrightarrow{\mathbb{P}} 0$, uniformly in j :

$$\forall \epsilon > 0, \quad \lim_{i \rightarrow \infty} \max_j \mathbb{P}(|X_{ij}| > \epsilon) = 0. \quad (12.4)$$

An array with property (12.4) is said to be *uniformly asymptotically negligible (UAN)*, and there is a striking converse to Lindeberg's Theorem:

Theorem 12.2 (Feller's Theorem) *If a triangular array satisfies the Triangular Array Conditions and is UAN, then $S_i \xrightarrow{d} \mathcal{N}(0, 1)$ (if and) only if Lindeberg's condition (12.1) holds.*

Proof: See Billingsley, Theorem 27.4, or Kallenberg, 5.12. ■

12.3 The Lyapounov Condition

A condition stronger (also often easier to consider and check) than Lindeberg's is the *Lyapounov condition*:

$$\exists \delta > 0 \text{ such that } \lim_{i \rightarrow \infty} \sum_j \mathbb{E}|X_{ij}|^{2+\delta} = 0 \quad (12.5)$$

Lemma 12.3 *Lyapounov's condition implies Lindeberg's condition.*

Proof: Fix any $\epsilon, \delta > 0$. For any random variable $|X| > \epsilon$, we have

$$X^2 = \frac{|X|^{2+\delta}}{|X|^\delta} \leq \frac{|X|^{2+\delta}}{\epsilon^\delta}$$

Thus for any random variable X we have

$$\mathbb{E}[X^2 \mathbf{1}(|X| > \epsilon)] \leq \frac{\mathbb{E}|X|^{2+\delta}}{\epsilon^\delta}$$

Take $X = X_{ij}$ to be the elements of our triangular array, and take δ to be the value from Lyapounov's condition. Then we can sum over j on the RHS and take the limit as $i \rightarrow \infty$ on both sides to get the Lindeberg's condition. ■

Theorem 12.4 (Lyapounov's Theorem) *If a triangular array satisfies the Triangular Array Conditions and the Lyapounov condition (12.5), then $S_i \xrightarrow{d} \mathcal{N}(0, 1)$.*

This follows from Lindeberg's Theorem, but we prove it with $\delta = 1$ below.

12.4 Preliminaries to the proof of Lyapounov's Theorem

We introduce two preliminaries to the proof. First:

Lemma 12.5 *If $X \sim \mathcal{N}(0, \sigma^2)$, $Y \sim \mathcal{N}(0, \tau^2)$ are independent, then $X + Y \sim \mathcal{N}(0, \sigma^2 + \tau^2)$.*

Proof Sketch: Either

1. use the formula for the convolution of densities, or
2. use characteristic or moment generating functions, or
3. use the radial symmetry of the joint density function of i.i.d. $\mathcal{N}(0, \sigma^2 + \tau^2)$ random variables U and V to argue that $U \sin \theta + V \cos \theta \sim \mathcal{N}(0, \sigma^2 + \tau^2)$. Take $\sin(\theta) = \left(\frac{\sigma^2}{\sigma^2 + \tau^2} \right)^{1/2}$.

To see how rotational invariance is unique to the normal distribution, see Kallenberg 13.2. ■

Second:

Lemma 12.6 $S_i \xrightarrow{d} Z$ if and only if $\lim_{i \rightarrow \infty} \mathbb{E}f(S_i) = \mathbb{E}f(Z)$ for all $f \in \mathbf{C}_b^3(\mathbb{R})$, the set of functions from $\mathbb{R}$ to $\mathbb{R}$ with three bounded, continuous derivatives.

Proof: See Durrett, Theorem 2.2, and use that $\mathbf{C}_b^3(\mathbb{R})$ is dense in $\mathbf{C}_b(\mathbb{R})$. ■

12.5 Proof of Lyapounov's Theorem

This proof illustrates the general idea of the proof of Lindeberg's theorem, and avoids a few tricky details which we will deal with later.

Proof: With n fixed, let $X_1, X_2, \dots, X_n$ be independent random variables, not necessarily identically distributed. Suppose $\mathbb{E}X_j = 0$ and let $\sigma_j^2 = \mathbb{E}(X_j^2) < \infty$. Then for $S = \sum_{j=1}^n X_j$ we have $\sigma^2 := \mathbf{Var}S = \sum_{j=1}^n \sigma_j^2$. Note:

1. If $\forall j, X_j \sim \mathcal{N}(0, \sigma_j^2)$, then $S \sim \mathcal{N}(0, \sigma^2)$ by Lemma 10.5.
2. Given independent random variables $X_1, X_2, \dots, X_n$ with arbitrary distributions, we can always construct a new sequence $Z_1, Z_2, \dots, Z_n$ of *normal* random variables with matching means and variances so that all of Z_i and X_i are mutually independent. This may involve changing the basic probability space, but that does not matter because the distribution of S is determined by the joint distribution of $(X_1, X_2, \dots, X_n)$, which remains the same.

Let

$$\begin{aligned} S &:= S_0 := X_1 + X_2 + X_3 + \dots + X_n, \\ S_1 &:= Z_1 + X_2 + X_3 + \dots + X_n, \\ S_2 &:= Z_1 + Z_2 + X_3 + \dots + X_n, \\ &\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ T &:= S_n := Z_1 + Z_2 + Z_3 + \dots + Z_n, \end{aligned}$$

We want to show that S is "close" in distribution to T , i.e., that $\mathbb{E}f(S)$ is close to $\mathbb{E}f(T)$ for all $f \in \mathbf{C}_b^3(\mathbb{R})$ with uniform bound K on f and its first three derivatives: $|f^{(i)}|$, $i = 1, 2, 3$.

By the triangle inequality,

$$|\mathbb{E}f(S) - \mathbb{E}f(T)| \leq \sum_{j=1}^n |\mathbb{E}f(S_j) - \mathbb{E}f(S_{j-1})|. \quad (12.6)$$

Let R_j be the sum of the common terms in S_{j-1} and S_j . Then $S_{j-1} = R_j + X_j$ and $S_j = R_j + Z_j$. Note that by construction, R_j and X_j are independent, as are R_j and Z_j .

We need to compare $\mathbb{E}f(R_j + X_j)$ and $\mathbb{E}f(R_j + Z_j)$. By the Taylor series expansion up to the third term,

$$f(R_j + X_j) = f(R_j) + X_j f^{(1)}(R_j) + \frac{X_j^2}{2!} f^{(2)}(R_j) + \frac{X_j^3}{3!} f^{(3)}(\alpha_j),$$

$$f(R_j + Z_j) = f(R_j) + Z_j f^{(1)}(R_j) + \frac{Z_j^2}{2!} f^{(2)}(R_j) + \frac{Z_j^3}{3!} f^{(3)}(\beta_j),$$

where α_j is a point between R_j and $R_j + X_j$ and β_j is a point between R_j and $R_j + Z_j$.

So, assuming that the X 's have a finite third moments, and noting that the Z 's do as well (see below), we can take expectations in each of these identities and subtract the resulting equations. Using independence and the fact that X and Z agree on their first and second moments, we see that everything below the third order cancels. Therefore,

$$|\mathbb{E}f(S_j) - \mathbb{E}f(S_{j-1})| = |\mathbb{E}f(R_j + X_j) - \mathbb{E}f(R_j + Z_j)| \quad (12.7)$$

$$= \left| \mathbb{E} \frac{X_j^3}{3!} f^{(3)}(\alpha_j) - \mathbb{E} \frac{Z_j^3}{3!} f^{(3)}(\beta_j) \right| \quad (12.8)$$

$$\leq \frac{K}{6} (\mathbb{E}|X_j|^3 + \mathbb{E}|Z_j|^3). \quad (12.9)$$

Let c be the third moment of a standard normal random variable. This is finite since,

$$c = 2 \int_0^\infty x^3 \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\} dx = 2 \cdot \frac{2}{\sqrt{2\pi}} < \infty$$

Therefore, $\mathbb{E}|Z_j|^3 = c\sigma_j^3$.

Jensen's inequality implies that $\|X\|_2 = (\mathbb{E}|X|^2)^{\frac{1}{2}} \leq (\mathbb{E}|X|^3)^{\frac{1}{3}} = \|X\|_3$, so $\sigma_j^3 \leq \mathbb{E}|X_j|^3$, and therefore $\mathbb{E}|Z_j|^3 = c\sigma_j^3 \leq c\mathbb{E}|X_j|^3$, for each j .

Applying this to (12.9), we get

$$\frac{K}{6} (\mathbb{E}|X_j|^3 + \mathbb{E}|Z_j|^3) \leq \frac{K(1+c)}{6} \mathbb{E}|X_j|^3.$$

Now, from (12.6), we get

$$|\mathbb{E}f(S) - \mathbb{E}f(T)| \leq \frac{K(c+1)}{6} \sum_{j=1}^n \mathbb{E}|X_j|^3, \quad (12.10)$$

So far we have only considered one row of the array, but (12.10) is in fact true for every row with K and c unchanged and T having the same distribution. For each i we have,

$$|\mathbb{E}f(S_i) - \mathbb{E}f(T)| \leq \frac{K(c+1)}{6} \sum_{j=1}^{n_i} \mathbb{E}|X_{ij}|^3, \quad (12.11)$$

Now, assuming Lyapounov's condition holds for $\delta = 1$, the RHS of (12.11) goes to zero as $i \rightarrow \infty$.

By Lemma 10.6, $S_i \xrightarrow{d} \mathcal{N}(0, 1)$ as $i \rightarrow \infty$. ■

12.6 Proof of Lindeberg's Central Limit Theorem

For Lyapounov's version of the CLT, we looked at a triangular array $\{X_{ij}\}$ with $\mathbb{E}X_{ij} = 0$, $\mathbb{E}X_{ij}^2 = \sigma_{ij}^2$, $\sum_{j=1}^{n_i} \sigma_{ij}^2 = 1$. Taking $S_i = X_{i1} + X_{i2} + \dots + X_{in_i}$, we saw that we could prove $S_i \xrightarrow{d} \mathcal{N}(0, 1)$ assuming that $\lim_{i \rightarrow \infty} \sum_{k=1}^{n_i} \mathbb{E}|X_{ik}|^3 = 0$.

This is a condition on third moments - we would like to see if a weaker condition will suffice. We used third moments in a Taylor series expansion as follows:

$$f(R + X) = f(R) + Xf^{(1)}(R) + \frac{X^2}{2!}f^{(2)}(R) + \frac{X^3}{3!}f^{(3)}(\alpha), \quad (12.12)$$

where α is a point between R and $R + X$.

Roughly, without the third moments assumption, the above expression is bad when X is large - although the first two moments exist, we might have $\mathbb{E}|X|^3 = \infty$. The idea now is to use the form in equation (12.12) when X is small and to make use of

$$f(R + X) = f(R) + Xf^{(1)}(R) + \frac{X^2}{2!}f^{(2)}(\gamma) \quad (12.13)$$

where γ is a point between R and $R + X$, when X is large.

Equating these expansions (12.12) and (12.13) for $f(R + X)$, we get an alternative

form for the remainder in (12.12):

$$\frac{X^3}{6}f^{(3)}(\alpha) = \frac{X^2}{2}f^{(2)}(\gamma) - \frac{X^2}{2}f^{(2)}(R) \quad (12.14)$$

$$= \frac{X^2}{2}[f^{(2)}(\gamma) - f^{(2)}(R)]\mathbf{1}(|X| > \epsilon) \quad (12.15)$$

$$+ \frac{X^3}{6}f^{(3)}(\alpha)\mathbf{1}(|X| \leq \epsilon) \quad (12.16)$$

for $\epsilon > 0$. Thus, for f with $|f^{(i)}| \leq K$ for $i = 2, 3$, we get

$$\left| \frac{X^3}{6}f^{(3)}(\alpha) \right| \leq KX^2\mathbf{1}(|X| > \epsilon) + \frac{K}{6}|X|^3\mathbf{1}(|X| \leq \epsilon) \quad (12.17)$$

$$\leq KX^2\mathbf{1}(|X| > \epsilon) + \frac{K}{6}\epsilon X^2, \quad (12.18)$$

an alternative to the upper bound $\frac{K}{6}|X|^3$, which we used in (12.9).

Now we return to the setup of section 10.5 and use our new result to get more refined bounds. From (12.6) and (12.8), we had

$$|\mathbb{E}f(S) - \mathbb{E}f(T)| \leq \sum_{j=1}^{n_j} \left| \mathbb{E} \frac{X_j^3}{6} f^{(3)}(\alpha_j) - \mathbb{E} \frac{Z_j^3}{6} f^{(3)}(\beta_j) \right|$$

Using (12.6), the new bound for X_j^3 (12.18), the assumption that $|f^{(3)}| < K$, and since $\mathbb{E}|Z_j|^3 = c\sigma_j^3$, we get

$$|\mathbb{E}f(S) - \mathbb{E}f(T)| \leq \sum_{j=1}^n \left[K\mathbb{E}X_j^2\mathbf{1}(|X_j| > \epsilon) + \frac{K}{6}\epsilon\mathbb{E}X_j^2 \right] + \sum_{j=1}^n \frac{K}{6}c\sigma_j^3 \quad (12.19)$$

$$= K \sum_{j=1}^n \mathbb{E}X_j^2\mathbf{1}(|X_j| > \epsilon) + \frac{K}{6}\epsilon\sigma^2 + \frac{cK}{6} \sum_{j=1}^n \sigma_j^3 \quad (12.20)$$

As $i \rightarrow \infty$ (going down the rows of the triangular array), the first term goes to zero by the Lindeberg condition. The last term goes to zero since

$$\sum_{j=1}^{n(i)} \sigma_{ij}^3 \leq \left(\max_{1 \leq j \leq n(i)} \sigma_{ij} \right) \sum_{j=1}^{n(i)} \sigma_{ij}^2 = \sigma^2 \max_{1 \leq j \leq n(i)} \sigma_{ij},$$

which tends to zero by (12.3). Only $\frac{K}{6}\epsilon\sigma^2$ remains, and letting $\epsilon \rightarrow 0$ finishes the argument. ■