Lecture 10 : Setup for the Central Limit Theorem

STAT205 *Lecturer: Jim Pitman* Scribe: *David S. Rosenberg* <drosen@stat.berkeley.edu>

See Durrett 2nd ed pages 116-118 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 the Helly 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 smooth function of a sum of independent variables using a Taylor expansion with error bound.

10.1 Triangular Arrays

Roughly speaking, a sum of many small independent random variables will be nearly normally distributed. To formulate a limit theorem of this kind, we must consider 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

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

It will be assumed throughout that triangular arrays satisfy 3 *Triangular Array Conditions*[1]:

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

Here the row index i should always be taken to range over $1, 2, 3, \ldots$, while the column index j ranges from 1 to n_i. It is *not* assumed that the r.vs in each row are identically distributed. And it is *not* assumed that different rows are independent. (Different rows could even be defined on different probability spaces.) For motivation, see section ********** below for how such a triangular array is set up in the most important application to partial sums $X_1 + X_2 + \cdots + X_n$ obtained from a sequence of independent r.v.s $X_1, X_2, \ldots$

It will usually be the case that $n_1 < n_2 < \cdots$, whence the term triangular. It is not necessary to assume this however.

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

10.2 The Lindeberg Condition and Some Consequences

We will write $\mathbf{L}(X)$ to denote the *law* or *distribution* of a random variable X. $\mathcal{N}(0,\sigma^2)$ is the normal distribution with mean 0 and variance σ^2.

Theorem 10.1 (Lindebergs Theorem) *Suppose that in addition to the Triangular Array Conditions, the triangular array satisfies* Lindebergs Condition*:*

$$\forall \epsilon > 0,\ \lim_{i\to\infty} \sum_{j=1}^{n_i} \mathbb{E}[X_{ij}^2 \mathbf{1}(|X_{ij}| > \epsilon)] = 0 \tag{10.1}$$

Then, as $i \to \infty$, $\mathbf{L}(S_i) \to \mathcal{N}(0,1)$.

This theorem will be proved in Section ********* below. For an alternative proof using characteristic functions, see Billingsley Sec. 27.

The Lindeberg condition makes precise the sense in which the r.v.s must be small. 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. The Lindeberg condition implies that the maximum contribution to the variance from any of the individual terms in a row becomes negligible as you go down the rows. We see this as follows:

$$\begin{aligned}
X_{ij}^2 &\le \epsilon^2 + X_{ij}^2 \mathbf{1}(|X_{ij}| > \epsilon) && (10.2)\\
\mathbb{E}X_{ij}^2 &\le \epsilon^2 + \mathbb{E}[X_{ij}^2 \mathbf{1}(|X_{ij}| > \epsilon)] && (10.3)\\
\mathbb{E}X_{ij}^2 &\le \epsilon^2 + \sum_j \mathbb{E}[X_{ij}^2 \mathbf{1}(|X_{ij}| > \epsilon)], \text{ which is independent of } j, \text{ so...} && (10.4)\\
\max_j \mathbb{E}X_{ij}^2 &\le \epsilon^2 + \sum_j \mathbb{E}[X_{ij}^2 \mathbf{1}(|X_{ij}| > \epsilon)] && (10.5)
\end{aligned}$$

The Lindeberg condition says that, as we go down the rows (i.e. $i \to \infty$), the summation on the RHS tends to zero. Since inequality (10.5) holds for all $\epsilon > 0$, we get

$$\lim_{i\to\infty} \max_j \mathbb{E}X_{ij}^2 = 0, \tag{10.6}$$

A consequence of (10.6) and condition 3 ($\sum_j \mathbb{E}X_{ij}^2 = 1$ for all i) is that $n_i \to \infty$ as $i \to \infty$. Another consequence follows from the application of (10.6) to Chebychevs inequality. We have for all $\epsilon > 0$,

$$\mathbb{P}(|X_{ij}| > \epsilon) \le \frac{\mathbb{E}(X_{ij}^2)}{\epsilon^2}$$

Taking the maximum over j and $i \to \infty$, we get

$$\forall \epsilon > 0,\ \lim_{i\to\infty} \max_j \mathbb{P}(|X_{ij}| > \epsilon) = 0. \tag{10.7}$$

An array with property (10.7) is said to be *uniformly asymptotically negligible (UAN)*, and there is a striking converse to Lindebergs theorem:

Theorem 10.2 (Fellers Theorem) *If a triangular array satisfies the the Triangular Array Conditions and is UAN, then $\mathbf{L}(S_i) \to \mathcal{N}(0,1)$ [if and] only if Lindebergs condition (10.1) holds.*

Proof: See Billingsley, Theorem 27.4. ■

For UAN arrays there is a more elaborate CLT with infinitely divisible laws as limits - well return to this in later lectures.

Just note for now that

1. it is possible to get normal limits from UAN triangular arrays with infinite variances, and that
2. it is possible to get a $\mathcal{N}(0, \sigma^2)$ limit with $\sigma^2 < 1$ for an array satisfying the Triangular Array Conditions.

10.3 The Lyapounov Condition

A condition stronger than Lindebergs that is often easier to check is the *Lyapounov condition*:

$$\exists \delta > 0 \text{ such that } \lim_{i \to \infty} \sum_j \mathbb{E}|X_{ij}|^{2+\delta} = 0 \tag{10.8}$$

Lemma 10.3 *Lyapounovs condition implies Lindebergs condition.*

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

$$X^2 = \frac{|X|^{2+\delta}}{|X|^\delta} \leq \frac{|X|^{2+\delta}}{\epsilon^\delta} \tag{10.9}$$

Thus for any r.v. X we have

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

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

Theorem 10.4 (Lyapounovs Theorem) *If a triangular array satisfies the Triangular Array Conditions and the Lyapounov condition (10.8), then* $\mathbf{L}(S_i) \to \mathcal{N}(0, 1)$.

10.4 Preliminaries to the proof of Lindebergs Theorem

The key property of the normal distribution is

Theorem 10.5 *If X and Y are independent with $\mathbf{L}(X) = \mathcal{N}(0, \sigma^2)$ and $\mathbf{L}(Y) = \mathcal{N}(0, \tau^2)$, then $\mathbf{L}(X+Y) = \mathcal{N}(0, \sigma^2 + \tau^2)$.*

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. $\mathcal{N}(0, \sigma^2 + \tau^2)$ r.v.s U and V to argue that $\mathbf{L}(U \sin\theta + V \cos\theta) = \mathbf{L}(U)$. Take $\sin(\theta) = \frac{\sigma^2}{\sigma^2+\tau^2}$

■

The key characterization of convergence in distribution is

Theorem 10.6 *$\mathbf{L}(S_i) \to \mathbf{L}(Z)$ if and only if $\lim_{i\to\infty} \mathbb{E}f(S_i) = \mathbb{E}f(Z)$ for all $f \in \mathbb{C}^3(-\infty, \infty)$, the set of functions from* reals *to* reals *with three bounded continuous derivatives.*

Proof: Mimic the proof of Theorem 6 on page 4.2 of notes on the convergence of probability laws. ?? ■

10.5 Proof of Lyapounovs Theorem for $\delta = 1$

This illustrates the general idea and avoids a few tricky details. With n fixed, let $X_1, X_2, \ldots, X_n$ 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 $S = X_1 + \cdots + X_n$ we have $\mathbf{V}arS = \sum_{j=1}^n \sigma_j^2$. Let $\sigma^2 = \mathbf{V}arS$. Note:

1. If $\mathbf{L}(X_j)$ is $\mathcal{N}(0, \sigma_j^2)$, then $\mathbf{L}(S)$ is $\mathcal{N}(0, \sigma^2)$ by Theorem 10.5.
2. Given independent r.v.s $X_1, \ldots, X_n$ with arbitrary distributions, we can always construct a new sequence $Z_1, \ldots, Z_n$ of *normal* r.v.s with matching means and variances so that

$$Z_1, Z_2, \ldots, Z_n, X_1, X_2, \ldots, X_n$$

are mutually independent. This may involve changing the basic probability space, but that doesnt matter because the distribution of S is determined by the joint distribution of $(X_1, \ldots, X_n)$, which remains the same.

Let

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

We want to show that $\mathbf{L}(S)$ is "close to $\mathbf{L}(T)$, which is $\mathcal{N}(0, \sigma^2)$, i.e., that $\mathbb{E}f(S)$ is "close to $\mathbb{E}f(T)$ for all $f \in \mathbb{C}^3(-\infty, \infty)$ with uniform bound K on $|f^{(i)}|$, $i = 0, 1, 2, 3$.

Clearly,

$$|\mathbb{E}f(S) - \mathbb{E}f(T)| \leq \sum_{j=1}^n |\mathbb{E}f(S_j) - \mathbb{E}f(S_{j-1})|. \tag{10.10}$$

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

$$R_j \text{ and } X_j \text{ are independent, as are } R_j \text{ and } Z_j \tag{10.11}$$

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,

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

where $\alpha \in (R_j, R_j + X_j)$. And the same is true with Z_j instead of X_j. So, assuming that the Xs have third moments, we can take expectations in each of these identites and subtract the resulting equations. We get the following:

1. Since $\mathbb{E}X_j = 0 = \mathbb{E}X_j$, and by the independence of X_j, R_j, and Z_j (10.11), we have $\mathbb{E}(X_j f^{(1)}(R_j)) = 0 = \mathbb{E}(Z_j f^{(1)}(R_j))$.
2. Since $\mathbf{V}arX_j = \mathbf{V}arZ_j$, and by (10.11), we have $\mathbb{E}(X_j^2 f^{(2)}(R_j)) = \mathbb{E}(Z_j^2 f^{(2)}(R_j))$.

Thus the first and second order terms cancel, so we are left with the last inequality below (the first two equalities summarize the previous paragraphs):

$$\begin{aligned} |\mathbb{E}f(S_j) - \mathbb{E}f(S_{j-1})| &= |\mathbb{E}f(R_j + X_j) - \mathbb{E}f(R_j + Z_j)| & (10.12)\\ &= \left| \mathbb{E}\frac{X_j^3}{3!} f^{(3)}(\alpha) - \mathbb{E}\frac{Z_j^3}{3!} f^{(3)}(\alpha) \right| & (10.13)\\ &\le \frac{K}{6}(\mathbb{E}|X_j|^3 + \mathbb{E}|Z_j|^3) & (10.14) \end{aligned}$$

where K is the bound on the derivatives of f. Now

$$\begin{aligned} \mathbb{E}|Z_j|^3 &= 2\int_0^\infty z^3 \frac{1}{\sqrt{2\pi}\sigma_j} \exp\{-z^2/(2\sigma_j^2)\}\, dz & (10.15)\\ &= 2\int_0^\infty \sigma_j^3 x^3 \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\}\, dx & (10.16)\\ &= c\sigma_j^3 & (10.17) \end{aligned}$$

where

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

and since $(\mathbb{E}|X|^2)^{\frac{1}{2}} \le (\mathbb{E}|X|^3)^{\frac{1}{3}}$ for any random variable X, we have $\sigma_j^3 \le \mathbb{E}|X_j|^3$ for each j. Thus $\mathbb{E}|Z_j|^3 = c\sigma_j^3 \le c\mathbb{E}|X_j|^3$, for each j. Applying this to (10.14), we get

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

Now, from (10.10), we get

$$|\mathbb{E}f(S) - \mathbb{E}f(T)| \le \frac{(c+1)K}{6} \sum_{j=1}^n \mathbb{E}|X_j|^3. \tag{10.18}$$

To summarize, we have proved:

Lemma 10.7 *Let $X_1, \ldots, X_n$ be independent random variables with $\mathbb{E}X_j = 0$ and $\mathbb{E}|X_j|^3 < \infty$. Let $S = X_1 + \cdots + X_n$ and let T be $\mathcal{N}(0, \sigma^2)$, where σ^2 is the variance of S. Then (10.18) holds for every function f with three continuous derivatives bounded by $\pm K$.*

Now check using 10.6 that Lyapounovs theorem for $\delta = 1$ is obtained by applying the Lemma to the rows of a triangular array satisfying the the Triangular Array Conditionsand the Lyapounov condition (10.8) for $\delta = 1$.

10.6 Proof of Lindebergs Central Limit Theorem

For Lyapounovs 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 $S_i = X_{i1} + X_{i2} + \cdots + X_{in}$, we saw that we could prove $\mathbf{L}(S_i) \to \mathcal{N}(0,1)$ assuming that $\lim_{i\to\infty} \sum_{k=1}^{n_i} \mathbb{E}|X_{ij}|^3 = 0$

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:

$$f(R+X) = f(R) + Xf^{(1)}(R) + \frac{X^2}{2!}f^{(2)}(R) + \frac{X^3}{3!}f^{(3)}(\alpha), \tag{10.19}$$

where $\alpha \in (R, R+X)$.

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 (10.19) when X is small and to make use of

$$f(R+X) = f(R) + Xf^{(1)}(R) + \frac{X^2}{2!}f^{(2)}(\beta) \tag{10.20}$$

where $\beta \in (R, R+X)$, when X is large.

Equating these expansions (10.19) and (10.20) for $f(R+X)$, we get an alternative form for the remainder in (10.19):

$$\begin{aligned} \frac{X^3}{6}f^{(3)}(\alpha) &= \frac{X^2}{2}f^{(2)}(\beta) - \frac{X^2}{2}f^{(2)}(R) && (10.21)\\ &= \frac{X^2}{2}[f^{(2)}(\beta) - f^{(2)}(R)]\mathbf{1}\left(|X| > \epsilon\right) + \frac{X^3}{6}f^{(3)}(\alpha)\mathbf{1}\left(|X| \le \epsilon\right) && (10.22)\end{aligned}$$

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

$$\begin{aligned} \left|\frac{X^3}{6}f^{(3)}(\alpha)\right| &\le KX^2\mathbf{1}\left(|X| > \epsilon\right) + \frac{K}{6}|X|^3\mathbf{1}\left(|X| \le \epsilon\right) && (10.23)\\ &\le KX^2\mathbf{1}\left(|X| > \epsilon\right) + \frac{K}{6}\epsilon X^2, && (10.24)\end{aligned}$$

an alternative to the upper bound $\frac{K}{6}|X|^3$, which we used in (10.14).

Now we return to the setup of section 10.5 and use our new result to get more refined bounds. From (10.10) and (10.13), we had

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

Using the triangle inequality, the new bound for X_j^3 (10.24), the assumption that $|f^{(3)}| < K$, and $\mathbb{E}|Z_j|^3 = c\sigma_j^3$ (10.17), we get

$$\begin{aligned} |\mathbb{E}f(S) - \mathbb{E}f(T)| &\le \sum_{j=1}^{n}\left[K\mathbb{E}X_j^2\mathbf{1}\left(|X_j| > \epsilon\right) + \frac{K}{6}\epsilon\mathbb{E}X_j^2\right] + \sum_{j=1}^{n}\frac{K}{6}c\sigma_j^3 && (10.25)\\ &= K\sum_{j=1}^{n}\mathbb{E}X_j^2\mathbf{1}\left(|X_j| > \epsilon\right) + \frac{K}{6}\epsilon\sigma^2 + \frac{cK}{6}\sum_{j=1}^{n}\sigma_j^3 && (10.26)\end{aligned}$$

As $i \to \infty$ (i.e. we go 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 (10.6). Only $\frac{K}{6}\epsilon\sigma^2$ remains, and letting $\epsilon \to 0$ finishes the argument.

10.7 Applications

Let $S_n = X_1 + X_2 + \cdots + X_n$ where $X_1, X_2, \ldots$ is a sequence of independent, possibly non-identically distributed r.v.s, each with mean 0. Let $\mathbf{V}arX = \sigma_j^2$ and $\mathbf{s}_n^2 = \sigma_{j=1}^n \sigma_j^2$. We want to know when $\mathbf{L}(S_i/\mathbf{s}_i) \to \mathcal{N}(0,1)$. Check Lindebergs condition for the triangular array $X_{ij} = X_j/\mathbf{s}_i$, $j = 1, 2, \ldots, i$. Then S_i in the Lindeberg CLT is replaced by $S_i/\mathbf{s}_i$, and the Lindeberg condition becomes

$$\lim_{n \to \infty} \sum_{j=1}^{n} \mathbb{E}\left[\frac{X_j^2}{\mathbf{s}_n^2} \mathbf{1}\left(\left| \frac{X_j}{\mathbf{s}_n} \right| > \epsilon \right) \right] = 0, \text{ for all } \epsilon > 0, \tag{10.27}$$

$$\text{i.e. } \lim_{n \to \infty} \frac{1}{\mathbf{s}_n^2} \sum_{j=1}^{n} \mathbb{E}\left[X_j^2 \mathbf{1}\left(|X_j| > \epsilon \mathbf{s}_n \right) \right] = 0, \text{ for all } \epsilon > 0, \tag{10.28}$$

Examples where the Lindeberg condition holds:

1. The i.i.d. case where $\mathbf{s}_n^2 = n\sigma^2$:

$$\frac{1}{n\sigma^2} \sum_{j=1}^{n} \mathbb{E}[X_j^2 \mathbf{1}\left(|X_j| > \epsilon\sigma\sqrt{n}\right)] = \frac{1}{\sigma^2} \mathbb{E}[X_1^2 \mathbf{1}\left(|X_1| > \epsilon\sigma\sqrt{n}\right)],$$

   and since $\mathbb{E}X_1^2 < \infty$, we can use the dominated convergence theorem to conclude that the Lindeberg condition holds.

2. Lyapounovs condition

$$\lim_{n \to \infty} \frac{1}{\mathbf{s}_n^{2+\delta}} \sum_{j=1}^{n} \mathbb{E}|X_j|^{2+\delta} = 0 \text{ for some } \delta > 0$$

   implies Lindebergs condition. The proof of this is given (essentially) in Lemma 10.3.

3. If $X_1, X_2, \ldots$ are uniformly bounded: $|X_j| \leq M$ for all j, and $\mathbf{s}_n \uparrow \infty$. Fix $\epsilon > 0$. For n so large that $\mathbf{s}_n \geq M/\epsilon$, we have

$$\mathbf{1}\left(|X_j| > \epsilon \mathbf{s}_n\right) = \mathbf{1}\left(|X_j| > M\right) = 0 \text{ for all } j.$$

   Hence the Lindeberg condition is satisfied.