Lindeberg-Feller central limit theorem

Patrick Breheny

September 23
Introduction

- Last time, we proved the central limit theorem for the iid case.
- Obviously, this is very useful, but at the same time, has clear limitations – the majority of practical applications of statistics involve modeling the relationship between some outcome $Y$ and a collection of potential predictors $\{X_j\}_{j=1}^d$.
- Those predictors are not the same for each observation; hence, $Y$ is not iid and the ordinary CLT does not apply.
Introduction (cont’d)

• Nevertheless, we’d certainly hope it to be the case that \( \sqrt{n}(\hat{\beta} - \beta) \) converges to a normal distribution even if the errors are not normally distributed.

• Does it? If so, under what circumstances?

• Before getting to this question, let’s first introduce the concept of a “triangular array” of variables.
A triangular array of random variables is of the form

\[
\begin{align*}
X_{11} \\
X_{21} & \quad X_{22} \\
X_{31} & \quad X_{32} & \quad X_{33} \\
\ldots
\end{align*}
\]

where the random variables in each row (i) are independent of each other, (ii) have zero mean and (iii) have finite variance.

The requirement that the variables have zero mean is only for convenience; we can always construct zero-mean variables by considering

\[X_{ni} = Y_{ni} - \mu_{ni}\]

I’ve stated the definition here in terms of scalar variables, but the entries in this triangle can also be random vectors \(x_{ni}\).
• We are going to be concerned with $Z_n = \sum_{i=1}^{n} X_{ni}$, the row-wise sum of the array.
• Since the elements of each row are independent, we have

$$s_n^2 = \mathbb{V}Z_n = \sum_{i=1}^{n} \mathbb{V}X_{ni} = \sum_{i=1}^{n} \sigma_{ni}^2$$

or, if the elements in the array are random vectors,

$$\mathbf{V}_n = \mathbb{V}\mathbf{z}_n = \sum_{i=1}^{n} \mathbb{V}\mathbf{x}_{ni} = \sum_{i=1}^{n} \Sigma_{ni}$$
• There are a few different ways of extending the central limit theorem to non-iid random variables; the most general of these is the Lindeberg-Feller theorem

• This version of the CLT involves a new condition known as the *Lindeberg condition*: for every $\epsilon > 0$,

$$
\frac{1}{s_n^2} \sum_{i=1}^{n} \mathbb{E}\{X_{ni}^2 1(|X_{ni}| \geq \epsilon s_n)\} \to 0
$$

as $n \to \infty$

• We’ll discuss the multivariate version of this condition a bit later
The Lindeberg condition is a bit abstract at first, so let’s see how it works, starting with the simplest case: iid random variables.

**Theorem:** Suppose $X_1, X_2, \ldots$ are iid with mean zero and finite variance. Then the Lindeberg condition is satisfied.

There are three key steps in this proof:

1. Replacing the infinite sum with a single quantity $\propto \mathbb{E}T_n$
2. $T_n \xrightarrow{P} 0$ (which happens if $s_n \to \infty$)
3. $\mathbb{E}T_n \to 0$ by the Dominated Convergence Theorem (requires finite variance)
Non-iid case

- The last two steps work out essentially the same way in non-iid settings.
- The first step, however, requires some resourcefulness.
- Typically, the proof proceeds along the lines of bounding the elements of the sum by their “worst-case scenario”; this eliminates the sum, but requires a condition requiring that the worse-case scenario can’t be too extreme.
- We’ll see a specific example of this later as it pertains to regression.
Lindeberg’s theorem

- We are now ready to present the Lindeberg-Feller theorem, although we won’t be proving it in this course.

**Theorem (Lindeberg):** Suppose \( \{X_{ni}\} \) is a triangular array with \( Z_n = \sum_{i=1}^{n} X_{ni} \) and \( s_n^2 = \sqrt{\text{Var} Z_n} \). If the Lindeberg condition holds: for every \( \epsilon > 0 \),

\[
\frac{1}{s_n^2} \sum_{i=1}^{n} \mathbb{E}\{X_{ni}^2 \cdot 1(|X_{ni}| \geq \epsilon s_n)\} \to 0,
\]

then \( Z_n/s_n \xrightarrow{d} \mathcal{N}(0, 1) \).
Lindeberg’s theorem, alternate statement

- The preceding theorem is expressed in terms of sums; it is often more natural to think about Lindeberg’s theorem in terms of means.

- **Theorem (Lindeberg):** Suppose \( \{X_{ni}\} \) is a triangular array such that \( Z_n = \frac{1}{n} \sum_{i=1}^{n} X_{ni}, \) \( s^2_n = \frac{1}{n} \sum_{i=1}^{n} \text{Var} X_{ni}, \) and \( s^2_n \rightarrow s^2 \neq 0. \) If the Lindeberg condition holds: for every \( \epsilon > 0, \)

\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\{X_{ni}^2 1(|X_{ni}| \geq \epsilon \sqrt{n})\} \rightarrow 0,
\]

then \( \sqrt{n}Z_n \xrightarrow{d} \mathcal{N}(0, s^2). \)

- Note: we’ve added an assumption that \( s^2_n \rightarrow s^2, \) but made the Lindeberg condition easier to handle (\( s_n \) no longer appears).
Feller’s Theorem

• The preceding theorem(s) show that the Lindeberg condition is sufficient for asymptotic normality

• Feller showed that it was also a necessary condition, if we introduce another requirement:

$$\max_i \frac{\sigma_{ni}^2}{\sum_{j=1}^{n} \sigma_{nj}^2} \to 0$$

as $n \to \infty$; i.e., no one term dominates the sum

• **Theorem (Feller):** Suppose $\{X_{ni}\}$ is a triangular array with $Z_n = \sum_{i=1}^{n} X_{ni}$ and $s_n^2 = \text{VAR}(Z_n)$. If $Z_n/s_n \overset{d}{\to} \text{N}(0, 1)$ and $\max_i \frac{\sigma_{ni}^2}{s_n^2} \to 0$, then the Lindeberg condition holds.
Putting these two theorems together, the Lindeberg-Feller Central Limit Theorem says that if no one term dominates the variance, then we have asymptotic normality if and only if the Lindeberg condition holds.

The forward (Lindeberg) part of the theorem is the most important part in practice, as our goal is typically to prove asymptotic normality.

However, it is worth noting that the Lindeberg condition is the minimal condition that must be met to ensure this.

For example, there is another CLT for non-iid variables called the Lyapunov CLT, which requires a “Lyapunov condition”; not surprisingly, this implies the Lindeberg condition, as it is a stronger condition than necessary for asymptotic normality.
• Now let’s look at the multivariate form of the Lindeberg-Feller CLT, which I’ll give in the “mean” form

• **Theorem (Lindeberg-Feller CLT):** Suppose \( \{x_{ni}\} \) is a triangular array of \( d \times 1 \) random vectors such that 
  \[ z_n = \frac{1}{n} \sum_{i=1}^{n} x_{ni} \text{ and } V_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{V} x_{ni} \rightarrow V, \]
  where \( V \) is positive definite. If for every \( \epsilon > 0 \),

  \[
  \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\{\|x_{ni}\|^2 1(\|x_{ni}\| \geq \epsilon \sqrt{n})\} \rightarrow 0,
  \]

  then \( \sqrt{n}z_n \xrightarrow{d} N(0, V) \).

• Or equivalently, \( \sqrt{n}V_n^{-1/2}z_n \xrightarrow{d} N(0, I) \)
Similar to the univariate case, the Lindeberg condition is both necessary and sufficient if we add the condition that no one term dominates the variance.

In the multivariate setting, this means that
\[
\frac{\sum_{i=1}^{n} \mathbf{V} \mathbf{x}_i}{\sum_{i=1}^{n} \mathbf{V} \mathbf{x}_i} \rightarrow 0_{d \times d}
\]
for all \( i \); the division here is element-wise.
• OK, now let’s take what we’ve learned and put it into practice, answering our question from the beginning of lecture: do we have a central limit theorem for linear regression?

• **Theorem:** Suppose \( y = X\beta + w \), where \( w_i \overset{iid}{\sim} (0, \sigma^2) \). Suppose \( \frac{1}{n} X^\top X \to \Sigma \), where \( \Sigma \) is positive definite, and let \( \gamma_n = \max_i \|x_i\| \), where \( x_i \) denotes the \( d \times 1 \) vector of covariates for subject \( i \) (taken to be fixed, not random). If \( \gamma_n \) is bounded and \( \gamma_n^2/n \to 0 \), then

\[
\frac{1}{\sigma} (X^\top X)^{1/2} (\hat{\beta} - \beta) \xrightarrow{d} N(0, I).
\]

• In other words, \( \hat{\beta} \sim N(\beta, \sigma^2 (X^\top X)^{-1}) \)
Remarks

- Note that in proving this result, we needed two key conditions
  - \( \frac{1}{n} \mathbf{X}^\top \mathbf{X} \) converging to a p.d. matrix; this seems obvious since if \( \mathbf{X}^\top \mathbf{X} \) was not invertible, \( \hat{\beta} \) isn't even well-defined
  - \( \gamma_n^2 / n \to 0 \); this is less obvious, but is connected to the idea of influence in regression

- In iid data, all observations essentially carry the same weight for the purposes of estimation and inference

- In regression, however, observations far from the mean of the covariate have much greater influence over the model fit

- This is essentially what \( \gamma_n \) is measuring: in words, \( \gamma_n^2 / n \to 0 \) is saying that no one observation can exhibit too great an influence
• This is one of those situations where theory helps to guide intuition and practice
• Let’s carry out a simulation to illustrate
• We will challenge the central limit theorem in two ways:
  ◦ $w$ will follow a $t$ distribution with $\nu$ degrees of freedom
  ◦ The elements of $X$ will be uniformly distributed (from 0 to 1) except for the first two elements of column 1, which will be set to $\pm a$
• In what follows, $n = 100$ unless otherwise noted; 1000 simulations were run for each example
As we will see, the more comfortably the Lindeberg condition holds, the faster the rate of convergence to normality.
Results: $\nu = 50$, $a = 5$

Influential observations, but $\varepsilon$ close to normal
Results: $\nu = 3, \alpha = 1$

Heavy tails, but no influential observations
Results: $\nu = 3, a = 5$

Heavy tails and influential observations
Results: $\nu = 3, \alpha = 5$

Heavy tails and influential observations, but $n = 1000$