

Lindeberg-Feller central limit theorem

Patrick Breheny

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 notation for a “triangular array” of variables

Triangular array

- A triangular array of random variables is of the form

$$\begin{array}{ccc} X_{11} & & \\ X_{21} & X_{22} & \\ X_{31} & X_{32} & X_{33} \\ \dots & & \end{array}$$

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 $\mathbf{x}_{ni}$

Triangular array (cont'd)

- 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,

$$\mathbf{V}_n = \mathbb{V} \mathbf{z}_n = \sum_{i=1}^n \mathbb{V} \mathbf{x}_{ni} = \sum_{i=1}^n \boldsymbol{\Sigma}_{ni}$$

Non-IID laws of large numbers

- Before moving on to central limit theorems, it's worth mentioning how the law of large numbers extends to the non-iid case
- **Theorem (Law of Large Numbers, non-IID):** Suppose $\mathbf{x}_1, \mathbf{x}_2, \dots$ are independent random variables with $\frac{1}{n} \sum_i \boldsymbol{\mu}_i \rightarrow \boldsymbol{\mu}$ and $\frac{1}{n} \sum \mathbb{V} \mathbf{x}_i$ is bounded. Then $\bar{\mathbf{x}} \xrightarrow{P} \boldsymbol{\mu}$.
- Note that if there is a uniform bound on the individual variances, meaning that $(\mathbb{V} \mathbf{x}_i)_{jk} < M$ for all i, j, k , then $\frac{1}{n} \sum \mathbb{V} \mathbf{x}_i$ is bounded as well

Non-iid conditions

- The fundamental idea of the central limit theorem is that the normal distribution arises from the accumulation of a large number of small contributions
- This happens automatically when random variables are iid
- When they are not, then the contribution from each variable differs and there is the possibility that one or two variables dominate the sum and get in the way of normality
- Extending the central limit theorem to non-iid random variables, then, requires some sort of condition that prevents this from happening

Feller condition

- Let us first consider the Feller condition:

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

as $n \rightarrow \infty$

- This is a very simple, direct expression of the main idea: no one term can dominate the total variance
- The Feller condition, however, is not sufficient to ensure a limiting normal distribution

Feller counterexample

- For example, suppose

$$Y_{ni} \stackrel{\text{i.i.d.}}{\sim} \text{Bern}(1/n) \quad X_{ni} = Y_{ni} - \frac{1}{n}$$

- In this case, the Feller condition does hold:

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

- However, the limiting distribution is Poisson, not Gaussian

Lindeberg condition

- Instead, we need a stronger (albeit less intuitive) condition known as the *Lindeberg condition*: for every $\epsilon > 0$,

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

as $n \rightarrow \infty$

- This condition is stronger in the sense that $L \implies F$ (Lindeberg implies Feller)
- As we'll see shortly, it also guarantees asymptotic normality

Is Lindeberg necessary?

- But first: is it necessary? I.e., is it possible to have asymptotic normality even if the Lindeberg condition does not hold?
- Of course it is: sums of Gaussian random variables are always Gaussian, even if one term dominates
- However, this is something of a “fragile” result: normality depends entirely on the original distribution — it doesn't arise from lots of small contributions

Is Lindeberg necessary? (cont'd)

- Feller (1935) showed that if you have asymptotic normality arising from infinite negligible contributions, then the Lindeberg condition is in fact necessary
- **Theorem (Feller):** Suppose $\{X_{ni}\}$ is a triangular array with $Z_n = \sum_{i=1}^n X_{ni}$ and $s_n^2 = \mathbb{V}Z_n$. If $Z_n/s_n \xrightarrow{d} N(0, 1)$ and $\max_i \sigma_{ni}^2/s_n^2 \rightarrow 0$, then the Lindeberg condition holds.
- In other words, the Lindeberg condition is the minimal requirement for a Gaussian limit
- For example, there is another CLT for non-iid variables called the Lyapunov CLT, which requires a “Lyapunov condition” (which implies the Lindeberg condition), a stronger condition than necessary for asymptotic normality

Lindeberg condition: Example

- 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 $X_1, X_2, \dots$ are iid with mean zero and finite variance. Then the Lindeberg condition is satisfied.
- There are three key steps in this proof:
 - Replacing the infinite sum with a single quantity $\propto \mathbb{E}T_n$
 - $T_n \xrightarrow{P} 0$ (which happens if $s_n \rightarrow \infty$)
 - $\mathbb{E}T_n \rightarrow 0$ 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 worst-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 $s_n^2 = \mathbb{V}Z_n$. If the Lindeberg condition holds: for every $\epsilon > 0$,

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

then $Z_n/s_n \xrightarrow{d} N(0, 1)$.

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 $Z_n = \frac{1}{n} \sum_{i=1}^n X_{ni}$, $s_n^2 = \frac{1}{n} \sum_{i=1}^n \mathbb{V}X_{ni}$, and $s_n^2 \rightarrow s^2 \neq 0$. 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 $\sqrt{n}Z_n \xrightarrow{d} N(0, s^2)$.

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

Multivariate CLT

- 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 $\{\mathbf{x}_{ni}\}$ is a triangular array of $d \times 1$ random vectors such that $\mathbf{z}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_{ni}$ and $\mathbf{V}_n = \frac{1}{n} \sum_{i=1}^n \mathbb{V} \mathbf{x}_{ni} \rightarrow \mathbf{V}$, where $\mathbf{V}$ is positive definite. If for every $\epsilon > 0$,

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}\{\|\mathbf{x}_{ni}\|^2 1(\|\mathbf{x}_{ni}\| \geq \epsilon\sqrt{n})\} \rightarrow 0,$$

then $\sqrt{n}\mathbf{z}_n \xrightarrow{d} N(\mathbf{0}, \mathbf{V})$.

- Or equivalently, $\sqrt{n}\mathbf{V}_n^{-1/2}\mathbf{z}_n \xrightarrow{d} N(\mathbf{0}, \mathbf{I})$

Multivariate Feller condition

- 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

$$\frac{\mathbb{V} \mathbf{x}_i}{\sum_{j=1}^n \mathbb{V} \mathbf{x}_j} \rightarrow \mathbf{0}_{d \times d}$$

for all i ; the division here is element-wise

CLT for linear regression

- 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 $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^* + \mathbf{w}$, where $w_i \stackrel{\text{iid}}{\sim} (0, \sigma^2)$. Suppose $\frac{1}{n}\mathbf{X}^\top \mathbf{X} \rightarrow \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is positive definite, and let $\mathbf{x}_i$ denote the $d \times 1$ vector of covariates for subject i (taken to be fixed, not random). If $\|\mathbf{x}_i\|$ is uniformly bounded, then

$$\frac{1}{\sigma}(\mathbf{X}^\top \mathbf{X})^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}).$$

- In other words, $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}^*, \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1})$

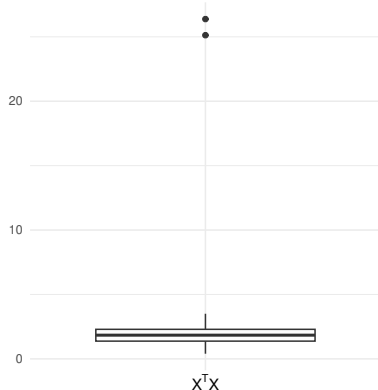
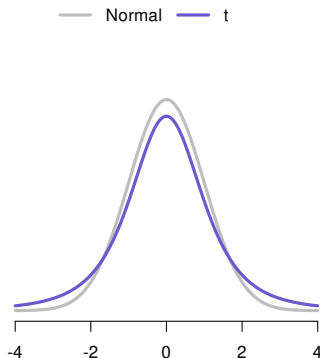
Remarks

- 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
 - $\|\mathbf{x}_i\|$ bounded; 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 $\|\mathbf{x}_i\|$ is measuring: in words, we are requiring that no one observation can exhibit too great an influence

Simulation

- 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:
 - $\mathbf{w}$ will follow a t distribution with ν degrees of freedom
 - The elements of $\mathbf{X}$ will be uniformly distributed (from -1 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

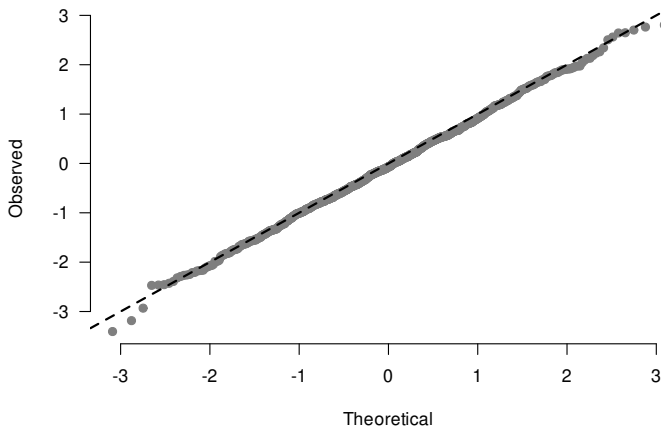
Illustration of the two conditions ($\nu = 3, a = 5$)



As we will see, the more comfortably the Lindeberg condition holds, the faster the rate of convergence to normality

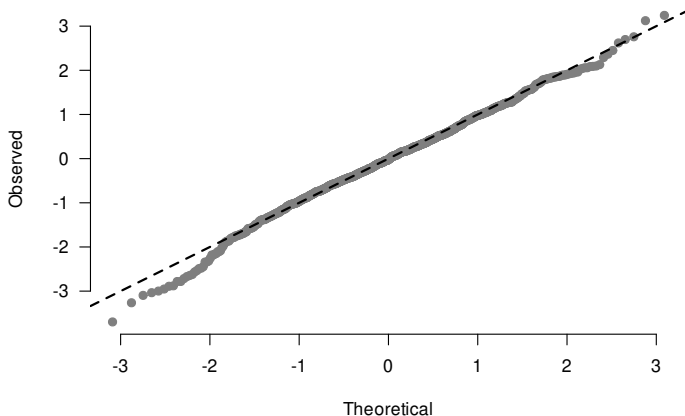
Results: $\nu = 50, a = 5$

Influential observations, but ε close to normal



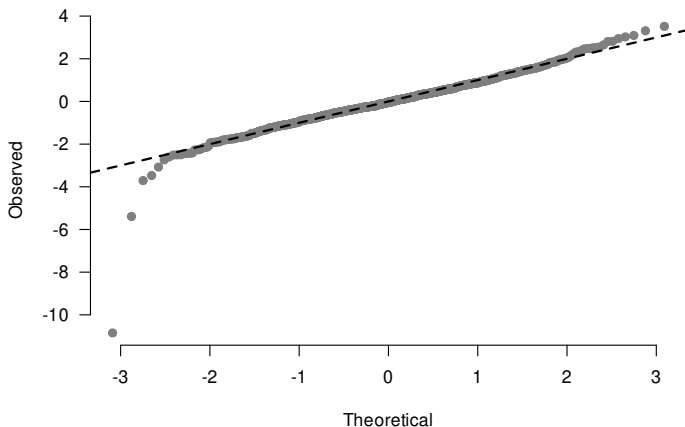
Results: $\nu = 3, a = 1$

Heavy tails, but no terribly influential observations



Results: $\nu = 3, a = 5$

Heavy tails and influential observations



Results: $\nu = 3, a = 5$

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

