STATS310A - Lecture 16

Persi Diaconis
Scribed by Michael Howes

11/16/21

Contents

1 Announcements

Final homeowork uploaded to Canvas and due November 30.

- Read chpaters 25,26.
- Do 25 / 1,3 and 26 1,3,12-14.
- The hints contain surprises.

2 The central limit theorem

We are in the process of proving Lindeberg's version of the central limit theorem. Recall that we have a triangular array of random variables $\{X_{n,i}\}$ where $i = 1, \ldots, k_n$ and $n = 1, 2, \ldots$. We assume that the array has independent rows. That is, for each n, $\{X_{n,i}\}_{i=1}^{k_n}$ are independent. Assume also that

$$\mathbb{E}[X_{n,i}] = 0 \quad \text{and} \quad \sigma_{i,n}^2 = \mathrm{Var}(X_{n,i}) < \infty.$$

Define $S_n = \sum_{i=1}^{k_n} X_{n,i}$ and $s_n^2 = \sum_{i=1}^{k_n} \sigma_{i,n}^2 = \mathrm{Var}(S_n)$.

Definition 1. A triangular array with independent rows $\{X_{n,i}\}$ is said to satisfy *Lindeberg's condtion* if for all $\varepsilon > 0$

$$\frac{1}{s_n^2} \sum_{i=1}^{k_n} \int_{\{|X_{n,i}| > \varepsilon s_n\}} |X_{n,i}|^2 d\mathbb{P} \xrightarrow{n} 0.$$

Lindeberge's version of the central limit theorem is:

Theorem 1 (Lindeberg)**.** *Let $\{X_{n,i}\}$ be a triangular array with independent rows. If $\{X_{n,i}\}$ satisfy Lindeberg's condtion, then for all $x \in \mathbb{R}$,*

$$\mathbb{P}\left(\frac{S_n}{s_n} \leq x\right) \to \Phi(x),$$

where $\Phi(x) = \mathbb{P}(Z \leq x)$ for $Z \sim \mathcal{N}(0, 1)$.

Proof. To prove this we will use the portmanteau theorem. Let $C_c^\infty(\mathbb{R})$ be the class of infinitely differentiable functions on $\mathbb{R}$ with compact support. By the portmanteau theorem it suffices to show that for all $f \in C_c^\infty(\mathbb{R})$, $\mathbb{E}[f(S_n/s_n)] \overset{n}{\to} \mathbb{E}[f(Z)]$ where $Z \sim \mathcal{N}(0,1)$. Thus fix such an f. Define $Z_{n,i}$ is be independent random variables such that $Z_{n,i} \sim \mathcal{N}(0, \sigma_{n,i}^2)$. Let $Z_n = \sum_{i=1}^{k_n} Z_{n,i}$, Then

$$Z = \frac{1}{s_n} Z_n \sim \mathcal{N}(0,1).$$

The idea behind the proof is to swap out $X_{n,i}$ for $Z_{n,i}$ one at a time. With this in mind, define

$$T_{n,i} = X_{n,1} + \ldots + X_{n,i-1} + Z_{n,i} + \ldots + Z_{n,k_n}.$$

Note that X_i, Z_i are independent of $T_{n,i}$ for each i. Furthermore we have

$$S_n = T_{n,k_n} + Z_{n,k_n} \quad \text{and} \quad Z_n = T_{n,1} + Z_{n,1}.$$

And also

$$T_{n,i} + Z_{n,i} = T_{n,i-1} + X_{n,i-1},$$

for $i = 2, \ldots, k_n$. Thus by telescoping we have

$$f\left(\frac{S_n}{s_n}\right) - f\left(\frac{Z_n}{s_n}\right) = \sum_{i=1}^{k_n} f\left(\frac{T_{n,i} + X_{n,i}}{s_n}\right) - f\left(\frac{T_{n,i} + Z_{n,i}}{s_n}\right).$$

And so

$$\left|\mathbb{E}\left[f\left(\frac{S_n}{s_n}\right)\right] - \mathbb{E}\left[f(Z)\right]\right| \leq \sum_{i=1}^{k_n} \left|\mathbb{E}\left[f\left(\frac{T_{n,i} + X_{n,i}}{s_n}\right)\right] - \mathbb{E}\left[f\left(\frac{T_{n,i} + Z_{n,i}}{s_n}\right)\right]\right| \tag{1}$$

We will now use Taylor's approximation to bounded each of the terms in the above sum. For $h \in \mathbb{R}$, define

$$g(h) = \left|f(x+h) - f(x) - hf'(x) - \frac{h^2}{2} f''(x)\right|.$$

Since all deriviates of f are bounded, Taylor's approximation with remainder says that there exists $k > 0$ such that for all h and x

$$g(h) \leq k \min\{|h|^3, |h|^2\}.$$

Thus for all x, h_1, h_2 we have

$$\left|f(x+h_1) - f(x+h_2) - f'(x)(h_1 - h_2) - \frac{1}{2} f''(x)(h_1^2 - h_2^2)\right| = |g(h_1) - g(h_2)| \leq |g(h_1)| + |g(h_2)|.$$

We wish to apply this to equation (1) with $x = \frac{T_{n,i}}{s_n}$, $h_1 = \frac{X_{n,i}}{s_n}$ and $h_2 = \frac{Z_{n,i}}{s_n}$. Thus we need to add the high order terms $f'(x)(h_1 - h_2)$ and $\frac{1}{2} f''(x)(h_1^2 - h_2^2)$. Since $X_{n,i}$ and $Z_{n,i}$ have the same mean and variance and $X_{n,i}, Z_{n,i}$ are independent of $T_{n,i}$, we have

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{T_{n,i} + X_{n,i}}{s_n}\right)\right] - \mathbb{E}\left[f\left(\frac{T_{n,i} + Z_{n,i}}{s_n}\right)\right] \\
=&
\end{aligned}$$

$\square$