Lecture 13 : Lindeberg's Theorem and the Helly-Bray Selection

STAT205 *Lecturer: Jim Pitman* Scribe: *Timothy Wu* <tcb6402@berkeley.edu>

This set of notes is a revision of the work of Lawrence Christopher Evans and David S. Rosenberg.

Abstract

In this lecture we begin with a review of Lindeberg's Theorem and its applications. We then build up the tools used in the Helly-Bray Selection Principle and we finish with its proof. The lecture provided here corresponds with sections 2.2 and 2.4 of [1].

13.1 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;

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

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

13.2 Lindeberg's Theorem

We will write $\mathcal{L}(X)$ 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. Recall:

Theorem 13.1 (Lindeberg's Theorem) *Suppose that in addition to the Triangular Array Conditions, the triangular array satisfies* Lindeberg's Condition*:*

$$\forall \epsilon > 0, \lim_{i \to \infty} \sum_{j=1}^{n_i} \mathbb{E}[X_{ij}^2 \mathbf{1}(|X_{ij}| > \epsilon)] = 0 \tag{13.1}$$

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

13.2.1 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 $\mathbf{s}_n^2 = \sum_{j=1}^n \sigma_j^2$. We want to know when $\mathcal{L}(S_i/\mathbf{s}_i) \to \mathcal{N}(0, 1)$. To this end, we check Lindeberg's 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

$$\lim_{n \to \infty} \sum_{j=1}^{n} \mathbb{E}\left[\frac{X_j^2}{\mathbf{s}_n^2} \mathbf{1}\left(\left|\frac{X_j}{\mathbf{s}_n}\right| > \epsilon\right)\right] = 0, \text{ for all } \epsilon > 0, \tag{13.2}$$

$$\text{i.e. } \lim_{n \to \infty} \frac{1}{\mathbf{s}_n^2} \sum_{j=1}^{n} \mathbb{E}\left[X_j^2 \mathbf{1}(|X_j| > \epsilon \mathbf{s}_n)\right] = 0, \text{ for all } \epsilon > 0. \tag{13.3}$$

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. Lyapounov's 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 Lindeberg's condition. The proof of this is given (essentially) in the previous lecture.

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.

13.3 Extended Distribution Functions

Extended distribution functions are an extension of distribution functions to the case where we allow mass to exist at $\pm\infty$. In the case of a cumulative distribution function F, we require that $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$. For the extended distribution function we relax this condition. This is convenient, as the limit of distribution functions is often not a proper cumulative distribution function.

Example 13.2 *Let $F_n = \delta_n$, the delta measure at n. Then as $n \to \infty$, $F_n \Rightarrow 0$. However, 0 is not a cumulative distribution function since* $\lim_{x\to\infty} 0 \neq 1$

So to deal with the case of mass at $\pm\infty$, we define the extended distribution function.

Definition 13.3 (Extended Distribution Function) *A function $F : \Re \to [0, 1]$ which is right continuous and nondecreasing is called an* extended distribution function. *We define $F(-\infty) := \lim_{x\downarrow-\infty} F(x)$ and $F(\infty) := \lim_{x\uparrow\infty} F(x)$ and thus there is a bijection between extended distribution functions and probability measures on $[-\infty, \infty]$ by the relation $\mu[-\infty, x] = F(x)$ for all $x < \infty$, and $\mu[x, \infty] = 1 - F(x)$.*

Note here that if Y has extended distribution function F, $F(-\infty) = P(Y = -\infty)$ but $F(\infty) = 1 - P(Y = \infty)$. Also note that for any e.d.f. F, if $F(\infty) = 1$ and $F(-\infty) = 0$, F is simply a c.d.f. We see now that although the function $F \equiv 0$ that we encountered above is not a c.d.f., it is an e.d.f.(extended distribution function) with $F(\infty) = 0$. Next, we look at theorems dealing with the limits of sequences of cumulative distribution functions and extended distribution functions.

13.4 The Helly-Bray Selection Principle

Theorem 13.4 (Helly-Bray Selection Principle) *Every sequence of extended distribution functions F_n has a subsequence $F_{n(k)}$ such that $F_{n(k)} \to F(x)$ for all continuity points x of F for some extended distribution function F.*

Before we prove the main theorem, we introduce the following Lemma:

Lemma 13.5 *Let $D \subset \mathbb{R}$ be dense. Let F_n be a sequence of e.d.f.s such that $\lim_{n\to\infty} F_n(d) = F_\infty(d)$, $\forall\, d \in D$. Then, $F_n \Rightarrow F_\star$ where $F_\star(x) := \inf_{x<d\in D} F_\infty(d)$.*

The proof of this lemma is left as an exercise to the reader. We now proceed to the proof of the Helly-Bray selection principle.

Proof: Let F_n be a sequence of e.d.f.s, and let $D = \{d_1, d_2, \ldots\}$ be any countable, dense set. Using Cantor's diagonal argument, we can find a subsequence $n(k)$ such that $F_{n(k)}(d) \to F(d)$ for all $d \in D$. We then simply apply the previous lemma and see that $F_{n(k)} \Rightarrow F_\star$. ■

The Helly-Bray selection principle as stated above begs the question: under what conditions can we find a subsequence converging to a *cumulative distribution function*? To answer this we first introduce the notion of tightness:

Definition 13.6 (Tightness) *A collection B of proper distributions (c.d.f.s) F is called* tight *if* $\lim_{x\to\infty} \sup_{F\in B} F(-x,x)^c = 0$. *In other words, if* $F(-x,x)^c \to 0$ *as* $x \to \infty$ uniformly over $F \in B$.

Combining the property of tightness with the Helly-Bray selection principle, we get:

Theorem 13.7 (Variation of the Helly-Bray Selection Principle) *Every* tight *sequence of proper distribution functions F_n has a subsequence $F_{n(k)}$, such that $F_{n(k)} \Rightarrow F$ where F is a* proper *distribution function.*

The proof of this theorem is immediate from the Helly-Bray selection principle and the definition of tightness.

13.5 Application of Helly-Bray

We use the Helly-Bray selection principle for the following important application:

Definition 13.8 *A collection D of bounded, continuous functions $f : \mathbb{R} \to \mathbb{R}$ is called a* determining class *if every probability measure P on $\mathbb{R}$ is characterized by the integrals $\int f\,dP$ for $f \in D$. (i.e. $\int f\,dP = \int f\,dQ, \forall\, f \in D \Leftrightarrow P(-\infty, x] = Q(-\infty, x], \forall\, x \in \mathbb{R}$).*

Theorem 13.9 *Let D be a determining class and let F_n be a sequence of probability measures on the line. If $\int f\,dF_n$ converges to some limit for all $f \in D$ and (F_n) is tight, then $F_n \Rightarrow F$, where F is the unique distribution determined by $\int f\,dF = \lim_{n\to\infty} \int f\,dF_n$ for all $f \in D$.*

Before we prove this theorem let us recall a lemma:

Lemma 13.10 *Let x_n be a sequence in a metric space S. $\exists\, x \in S$ such that $x_n \to x \Leftrightarrow \forall$ subsequence $x_{n(k)}$ there is a further subsequence $x_{n(k(l))}$ such that $x_{n(k(l))}$ converges to x.*

The proof of this lemma is an easy exercise using proof by contradiction. Now we prove the theorem:

Proof: For any subsequence $F_{n(k)}$ tightness and the Helly-Bray selection principle tells us that there is a further subsequence $F_{n(k(l))}$ such that $F_{n(k(l))} \Rightarrow G$ for some distribution function G. $F_{n(k(l))} \Rightarrow G$ in turn implies that $\int f\,dG = \lim_{l\to\infty} \int f\,dF_{n(k(l))}$ for all bounded, countinuous f. So in particular the equality holds for all $f \in D$.

But by our assumption, we know that $\lim_{l\to\infty} \int f\,dF_{n(k(l))} = \int f\,dF$. These two statements together imply that $\int f\,dG = \int f\,dF$ for all $f \in D$. But D is a determining class and so $\int f\,dG = \int f\,dF$ for all bounded, countinuous f. Thus we have that $\int f\,dF = \lim_{l\to\infty} \int f\,dF_{n(k(l))}$ for all bounded, countinuous f. In other words, we know that $F_{n(k(l))} \Rightarrow F$.

Now we recall that weak convergence is equivalent to convergence in the Lévy metric ([1], chapter 2.3, exercise 2.15). We have thus shown that every subsequence of F_n has a further subsequence converging to F. Since the Lévy metric is a metric, our lemma gives us that $F_n \Rightarrow F$.

∎

References

[1] Richard Durrett. *Probability: theory and examples, 3rd edition.* Thomson Brooks/Cole, 2005.