Central Limit Theorems
ECO2402F

1 Introduction

The fact that distributions of sums of random variables can be approximated by Normal distributions has been known to various degrees since the 18th century. Since DeMoivre proved that the Binomial distribution could be approximated by a Normal distribution, central limit theorems (CLTs) have been proved in a variety of problems.

The purpose of this handout is not to give an exhaustive survey of CLTs (as this would be impossible) but rather to outline some of the main CLTs for sums of independent and dependent random variables.

A very nice survey on the early (pre-1935) history of CLTs can be found in Le Cam (1986).

2 Sums of independent random variables

We will begin by stating (and proving) the most general CLT for sums of independent random variables. This result was established by Lindeberg (1922) and effectively has all the other CLTs for sums of independent random variables as special cases.

LINDEBERG CLT. Let $\{X_{ni} : n \geq 1; i = 1, \cdots, k_n\}$ be a triangular array of random variables with $X_{n1}, \cdots, X_{nk_n}$ independent for each n, $E(X_{ni}) = 0$ and

$$\sum_{i=1}^{k_n} E(X_{ni}^2) = 1.$$

If for each fixed $\epsilon > 0$,

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} E\left[X_{ni}^2 I(|X_{ni}| > \epsilon)\right] = 0$$

then

$$Z_n = \sum_{i=1}^{k_n} X_{ni} \to_d Z \sim N(0, 1).$$

Before giving a proof of the Lindeberg CLT, note that the CLT for i.i.d. random variables is a special case of the Lindeberg CLT. More precisely, if $X_1, X_2, \cdots$ is a sequence of i.i.d. random variables with mean 0 and variance σ^2 then the asymptotic normality of $Z_n =$

$n^{-1/2}\sum_{i=1}^n X_i$ follows by setting $k_n = n$ and $X_{ni} = X_i/\sqrt{n}$; note that

$$\begin{aligned}
\sum_{i=1}^{k_n} E\left[X_{ni}^2 I(|X_{ni}| > \epsilon)\right] &= \frac{1}{n}\sum_{i=1}^n E[X_i^2 I(|X_i| > \epsilon\sqrt{n})] \\
&= E[X_1^2 I(|X_1| > \epsilon\sqrt{n})] \\
&\to 0 \quad \text{as } n \to \infty
\end{aligned}$$

since $E(X_i^2) < \infty$.

Proof. There are a variety of approaches to proving the Lindeberg CLT; textbooks typically show convergence of the characteristic functions (or moment generating functions). We will show that $E[f(Z_n)] \to E[f(Z)]$ for all infinitely differentiable functions f with bounded derivatives. (Any bounded continuous function can be approximated by a sequence of such functions.) This proof is essentially the same as Lindeberg's (1922) proof and provides considerable insight as to why sums of random variables are approximately Normal under appropriate conditions.

For each "row" of the triangular array, define independent random variables Y_{ni} ($i = 1, \cdots, k_n$), independent of the X_{ni}'s, with $Y_{ni} \sim N(0, E(X_{ni}^2))$. Then define

$$S_j = \sum_{i=1}^{j-1} X_{ni} + \sum_{i=j+1}^{k_n} Y_{ni}$$

and note that

(a) $S_{k_n} + X_{nk_n} = Z_n$;

(b) $S_1 + Y_{n1} \sim N(0, 1)$;

(c) $S_j + X_{nj} = S_{j+1} + Y_{n,j+1}$ for $j = 1, \cdots, k_n - 1$.

Set $Z = S_1 + Y_{n1} \sim N(0, 1)$; then we have

$$f(Z_n) - f(Z) = \sum_{j=1}^{k_n} [f(S_j + X_{nj}) - f(S_j + Y_{nj})]$$

and so

$$|E[f(Z_n)] - E[f(Z)]| \le \sum_{j=1}^{k_n} |E[f(S_j + X_{nj})] - E[f(S_j + Y_{nj})]|\,.$$

By a Taylor series expansion,

$$\begin{aligned}
f(S_j + X_{nj}) &= f(S_j) + X_{nj} f'(S_j) + \frac{1}{2} X_{nj}^2 f''(S_j) + R(S_j, X_{nj}) \\
f(S_j + Y_{nj}) &= f(S_j) + Y_{nj} f'(S_j) + \frac{1}{2} Y_{nj}^2 f''(S_j) + R(S_j, Y_{nj})
\end{aligned}$$

where the remainder term $R(x,y)$ has the bounds

$$\begin{aligned}|R(x,y)| &\leq \frac{1}{6}M_3|y|^3\\ |R(x,y)| &\leq M_2 y^2\end{aligned}$$

with M_2 and M_3 bounds on $|f''|$ and $|f'''|$. Taking expectations term by term, we get

$$|E[f(Z_n)] - E[f(Z)]| \leq \sum_{j=1}^{k_n} \{E[|R(S_j, X_{nj})|] + E[|R(S_j, Y_{nj})|]\}.$$

Now let $C = \max(M_2, M_3/6)$; then for $0 < \epsilon < 1$,

$$\begin{aligned}|R(S_j, X_{nj})| &\leq C|X_{nj}|^3 I(|X_{nj}| \leq \epsilon) + CX_{nj}^2 I(|X_{nj}| > \epsilon)\\ &\leq C\epsilon X_{nj}^2 + CX_{nj}^2 I(|X_{nj}| > \epsilon)\end{aligned}$$

which leads to

$$\begin{aligned}\limsup_{n\to\infty} \sum_{j=1}^{k_n} E[|R(S_j, X_{nj})|] &\leq \limsup_{n\to\infty} \left\{C\epsilon \sum_{j=1}^{k_n} E(X_{nj}^2) + C\sum_{j=1}^{k_n} E\left[X_{nj}^2 I(|X_{nj}| > \epsilon)\right]\right\}\\ &= C\epsilon\end{aligned}$$

and so taking $\epsilon \to 0$, it follows that

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} E[|R(S_j, X_{nj})|] = 0.$$

Likewise, it follows that

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} E[|R(S_j, Y_{nj})|] = 0$$

which completes the proof of the theorem. □

The condition

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} E\left[X_{ni}^2 I(|X_{ni}| > \epsilon)\right] = 0$$

is known as the **Lindeberg condition**. If, in addition to the assumptions given above, we assume that

$$\lim_{n\to\infty} \max_{1\leq i\leq k_n} E(X_{ni}^2) = 0$$

then the Lindeberg condition is a necessary as well as sufficient condition for asymptotic normality.

We indicated above that the Lindeberg CLT applies trivially to sums of i.i.d. random variables. Next we will consider weighted sums of i.i.d. random variables. Suppose that

$X_1, X_2, \cdots$ are i.i.d. random variables (with mean 0 and variance 1), $\{c_{ni}; n \geq 1, i = 1, \cdots, n\}$ are constants and define

$$Z_n = \frac{1}{\sigma_n} \sum_{i=1}^{n} c_{ni} X_i$$

where $\sigma_n^2 = c_{n1}^2 + \cdots + c_{nn}^2$; note that $E(Z_n) = 0$ and $\text{Var}(Z_n) = 1$. To determine if Z_n converges to a Normal distribution, we need to check the Lindeberg condition with $X_{ni} = c_{ni} X_i / \sigma_n$; we have

$$\sum_{i=1}^{k_n} E\left[X_{ni}^2 I(|X_{ni}| > \epsilon)\right] = \frac{1}{\sigma_n^2} \sum_{i=1}^{n} c_{ni}^2 E[X_i^2 I(|X_i| > \epsilon\sigma_n/|c_{ni}|)].$$

We must determine conditions on the c_{ni}'s to guarantee that the right hand side above tends to 0 as $n \to \infty$; a sufficient condition for this is $\sigma_n/|c_{ni}| \to \infty$ for all $1 \leq i \leq n$ or, equivalently,

$$\max_{1 \leq i \leq n} \frac{c_{ni}^2}{\sigma_n^2} \to 0$$

as $n \to \infty$.

A somewhat weaker CLT (than the Lindeberg CLT) was proved by Lyapounov who used (in place of the Lindeberg condition) the condition

$$\sum_{i=1}^{k_n} E[|X_{ni}|^{2+\delta}] \to 0 \quad \text{as } n \to \infty$$

where as before, we assume that

$$\sum_{i=1}^{k_n} E(X_{ni}^2) = 1.$$

It is easy to show that Lyapounov's condition implies Lindeberg's condition; however, in many applications, Lyapounov's condition is easier to verify (particularly if $E[|X_{ni}|^3] < \infty$) and so is quite useful in practice.

3 Martingales

CLTs for sums of independent random variables can easily be extended to martingales (see definition below). The intuition for this becomes clear if one looks at the proof given above of the Lindeberg CLT; it is not the independence of the X_{ni}'s that is particularly important but rather the fact that independence implies that the random variables are uncorrelated. It turns out that martingale theory provides the necessary structure to exploit a lack of correlation.

Martingales are typically defined with respect to an increasing sequence of σ-fields (or σ-algebras). σ-fields are simply collections of events satisfying some conditions; in practical terms, they represent what is known about a certain process (or experiment) at a given point in time. For example, suppose we observe values of random variables $X_0, X_1, \cdots, X_n$. Then we could define a σ-field $\mathcal{A}_n$ to represent the accumulated information about $X_0, \cdots, X_n$:

$$\mathcal{A}_n = \sigma(X_0, \cdots, X_n).$$

In this case, we could define the conditional expected value $E(Y|\mathcal{A}_n)$ to be

$$E(Y|\mathcal{A}_n) = E(Y|X_0, \cdots, X_n),$$

that is, the expected value of Y given the values of $X_0, \cdots, X_n$.

DEFINITION. Let $\{\mathcal{A}_n\}$ be an increasing sequence of σ-fields ($\mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots$) and let $\{X_n\}$ be a sequence of random variables with X_n $\mathcal{A}_n$-measurable. Then $\{X_n\}$ is a **martingale** with respect to $\{\mathcal{A}_n\}$ if $E(X_n|\mathcal{A}_{n-1}) = X_{n-1}$.
If $\{X_n\}$ is a martingale then the sequence $X_1, X_2 - X_1, X_3 - X_2, \cdots$ is called a **martingale difference sequence.**

In most cases, the σ-field $\{\mathcal{A}_n\}$ is simply taken to be $\sigma(X_n, X_{n-1}, X_{n-2}, \cdots)$, the σ-field generated by $X_n, X_{n-1}, X_{n-2}, \cdots$. If we say that $\{X_n\}$ is a martingale without reference to σ-fields $\{\mathcal{A}_n\}$ then it is implicitly understood that $\mathcal{A}_n = \sigma(X_n, X_{n-1}, X_{n-2}, \cdots)$. It is easy to verify that martingale differences are uncorrelated random variables although they are not necessarily independent.

The Lindeberg CLT can be extended to martingales in a fairly straightforward manner.

MARTINGALE CLT. Let $\{X_{ni} : n \geq 1; i = 1, \cdots, k_n\}$ be a martingale difference array; that is, for each fixed n, $\{X_{ni}\}$ is a martingale difference sequence. If

(a) $\sum_{i=1}^{k_n} E(X_{ni}^2|\mathcal{A}_{n,i-1}) \to_p \sigma^2$, and

(b) $\sum_{i=1}^{k_n} E\left[X_{ni}^2 I(|X_{ni}| > \epsilon)|\mathcal{A}_{n,i-1}\right] \to_p 0$ for each $\epsilon > 0$

then

$$Z_n = \sum_{i=1}^{k_n} X_{ni} \to_d Z \sim N(0, \sigma^2)$$

A proof of this CLT for martingales can be found in Pollard (1984). Note that condition (b) is a martingale analogue of the Lindeberg condition. Applications of this CLT for martingales can also be found in Pollard (1984).

4 Sums of dependent random variables

Finally, we will consider CLTs for sums of dependent random variables. To simplify the situation somewhat, we will focus on sums of stationary random variables (see definition below).

DEFINITION. A stochastic process $\{X_t : t = 0, \pm 1, \pm 2, \cdots\}$ is **strongly stationary** if the joint distribution of $(X_{t+1}, X_{t+2}, \cdots, X_{t+s})$ is independent of t for each s. $\{X_t\}$ is **weakly stationary** (or stationary) if $E(X_t^2) < \infty$, $E(X_t)$ is independent of t and $\text{Cov}(X_t, X_{t+s})$ depends only on $|s|$ (and not on t).

Suppose that $\{X_t\}$ is a stationary (strongly or weakly) process (with mean 0) and define

$$Z_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t$$

In order for Z_n to converge in distribution to a Normal random variable, we need to insure that the dependence between X_t and X_{t+s} becomes negligible as $s \to \infty$. There are several ways to do this as will be seen below; these approaches will allow us to approximate Z_n by

$$Z_n' = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t$$

(in the sense that $Z_n - Z_n' \to_p 0$) where $\{Y_t\}$ are either independent random variables or martingale differences. The idea here is to write X_t as

$$X_t = Y_t + V_t$$

where $\{Y_t\}$ are i.i.d. or martingale differences, and $\{V_t\}$ is a stationary process whose variability is negligible at "low" frequencies; we usually write $V_t = \eta_t - \eta_{t-1}$ where $\{\eta_t\}$ is stationary so that

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t + \frac{1}{\sqrt{n}} (\eta_n - \eta_0) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t + o_p(1). \end{aligned}$$

One way to describe the weakening of dependence between X_t and X_{t+s} as s increases is through **mixing conditions**. It turns out that there are several types of mixing although here we will concentrate on strong mixing, which is perhaps the simplest type of mixing.

DEFINITION. Let $\{X_t\}$ be a strongly stationary stochastic process. Define

$$\alpha_k = \sup_{A,B} |P(A \cap B) - P(A)P(B)|$$

where the supremum is taken over $A \in \sigma(\cdots, X_{t-2}, X_{t-1}, X_t)$ and $B \in \sigma(X_{t+k}, X_{t+k+1}, \cdots)$. $\{X_t\}$ is **strong mixing** if $\alpha_k \to 0$ as $k \to \infty$; the α_k's are called the mixing coefficients.

If a process $\{X_t\}$ is strong mixing then the dependence between X_t and X_{t+s} becomes negligible as $s \to \infty$. In order to prove CLTs for sums of stationary strong mixing random variables, we need to assume some sort of summability condition for the mixing coefficients, which implies that these coefficients tend to 0 at a certain rate. Generally speaking, it is difficult to verify that a given stochastic process is strong mixing although in many practical situations, it does seem to be a reasonable condition; for this reason, strong mixing is often assumed as a way of grossly describing the dependence structure of a stochastic process.

THEOREM (Gordin, 1969). Suppose that $\{X_t\}$ is strongly stationary and strong mixing with $E(X_t) = 0$, $E\left[|X_t|^{2+\delta}\right] < \infty$ for some $\delta > 0$ and $\sum_{k=1}^{\infty} \alpha_k^{\delta/(2+\delta)} < \infty$. Then there exist stationary processes $\{Y_t\}$ and $\{\eta_t\}$ such that

$$X_t = Y_t + \eta_t - \eta_{t-1}$$

where

(i) $E(Y_t|X_{t-1}, X_{t-2}, \cdots) = 0$,

(ii) $E(Y_t^2) < \infty$,

(iii) $E(\eta_t^2) < \infty$,

(iv) $n^{-1} \sum_{t=1}^{n} E(Y_t^2|X_{t-1}, X_{t-2}, \cdots) \to_p \mathrm{Var}(X_t) + 2\sum_{s=1}^{\infty} \mathrm{Cov}(X_t, X_{t+s})$.

With this result in hand, we can write

$$\begin{aligned} Z_n &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (Y_t + \eta_t - \eta_{t-1}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t + \frac{1}{\sqrt{n}} (\eta_n - \eta_0) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t + o_p(1) \end{aligned}$$

Note that condition (i) above implies that $\{Y_t\}$ is a martingale difference sequence; the stationarity allows us to apply the CLT for martingales given in the previous section.

Next we will assume that $\{X_t\}$ is a **linear process**; that is, X_t can be written as

$$X_t = \sum_{u=0}^{\infty} c_u \varepsilon_{t-u}$$

where $\{\varepsilon_t\}$ is an i.i.d. (or stationary martingale difference) sequence with $E(\varepsilon_t) = 0$ and $\mathrm{Var}(\varepsilon_t) = 1$. Examples of linear processes include autoregressive-moving average (ARMA) processes as well as some long memory processes. Define

$$\tilde{c}_u = \sum_{s=u+1}^{\infty} c_s$$

and assume that $\sum_{u=0}^{\infty} \tilde{c}_u^2 < \infty$. (ARMA processes satisfy this summability condition.) Then we can write

$$X_t = \left(\sum_{u=0}^{\infty} c_u\right) \varepsilon_t + \eta_t - \eta_{t-1}$$

where $\eta_t = \sum_{u=0}^{\infty} \tilde{c}_u \varepsilon_{t-u}$. This representation of $\{X_t\}$ is called the **Beveridge-Nelson decomposition** (Beveridge and Nelson, 1981; see also Phillips and Solo, 1992) and, although its proof is simply high school algebra, it is very useful in practice. Again, we have

$$\begin{aligned} Z_n &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t = \left(\sum_{u=0}^{\infty} c_u\right) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t + \frac{1}{\sqrt{n}} (\eta_n - \eta_0) \\ &= \left(\sum_{u=0}^{\infty} c_u\right) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t + o_p(1) \end{aligned}$$

References

Beveridge, S. and Nelson, C.R. (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle". *Journal of Monetary Economics.* **7**, 151-174.

Gordin, M.I. (1969) The central limit theorem for stationary processes. *Soviet Mathematics Doklady.* **10**, 1174-1176.

Le Cam, L. (1986) The central limit theorem around 1935 (with discussion). *Statistical Science.* **1**, 78-96.

Lindeberg, J.W. (1922) Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematiche Zeitschrift.* **15**, 211-225.

Phillips, P.C.B. and Solo, V. (1992) Asymptotics for linear processes. *Annals of Statistics.* **20**, 971-1001.

Pollard, D. (1984) *Convergence of Stochastic Processes.* New York: Springer.