Central Limit Theorem

We'll prove the central limit theorem in two different ways. The first using Lindenberg's swapping trick, and the other using characteristic functions.

Central Limit Theorem. *Let $X_1, X_2, ...$ be iid random variables with $\mathbb{E}[X_1^2] < \infty$. Let $S_n := X_1 + ... + X_n$. Then*

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{Var(S_n)}} \xrightarrow{d} \mathcal{N}(0,1).$$

Lindenberg's swapping trick

Proof. For this proof, we do not assume the random variables are iid, just mutually independent. Assume that $\mathbb{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = 1$ (otherwise take $Y_i := \frac{X_i - \mathbb{E}[X_i]}{\sqrt{\mathrm{Var}(X_i)}}$ and proceed as follows). Let $Z \sim \mathcal{N}(0,1)$. Assuming a finite third absolute moment i.e. $\mathbb{E}[|X_i|^3] < \infty$, it is enough to show that for every continuous, bounded, three-times differentiable function g, we have

$$\mathbb{E}\left[g\left(\frac{S_n}{\sqrt{n}}\right)\right] \xrightarrow{n\to\infty} \mathbb{E}[g(Z)] \quad (**).$$

Observation 1. If $Z_1, ..., Z_n$ are iid standard normal, then $\frac{Z_1+...+Z_n}{\sqrt{n}} \sim \mathcal{N}(0,1)$. Now define $T_n := Z_1 + ... + Z_n$ and $S_n := X_1 + ... + X_n$.
Observation 2. We note that (**) is equivalent to showing

$$\mathbb{E}\left[g\left(\frac{S_n}{\sqrt{n}}\right)\right] - \mathbb{E}\left[g\left(\frac{T_n}{\sqrt{n}}\right)\right] \xrightarrow{n\to\infty} 0.$$

Now, the idea is to swap the X_i's to Z_i's one by one. And when n is large, the difference in expectations for the given sums tends to 0. We define auxillary random variables

$$\begin{aligned}
S_n^{(0)} &:= X_1 + X_2 + ... + X_n \\
S_n^{(1)} &:= Z_1 + X_2 + ... + X_n \\
S_n^{(2)} &:= Z_1 + Z_2 + X_3 + ... + X_n \\
&\vdots \\
S_n^{(j)} &:= Z_1 + ... + Z_j + X_{j+1} + ... + X_n \\
&\vdots \\
S_n^{(n)} &:= Z_1 + ... + Z_n.
\end{aligned}$$

Note that $S_n^{(0)} = S_n$ and $S_n^{(n)} = T_n$. Now we can rewrite the expression that we want to tend towards 0 as a telescoping sum. We have

$$\begin{aligned}
\mathbb{E}\left[g\left(\frac{S_n}{\sqrt{n}}\right)\right] - \mathbb{E}\left[g\left(\frac{T_n}{\sqrt{n}}\right)\right] &= \sum_{j=1}^{n}\left(\mathbb{E}\left[g\left(\frac{S_n^{(j-1)}}{\sqrt{n}}\right)\right] - \mathbb{E}\left[g\left(\frac{S_n^{(j)}}{\sqrt{n}}\right)\right]\right) \\
&= \sum_{j=1}^{n}\mathbb{E}\left[g\left(\frac{S_n^{(j-1)}}{\sqrt{n}}\right) - g\left(\frac{S_n^{(j)}}{\sqrt{n}}\right)\right].
\end{aligned}$$

If we let $R_j := Z_1 + ... + Z_{j-1} + X_{j+1} + ... + X_n$ i.e. leaving out the jth random variable, then $S_n^{(j-1)} = R_j + X_j$, and $S_n^{(j)} = R_j + Z_j$. Since g is bounded and three-times differentiable, we can use its Taylor expansion centered around the point r. By Taylor's Theorem, we have

$$g(r+x) = g(r) + xg'(r) + \frac{x^2}{2}g''(r) + \frac{x^3}{6}g'''(r')$$

for some $r' \in (r, r+x)$. If we take the expectation of the Taylor approximation and let r and x be independent, we get

$$\begin{aligned}\mathbb{E}[g(r+x)] &= \mathbb{E}[g(r) + xg'(r) + \frac{x^2}{2}g''(r) + \frac{x^3}{6}g'''(r')] \\ &= \mathbb{E}[g(r)] + \mathbb{E}[x]\mathbb{E}[g'(r)] + \frac{\mathbb{E}[x^2]}{2}\mathbb{E}[g''(r)] + \mathbb{E}\left[\frac{x^3}{6}g'''(r')\right].\end{aligned}$$

Note that we can't split up the expectation in the cubic term because x and r' depend on each other. Now, letting $r = \frac{R_j}{\sqrt{n}}$ and $x = \frac{X_j}{\sqrt{n}}$ which are independent, we have

$$\begin{aligned}\mathbb{E}\left[g\left(S_n^{(j-1)}\right)\right] &= \mathbb{E}\left[g\left(\frac{R_j}{\sqrt{n}}\right)\right] + \mathbb{E}[X_j]\mathbb{E}\left[g'\left(\frac{R_j}{\sqrt{n}}\right)\right] + \frac{1}{2}\mathbb{E}[X_j^2]\mathbb{E}\left[g''\left(\frac{R_j}{\sqrt{n}}\right)\right] + \frac{1}{6}\mathbb{E}\left[X_j^3 g'''\left(\frac{R_j'}{\sqrt{n}}\right)\right] \\ &= \mathbb{E}\left[g\left(\frac{R_j}{\sqrt{n}}\right)\right] + \frac{1}{2n}\mathbb{E}\left[g''\left(\frac{R_j}{\sqrt{n}}\right)\right] + \frac{1}{6n^{3/2}}\mathbb{E}\left[X_j^3 g'''\left(\frac{R_j'}{\sqrt{n}}\right)\right].\end{aligned}$$

Note the first order term disappears because X_j has mean zero. Applying the same technique to $S_n^{(j)}$, we get

$$\mathbb{E}\left[g\left(\frac{S_n^{(j)}}{\sqrt{n}}\right)\right] = \mathbb{E}\left[g\left(\frac{R_j}{\sqrt{n}}\right)\right] + \frac{1}{2n}\mathbb{E}\left[g''\left(\frac{R_j}{\sqrt{n}}\right)\right] + \frac{1}{6n^{3/2}}\mathbb{E}\left[Z_j^3 g'''\left(\frac{\tilde{R}_j'}{\sqrt{n}}\right)\right].$$

The zero, first, and second order terms are all the same, so they cancel when we consider the difference in the telescoping sum above i.e.

$$\mathbb{E}\left[g\left(\frac{S_n^{(j-1)}}{\sqrt{n}}\right)\right] - \mathbb{E}\left[g\left(\frac{S_n^{(j)}}{\sqrt{n}}\right)\right] = \frac{1}{6n^{3/2}}\left(\mathbb{E}\left[X_j^3 g'''\left(\frac{R_j'}{\sqrt{n}}\right)\right] - \mathbb{E}\left[Z_j^3 g'''\left(\frac{\tilde{R}_j'}{\sqrt{n}}\right)\right]\right).$$

Recall that $\mathbb{E}[|X_j^3|] < C_1$ and $|g'''| \leq C_2$ for some $C_1, C_2 \in \mathbb{R}^+$ by assumption, and it's also true that $\mathbb{E}[|Z_j^3|] \leq C_3$ for some $C_3 \in \mathbb{R}^+$. Hence,

$$\mathbb{E}\left[g\left(\frac{S_n^{(j-1)}}{\sqrt{n}}\right)\right] - \mathbb{E}\left[g\left(\frac{S_n^{(j)}}{\sqrt{n}}\right)\right] \leq \frac{1}{6n^{3/2}}[C_2(C_1 + C_3)].$$

Thus,

$$\left|\sum_{j=1}^{n}\left(\mathbb{E}\left[g\left(\frac{S_n^{(j-1)}}{\sqrt{n}}\right)\right] - \mathbb{E}\left[g\left(\frac{S_n^{(j)}}{\sqrt{n}}\right)\right]\right)\right| \leq \sum_{j=1}^{n}\frac{c'}{n^{3/2}} = \frac{c'}{\sqrt{n}}$$

for some $c' \in \mathbb{R}$, and

$$\frac{c'}{\sqrt{n}} \xrightarrow{n \longrightarrow \infty} 0.$$

We have proved the central limit theorem. □

Using characteristic functions

We show that the characteristic function for the scaled sum converges to the characteristic function for a standard normal random variable as $n \longrightarrow \infty$. It is known that when a characteristic function converges to another characteristic function, we have corresponding weak convergence.

Proof. Let $X_1, ..., X_n$ be iid random variables with $\mathbb{E}[X_1] = 0$ and $\mathrm{Var}(X_1) = 1$. Let $S_n := X_1 + ... + X_n$. We'll take the Taylor expansion of φ_{X_1} centered around $t = 0$ up to order m assuming it's m-times differentiable (X has finite absolute mth moment). We have

$$\varphi_{X_1}(t) = \sum_{k=0}^{m}\frac{\varphi_{X_1}^{(k)}(0)}{k!}t^k + \mathcal{O}(t^m) = \sum_{k=0}^{m}\frac{\mathbb{E}[X_1^k]}{k!}(it)^k + \mathcal{O}(t^m).$$

Now fix $t \in \mathbb{R}$. Then

$$\begin{aligned}
\varphi_{\frac{S_n}{\sqrt{n}}}(t) &= \varphi_{S_n}(\frac{t}{\sqrt{n}}) \\
&= \prod_{i=1}^{n} \varphi_{X_i}(\frac{t}{\sqrt{n}}) \\
&= (\varphi_{X_1}(\frac{t}{\sqrt{n}}))^n \\
&= \left(1 + \frac{1}{\sqrt{n}}\mathbb{E}[X_1]\; it - \frac{\mathbb{E}[X_1^2]}{2n}\; t^2 + \mathcal{O}(\frac{t^2}{n})\right)^n \\
&= \left(1 - \frac{t^2}{2n} + \mathcal{O}(\frac{t^2}{n})\right)^n && (X_1 \text{ has mean zero and unit variance}).
\end{aligned}$$

As $n \longrightarrow \infty$, we have

$$\lim_{n\to\infty} \left(1 - \frac{t^2}{2n} + \mathcal{O}(\frac{t^2}{n})\right)^n = e^{-t^2/2}$$

which is the characteristic function for a standard normal random variable. □