We are almost ready to prove the central limit theorem for i.i.d. sequences, but first we need a few more elementary facts.

Lemma 13.1. *Let $z_1, ..., z_n$ and $w_1, ..., w_n$ be complex numbers, each having modulus at most θ. Then*

$$\left|\prod_{k=1}^{n} z_k - \prod_{k=1}^{n} w_k\right| \leq \theta^{n-1} \sum_{k=1}^{n} |z_k - w_k|.$$

Proof. The inequality holds trivially for $n = 1$.

Now assume that it is true for $1 \leq m < n$. Then

$$\begin{aligned}\left|\prod_{k=1}^{n} z_k - \prod_{k=1}^{n} w_k\right| &\leq \left|z_1 \prod_{k=2}^{n} z_k - z_1 \prod_{k=2}^{n} w_k\right| + \left|z_1 \prod_{k=2}^{n} w_k - w_1 \prod_{k=2}^{n} w_k\right| \\ &\leq \theta \left|\prod_{k=2}^{n} z_k - \prod_{k=2}^{n} w_k\right| + |z_1 - w_1| \theta^{n-1} \\ &\leq \theta \cdot \theta^{n-2} \sum_{k=2}^{n} |z_k - w_k| + \theta^{n-1} |z_1 - w_1| = \theta^{n-1} \sum_{k=1}^{n} |z_k - w_k|\end{aligned}$$

and the result follows by the principle of induction. □

Lemma 13.2. *If $z \in \mathbb{C}$ has $|z| \leq 1$, then $|e^z - (1 + z)| \leq |z|^2$.*

Proof. Expanding the analytic function e^z in a power series about 0 gives

$$\begin{aligned}|e^z - (1 + z)| &= \left|\sum_{k=2}^{\infty} \frac{z^k}{k!}\right| \leq \sum_{k=2}^{\infty} \frac{|z|^k}{k!} \\ &= |z|^2 \sum_{k=2}^{\infty} \frac{|z|^{k-2}}{k!} \leq |z|^2 \sum_{k=1}^{\infty} \frac{1}{2^k} = |z|^2.\end{aligned}$$

□

Theorem 13.1. *If $\{c_n\}_{n=1}^{\infty}$ is a sequence of complex numbers which converges to c, then*

$$\lim_{n\to\infty} \left(1 + \frac{c_n}{n}\right)^n = e^c.$$

Proof. Choose n large enough that $|c_n| < 2|c|$ and $\frac{|c_n|}{n} \leq 1$. Then $\left|1 + \frac{c_n}{n}\right| \leq 1 + \frac{|c_n|}{n} \leq e^{\frac{|c_n|}{n}} \leq e^{\frac{2|c|}{n}}$, so taking $z_m = 1 + \frac{c_n}{n}$, $w_m = e^{\frac{c_n}{n}}$, $\theta = e^{\frac{2|c|}{n}}$ in the statement of Lemma 13.1 and then appealing to Lemma 13.2 gives

$$\begin{aligned}\left|\left(1 + \frac{c_n}{n}\right)^n - e^{c_n}\right| &\leq \left(e^{\frac{2|c|}{n}}\right)^{n-1} n \left|e^{\frac{c_n}{n}} - \left(1 + \frac{c_n}{n}\right)\right| \leq e^{2|c|\frac{n-1}{n}} n \left(\frac{c_n}{n}\right)^2 \\ &\leq e^{2|c|} n \frac{4|c|^2}{n^2} = \frac{4|c|^2 e^{2|c|}}{n} \to 0,\end{aligned}$$

hence

$$\left|\left(1 + \frac{c_n}{n}\right)^n - e^c\right| \leq \left|\left(1 + \frac{c_n}{n}\right)^n - e^{c_n}\right| + |e^{c_n} - e^c| \to 0.$$

□

After all of the work of the past two sections, we are finally able to prove the classical central limit theorem!

Theorem 13.2 (Central Limit Theorem)**.** *If $X_1, X_2, \ldots$ are i.i.d. with $E[X_1] = \mu$ and $\mathrm{Var}(X_1) = \sigma^2 \in (0, \infty)$, then*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow Z \sim N(0,1).$$

Proof. By considering $X_k' = X_k - \mu$ if necessary, it suffices to prove the result for $\mu = 0$.

We have seen that the standard normal has characteristic function $\varphi_Z(t) = e^{-\frac{t^2}{2}}$, which is continuous at $t = 0$, so Theorem 12.3 shows that we only need to demonstrate that the characteristic functions of $\frac{S_n}{\sigma\sqrt{n}}$ converge pointwise to $\varphi_Z(t)$.

Since X_1 has ch.f. $\varphi(t) = 1 - \frac{1}{2}\sigma^2 t^2 + o(t^2)$ by Corollary 12.1, it follows from Theorem 13.1 and the basic properties of characteristic functions that $\frac{S_n}{\sigma\sqrt{n}}$ has ch.f.

$$\begin{aligned}
\varphi_n(t) &= E\left[\exp\left(i\frac{t}{\sigma\sqrt{n}}\sum_{k=1}^{n} X_k\right)\right] = E\left[\prod_{k=1}^{n}\exp\left(i\frac{t}{\sigma\sqrt{n}}X_k\right)\right] = \prod_{k=1}^{n} E\left[\exp\left(i\frac{t}{\sigma\sqrt{n}}X_k\right)\right] \\
&= E\left[\exp\left(i\frac{t}{\sigma\sqrt{n}}X_1\right)\right]^n = \varphi\left(\frac{t}{\sigma\sqrt{n}}\right)^n = \left(1 - \frac{1}{2}\sigma^2\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)\right)^n \\
&= \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{-\frac{t^2}{2}}.
\end{aligned}$$

□

Multiplying the standardized sum by $\frac{\left(\frac{1}{n}\right)}{\left(\frac{1}{n}\right)}$ gives $\dfrac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} \Rightarrow Z$ where $\overline{X_n} = \frac{1}{n}S_n$ is the sample mean and $\frac{\sigma}{\sqrt{n}}$ is the standard error. Thus the CLT can be interpreted as a statement about how sample averages fluctuate about the population mean.

The following poetic description of the CLT is due to Francis Galton:

> I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error." The Law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and complete self-effacement amidst the wildest confusion. The huger the mob and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of unreason.

The most common application of the central limit theorem is to provide justification for approximating a sum of i.i.d. random variables (possibly with unknown distributions) with a normal random variable (for which probabilities can be read off from a table).

However, one should keep in mind that, *a priori*, the identification is only valid in the $n \to \infty$ limit. It is truly amazing that it holds at all, and often it is remarkably accurate even for small values of n, but one needs empirical evidence or more advanced theory to justify the approximation for finite sample averages.

The issue of convergence rates is addressed in a subsequent section.

Our next order of business is to adapt the argument from Theorem 13.2 to prove one of the most well-known generalizations of the central limit theorem, which applies to triangular arrays of independent (but not necessarily identically distributed) random variables.

We will use the notation $E[Y;A] := E[Y 1_A(Y)]$ for the expectation of the random variable Y restricted to the event A.

Theorem 13.3 (Lindeberg-Feller)**.** *For each $n \in \mathbb{N}$, let $X_{n,1}, ..., X_{n,n}$ be independent random variables with $E[X_{n,m}] = 0$. If*

(1) $\lim_{n\to\infty} \sum_{m=1}^{n} E\left[X_{n,m}^2\right] = \sigma^2 \in (0, \infty)$,

(2) $\lim_{n\to\infty} \sum_{m=1}^{n} E\left[X_{n,m}^2; |X_{n,m}| > \varepsilon\right] = 0$ *for all* $\varepsilon > 0$,

then $S_n := \sum_{m=1}^{n} X_{n,m} \Rightarrow \sigma Z$ *where Z has the standard normal distribution.*

Proof.

Let $\varphi_{n,m}(t) = E\left[e^{itX_{n,m}}\right]$, $\sigma_{n,m}^2 = E\left[X_{n,m}^2\right]$. By Theorem 12.3, it suffices to show that

$$\prod_{m=1}^{n} \varphi_{n,m}(t) \to e^{-\frac{t^2\sigma^2}{2}}.$$

From Corollary 12.2, we have

$$\begin{aligned}\left|\varphi_{n,m}(t) - \left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right)\right| &= \left|\varphi_{n,m}(t) - \sum_{k=0}^{2} \frac{(it)^k}{k!} E[X_{n,m}^k]\right| \le E\left[\min\left(\frac{|tX_{n,m}|^3}{3!}, \frac{2\,|tX_{n,m}|^2}{2!}\right)\right] \\ &\le |t|^3 E\left[|X_{n,m}|^3; |X_{n,m}| \le \varepsilon\right] + t^2 E\left[X_{n,m}^2; |X_{n,m}| > \varepsilon\right] \\ &\le \varepsilon |t|^3 E\left[|X_{n,m}|^2\right] + t^2 E\left[X_{n,m}^2; |X_{n,m}| > \varepsilon\right].\end{aligned}$$

Summing over $m \in [n]$, taking limits, and appealing to assumptions 1 and 2 gives

$$\begin{aligned}\limsup_{n\to\infty} \sum_{m=1}^{n} \left|\varphi_{n,m}(t) - \left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right)\right| &\le \varepsilon |t|^3 \limsup_{n\to\infty} \sum_{m=1}^{n} E\left[|X_{n,m}|^2\right] \\ &\quad + t^2 \limsup_{n\to\infty} \sum_{m=1}^{n} E\left[X_{n,m}^2; |X_{n,m}| > \varepsilon\right] = \varepsilon |t|^3 \sigma^2,\end{aligned}$$

hence $\lim_{n\to\infty} \sum_{m=1}^{n} \left|\varphi_{n,m}(t) - \left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right)\right| = 0$ as ε can be taken arbitrarily small.

We now observe that $\sigma_{n,m}^2 \le \varepsilon^2 + E\left[|X_{n,m}|^2; |X_{n,m}| > \varepsilon\right]$ for all $\varepsilon > 0$ and the latter term goes to 0 as $n \to \infty$ by the second assumption.

Accordingly, for any fixed t, we can find n large enough that $1 \ge 1 - \frac{t^2\sigma_{n,m}^2}{2} \ge -1$. Since $|\varphi_{n,m}(t)| \le 1$ as well, $z_m = \varphi_{n,m}(t)$ and $w_m = 1 - \frac{t^2\sigma_{n,m}^2}{2}$ satisfy the assumptions of Lemma 13.1 with $\theta = 1$ for large n and thus

$$\limsup_{n\to\infty} \left|\prod_{m=1}^{n} \varphi_{n,m}(t) - \prod_{m=1}^{n}\left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right)\right| \le \limsup_{n\to\infty} \sum_{m=1}^{n} \left|\varphi_{n,m}(t) - \left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right)\right| = 0.$$

Finally, since $\lim_{n\to\infty} \sum_{j=1}^{n} \frac{t^2\sigma_{n,j}^2}{2} = -\frac{t^2\sigma^2}{2}$ it follows from Fact 11.2 that $\prod_{m=1}^{n}\left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right) \to e^{-\frac{t^2\sigma^2}{2}}$ as $n \to \infty$ and the proof is complete. □

Roughly, Theorem 13.3 says that a sum of a large number of small independent effects has an approximately normal distribution.

Example 13.1. Let π_n be a permutation chosen from the uniform distribution on S_n and let $K_n = K(\pi_n)$ be the number of cycles in π_n. For example, if π is the permutation of $\{1, 2, \ldots, 6\}$ written in one-line notation as 532146, then π can be expressed in cycle notation as (154)(23)(6), so $K(\pi) = 3$.

* Observe that there are $\dfrac{n!}{\prod_{k=1}^{n} k^{\lambda_k}\lambda_k!}$ ways to write a permutation having λ_k k-cycles, $k = 1, ..., n$.

Indeed, once we have fixed the placement of parentheses dictated by the cycle type - say beginning with λ_1 pairs of parentheses having room for 1 symbol, followed by λ_2 pairs of parentheses having room for 2 symbols, and so forth - there are $n!$ ways to distribute the n symbols amongst the parentheses.
But this overcounts since we can permute each of the λ_k k-cycles amongst themselves and we can write each k-cycle in k different ways.
For this reason, it is sometimes helpful to use the canonical cycle notation wherein the largest element appears first within a cycle and cycles are sorted in increasing order of their first element.
For example, we would write $\pi = (32)(541)(6)$.

** Note that the map which drops the parentheses in the canonical cycle notation of σ to obtain σ' in one-line notation (so that $\pi' = 325416$, for example) gives a bijection between permutations with k cycles and permutations with k record values.
(A record value of $\sigma \in S_n$ is a number $j \in [n]$ such that $\sigma(j) > \sigma(i)$ for all $i < j$. Here we are thinking of $\sigma(j)$ as the ultimate ranking of the jth competitor.)

Now the number of permutations of $[n]$ having k cycles is the unsigned Stirling number of the first kind, denoted $c(n, k)$.
These numbers can be computed using the recurrence $c(n + 1, k) = nc(n, k) + c(n, k - 1)$.
This is because every permutation of $[n + 1]$ having k cycles either has $n + 1$ as a fixed point (that is, in a cycle of size 1) or not. The number of the former is just $c(n, k - 1)$ and the number of the latter is $nc(n, k)$ as $n + 1$ can follow any of the first n symbols divided into k cyclically ordered groups.
Thus, in principle, one can explicitly compute $P(K_n = k) = \dfrac{c(n, k)}{n!}$, but this is computationally prohibitive for large n.

We will show that when suitably standardized, K_n is asymptotically normal. To do so, we will construct random permutations using the Chinese Restaurant Process:
In a restaurant with many large circular tables, Person 1 enters and sits at a table. Then Person 2 enters and either sits to the right of Person 1 or at a new table with equal probability. In general, when person k enters, they are equally likely to sit to the right of any of the $k - 1$ seated customers or to sit at an empty table. We associate the seating arrangement after n people have entered with the permutation whose cycles are the tables with occupants read off clockwise.
That this generates a permutation from the uniform distribution follows by induction: It is certainly true when $n = 1$, and if we have a seating arrangement corresponding to a uniform permutation of $[n - 1]$ before person n sits down, then the rules of the process ensure that we have a uniform permutation of n afterward by the same line of reasoning used to establish the recursion for $c(n, k)$.

If we let $X_{n,k}$ be the indicator that Person k sits at an unoccupied table, then $K_n = \sum_{k=1}^n X_{n,k}$. Since the $X_{n,k}$'s are clearly independent, we have

$$E[K_n] = \sum_{k=1}^n E[X_{n,k}] = \sum_{k=1}^n P(k \text{ sits at a new table}) = \sum_{k=1}^n \frac{1}{k} \approx \log(n)$$

and

$$\text{Var}(K_n) = \sum_{k=1}^n \text{Var}(X_{n,k}) = \sum_{k=1}^n \left(\frac{1}{k} - \frac{1}{k^2}\right) \approx \log(n).$$

More precisely, if we set $Y_{n,k} = \frac{X_{n,k} - \frac{1}{k}}{\sqrt{\log(n)}}$, then $E\left[Y_{n,k}\right] = 0$ and

$$\sum_{k=1}^n E\left[Y_{n,k}^2\right] = \frac{1}{\log(n)} \text{Var}\left(K_n\right) = \frac{1}{\log(n)} \sum_{k=1}^n \left(\frac{1}{k} - \frac{1}{k^2}\right) \to 1.$$

Also,

$$\sum_{k=1}^n E\left[Y_{n,k}^2; |Y_{n,k}| > \varepsilon\right] \to 0$$

since the sum is 0 once $\log(n)^{-\frac{1}{2}} < \varepsilon$.

Therefore, Theorem 13.3 implies that $\sum_{k=1}^n Y_{n,k} \Rightarrow Z \sim N(0,1)$.

Because $\sum_{k=2}^n \frac{1}{k} \leq \int_1^n \frac{dx}{x} \leq \sum_{k=1}^{n-1} \frac{1}{k}$, the conclusion can be written as $\frac{K_n - \log(n)}{\sqrt{\log(n)}} \Rightarrow Z$.

In terms of sequences of independent random variables, Theorem 13.3 specializes to

Corollary 13.1. *Suppose that $X_1, X_2, \ldots$ are independent, random variables with $E[X_k] = 0$ and $\text{Var}(X_k) = \sigma_k^2 \in (0, \infty)$ for all k. Let $S_n = \sum_{k=1}^n X_k$ and $s_n^2 = \sum_{k=1}^n \sigma_k^2$. If the sequence satisfies Lindeberg's condition:*

$$\lim_{n \to \infty} \frac{1}{s_n^2} \sum_{k=1}^n E[X_k^2; |X_k| > \varepsilon s_n] = 0$$

for every $\varepsilon > 0$, then

$$\frac{S_n}{s_n} \Rightarrow Z \sim N(0,1).$$

Proof. Take $X_{n,m} = \frac{X_m}{s_n}$ in Theorem 13.3. □

Of course the mean zero condition is just a matter of convenience since finite variance implies finite mean and we can always consider $X_k' = X_k - E[X_k]$.

Note that the classical central limit theorem is an immediate consequence of Corollary 13.1 since $X_1, X_2, \ldots$ i.i.d. with mean zero and finite variance σ^2 gives $s_n = \sigma\sqrt{n}$ and

$$\begin{aligned} \lim_{n \to \infty} \frac{1}{s_n^2} \sum_{k=1}^n E[X_k^2; |X_k| > \varepsilon s_n] &= \sigma^2 \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^n E[X_k^2; |X_k| > \varepsilon\sigma\sqrt{n}] \\ &= \sigma^2 \lim_{n \to \infty} E[X_1^2; |X_1| > \varepsilon\sigma\sqrt{n}] = 0 \end{aligned}$$

by the DCT and finite variance.

We conclude with an example showing that one can have normal convergence even when the Lindeberg condition is not satisfied.

Example 13.2. Let $X_1, X_2, ...$ be independent with $X_1 \sim N(0,1)$ and $X_k \sim N(0, 2^{k-2})$ for $k \geq 2$. Setting $S_n = \sum_{k=1}^n X_k$, we have

$$s_n^2 = \sum_{k=1}^n \mathrm{Var}(X_k) = 1 + \sum_{k=2}^n 2^{k-2} = 1 + \sum_{k=0}^{n-2} 2^k = 1 + \frac{2^{n-1}-1}{2-1} = 2^{n-1}.$$

For any $\varepsilon \in \left(0, \frac{1}{\sqrt{2}}\right)$, $n \geq 2$,

$$E\left[X_n^2; |X_n| > \varepsilon s_n\right] \geq E[X_n^2] - E\left[X_n^2; |X_n| \leq \varepsilon s_n\right] \geq E[X_n^2] - \varepsilon^2 s_n^2 P\left(|X_n| > \varepsilon s_n\right) \geq 2^{n-2} - \varepsilon^2 2^{n-1},$$

thus

$$\frac{1}{s_n^2}\sum_{k=1}^n E[X_k^2; |X_k| > \varepsilon s_n] \geq \frac{E\left[X_n^2; |X_n| > \varepsilon s_n\right]}{s_n^2} \geq \frac{2^{n-2} - \varepsilon^2 2^{n-1}}{2^{n-1}} = \frac{1}{2} - \varepsilon^2 > 0$$

for all $n \geq 2$.

However, we observe that if W_1 and W_2 are independent with $W_k \sim N(\mu_k, \sigma_k^2)$, then W_k has ch.f. $\varphi_k(t) = e^{i\mu_k t - \frac{\sigma_k^2 t^2}{2}}$, hence $W_1 + W_2$ has ch.f. $\varphi_{W_1+W_2}(t) = \varphi_1(t)\varphi_2(t) = e^{i(\mu_1+\mu_2)t - \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}$, so $W_1 + W_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

By induction, we have "Sums of independent normals are normal and the means and variances add."

Applied to the case at hand, we see that $S_n = \sum_{k=1}^n X_k \sim N(0, s_n^2)$, hence $\frac{S_n}{s_n} \sim N(0,1)$ for all n and thus in the $n \to \infty$ limit.