1 Central Limit Theorem Proof

The central limit theorem (CLT) posits that means are approximately normally distributed for large sample sizes. This is crucial for many statistical applications in many fields of research that deal with randomness. Any measure that we are interested in that has inherent noise (blood measures, neurophysiological activity, social behavior) can be analyzed by knowing how a mean of a sample of such measures is distributed.

I couldn't find a self-contained proof of the central limit theorem. So I made one myself. Maybe it helps someone else at some point. I took parts from various sources, mostly from

- `https://sas.uwaterloo.ca/ dlmcleis/s901/chapt6.pdf` and
- `https://www.youtube.com/c/papaflammy/`.

So how does the CLT look like? It requires multiple independently and identically distributed (iid.) measures X_i with $i \in 1..n$. The iid. assumption relates to a situation in which the observations X_i come from the same randomness mechanism (e.g., many participants receiving the same medication, it is not the case that one gets a higher dosage than another unless that is the random process we are interested in) and are independent draws from this mechanism. Then, if we were to repeat this experiment — draw n observations and compute the mean — over and over again, the computed means will be approximately normally distributed.

Theorem 1 (Central Limit Theorem)**.** *Let X_1, X_2, ..., X_n iid. as X. Assume that all moments are finite and, in particular, assume without loss of generality that $E[X] = 0$ and $Var[X] = 1$. Let $Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$. Then*

$$\lim_{n\to\infty} Z_n \sim \mathcal{N}(0,1).$$

We will first look at a short overview of the proof. In this, we will use technicalities about characteristic functions, which I explain later. If you are a mathematician, this structure may appear usual to you because most of the time, people write all the required lemmas before the important theorem. But here, I focus on the overview first and only then add the details.

To start with, note that we multiply the sum of X_is by $\frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}}$ to make Z_n have variance 1.

Next, we need so-called characteristic functions as the main work-horse of the proof. The characteristic function of a random variable X is

$$\psi_X(t) = E[e^{itX}],$$

which describes the moments of X. Remember that the series definition of e is

$$e^c = \frac{c^0}{0!} + \frac{c^1}{1!} + \frac{c^2}{2!} + \frac{c^3}{3!} + ...$$

With this, we can see that the characteristic function tells us the m-th moment of the random variable X by taking the m-th derivative of $\psi_X(t/i)$ at $t = 0$. Remember that the 0th moment is constant 1, the 1st moment is the expected value of the random variable, the 2nd moment is related to the variance, the 3rd to the skewness etc.

$$\begin{aligned}
\frac{d^m}{dt^m}\psi_X(-t/i)\bigg|_{t=0} &= \frac{d^m}{dt^m}E\left[e^{tX}\right]\bigg|_{t=0} \\
&= \frac{d^m}{dt^m}E\left[\frac{(tX)^0}{0!} + \frac{(tX)^1}{1!} + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \;...\right]\Bigg|_{t=0} \\
&= \frac{d^m}{dt^m}\left[\frac{t^0E[X^0]}{0!} + \frac{t^1E[X^1]}{1!} + \frac{t^2E[X^2]}{2!} + \frac{t^3E[X^3]}{3!} + \;...\right]\Bigg|_{t=0} \\
&= E[X^m]
\end{aligned}$$

This is so because all terms with a lower exponent than m vanish due to the derivatives and all terms with a higher exponent are set to zero by $t = 0$.

The core idea of the proof is showing that the moments vanish when taking the mean over multiple independent draws of the same random variable. In fact, all moments but the 0th and the 1st (expected value) vanish but the higher the moment the faster it vanishes. Thus, the shape of the distribution of that random variable (determined by its moments) gets washed out and approaches that of a normal one.

Proof. Z_n is a sum of scaled random variables $\frac{X_i}{\sqrt{n}}$. The sum of multiple random variables has a characteristic function that is the product of individual characteristic functions (see Lemma 2). Thus, z_n has the following characteristic function.

$$
\begin{aligned}
\psi_{Z_n}(t) &= \prod_{i=1}^{n} \psi_{\frac{X_i}{\sqrt{n}}}(t) \\
&= [E[e^{itX/\sqrt{n}}]]^n \\
&= \left[E\left[\frac{(itX/\sqrt{n})^0}{0!} + \frac{(itX/\sqrt{n})^1}{1!} + \frac{(itX/\sqrt{n})^2}{2!} + \frac{(itX/\sqrt{n})^3}{3!} + \dots\right]\right]^n \\
&= \left[1 + \frac{(itE[X])}{\sqrt{n}} - \left(\frac{t^2E[X^2]}{2n}\right) + O\left(\frac{1}{n^{3/2}}\right) + \dots\right]^n \\
&= \left[1 - \left(\frac{t^2}{2n}\right) + O\left(\frac{1}{n^{3/2}}\right)\right]^n
\end{aligned}
$$

We consider the limit of this characteristic function as $n \to \infty$.

$$
\begin{aligned}
\lim_{n\to\infty} \psi_{Z_n}(t) &= \lim_{n\to\infty} \left[1 - \left(\frac{t^2}{2n}\right) + O\left(\frac{1}{n^{3/2}}\right)\right]^n \\
&= \lim_{n\to\infty} \left[\left(1 + \frac{(-t^2/2)}{n}\right)^n + \sum_{k=1}^{n} \binom{n}{k} \left(1 + \frac{1}{n}\right)^{n-k} O\left(\frac{1}{n^{3/2}}\right)^k\right] \\
&= e^{-\frac{t^2}{2}} + \lim_{n\to\infty} \left[\sum_{k=1}^{n} \binom{n}{k} \left(1 + \frac{1}{n}\right)^{n-k} O\left(\frac{1}{n^{3/2}}\right)^k\right]
\end{aligned}
$$

The first part produces $e^{-t^2/2}$ and the second part vanishes as n approaches infinity because

$$
\begin{aligned}
\left|\lim_{n\to\infty} \left[\sum_{k=1}^{n} \binom{n}{k} \left(1 + \frac{1}{n}\right)^{n-k} O\left(\frac{1}{n^{3/2}}\right)^k\right]\right| &\le \left|\lim_{n\to\infty} \left[\sum_{k=1}^{n} n^k \cdot e \cdot O\left(\frac{1}{n^{k3/2}}\right)\right]\right| \\
&= \left|\lim_{n\to\infty} \left[O\left(\frac{1}{n^{1/2}}\right)\right]\right| \\
&= 0
\end{aligned}
$$

As a result, what is left over is the characteristic function

$$
\lim_{n\to\infty} \psi_{Z_n}(t) = e^{-t^2/2}
$$

which turns out to be that of the standard normal distribution (Lemma 3). Since characteristic functions uniquely determine the probability distribution (Lemma 4), Z_n approaches no other but the normal distribution as $n \to \infty$. $\square$

What is left to show is that

- the product of characteristic functions is the characteristic function of the sum of the corresponding random variables,
- $e^{-t^2/2}$ is the characteristic function of the standard normal distribution, and
- that the characteristic function uniquely determines the distribution of a random variable.

1.1 Characteristic Functions

Lemma 2 (Sum of random variables, product of characteristic functions)**.** *The sum of two random variables* $Z = X + Y$ *has the characteristic function of the product of the two variable's characteristic functions,* $\psi_Z(t) = \psi_X(t)\psi_Y(t)$*.*

Proof.

$$\begin{aligned}
\psi_Z(t) &= E\left[e^{itZ}\right] \\
&= E\left[e^{it(X+Y)}\right] \\
&= E\left[e^{itX}e^{itY}\right] \\
&= E\left[e^{itX}\right]E\left[e^{itY}\right] \\
&= \psi_X(t)\psi_Y(t)
\end{aligned}$$

□

Lemma 3 (Characteristic functions of the standard normal distribution)**.** *A random variable X with the standard normal probability distribution $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ has the characteristic function $\psi_X(t) = e^{-t/2}$.*

Proof.

$$\begin{aligned}
\psi_X(t) &= E\left[e^{itX}\right] \\
&= \int_{-\infty}^{\infty} e^{itx} f(x) dx \\
&= \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2 - 2itx)} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2 - 2itx + (it)^2 - (it)^2)} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2} e^{-\frac{t^2}{2}} dx \\
&= e^{-\frac{t^2}{2}} \int_{-\infty}^{\infty} f(x - it) dx \\
&= e^{-\frac{t^2}{2}} \int_{-\infty}^{\infty} f(y) dy \\
&= e^{-\frac{t^2}{2}}
\end{aligned}$$

The integral at the end is equal to 1 because we can substitute $y = x - it$, $\frac{dy}{dx} = 1 \iff dx = dy$.

□

Lemma 4 (Probability distributions are uniquely determined by their characteristic function)**.** *Let $\psi_X(t)$ be the characteristic function of X, then the probability mass of X in the interval $[a,b]$ is recovered by*

$$\lim_{T\to\infty} \frac{1}{2\pi} \int\limits_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \psi_X(t) dt = P(a < X < b) + \frac{P(X=a) + P(X=b)}{2}.$$

This lemma is not as straight forward to proof as the others. It makes intuitive sense that all moments (from 0 to infinity) together fully determine a random variable. On a high level, we show this by showing how to recover the exact distribution $P(X)$ from the characteristic function $\psi_X(t)$. This requires a relatively complicated integral. We will again show the overview of the proof first and assume same technical lemmas in doing so. After, we proof the details in further lemmas.

Proof. Using $e^{itc} = cos(tc) + i \cdot sin(tc)$, the two exponential terms in the integral can be rewritten as

$$\begin{aligned}
\int\limits_{-T}^{T} \frac{e^{itc}}{2it} dt &= \int_0^T \frac{\cos(tc) + i \cdot \sin(tc)}{2it} dt + \int_{-T}^0 \frac{\cos(tc) - i \cdot \sin(tc)}{2it} dt \\
&= \int_0^T \frac{\cos(tc) + i \cdot \sin(tc)}{2it} dt + \int_0^T \frac{\cos(tc) + i \cdot (-1)\sin(tc)}{2i(-t)} dt \\
&= \int_0^T \frac{\cos(tc)}{2it} - \frac{\cos(tc)}{2it} + \frac{\sin(tc)}{2t} + \frac{\sin(tc)}{2t} dt \\
&= \int_0^T \frac{\sin(tc)}{t} dt.
\end{aligned}$$

Coming back to the equation we want to prove here, we replace the characteristic function by its definition and obtain:

$$\lim_{T\to\infty}\frac{1}{2\pi}\int\limits_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\psi_X(t)dt = \lim_{T\to\infty}\frac{1}{2\pi}\int\limits_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\int\limits_{-\infty}^{\infty}e^{itx}f(x)dxdt$$
$$= \int\limits_{-\infty}^{\infty}\lim_{T\to\infty}\frac{1}{2\pi}\int\limits_{-T}^{T}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}f(x)dtdx$$
$$= \int\limits_{-\infty}^{\infty}\frac{1}{\pi}\lim_{T\to\infty}\int\limits_{0}^{T}\frac{\sin(t(x-a))}{t}-\frac{\sin(t(x-b))}{t}dtf(x)dx$$
$$= \int\limits_{(-\infty,a)} 0\cdot f(x)dx + \int\limits_{[a,a]}\frac{1}{2}\cdot f(x)dx + \int\limits_{(a,b)} 1\cdot f(x)dx + \int\limits_{[b,b]}\frac{1}{2}\cdot f(x)dx + \int\limits_{(b,\infty)} 0\cdot f(x)dx$$
$$= P(a < X < b) + \frac{1}{2}\left(P(X=a)+P(X=b)\right)$$

Where we the solution to the inner integral from Lemma 5.

With that, we have shown that the clever computation retrieves the probability mass for any interval. This is done by ensuring that all probabilities before a and after b are set to 0, those within the interval are preserved as $P(a < X < b)$ and the probabilities of the edges are halved (for compatibility). Therefore, characteristic functions can be translated into probability functions and these two functions therefore have a one-to-one relationship. □

1.2 Technical Lemmas

How does the integral in Lemma 4 solve exactly in the way that allows us to recover the probabilities? For this, we need more technical results.

Lemma 5 (Limit of the intregral of $\sin(tc)/t$).

$$\lim_{T\to\infty}\int_0^T \frac{\sin(tc)}{t}dt = \begin{cases} -\frac{\pi}{2}, & \textit{for } c<0 \\ +\frac{\pi}{2}, & \textit{for } c>0 \\ 0, & \textit{for } c=0 \end{cases}$$

Proof. For this proof, we need the Laplace transform of functions $g(t)$.

$$\mathcal{L}\{g\}(s) = \int_s^\infty g(t)e^{-tx}dt$$

We consider the special case of functions $g(t) = h(t)/t$.

$$\begin{aligned}
\mathcal{L}\left\{\frac{h(t)}{t}\right\}(s) &= \int_s^\infty \frac{h(t)}{t}e^{-st}dt \\
&= \int_s^\infty \frac{h(t)}{t}\left(\int_\infty^s -te^{-rt}dr\right)dt \\
&= \int_s^\infty h(t)\left(\int_s^\infty e^{-rt}dr\right)dt \\
&= \int_s^\infty \int_s^\infty h(t)e^{-rt}dtdr \\
&= \int_s^\infty \mathcal{L}\{h(t)\}(r)dr
\end{aligned}$$

We now consider $g(t) = h(t)/t = \sin(t)/t$ where we solve the Laplace transform $\mathcal{L}(\sin(t))$ in Lemma 6.

$$\begin{aligned}
\mathcal{L}\left\{\frac{\sin(t)}{t}\right\}(s) &= \int_s^\infty \mathcal{L}\{\sin(t)\}(r)\,dr \\
&= \int_s^\infty \frac{1}{1+r^2}\,dr \\
&= \tan^{-1}(r)\Big|_{r=s}^{r=\infty} \\
&= \frac{\pi}{2} - \tan^{-1}(s).
\end{aligned}$$

With this, we can solve the initial integral. We get rid fo the constant c in $\sin(ct)$ with a change of variables: We substitute $u = tc$, $t = u/c$ and $du = c \cdot dt$. We also add the vanishing term e^{-st} with the limit of $s \to \infty$ to write it as a Laplace transform.

$$\begin{aligned}
\lim_{T\to\infty} \int_0^T \frac{\sin(tc)}{t} dt &= \lim_{T\to\infty} \int_0^{T/c} \frac{\sin(u)}{u}\frac{c}{c} dt \\
&= \int_0^\infty \frac{\sin(u)}{u} du \\
&= \lim_{s\to 0} \int_s^\infty \frac{\sin(u)}{u} e^{-st} du \\
&= \lim_{s\to 0} \mathcal{L}\left\{\frac{\sin(u)}{u}\right\}(s) \\
&= \lim_{s\to 0} \int_s^\infty \mathcal{L}\left\{\sin(u)\right\}(r) dr \\
&= \int_0^\infty \frac{1}{1+r^2} dr \\
&= \tan^{-1}(r)|_{r=0}^{r=\infty} \\
&= \frac{\pi}{2} - 0 \\
&= \frac{\pi}{2}
\end{aligned}$$

We still need to proof the solution to the Laplace transform (Lemma 6) and the integral (Lemma 7) to arrive at the desired statement.

Since we consider the limit to infinity, this constant simply drops if it is positive which we assume for now ($c > 0$). For negative constants, $c < 0$, the integral goes from $-\infty$ to 0 and the proof can be written analogously with an additional minus in the end. For $c = 0$, the integral is simply 0. $\square$

Lemma 6 (Laplace transform of $\sin(t)$)**.**

$$\mathcal{L}\{\sin(t)\}(s) = \frac{1}{1+s^2}$$

Proof.

$$\begin{aligned}
\mathcal{L}\{e^{at}\}(s) &= \int_0^\infty e^{-st}e^{at}dt \\
&= \int_0^\infty e^{(a-s)t}dt \\
&= \frac{1}{a-s}e^{(a-s)t}\Big|_{t=0}^{t=\infty} \\
&= \frac{1}{a-s}e^{(a-s)\infty} - \frac{1}{a-s}e^{(a-s)\cdot 0} \\
&\overbrace{=}^{s>a} 0 - \frac{1}{a-s} \\
&= \frac{1}{s-a}
\end{aligned}$$

Now we can solve the Laplace transform of $\sin(t)$ using the equality

$$\frac{e^{it} - e^{-it}}{2it} = \frac{\cos(t) + i\sin(t) - \cos(-t) - i\sin(-t)}{2i} = \frac{2i\sin(t)}{2i} = \sin(t).$$

$$
\begin{aligned}
\mathcal{L}\{\sin(t)\}(s) &= \mathcal{L}\left\{\frac{e^{it} - e^{-it}}{2i}\right\}(s) \\
&= \frac{1}{2i}\mathcal{L}\{e^{it} - e^{-it}\} \\
&= \frac{1}{2i}\left(\frac{1}{s-i} - \frac{1}{s-(-i)}\right) \\
&= \frac{1}{2i}\left(\frac{(s+i)}{(s+i)(s-i)} - \frac{s-i}{(s+i)(s-i)}\right) \\
&= \frac{1}{2i}\left(\frac{2i}{s^2 - i^2}\right) \\
&= \frac{1}{s^2+1}
\end{aligned}
$$

□

Lemma 7 (Derivative of the inverse tangens)**.**

$$\int \frac{1}{1+s^2} ds = \tan^{-1}(s)$$

Proof. The derivative of the tangens is

$$\begin{aligned}
\frac{d}{ds}\tan(s) &= \frac{d}{ds}\frac{\sin(s)}{\cos(s)} \\
&= \frac{\sin'(s)\cos(s) - \sin(s)\cos'(s)}{\cos^2(s)} \\
&= \frac{\cos^2(s) + \sin^2(s)}{\cos^2(s)} \\
&= 1 + \tan^2(s),
\end{aligned}$$

and the derivative of the inverse therefore is

$$\begin{aligned}
\frac{d}{ds}\tan^{-1}(s) &= \frac{1}{\tan'(\tan^{-1}(s))} \\
&= \frac{1}{1+\tan^2(\tan^{-1}(s))} \\
&= \frac{1}{1+s^2}.
\end{aligned}$$

□