Central Limit Theorem: Proof

Using Cumulants

The following is adapted from a lecture given by Dr. Yuval Filmus at the University of Toronto.

The Central Limit Theorem

Recall that the Central Limit Theorem (or CLT) states that the sum of independent, normalized random variables converges in distribution to a Gaussian (or "Normal") distribution.

Specifically, if $X \sim \mathcal{N}(\mu, \sigma^2)$ with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}^+$, then the Probability Density Function (PDF) of X is given by:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Now, we consider the so-called "weak" CLT. Suppose that $\{X_i\}_{i=1}^n$ are *iid* random variables with mean zero and fixed variance σ^2 for $n \in \mathbb{N}$. Then...

$$\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} = \frac{\sum_{i=1}^n X_i}{\sqrt{n}} \xrightarrow{\text{dist.}} \mathcal{N}(0, \sigma^2)$$

For random variables A and B, $A \xrightarrow{\text{dist.}} B$ indicates the convergence in distribution of A to B. We say that $A \xrightarrow{\text{dist.}} B$ iff $\mathbb{P}(A \in [c,d]) \to \mathbb{P}(B \in [c,d])$ for every closed interval $[c,d] \subseteq \mathbb{R}$.

Properties of the Normal Distriubution

Recall the scaling and convolution properties of the normal distribution.

Specifically, if X_1 and X_2 are independent normal variables with mean μ_1 and μ_2 (respectively) and variance σ_1^2 and σ_2^2, we can write

$$X_1 \perp X_2 \implies X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Further, if $c \in \mathbb{R}$, we can write that

$$cX \sim \mathcal{N}(c\mu, c^2\sigma^2)$$

From this, it should be obvious why the CLT holds in the 'trivial' case of normally distributed variables. The more interesting case involves general random variables, which we explore via moment generating functions.

Moment Generating Functions

Recall the moment-generating function (MGF) of a random variable X is given by

$$M_X(t) = \mathbb{E}[e^{tX}] \text{ for } t \in \mathbb{R}$$

Recalling the Taylor series expansion of e^{tX}, we see that the MGF lives up to its name of uncovering the moments of random variables:

$$M_X(t) = 1 + t\mathbb{E}(X) + \frac{t^2}{2!}\mathbb{E}(X^2) + \frac{t^3}{3!}\mathbb{E}(X^3) + \cdots = \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbb{E}(X^n)$$

Now, if we centralize these moments on $t = 0$, we extract key components of the mean, variance, skew, kurtosis, etc. Therefore, the n-th derivative of the MGF at $t = 0$ allows us to extract the *raw* moments; specifically, $M_X^{(n)}(0) = \mathbb{E}(X^n)$.

We assume that the MGF exists (i.e. the integral $\mathbb{E}(e^{tX})$ converges for some nonzero t.) So, we assume that X has moments of all orders n and thus $M_X(t)$ is well-defined.

Properties of Moment Generating Functions

Before we continue, we cite some key properties of MGFs. Firstly, for $X \perp Y$ we have:

$$M_{X+Y}(t) = \mathbb{E}(e^{t(X+Y)}) = \mathbb{E}(e^{tX}e^{tY}) = M_X(t)M_Y(t)$$

Further, for a constant $c \in \mathbb{R}$ and random variable cX, we have

$$M_{cX}(t) = \mathbb{E}[e^{t(cX)}] = M_X(ct)$$

Cumulants

We are nearly at the point where we can present the proof. However, we must first define the *cumulant* generating function (CGF) of a random variable X, which is just the log-MGF. We assume that both the CGF and MGF exist.

$$K_X(t) = \log M_X(t) = \log\Big(\sum_{n=0}^{\infty}\frac{t^n}{n!}\mathbb{E}[X^n]\Big) = \log\Big(1 + \underbrace{\sum_{n=1}^{\infty}\frac{t^n}{n!}\mathbb{E}[X^n]}_{m(t)}\Big)$$

While somewhat unintuitive, the cumulant comes from the observation above that the MGF (from it's Taylor series expansion) is $M_X(t) = 1 + \mathcal{O}(t)$, so taking the log is helpful. We retain the summation component and refer to it as $m(t)$ for the expansion to follow

Now, the Taylor series expansion of $\log(1+x)$ is given by:

$$\log(1+x) = 0 + x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{x^n}{n}$$

So, we can expand the cumulants $K_X(t)$ as a power series of $M_X(t)$, neat! Recalling $\log(1) = 0$ and allowing $x = m(t) = \mathbb{E}[X]t + \frac{t^2}{2}\mathbb{E}[X^2] + ...$, we see that...

$$K_X(t) = \Big(\mathbb{E}[X]t + \frac{t^2}{2}\mathbb{E}[X^2] + ... \Big) - \frac{(\mathbb{E}[X]t + ...)^2}{2} + ...$$

Combining like terms, we notice that

$$K_X(t) = \mathbb{E}[X]t + \frac{t^2}{2}(\mathbb{E}[X^2] - \mathbb{E}[X]^2) + ...$$

Crucially, the first two coefficients of $K_X(t)$, disregarding the $1/n!$ factors, are the expectation and variance! These coefficients are the *cumulants*. Formally, we can write the nth cumulant similarly to how we wrote the nth moment, that is:

$$K_n[X] = K_X^{(n)}(0)$$

From the above, it is direct to show that $K_1[X]$ is the expectation $\mathbb{E}[X]$ and the second cumulant $K_2[X]$ is the variance $\text{var}(X)$. One could follow this procedure to derive the mean and variance of common distributions rather than a general variable X. For interest's sake, we show the methodology for $K_1[X]$:

$$\begin{aligned}
K_1[X] &= \frac{d}{dt}K_X(t)\Big|_{t=0} \\
&= \frac{d}{dt}\Big(t\mathbb{E}[X] + \frac{t^2}{2}(\mathbb{E}[X^2] - \mathbb{E}[X]^2) + t^3(...)\Big)\Big|_{t=0} \\
&= \mathbb{E}[X] + t(...) + t^2(...) + t^3(...)\Big|_{t=0} \\
&= \mathbb{E}[X]
\end{aligned}$$

We will follow this procedure and generalize it in the next section.

Limit Cumulant Generating Function

We have now done most of the hard work. Notice that the CGF retains the scaling and convolution properties of the MGF. Namely, for $X \perp Y$ and constant c we have

$$K_{K+Y}(t) = K_X(t) + K_Y(t), \text{ and } K_{cX}(t) = K_X(ct)$$

Now, suppose $\{X_i\}_{i=1}^{n}$ are *iid* random variables with zero mean. It follows that the CGF is:

$$K_{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i} = K_{X_1}\left(\frac{t}{\sqrt{n}}\right) + K_{X_2}\left(\frac{t}{\sqrt{n}}\right) + \cdots + K_{X_n}\left(\frac{t}{\sqrt{n}}\right)$$

How do we express this in terms of the cumulants? Consider a single X_i. We take the m-th derivative evaluated at $t = 0$, and all terms will cancel out except for the m-th term (recall the process for $m = 1$ we did earlier!) Consider $K_m[X_i]$ for $m = 2$:

$$
\begin{aligned}
K_2[X_i] &= \frac{d^2}{d^2 t} K_{X_i}\left(\frac{t}{\sqrt{n}}\right)\Big|_{t=0} \\
&= \frac{d^2}{d^2 t}\left(\frac{t}{\sqrt{n}}\mathbb{E}[X] + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) + t^3(\dots)\right)\Big|_{t=0} \\
&= 0 + \frac{1}{\sqrt{n}^2}(\mathbb{E}[X^2] - \mathbb{E}[X]^2) + 0 \\
&= \frac{1}{n}K_2[X]
\end{aligned}
$$

Generalizing this, we have the following. The ellipses on the left hand side get "derivated out," since they have order less than m. Contrarily, the ellipses on the right-hand side have order higher than m, thus their derivatives contain a factor of t and and are eliminated when evaluated at $t = 0$. This is one of the many ways where the "central" aspect of the central limit theorem is so important! (The other, of course, is the centering of the mean of the random variables at zero.)

$$K_m[X_i] = \frac{d^m}{d^m t} K_{X_i}\left(\frac{t}{\sqrt{n}}\right)\Big|_{t=0} = \frac{d^m}{d^m t}\left(\cdots + \frac{1}{m}\left(\frac{t}{\sqrt{n}}\right)^m(\mathbb{E}[X^m] + \dots) + \dots\right) = \frac{K_m[X]}{\sqrt{n}^m}$$

Expressing this as a sum of the cumulants, we have:

$$K_m\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right] = \frac{K_m[X_1] + K_m[X_2] + \cdots + K_m[X_n]}{n^{m/2}}$$

For $m = 1$, we have $K_1[X_i] = \mathbb{E}[X_i] = 0$, by the assumption that the variables have mean zero. Then, since the cumulants are bounded by assumption, there exists some constant $C \in \mathbb{R}$ such that $K_m[X_i] \leq C$ for all $i \in [1, n]$ and for all m. Thus, for $m > 2$ we have

$$K_{m>2}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right] \leq \frac{nC}{n^{m/2}} = \frac{C}{n^{(m-2)/2}} \to 0, \text{ as } n \to \infty$$

Further, for $m = 2$, recall that the X_i are identically distributed, thus $\forall i, j, K_2[X_i] = K_2[X_j]$. We utilize this in the $m = 2$ case:

$$K_2\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\right] = \frac{K_2[X_1] + \cdots + K_2[X_n]}{n^{2/2}} = \frac{nK_2[X]}{n} = K_2[X]$$

So, as $n \to \infty$ <u>only</u> the K_2 component remains, since higher order cumulants decay at at most $\mathcal{O}(n^{-1/2})$ and the first-order cumulant is zero (by construction.) We can use this fact to determine the limit cumulant generating function, which only contains the K_2 component. Recall from the earlier expansion of $K_X(t)$ that this component is

$$K_X(t) = \frac{t^2}{2}(\mathbb{E}[X^2] - \mathbb{E}[X]^2) = \boxed{\frac{1}{2}\sigma^2 t^2} \tag{$*$}$$

So, this is the CGF of the normalized random variables as $n \to \infty$. What remains is to show that it is equivalent to the Normal CGF, and is therefore equal in distribution to a normal random variable.

The Normal CGF

We now derive the normal CGF to demonstrate its equivalency to Equation $(*)$. Recall that the CGF (or MGF) <u>uniquely</u> determines the distribution of a random variable if it exists in a neighborhood of 0. Our random variable in the previous section was centered at zero (it is the "central" limit theorem after all.) We hence show it is equivalent a normal distribution centered at 0 with variance equal to the variance of X.

Let $Y \sim \mathcal{N}(0, \sigma^2)$. Thus the probability density function $f_Y(y)$ is given by

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{y^2}{2\sigma^2}\right)$$

Now, recalling that the MGF is $M_Y(t) = \mathbb{E}[e^{tY}]$, by the Law of the Unconscious Statistician we have

$$\begin{aligned}
M_Y(t) &= \mathbb{E}[e^{tY}] \\
&= \int_{-\infty}^{\infty} e^{ty}f_Y(y)\mathrm{d}y \\
&= \int_{-\infty}^{\infty} e^{ty}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-y^2/2\sigma^2}\mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{ty-y^2/2\sigma^2}\mathrm{d}y
\end{aligned}$$

Now, we isolate the exponent and simplify by completing the square:

$$ty - \frac{y^2}{2\sigma^2} = -\frac{1}{2\sigma^2}(y^2 - 2t\sigma^2 y)$$
$$= -\frac{1}{2\sigma^2}((y - t\sigma^2)^2 - t^2\sigma^4)$$
$$= -\frac{(y - t\sigma^2)^2}{2\sigma^2} + \frac{1}{2}t^2\sigma^2$$

Plugging this back into our expectation, we have that

$$M_Y(t) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{-\frac{(y-t\sigma^2)^2}{2\sigma^2}} e^{\frac{1}{2}t^2\sigma^2}\,\mathrm{d}y$$
$$= e^{\frac{1}{2}t^2\sigma^2}\underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-t\sigma^2)^2}{2\sigma^2}}\,\mathrm{d}y}_{\text{PDF of } \mathcal{N}(t\sigma^2,\sigma^2),\ \text{integrates to 1}}$$
$$= e^{\frac{1}{2}t^2\sigma^2}$$

Then, we recover the CGF directly by taking the natural logarithm of the MGF

$$K_X(t) = \log M_X(t) = \log(e^{\frac{1}{2}t^2\sigma^2}) = \frac{1}{2}t^2\sigma^2$$

Crucially, this is **the same** as equation $(*)$! Since both functions are defined on an open interval containing zero and have the same CGF, this implies their MGFs are the same and thus they are equal in distribution as $n \to \infty$. Thus, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \xrightarrow{\text{dist}} \mathcal{N}(0, \sigma^2)$, as required. This proves the Central Limit Theorem. $\square$