Proving the Central Limit Theorem

Aditya Sengupta

30 September, 2019

Theorem (Central Limit Theorem)**.** *Given a collection of i.i.d. random variables $X_1, \ldots, X_n$ all of which have mean μ and variance σ^2, the distribution of the sample mean $S_n = \frac{\sum_{i=1}^n X_i}{n}$ in the limit $n \to \infty$ is approximately $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.*

We first prove a couple of intermediate results.

Lemma 1. *For i.i.d. random variables $X_1, \ldots, X_n$ each having MGF $M(s) = \mathbb{E}[e^{sx}]$, the MGF of their sum is given by $(M(s))^n$.*

Proof. of Lemma 1. Let $S = \sum_{i=1}^n X_i$.

$$M_S(s) = \mathbb{E}[e^{sS}] = \mathbb{E}[e^{s(X_1+X_2+\cdots+X_n)}] \tag{1}$$

Since the X_is are all independent, we can split them to get

$$M_S(s) = \mathbb{E}[e^{sX_1}] \cdot \mathbb{E}[e^{sX_2}] \cdot \cdots \cdot \mathbb{E}[e^{sX_n}] \tag{2}$$

Since the X_is are all identically distributed, this gives us

$$M_S(s) = \left(\mathbb{E}[e^{sX_i}]\right)^n = (M(s))^n. \tag{3}$$

□

Lemma 2. *The Fourier transform of a Gaussian PDF is a Gaussian in frequency space:*

$$\mathcal{F}\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}}\right)(\omega) = e^{-\frac{\omega^2\sigma^2}{2}} \tag{4}$$

Proof. of Lemma 2. Presented better than I can at
`http://www.cse.yorku.ca/~kosta/CompVis_Notes/fourier_transform_Gaussian.pdf`. (This is something I just sort of remembered when I had to use it in the proof.) □

Proof. of the Central Limit Theorem.

We're first going to make the simplifying assumption that $\mu = 0$. If it isn't, we can rescale the X_is so that it is.

$$\text{To show: } S_n \text{ approx.} \sim \mathcal{N}(0, \sigma^2/n)$$

These distributions are approximately equal if the pdf of S_n converges pointwise to that of $\mathcal{N}(0, \sigma^2/n)$. I'd first like to rescale this so that we only have n dependence on the left (so that I can apply a limit), so we can equivalently state this as the pdf of $\sqrt{n}S_n = \frac{\sum X_i}{\sqrt{n}}$ converging pointwise to the pdf of $\mathcal{N}(0, \sigma^2)$.

$$\text{To show: } \lim_{n\to\infty} f_{\sum X_i/\sqrt{n}}(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2} \tag{5}$$

We have to construct a pdf of a sum of random variables. We can deal with this through convolution, or through MGFs; since they're i.i.d., MGFs will be easier because that'll just reduce to a power on the MGF of a single variable, by Lemma 1. Therefore, we can start looking at the MGF of $X_i/\sqrt{n}$.

In general, we don't know anything about what the MGF of X_i is, but we do know that n is large which means that $\frac{1}{\sqrt{n}}$ is small, so we can Taylor expand the MGF of an arbitrary $X_i/\sqrt{n}$ and consider all terms of order less than $\frac{1}{n}$ to be negligible. This is the heart of the proof: as n is driven high, the sample mean gets closer to the true mean and moments above second-order contribute less and less to its distribution around the true mean.

$$M_{X_i}\left(\frac{s}{\sqrt{n}}\right) \approx M_{X_i}(0) + \frac{M'_{X_i}(0)}{1!}\frac{s}{\sqrt{n}} + \frac{M''_{X_i}(0)}{2!}\frac{s^2}{n} + \dots \tag{6}$$

If you'd like you can specify X_i as X_1 for clarity, because they're all the same. We know that the nth derivative of an MGF evaluated at $s = 0$ is the nth moment of a distribution. The zeroth moment is 1, the first moment is the mean which we've set to zero, and the second moment is the variance which we know. Therefore

$$M_{X_i}\left(\frac{s}{\sqrt{n}}\right) \approx 1 + \frac{\sigma^2 s^2}{2n} \tag{7}$$

We still have the problem of translating this MGF to a pdf. The MGF is essentially a frequency-domain way of looking at a distribution, so it's related to the pdf by a Fourier transform:

$$M_X(s) = M_X(i\omega) = \mathcal{F}(f_X(x))(\omega) = \int_{-\infty}^{\infty} f_X(x)e^{i\omega x}dx = \mathbb{E}[e^{i\omega x}] \tag{8}$$

(The exact relation may be a Laplace transform, but we'll get a nice result out of the Fourier version soon). You can see that the above equation reduces to the usual integral for the MGF, just with the substitution $s \to i\omega$.

What we actually want to use now is the *inverse* Fourier transform, which gives us the pdf from the MGF.

$$f_X(x) = \mathcal{F}^{-1}(M_X(i\omega))(x) = \int_{-\infty}^{\infty} M_X(i\omega)e^{-i\omega x}d\omega \tag{9}$$

We substitute in the MGF that we're interested in, that of $\sum X_i/\sqrt{n}$, and apply Lemma 1:

$$M_{\sum X_i/\sqrt{n}}(s) = \left(M_{X_i}\left(\frac{s}{\sqrt{n}}\right)\right)^n = \left(M_{X_i}\left(\frac{s}{\sqrt{n}}\right)\right)^n = \left(1 + \frac{\sigma^2 s^2}{2n}\right)^n \tag{10}$$

We send $s \to i\omega$ to match the form of a Fourier transform.

$$M_{\sum X_i/\sqrt{n}}(i\omega) = \left(1 - \frac{\sigma^2\omega^2}{2n}\right)^n \tag{11}$$

This is the limit definition for e^x, and since we're working in the limit $n \to \infty$, this is valid for us to use.

$$M_{\sum X_i/\sqrt{n}}(i\omega) = e^{-\frac{\sigma^2\omega^2}{2}} \tag{12}$$

As we saw in Lemma 2, this is the Fourier transform of a Gaussian $\mathcal{N}(0, \sigma^2)$, so the pdf corresponding to this MGF must be that of a normal distribution with variance σ^2. Therefore, $\lim_{n\to\infty} f_{\sum X_i/\sqrt{n}} = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}$, which is what we wanted! $\square$

Comments

I first tried to prove this using an $\epsilon - N$ argument, i.e. if we choose the cdf $F_{S_n}(z)$ of the sample mean of any distribution, then for any $\epsilon > 0$ there exists an $N \in \mathbb{R}$ such that $n > N$ implies $|F_{S_n}(z) - \Phi(z \mid \mu, \sigma^2/n)| < \epsilon$. Unfortunately, this was awful, because the cleanest way to define the cdf sample mean was by integrating a pdf that I got through an MGF argument similar to what's in the eventual proof, which produces a double or triple integral whose exact form isn't known and so can't be readily simplified or approximated.

Further, after I'd finished the proof below, I was shown a counterexample to any possible $\epsilon - N$ argument: for any N, select $X_i \sim \text{Gamma}\left(\frac{1}{N}, 1\right)$. (The Gamma$(k, \theta)$ distribution has PDF $\frac{x^{k-1}e^{-x/\theta}}{\theta^k\Gamma(k)}$; I haven't yet verified that this counterexample works.) The sample mean then goes to an exponential distribution. This means it's impossible to come up with a generalized bound that works for any distribution, i.e. there does not exist a function $N(\epsilon)$ that is guaranteed to give us a certain confidence interval on a sample-mean distribution regardless of the underlying distribution from which we drew the X_is.

I also presented the FT and IFT equations even though I didn't end up using them (the Fourier work was done in Lemma 2 which I didn't prove myself); I figured this would be nicer than a black-box transform from an MGF that looks roughly right. You could also recognize the final expression as the usual MGF for a Gaussian, which wouldn't require the $s \to i\omega$ transform but which I would find less satisfying.