Comments on Lecture 8 - Observations about the Central Limit Theorem

Luisa Cutillo

November 5, 2018

1 Central Limit Theorem

The Central Limit Theorem (CLT) deals with the long-run behaviour of the sample mean as n grows. In general, the colloquial result is that *everything becomes Normal eventually.* During the lecture we formalized this concept anf proved the theorem. In the following I'll try to give you more insight into this incredibly powerful result.

Consider i.i.d. random variables $X_1, X_2, \ldots, X_n$ each with meanμ and variance σ^2 (NOTE WELL: it doesn't necessarily have to be a Normal distribution! For example, if we had that each X was distributed Unif(0,1), then μ=1/2 and σ^2=1/12).

Again, define $\bar{X}_n$ as the *sample mean* of X. We can write this out as:

$$\bar{X}_n = \frac{X_1 + \ldots + X_n}{n}$$

As discussed above, we know that the sample mean is itself a random variable, and we know that it approaches the true mean in the long run by the Law of Large Numbers (LLN). However, are we able to nail down a specific distribution for this random variable as we approach the long run, not just the value that it converges to? A good place to start is to find the mean and variance; these parameters won't tell us what the distribution is, but they will be useful once we determine the distribution. To find the expectation and variance, we can just *brute force* our calculations. First, for the expectation, we take the expectation of both sides:

$$E(\bar{X}_n) = E(\frac{X_1 + \ldots + X_n}{n})$$

By linearity:

$$= E(\frac{X_1}{n}) + \ldots + E(\frac{X_n}{n})$$

Since n is a known constant, we can factor it out of the expectation:

$$= \frac{1}{n}E(X1) + \ldots + \frac{1}{n}E(Xn)$$

Now, we are left with the expectation, or mean, of each X. Do we know these values? Well, recall above that, by the set-up of the problem, each X is a random variable with mean μ. That is, the expectation of each X is μ. We get:

$$= \mu_n + ... + \mu_n$$

We have n of these terms, so they sum to:

$$= \mu$$

Hence, we get that $E(\bar{X}_n) = \mu$. Think about this result: it says that the average of the sample mean is equal to μ, where μ is the average of each of the random variables that make up the sample mean. This is intuitive; the sample mean should have an average of μ (you could say that $\bar{X}_n$ is unbiased for μ, since it has expectation μ; this is a concept that you will explore more in detail). Let's now turn to the Variance:

$$Var(\bar{X}_n) = Var(\frac{X_1 + ... + X_n}{n})$$

We know that the X terms are independent, so the variance of the sum is the sum of the variances:

$$= Var(\frac{X_1}{n}) + ... + Var(\frac{X_n}{n})$$

Since n is a constant, we factor it out (remembering to square it):

$$= \frac{1}{n^2}Var(X_1) + ... + \frac{1}{n^2}Var(X_n)$$

Do we know the variance of each X term? In the set-up of the problem, it was defined as ?2, so we can simply plug in ?2 for each variance:

$$= \frac{\sigma^2}{n^2} + ... + \frac{\sigma^2}{n^2}$$

We have n of these terms (since we have n random variables and thus n variances) so this simplifies to:

$$= \frac{\sigma^2}{n}$$

Consider this result for a moment. First, consider when n=1. In this case, the sample mean is just X_1 (since, by definition, we would have $\frac{X_1}{1} = X_1$). The variance that we calculated above, $\frac{\sigma^2}{n}$, comes out to σ^2 when $n = 1$, which makes sense, since this is just the variance of X_1. Next, consider what happens to this variance as n grows; it gets smaller and smaller, since n is in the denominator. Does this make sense? As n grows, we are essentially adding up more and more random variables in our sample mean calculation.

It makes sense, then, that the overall sample mean will have less variance; among other things, adding up more random variables means that the effect of *outlier* random variables is lessened (i.e., if we observe an extremely large value for X_1, it is mediated by the sheer number of random variables).

So, we found that the sample mean has mean μ and variance $\frac{\sigma^2}{n}$, where ? is the mean of each underlying random variable, σ^2 is the variance of each underlying random variable, and n is the total number of random variables. Now that we have the parameters, we are ready for the main result of the CLT.

The CLT states that, for large n, the distribution of the sample mean approaches a Normal distribution. This is an extremely powerful result, because it holds no matter what the distribution of the underlying random variables (i.e., the X's) is. We know that Normal random variables are governed by the mean and variance (i.e., these are the two parameters), and we already found the mean and variance of $\bar{X}_n$, so we can say:

$$\bar{X_n} \to^D N(\mu, \frac{\sigma^2}{n})$$

Where $\to^D$ means *converges in distribution*; it's implied here that this convergence takes place as n, or the number of underlying random variables, grows.

Think about this distribution as n gets extremely large. The mean, μ, will be unaffected, but the variance will be close to 0, so the distribution will essentially be a constant (specifically, the constant μ with no variance). This makes sense: if we take an extremely high number of draws from a distribution, we should get that this sample mean is at the true mean, with very little variance. It?s also the result we saw from the LLN, which said that the sample mean approaches a constant: as n grows here, we approach a variance of 0, which essentially means we have a constant (since constants have variance 0). The CLT just describes the distribution *on the way* to the LLN convergence.

Hopefully this brings some clarity to the statement "everything becomes Normal": taking the sum of i.i.d. random variables (we worked with the sample mean here, but the sample mean is just the sum divided by a constant n), regardless of the underlying distribution of the random variables, yields a Normal distribution. Please check out the Notebook 4, 04-CentralLimitTheorem.ipynb, on my Microsoft Azure MATH2715 page:
`https://notebooks.azure.com/luisacutillo78/libraries/Luisa0`
for a graphical interpretation of the CLT.
Further Reference: `https://bookdown.org/probability/beta/`