Some Comments on The Law of Large Numbers and The Central Limit Theorem

1 Proofs using the MGF

The standard proof of the "weak" LLN uses the Chebyshev Inequality, which is a useful inequality in its own right. However, we can also prove it by the same method as the CLT is. While this approach has a big drawback – we have to rely on a theorem we cannot prove here, that is the fact that convergence of Moment Generating Functions and/or Characteristic Functions implies "weak convergence" of the distributions – it dovetails with the CLT proof to highlight an interesting detail.

1.1 The LLN

Suppose the i.i.d. random variables X_i have a well-defined MGF. We have

$$E\left[e^{tX_i}\right] = M(t) \qquad E\left[e^{\frac{1}{n}X_i t}\right] = M\left(\frac{t}{n}\right)$$

Hence,

$$E\left[e^{t\frac{1}{n}\sum_{i=1}^{n} X_i}\right] = E\left[e^{\sum_{i=1}^{n}\frac{t}{n}X_i}\right] = \left(M\left(\frac{t}{n}\right)\right)^n$$

Now, M has continuous derivatives, hence can be expanded using Taylor's formula and

$$\left(M\left(\frac{t}{n}\right)\right)^n = \left(M(0) + \frac{t}{n}M'(0) + o\left(\frac{t}{n}\right)\right)^n$$

Since $M(0) = 1$, by a well-known basic fact, the right hand side converges, as $n \to \infty$ to

$$e^{tM'(0)} = e^{t\cdot EX_i}$$

You can easily see that this is the MGF of a *degenerate* RV, equal to EX_i with probability one. Thus, we have that the distribution of our sum converges weakly to the distribution of such a variable.

The statement is not quite the same as the usual LLN, but it can be proved that weak convergence to a constant implies convergence in probability.

1.2 The CLT

The proof of the CLT for variables admitting a MGF is on page 434-435 of the book. Note that it is equivalent to the use of L'Hospital's Theorem to use Taylor's formula. Here is a short summary.

We first change our variables to $Y_i = \frac{X_i - EX_i}{\sqrt{Var(X_i)}}$. This entails that there is no loss of generality, of course, if we assume that our variables have mean zero, and variance 1. Hence, their MGF can be written as

$$M(t) = M(0) + tM'(0) + \frac{t^2}{2}M''(0) + o\left(t^2\right) = 1 + \frac{t^2}{2} + o\left(t^2\right)$$

We now have

$$E\left[e^{t\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i}\right] = E\left[e^{\sum_{i=1}^{n} \frac{t}{\sqrt{n}} Y_i}\right] = \left(M\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 + \frac{t^2}{2}\cdot\frac{1}{n} + o\left(\frac{t^2}{n}\right)\right)^n \to e^{\frac{t^2}{2}}$$

1.3 The Point of These Proofs

As you can see, both proofs can rely on a Taylor expansion and

- The LLN corresponds to a first order expansion
- The CLT corresponds to a second order expansion

You may or may not be helped by these statements, but they do suggest that, in some intuitive way, "expected values are a first approximation to random variables", and, more interestingly, "Gaussian distributions are the first non-trivial approximation to general distributions (that have enough moments)".

2 Why Not Use Characteristic Functions?

Actually, the "real" proofs do use characteristic functions. The problem in this course is that characteristic functions are complex-valued functions, so that, though we may be tempted to proceed formally as above, just plugging a "*it*" wherever we see "*t*", this is not quite legitimate. It turns out that the result we would get is correct, but the proof would not be rigorous.

In short, we have to deal with powers of complex numbers, which is done precisely via stuff like

$$C(t) = E\left[e^{itX}\right] = e^{\phi(t)}$$

where $\phi(t) = \log C(t)$ is the *principal determination of the complex logarithm of* C. We would then expand $\phi\left(\frac{t}{n}\right)$, and $\phi\left(\frac{t}{\sqrt{n}}\right)$ using the complex Taylor formula. To deal correctly with all this, we do need some complex analysis.

Additionally, the CLT requires us to show that the characteristic function of a $N(0,1)$ variable is $e^{-\frac{t^2}{2}}$. While this can be proved, somewhat unintuitively, without serious complex analysis (one such proof is in the very nice, but somewhat advanced, book by J.Jacod, P.Protter: *Probability Essentials*. Springer Universitext 2000, p. 103), the usual proofs go through "contour integration" of functions of complex variable or "analytic continuation" of such functions.