probability, $Z_N \to_P Z$, if

$$P(|Z_N - Z| > \epsilon) \to 0$$

as $N \to \infty$. (The weak LLN is called "weak" because it asserts convergence in probability, which turns out to be a somewhat "weak" sense of stochastic convergence, in the mathematical sense that there are "stronger" forms of convergence — that is, it's possible to find sequences of r.v.'s which converge in probability but not in these stronger senses. In addition, it's possible to prove the LLN without assuming that the variance exists; existence of the mean turns out to be sufficient. But discussing these stronger concepts of convergence would take us too far afield[13]; convergence in probability will be plenty strong enough for our purposes.)

We discussed convergence of r.v.'s above; it's often also useful to think about convergence of distributions. We say a sequence of r.v.'s with cdf's $F_N(u)$ "converge in distribution" if

$$\lim_{N\to\infty} F_N(u) \to F(u)$$

for all u such that F is continuous at u (here F is itself a cdf). **Exercise 35**: Explain why do we need to restrict our attention to continuity points of F. (Hint: think of the following sequence of distributions: $F_N(u) = 1(u < 1/N)$, where the "indicator" function of a set A is one if $x \in A$ and zero otherwise.)

It's worth emphasizing that convergence in distribution — because it only looks at the cdf — is in fact weaker than convergence in probability. For example, if p_X is symmetric, then the sequence $X, -X, X, -X, ...$ trivially converges in distribution to X, but obviously doesn't converge in probability. **Exercise 36**: Prove that convergence in probability actually is stronger, that is, implies convergence in distribution.

Central limit theorem

The second fundamental result in probability theory, after the LLN, is the CLT: if X_i are i.i.d. with mean zero and variance 1, then

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \to_D \mathcal{N}(0, 1),$$

[13] Again, see e.g. Breiman '68 for more information.

Figure 9: De Moivre and Laplace.

where $\mathcal{N}(0,1)$ is the standard normal density. More generally, the usual rescalings tell us that

$$\frac{1}{\sigma(X)\sqrt{N}}\sum_{i=1}^{N}(X_i - E(X)) \to_D \mathcal{N}(0,1).$$

Thus we know not only that (from the LLN) the distribution of the sample mean approaches the degenerate distribution on $E(X)$, but moreover (from the CLT) we know exactly what this distribution looks like, asymptotically, if we take out our magnifying glass and zoom in on $E(X)$, to a scale of $N^{-1/2}$. In this sense the CLT is a stronger result than the WLLN: it gives more details about what the asymptotic distribution actually looks like.

One thing worth noting: keep in mind that the CLT really only tells us what's going on in the local neighborhood $(E(X) - N^{-1/2}c, E(X) + N^{-1/2}c)$ — think of this as the mean plus or minus a few standard deviations. But this does *not* imply that, say,

$$P(\frac{1}{N}\sum_{i=1}^{N} X_i \leq -\epsilon) \sim \int_{\infty}^{-\epsilon} \mathcal{N}(0, \frac{1}{N})(x)dx = \int_{\infty}^{-\sqrt{N}\epsilon} \mathcal{N}(0,1)(x)dx \quad \textbf{not true};$$

a different asymptotic approximation typically holds for the "large deviations," the tails of the sample mean distribution[14].

More on stochastic convergence

So, as emphasized above, convergence in distribution can drastically simplify our lives, if we can find a simple approximate (limit) distribution to substitute for our original complicated distribution. The CLT is the canonical example of this; the Poisson theorem is another. What are some general methods to prove convergence in distribution?

Delta method

The first thing to note is that if X_N converge in distribution or probability to a constant c, then $g(X_N) \rightarrow_D g(c)$ for any continuous function $g(.)$. **Exercise 37**: Prove this, using the definition of continuity of a function: a function $g(u)$ is continuous at u if for any possible fixed $\epsilon > 0$, there is some (possibly very small) δ such that $|g(u+v) - g(u)| < \epsilon$, for all v such that $-\delta < v < \delta$. (If you're having trouble, just try proving this for convergence in probability.)

So the LLN for sample means immediately implies an LLN for a bunch of functions of the sample mean, e.g., if X_i are i.i.d. with $V(X) < \infty$, then

$$\left(\prod_{i=1}^{N} e^{X_i}\right)^{1/N} = e^{\frac{1}{N}\sum_{i=1}^{N} X_i} \rightarrow_P e^{E(X)},$$

(which of course should not be confused with $E(e^X)$; in fact, **Exercise 38**: Which is greater, $E(e^X)$ or $e^{E(X)}$? Give an example where one of $E(e^X)$ or $e^{E(X)}$ is infinite, but the other is finite).

We can also "zoom in" to look at the asymptotic distribution (not just the limit point) of $g(Z)$, whenever g is sufficiently smooth. For example, let's say $g(.)$ has a Taylor expansion at u,

$$g(z) = g(u) + g'(u)(z-u) + o(|z-u|), \;\; |z-u| \rightarrow 0,$$

where $|g'(u)| > 0$ and $z = o(y)$ means $z/y \rightarrow 0$. Then if

$$a_N(z_N - u) \rightarrow_D q,$$

[14]See e.g., Large deviations techniques and applications, Dembo and Zeitouni '93, for more information.

for some limit distribution q and a sequence of constants $a_N \to \infty$ (think $a_N = N^{1/2}$, if Z_N is the sample mean), then

$$a_N \frac{g(Z_N) - g(u)}{g'(u)} \to_D q,$$

since

$$a_N \frac{g(Z_N) - g(u)}{g'(u)} = a_N(Z_N - u) + o\left(a_N \frac{|Z_N - u|}{g'(u)}\right);$$

the first term converges in distribution to q (by our assumption) and the second one converges to zero in probability (**Exercise 39**: Prove this; i.e., prove that the remainder term

$$a_N \frac{g(Z_N) - g(u)}{g'(u)} - a_N(Z_N - u)$$

converges to zero in probability, by using the Taylor expansion formula). In other words, limit distributions are passed through functions in a pretty simple way. This is called the "delta method" (I suppose because of the deltas and epsilons involved in this kind of limiting argument), and we'll be using it a lot. The main application is when we've already proven a CLT for Z_N,

$$\sqrt{N} \frac{Z_N - \mu}{\sigma} \to_D N(0, 1),$$

in which case

$$\sqrt{N}(g(Z_N) - g(\mu)) \to_D N(0, \sigma^2(g'(\mu))^2).$$

Exercise 40: Assume $N^{1/2} Z_N \to_D \mathcal{N}(0, 1)$. Then what is the asymptotic distribution of 1) $g(Z_N) = (Z_N - 1)^2$? 2) what about $g(Z_N) = Z_N^2$? Does anything go wrong when applying the delta method in this case? Can you fix this problem?

Mgf method

What if the r.v. we're interested in, Y_N, can't be written as $g(X_N)$, i.e., a nice function of an r.v. we already know converges? Are there methods to prove limit theorems directly?

Here we turn to our old friend the mgf. It turns out that the following generalization of the mgf invertibility theorem we quoted above is true:

Theorem 2. *The distribution functions F_N converge to F if:*

- *the corresponding mgf's $M_{X_N}(s)$ and $M_X(s)$ exist (and are finite) for all $s \in (-z, z)$, for all N, for some positive constant z.*

- *$M_{X_N}(s) \to M_X(s)$ for all $s \in (-z, z)$.*

So, once again, if we have a good handle on the mgf's of X_N, we can learn a lot about the limit distribution. In fact, this idea provides the simplest way to prove the CLT.

Proof: assume X_i has mean zero and unit variance; the general case follows easily, by the usual rescalings.

Now let's look at $M_N(s)$, the mgf of $\frac{1}{\sqrt{N}} \sum_{i=1}^N X_i$. If X_i has mgf $M(s)$, then $\frac{1}{\sqrt{N}} \sum_{i=1}^N X_i$ has mgf

$$M(s/\sqrt{N})^N.$$

Now let's make a Taylor expansion. We know that $M(0) = 1, M'(0) = 0$, and $M''(0) = 1$. (Why?) So we can write

$$M(s) = 1 + s^2/2 + o(s^2).$$

Now we just note that $M_N(s)$ converges to $e^{s^2/2}$, recall the mgf of a standard normal r.v., and then appeal to our general convergence-in-distribution theorem for mgf's. □