STA205 Probability & Measure

CLT Variations

2002 April 8

The Central Limit Theorem tells us that $(S_n - n\mu)/\sigma\sqrt{n} \Rightarrow \mathsf{No}(0,1)$ whenever $S_n = \sum X_j$ for *i.i.d.* $X_j \in L^2$; informally, it says that sums and averages of a large number of independent random variables have approximately a normal distribution.

Yet it is well-known that the partial sum $S_n = \sum X_j$ of Bernoulli random variables $X_j \sim \mathsf{Bi}(1, p_n)$ with small success probabilities $n\,p_n \to \lambda$ has an approximate Poisson $\mathsf{Po}(\lambda)$ distribution, not Normal, as $n \to \infty$. What sort of distributional limits are possible for partial sums of independent random variables?

The answer to two specific forms of this question were given in the 1920's by by French probabilist Paul Lévy and Russian Alexander Khinchine; Kolmogorov and de Finetti also made contributions. I'll state the results and sketch the idea of the proof.

Theorem 1 (Lévy-Khinchine) *Let S_n be the partial sum of independent identically distributed random variables ξ_j, and suppose that for some sequences of real numbers A_n, C_n the sequence $S_n/C_n - A_n$ converges in distribution. Then the limit is necessarily a **stable distribution** $\mathsf{St}(\alpha, \beta, \gamma, \delta)$ with characteristic function of the form*

$$\log \phi(\omega) = \begin{cases} i\delta\omega - |\gamma\omega|^\alpha + i\beta \tan\frac{\pi\alpha}{2}\{|\gamma\omega|^\alpha \operatorname{sgn}\omega - \gamma\omega\} & \alpha \neq 1 \\ i\delta\omega - |\gamma\omega| \quad - \frac{2i\beta}{\pi}\gamma\omega\,\log|\gamma\omega| & \alpha = 1 \end{cases}$$

for some numbers $0 < \alpha \le 2$, $-1 \le \beta \le 1$, $0 < \gamma < \infty$, and $\delta \in \mathbb{R}$. The symmetric case ($\beta = \delta = 0$) is simpler,

$$\log \phi(\omega) = -|\gamma\omega|^\alpha,$$

and includes the standard Normal and Cauchy distributions.

If the summands ξ_j are allowed to have different distributions as n increases, then the class of limits enlarges:

Theorem 2 (Lévy-Khinchine) *Let S_n be the partial sum of independent identically distributed random variables $\xi_j^{(n)}$, and suppose that sequence S_n converges in distribution. Then the limit is necessarily an **infinitely divisible distribution** with characteristic function of the form*

$$\log \phi(\omega) = i\delta\omega - \sigma^2\omega^2/2 + \int_{\mathbb{R}} \{e^{i\omega u} - 1 - i\omega h(u)\}\nu(du)$$

for some numbers $\delta \in \mathbb{R}$ and $0 \le \sigma^2 < \infty$, some bounded function $h(u)$ satisfying $h(u) = u + O(u^2)$ near $u = 0$, and some nonnegative measure $\nu(du)$ satisfying the integrability condition

$$\int_{\mathbb{R}} (1 \wedge u^2)\, \nu(du) < \infty.$$

If the limit has finite variance this can be written in the form

$$\log \phi(\omega) = i\delta\omega - \sigma^2\omega^2/2 + \int_{\mathbb{R}} \{e^{i\omega u} - 1 - i\omega u\}\nu(du)$$

with mean δ and variance $\sigma^2 + \int u^2\, \nu(du)$.

Finally, an interesting variation on the Central Limit Theorm is

Theorem 3 (Berry-Esseen) *Let $X_j \in L^3$ be i.i.d. with mean μ and variance σ^2. Let $F_n(z) \equiv \Pr\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \le z\right]$ for $z \in \mathbb{R}$ be the CDF of the standardized partial sums, and let $\Phi(z)$ be the CDF of the standard normal distribution. Then*

$$\sup_{z\in\mathbb{R}} |F_n(z) - \Phi(z)| \le C \frac{\mathsf{E}|X_1 - \mu|^3}{\sigma^3\sqrt{n}}$$

for an absolute constant C known not to exceed $33/4$.

This gives an answer to the inevitable question, *How large does n have to be?* concerning the CLT; for Bernoulli random variables, for example, where $\mathsf{E}[|\xi - p|^3] = pq(p^2 + q^2)$ and $\sigma^2 = pq$, the bound is

$$|F_n(z) - \Phi(z)| \le C \frac{p^2 + q^2}{\sqrt{n\,p\,q}},$$

or $C/\sqrt{n}$ for $p = q = 1/2$ and approximately $C/\sqrt{n\,p}$ for small $p \approx 0$. Notice that in our original example with $n\,p_n \to \lambda$ this bound does not shrink to zero, and the Central Limit Theorem does not apply.