1 Five ways toward CLT

I mean the central limit theorem (CLT) for independent identically distributed (i.i.d.) random variables with second moments. See 'Results formulated', Th. 3b1.

Many generalizations and sharpenings are well-known. Various proofs sketched below differ in their suitability for generalizations and sharpenings.

We denote $S_n = X_1 + \dots + X_n$.

1a Moment method

(a) The limit

$$\lim_{n\to\infty} \mathbb{E} \Big(\frac{S_n}{\sqrt{n}} \Big)^m$$

exists for each $m = 1, 2, 3, \dots$ and does not depend on the distribution of X_1 (provided that X_1 has all moments).

(b) Convergence of moments implies convergence of distributions (provided that moments do not grow too fast).

(c) A distribution of X_1 without higher moments is approximated by distributions with all moments (just bounded).

1b Fourier transform (characteristic functions)

(a) For $\lambda \in \mathbb{R}$,

$$\mathbb{E} \exp \Big(\mathrm{i}\lambda \frac{S_n}{\sqrt{n}} \Big) \to \exp \Big(-\frac{\lambda^2}{2} \Big) \quad \text{as } n \to \infty$$

uniformly on bounded intervals.

(b) Convergence of distributions follows.

1c Smooth test functions

(a)

$$\mathbb{E}\, f\Big(\frac{S_n}{\sqrt{n}} \Big) - \mathbb{E}\, f\Big(\frac{\tilde S_n}{\sqrt{n}} \Big) \to 0 \quad \text{as } n \to \infty$$

for every function $f : \mathbb{R} \to \mathbb{R}$ having continuous bounded derivatives f, f', f'', f'''. Here $\tilde S_n = \tilde X_1 + \dots + \tilde X_n$ where $\tilde X_1, \tilde X_2, \dots$ satisfy the same conditions as $X_1, X_2, \dots$ (and are arbitrary otherwise).

(b) Convergence of distributions follows easily.

1d Using Poisson distributions

(a) S_n is close to S_{N_n} where N_n is a Poisson random variable with $\mathbb{E}\, N_n = n$ (independent of $X_1, X_2, \dots$).

(b) Assuming that the distribution of X_1 is concentrated on a finite set, one represents S_{N_n} as a linear combination of *independent* Poisson random variables.

(c) Convergence of distributions follows.

(d) A distribution of X_1 is approximated by distributions concentrated on finite sets.

1e Using multinomial distributions

(a) Assuming that the distribution of X_1 is concentrated on a finite set, one represents S_n as a linear function of a multinomial random vector.

(b) The multinomial distribution converges to a multinormal distribution.

(c) A distribution of X_1 is approximated by distributions concentrated on finite sets.

2 A proof of CLT

Given a probability measure ν on $\mathbb{R}$ and a bounded continuous function $f : \mathbb{R} \to \mathbb{R}$, their *convolution* $\nu * f$ is a function $\mathbb{R} \to \mathbb{R}$ defined by

$$(\nu * f)(x) = \int f(x+y)\, \nu(\mathrm{d}y)\,.$$

Note that $(\nu * f)(x) = \mathbb{E}\, f(x+X)$ if $X \sim \nu$, and

$$\mathbb{E}\, f\Big(\frac{S_n}{\sqrt{n}}\Big) = (\mu_n * \dots * \mu_n * f)(0) = (\mu_n^{*n} * f)(0) \tag{2.1}$$

where μ_n is the distribution of $X_1/\sqrt{n}$.

2.2 Exercise. $\nu * f$ is bounded and continuous.

Prove it.

2.3 Exercise. Let f have a continuous derivative f', and f, f' be bounded. Then $\nu * f$ has a continuous derivative, and

$$(\nu * f)' = \nu * f'\,.$$

Prove it.

Hint: $(\nu * f)(x+h) - (\nu * f)(x) = \int_x^{x+h} (\nu * f')(u)\, \mathrm{d}u$.

We have $\int x\,\mu(\mathrm{d}x) = 0$, $\int x^2\,\mu(\mathrm{d}x) = 1$. Let $\tilde\mu$ be another probability measure on $\mathbb{R}$ satisfying $\int x\,\tilde\mu(\mathrm{d}x) = 0$, $\int x^2\,\tilde\mu(\mathrm{d}x) = 1$. Taking into account that $\int f(x)\,\mu_n(\mathrm{d}x) = \int f(x/\sqrt{n})\,\mu(\mathrm{d}x)$ for any f, we introduce $\tilde\mu_n$ by $\int f(x)\,\tilde\mu_n(\mathrm{d}x) = \int f(x/\sqrt{n})\,\tilde\mu(\mathrm{d}x)$. Note that $\int(a+bx+cx^2)\,(\mu-\tilde\mu)(\mathrm{d}x) = 0$ for all a, b, c; also $\int(a+bx+cx^2)\,(\mu_n-\tilde\mu_n)(\mathrm{d}x) = 0$.

Let $f : \mathbb{R} \to \mathbb{R}$ have continuous derivatives f', f'', f'''; assume that f, f', f'', f''' are bounded. We define g by

$$f(x) = f(0) + f'(0)x + \frac{1}{2}f''(0)x^2 + g(x)\,,$$

then

$$\int f\,\mathrm{d}\mu_n - \int f\,\mathrm{d}\tilde\mu_n = \int g\,\mathrm{d}\mu_n - \int g\,\mathrm{d}\tilde\mu_n\,.$$

However,

$$|g(x)| \le \|f'''\| \cdot \frac{1}{6}|x|^3$$

($\|\cdot\|$ stands for the supremum norm). If X_1 is bounded, that is, μ is concentrated on some $[-M, M]$, then

$$\Big|\int g\,\mathrm{d}\mu_n\Big| = \Big|\int_{-M}^{M} g\Big(\frac{x}{\sqrt{n}}\Big)\,\mu(\mathrm{d}x)\Big| \le 2M \cdot \|f'''\| \cdot \frac{1}{6} \cdot \Big(\frac{M}{\sqrt{n}}\Big)^3 = o\Big(\frac{1}{n}\Big)\,.$$

For unbounded X_1 an additional argument is needed:

$$|g(x)| \le \|f''\| \cdot |x|^2\,,$$

thus,

$$\begin{aligned}\Big|\int g\,\mathrm{d}\mu_n\Big| = \Big|\int g\Big(\frac{x}{\sqrt{n}}\Big)\,\mu(\mathrm{d}x)\Big| \le \int_{|x|<n^{1/12}} \dots + \int_{|x|>n^{1/12}} \dots \le \\ \le \frac{1}{6}\|f'''\|n^{-5/4} + \|f''\|\frac{1}{n}\int_{|x|>n^{1/12}} x^2\,\mu(\mathrm{d}x) = o\Big(\frac{1}{n}\Big)\,.\end{aligned}$$

The same holds for $\int g\,\mathrm{d}\tilde\mu_n$, and we get

$$\Big|\int f\,\mathrm{d}\mu_n - \int f\,\mathrm{d}\tilde\mu_n\Big| \le \frac{1}{3}\|f'''\|n^{-5/4} + \|f''\|\frac{1}{n}\int_{|x|>n^{1/12}} x^2\,(\mu(\mathrm{d}x)+\tilde\mu(\mathrm{d}x)) = o\Big(\frac{1}{n}\Big)\,.$$

In other words, $(\mu_n * f - \tilde\mu_n * f)(0) \le \dots = o(1/n)$. Similarly, $(\mu_n * f - \tilde\mu_n * f)(x) \le \dots = o(1/n)$ uniformly in x, therefore

$$\|\mu_n * f - \tilde\mu_n * f\| = o\Big(\frac{1}{n}\Big)\,. \tag{2.4}$$

The right-hand side depends on f only via $\|f''\|$ and $\|f'''\|$. Thus,

$$\sup_{\|f''\|\le C_2, \|f'''\|\le C_3} \|\mu_n * f - \tilde\mu_n * f\| = o\Big(\frac{1}{n}\Big) \tag{2.5}$$

for all C_2, C_3.

We use the sum

$$\mu_n^{*n} * f - \tilde\mu_n^{*n} * f = \sum_{k=0}^{n-1} \big(\mu_n^{*(n-k)} * \tilde\mu_n^{*k} * f - \mu_n^{*(n-k-1)} * \tilde\mu_n^{*(k+1)} * f \big) = \\ = \sum_{k=0}^{n-1} \mu_n^{*(n-k-1)} * (\mu_n - \tilde\mu_n) * \tilde\mu_n^{*k} * f \,,$$

note that $\|(\tilde\mu_n^{*k} * f)''\| \le \|f''\|$, $\|(\tilde\mu_n^{*k} * f)'''\| \le \|f'''\|$ and conclude that $\|(\mu_n - \tilde\mu_n) * \tilde\mu_n^{*k} * f\| = o(1/n)$ uniformly in k. Taking into account that $\|\mu_n^{*(n-k-1)} * (\dots)\| \le \|(\dots)\|$ we get

$$\|\mu_n^{*n} * f - \tilde\mu_n^{*n} * f\| \to 0 \quad \text{as } n \to \infty \,.$$

By (2.1),

$$\mathbb{E}\, f\Big(\frac{S_n}{\sqrt n}\Big) - \mathbb{E}\, f\Big(\frac{\tilde S_n}{\sqrt n}\Big) \to 0 \quad \text{as } n \to \infty \,,$$

which completes the first part of the proof (part (a) of 1c).

2.6 Exercise. The following two conditions are equivalent for every sequence of probability measures $\nu_1, \nu_2, \dots$ on $\mathbb{R}$:

(a) $\nu_n\big((-\infty, x]\big) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \mathrm{e}^{-u^2/2}\, \mathrm{d}u$ (as $n \to \infty$) for all $x \in \mathbb{R}$;

(b) $\int f \,\mathrm{d}\nu \to \frac{1}{\sqrt{2\pi}} \int f(u) \mathrm{e}^{-u^2/2}\, \mathrm{d}u$ (as $n \to \infty$) for every function $f : \mathbb{R} \to \mathbb{R}$ having continuous bounded derivatives f, f', f'', f''' and the limits $f(-\infty)$, $f(+\infty)$.

Prove it.

Hint: (a) $\Longrightarrow$(b): approximate f uniformly by step functions; (b) $\Longrightarrow$(a): construct smooth $f : \mathbb{R} \to [0,1]$ that vanishes on $[x+\varepsilon, \infty)$ and equals 1 on $(-\infty, x-\varepsilon]$.

Choosing $\tilde X_1$ as in the De Moivre-Laplace theorem ('Results formulated', Th. 2b2) and using that theorem we see that $\tilde\mu_n$ satisfy 2.6(a), therefore 2.6(b), too. Using the result of the first part we see that μ_n satisfy 2.6(b), therefore 2.6(a).