A5 Central limit theorem

While the central limit theorem can be found in almost all textbooks on statistics, its proof is almost never given. The best reference in the accessible literature that includes a discussion of the validity limitations can be found in Cramér (1946).[1] Van Kampen (1981)[2] gives a more intuitive discussion. You will need to know what the characteristic function of a probability distribution is (see Appendix A3 on page 141). Cramér states the central limit theorem as follows:

> *Whatever be the distributions of the independent variables x_i – subject to certain very general conditions – the sum $x = x_i + \cdots + x_n$ is asymptotically normal (m, σ), where m is the sum of means and σ^2 is the sum of variances.*

"Asymptotically normal" means that the distribution of x tends to the normal distribution $N(m, \sigma)$ for large n. The "certain very general conditions" include the requirement that every contributing distribution has a finite variance; in addition the sum of third moments, divided by the 3/2 power of the total variance, must tend to zero for large n. The latter is of course always true for symmetric distributions, but it is also true for a sum of equivalent distributions. It is false only in pathological cases.

Consider a large number n of independent continuous random variables $x_1, x_2, \ldots, x_n$ with sum x:

$$x = \sum_{i=1}^{n} x_i, \tag{A5.1}$$

each with a probability density function $f_i(x)$. Let each pdf have a finite mean m_i and variance σ_i^2. We now ask the question what can we say about the probability density function $f(x)$, when n tends to infinity.

First eliminate the mean. Since

$$\sum_i (x_i - m_i) = \sum_i x_i - \sum_i m_i = x - m, \tag{A5.2}$$

[1] See reference list on page 123.
[2] See reference list on page 124.

the mean of x is the sum of the means m_i. So by considering $x_i - m_i$ instead of x_i, all contributing variables and the resulting sum have zero mean. Now consider the density function of the sum $f(x)$. This is a *convolution* of all f_i and hence the characteristic function $\Phi(t)$ of $f(x)$ is the *product* of the characteristic functions $\Phi_i(t)$ of $f_i(x_i)$:

$$\Phi(t) = \Pi_{i=1}^{n} \Phi_i(t), \tag{A5.3}$$

or

$$\ln \Phi(t) = \sum_{i=1}^{n} \ln \Phi_i(t), \tag{A5.4}$$

where

$$\Phi_i(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \mathrm{e}^{ixt} f_i(x)\, dx. \tag{A5.5}$$

We know that $\Phi(0) = 1$, but for $t \neq 0$ each $\Phi_i(t) < 1$ (because at $t = 0$ the first derivative is zero and the second derivative is negative) and hence the product $\Phi(t)$ tends to zero. So $\Phi_i(t)$ is a rapidly decaying function of t. How does it behave for small t?

Consider the expansion of $\ln \Phi_i(t)$ in powers of t, which follows from the expansion (A3.4) on page 141 of $\Phi_i(t)$:

$$\ln \Phi_i(t) = -\frac{1}{2}\sigma_i^2 t^2 - \frac{i}{6}\mu_{3i} t^3 + \frac{1}{24}(\mu_{4i} - 3\sigma_i^4)t^4 + \cdots \tag{A5.6}$$

From this we find, denoting $\sum_i \sigma_i^2$ by σ^2:

$$\ln \Phi(t) = -\frac{1}{2}\sigma^2 t^2 \left[1 + \frac{i}{3}\frac{\sum_i \mu_{3i}}{\sigma^3}\sigma t - \frac{1}{12}\left(\frac{\sum_i \mu_{4i}}{\sigma^4} - 3\frac{\sum_i \sigma_i^4}{\sigma^4}\right)\sigma^2 t^2 \dots\right] \tag{A5.7}$$

Since – under mild conditions – terms such as $\sum_i \sigma_i^2, \sum_i \mu_{3i}$, etc. scale proportional to n, the factor $\sum_i \mu_{3i}/\sigma^3$ in the second term scales as $n^{-1/2}$ and the factor in the third term scales as n^{-1}. So, for large n, $\ln \Phi(t)$ approaches $-\sigma^2 t^2/2$:

$$\lim_{n\to\infty} \Phi(t) = \mathrm{e}^{-(\sigma^2 t^2/2)}, \tag{A5.8}$$

which implies that the probability density function is normal:

$$\lim_{n\to\infty} f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{A5.9}$$

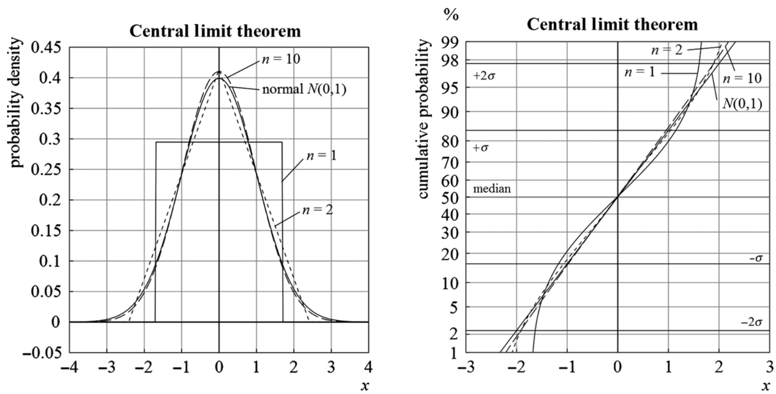

Figure A5.1 The probability distribution of the sum of n random numbers ($n = 1, 2, 10$), chosen from a uniform distribution between $-a$ and $+a$, compared with the normal distribution $N(0, 1)$. Each distribution has unit variance. Left: probability density, right: cumulative distribution on a probability scale.

Summarizing we can say that, with increasing n, the higher powers of t in $\ln \Phi(t)$ die out relative to the t^2-term; the higher the power, the faster it dies out. The most persistent is the third power (related to the skewness), which diminishes only slowly with the inverse square root of n.

An example, related to Exercise 4.10 on page 51, is the distribution function of the sum of n random numbers, sampled from a uniform distribution in the domain $[-a, a\rangle$. Figure A5.1 gives the distribution functions for $n = 1, 2, 10$, compared with the normal distribution $N(0, 1)$. In each case a has been chosen such that the distribution function of the resulting sum variable has a standard deviation of 1.

See **Python code** A5.1 on page 194 for a Python code to generate the distribution function for arbitrary n using Fourier transforms.

It is clear that the derivation fails completely when one or more of the contributing distributions has an undefined (infinite) variance, such as the Lorentz distribution (see page 43) has. In fact, the sum of Lorentz-distributed random variables remains Lorentz-distributed!