Introduction to Central Limit Theorem + α

stma

April 24, 2025

Contents

1 Introduction

There seemingly quite a few recognized applications found with Central Limit Theorem, as well being foundation of an experimental structure in science.

The intuition gives us the way of transition from "sample" to "population", where finiteness and practicality characterize the *sample* that falls in as the hands-on data of analysis at a laboratory; whereas infinite (in the sense of obscurity), theoretical features are the key for *population*.

Although what sample represents vary in the experimental setting, the basic idea claims that **sufficiently large** number of samples are averaged to a law of so called *normal distribution*, as in the figure 1, where the averaged value $\frac{\sum_i^N X_i}{N}$ of N dices have "regulated" probability that takes real numbers between 1 and 6 (i.e. called *probability density*), **no matter what probability law ordered the distribution of original value X_i to follow** (i.e. for dice roll, we assume each X_i obeys the discrete uniform distribution).

Note. While in theory, the distribution of averaged N-dices must be centred at zero by transition to attain genuine normal distribution, we can still see some normalizing phenomena (around the mean) without centralization as in figure 2. In practice, the actual distribution may turn out to be ambiguous, depending on the randomized process.

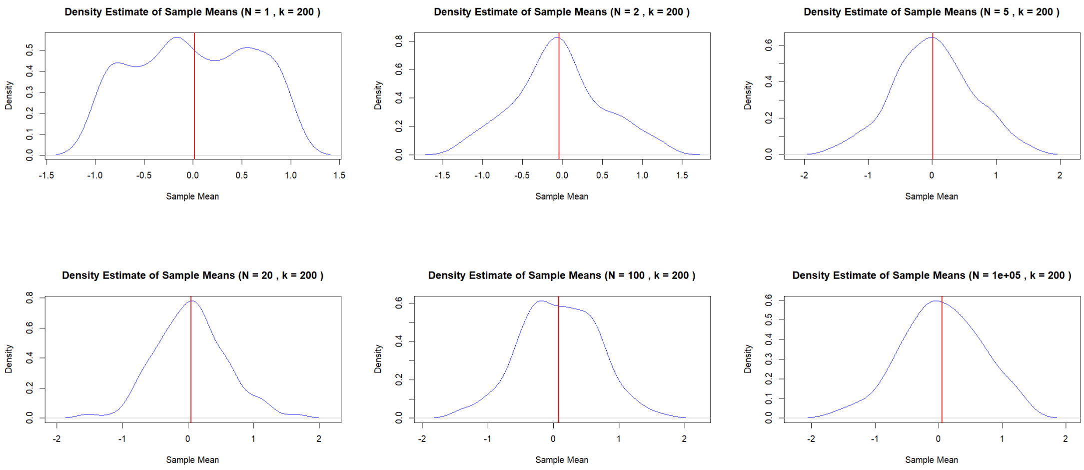


Figure 1: The mean value of N dices tends to normal distribution after translation $\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}$

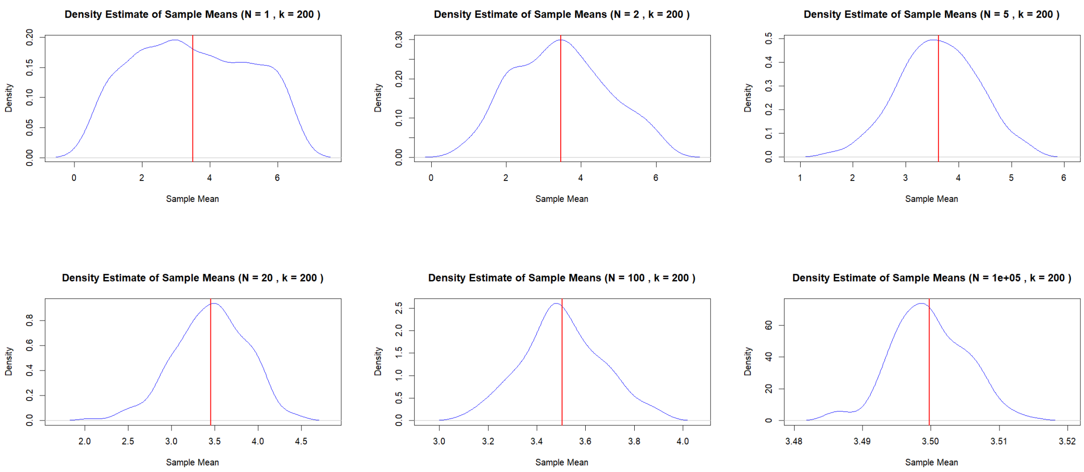


Figure 2: The mean value of N dices tends to non-normal distribution without zero-centred assumption

2 Definitions

2.1 Triangular Array

Definition. (triangular array) For each $n \in \mathbb{N}$, a sequence $X_{n1}, \ldots, X_{nr_n}$ of independent random variables is called *triangular array* of random variables, when they have distinct probability space for the sequence depending on n.

Note. (mean finiteness implication) In convention, a random variable X with finite variance admits finite mean, namely

$$\mathrm{Var}[X] < \infty \implies \mathrm{E}[X] < \infty.$$

In a probability space $(\Omega, \mathcal{F}, P)$, a random variable X is defined as a $\mathcal{F}$-measurable real-valued function; accordingly, the expected value is defined as

$$E[X] = \int_\Omega X dP = \int_{\omega \in \Omega} X(\omega) P(d\omega).$$

Therefore, when we say X has finite variance, it is determined as a finite value

$$\mathrm{Var}[X] = E[(X - E[X])^2] = \int_\Omega (X - \mu)^2 dP = \int_{\omega \in \Omega} (X(\omega) - \mu)^2 P(d\omega),$$

provided that $E[X] = \mu$ is finite; otherwise we have

$$\mathrm{Var}[X] = E[X^2] - \mu^2 = E[X^2] - \infty < \infty,$$

which contradicts a widely recognized convention that $(\infty - \infty)$ is undefined.

Note that this is not a matter of triviality, but is thought of a foundational convention in the domain (c.f. divided by zero convention).

2.2 Independent and identically distributed Random Variables

Definition. (i.i.d. random variables) A sequence of random variables are said to be *independently distributed* if the corresponding distributions do not depend on each other. They are said *identically distributed* if the distributions are identical.

2.3 Convergence Concepts

There are three of widely used convergence concepts within the statistical context: *almost surely* (a.k.a. strong), *in probability* and *in distribution* (a.k.a. weak), from stronger to weaker (i.e. the former implies later, accordingly).

Definition. (convergence in distribution) A sequence of random variables $X_1, \ldots$ *converges in distribution* (*converges weakly*) to a random variable X if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

at all continuous points of $F_X(x)$. We denote the convergence by $F_{X_n} \Rightarrow F_X$, or even $X_n \Rightarrow X$ if there is no confusion.

This is a special case of weak convergence of measures, where F_{X_n}, F_X account for the (probability) measures.

Example. Let $U_1, U_2, \ldots, U_n$ be a series of i.i.d. random variables with standard uniform distribution (i.e. $U_i \sim \mathrm{unif}(0,1)$).

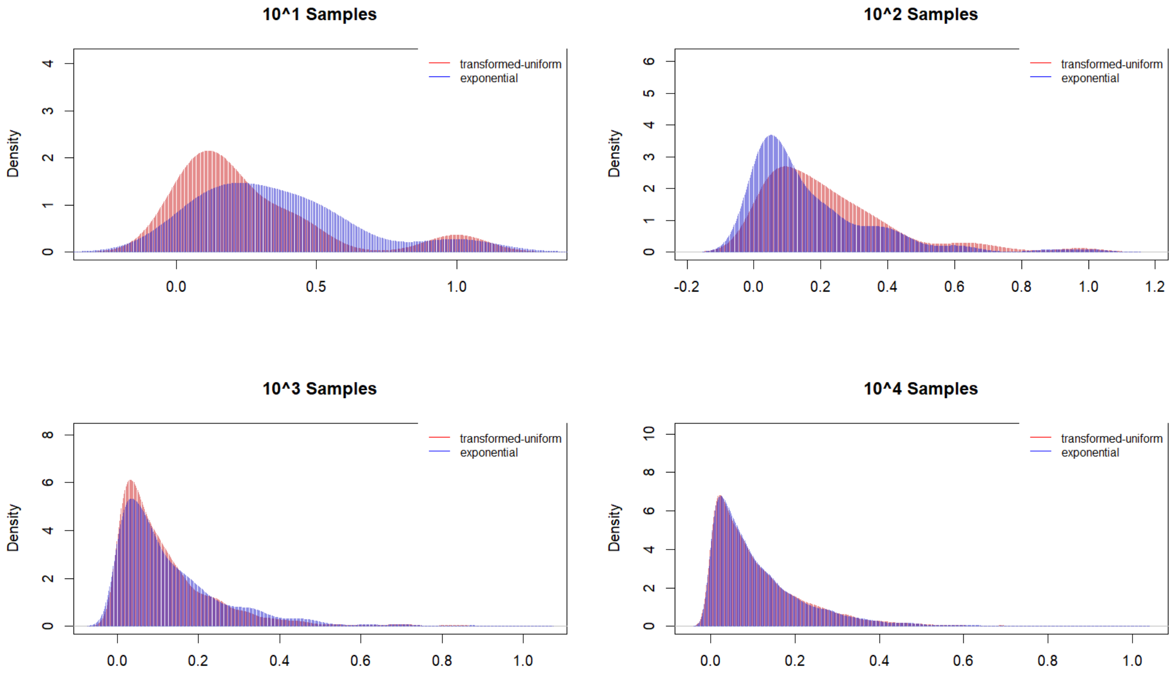
Define

$$\hat{Z}_{n,t} = \frac{1}{n} \sum_{i=1}^{n} I_{Y_i \le t}, \quad t \ge 0,$$

where $\lambda > 0$ fixed and $Y_i = -\lambda \log(1 - U_i)$.

Then we see that $\lim_{n \to \infty} \hat{Z}_{n,t} \sim \exp(\lambda)$ as in the figure 3 shows (this actually converges almost surely, and hence converges in probability).

Figure 3: A process of fitting empirical density to exp(λ)



Definition. (convergence in probability) A sequence of random variables, $X_1, X_2, \ldots$ *converges in probability* to a random variable X if, for all $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0 = 1 - \lim_{n\to\infty} P(|X_n - X| < \epsilon)$$

Definition. (almost sure convergence) A sequence of random variables, $X_1, X_2, \ldots$, *converges almost surely* to a random variable X if, for all $\epsilon > 0$,

$$P(\lim_{n\to\infty} |X_n - X| < \epsilon) = 1 = 1 - P(\lim_{n\to\infty} |X_n - X| \geq \epsilon)$$

Example. (converge in probability, not almost surely) For each $1 \leq m \in \mathbb{N}$, we can uniquely assign an indicator function I_m defined by

$$I_m = I_{[\frac{m \mod \tau(n_m)}{n_m}, \frac{1+(m \mod \tau(n_m))}{n_m}]}$$

such that $\tau(n_m) \leq m < \tau(n_m + 1)$ holds, where $\tau(n) = (n^2 - n + 2)/2$.

For example, $I_1 = I_{[0,1]}$, $I_2 = I_{[0,1/2]}$, $I_3 = I_{[1/2,1]}$, $I_4 = I_{[0,1/3]}$, and so on.

Because n_m can be defined by $n_m = \text{argmax}_{n \in L_m} \tau(n)$ with $L_m = \{n \in \mathbb{N} \mid \tau(n) \leq m\}$, the bigger m becomes the smaller the width $1/n_m$ of interval on which I_m takes non-zero value becomes.

Hence the sequence of random variables $X_m(s) = s + I_m(s)$ converges to $X(s) = s$ in probability, but not almost surely (i.e. we can always find some m' such that $I_{m'}(s) = 1$, periodically).

If we take the subsequence of random variables, say $X'_m(s) = s + I_{\tau(m)}(s)$, this indeed converges almost surely to X.

2.4 Stable distribution

The concept of stability gives a notion of algebraic operations on random variables closed in a probabilistic model.

Definition. (degenerate distribution) A distribution is *degenerate* if the support is singleton.

A *non-degenerate distribution* is that of not degenerate.

Table 1: required conditions for each version

conditions							
	Independence	Identical distribution	mean finiteness	variance finiteness	Lindeberg condition	Generalized condition
Classical setting	o	o	canonical	o	canonical	-
Lindeberg version	o	x	canonical	o	o	-
Generalized version	o	o	o	x	-	o

Note. Since we are concerned only with univariate distribution, this definition is sufficient.

Definition. (stable distribution, [**VI-1**, **Definition 1**, 2]) Let $X, X_1, X_2, \ldots$ be i.i.d. random variables with a common distribution F and $S_n = X_1 + \ldots + X_n$.

The distribution F is *stable* if F is non-degenerate, and for all $n \in \mathbb{N}$, there exists $c_n, \gamma_n \in \mathbb{R}$ with $c_n > 0$ such that

$$S_n \stackrel{d}{=} c_n X + \gamma_n,$$

where $\stackrel{d}{=}$ means the equivalence of the distribution.

The distribution is said to be *strictly stable* if this holds with $\gamma_n = 0$.

Note. How pairs (c_n, γ_n) are assigned is worth to pay attention — it says that the pair exists for each n (and the corresponding S_n, not individual X_i).

Theorem 1. ([**VI-1**, **Theorem 1**, 2]) Let F be a stable distribution. The norming constants are of the form $c_n = n^{1/\alpha}$ with $0 < \alpha \leq 2$. The constant α will be called *the characteristic exponent of F*.

One can call a stable distribution of the characteristic exponent α by *α-stable distribution* for short.

Proof. We omit the proof. Follow the reference. □

2.5 Domain of attraction

Definition. Let $X_1, X_2, \ldots$ are sequence of i.i.d. random variables with common distribution F.

The distribution F belongs to *the domain of attraction* of non-degenerate distribution G if there exist constants $a_n > 0$ and b_n such that the distribution of $a_n^{-1} \sum_j^n X_j - b_n$ tends to G; in other words

$$a_n^{-1} \sum_j^n X_j - b_n \Rightarrow G$$

holds.

3 The Weak Law of Large Numbers (WLLN)

Theorem 2. Let $X_1, X_2, \ldots$ be i.i.d. random variables with $EX_i = \mu$ and $\mathrm{Var}X_i = \sigma^2 < \infty$, then for all $\epsilon > 0$, sample mean converges to population mean in probability, namely:

$$\lim_{n \to \infty} P(|\overline{X_n} - \mu| < \epsilon) = 1.$$

4 Conditions for the CLT

In classical setting, *Lindeberg condition* is derived from an identically distributed sequence (i.e. not necessarily triangular array) of random variables, where we assume

$$X_{nk} = X_k,\ r_n = n,\ 1 \leq k \leq n.$$

Then the condition is reduced to

$$\forall \epsilon > 0, \quad \sum_{k=1}^{n} \frac{1}{\sigma^2} \int_{|X_k| \geq \epsilon \sqrt{n} \sigma} X_k^2 \, dP \leq \int_{|X_k| \geq \epsilon \sqrt{n} \sigma} dP \xrightarrow{n \to \infty} 0,$$

which is canonical because X_k is constant against n.

In contrast, when we relax the previous condition by setting $r_n = n$, $1 \leq k \leq n$ as X_{nk} may belong to its own probability space, being understood that X_{nk} are i.i.d. over $k \leq n$ for each n, we see that *Lindeberg* is not canonical anymore.

In this case, the Lindeberg condition is expressed as

$$\forall \epsilon > 0, \quad \int_{|X_{nk}| \geq \epsilon \sqrt{n} \sigma_n} dp \xrightarrow{n \to \infty} 0,$$

or more concisely $|X_{nk}| = o(\sqrt{n}\sigma_n)$ in Landau's notation, implying that the ratio $|X_{nk}|/\sigma_n$ is $o(\sqrt{n})$.

This is the case when for sufficiently large n, any individual contribution of X_{nk} to the variance σ_n are uniformly suppressed.

For classical and *Lindeberg*, the following condition must be met in common.

$$E[X_{nk}] = 0, \quad \sigma_{nk}^2 = E[X_{nk}^2], \quad s_n^2 = \sum_{k=1}^{r_n} \sigma_{nk}^2 \tag{1}$$

For a concise overview of conditions, see table 1.

4.1 Independence and identity of distributions

Only under Lindeberg assumption, do the random variables require to be identically distributed; namely, it accepts the case when $F_{X_{nk}} \neq F_{X_{n'k'}}$, for some n, n', k, k'.

4.2 Finite Variance

Only the Generalized condition tolerate infinite variance.

4.3 Lindeberg Condition

Definition. (Lindeberg condition) *Lindeberg* version of CLT requires that the following equation holds in addition to the condition (1).

$$\lim_{n \to \infty} \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{|X_{nk}| \geq \epsilon s_n} X_{nk}^2 \, dP = 0$$

for any $\epsilon > 0$.

The Lindeberg condition is important since it ensures asymptotically tamed behaviour in a way that the sum of independent random variables doesn't "blow up", preventing the condition of CLT from violated.

4.4 Generalized Condition

Although Generalized Central Limit Theorem can be regarded as a variant of CLT, unlike Lindeberg's, it is not straightforward to interpret the GCLT in terms of "condition-statement" format where the statement requires standard normal distribution to be attained; that is where GCLT generalizes.

Rather, since normal distribution corresponds to a Lévy 2-stable, we offer the following as the condition.

Definition. (Generalized condition) Let $X_1, \ldots$ be i.i.d. sequence of random variables and Z be a non-degenerate random variable.

The GCLT requires Z to be α-stable for some $0 < \alpha \leq 2$.

5 The Central Limit Theorem

5.1 Statement of the Theorem

Theorem 3. (classical CLT) Suppose that $\{X_n\}$ is an independent sequence of random variables having the same distribution with mean c and finite positive variance σ^2. If $S_n = X_1 + \cdots + X_n$, then

$$\frac{S_n - nc}{\sigma\sqrt{n}} \Rightarrow N(0, 1).$$

where $\Rightarrow$ denotes the convergence in distribution.

Theorem 4. (Lindeberg CLT,[**Theorem 27.2**, 1])
Suppose that for each n the sequence $X_{n1}, \ldots, X_{nr_n}$ is independent and satisfies 1. If 4.3 holds for all positive ϵ, then

$$S_n/s_n \Rightarrow N(0, 1).$$

Theorem 5. (Generalized CLT, [4]) Let $X_1, \ldots$ be i.i.d. sequence of random variables and Z be a non-degenerate random variable. For some constants $a_n > 0$ and $b_n \in \mathbb{R}$, $a_n \sum_i^n x_i - b_n$ converges in distribution to Z if and only if Z is α-stable for some $0 < \alpha \leq 2$.

5.2 Relationship to WLLN

Theorem 6. (mapping theorem, [**Theorem 25.7**, 1]) Suppose that $h : \mathbb{R} \to \mathbb{R}$ is measurable and that the set D_n of its discontinuities is measurable. If $\mu_n \Rightarrow \mu$ and $\mu(D_n) = 0$, then $\mu_n h^{-1} \Rightarrow \mu h^{-1}$.

Corollary 1. CLT implies WLLN.

Proof. Assuming classical CLT, theorem 6 implies that $\sqrt{n}(\frac{S_n}{n} - c) \Rightarrow \mathcal{N}(0, \sigma)$. For an arbitrary $\epsilon > 0$, take a natural number $N \in \mathbb{N}$ such that

$$\forall t \in \mathbb{R}, \forall n \geq N, \left|F_{\sqrt{n}(\overline{X_n} - c)} - F_Z(t)\right| < \frac{1}{2n}.$$

For some random variable $Z \sim \mathcal{N}(0, 1)$, it follows that

$$\begin{aligned} P(|\overline{X_n} - c| \geq \epsilon) &= P(\sqrt{n}|\overline{X_n} - c| \geq \sqrt{n}\epsilon) \\ &= P(\sqrt{n}(\overline{X_n} - c) \geq \sqrt{n}\epsilon) + P(\sqrt{n}(\overline{X_n} - c) \leq -\sqrt{n}\epsilon) \\ &\leq 1 - (F_Z(\sqrt{n}\epsilon) - \frac{1}{2n}) + F_Z(-\sqrt{n}\epsilon) + \frac{1}{2n} \\ &= (1 - F_Z(\sqrt{n}\epsilon)) + \frac{1}{n} + F_Z(-\sqrt{n}\epsilon), \end{aligned} \tag{2}$$

where all the terms vanish in the last equation 2, when $n \to \infty$. □

References

[1] Patrick Billingsley. *Probability and measure.* 3. ed. A Wiley-Interscience publication. New York [u.a.]: Wiley, 1995. XII, 593. ISBN: 0471007102.

[2] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. New York: John Wiley & Sons Inc., 1971, pp. xxiv+669.

[3] Takahiro Hasebe and Alexey Kuznetsov. "On free stable distributions". In: *Electronic Communications in Probability* 19.none (Jan. 2014). ISSN: 1083-589X. DOI: 10.1214/ecp.v19-3443. URL: http://dx.doi.org/10.1214/ECP.v19-3443.

[4] J.P. Nolan. *Univariate Stable Distributions: Models for Heavy Tailed Data*. Springer Series in Operations Research and Financial Engineering. Springer International Publishing, 2020. ISBN: 9783030529154.