IS NORMAL DISTRIBUTION NORMAL?

Bofan Chen
Department of pure mathematics and mathematical statistics
University of Cambridge
Cambridge, UK
cbfcbf.byron@gmail.com

ABSTRACT

Normal distribution is known as the most common distribution in nature. Here, we explore two possible explanations for the commonness of normal distribution. We also provide some critiques for each explanation and show that the commonness of normal distribution still cannot be fully justified and needs further research.

***K*eywords** Normal distribution · Central limit theorem · Entropy

1 Introduction

Scientists always use normal distribution to model certain physical quantities, such as weights of bread, scores of students, heights of people, measure errors. This is because we always naturally consider that among all the distributions, normal distribution is the most special. We can observe it in many natural phenomena. But why does god choose this particular distribution to be special? Is the normal distribution really normal?

Over the decades, many scientists has explored the particularity of normal distribution. Its particularity first appears at the central limit theorem [1], which says that the sum of random variables finally converges to the normal distribution in a distribution sense. If we divide our interest quantities into many small part random variables, we can obtain the normal distribution. Besides, researchers also prove that the normal distribution arrives at the maximal entropy under the constraint of certain variance, so the normal distribution is consistant with the second law of thermodynamics.[2]

However, this problem is actually still a mystery. So far, no research has fully supported and proved the commonness of normal distribution. In this paper, we give comments and critiques for previous explanations. We also show that the normal distribution is not that normal from a philosophical perspective. It is the way of looking at the world that truly influences the particularity of distribution.

2 Explanation given by central limit theorem

One of the most popular explanation of the commonness of normal distribution is given by central limit theorem (CLT). In this section, we discuss certain versions of CLT, and show how reliable it is to support the commonness of normal distribution. We give the critique of this explanation at the end of this section.

2.1 CLT(Lindeberg-Levy)

CLT(Lindeberg-Levy) is the most common version of CLT. It shows that the sum of a sequence of i.i.d (Independent and Identically Distributed) random variables will finally converge to a normal distribution with the number of participants increase.

Theorem 1 (Lindeberg-Lévy CLT [1])**.** *Suppose* $\{X_1, \ldots, X_n\}$ *is a sequence of i.i.d. random variables with* $\mathrm{E}[X_i] = \mu$ *and* $\operatorname{Var}[X_i] = \sigma^2 < \infty$*. Then, as* n *approaches infinity, the random variables* $\sqrt{n}\left(\bar{X}_n - \mu\right)$ *converge in distribution to a normal* $\mathcal{N}\left(0, \sigma^2\right)$:

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right).$$

Based on this classical version of CLT, one may explain the commonness of normal distribution as follows: The majority of natural physical quantities consist of many small parts of components which can be seen as random variables. Thus the macrovariable X we observe is the sum of these random variables X_i, *i.e.*

$$X = \sum_{i=1}^{\infty} X_i.$$

The CLT tells us it should follow normal distribution.[3] For example, the weights of bread consist of the weights of flour, milk, and other ingredients; the heights of students consist of the contributions of different genes.

At first glance, that explanation makes sense. However, when we carefully check the condition of classical version of CLT, we find that explanation unreasonable. The condition of Lindeberg-Lévy CLT requires that the sequence of random variables should be independent and identically distributed, but it seems that the elements that make up macrovariables are not identically distributed in most cases. The weights of ingredients to make bread are totally different. Moreover, there are still certain cases that the elements are not independent as well. The contribution of each gene to one's height is extremely complex, because they may influence each other and be dependent .

For the condition of identical distribution, we can use certain methods to make this condition weak. Since we observe that

$$X - E[X] = \sum_{n=1}^{\infty} (X_i - E[X_i]),$$

we do not require each element have the same expectation. Besides, we do not require each element have the same variance, which will be further discussed detailly in the next subsection.

2.2 Other CLT versions

There are other CLT versions which makes the condition weaker that do not require the identical distribution of elements in random variable sequence. One of the important CLT versions is proposed by Lyapunov as follows.

Theorem 2 (Lyapunov CLT [1])**.** *Suppose $\{X_1, \ldots, X_n, \ldots\}$ is a sequence of independent random variables, each with finite expected value μ and variance σ_i^2. Define*

$$s_n^2 = \sum_{i=1}^{n} \sigma_i^2$$

If for some $\delta > 0$, Lyapunov's condition

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathrm{E}\left[|X_i - \mu_i|^{2+\delta}\right] = 0$$

is satisfied, then a sum of $\frac{X_i - \mu_i}{s_n}$ converges in distribution to a standard normal random variable, as n goes to infinity:

$$\frac{1}{s_n} \sum_{i=1}^{n} (X_i - \mu_i) \xrightarrow{d} \mathcal{N}(0, 1)$$

We can observe that this theorem replaces the identical condition by Lyapunov's condition in CLT. Besides, there are other similar conditions discovered as well, like Lindeberg's condition.

Theorem 3 (Lindeberg's condition [1])**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $X_k : \Omega \to \mathbb{R}, k \in \mathbb{N}$, be independent random variables defined on that space. Assume the expected values $\mathbb{E}[X_k] = \mu_k$ and variances $\operatorname{Var}[X_k] = \sigma_k^2$ exist and are finite. Also let $s_n^2 := \sum_{k=1}^{n} \sigma_k^2$. If this sequence of independent random variables X_k satisfies Lindeberg's condition:*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}\left[(X_k - \mu_k)^2 \cdot \mathbf{1}_{\{|X_k - \mu_k| > \varepsilon s_n\}}\right] = 0$$

for all $\varepsilon > 0$, where $1_{\{\ldots\}}$ is the indicator function, then the central limit theorem holds, i.e. the random variables

$$Z_n := \frac{\sum_{k=1}^{n} (X_k - \mu_k)}{s_n}$$

converge in distribution to a standard normal random variable as $n \to \infty$.

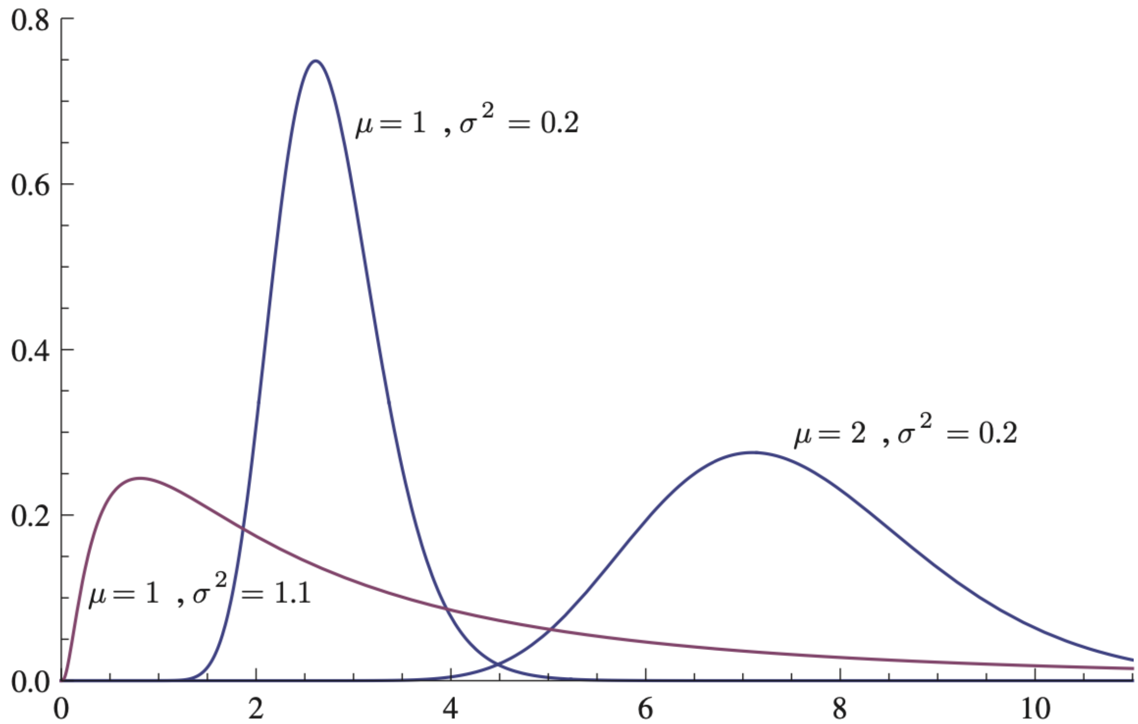

Figure 1: log-normal distribution with different parameters.

It can be proved that Lindeberg's condition is sufficient, but not necessary. However, if the sequence of independent random variables in question satisfies

$$\max_{k=1,\ldots,n} \frac{\sigma_k^2}{s_n^2} \to 0, \quad \text{as } n \to \infty,$$

then Lindeberg's condition can be both sufficient and necessary, i.e. it holds if and only if the result of central limit theorem holds.

Based on the previous two similar conditions, intuitively, it means that the central limit theorem requires the variance of every element in the random variable sequence should be small enough comparing to the sum of all elements' variances. Considering the example of bread, we can now apply these new CLT versions to explain the weights of bread with no identical distribution conditions for the weights of ingredients.

2.3 Critique of CLT explanation

In fact, previous explanation of commonness of normal distribution is still far from convincing. This is because not all natural quantities which follows normal distribution can be seen as the sum of lots of random variables. For instance, the rate of return in the stock market cannot be divided into many small parts. Conversely, it is more like a multiplicity of factors, such as policy factors and emotional factors, rather than the sum:

$$Y = \prod_{i=1}^{\infty} Y_i$$

How to explain this? Actually, based on CLT, it should follow log-normal distribution. Since sometimes $\log Y$ can be expressed as the sum of independent random variables $\log Y_i$

$$\log Y = \sum_{i=1}^{\infty} \log Y_i.$$

The interesting thing is log-normal distribution is quite similar to normal distribution under certain conditions.[2] Figure 1 shows the log-normal distribution with different parameters from which we can observe that when $\mu \gg \sigma^2$ log-normal distribution is quite like normal distribution. Additionally, in empirical datasets, so far there have been also no examples which fit normal distribution but does not fit log-normal distribution.[4] In this case, log-normal distribution is also common, because we cannot distinguish it from normal distribution in the empirical datasets. And to some extent, we should think that the distribution of heights or scores should be more like log-normal as these quantities cannot be negative which is against the assumption of normal distribution.

Besides, as previous mentioned at the beginning of this section, there are also lots of natural quantities not satisfying the condition of independence in CLT. Many quantities in the nature are like network. Like dominoes, a slight move in one part may affect the situation as a whole. Therefore, it is unreasonable to put forward such a strong assumption.

In addition to the independent condition and sum form mentioned before, there are still other problems raised from CLT explanation. For example, some quantities in nature may not have finite variance. This again leads to invaildation of CLT. The theorem that should replace the CLT is generalized central limit theorem, which shows that the sum of power-law distribution will converge to a cluster of stable distribution rather than normal distribution.[5, 6]

3 Explanation given by maximal entropy

Besides the explanation given by CLT, scientists also proposed an other explanation based on maximal entropy. In this section, we explore the particularity of normal distribution from the perspective of entropy and also provide some opposite evidence.

3.1 Maximal entropy characteristic in normal distribution

Entropy is first proposed in thermodynamics, which is a quantity to measure the molecular disorder of a system. Based on the second law of thermodynamics, any system intend to develop toward the direction which has a higher entropy. By considering this, we can use the concept of entropy to explain the formation of normal distribution directly.

The definition of entropy for a continuous random variables can be expressed as follows.

Definition 1 (Entropy)**.** *If X is a continuous random variable with probability density $p(x)$, then the differential entropy of X is defined as*

$$H(X) = -\int_{-\infty}^{\infty} p(x) \log p(x) dx.$$

The quantity $p(x) \log p(x)$ is understood to be zero whenever $p(x) = 0$.

Based on this definition, we can further define the maximal entropy distribution under center constraints as follows.

Theorem 4 (Maximal entropy distribution [7])**.** *Suppose S is a closed subset of the real numbers $\mathbb{R}$ and we choose to specify n measurable functions $f_1, \cdots, f_n$ and n numbers $a_1, \ldots, a_n$. We consider the class C of all real-valued random variables which are supported on S (i.e. whose density function is zero outside of S) and which satisfy the n moment conditions:*

$$\mathbb{E}\left[f_j(X)\right] \geq a_j \quad \textit{for } j = 1, \ldots, n$$

If there is a member in C whose density function is positive everywhere in S, and if there exists a maximal entropy distribution for C, then its probability density $p(x)$ has the following form:

$$p(x) = \exp\left(\sum_{j=0}^{n} \lambda_j f_j(x)\right) \quad \textit{for all } x \in S$$

where we assume that $f_0(x) = 1$. The constant λ_0 and the n Lagrange multipliers $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ solve the constrained optimization problem with $a_0 = 1$ (this condition ensures that p integrates to unity):

$$\max_{\lambda_0; \boldsymbol{\lambda}} \left\{ \sum_{j=0}^{n} \lambda_j a_j - \int \exp\left(\sum_{j=0}^{n} \lambda_j f_j(x)\right) dx \right\} \quad \textit{subject to } : \boldsymbol{\lambda} \geq \mathbf{0}$$

Using the Karush-Kuhn-Tucker conditions, it can be shown that the optimization problem has a unique solution because the objective function in the optimization is concave in $\boldsymbol{\lambda}$.

Proof. In the case of equality constraints, this theorem is proved with the calculus of variations and Lagrange multipliers. The constraints can be written as

$$\int_{-\infty}^{\infty} f_j(x)p(x)dx = a_j$$

We consider the functional

$$J(p) = \int_{-\infty}^{\infty} p(x)\ln p(x)dx - \eta_0 \left(\int_{-\infty}^{\infty} p(x)dx - 1\right) - \sum_{j=1}^{n} \lambda_j \left(\int_{-\infty}^{\infty} f_j(x)p(x)dx - a_j\right)$$

where η_0 and $\lambda_j, j \geq 1$ are the Lagrange multipliers. The zeroth constraint ensures the second axiom of probability. The other constraints are that the measurements of the function are given constants up to order n. The entropy attains an extremum when the functional derivative is equal to zero:

$$\frac{\delta J}{\delta p}(p) = \ln p(x) + 1 - \eta_0 - \sum_{j=1}^{n} \lambda_j f_j(x) = 0$$

Therefore, the extremal entropy probability distribution in this case must be of the form ($\lambda_0 := \eta_0 - 1$),

$$p(x) = e^{-1+\eta_0} \cdot e^{\sum_{j=1}^{n} \lambda_j f_j(x)} = \exp\left(\sum_{j=0}^{n} \lambda_j f_j(x)\right)$$

remembering that $f_0(x) = 1$. It can be verified that this is the maximal solution by checking that the variation around this solution is always negative. □

Corollary 1. *For the specifical constraint* $\mathrm{Var}[X] \leq \sigma^2 < \infty$*, the maximal entropy distribution is normal distribution.*

Proof. Taking $f(x) = -x^2$ for this constraint, we can obtain $p(x) \propto e^{-x^2}$, which reaches normal distribution. □

We observe that the normal distribution is the maximal entropy distribution under the constraint of variance. Therefore, it seems that we can explain normal distribution as follows: God will control the variance and expectation of certain nature quantities through nature rules. Under these external constraints, normal distributon becomes the common distribution in nature. For example, those people whose heights are far from optimal human height will be gradually eliminated by natural selection. This provides a variance constraint for human heights distribution. With the time goes by, the entropy of height distribution will gradually converge to its maximum. Eventually, the height distribution follow the normal distribution.

3.2 Critique of maximal entropy explanation

This explanation seems to be more convincing than the CLT one. It does not requires complex conditions and the maximal entropy law also very natural. However, there are still two points that need more discussion.

The first point is why the variance constraint is special in nature. According to the definition of maximal entropy distribution, actually every distribution is the maximal entropy distribution under certain constraint. Because we cannot justify the particularity of variance constraint, the commonness of normal distribution can still be questioned.

The second point is about the way we measure the world. To some extend, the way we measure the world is random and baseless. For example, we alway use Decibel(dB), the logarithm of altitude, to measure the loudness of sound, and we also use Kelvin, that is proportional to the square of molecular speed, to measure the temperature. The ruler of measure seems baseless. When we change the ruler, the constraint will also change and normal distribution is no longer normal again. We cannot guarantee that we use the same ruler as god, so it is not proper to say normal distribution is normal.

4 Conclusion

In this paper, we have discussed two possible explanations for the commonness of normal distribution. For the explanation given by CLT, we show that the condition of independence and sum form cannot be easily satisfied in natural quantities. For the explanation given by maximal entropy, we point that the perticularity of variance constraint and the measure ruler of the god still need to be justified. Therefore, the commonness of normal distribution is still a mystery to humans and needs more further research to discuss.

References

[1] Patrick Billingsley. *Probability and Measure (3rd ed.)*. 1995.

[2] Aidan Lyon. Why are normal distributions normal? *The British Journal for the Philosophy of Science*, 2014.

[3] Erik Gregersen et al. *The Britannica guide to statistics and probability*. Britannica Educational Publishing, 2010.

[4] Eckhard Limpert, Werner A Stahel, and Markus Abbt. Log-normal distributions across the sciences. *BioScience*, 51(5):341–352, 2001.

[5] D. V. Lindley. Limit distributions for sums of independent random variables. *Journal of the Royal Statistical Society. Series A (General)*, 118(2):248–249, 1955.

[6] Masaru Shintani and Ken Umeno. Super generalized central limit theorem—limit distributions for sums of non-identical random variables with power laws—. *Journal of the Physical Society of Japan*, 87(4):043003, 2018.

[7] Zdravko I Botev and Dirk P Kroese. The generalized cross entropy method, with applications to probability density estimation. *Methodology and Computing in Applied Probability*, 13:1–27, 2011.