CENTRAL LIMIT THEOREM: HAVE A TASTE

XIAODONG WANG

Abstract. We formulate a proof of the Central Limit Theorem by moment generating functions. As an application, we discuss how to use the Central Limit Theorem to estimate the probability of an unknown distribution.

1. Introduction

According to the *Central Limit Theorem*, a large sample of independent random variables after normalization converge to the Gauss distribution. It is noteworthy that the distribution of such random variables may be arbitrary or even unknown. Nevertheless, we can estimate the distribution of the random variables using the theorem.

The history of the Central Limit Theorem originated in the 18th century. Abraham de Moivre, a French mathematician, discovered the initial version of binomial approximation. However, the theorem did not own a formal name until 1930, when Hungarian mathematician George Pólya named it the Central Limit Theorem. Regarding its name, some believe that *central* means very important as it was a central problem in probability for many decades, while others hold the view that *central* means fluctuations around the center.

This article aims to prove the following theorem.

Theorem 1.1 (Central Limit Theorem)**.** *Suppose* $\{X_1, X_2, ..., X_n, ...\}$ *are independent and identically distributed random variables with common mean* $\mathbb{E}(X_i) = \mu < +\infty (\forall i \in \mathbb{N})$ *and common variance* $Var(X_i) = \sigma^2 < +\infty (\forall i \in \mathbb{N})$.
Let

$$S_n = \sum_{i=1}^{n} X_i$$

and

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}. \tag{1.1}$$

Then for any $x \in \mathbb{R}, (-\infty < x < +\infty)$, *we have*

$$\lim_{n\to\infty} \mathbb{P}\{Z_n \le x\} = \Phi(x),$$

that is

$$\lim_{n\to\infty} \mathbb{P}\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\} = \Phi(x),$$

Date: November 2020.
2010 *Mathematics Subject Classification.* Primary.
Key words and phrases. Normal distribution, Probability, Statistical distributions, Statistics.

where $\Phi(x)$ *is the cumulative distribution function of the standard normal distribution, i.e.*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy.$$

In the next section, we will develop an arsenal of different tools to prove the Central Limit Theorem in Section 3. As a utilization, we shall, in Section 4, introduce the normal approximation method to estimate the probability of an unknown distribution.

2. Preliminary

Above all, we state the definition of moment generating function here.

Definition 2.1. Suppose X is a random variable, then for any real number t, e^{tX} is a new random variable as a function of X. As a result, $\mathbb{E}(e^{tX})$ is a real value and we obtain a real function as t varies.

Such function is called the moment generating function of the given random variable X, and denoted by $M_X(t)$.

As an exercise, let us apply the definition to solve the following problem, whose result will come in handy later.

Exercise 2.2. The probability density function of $N(0,1)$ is $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} (-\infty < x < +\infty)$. By definition, its moment generating function could be calculated:

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \int_{-\infty}^{+\infty} e^{tx} f(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2-2tx}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2-2tx+t^2}{2}} \cdot e^{\frac{t^2}{2}} dx \\ &= e^{\frac{t^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-t)^2}{2}} dx \\ &= e^{\frac{t^2}{2}} \cdot 1 = e^{\frac{t^2}{2}}. \end{aligned}$$

Furthermore, we can obtain the following corollaries from the definition. The proofs are straightforward and hence left to the reader.

Corollary 2.1. *For any random variable* X*, we have* $M_X(0) = 1$.

Corollary 2.2. *If* $Y = X_1 + X_2 + ... + X_n$ *and* $\{X_1, X_2, ..., X_n\}$ *are (mutually) independent, then*

$$M_Y(t) = M_{X_1}(t) M_{X_2}(t) ... M_{X_n}(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

The lemma below will play a significant role in our proof of the Central Limit Theorem, but its proof is omitted here. Interested readers may consult the book[6].

Lemma 2.3. *The moment generating function uniquely determines the distribution. In other words, there exists a one-to-one correspondence between the moment generating function and the distribution of a random variable.*

3. Proof of the main theorem

After we have built up a versatile arsenal of implements in Section 2, we can now prove Theorem 1.1.

Proof of Theorem 1.1. Since $\{X_1, X_2, ..., X_n, ...\}$ are independent and identically distributed random variables with common mean $\mathbb{E}(X_i) = \mu(\forall i \in \mathbb{N})$ and common variance $Var(X_i) = \sigma^2 < +\infty(\forall i \in \mathbb{N})$, we know the random variables $\{Y_1, Y_2, ..., Y_n, ...\}$ are also independent and identically distributed where

$$Y_i = \frac{X_i - \mu}{\sigma}(\forall i \in \mathbb{N}).$$

It follows that all the random variables $\{Y_i : i \in \mathbb{N}\}$ have the same moment generating function, denoted by $M_Y(t)$, say.

In addition, since $\mathbb{E}(X_i) = \mu < +\infty$ and $Var(X_i) = \sigma^2 < +\infty$ we could easily obtain that for any $i \in \mathbb{N}$

$$\mathbb{E}(Y_i) = 0, \tag{3.1}$$

and

$$Var(Y_i) = 1. \tag{3.2}$$

And we could expand the $M_Y(t)$ as

$$M_Y(t) = M_Y(0) + \frac{M'_Y(0)}{1!}t + \frac{M''_Y(0)}{2!}t^2 + o(t^2), \tag{3.3}$$

where $o(t^2)$ means

$$\lim_{t\to 0} \frac{o(t^2)}{t^2} = 0.$$

Note that $Var(Y_i) < \infty$ guarantees that the form (3.3) holds at least for a small interval $0 \leq t < \epsilon$.

Accommodate the fact that $\mathbb{E}(Y_i) = 0$ and $Var(Y_i) = 0$ by (3.1) and (3.2), which implies

(i) $M_Y(0) = 1$, due to Corollary 2.1;
(ii) $M'_Y(0) = \mathbb{E}(Y_i) = 0(\forall i \in \mathbb{N})$;
(iii) $M''_Y(0) = \mathbb{E}(Y_i^2) = Var(Y_i) + (\mathbb{E}(Y_i))^2 = 1(\forall i \in \mathbb{N})$.

Then we can reduce the Taylor expansion (3.3) into

$$M_Y(t) = 1 + \frac{1}{2}t^2 + o(t^2),$$

where

$$\lim_{t\to 0} \frac{o(t^2)}{t^2} = 0. \tag{3.4}$$

By (3.4), we may assume, without loss of generality, that

$$o(t^2) = t^3 \cdot R(t),$$

where $R(t)$ is a continuous, bounded function of t near $t = 0$. Therefore, the moment generating function of $\{Y_i : i \in \mathbb{N}\}$ i.e. $M_Y(t)$ can be written as

$$M_Y(t) = 1 + \frac{1}{2}t^2 + t^3 R(t). \tag{3.5}$$

Now by (1.1) we know that

$$\begin{aligned} Z_n &= \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \\ &= \frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i - \mu}{\sigma} \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i. \end{aligned}$$

Since $\{Y_1, Y_2, ..., Y_n\}$ are independent, the moment generating function of Z_n, denoted by $M_{Z_n}(t)$, is

$$M_{Z_n}(t) = \prod_{i=1}^n M_{\frac{1}{\sqrt{n}}Y_i}(t) = \left(M_Y(\frac{t}{\sqrt{n}})\right)^n.$$

Based on (3.5), we have $M_Y(t) = 1 + \frac{1}{2}t^2 + t^3 R(t)$ and hence

$$M_{Z_n}(t) = [1 + \frac{1}{2}(\frac{t}{\sqrt{n}})^2 + (\frac{t}{\sqrt{n}})^3 \cdot R(\frac{t}{\sqrt{n}})]^n. \tag{3.6}$$

At this point, we consider the term $(\frac{t}{\sqrt{n}})^3 \cdot R(\frac{t}{\sqrt{n}})$ in (3.6). For any fixed t, $\lim_{n\to\infty} t^3 \cdot R(\frac{t}{\sqrt{n}}) = t^3 \cdot R(0)$, which is bounded. But $\lim_{n\to\infty}(\frac{1}{\sqrt{n}})^3 = 0$ and thus when considering the large n, we know that for any fixed t, this term is very small comparing with $\frac{t^2}{n}$ for both n and t.

Hence, by (3.6), we actually have

$$\begin{aligned} M_{Z_n}(t) &\approx \left(1 + \frac{1}{2}(\frac{t}{\sqrt{n}})^2\right)^n \\ &= (1 + \frac{t^2}{2n})^n. \end{aligned}$$

Then we could write, by ignoring the small fluctuation, that for large n

$$M_{Z_n}(t) = (1 + \frac{t^2}{2n})^n.$$

Note that for any $x \in \mathbb{R}$, we know $\lim_{n\to\infty}(1 + \frac{x}{n})^n = e^x$ which by letting $x = \frac{t^2}{2}$ implies

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty}\left(1 + \frac{(\frac{t^2}{2})^2}{n}\right)^n = e^{\frac{t^2}{2}}.$$

In short,

$$\lim_{n\to\infty} M_{Z_n}(t) = e^{\frac{t^2}{2}}.$$

Since $e^{\frac{t^2}{2}}$ is the moment generating function of the standard normal distribution $N(0,1)$. According to Lemma 2.3, we conclude that as $n \to \infty$ the cumulative distribution function of Z_n tends to cumulative distribution of $N(0,1)$. Therefore, for $\forall x \in \mathbb{R}$, we have

$$\lim_{n\to\infty} \mathbb{P}\{Z_n \leq x\} = \Phi(x)$$

or

$$\lim_{n\to\infty} \mathbb{P}\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\} = \Phi(x)$$

by denoting the cumulative distribution of $N(0,1)$ as $\Phi(x)$.

The proof of the Central Limit Theorem is now complete. $\square$

4. Application

In statistical problems, it usually assumes that we have a large sample of random variables $\{X_1, X_2, ..., X_n\}$ which are independent and identically distributed.

Let

$$Y = X_1 + X_2 + ... + X_n. \tag{4.1}$$

Our interest is trying to find probabilities such as

$$\mathbb{P}\{a < Y \leq b\} \tag{4.2}$$

for constants $a < b$.

If we know the cumulative distribution function of Y, denoted by $F_Y(y)$, then, of course

$$\mathbb{P}\{a < Y \leq b\} = F_Y(b) - F_Y(a).$$

However, it is usually hard work to find the cumulative distribution function of Y. Another method to calculate the value in (4.2) is to use the Central Limit Theorem. This method is called the *normal approximation method*. When the n in (4.1) is large, the usual procedure is as follows:

Assume that the random sum Y in (4.1) satisfies $\mathbb{E}(Y) = \mu < \infty$ and $Var(Y) = \sigma^2 < \infty$, then

$$\begin{aligned}\mathbb{P}\{a < Y \leq b\} &= \mathbb{P}\{\frac{a-\mu}{\sigma} < \frac{Y-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\} \\ &\approx \mathbb{P}\{\frac{a-\mu}{\sigma} < Z \leq \frac{b-\mu}{\sigma}\}\end{aligned}$$

(where $Z \sim N(0,1)$)
i.e.

$$\mathbb{P}\{a < Y \leq b\} \approx \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma}). \tag{4.3}$$

when the n in (4.1) is large.

For how large of n, will the normal approximation method in (4.3) be appropriate? This will depend upon whether the probability density function of Y is symmetric or not. If the probability density function of Y is symmetric, (such as t-distribution), then $n \geq 30$ is enough; otherwise, we need larger n.

5. Conclusions

In conclusion, the Central Limit Theorem is extremely powerful in the world of mathematics owing to numerous applications in probability theory and statistics. We have stated the Central Limit Theorem, proved it by moment generating function, and finally demonstrated an application of the theorem in statistics. In the future, it would be compelling and provoking to study other applications of the Central Limit Theorem as well as its other properties, such as convergence rates.

6. Acknowledgments

This work was one of the assignments for the course *Mathematical Writing in English* in fall 2020. The author would like to express his special thanks of gratitude to Professor Xiaoming Wang's guidance and Professor Anyue Chen for his notes on *Probability Theory*. Besides, we thank all the graders or classmates for carefully checking the article and appreciate their helpful comments.

References

[1] Edwin Thompson Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.
[2] Kai Lai Chung. *A Course in Probability Theory*. Academic Press, 2000.
[3] Henk Tijms. *Understanding Probability: Chance Rules in Everyday Life*. Cambridge University Press, 2007.
[4] Sheldon Ross. *A First Course in Probability*. Pearson Education, Inc., 2010.
[5] Hans Fischer. *A History of the Central Limit Theorem*. Springer, 2011.
[6] William Feller. *An Introduction to Probability and Its Applications, Volume 2*. John Wiley & Sons, Inc., 1991.

Department of Mathematics, Southern University of Science and Technology, Shenzhen, Guangdong, China
Email address: 11610629@mail.sustech.edu.cn