ORIGINAL RESEARCH ARTICLE

The Application of Central Limit Theorem

Qianju Cheng,Zongming Yang,Xifan Lu

Mathematics Institute, Guilin University of Technology, Guangxi, China

ABSTRACT

The central limit theorem is a kind of important theorem in the probability theory that the distribution of the sequence of random variables is asymptotically normal. In this paper, we first give some common central limit theorems from the content of the central limit theorem, and then prove the central limit theorem in the supply of electricity, device prices, shopping malls management, cigarette manufacturing, social life, military problem and so on. Finally, the advantages and disadvantages of the central limit theorem are analyzed.

KEYWORDS: random variable; central limit theorem; normal distribution; probability theory; approximate calculation

1. Introduction

Probability theory discusses the distribution of random variables and the distribution of parts is asymptotically a class of theorems of normal distribution. Probability theory of the most important class a theorem, a wide range of practical application background. In nature and production, some phenomena are influenced by many independent random factors, and if the influence of each factor is very small, the overall effect can be seen as obeying the normal distribution. The central limit theorem proves this phenomenon mathematically. The earliest central limit theorem is to discuss the problem of the number of occurrences of the incident asymptomatic in the heavy Bernoulli experiment. Before and after 1716, Diomofer discussed the probabilities of each trial event in the Bernoulli trial, followed by the promotion and improvement of P.-S. Laplace and AM Yapunov Since P. Levi established the feature function theory systematically from 1919 to 1925, the study of the central limit theorem has developed rapidly, and has produced general limit theorem and local limit theorem. The limit theorem is an important part of probability theory and one of the cornerstones of mathematical statistics. Its theoretical results are perfect. For a long time, the probabilistic analysis method for the study of the limit theorem has influenced the development of probability theory. At the same time the new limit theory problem is also in reality continue to produce.

2. Common central limit theorem

2.1. The formulation of the central limit theorem

The limit distribution of the sum of the random variables is subject to the normal distribution under certain conditions. In the probability theory, it is called the central limit theorem. Specifically, the central limit theorem answers the limit distribution of the sum of the random variables under what conditions is a normal distribution.

Intuitively, if a random variable is determined by the sum of a large number (or even) of random factors, where the individual effects of each random factor are negligible and the effect of each factor is relatively uniform, then it follows a normal distribution, In many cases, a random variable can be expressed as the sum of a large number of independent random variables,

$$X = \xi_1 + \xi_2 + \cdots \xi_n$$

Here, each ξ_i of the effects of a random factor, if the above formula contains a sufficient number of random factors to determine the effect (n large), then $\sum_{i=1}^{n}\xi_i$ the distribution of the approximate distribution of the central limit theorem to explain what conditions. The sum of a large number of independent random variables approximates the normal distribution, that is, under what conditions, then $n \to +\infty$ the sum of the independent random variables is subject to the normal distribution.

2.2. Common central limit theorem

The central limit theorem has been very rich in its content, but the most common theorem is as follows

2.2.1 Di Moff - Laplace Theorem

Is the μ_n number of n occurrences of events in the Bernoulli test, and the probability of $p(0 \le p \le 1)$ occurrence in each test is $\varepsilon > 0$, for any, $\lim_{n\to+\infty} P\left(\left|\frac{\mu_n}{n} - p\right| < \varepsilon\right) = 1$

$\xi_i = \begin{cases}1, \\ 0,\end{cases}$ Proof of order $1 \le i \le n$

It is a separate random variable, and

$$E\xi_i = p, D\xi_i = p(1-p) = pq \; (i = 1, \cdots, n)$$

$$\frac{\mu_n}{n} - p = \frac{\mu_n - np}{n} = \frac{\sum_{i=1}^{n}\xi_i - E\left(\sum_{i=1}^{n}\xi_i\right)}{n}$$

By the Chebyshev inequality

$$P\left(\left|\frac{\mu_n}{n} - p\right| \ge \varepsilon\right) = \left(P\left|\sum_{i=1}^{n}\xi_i - E\left(\sum_{i=1}^{n}\xi_i\right)\right| \ge n\varepsilon\right) \le \frac{D\left(\sum_{i=1}^{n}\xi_i\right)}{n^2\varepsilon^2}$$

Also known by independence

$$D\left(\sum_{i=1}^{n}\xi_i\right) = \sum_{i=1}^{n} D\xi_i = npq$$

So there is

$$P\left(\left|\frac{\mu_n}{n} - p\right| \ge \varepsilon\right) \le \frac{npq}{n^2\varepsilon^2} = \frac{1}{n}\frac{pq}{\varepsilon^2} \to 0(n \to +\infty)$$

This also proves the theorem.

The theorem is the earliest central limit theorem. In 1733, Diemuofu proved the $p=\frac{1}{2}$ above theorem, and then Laplacus extended it to any positive number less than one. The theorem shows that the normal distribution is the limit distribution of the binomial distribution. When sufficient, we can use the conclusion of the theorem to calculate the probability of binomial distribution.

The theorem mainly applies two aspects

1 approximates the probability that a random variable obeying a binomial distribution takes a range

2 It is known that the probability of taking a random variable of a binomial distribution within a range is estimated by the range (or the maximum of the range).

2.2.2 Lyapunov Center Limit Theorem

Set $\xi_1,\xi_2,\cdots$ the sequence of independent random variables $E\xi_k=a_k, D\xi_k=\sigma_k^2\left(k=1,2,\cdots\right)$, '

$$B_n^2=\sum_{k=1}^{n}\sigma_k^2,$$

If $\delta>0$ there is, there is

$$\frac{1}{B_n^{2+\delta}}E\left|\xi_k-a_k\right|^{2+\delta}\to 0, n\to+\infty$$

Then there is x any

$$\lim_{n\to+\infty}P\left(\frac{1}{B_n}\sum_{k=1}^{n}\left(\xi_k-a_k\right)\le X\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$$

It is proved that the continuous random variable is a density function

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}^{\infty}\left(x-a_k\right)^2 p_k\left(x\right)d\left(x\right)$$

$$\le\frac{1}{B_n^2\left(\tau B_n\right)^{\delta}}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}^{\infty}\left|x-a_k\right|^{2+\delta}p_k\left(x\right)d\left(x\right)$$

$$\le\frac{1}{\tau^{\delta}}\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}\int_{-\infty}^{+\infty}\left|x-a_k\right|^{2+\delta}p_k\left(x\right)d\left(x\right)$$

$$=\frac{1}{\tau^{\delta}}\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}E\left|\xi_k-a_k\right|^{2+\delta}\to 0, n\to+\infty$$

Similarly, to verify the case of discrete, can prove this theorem.

This theorem was presented by Lyapunov in 1900. It is shown that in the theorem, the random variable, $Z_n = \frac{1}{B_n}\sum_{k=1}^{n}(\xi_k - a_k)$, when n is large, approximates obey the $N\left(\sum_{k=1}^{n} a_k, B_n^2\right)$ normal distribution regardless of the distribution of the random variables $\xi_k \ (k = 1, 2, \cdots, n)$, as long as the conditions of the theorem are satisfied as $\sum_{k=1}^{n} a_k$. To obey the normal distribution, which is why the normal random variable in the probability theory plays an important role in a basic reason.

In many cases, the random variables considered can be expressed as the sum of many independent random variables. For example, at any given time, the power consumption of a city is the sum of the power consumption of a large number of users; the measurement error of a physical experiment is synthesized by a number of unobservable and addable minor errors, which tend to obey Normal distribution.

2.3. The Similarities and Differences of Center Limit Theorem

The above three central theorem is to study the distribution functions of independent random variables. Under certain conditions, when they are sufficiently large, they are transformed into normal distributions, and their differences are only different. In addition to the central limit theorem, Chebyshev's law of large numbers can be used for approximate calculations.

When $E(X_i) = \mu, D(X_i) = \sigma^2 > 0$, there is Chebyshev tree theorem can be seen and $\varepsilon > 0$, hence

From the limit theorem of the Lindbergh-Levy center,

$$\lim_{n\to+\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}(X_i - \mu)\right| \le \varepsilon\right\} = \lim_{n\to+\infty} P\left\{\left|\frac{\sum_{k=1}^{n}(X_i - \mu)}{\sigma\sqrt{n}}\right| < \frac{\sqrt{n}\varepsilon}{\sigma}\right\} \approx \Phi\left|\frac{\sqrt{n}\varepsilon}{\sigma}\right| - \Phi\left|-\frac{\sqrt{n}\varepsilon}{\sigma}\right| = 2\Phi\left|\frac{\sqrt{n}\varepsilon}{\sigma}\right| - 1$$

Thus, under the conditions set, the central limit theorem is more accurate than the law of large numbers in the above approximation. Therefore, the central limit theorem has some advantages.

3. Application of central limit theorem

Through the study, it is found that the central limit theorem is significant and the application is quite extensive.

3.1. Application of central limit theorem in the supply of electricity

There are 200 lathes in a workshop, each lathe for a variety of reasons there parking, set the probability of starting each lathe 0.6, each lathe power consumption of 1 kilowatt, and set each lathe start or stop is independent of each other. Seeking at least how much power the plant should supply to 99.9% of the probability of ensuring that the workshop will not be affected by insufficient power supply?

Solution ξ for a certain time to start the number of lathes $\xi \sim B(200, 0.6)$ $np = 120$, $np(1-p) = 48$.

From this limits of lathes,

$$P\{0 \le \xi \le N\} \approx \Phi\left(\frac{N-120}{\sqrt{48}}\right) - \Phi\left(\frac{-120}{\sqrt{48}}\right) \approx \Phi\left(\frac{N-120}{\sqrt{48}}\right) \ge 0.999$$

Referring to the table,

$$\frac{N-120}{\sqrt{48}}=3.1, N=141.$$

Therefore, at least 141 kilowatts of electricity in this workshop should be supplied with probability of 99.9% to ensure that the plant will not be affected by insufficient power supply.

3.2. Application of central limit theorem in device price

The average service life of a device (in hours) is a normal distribution with an average service life of 20 hours. The specific use is to replace another new device as soon as a device is damaged. As a result, it is known that each device Purchase price for. How much budget should be made for this device in the annual plan? Only 95% of the money is available for one year (assuming 2000 working hours a year)?

Solve the first device life, due to obey the parameters for the exponential distribution $E\left(X_k\right)=\frac{1}{\lambda}$, $\lambda=0.05$, and

$$D\left(X_k\right)=\frac{1}{\lambda^2}=400$$

Assuming that at least one year to prepare spare parts to have 95% mutual opposition is $k=1,2,3,\cdots,n$, remember $X_{k1},X_{k2},\cdots,X_{kn}$

On the Limit Theorem of Lyapunov Center

$$P\left(Y_n \geq 2000\right)=0.95,$$

Which is

$$0.05=P\left\{Y_n<2000\right\}=P\left\{\frac{Y_n-20n}{20\sqrt{n}}<\frac{2000-20n}{20\sqrt{n}}\right\}\approx\Phi\left(\frac{2000-20n}{20\sqrt{n}}\right)$$

Also,

$$\Phi\left(\frac{2000-20n}{20\sqrt{n}}\right)=\Phi\left(\frac{n-100}{\sqrt{n}}\right)\approx 0.95$$

Referring at the table,

$$\frac{n-100}{\sqrt{n}}=1.64, n\approx 118$$

Therefore, in the annual plan should be done during this period to do the budget of 118 yuan may have 95% of the grasp of a year enough.

3.3. Application of central limit theorem in shopping mall management

In the practical problem, if the sample study is a large sample problem, then we can approximate the calculation by the central theorem, the general parameters of the estimation and estimation of shopping malls in the order of goods and sampling test two aspects of the use of To the central limit theorem.

(1) Order of goods

A shop is responsible for the supply of goods. A certain product in a period of time each person needs to use a probability of 0.6. Assuming that at this time each person to buy or not independent of each other, ask the store how many pieces of this product can be 0.997 probability to ensure that no out of stock?

The sequence of random variables is independent of each other. Set the store should be prepared for such goods, then

$$X = \sum_{i=1}^{1000} X_i$$

Obey the parameters $n = 1000, p = 0.6$ of the binomial distribution, according to the meaning of , so $p\{\xi_k = 1\} = 0.6, p\{\xi_k = 0\} = 0.4$ its mathematical expectations and standard variance will be

$$E(X) = np = 1000 \times 0.6, D(X) = \sqrt{np(1-p)} = \sqrt{1000 \times 0.6 \times 0.4} = 4\sqrt{15}$$

By the central limit theorem,

$$P\left\{\sum_{i=1}^{1000} \xi_k \le m\right\} \approx \Phi\left(\frac{m-600}{4\sqrt{15}}\right) \ge 0.997$$

Check the standard normal distribution

$$\frac{m-600}{4\sqrt{15}} \approx 2.75$$

Therefore $m = 643$

Therefore, the store should be prepared at least 643 of this product to 99.7% probability to ensure that no out of stock.

(2) Sampling inspection problems

If the number of defective products is found to be more than 10, the product will be rejected and 10% of the product will be asked. If the number of defects is at least less, Probability to 90%?

Because of the mathematical expectation and variance of random variables

$$E\xi_i = 10, D\xi_i = p(1-p) = 0.1 \times 0.9 = 0.09$$

So its mathematical expectations and variance

$$E(\xi) = n\xi_i = 0.1n; D(\xi) = 0.9n$$

By the central limit theorem

$$P\{10 \le \xi \le n\} \approx \Phi\left(3\sqrt{n}\right) - \Phi\left(\frac{100-n}{3\sqrt{n}}\right)$$

Due to $\Phi\left(3\sqrt{n}\right)\approx 1$,

Which is

$$P\{10\le\xi\le n\}\approx 1-\Phi\left(\frac{n-100}{3\sqrt{n}}\right)\ge 0.9$$

Referring the table , solutions have to be $\frac{n-100}{3\sqrt{n}}\ge 1.28\ n\ge 147$

So at least 147 products should be tested in order to ensure that the probability of rejection of product to 0.9.

In this paper, we mainly study the order of ordering, sampling and testing of commercial items in shopping malls, turning them into mathematical problems, establishing mathematical models, and then using the central limit theorem to infer. Finally, we find the best solution for these practical problems. The decision-making in business management provides the theoretical basis for the test.

3.4. In the cigarette manufacturing industry in the application

From the production line randomly removed cigarettes $X_1, X_2 \ldots, X_n$, their quality characteristics of the suction resistance is a separate random variable $N\left(\mu,\sigma^2\right)$, similar to the normal distribution $\bar{X}=\frac{X_1+X_2+\ldots+X_n}{n}$, the sample mean similar to the normal distribution $N\left(\mu,\frac{\sigma^2}{n}\right)$

Case of cigarettes weight, circumference and suction resistance is an important indicator of the quality of cigarettes logistics indicators, but also an assessment of the industry an essential indicator. Weight and cigarette smoke resistance, hardness. The amount of tar and sensory quality and the consumption of tobacco is closely related; the difficulty of suction and suction and the amount of tar, smoke nicotine and flue gas carbon monoxide and other safety and health indicators are closely related. Therefore, the manufacturing enterprises from their own point of view in line with the weight, circumference and suction standards to determine the important factors of cigarettes

In the case of equipment, material $N\left(\mu,\sigma^2\right)$, operation level and so on, the normal distribution variance is also basically constant, so that the normal distribution mean To maintain a certain interval in order to make most of the numerical distribution within the tolerance, that is, to meet the maximum standards, and the normal distribution of the mean and similar to the normal distribution, so as to solve the problem provides a theoretical basis.

Table 1 for the 'real dragon' brand of physical indicators of the test data. Through the simple operation of the table, the practical application of the introduction (generally $n\ge 30$).

Table 1.

Number	weight/g	Resistance/KPa	Circumference/mm
1	0.901	1.187	24.18
2	0.869	1.188	24.12
3	0.876	1.162	24.23
4	0.933	1.231	24.21
5	0.899	1.169	24.27
6	0.879	1.125	24.21
7	0.891	1.174	24.21
8	0.907	1.197	24.22
9	0.881	1.183	24.19
10	0.928	1.242	24.22
11	0.876	1.087	24.27
12	0.909	1.17	24.19
13	0.894	1.113	24.21
14	0.912	1.212	24.21
15	0.903	1.122	24.18
16	0.91	1.217	24.21
17	0.884	1.091	24.18
18	0.895	1.173	24.23
19	0.819	1.059	24.14
20	0.867	1.098	24.16
21	0.867	1.118	24.11
22	0.888	1.102	24.17
23	0.852	1.136	24.18
24	0.877	1.163	24.29
25	0.927	1.159	24.24
26	0.908	1.191	24.16
27	0.889	1.094	24.16
28	0.866	1.161	24.21
29	0.866	1.138	24.22
30	0.866	1.137	24.18
Mean	0.888	1.188	24.2
Standard deviation	0.025	1.188	24.2

Take the suction as an example, the 'true dragon' brand smoke single smoking resistance standard $(1.210 \pm 0.150)kPa$, the average resistance standard for the suction , from the table can be 'real dragon' suction resistance similar to the normal distribution $N(1.153, 0.047^2)$, and suction mean approximate obey $N\left(1.153, \frac{0.047^2}{30}\right)$. Operators want to make a single suction resistance to meet the standard, first of all to maximize the average resistance to meet the standard and the operator in the actual operation has been to the average resistance to control the object $(1.210 \pm 0.070)kPa$. Therefore, the suction resistance mean normal distribution to $N\left(1.153, \frac{0.047^2}{30}\right)$ the left (want smaller) or to the right (to the larger value) cannot be more than the average resistance line tolerance line $(\pm 0.070kPa)$. Assuming that the normal distribution pattern encounters the tolerance line, then the value of the mean value of the suction can be obtained. If the normal distribution map is translated to the left $=\left(99.73\%+x=\frac{(1-99.73\%)}{2}\right)$, the minimum value of the suction resistance can be obtained; otherwise, Maximum suction resistance. In both cases, the average resistance rate of suction resistance so that not only can better meet the criteria of the average $(1.210\pm0.070)kPa$ suction resistance, but also better meet the smoke $(1.210\pm0.150)kPa$ resistance standards

Minimum value of suction resistance

$$\mu_L=\mu_0-\frac{T}{2}+3\frac{\sigma}{\sqrt{n}}$$

Maximum suction resistance

$$\mu_U=\mu_0-\frac{T}{2}-3\frac{\sigma}{\sqrt{n}}$$

To promote the minimum value of suction resistance

$$\mu_L=\mu_0-\frac{T}{2}+K\frac{\sigma}{\sqrt{n}}$$

Maximum suction resistance,

$$\mu_U=\mu_0-\frac{T}{2}-K\frac{\sigma}{\sqrt{n}}$$

This is two important derivation formula, μ_0 is the standard value of the suction resistance standard, the T tolerance of the mean value of the suction resistance, the K multiple of the σ suction resistance and is the $\frac{\sigma}{\sqrt{n}}$ standard deviation of the mean value of the suction resistance. Take $\mu_0=1.210kPa,\frac{T}{2}=0.070kPa,\frac{\sigma}{\sqrt{n}}=\sqrt{\frac{0.047^2}{30}}=0.0086,K=3$ the formula into the formula that is $\mu_L=1.166kPa,\mu_U=1.254kPa,$, the $(1.166,1.254)kPa$ average resistance between, $(1.210\pm0.070)kPa$ and $(1.140\sim1.280)kPa$ is reduced and is more effective as scientific control of the suction resistance indicators.

In summary, in the tobacco processing enterprises, the use of the mean distribution of the sample that is the central limit theorem, the weight, circumference and suction resistance can be more effective, scientific quality control, and narrow control range, more accurate control.

3.5. The Application of Central Limit Theorem in Social Life

As a result of the continuing population growth and the serious imbalance between men and women, government departments have slowly begun to take various measures to prevent, before this, the sex of newborn babies to determine and statistics is necessary, and the center limit theorem in this respect can reflect its unique role.

Set the boy birth rate of 0.515, seeking 10,000 newborn babies in the number of girls is not less than the probability of how much?

The number of boys in the 10,000 infants is the number of boys, then X (1000, 0.515), the number of girls required is not less than the probability of the number of boys $P\{x \leq 5000\}$, by the Di Moore - Laplas theorem

$$P\{x \leq 5000\} = \Phi\left(\frac{5000 - 1000 \times 0.515}{\sqrt{1000 \times 0.515 \times 0.485}}\right) = \Phi(-3) = 1 - \Phi(3) = 0.00135$$

4. Conclusion

In this paper, we can see that the central limit theorem can play a great role both in the field of theoretical mathematics and in practical application. When these examples are mainly applied to the three general forms of the central limit theorem. In the practical problem, if the problem studied is a large sample problem, we can also use the central limit theorem to carry on the statistical analysis to it, to carry on the inferred trajectory to some of the general parameters, we can see that the central limit theorem in real life Application is very broad, is the core of probability theory, learn to use the central limit theorem for our study and life is very helpful. Therefore, it is necessary to use the central limit theorem of Lindeberg-Levi's limit theorem, Diemo-Laplacian center limit theorem, and the Lyapunov center limit theorem, and use them to solve the problem of real life Is very necessary.

The central limit theorem applied to the sample is large enough that the sample obeys the distribution of the whole state, which embodies the characteristics of the variables in the distribution. In the occasional existence of the inevitability, when the sample obeys the normal distribution, the use of the central limit theorem approximation calculation, more accurate. Under certain conditions, even if the original does not obey the overall distribution of some of the machine variables may also be a normal distribution for the limit.

References

1. Xie Yongqin. Probability Theory and Mathematical Statistics [M]. Beijing: Beijing University of Posts and Telecommunications Press, 2009: 18-36.
2. Wei Wenyuan. Probability theory and mathematical statistics [M]. Tianjin: Tianjin Education Press .2006: 89-94.
3. Su Chun. Probability theory [M]. Beijing: Science Press, 2010: 34-36.
4. Gong Yourong. Statistics [M]. Beijing: Mechanical Industry Press, 2004
5. Yan Keyu. Cigarette smoke chemistry [M]. Zhengzhou: Zhengzhou University Press, 2002.
6. Ding Zhengsheng. Probability theory and mathematical statistics concise tutorial [M]. Beijing: Higher Education Press, 2005: 27-49.
7. Fan Jincheng, Zhou Jia, Qian Changben. Probability theory [M]. Xi'an: Xi'an Jiaotong University Press .1987: 105-130.
8. Wei Zongshu. Probability theory and mathematical statistics tutorial [M]. Beijing: Higher Education Press .2008: 45-99.
9. Song Qinglong. The central limit theorem in the mall management role [J]. Shopping Mall Modernization. 2006, (482): 70-71
10. Fisher R. A. Theory of statistical estimation. Proceedings of the Cambridge Philosophical Society, 1925 (22): 700-715
11. Finklestein J A, Byar D P. Percentage points of the asympotic distributions of one and two sample K-S statistics for truncated or censored data. Technologist, 1975 (17): 507-510