Journal of Sociology and Ethnology (2021) 3: 84-90
Clausius Scientific Press, Canada
DOI: 10.23977/jsoce.2021.030211
ISSN 2616-2318

On the Application of Central Limit Theorem in Practical Problems

Yong He*

Foundational Course Department, Wuhan Donghu University, Wuhan 430212, Hubei, Peoples R China

274708721@qq.com

**corresponding author*

Keywords: Central limit theorem, Term life insurance profit and loss, Statistical inference

Abstract: The central limit theorem is the bridge between probability theory and mathematical statistics, and also the bridge between statistical methods and practical problems. It provides a theoretical basis for us to solve practical problems. This paper discusses the application of the central limit theorem in real life from two aspects: profit and loss of term life insurance in economics, determining production task according to demand and statistical inference in statistics. These cases and methods can not only be used as a supplement to the course teaching, but also can be used as a discussion of course ideology and politics.

1. Introduction

In practical problems, many random variables obey normal distribution. Even if some random variables do not obey normal distribution, they are formed by the combined influence of a large number of independent random factors. Such random variables usually obey normal distribution approximately. In probability theory and mathematical statistics, this theorem describing the sum of a large number of independent random variables approximately obeying the normal distribution is the central limit theorem. The central limit theorem paves the way for the application of mathematical statistics in social statistics. The key to inferring a population with samples is to master the sampling distribution of samples. The central limit theorem shows that as long as the sample size is large enough, the samples of an unknown population will approximately obey normal distribution. As long as a large number of observations are used to obtain enough random sample data, the methods used to deal with problems in mathematical statistics can be applied to statistics. Therefore, the central limit theorem provides a theoretical basis for us to use the normal distribution to solve practical problems, and it has been widely used in the fields of economic management and medical engineering. The application of the central limit theorem is particularly prominent in the field of finance and insurance. In 2005, Wang Donghong studied the important application of the central limit theorem in insurance [1]. In 2011, Wang Bingcan et al. studied the application of the central limit theorem in the formulation of insurance premium and the amount of reserve, the formulation of the number of insurance policies and the reduction of the average risk value of the insured individual [2].

But the central limit theorem in the current university undergraduate probability and mathematical statistics teaching materials involved in the actual application, this is not conducive to the education idea, the teaching practice of course in this paper, from the economics of term life insurance profit and loss, according to the demand to determine the production task, and statistics in statistical inference two aspects to discuss the application of central limit theorem in real life, These cases can not only serve as the supplement of course teaching, but also serve as the ideological and political discussion of the course.

2. Two Important Central Limit Theorems

2.1. Lindberg-Levy Central Limit Theorem

Assuming that $(X_1, X_2, \cdots, X_n)$ is a simple random sample from the population X, $E(X) = \mu, D(X) = \sigma^2$, when n is sufficiently large

$$\sum_{i=1}^{n} X_i \overset{\cdot}{\sim} N(n\mu, n\sigma^2), \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{\cdot}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right), \quad \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \overset{\cdot}{\sim} N(0,1)$$

2.2. DeMoivre - Laplace Central Limit Theorem

Assuming that $(X_1, X_2, \cdots, X_n)$ is a simple random sample from the population $X \sim B(1, p)$, when n is sufficiently large

$$\sum_{i=1}^{n} X_i \overset{\cdot}{\sim} N(np, np(1-p)), \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{\cdot}{\sim} N\left(p, \frac{p(1-p)}{n}\right), \quad \frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{\cdot}{\sim} N(0,1)$$

When n is sufficiently large, $X = \sum_{i=1}^{n} X_i \sim B(n, p) \overset{\cdot}{\sim} N(np, np(1-p))$, so

$$P\{a < X \le b\} = F(b) - F(a) \approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right)$$

3. Application of the Central Limit Theorem in Economics

3.1. Term Life Insurance Profits and Losses

We know that the purpose of life insurance company operation is to make profits, and the profit and loss of the insurance company can be estimated and predicted by using the central limit theorem. Assuming that n people in a life insurance company take out one-year term life insurance in a period of time, whether they die within a year is independent of each other, and no new policyholders join the insurance business within a year, and no one resigns, we use the central limit theorem to estimate the profit and loss probability of these policies [3]. Assume that the premium per policy is M, the amount insured is Q, and the mortality rate of the policyholder is p, define:

$$X_i = \begin{cases} 1, & \text{The ith people dead} \\ 0, & \text{The ith people alive} \end{cases}, i = 1, 2, \cdots n$$

When n is sufficiently large, $X = \sum_{i=1}^{n} X_i \sim B(n,p) \overset{\cdot}{\sim} N(np, np(1-p))$, according to the central limit theorem, the loss probability of the insurance company is

$$P(nM < XQ) = P\left(X > \frac{nM}{Q}\right) = 1 - P\left(X \le \frac{nM}{Q}\right) = 1 - F\left(\frac{nM}{Q}\right) = 1 - \Phi\left(\frac{\frac{nM}{Q} - np}{\sqrt{np(1-p)}}\right) = \beta$$

If β is small, it will be beneficial to the company's profits. When β is large, in order to make profits, life insurance companies can reduce the loss rate by increasing premiums and other means.

3.2. Determine Production Tasks According to Demand (Sample Size)

Suppose that the factory is responsible for the supply of goods for n individuals in a certain area, and the probability that each person will need one piece of this commodity in a certain period of time is p. Assuming that each person will buy or not independently during this period of time, the factory has at least β assurance to meet the social demand, and the number of goods that the factory needs to produce is at least M [3], define:

$$X_i = \begin{cases} 1, & \text{The ith people buys the goods} \\ 0, & \text{The ith people doesn't buys the goods} \end{cases}, i = 1, 2, \cdots n$$

When n is sufficiently large, $X = \sum_{i=1}^{n} X_i \sim B(n,p) \overset{\cdot}{\sim} N(np, np(1-p))$, according to the central limit theorem,

$$P(X \le M) = F(M) = \Phi\left(\frac{M - np}{\sqrt{np(1-p)}}\right) \ge \beta$$

Define $\Phi(x_\beta) = \beta$, so $M \ge np + x_\beta\sqrt{np(1-p)}$, the factory needs to produce at least $np + x_\beta\sqrt{np(1-p)}$ goods.

4. Application of Central Limit Theorem in Statistical Inference

4.1. Application of Central Limit Theorem to Parameter Estimation

Mathematical statistics generally studies parameter estimation and hypothesis testing of the mean value and variance of the normal population, but the random variable sequences obtained by random sampling are independent of each other and subject to the same distribution. Therefore, in practical problems, no matter what distribution the population is subject to, as long as the sample size is large enough, We can use the central limit theorem to treat the sum of random variables as being normally distributed, $(X_1, X_2, \cdots, X_n)$ is a simple random sample from the population X, $E(X)=\mu, D(X)=\sigma^2$, when n is sufficiently large

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \dot{\sim} N(0,1)$$

When the confidence level is equal to $1-\alpha$, the confidence interval of the population mean μ is

$$\left(\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}, \overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$$

When the population variance σ^2 is unknown, since the sample variance s^2 is the unbiased estimator of the population variance σ^2, the sample variance s^2 is used to replace the population variance σ^2, which can be known from the central limit theorem: when n is sufficiently large

$$\frac{\overline{X}-\mu}{s/\sqrt{n}} \dot{\sim} N(0,1)$$

When the confidence level is equal to $1-\alpha$, the confidence interval of the population mean μ is

$$\left(\overline{X}-\frac{s}{\sqrt{n}}z_{\alpha/2}, \overline{X}+\frac{s}{\sqrt{n}}z_{\alpha/2}\right)$$

4.1.1. For special cases, the population $X \sim B(1,p)$, where p is unknown

Assuming that $(X_1, X_2, \cdots, X_n)$ is a simple random sample from the population X, $E(X)=p$, $D(X)=p(1-p)$, when n is sufficiently large

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}=\frac{\overline{X}-p}{\sqrt{\frac{1}{n}p(1-p)}} \dot{\sim} N(0,1)$$

When the confidence level is equal to $1-\alpha$, the confidence interval of the population mean p is

$$\left(\overline{X}-\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}z_{\alpha/2}, \overline{X}+\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}z_{\alpha/2}\right)$$

Example 1: a certain region investigated the proportion of women among laid-off workers, and randomly selected 36 laid-off workers, 20 of whom were women. When the confidence level is equal to 95%.The confidence interval is obtained for the proportion of female laid-off workers in this region.

Solution: $n=36$, $\bar{x}=\frac{20}{36}$, $\alpha=0.05$, $z_{\alpha/2}=z_{0.025}=1.96$, $\left(\bar{X}-\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}z_{\alpha/2}, \bar{X}+\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}z_{\alpha/2}\right)=(0.40,0.72)$

4.1.2. For special cases, the population $X\sim P(\lambda)$, where λ is unknown

Assuming that $(X_1,X_2,\cdots,X_n)$ is a simple random sample from the population X, $E(X)=\lambda$, $D(X)=\lambda$, when n is sufficiently large

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}=\frac{\bar{X}-\lambda}{\sqrt{\frac{\lambda}{n}}}\ \dot{\sim}\ N(0,1)$$

When the confidence level is equal to $1-\alpha$, the confidence interval of the population mean λ is

$$\left(\bar{X}-\sqrt{\frac{\bar{X}}{n}}z_{\alpha/2}, \bar{X}+\sqrt{\frac{\bar{X}}{n}}z_{\alpha/2}\right)[4]$$

Example 2: Arrived at the bus stop in a unit time the number of passengers to $P(\lambda)$, for different station, the difference is only a value of different parameters λ, one of the city bus stop for the investigation of 100 time units, per unit time is 20 minutes, calculate every 20 minutes to the passenger Numbers of average $\bar{X}=15.2$ people at the station. When the confidence level is equal to 95%, the confidence interval of the population mean λ is obtained.

Solution: $n=100$, $\bar{x}=15.2$, $\alpha=0.05$, $z_{\alpha/2}=z_{0.025}=1.96$, $\left(\bar{X}-\sqrt{\frac{\bar{X}}{n}}z_{\alpha/2}, \bar{X}+\sqrt{\frac{\bar{X}}{n}}z_{\alpha/2}\right)=(14.436,15.964)$

4.2. Application of Central Limit Theorem in Hypothesis Testing

In many practical problems, studies are conducted on almost all large samples. According to the central limit theorem, when the sample size is large, hypothesis test is carried out on the mean of the population. Come up with the null hypothesis: $H_0: E(X)=\mu_0$, when the null hypothesis holds, the test statistics

$$U=\frac{\bar{X}-\mu_0}{\sqrt{D(X)}/\sqrt{n}}\dot{\sim}N(0,1) \text{ or } U=\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\dot{\sim}N(0,1)$$

When the significance level is α, the rejection domain of H_0: two-sided test $|u|>z_{\alpha/2}$, right side test $u>z_\alpha$, left side test $u<-z_\alpha$.

Example 3: According to statistics, the infection rate at the beginning of COVID-19 epidemic in 2020 was 5%. In order to meet the needs of epidemic prevention, 500 people were randomly selected from a community for quarantine, and the result showed that the number of infected people was no more than 1. Is the infectious disease already prevalent in the community?

Solution: Assuming that the COVID-19 infection rate in this community is p, the number of people infected is X, so $X \sim B(n,p)$, because $n=500$ is big, by the central limit theorem $X \overset{\cdot}{\sim} N(np,npq)$, so

$$\frac{X}{n} \overset{\cdot}{\sim} N\left(p,\frac{1}{n}pq\right), \frac{\frac{X}{n}-p}{\sqrt{\frac{1}{n}pq}} \overset{\cdot}{\sim} N(0,1)$$

The fact that the disease is already prevalent in the community means that the COVID-19 infection rate in the community is $p \geq p_0 = 5\%$, this is essentially asking for the following left-sided hypothesis test:

The null hypothesis: $H_0 : p \geq p_0 = 5\%$; alternative hypothesis: $H_1 : p < p_0 = 5\%$

when the null hypothesis holds, the test statistics:

$$U = \frac{\frac{X}{n}-p_0}{\sqrt{\frac{1}{n}p_0(1-p_0)}} \overset{\cdot}{\sim} N(0,1)$$

So

$$P\left\{\frac{\frac{X}{n}-p_0}{\sqrt{\frac{1}{n}p_0(1-p_0)}} < -z_\alpha\right\} = \alpha$$

When the significance level is 0.01, the rejection domain of $H_0 : \{u < -z_\alpha\} = \{u < -z_{0.01}\} = \{u < -2.33\}$,

When $n=500, X=1$, so the test statistics

$$u = \frac{\frac{1}{500}-0.05}{\sqrt{\frac{0.05 \times 0.95}{500}}} \approx -4.8 < -2.33$$

Therefore, H is rejected. So, it can be considered that the epidemic disease has not been prevalent in this community, and it is completely in time to take preventive measures.

5. Conclusion

Due to the practical problems involved are large sample, whether or not known is not a necessary condition for the overall distribution, when the sample size is large enough can approximate as normal distribution, which is central limit theorem can solve a lot of practical problems in the, this also allows us to the central limit theorem research has certain theoretical significance and practical value, These research methods and cases are also a powerful supplement to our classroom teaching, which makes this course practice the ideological and political ideas of the course into practice!

References

[1] Wang Donghong. Important Applications of Law of Large Numbers and Central Limit Theorem in Insurance [J]. Mathematics in Practice and Theory,2005,35(10): 128-133.

[2] Wang Bingshen, Wei Yanhua, Lin Zhu. Application of Law of Large Numbers and Central Limit Theorem in Insurance [J]. Journal of Tonghua Normal University,2011,32(12):8-10.

[3] Pull poor, side ba, side ba zola. Application of Probability Limit Theory in Economic Insurance [J]. Mathematics in Practice and Theory,2018,48(12): 276-281.

[4] He Yong, Ming Jiexiu. Probability Theory and Mathematical Statistics [M]. Wuhan: Wuhan University Press,2012.

[5]Zhang zhicheng, zhang yan. Application of central limit theorem in practice [J]. Journal of henan institute of technology,2020,8(4): 28-30.