# Application of Central Limit Theorem and Law of Large Numbers in Insurance Industry

## **Ruoyu Huang**

Shanghai Pinghe School, Shanghai, China huangruoyu22222@outlook.com

*Abstract:* Insurance safeguards property and health while mitigating economic risks, driving research to optimize mechanisms through mathematical tools. This paper examines the Central Limit Theorem (CLT) and Law of Large Numbers (LLN) in insurance, highlighting their complementary roles: CLT models loss distributions (e.g., approximating total risks as normal for capital reserve estimation) and LLN stabilizes long-term claims (e.g., balancing premium affordability and profitability through risk pooling). Empirical analyses across auto, health, and life insurance demonstrate CLT quantifying extreme payout probabilities (95% confidence intervals for catastrophic losses) and LLN optimizing premium structures via historical claims convergence. Copula theory extends traditional independence assumptions for correlated risks (e.g., regional disaster impacts on crop insurance), while machine learning (e.g., random forests) enhances personalized risk predictions through training on multi-source data (demographics, telematics). The synergy of CLT and LLN strengthens capital allocation and solvency frameworks, offering a robust mathematical foundation for sustainable, datadriven risk governance. By integrating classical theory with modern techniques like highdimensional Copula models and reinforcement learning, this study advances dynamic premium adjustments, real-time capital optimization, and compliance strategies. These innovations address emerging challenges in InsurTech, such as managing climate-induced risks (e.g., flood insurance) and ensuring transparency in AI-driven underwriting systems.

*Keywords:* Central limit theorem, Law of large numbers, Insurance actuarial, Risk assessment, Normal approximation

## 1. Introduction

The core of the insurance industry lies in risk transfer and allocation, and its mathematical foundation relies on the two pillars of probability theory: the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT). LLN guarantees the stability of long-term expected losses through convergence, while CLT provides a quantitative tool for short-term fluctuations through a normal distribution. Traditional insurance models are mostly based on independent and equally distributed assumptions, while actual risks are often correlated (for example, natural disasters lead to simultaneous claims in multiple regions). This paper explores the adaptability of classical theories in modern insurance scenarios by extending independent risk models and machine learning techniques.

Wang showed that the maximum amount of third-party liability insurance in auto insurance is set at 200,000 yuan, which is behind the joint effect of CLT and LLN [1]. When the number of insured subjects is large enough, the insurance company calculates the extreme probability of payment

<sup>@</sup> 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

through CLT, and combines the convergence of LLN to ensure long-term profit stability. This conclusion provides a theoretical basis for the insurance industry's premium limitation. In addition, Tang verified the universality of the law of large numbers through lottery examples, revealing the regularity of random events under large samples, and further supporting the feasibility of insurance risk pooling [2].

The rest of this paper is organized as following. In section 2, the author will introduce the method and theory. Next, section 3 is devoted to results and applications. Finally, the last section gives a conclusion.

## 2. Method and Theory

## 2.1. Law of Large Numbers and Central Limit Theorem

By setting independent identically distributed (i.i.d.) random variable sequence  $X_1, X_2, ..., X_n$ , whose expectation is  $\mu$ , then the sample mean converges to the expectation with probability:

$$\lim_{n \to \infty} P\left( \left| \frac{1}{n} \sum_{i=1}^{n} X_{i} - \mu \right| < \varepsilon \right) = 1.$$
(1)

This is the weak law of large numbers. Sinchin's Law of large numbers further relaxes the conditions, only the expectation exists, and is applicable to the thick tail distribution scenario.

For insurance applications, the first is about premium pricing, estimating expected losses from historical claims data E[X], to ensure that premiums cover long-term costs. For example, in the case of auto insurance, Wang predicted the average payout amount through LLN and designed the premium structure combined with the safety additional amount [3]. Second is about risk pooling, the number of subjects when insured n is large enough, the actual loss rate approaches the theoretical probability (such as mortality).

Sequence of random variables for i.i.d.  $X_1, X_2, ..., X_n$ , whose mean is  $\mu$ , the variance is  $\sigma^2$ , Then the standardized sample mean converges to the standard normal distribution according to the distribution:

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \stackrel{d}{\to} N(0, 1)$$
(2)

The Lindeberg-Levy CLT is its classical form, which requires finite variance. For insurance applications, the first is about reserve calculation, total claims paid  $S = \sum X_i$ , approximate obedience  $N(n\mu, n\sigma^2)$ , to estimate the probability of extreme losses. Wang proposed that CLT was used to calculate the additional safety factor  $\lambda$ , which ensures that insurance companies avoid losses with a 95% probability [4]. The second is about confidence interval construction, calculating the margin of safety for premium adjustment based on normal distribution.

## 2.2. Contrast of similarities and differences

The Comparison of theorem of large numbers and central limit theorem is shown in Table 1. As two cornerstones of probability theory, Central Limit theorem (CLT) and Law of large Numbers (LLN) show remarkable commonality in both theory and application. First, both require that the sequence of random variables satisfy the basic assumption of an independent identical distribution (i.i.d.), a condition that provides mathematical rigor for statistical inference. Second, they are both suitable for analyzing the collective behavior of a large number of random variables: LLN characterizes long-term stability through the convergence of the sample mean, while CLT quantifies short-term volatility

through a normal approximation. In practice, they have irreplaceable value in risk modeling and data analysis in insurance, finance, medicine and other fields. For example, insurance companies use LLN to predict the long-term average loss ratio to set premiums, while CLT to estimate the extreme loss probability to determine the size of reserves. The synergistic effect of the two forms the theoretical basis of modern actuarial science [3, 5].

Table 1	: Com	parison	of theorem	of large	numbers	and	central	limit	theorem
I dole I	. 00111	puliboli	or meorem	or large	numbers	unu	contrat	1111110	uncorenn

Peculiarity	Theorem of Large Numbers (LLN)	Central Limit Theorem (CLT)			
	The mean of the random variable	The distribution of the mean (or			
Core content	approaches the mathematical	sum) of random variables tends			
	expectation.	to be normal.			
Focus	Convergence (approximate to	Distribution pattern (normal			
Focus	expected value).	distribution).			
Mathematical expression	$\lim_{n \to \infty} P\left( \left  \frac{1}{n} \sum_{i=1}^{n} X_{i} - \mu \right  < \varepsilon \right) = 1$ (Converges almost everywhere or with probability)	$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \stackrel{d}{\to} N(0,1).$			
Convergent type	Converges almost everywhere or with	Distributional convergence			
Convergent type	probability.	(weak convergence).			
	Probability estimation, expected	Statistical inference, risk			
Application scenario	value calculation, Monte Carlo	assessment, distribution			
	simulation.	approximation.			
The requirement for	Independent identically distributed	Independent identically			
random variables	expectation exists	distributed, with limited			
		variance.			

Although both CLT and LLN are based on the assumption of independent co-distribution, their core concerns and application scenarios are fundamentally different. From the perspective of mathematical connotation, LLN emphasizes that the sample mean converges to the expected value according to probability  $\overline{X}_{n\to}^{\ \ p}\mu$ , while CLT reveals that the distribution of the standardized sample mean converges to the normal distribution  $\overline{Z}_{n\to}^{\ \ d}N(1,0)$ . This difference leads directly to the divergence of its application goals: LLN is used primarily to verify theoretical expected values (e.g., mortality, frequency of claims), while CLT is used to construct confidence intervals or calculate the probability of extreme events (e.g., bankruptcy risk). Moreover, the mathematical requirements for the two are different - CLT requires a finite variance to guarantee the validity of a normal approximation, whereas LLN requires only the expected existence (as in Sinchin's Law of large numbers). Taking insurance solvency analysis as an example, LLN can predict the average number of deaths, while CLT further calculates the probability that the number of deaths exceeds the threshold, thus providing a quantitative basis for capital adequacy regulation [4, 6].

Combining the advantages of CLT and LLN, the multi-dimensional analysis of random phenomena can be realized. In the insurance industry, LLN provides a long-term equilibrium benchmark for premium pricing, while CLT quantifies risk exposure through a normal distribution, and the combination of the two can optimize risk management strategies (such as safety add-on factor design). Further, modern risk theories have broken through the traditional assumption of independent and same distribution. For example, Copula model extends the application boundary of CLT by describing the correlation between variables [7], while machine learning techniques (such as random

forest) integrate the convergence idea of LLN to improve the accuracy of individual risk assessment. Future research could explore the adaptability of non-parametric CLT in high-dimensional data, or combine deep learning models to enhance the robustness of LLN in non-stationary environments. These advances will promote the application of probability theory in complex systems such as artificial intelligence and climate prediction, demonstrating its interdisciplinary value.

## **2.3.** Extension of independent risk model (Copula theory)

The traditional CLT assumes that risks are independent, but in practice (such as regional disasters) risks are often correlated. Copula model provides a flexible risk modeling tool by separating the correlation structure and edge distribution between variables [8].

Let the marginal distributions of two risk variables X and Y be  $F_X(x)$  and  $F_Y(y)$ , respectively, and their joint distributions can be expressed as:

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y))$$
(3)

where C(u, v) is the Copula function. For a Gaussian Copula, the form is:

$$C_{\rho}(u, v) = \Phi_{\rho}(\Phi^{-1}(u), \Phi^{-1}(v))$$
(4)

where  $\rho$  is the correlation coefficient and  $\Phi_{\rho}$  is the bivariate normal distribution function. For application significance, Copula was used to generate relevant risk samples and analyze the distribution of total loss S=X+Y. And the value at risk (VaR) under independent and non-independent hypotheses is compared to reveal the amplification effect of correlation on extreme events.

## 3. **Results and Application**

## 3.1. Auto Insurance Premium Pricing and Risk Reserve

For the problem setting, an insurance company covers 10,000 vehicles, each vehicle annual loss  $X_i \sim Exp(\lambda)$ , needs to determine the total reserve H so that the probability of underpayment is less than 5%. For the LLN applications, the expected total loss is  $E[S] = n/\lambda$ . When n is large enough, the actual payout S fluctuates around E[S] [9]. For the CLT applications, By CLT,  $S \approx N(n/\lambda, n/\lambda^2)$ . If  $Z = \frac{S-n/\lambda}{\sqrt{n/\lambda}}$ , then,

$$P(S > H) = P\left(Z > \frac{H - n/\lambda}{\sqrt{n/\lambda}}\right) = 0.05 \Rightarrow H = \frac{n}{\lambda} + 1.645 \cdot \sqrt{\frac{n}{\lambda}}$$
(5)

As a calculation example, if n = 10,000,  $\lambda = 0.01$ , then  $H = 1,000,000 + 1.645 \times 316.23 \approx 1,052,000$ .

Wang further analyzed the limitation of the amount of third-party liability insurance [6]. Assuming that an insurance company underwrites 10,000 policies and the probability of death p = 0.001, the CLT calculation shows that when the insured amount exceeds 200,000 yuan, the profit of the insurance company decreases rapidly with the increase of the number of deaths. For example, when the number of deaths is 13, the profit corresponding to 200,000 yuan of insurance is 7.61 million yuan, while the loss of 1 million yuan of insurance is 130,000 yuan. This result explains why insurance companies limit the maximum coverage to 200,000 yuan.

## 3.2. Prediction of Medical Expenses in Health Insurance

For problem setting, 100,000 enrollees in certain area  $X_i \sim \Gamma(k = 2, \theta = 1000)$ , each annual medical cost A, need to assess the probability that the total cost exceeds 210 million yuan. For LLN

Perspective, the long-term average total cost is approaching  $E[S] = nk\theta = 2 \times 10^8$  [10]. For the CLT calculation, by CLT, total cost  $S \approx N(2 \times 10^8, 2 \times 10^{11})$ , then

$$P(S > 2.1 \times 10^8) = 1 - \Phi\left(\frac{2.1 \times 10^8 - 2 \times 10^8}{\sqrt{2 \times 10^{11}}}\right) \approx 1 - \Phi(2.236) = 1.27\%$$
(6)

In conclusion, the probability of extreme overspending is low, which verifies the robustness of the insurance pool.

For simulation method, Gaussian Copula is used to generate random variables X and Y with correlation coefficient  $\rho = 0.7$ , whose edge distribution is exponential. A typical example is shown in Figure 1.



Figure 1: Total loss analysis under the independent risk model.

The results that the right tail of the total loss distribution is thicker than the normal distribution, indicating that the correlation exacerbates the extreme risk. Reserve strategies need to be adjusted to cope with potentially high payouts.

#### 4. Conclusion

The Central Limit Theorem (CLT) and the Law of Large Numbers (LLN) serve as dual pillars in the insurance industry, with LLN ensuring long-term stability through expectation convergence and CLT quantifying short-term risks via normal distribution approximations. By integrating Copula models to address correlated risks and machine learning techniques like random forests for individualized claim predictions, this study bridges classical probability theory with modern data-driven approaches. For example, random forests refine granular risk assessments, while CLT aggregates total risk distributions to calculate extreme payout probabilities, enabling dynamic premium adjustments and efficient capital allocation. Wang's work further validates that CLT-enhanced safety load calculations and expanded underwriting scales significantly improve solvency, demonstrating the synergy of theoretical rigor and practical adaptability in actuarial science.

Future research should focus on three frontiers, the first one is leveraging Generative Adversarial Networks (GANs) to simulate loss data, enhancing predictive accuracy in low-data scenarios; the second is developing high-dimensional Copula models to unravel complex risk interdependencies (e.g., climate-economic-health linkages); and the final one is deploying reinforcement learning for real-time, adaptive premium strategies. Additionally, as InsurTech advances, ensuring the robustness

of CLT and LLN in distributed and heterogeneous data environments will be critical. These innovations, coupled with ethical AI frameworks, could redefine risk governance, enabling insurers to balance profitability with societal resilience. Ultimately, fusing classical theorems with cutting-edge technologies promises transformative breakthroughs in sustainable insurance solutions.

#### References

- [1] Wang, D. H. (2005). The important application of the law of large numbers and the central limit theorem in the insurance industry. Mathematics in Practice and Theory, 35(10), 128–133.
- [2] Tang, L., & Li, Y. R. (2005). Practical applications of the law of large numbers and the central limit theorem. Journal of Guangdong Polytechnic Normal University, (6), 75–76.
- [3] Wang, B. C., Wei, Y. H., & Lin, Z. (2011). The application of the law of large numbers and the central limit theorem in insurance. Journal of Tonghua Normal University, 32(12), 8–10.
- [4] Wang, J. Z. (2003). Application of the central limit theorem in analyzing insurance solvency. Statistical Education, (2), 33–34.
- [5] McNeil, A. J., Frey, R., & Embrechts, P. (2015). Quantitative risk management. Princeton University Press.
- [6] Yue, J. J. (2007). Application of the central limit theorem and the law of large numbers in insurance. Journal of Jiangsu Institute of Education (Natural Sciences), 24(4), 1-3.
- [7] Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). Modelling extremal events for insurance and finance. Springer.
- [8] Bai, Y. F., & Liu, L. (2013). Relationship and applications between the law of large numbers and central limit theorem. Undergraduate Thesis, Qufu Normal University.
- [9] Kaas, R., Goovaerts, M., Dhaene, J., & Denuit, M. (2008). Modern actuarial risk theory: Using R. Springer.
- [10] Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). Loss models: From data to decisions. Wiley.