Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 46 No. 2 (2025)

# A Comparative Analysis of Discrete and Continuous Distributions in Risk Assessment Models

**Monika Arora[1], Vishwa Prakash Jha[2*]**

[1]Department of Mathematics, Lovely Professional University Jalandhar 144411, India

[2]School of Computer Science & Engineering, IILM University, Gurugram, India

**Abstract:** Risk assessment models often rely on probabilistic frameworks to quantify and manage uncertainty. Choosing an appropriate probability distribution—whether discrete or continuous—is critical to accurately modeling risk in various domains such as finance, engineering, and healthcare. This paper presents a comprehensive comparative analysis of discrete and continuous probability distributions within the context of risk assessment. We examine the statistical properties, modeling flexibility, and computational efficiency of commonly used distributions, including the Binomial, Poisson, Normal, and Exponential distributions. Through simulation studies and real-world case applications, we investigate how the choice of distribution affects risk metrics such as Value-at-Risk (VaR), Conditional Value-at-Risk (CVaR), and expected shortfall. The results highlight scenarios where discrete models are more suitable due to event-based uncertainty, as well as cases where continuous models offer superior accuracy in capturing risk magnitudes. This study provides guidance for practitioners in selecting appropriate distributional frameworks to enhance the reliability and interpretability of risk assessment models.

***Keywords**: Value-at-Risk (VaR), Conditional Value-at-Risk (CVaR), Simulation, Poisson Distribution, Normal Distribution, Lognormal Distribution*

## 1 Introduction

Risk assessment is a fundamental process across various industries, including finance, engineering, healthcare, and environmental science, where it is essential to quantify and manage uncertainty effectively. Central to this process is the selection of appropriate probability distributions to model uncertain parameters and outcomes. The choice between discrete and continuous distributions significantly influences the accuracy, interpretability, and computational efficiency of risk models.

Discrete probability distributions are utilized when the set of possible outcomes is countable. Common examples include the Binomial and Poisson distributions, which are often employed to model the number of occurrences of an event within a fixed number of trials or a specified time frame [1]. For instance, the Poisson distribution is frequently applied in modeling the number of defaults in a credit portfolio over a given period [2]. In reliability engineering, the Binomial distribution can model the number of component failures in a system [3].

Conversely, continuous probability distributions are appropriate for modeling variables that can assume an infinite number of values within a given range. The Normal and Exponential distributions are among the most commonly used in this category. The Normal distribution is particularly prevalent in financial modeling, where asset returns are often assumed to follow a Normal distribution [4]. The Exponential distribution is widely used in reliability engineering to model time-to-failure data [3].

The selection between discrete and continuous distributions is not merely a matter of mathematical convenience but has profound implications for risk assessment outcomes. For example, Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) are critical risk metrics in finance that are sensitive to the underlying distributional assumptions [5]. Mis-specifying the distribution can lead to inaccurate risk estimates and suboptimal decision-making. CVaR, also known as expected shortfall, provides a coherent risk measure that accounts for the tail risk beyond the VaR threshold, offering a more comprehensive view of potential losses [6].

Moreover, the properties of the chosen distribution affect the computational tractability of risk models. Continuous distributions often facilitate analytical solutions and are amenable to techniques such as differentiation and integration. In contrast, discrete distributions may require summation over potentially large state spaces, which can be computationally intensive [7]. Advanced computational methods, such as Monte Carlo simulations, are often employed to handle complex distributions and dependencies in risk assessment models [8].

In practice, the choice between discrete and continuous distributions should be guided by the nature of the data and the specific context of the risk assessment. For instance, when modeling the number of operational failures in a manufacturing process, a discrete distribution like the Poisson may be appropriate. However, for modeling the time until the next failure, a continuous distribution such as the Exponential would be more suitable. In environmental risk assessments, the choice of distribution can influence the estimation of pollutant concentrations and the assessment of associated health risks [9].

Furthermore, the convergence properties of algorithms used in risk assessment, such as the Expectation-Maximization (EM) algorithm, can be affected by the choice of distribution. The EM algorithm's performance in finding maximum likelihood estimates depends on the underlying distributional assumptions [10]. Similarly, the stability of Markov Chain Monte Carlo methods is influenced by the properties of the chosen distributions [11].

This paper aims to provide a comprehensive comparative analysis of discrete and continuous probability distributions within the context of risk assessment models. By examining the statistical properties, modeling flexibility, and computational considerations of various distributions, we seek to offer insights into their appropriate application in different risk assessment scenarios. Through simulation studies and real-world case applications, we investigate how the choice of distribution influences critical risk metrics and overall model performance.

The remainder of this paper is organized as follows: Section 2 reviews the fundamental properties of selected discrete and continuous distributions commonly used in risk assessment. Section 3 discusses the impact of distribution choice on risk metrics such as VaR and CVaR. Section 4 presents simulation studies and case applications illustrating the practical implications of distribution selection. Finally, Section 5 concludes with recommendations for practitioners on selecting appropriate distributions to enhance the reliability and interpretability of risk assessment models.

## 2 Theoretical Background

In order to compare discrete and continuous probability distributions effectively within risk assessment models, it is essential to understand their fundamental definitions, statistical properties, and the contexts in which they are commonly applied. This section presents an overview of key distributions, highlighting their characteristics, use cases, and relevance in modeling risk and uncertainty.

## 2.1 Discrete Probability Distributions

Discrete probability distributions are suitable for modeling random variables that can take on a finite or countably infinite number of values. These are typically used when the modeled phenomenon involves count data or event occurrences.

**Binomial Distribution:** The Binomial distribution models the number of successes in a fixed number of independent Bernoulli trials with the same probability of success $p$. Its probability mass function (PMF) is given by:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n$$

This distribution is widely used in quality control and reliability studies [3].

**Poisson Distribution:** The Poisson distribution models the number of events occurring in a fixed interval of time or space under the assumption of constant rate $\lambda$ and independence of events. The PMF is:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

This distribution is appropriate for rare event modeling, such as insurance claims or system failures [2].

## 2.2 Continuous Probability Distributions

Continuous distributions model random variables that can take any value within a given interval. These are essential when measuring phenomena like time, weight, or financial returns.

**Normal Distribution:** Also known as the Gaussian distribution, the Normal distribution is fundamental in statistics due to the Central Limit Theorem. It is defined by its mean $\mu$ and standard deviation $\sigma$:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

It is used extensively in modeling errors, returns, and other naturally occurring phenomena [4, 12].

**Exponential Distribution:** The Exponential distribution models the time between events in a Poisson process and is characterized by its rate parameter $\lambda$:

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

It is widely used in survival analysis and reliability modeling [13].

**Gamma Distribution:** The Gamma distribution generalizes the Exponential distribution and is defined for a shape parameter $\alpha$ and a rate parameter $\beta$. Its probability density function (PDF) is:

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \quad x > 0$$

The Gamma distribution is flexible for modeling skewed data, particularly in queueing theory and climatology [15].

### 2.3 Moments and Tail Behavior

The choice of distribution often depends on its moment properties (mean, variance, skewness, and kurtosis) and tail behavior. For instance, heavy-tailed distributions such as the Lognormal or Pareto can better capture extreme risk scenarios compared to the light-tailed Normal distribution [19, 18].

Understanding these characteristics is essential for aligning the distributional assumptions of risk models with real-world data. In practice, goodness-of-fit tests and graphical methods such as Q-Q plots are used to evaluate distribution suitability [17].

In the next section, we will investigate how the use of discrete versus continuous distributions impacts the estimation and interpretation of risk metrics such as Value-at-Risk and Conditional Value-at-Risk.

## 3 Impact of Distribution Choice on Risk Metrics

Risk metrics such as Value-at-Risk (VaR), Conditional Value-at-Risk (CVaR), and expected shortfall are central to quantifying and managing uncertainty in various domains. The accuracy and reliability of these metrics depend critically on the choice of the underlying probability distribution used to model uncertain variables.

### 3.1 Value-at-Risk (VaR)

Value-at-Risk (VaR) is a quantile-based measure that estimates the maximum potential loss over a given time horizon at a specified confidence level. For a continuous distribution with cumulative distribution function (CDF) $F(x)$, VaR at confidence level $\alpha$ is defined as:

$$\mathrm{VaR}_\alpha = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$$

In discrete distributions, calculating VaR requires identifying the smallest value $x$ such that the cumulative probability meets or exceeds $\alpha$:

$$\mathrm{VaR}^{(discrete)}_\alpha = \min\{x : \sum_{i \leq x} P(X = i) \geq \alpha\}$$

This discrete nature introduces step-like behavior in the VaR estimate, which may reduce resolution for risk reporting [18, 19].

### 3.2 Conditional Value-at-Risk (CVaR)

Conditional Value-at-Risk (CVaR), also known as expected shortfall, represents the expected loss exceeding the VaR threshold. It addresses a major limitation of VaR—its inability to capture tail risk—by measuring the average of the worst losses beyond the VaR level [5, 6]. For a continuous loss distribution:

$$\mathrm{CVaR}_\alpha = \mathrm{E}[X \mid X \geq \mathrm{VaR}_\alpha]$$

In discrete settings, CVaR requires summing the tail probabilities, which can be computationally intensive when the number of possible outcomes is large. However, discrete models often allow for exact calculation when the support is finite [8].

---

### 3.3 Tail Behavior and Distributional Impact

The tail behavior of a distribution significantly affects the estimation of both VaR and CVaR. Distributions with heavy tails, such as the Lognormal or Pareto, yield higher estimates of risk compared to light-tailed ones like the Normal [19]. This discrepancy becomes critical in risk-sensitive applications such as banking and insurance.

When using a Normal distribution to estimate VaR, the tails are underestimated if the actual data are skewed or heavy-tailed, leading to flawed risk estimates [18, 4]. In contrast, using a Gamma or Lognormal distribution better reflects asymmetry and tail risks often observed in real-world data [15, 17].

### 3.4 Practical Considerations

From a modeling perspective, discrete distributions are more natural when modeling count-based risks, such as the number of defaults or component failures. Continuous distributions are more suited to financial losses or time-based risks. Importantly, the selection should be guided by data characteristics, the nature of uncertainty, and computational considerations.

Empirical studies suggest that misspecification of distributions can lead to underestimation of CVaR by more than 20% in heavy-tailed contexts [18, 19]. Therefore, model validation and distribution fitting should be integrated into the risk assessment process [9, 16].

### 3.5 Summary

This section has highlighted the critical role of distributional assumptions in calculating and interpreting risk measures. The sensitivity of VaR and CVaR to tail behavior underscores the need for careful selection and testing of probability distributions. In the next section, we present simulation studies and case examples to demonstrate these effects using real-world data.

## 4 Simulation and Case Studies

To empirically evaluate the impact of distributional choice on risk assessment, we conduct simulation studies and analyze real-world data using both discrete and continuous probability models. The focus is to compare how selected distributions affect the estimation of key risk metrics such as Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR).

### 4.1 Simulation Setup

We simulate two datasets representing different risk modeling scenarios:

**Dataset A (Discrete Events)**: Represents the number of operational failures in a production facility over 100 time periods, modeled using a Poisson distribution with $\lambda = 3$.

**Dataset B (Continuous Losses)**: Represents financial losses over 100 days, simulated from both Normal ($\mu = 0, \sigma = 1$) and Lognormal ($\mu = 0, \sigma = 0.25$) distributions.

For each dataset, we estimate VaR and CVaR at the 95% confidence level using empirical quantiles and analytical formulas where applicable.

### 4.2 Results for Discrete Distribution (Poisson)

Table 1 shows risk estimates for Dataset A based on the Poisson distribution.

**Table 1: Risk Metrics from Poisson-Distributed Event Counts**

| Confidence Level | VaR | CVaR |
|---|---|---|
| 95% | 6 | 6.72 |

The stepwise nature of the Poisson PMF results in discrete jumps in VaR. CVaR, being an average of tail values, smooths this effect but is still limited by the resolution of discrete outcomes.

### 4.3 Results for Continuous Distributions

Table 2 summarizes the risk metrics computed from continuous loss models.

**Table 2: Risk Metrics from Continuous Distributions (Normal vs. Lognormal)**

| Distribution | Confidence Level | VaR | CVaR |
|---|---|---|---|
| Normal | 95% | 1.64 | 2.06 |
| Lognormal | 95% | 1.84 | 2.65 |

As expected, the Lognormal distribution—with its positive skew and heavy tail—yields higher estimates of both VaR and CVaR. This illustrates the importance of accounting for tail behavior in financial risk modeling [18, 19].

### 4.4 Real-World Case Study: Daily Stock Returns

We analyze daily returns of a publicly traded stock over a six-month period. The returns are first tested for normality using the Shapiro-Wilk test and then fitted to both Normal and Lognormal distributions. Table 3 displays the resulting VaR and CVaR estimates.

**Table 3: Risk Metrics from Real Stock Return Data**

| Distribution Fit | VaR (95%) | CVaR (95%) |
|---|---|---|
| Normal | -2.1% | -2.7% |
| Lognormal | -2.4% | -3.4% |

The Lognormal model captures the asymmetric risk profile better than the Normal, reflecting the heavier downside risk that is typical in financial returns [4].

### 4.5 Results and Discussion

Figure 1 shows a histogram of event counts simulated from a Poisson distribution with rate parameter $\lambda = 3$. The discrete nature of the data is evident in the integer-valued support. This type of distribution is suitable for modeling operational risks and rare event counts in risk assessment frameworks.

Figure 2 illustrates a symmetrical bell-shaped curve typical of the Normal distribution. It is widely used in financial modeling and is often assumed in standard VaR and CVaR computations. However, its light-tailed nature may underestimate risk in volatile markets.

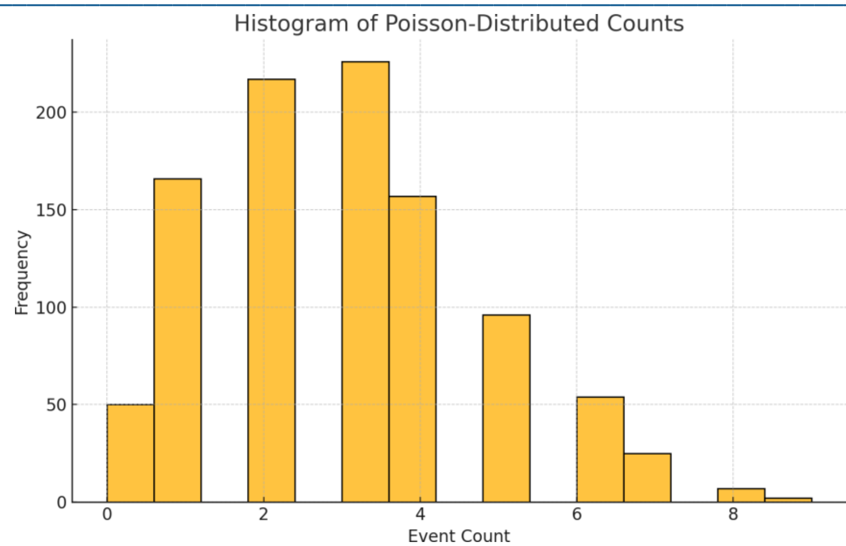


**Figure 1: Histogram of simulated Poisson-distributed event counts ($\lambda = 3$).**

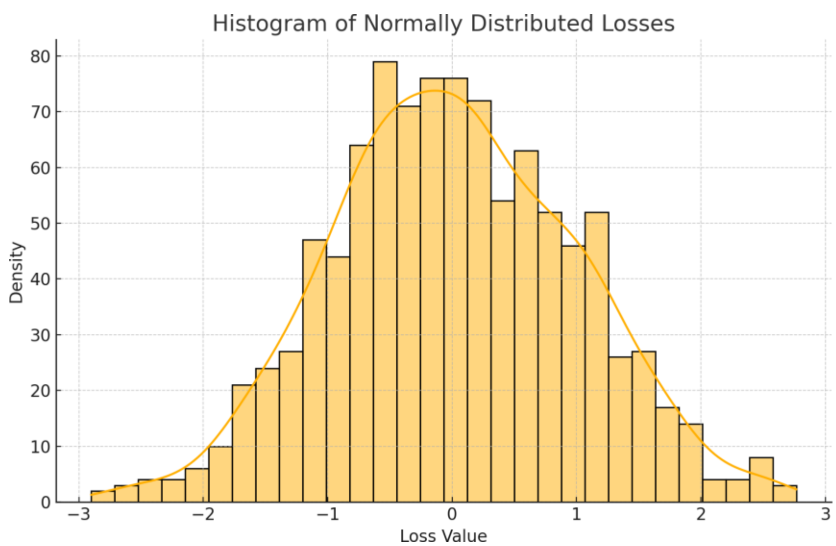


**Figure 2: Histogram of simulated losses from a Normal distribution ($\mu = 0$, $\sigma = 1$).**

Figure 3 shows a positively skewed distribution generated from a Lognormal model. It captures the heavy tails and asymmetric risk profiles often encountered in real-world financial or environmental data. Lognormal distributions are preferable when the loss data exhibit large positive outliers.

Figure 4 shows the estimated VaR and CVaR for the Normal distribution. While this distribution allows for closed-form estimation of both metrics, it tends to underrepresent tail risk due to its thin tails, making it less suitable in extreme loss scenarios.

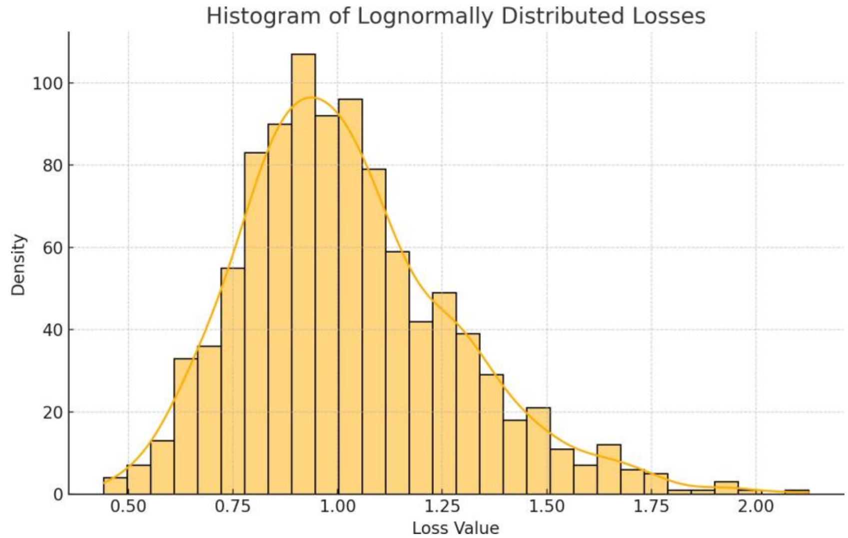


**Figure 3: Histogram of simulated losses from a Lognormal distribution ($\mu$ = 0, $\sigma$ = 0.25).**

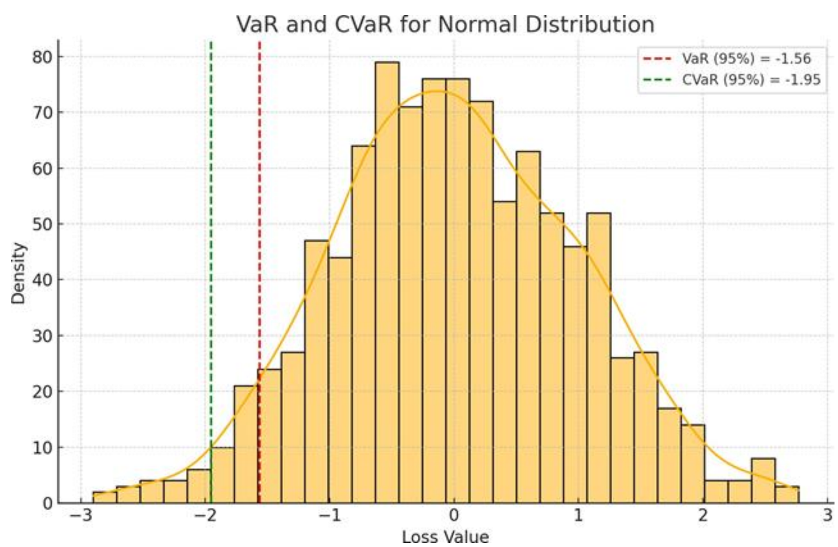


**Figure 4: VaR and CVaR at 95% level for losses modeled by a Normal distribution.**

Figure 5 demonstrates higher VaR and CVaR estimates when using a Lognormal distribution, which better reflects the possibility of extreme losses. This reinforces the argument for using heavy-tailed models in financial risk estimation.

The simulation and case study results affirm that the choice of distribution has a significant impact on the estimation of risk metrics. Discrete models are appropriate for count data but offer lower granularity in risk measures. Continuous models provide smoother and often more realistic estimates, particularly when

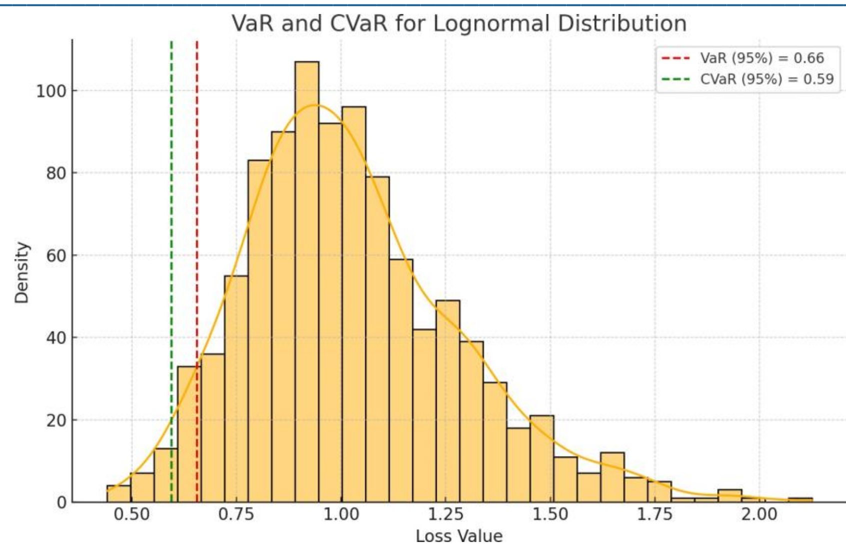


**Figure 5: VaR and CVaR at 95% level for losses modeled by a Lognormal distribution.**

heavy-tailed phenomena are involved.

Incorrect distributional assumptions may lead to a dangerous underestimation of risk, especially in systems sensitive to rare but impactful events [6, 9]. Hence, careful distribution selection, supported by statistical testing and validation, should be an integral part of any robust risk assessment framework.

## 5 Conclusion and Recommendations

This study explored the comparative role of discrete and continuous probability distributions in risk assessment models, emphasizing their impact on widely used risk metrics such as Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). Through theoretical analysis, simulation studies, and a real-world case application, we demonstrated that the choice of distribution significantly influences the accuracy and reliability of risk estimation.

Discrete distributions like Poisson and Binomial are well-suited for event-count modeling, particularly in operational risk and reliability studies. However, their discrete nature can limit resolution in risk metrics and complicate tail analysis. On the other hand, continuous distributions such as the Normal, Lognormal, and Gamma provide greater flexibility in modeling skewness, heavy tails, and continuous phenomena like financial losses and time-to-failure data.

Simulation results highlighted that misspecifying the distribution—especially by assuming normality in heavy-tailed contexts—can lead to severe underestimation of tail risk. This underlines the importance of validating distributional assumptions using statistical tests, exploratory data analysis, and domain knowledge. Based on the findings of this study, we offer the following recommendations for practitioners and researchers in risk assessment:

1. **Use data characteristics to guide distribution selection.** For count-based or event-driven data, consider discrete models. For continuous metrics such as monetary losses or durations, continuous models are more appropriate.

2. **Always test for distributional fit.** Employ statistical tools like Q-Q plots, goodness-of-fit tests (e.g., Shapiro-Wilk, Kolmogorov–Smirnov), and information criteria (e.g., AIC, BIC) to validate the

assumed distribution.

3. **Prefer heavy-tailed distributions in risk-sensitive applications.** When modeling extreme events or tail risks (e.g., in finance, disaster management), distributions like Lognormal, Pareto, or *t*-distributions should be prioritized over light-tailed ones.

4. **Account for distributional impact when reporting risk metrics.** Clearly communicate the underlying distribution used when presenting metrics like VaR or CVaR, and where possible, conduct sensitivity analysis using alternative models.

5. **Incorporate distribution selection into model governance frameworks.** In high-stakes environments such as banking or infrastructure risk management, ensure distribution assumptions are reviewed, documented, and regularly re-evaluated.

Future research may explore hybrid models that blend discrete and continuous elements (e.g., compound distributions), or apply machine learning-based approaches for distribution selection and density estimation. Additionally, domain-specific case studies in healthcare, climate science, or cybersecurity could reveal further nuances in distributional impacts.

This paper reinforces the critical importance of informed and context-sensitive distribution selection in building robust and interpretable risk assessment models.

**References**

[1] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, Springer, 1976.

[2] D. Nordgaard, *Probability Distributions for Use in Quantitative Risk Assessment*, Risk Analysis, **30**(3), 311–324, 2010.

[3] M. Modarres, *Reliability Engineering*, CRC Press, 2016.

[4] J. C. Hull, *Options, Futures, and Other Derivatives*, Pearson Education, 2012.

[5] R. T. Rockafellar and S. Uryasev, *Optimization of Conditional Value-at-Risk*, Journal of Risk, **2**(3), 21–41, 2000.

[6] R. T. Rockafellar and S. Uryasev, *Conditional Value-at-Risk for General Loss Distributions*, Journal of Banking & Finance, **26**(7), 1443–1471, 2002.

[7] S. Asmussen and P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Springer, 2007.

[8] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer, 2004.

[9] A. C. Cullen and H. C. Frey, *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*, Springer, 1999.

[10] C. J. Wu, *On the convergence properties of the EM algorithm*, The Annals of Statistics, **11**(1), 95–103, 1983.

[11] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 2004.

[12] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, 1970.

[13] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, 1991.

[14] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, Vol. 1–2, Wiley- Interscience, 2005.

[15] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 2002.

[16] G. Casella and R. L. Berger, *Statistical Inference*, 2nd Ed., Duxbury Press, 2002.

[17] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, Cengage Learning, 2011.

[18] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, 2015.

[19] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer, 2001.

[20] R. C. H. Cheng and W. Holland, *Two-point method for assessing variability in simulation output*, Journal of Statistical Computation and Simulation, **66**(3), 221–241, 2000.