

Application of Normal Distribution to Stock Exchange and Binomial Tree Model

Mengrui Han *

Victoria World Academy, Singapore, Singapore

* Corresponding Author Email: hanm96924@gmail.com

Abstract. Stock price modeling is a cornerstone of financial economics, in which a correct representation of price behavior is critical to asset pricing and risk management. This paper aims at discussing the use of the lognormal distribution in stock price models and provides the aim of elucidating its use in both discrete-time and continuous-time models. The study employs a probabilistic modeling methodology that involves theoretical study and introduces a geometric Brownian motion as the stochastic behavior of the stock prices. The work studies how the changes in the stock prices can be modeled as a multiplicative process over time by investigating the characteristics of lognormal distribution. The findings show that the stock prices are lognormally distributed, given the usual modeling assumptions, which assure of stock prices being non-negative and reflecting observed market behavior. Additionally, the discussion shows that the geometric Brownian motion is a sound theoretical basis of popular financial models. To sum up, it is necessary to state that the paper proves the usefulness and feasibility of the use of lognormal distribution in the financial models, especially in the valuation of assets and options.

Keywords: Normal distribution, likelihood function, stock exchange, binomial tree model.

1. Introduction

Normal distribution is one of the most significant and significant distributions in statistics. Its importance is primarily expressed in three ways; theoretical basis, practical implementation and analysis of data. The normal distribution is, theoretically, closely connected with the Central Limit Theorem [1]. According to this theorem, in large samples, no matter what the original data may be distributed like, the distribution of the sample mean will be close to a normal distribution. This renders the normal distribution to be the fundamental basis of statistical inference, including confidence intervals and hypothesis testing. Secondly, in the real world, normal distribution is predominantly found in natural and social phenomena. An example of such things is human height, test scores, measurement errors and returns on financial assets, which tend to be approximately normally distributed [2]. Accordingly, the normal distribution is an important tool in modeling and decision-making in other areas like finance (risk modeling), medicine (measurement error analysis), and engineering (quality control). Again, when it comes to data analysis methods, most traditional statistical methods (t-tests, regression analysis, ANOVA, etc.) assume that the data are more or less normally distributed. Normal distribution is symmetric and it is fully deferred by the mean and standard deviation thus the calculations and interpretations are a lot easier and more intuitive. Lastly, normal distribution too has great mathematical qualities (additive, stable, etc.) making it a cornerstone of probability theory and statistics [3].

The literature has extensively used Normal Distribution in diverse applications including crossing of electrical engineering and intelligent optimization algorithms, statistical modeling and its overlap with other areas including environmental science [4]. As an illustration, researchers suggested a multi-objective optimization algorithm, which is based on generalized normal distribution (MOGENDO), and implemented it to engineering optimization problems. Moreover, the statistical test (Wilcoxon test) is used to point out that the results have a great benefit. People also introduced a new probability distribution model - Slash Power Normal Distribution (SPND) and tested it on the analysis of environmental pollution data. It has been confirmed to work in the fitting of polluted data [5].

This paper will analyze the theory and applications of normal distribution. Section 2 will give the history and derivation of the normal distribution. In section 3, the applications of normal distribution will be provided. The conclusion will be given the last Section.

2. Normal Distribution

2.1. History of Normal Distribution

Normal distribution dates back to the 18-th century when Carl Friedrich Gauss in 1795 found the prime number theorem and the least squares method. Based on these, he constructed the Gaussian bell curve, which was later to be known as the normal distribution. The physical example of this is the Galton board experiment, where little balls are dropped through a stack of pins, where they will randomly hit at each level either to the left or the right [6]. In numerous experiments the balls will pile up in compartments to form a bell-shaped curve, symmetrical about the middle, which is the most common value, and the extremes of which are decreasing. This experiment is a graphic illustration of how a combination of many independent random processes can create a normal distribution, which can be seen in many natural and social phenomena, including exam scores or the distribution of height of a population.

The normal distribution is identified by being a symmetric, bell-shaped curve which has two parameters, one being the mean (μ), which is a measure of central tendency, and the other standard deviation (σ), which is a measure of the spread of data. The probability density function is zero at the end, and the area under the curve is equal to one. One important property is the empirical rule: about 68.3% of data are within one standard deviation of the mean ($\mu \pm \sigma$), 95.5% within two ($\mu \pm 2\sigma$), and 99.7% within three ($\mu \pm 3\sigma$). This enables effective making of probabilistic predictions regarding the outcome of measurements [7]. These properties play a crucial role in metrology to determine uncertainty in measurements and to measure compliance to specifications. As an example, the application of the normal distribution to measurement science is demonstrated by the fact that the probability that a product conforms to regulatory standards can be determined by how the data concerning the ratio of the packaging voids are distributed, as seen in the example given in the paper.

2.2. Derivation of Normal Distribution

Normal Distribution can be derived as follows. It is derived out of the analysis of random errors in measurements. Noting that the arithmetic mean of repeated measurements is generally the best estimate of the true value, Gauss tried to determine the error distribution in which the sample mean is the best estimator. Rather than making an assumption about distribution he obtained it using this optimality condition [8].

The former is to establish the likelihood function. Let the true value be θ and the error density function be $f(x)$. With a set of independent error terms $(x_1, x_2, \dots, x_n)$, the likelihood function is:

$$L(\theta) = \prod_{i=1}^n f(x_i - \theta) \quad (1)$$

Gauss supposed that the sample mean $\bar{x}$, is the best estimator $\hat{\theta}$ that maximizes this probability.

Secondly, the Likelihood should be maximized. The objective of maximizing $L(\theta)$ maximizing the logarithm. Differentiating the derivative with respect to and setting it to zero at yields:

$$\sum_{i=1}^n \frac{f'(x_i - \hat{\theta})}{f(x_i - \hat{\theta})} = 0 \quad (2)$$

Define $g(x) = f'(x)/f(x)$. The condition becomes $\sum_{i=1}^n g(x_i - \bar{x}) = 0$.

Thirdly, the Form of $g(x)$ has to be determined. Even function $f(x)$, with the symmetry of errors. By Lemma 1, $g(x)$ is odd: $g(-x) = -g(x)$, and $g(0) = 0$.

Gauss then chose a specific set of n error values: $x_1 = x_2 = \dots = x_m = -t$ and $x_{m+1} = mt$, with $n = m + 1$. For this set, $\bar{x} = 0$. Substituting into the summation yields:

$$m \cdot g(-t) + g(mt) = 0 \Rightarrow g(mt) = mg(t) \quad (3)$$

This functional equation is valid at all natural numbers, m , as well as real, t . The assumptions of continuity and differentiability mean that $g(x)$ is a linear function: $g(x) = cx$. Fourthly, Solving for $f(x)$. From $g(x) = f'(x)/f(x) = cx$, one can integrate:

$$\frac{d \ln f(x)}{d x} = cx \Rightarrow \ln f(x) = \frac{1}{2} cx^2 + C \quad (4)$$

Thus,

$$f(x) = M e^{\frac{1}{2} cx^2}, \quad \text{where } M = e^C \quad (5)$$

For $f(x)$ to be a valid density, it must integrate to 1, requiring $c < 0$ with $c = -1/\sigma^2$.

Finally, Normalizing the Density. Using the well-known Gaussian integral $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$, the constant is determined

$$\int_{-\infty}^{\infty} M e^{-x^2/(2\sigma^2)} dx = 1 \Rightarrow M = \frac{1}{\sigma\sqrt{2\pi}} \quad (6)$$

The last density function is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \quad (7)$$

It is noteworthy that this derivation uses reversed logic: rather than deriving a distribution by fitting the distribution to data, Gauss derived the distribution by a property of estimators he wanted to be satisfied with, namely that the sample mean should maximize the likelihood. This work formed the basis of the maximum likelihood estimation and made the normal distribution the core of statistical theory especially in the science of measurements and error analysis.

3. Applications

3.1. Stock Exchange

Stock price dynamics in financial markets is generally considered to be a stochastic process. The empirical research has revealed that stock prices are not allowed to be negative and that the changes in relative prices are randomly distributed. These properties match the lognormal distribution that is a distribution which is characterized by the fact that the natural logarithm of the random variable is a normal distribution. In particular when the logarithmic form of S , i.e., $\ln S$, is normally distributed $N(\mu, \sigma^2)$ is lognormally distributed. This is the reason why the lognormal distribution is especially appropriate to model asset prices: it inherently possesses positivity, and the fact that price changes are multiplicative is well represented [9].

Geometric Brownian motion, a special case of the Ito process, is usually used to model the behavior of stock prices. The stochastic different equation can be expressed as:

$$dS = \mu S dt + \sigma S dz \quad (8)$$

where S is the stock price, μ is the expected annual return, σ is the volatility, and z is a Wiener process. Using the Ito lemma, the price of the logarithmic distribution is normally distributed:

$$\ln S_T \sim N\left(\ln S + \left(\mu - \frac{\sigma^2}{2}\right)(T - t), \sigma^2(T - t)\right) \quad (9)$$

This means that the stock price S_T in the future will follow a lognormal distribution. On this basis, the Black-Scholes option pricing model presumes that the price of assets is lognormally distributed to obtain the price of European call options in terms of risk-neutral valuation.

There are two common applications of the lognormal distribution. First, it is applied in building future stock price confidence intervals. Given price S , average return μ , volatility σ and horizon $T - t$, the logarithmic price interval is initially computed using the normal distribution; the result is exp, and so the price interval can be obtained by exponentiating. Second, Black-Scholes formula of a European call option is given as:

$$C = S\Phi(d_1) - Xe^{-r(T-t)}\Phi(d_2) \quad (10)$$

where

$$d_1 = \frac{\ln(S/X) + (r + \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}}, \quad d_2 = d_1 - \sigma\sqrt{T - t} \quad (11)$$

is the cumulative distribution function of the normal distribution. These estimates are based on the assumption that the prices of stocks have a lognormal distribution.

The practical use shows that when the stock price is 100 yuan, the volatility of the stock is 20 percent, and the expected stock rate is 15 percent, the 95 percent interval of the price of the stock at the end of one month would be (90.20, 113.30) yuan. This gives a clear direction in which to trade, i.e., sell around the upper bound and buy around the lower bound. In the option pricing, the stock price is 84 yuan, the strike price is 80 yuan, the risk-free rate is 10 percent, the volatility is 20 percent and the expiration period is six months, the price of the call option will be around 9.52 yuan. These kinds of valuations assist investors in detecting mispriced options. In general, the lognormal distribution is a key instrument in financial modeling as it allows forecasting prices, estimating risk, and making well-informed investment decisions.

3.2. Stock Prices

In financial markets, the price of an asset, like a stock, exchange rate, or index, cannot be negative, and their time dependence has a tendency to follow a multiplicative dynamic. These properties make the lognormal distribution the natural distribution since, when the logarithm of a variable is normally distributed, the variable has a lognormal distribution. This characteristic renders the lognormal distribution especially appropriate when it comes to describing the behavior of stock prices when the eventual price is the cumulative product of a large number of independent random factors. The assumption of a lognormal distribution of asset prices is commonly used in empirical work and theoretical models in the field of options pricing, such as the widely used Black-Scholes option pricing model [10].

There are usually two stock price models that are regarded: discrete-time and continuous-time models. In discrete time, the ratios of successive prices $S(n)/S(n - 1)$ are assumed to be independent and lognormally distributed. This implies that the logarithmic returns $\ln(S(n)/S(n - 1))$ are normally distributed. The stock prices are modeled by the geometric Brownian motion in continuous time:

$$S(t) = S(0)e^{\mu t + \sigma B(t)} \quad (12)$$

where $B(t)$ is a standard Brownian motion. Taking logarithms gives $\ln S(t) \sim N(\ln S(0) + \mu t, \sigma^2 t)$, confirming that $S(t)$ follows a lognormal distribution. This formulation ensures positivity and captures the stochastic nature of price movements.

Lognormal distribution allows analysis of stock prices on probability basis by transforming to normal distribution. In discrete time models, the normal distribution of logarithmic returns is used to calculate probabilities of future prices. As an example, the likelihood that the price will rise in several periods will be found by considering the cumulative distribution of the normal variable. To derive the mean and variance of future prices, the author will use some equations in continuous time. Also, the

binomial tree model converges to the lognormal distribution when the number of steps is high, illustrating those discrete approximations to continuous geometric Brownian motion hold.

The applications provide realistic quantitative results. In the discrete time case, where the price ratios between the two dynamics follow the lognormal distribution, there is a chance that two consecutive weekly price gains occur of about 33.56 percent and that the chance of the price after two weeks being greater than the price at the beginning is 61.13 percent. In the binomial tree model, where there are 1000 steps, the likelihood of the price of the stock to increase by a minimum of 30 percent is about 99.78 percent. Considering a stock that has a starting price of 100, the probability that the price of the stock will be 103.05 after a year, its variance is 214.51. These calculations show that the lognormal distribution offers a consistent model of pricing options, probability estimation, and analysis of investment plans in discrete and continuous time.

3.3. Binomial tree model

Two basic option pricing models are the Black-Scholes and the binomial tree. Although in the Black-Scholes model, the returns on assets are normally distributed, then empirical studies have indicated that the returns of financial assets tend to be skewed and excessively kurtotic, which are not normally distributed. This drift results in the familiar effect of the volatility smile, implied volatilities change with strike prices, and the assumption of constant volatility is broken. This limitation is passed on to the standard binomial tree model which models the normal distribution using the binomial distribution. In order to overcome this problem, non-normal extensions have been created to incorporate the skewness and kurtosis of the actual distribution of asset returns, thus enhancing the accuracy of prices [11].

Standard binomial option pricing model: The underlying asset price is assumed to follow a binomial process. The probability of a risk-neutral P and the up and down factors u and d are as follows:

$$u = e^{(r-d-\sigma^2/2)\Delta t + \sigma\sqrt{\Delta t}}, \quad d = e^{(r-d-\sigma^2/2)\Delta t - \sigma\sqrt{\Delta t}}, \quad P = \frac{e^{r\Delta t} - d}{u - d} \quad (13)$$

The European call option price is then:

$$C_0 = e^{-rT} \sum_{i=0}^n \binom{n}{i} P^i (1-P)^{n-i} \max(S_0 u^i d^{n-i} - K, 0) \quad (14)$$

In non-normal distributions, to convert the binomial probability density $b(x)$ to a non-normal density $f(x)$, the Edgeworth expansion is applied and skewness and kurtosis are included:

$$\frac{f(x)}{b(x)} = 1 + \frac{\gamma_1(x^3 - 3x)}{6} + \frac{(\gamma_2 - 3)(x^4 - 6x^2 + 3)}{24} + \frac{\gamma_1^2(x^6 - 15x^4 + 45x^2 - 15)}{72} \quad (15)$$

where $(x = (2i - n)/\sqrt{n})$, (γ_1) is skewness, and (γ_2) is kurtosis. This adjusted density replaces the binomial probabilities at the terminal nodes.

Under non-normal distributions of the European options, the pricing mechanism merely replaces the usual binomial probabilities with the transformed Edgeworth probabilities. In the case of American options, though, the recursive backward through the tree is somewhat more complicated since the transformed probabilities are not directly applicable to each node. One of the decomposition approaches is suggested: the terminal node probabilities are divided by the number of paths to the node. Namely, the terminal node probability is broken down into the contribution of its two predecessor nodes, enabling the backward induction. The value of options in each node is the best of the immediate exercise and the discounted future value of the continuation. This recursion gives the price of the American option in non-normal distributions. Moreover, a nonparametric method substitutes the theoretical distribution with historical return distributions, further distributional assumptions.

The data of the KOSPI200 index options (January to April 2005), 851 put option samples are empirically validated. The skewness (-0.29876) and kurtosis (5.17829) of non-normal

distribution are based on the five-years of historical returns. The standard binomial model is compared to the errors of pricing of the non-normal binomial model. Hypothesis testing shows that the non-normal model produces significantly smaller pricing errors. The non-normal distribution captures the volatility smile effect observed in market prices, thereby improving the standard binomial model. This shows that when the normality assumption is relaxed to allow skewness and kurtosis, the price of options is more realistic especially in a market where asset returns are very far away in a normal distribution.

4. Conclusion

This study explored the application of the lognormal distribution in stock price modeling under both discrete-time and continuous-time frameworks. By incorporating probabilistic analysis and geometric Brownian motion, the research demonstrated that stock prices can be effectively described as lognormally distributed processes. The findings confirm that this modeling approach ensures non-negativity and aligns well with observed financial market behavior. Moreover, the study highlights the theoretical connection between lognormal distribution and widely used financial models, such as those applied in option pricing. Overall, the results suggest that the lognormal distribution provides a robust and practical foundation for modeling stock price dynamics.

For future work, one could try using other distributions instead of just the lognormal one, like the student's t or something that captures skewness and kurtosis better, as real stock returns are often not perfectly normal. Also, one could bring in actual market data, maybe even high-frequency data, to test how well these models really work in practice. It would also be interesting to compare the standard binomial tree with the non-normal version one talked about, under different market situations. And maybe one could even try using some simple machine learning methods to let the distribution parameters change over time. That might give us a more realistic and flexible way to price options and manage risk.

References

- [1] Chen M, Ma J, Leung Y. The slash power normal distribution with application to pollution data. *Mathematical Problems in Engineering*, 2022, 2022: 7086747.
- [2] Pandya S B, Jangir P, Mahdal M, et al. Optimizing brushless direct current motor design: An application of the multi-objective generalized normal distribution optimization. *Heliyon*, 2024, 10: e26369.
- [3] Wang Z G. Lognormal distribution and its application in the securities. *Journal of Suzhou Vocational University*, 2012, 23(3): 62-64.
- [4] An Z Q. A binomial option pricing model under non-normal distribution. *Lilun Xintan*, 2007(6): 8-10.
- [5] Yu Y. The application of the lognormal distribution in the stock price models. *Journal of Langfang Teachers College (Natural Science Edition)*, 2012, 12(5): 69-72.
- [6] Lü J. Application of normal distribution in medical insurance. *Reader Forum*.
- [7] Tai F, Yuan Y, Li Z. Study of teaching supervision evaluation based on normal distribution theory. *Innovation and Entrepreneurship Theory Research and Practice*, 2024(9): 21-24.
- [8] Li X, Xiao J, Zhang H, et al. Normal distribution screening method to improve near-infrared spectral stability. *Spectroscopy and Spectral Analysis*, 2025, 45(6): 1566-1577.
- [9] Li D, Wang N. Analysis of the practical significance of normal distribution in metrology. *Weighing Instrument*, 2023(10): 12-13.
- [10] Sun Q, Wang P, Zhang Z. Research on double differential velocity ratio method based on normal distribution: A case study of Changdao earthquake swarm. *Technology for Earthquake Disaster Prevention*, 2024, 19(1): 119-129.
- [11] Deng J, Liu C, Lao X. Design analysis of dimensional chain based on normal distribution law. *Aeroengine*, 2024, 50(1): 164-169.