Price Distribution Model for the Crypto Kitties Market in 2018 Proposal

by

Niket Rayapeddi

231765

International Christian University・国際基督教大学

ECO382 Advanced Studies in Economics I・ECO382 経済学特別講義 I

Correspondence: Niket Rayapeddi , International Christian University, 3-10-2 Osawa, Mitaka, Tokyo 181-8585, Japan. Tel: 080-9671-5999. E-mail: c231765n@icu.ac.jp

TABLE OF CONTENTS

Introduction 3

Literature Review 4

Theoretical Framework- The Log-normal Distribution Model………………..6

Data and Empirical Methodology

3.1 Data ……………………………………………………………………………7

3.2 Empirical Methodology…………………………………………………………..8

Results 9

Conclusion 12

References 14

Introduction

In 2018, a certain collection of Non-Fungible Tokens (NFT) known as CryptoKitties collection had earned up to a total of 20 million dollars off of customer spending shortly after it launched. NFTs are digital assets stored on the blockchain which have gained significant popularity since 2017 (Kapilkov, 2020). While NFTs serve many purposes, such as serving as digital representation of existing physical objects and its removal of the need of intermediaries, it can also be seen as a form of investment. Hence, this leads to the question of how the distribution of price behaves in the NFT market.

Could it be assumed that the price of NFTs behaves similarly to other forms of assets if it can be seen as a form of investment? Certain assets, such as real estate and stocks, tend to have a log normal distribution model best represent the histogram of their respective price variables (Sax, 2022). Assuming the price of NFTs behave similarly, the log-normal distribution would be the best fitting model for the NFT market. To test this hypothesis, the CryptoKitties NFT market in the year 2018 was examined to see whether or not the log normal distribution fits it the best. The year 2018 was selected as it was the year the CryptoKitties were the most prevalent in terms of number of sales and customers, which enabled the research to utilize a substantial amount of data.

This paper will provide an overview of the distribution model of price for the Crypto Kitties market as follows. The literature review glances at existing literature which analyze the NFT market, the price of NFTs, and the distribution of the price variable of several markets. The Theoretical Framework will then explain the log-normal distribution, a common distribution model for many assets and the assumed best fit distribution model for the Crypto Kitties NFT market. Finally, Data and Empirical Methodology state the dataset used in this paper along with the methodology used to estimate the fitness of the model to the dataset. The results of the distribution fitting and the following conclusion conclude the paper.

Literature Review

Mekacher et al, (2022), examined the differences in price of NFTs while considering its scarcity. The research examines several collections involving 1,479,020 published NFT data, between January 23 2018 and June 6 2022. These data points were collected from OpenSea, a marketplace for NFTs. Rarity of NFT within a collection is calculated as the sum of the rarity of all the individual traits. Individual trait rarity was calculated as a fraction of NFTs within a collection possessing this trait. Schaar and Kampakis (2022) also explored rarity and its contribution to price through looking specifically at 11,000 transactions with the CryptoPunks collection and made use of the hedonic regression for analysis. Both studies revealed how the rarity attribute of NFT does have a positive effect on the price. Furthermore, Schnoering & Inzirillo (2022) built on the previously mentioned research and utilizes the hedonic framework to calculate indices based on fitted hedonic regression lines, which are commonly used in real estate. Hedonic indices are used as proxies for price indices such as Consumer Price Index. These indices allow for them to observe dynamics and performance of the NFT market, which possesses a strong upward trend.

Vidal-Tomás (2022) looks at the specific niche of NFTs, the play-to-earn and metaverse tokens. With the data extracted from the CoinGecko database, they were able to analyze the co-movement in the tokens with the crypto-market using both Pearson and Kendall correlation.

Using the short term and long term performance of tokens, calculated through equations provided in the report, it was possible to compare the dynamics/performance of the tokens pre and post covid. The post covid long term performance of the metaverse and play-to-earn tokens was revealed to be positive, therefore, concluding that covid-19 didn't negatively affect the market.

Nadini et al, (2021) aims at characterizing statistical properties of the market in addition to building a network of interactions between sellers and buyers. The data was collected from NonFungible Corporation, a company which tracks historical NFT sales data. The NFTs were categorized into several categories where it was revealed that the Art, Collectible, and Games category dominate the market in the respective order. Price, Sale, and Collection distribution for each category mostly display a power law distribution. Network of trades was considered where a trader's strength, the total number of purchases and sales made by the trader, was found to be distributed as a power law. Similarly the Network of NFTs revealed how the strength of NFTs, defined by the number of times NFT was sold, also possess power law distribution. The collected and cleaned panel dataset utilized in the research includes various transactions involving NFTs from various markets spanning from the year 2017 to 2021. This dataset involved the price of NFTs sold, which can be utilized to create a histogram to analyze the price distribution of specific NFT markets.

Takaai et al, (2011) look at the cross sectional price distribution of housing prices in greater Tokyo. The results reveal how the distribution follows what was hypothesized and has a log-normal distribution where prices are asymmetric and skewed to the right. On the topic of price distribution, Sax (2022) examined how price distribution looks in the whole NFT market. Specifically it looks into what patterns emerge across various NFTs markets. Price distributions were found to be very concentrated across collections and had a consistent general shape. Both mentioned research inspire the idea of a temporal analysis of the price distribution in a specific NFT market. Findings from both research also inspire the hypothesis that the lognormal distribution is the best fitting model to the data due to its shape and utilization in other financial markets.

Theoretical Framework: Log-normal distribution Model

The log-normal distribution is a continuous probability distribution of a random variable in probability theory. The logarithm of the random variable will have a normal distribution if it has a lognormal distribution. Its key features include the fact that it has a lower bound of zero and that its distribution is right skewed, meaning it has a long right tail. The log-normal distribution is parameterized through two parameters. The first parameter is the location μ, which informs where the model is located on the x axis. The second parameter is the scale σ, which affects the shape of the distribution. The probability density function, or the functions which returns the likelihood of a random variable value appearing within a given range, for the log-normal distribution is given as follows in figure 1:

Figure 1: Probability Density Function of Log-normal distribution

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

Source:Kissel, R., Poserina J. (2017) Log-Normal Distribution Statistics. sciencedirect.com. https://www.sciencedirect.com/topics/mathematics/lognormal-distribution where:

σ stands for the scale parameter

μ stands for the location parameter

x is the random variable, or price

The log-normal distribution is a commonly used distribution for financial assets to describe share prices (Sax, 2022) . As previously mentioned, a substantial reason being that prices can't be negative. Stock prices are well known to be modeled with log-normal distribution due to their low mean and high variance. Therefore, assuming NFT prices also behave similarly to that of other financial asset prices, the log-normal model would best fit the price distribution.

Data and Empirical Methodology

Data

This paper analyzed the price distribution of the panel data of a specific NFT market (CryptoKitties) throughout the whole year of 2018. The price data was then retrieved from each of the 285,000 transactions of NFTs made throughout the entire year. The dataset was then split to twelve datasets for each month of the year of 2018 for monthly analysis. For the sake of simplicity, summary statistics for the entire year, as opposed to monthly statistics, was calculated and provided below in figure 2.

Figure 2: Summary Statistics

2018 CryptoKitties Prices (USD)	
count	285,000
mean	31.26
standard deviation	489.71
min	8.48e-16
max	17,3559

Source: Calculations of 2018 CryptoKitties Price data, based on panel data collected from OpenSea which includes various NFT markets spanning from 2017 to 2021

Data was collected from Opensea, a NFT marketplace. The summary statistics for the dataset is provided above in figure 1 for the key variable, price. The prices for the NFT were measured in U.S. dollars. As observed in figure 1, a lot of variance exists within the dataset in regards to the mean. The large contrast between the minimum price and the maximum price implies a wide range of prices for the CryptoKitties NFTs. The mean for the entire dataset is shown to be significantly low, given the wide contrast between the minimum price and maximum price. While the histogram would center around the low mean, the dataset is well spread out based on the high standard deviation and the range of values. The observed low mean and high standard deviation of the data follow the characteristics of the presumed best fit lognormal distribution model.

Empirical Methodology

The Fitter Library from Python is utilized to find the best distribution for the prices in the data for each month. A Fitter instance is created as defined below:

Fitter Instance (dataset, provided distribution, method) where:

dataset: monthly Crypto Kitties price dataset

provided distribution: list of common distributions to fit to the dataset such as gamma distribution, exponential distribution, and the hypothesized best fitting distribution model, the lognormal distribution

method: econometric technique to test the fitness of the models to the datasets. Sum of Squares and Akaike Information Criteria (AIC) is both used

Executing the instance will return the histogram with the different fitted models. It will also return a ranking of the top 5 fitting graphs with the respective sum of squares values. Running get_best method of the instance will return the best fitting model along with the parameters of the model.

AIC was used for the "goodness of fit" measure as it accounts for some distribution models, which have more complex functions and variables than others. Other measures, such as the sum of squares, will tend to have a lower error score if the model has more variables which would result in an unfair comparison. AIC penalizes models for such complexity, making it a more suitable measure. Similar to that of sum of squares, the lower the estimated AIC value, the better the fit to the data. Given that the stated hypothesis is correct, the returned fitness estimator value for the log-normal model should be the lowest amongst all the distribution models.

Results

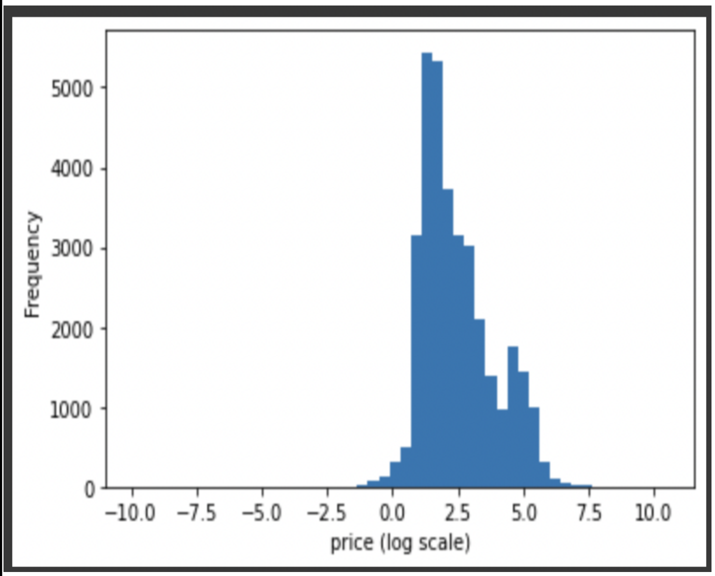
The implementation of Fitter Instance, the plotted histogram and distribution models, and the resulting best fit model based on two "goodness of fit" estimators for the month of January 2018 is reported below in figure 3.

Figure 3: January 2018 Price Histogram Distribution Fitting

January 2018 Price Histogram	Fitting Distribution Results

January 2018 Price Histogram

Histogram for Price Variable of January Dataset
x is log scale



Best Fitting model using AIC:

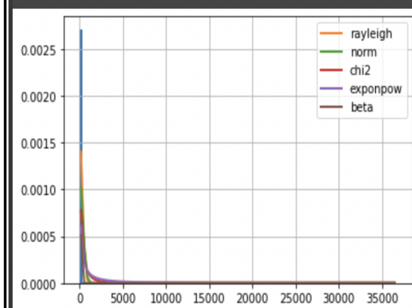
```
[52] f.get_best(method = 'aic')

{'uniform': {'loc': 0.0554568113425926, 'scale': 23511.919543188655}}
```

Fitting Distribution Results

Histogram with all fitted distribution models +
Sum Squared and AIC error for 5 of the distribution
models:

```
Fitting 12 distributions: 100%|██████████| 12/12 [00:14<00:00,  1.18s/it]
```

	sumsquare_error	aic	bic	kl_div	ks_statistic	ks_pvalue
rayleigh	0.000002	327155.134530	-804946.542434	inf	0.410812	0.0
norm	0.000003	331327.934571	-787606.412772	inf	0.454689	0.0
chi2	0.000004	4945.734788	-780028.116793	inf	0.323721	0.0
exponpow	0.000004	2692.202319	-774585.469202	inf	0.393010	0.0
beta	0.000005	46141.131314	-771944.612307	inf	0.190051	0.0



Best Fitting model using sum square:

```
f.get_best(method = 'sumsquare_error')

{'rayleigh': {'loc': -340.91762860269785, 'scale': 375.2705004828173}}
```

Source: Calculations of 2018 CryptoKitties Price data, based on panel data collected from OpenSea which includes various NFT markets spanning from 2017 to 2021

Figure 2 only displayed the analysis for January 2018. The same methodology was implemented for every month of the year. As displayed on figure 2, the histogram seems to display characteristics of the log-normal distribution, since both histogram on the left and right are visually similar to that of a log-normal distribution. However, when the goodness of fit measures were used to calculate the best fitting distribution, it returned rather astonishing results. Best fitting distributions for each month are displayed below in Figure 4:

Figure 4: Best Fitting Distributions for each Month using different estimators

Month	Using Sum of Squares	Using AIC
January 2018	rayleigh	uniform
February 2018	exponential	uniform
March 2018	beta	uniform
April 2018	beta	uniform
May 2018	exponential	uniform
June 2018	exponential	uniform
July 2018	beta	uniform
August 2018	exponential	uniform
September 2018	error	error
October 2018	rayleigh	uniform
November 2018	error	error
December 2018	rayleigh	uniform

Source: Implementations of methodology on 2018 CryptoKitties Price data, based on panel data collected from OpenSea which includes various NFT markets spanning from 2017 to 2021

The best fitting distribution model was calculated and recorded above in figure 3 for each month using both Sum of Squares and AIC estimator. September and November of 2018 results were not able to be returned due to the fitting of monthly data taking a significant amount of time which resulted in an error. Looking at both columns, it can be seen that the lognormal distribution does not seem to be the distribution of best fit for either of the estimators. The best distribution using the sum of squares are either rayleigh, beta, or exponential distribution, most of which are arguably more complex, in terms of function and number of parameters, than the lognormal distribution. While it is not the supposed best fit according to the sum of squares, the estimated fitness value for the lognormal was significantly low, coming under 0.001 for most of the months. This implies that while log-normal isn't the best fitting model for the monthly data, it is still a pretty accurate model due to its significantly low sum of square error score.

As mentioned previously, it was important to consider the fact that the fitness of the models to the data could be affected by the different complexities of the different distribution models. To account for such complexity, the AIC estimator is also used to calculate best fitting distribution. Unlike the sum of squares estimator, uniform distribution was returned as the best fitting model for most of the months. When the log-normal distribution AIC was compared to the uniform distribution AIC score for each month, a significant difference was observed, with there being at least a 1000 value difference. This would indicate that the uniform distribution serves as a better fit as the distribution model to the data compared to lognormal distribution by a large margin according to the AIC estimator.

Conclusion

This paper explored what the best fitting distribution model is to the price data of the CryptoKitties NFT market in the year 2018. The results revealed that the log-normal distribution is not the best fitting distribution to the histogram of each monthly data as revealed by the AIC estimator. This would conclude that the proposed hypothesis is rejected. An insightful finding was that the uniform distribution was shown to be the best fitting distribution model throughout most of the year. Further investigation needs to be made regarding why such is the case.

The results could have been limited by the errors in the months of September and November. However, it can be overlooked due to the uniform distribution being the best model for the rest of the months. It was also important to consider the possibility that the sample amount used to create histogram and distributions may have affected the results for the best fitting model. Using quarterly data as opposed to monthly data is a subject for follow-up research. Investigation in different NFT markets as well as different time periods can also be conducted and compared with the current results.

The applications of the findings rely on the fundamental concepts of probability distributions. In the field of finance, statistical assumptions are made on the price variable of different financial assets based on the type of probability distribution. For example, investors can predict the price of stocks based on the assumption that the price data takes the form of log-normal distribution. The results from this paper reveal that those same assumptions may not be applicable to the NFT market, given the fact that the best fitting distribution model turned out to be uniform and not log-normal.

List of References

Mekacher, A., Bracci, A., Nadini, M., Martino, M., Alessandretti, L., Aiello, L. M., & Baronchelli, A. (2022, October 16). *Heterogeneous rarity patterns drive price dynamics in NFT collections*. nature.com. https://www.nature.com/articles/s41598-022-17922-5

Nadini, M., Alessandretti, L., Giacinto, F. D., Martino, M., Aiello, L. M., & Baronchelli, A. (2021, October 22). *Mapping the NFT revolution: market trends, trade networks, and visual features*. nature.com. https://www.nature.com/articles/s41598-021-00053-8

Takaai, O., Takayuki, M., Chihiro, S. D., & Tsutomu, W. (2011, March 11). *Evolution of House Distribution*. DP. https://www.rieti.go.jp/jp/publications/dp/11e019.pdf

Sax, G. (2022, March 18). *Exploring NFT Price Distribution Across Collections*. medium.com. https://medium.com/@goblinsax/exploring-nft-price-distribution-across-collections-782193240f2c

Schaar, L., & Kampakis, S. (2022, January 18). *Non-fungible Tokens as an Alternative Investment:Evidence from CryptoPunks*. scholar.googleusercontent.com. https://scholar.googleusercontent.com/scholar?q=cache:wMGNE8eazjMJ:scholar.google.com/++journal+of+economic+perspective+nfts&hl=en&as_sdt=0,5

Schnoering, H., & Inzirillo, H. (2022, March 7). *Constructing a NFT Price Index and Applications*. https://arxiv.org/abs/2202.08966

Vasan, K., Janosov, M., & Barabási, A. L. (2022, February 17). *Quantifying NFT-driven networks in crypto art*. nature.com. https://www.nature.com/articles/s41598-022-05146-6

Vidal-Tomás, D. (2022, June 10). *The new crypto niche: NFTs, play-to-earn, and metaverse tokens*. sciencedirect. https://www.sciencedirect.com/science/article/abs/pii/S1544612322000630

Word count: 2930