Implementation of Monte-Carlo Method in Curriculum Efficiency, Cost Forecasting and Price Path Prediction

Qixuan Hu

Department of Mathematics, University College London, London WC1E 6BT, the United Kingdom

zcahqhu@ucl.ac.uk

Abstract. As a matter of fact, Monte-Carlo method has been widely implemented in various fields including distribution issues and stochastic process in recent years. With this in mind, this study mainly focuses the utilization of Monte-Carlo simulation onto some real-life scenarios. In retrospect, the Monte-Carlo simulation is a program-based algorithm which compute the possible outcome when the analysts input the samples. In general, it is used to give a portrait of the future for risk management in many fields. According to the analysis, the mechanics of using Monte-Carlo in some different cases is explained, analyzed and compared to judge the efficiency of using different distributions for an estimation. Moreover, some limits of using those are suggested and appealed for solutions to modify and improve the algorithm tool for future research. Overall, these results shed light on guiding further exploration of applications for Monte-Carlo simulation based on state-of-art techniques as well as broadening the implementation situations.

Keywords: Monte-Carlo method, computer science, statistics, finance.

1 Introduction

The world is filled with uncertainties. Every undertaking is the equivalence of a certain level of uncertainty [1]. The risk of an undertaking causes adverse impact, which may result in a loss of assets, credit, reputation, or potential customer of a company. Although it is not possible to eliminate the risk of a specific incident, prediction can be computed to avoid the inconsiderate outcome, or even control the loss for the inevitable consequence. Monte Carlo simulation is an effective tool of risk management to predict the possible outcomes for making decisions of an adequate amount of risk. Monte Carlo method is utilized in statistics to generate a model of variant results with different probabilities even if the process is not accurately predicted because random variables interfere with the task. This statistical tool was developed in 1940s by mathematicians to discover the effect of uncertainty on the Manhattan Project [2]. Monte Carlo simulation enable the analysts to model the risk or uncertainty, and then examine and evaluate the impact of the risk. To apply this method, a variable, or a multiple of variables are defined to represent the uncertainties. Then a random number is gen-

K. Elbagory et al. (eds.), *Proceedings of the 9th International Conference on Financial Innovation and Economic Development (ICFIED 2024)*, Advances in Economics, Business and Management Research 281,
https://doi.org/10.2991/978-94-6463-408-2_40

erated by the random number function using program commands in purpose to compute the outcome with respect to that specific number. The same simulation is executed for multiple times, and the average is taken for an accurate result.

Monte Carlo simulation is widely used in many fields such as finance, physics, and biology, computer science. In finance, Monte Carlo simulation is used to manage the risk of loss in an investment project. For a similar approach, theories in some fields are examined or demonstrated by using the Monte Carlo method. To extend the Monte Carlo simulation, some further algorithms such as random walk and Markov chain is developed and applied to improve the accuracy of the prediction.

The introduction of Monte Carlo simulation has raised the importance of uncertainty and accessed the impact of uncertainty in many daily life scenarios, such as project management, deep learning of artificial intelligence, forecasting prices of financial derivatives. Different from the normal way of forecasting, Monte Calo simulation usually computes a set of output by a given set of fixed inputting samples. To be more specific, the fixed sample is chosen by the analysts or even generated by the random number program, so that each sample has a certain and fixed value to be executed. To apply the Mone Carlo simulation, an appropriate statistical model is first chosen by identifying the independent variables (known as the input, usually carries a certain level of uncertainty) and dependent variables (known as the output). The statistical model should strike to the probability distribution such as uniform distribution, normal or exponential distribution (these are the basic mathematical distribution applied frequently in many fields of study). Then samples of fixed values are generated by a program, or even the historical data under the decision of the analyst is repetitively utilized to promote the simulation for a more accurate result. Sometimes both real historical and program generated samples are used for simulation, but a weight is set to judge the result. Normally, a single Monte Carlo simulation involves a hundred thousand samples, or even more, to make sure the accuracy of the result, due to the Law of Large Number. When the simulation is completed, a mean value is output as the result of prediction, and a program enable the analyst to further derive the variance (or standard deviation, commonly used to describe the spread or the convergence of the output) to give a portrait of the range where the realistic value might strike to.

There are many advantages in Monte Carlo simulation with fixed samples, such as the capacity to develop a sensitivity portrait of a given model. Sensitivity analysis explains how the outcome is shaped by different kinds of uncertainties of in a model. This regression method can be used to understand the relationship between input and output of a specific model, identify the vital factor which cause the risk, and further provides a modification of the model to reduce uncertainty. There are no constraints when simulating by Monte Carlo method [3]. According to professor Wade, the limitation of using Monte Carlo simulation is equivalent to the limitation of the tools used for the simulation, such as the distribution model. Finally, the only limitation of the Monte Carlo simulation is the pseudorandom sample generated by computer program. To deal with it, operation with quantum computing is vital to cover the shortage of the technology at this stage. Hence this addressed problem will not be discussed.

The motivation of this article is to review some applications of Monte Carlo simulation in different fields, and make a comparison of the condition of applying the simulation between different scenarios, and analyze the mechanism of the method. Finally, some limitation and potential improvements to the method are concluded. For the rest of this article, section 2 is the analysis of the several implementations of Monte Carlo; section 3 discusses the limitation of Monte Carlo simulation, and appeals some potential improvement for further outlooks; the last section draws a conclusion of the analysis suggested in this article.

2 Case Analysis

2.1 Curriculum Efficiency

In this case study, the focus is the curriculum and course exams contributed to the successful graduation for college students. The Monte Carlo simulation is used in this case to estimate the rate of successful graduation of students [4]. In this model, it is assumed that the final exam of all courses must be passed to graduate. Moreover, specific course is not an option to be signed up if the pre-course is failed. Then the incident such that a student pass or fail the exam can be generated by random numbers, and then it is compared to the pass rate among all the students, given by the institution (for example, a pass rate of 0.6 means 60% of students taking the course pass the exam) to determine whether a student pass or fail the exam. Analytically, the independent variable in this situation is the number representing the status of the exams taken for each course, and so, the dependent variables is the time taken, measured in semester, to graduate from the college.

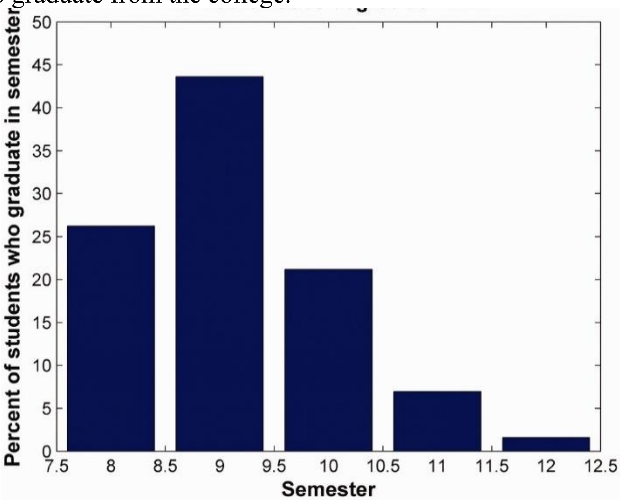

Fig. 1. A relative frequency distribution [4].

The Monte Carlo algorithm is as follows. First, a set of samples of the independent variable is generated using random number program. Next, each piece of sample is compared to the pass rate: if the number is less than the pass rate of the course, then the student has higher rank, so it is a pass, and vice versa; if there is a failure, then coursed is taken once again, and the fail rate is halved in this situation. When a course is passed, it soon gets into the next course. During proceeding of the algorithm, the time taken to pass each courser is recorded, and finally, a relative frequency distribution is output as a graph which presents the number of students graduate by semester. The result is as shown in Fig. 1 [4]. It is well noticed that a quarter of the students will graduate without any failure in the course exams; about a half (42.5%) of the students fail once in the exams; and the rest of the students fail at least two courses. To further understand a student academic performance, another algorithm is used to increase the passing probability for a retake of failed course [5]. Instead of a halved fail rate for the retake as in the first scenario, a new parameter is defined as the mean increment of passing the retake to adjust the probability for passing the course. The performance is simulated using Monte Carlo method with this parameter varying to check for the sensitivity. As a result, as the increment increase from 30% to 50%, the result varies by 2.2%; an increase from 50% to 80% produces a variation of 1.7% [5], which is not sensitive factor contribute to the performance of a student. Another algorithm uses the normal distribution to set up the model. A historical data is cited to generate histograms, and the performance of the exam is normalized so that the mean, the standard deviation may be derived. However, the result doesn't change much, which means the choice of mathematical model does not play an important role for accuracy [6] (Fig. 2).

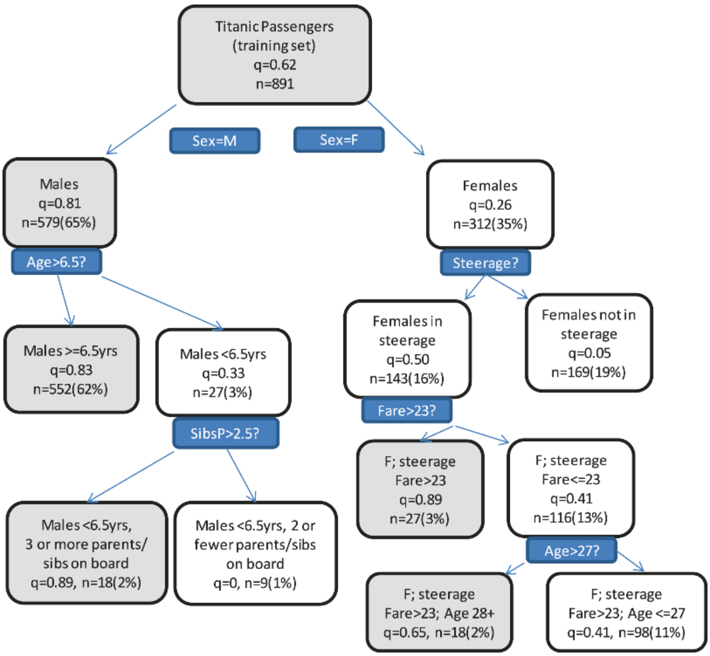

Fig. 2. A sketch of the tree results [7].

2.2 Forecasting Hospital cost

The second scenario focuses on the cost prediction. It is important to estimate the cost required to run a company in preventing the potential risk such as the liquidity risk. In the hospital, costs needed to proceed surgeries and surgical procedures are dependent on some factors such as the time taken of the surgery and a serious of treatment or care to the patients. The tool aiming to predict the cost of surgical care is not fully developed in the hospital administration perspective, thus it is vital to create and modify an appropriate model for forecasting the expenditure with a higher accuracy [7].

The patient data (age, surgery time, etc.) used in this estimation was a historical registry of patients at the Teaching Hospital of GZO in the period from 1st January 2013 to 31st December of 2019 [7]. The model used for forecasting is the random forest algorithm. To be more specific, it is an algorithm as a general-purpose classification and regression model approaching the estimation by combining random decision trees and sum up the prediction by averaging [8]. It is a branch of decision trees constructed and derivation towards the estimation, which to say, a further refinement of classification and regression tree (CART) models [9]. An example given by Steven J. Rigatti is the RMS titanic sinking incident [9]. See the following chart cited from the article 'random forest' [9]. It is clear that the 891 passengers are separated into different branches by status of each passenger representing by probabilities. The final estimation is generated using program with multiple executions. Back to be hospital scenario, a similar algorithm is created to separate the patients by their age, gender, time taken for surgery, etc. The different types of costs are then associated to them. Using python, it is possible to calculate the mean and standard deviation of the hospital cost (seen from Fig. 3 [7]).

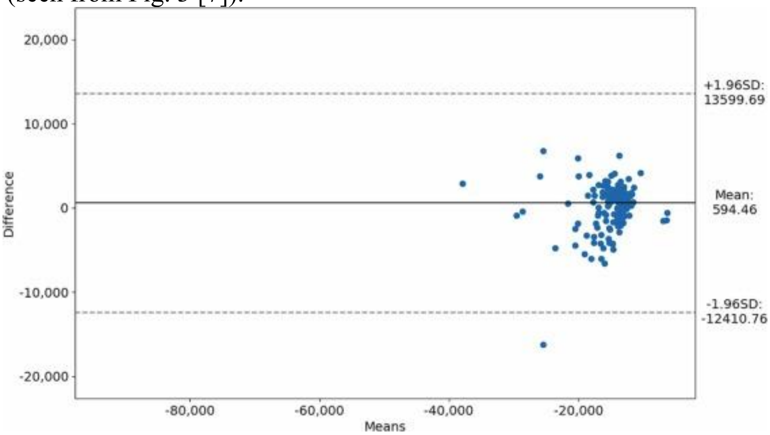

Fig. 3. The scatter diagram of the difference.

2.3 Price Path Prediction

The function of geometric Brownian motion in this part is to predict a short-term trend of the share price of firms. GBM is a random process with continuous time measure to forecast the option price. It is applied for an analysis of the scenario of the real stock prices in Bursa Malaysia through the maximum number of days of closing prices required for predicting the stock prices by the accuracy of the model [10], aiming predict a short-term investment, thus geometric Bronioan motion is used instead of the Markov-Fourier grey model that for long term investment. To collect the data used for simulation, the closing prices are recorded for each day in a period of a month. It is able to derive some useful parameters for the return distribution: the percentage growth of asset (which is the return on investment, measured by the difference of price over the price at the previous day), drift (measured by the mean of the return), volatility (measured by standard deviation of the return). A differential equation is set for this distribution and it is then integrated to derive the asset price equation which is the main equation involving in predict the asset value at a specific time. In another scenario, the same model is utilized to predict the dynamic path of the oil price. Some useful properties in geometric Brownian motion enables closed form solutions to evaluate the asset [11]. The historical data shapes the expected growing rate of oil price, follows by the instantaneous standard deviation, which measures the range of variation of price with respect to the horizontal time.

3 Limitations and Prospects

Although the scenarios in the previous part all generate reasonable result, there are still some limitations towards these models or method. In the algorithm that simulate the graduation of college students, the way to judge whether a student's pass or not is by a randomly generated number in the range 0 to 1 and compare it to the pass rate. The distribution for the algorithm is uniform distribution to some extent. However, the original way to judge the graduation is whether the mark earned by a student in the exam is above the minimum requirement. In this situation, the performance fits the normal distribution, so that the independent variable becomes the random percentage generated randomly, and thus it is possible to find the mean and standard deviation. This is kind of another approach for estimation. The main limitation of using Monte Carlo estimation is the requirement of the representative historical data which is vital in shaping the algorithm numerically. In the scenarios mention above, the accuracy depends the most on the data acquired from the previous time. In addition, the sample used should not be up-to date. Hence, the algorithm limits its accuracy for estimating if the quantity of samples is not enough.

4 Conclusion

The implementation of Monte Carlo simulation becomes popular on predicting the future in many fields. In this article, several scenarios were analyzed and different distributions, or models are especially useful in some condition, although there are limitations unavoidable such as the quantity and the timeliness of the historical samples used. The key point to improve the algorithm follows from the limitation: to compare the different model using the same sample and check for the accuracy to decide the most efficient model. For other approaches, it is also available to develop new mathematical tools for settling the Monte Carlo method.

References

1. Pascu, P.: Economic Risks for Project Management: a Monte Carlo Simulation Approach. USV Annals of Economics & Public Administration, 23(1), 128–133 (2023).
2. Goodman, D. H.: A Hybrid Approach to Determining Company-Specific Risk: Using Monte Carlo Simulation. Value Examiner, 12–17 (2022).
3. Blanchett, D., Pfau, W.: The Power and Limitations of Monte Carlo Simulations[J]. Advisor Perspectives, VettaFi, 8, 26 (2014).
4. Torres, D., Crichigno, J., Sanchez, C.: Assessing Curriculum Efficiency Through Monte Carlo Simulation. Journal of College Student Retention: Research, Theory & Practice, 22(4), 597-610 (2021).
5. Caro, E., González, C., Mira, J. M.: Student academic performance stochastic simulator based on the Monte Carlo method. Computers and Education, 76, 42–54 (2014).
6. Huang, S., Fang, N.: Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. Computers & Education, 61, 133-145 (2013).
7. Ochs, V., Tobler, A., Enodien, B., et al.: Development and validation of a predictive model of the hospital cost associated with bariatric surgery. Obesity Research & Clinical Practice (2023).
8. Biau, G., Scornet, E.: A random forest guided tour. Test, 25, 197-227 (2016).
9. Rigatti, S. J.: Random forest. Journal of Insurance Medicine, 47(1), 31-39 (2017).
10. Abidin, S. N. Z., Jaffar, M. M.: Forecasting share prices of small size companies in Bursa Malaysia using geometric Brownian motion. Applied Mathematics & Information Sciences, 8(1), 107 (2014).
11. Postali, F. A. S., Picchetti, P.: Geometric Brownian motion and structural breaks in oil prices: a quantitative analysis. Energy Economics, 28(4), 506-522 (2006).