ISSN 2616-7433 Vol. 4, Issue 9: 44-48, DOI: 10.25236/FSST.2022.040910

Broad View of Computational Statistics

Ruiying Li[1,*], Fan Yan[2]

[1]*School of WMG, University of Warwick, Coventry, CV4 7AL, United Kingdom*
[2]*School of Finance and Economics, Shanghai University, Shanghai, 200083, China*
[*]*Corresponding author email: ruiying.li129@gmail.com*

Abstract: *The emergence of computational statistics could be identified as the second statistical revolution. From this point, the development of statistics depends on the advancement of computer science. This paper tried to summarise computational statistics from four sides: computational statistics drivers and motivations, the difference between computational statistics and traditional statistics characteristics, essential computational statistical techniques, and application and use cases. In particular, the critical computational statistical techniques section discussed Monte Carlo, bootstrap, graphical, and randomization methods.*

Keywords: *Computational Statistics, Traditional Statistics, Monte Carlo Methods (MCM), Bootstrap Methods, Graphical Methods, Randomization Methods*

1. Introduction

The emergence of computer science and information technology is regarded as the second statistical revolution [1]. The evolution of statistics and computers are tightly intertwined. Especially recently, the majority of statistical developments rely on advances in computer science [2]. In light of the extensive numerical calculations and the increased need for data presentation, computational intensity is now the most appealing statistical strategy [2]. The topic of 'computational statistics' is an area of statistical approaches that grew as a result of the advent of contemporary computers and aims to identify information inside non-numerical and numerical data. In addition to computationally expensive statistical techniques, computing statistics have included statistical computing techniques [3,4]. The latter relates to statistical procedures made possible by computational technologies. For instance, software engineering and database methodology are examples of common statistical computing techniques. Computational statistics depend on statistical computing, applied statistics and mathematical statistics, which links to a vast array of contemporary statistical methods [3]. This article presents a comprehensive overview of computational statistics, including its motives, properties, fundamental methods, and applications.

2. Computational Statistics Drivers and Motivations

The emergence of advances in big data, computer science and statistical computing prompted the formation of Computational Statistics. The evolution of statistical theory has been assisted by developments in statistical computing. Symbolic computing programs have been developed specifically to decrease error creation and simplify mathematical deduction [3]. Additionally, simulation facilitates rapid investigation of viable solutions [2]. The concept of statistical computation has played a significant role in the evolution of several professions. In addition, advancements in computer science, particularly the creation of digital technology and upgrading in software and hardware, have altered the form and manner of the everyday job of statisticians. Researchers employ an increasing number of computers, and statisticians depend on them more and more. Therefore, computers play a greater role in statistical research, including data storage, processing, and publishing [3]. These technologies drastically increase the efficiency of regular analysis and decrease the amount of expertise statisticians must possess. According to Gentle, the advent of computers has caused a paradigm change in statistical analysis [4]. The abundance of information is one of the fundamental features of the digital era. During the study, statisticians must record and evaluate massive amounts of data, which considerably increases the complexity of statistical job. By enhancing the pace of data mining, computers aid statisticians in accelerating the processing and analysis of data. As a result of the aforementioned circumstances, Computational Statistics was formed.

ISSN 2616-7433 Vol. 4, Issue 9: 44-48, DOI: 10.25236/FSST.2022.040910

3. Differentiate from Traditional Statistics Characteristics

In classical statistical analysis, statisticians employ both knowledge from other domains and observational data to build various models. Through these models, people may comprehend and analyze the evolution of trends and the collection of data [4]. Statistical methods have pursued a paradigm that use data analysis to choose or revise models, define suitable variations, and extrapolate processes for millennia [3]. Even though the concept of computational mathematics has been greatly improved ever since 1930s, that it wasn't until 1980 that the field of statistics gained widespread popularity. This popularity is mostly due to the fast development of computer technique [2]. Due to increases in processing capacity and the growth of computers, it has been possible to employ increasingly complicated statistical procedures. The exponential development in processing power has benefited the statistical investigation. It has resulted in more accurate and precise models. Additionally, large-scale simulation testing has been permitted. This circumstance led to the development of statistical computing. Computers' immense processing capacity and ability to present findings graphically expand the scope of statistical analysis. On this basis, Computational Statistics was created. Unlike conventional statistics, computational statistics has the following two characteristics:

• It requires extensive computing effort: It is challenging for traditional statistical mathematics to cope with very precise models. Due to this constraint, statisticians must rely on approximation or estimating approaches to fulfil their objectives. The evolution of computing has altered this circumstance. Computation reasoning is the exact simulator that flawlessly resolves this issue and serves as a beneficial alternative strategy [4]. On the basis of conventional estimation methods, it can perform a multitude of computations fairly fast and present the results graphically.

• The essence of scientific instruments: The primary components of the scientific method are theories and experiment. The computer's capacity to quickly examine several situations represents a unique kind of instrument. One of the unique characteristics of computer engineering is the ability to simulate several possibilities and offer information to investigators [2]. It is comparable in some respects to the experiment results described above. However, instead of doing actual experiments, the observation inference is derived by the assumption model [5]. In addition to conducting computations and storing data, computers may also play a more embedded methods in computational statistics, such as recommending alternative theories and models for researchers [5].

4. Key Techniques

4.1 Monte Carlo Methods (MCM)

Monte Carlo is among the most well-known experimental approaches in computational statistics, which uses simulated random numbers to estimate the posterior distribution of specific functions [6]. Since it employs random numbers, execution of this method requires a random number source [7]. These numbers may be pseudo-random, meaning they are predictable yet seem randomized [7]. Due to their high operational viability, pseudorandom numbers are often used in scientific research. In addition, Monte Carlo Methods may be used to situations that lack stochastic elements. Using the MCM method, these issues may be handled by estimating future values, similar to random variable function problems [6]. It should be highlighted that Carlo study does not only replicate random sampling. Given that the overall purpose is to estimate a certain quantity associated with a particular random variable [8], it is essential that the chosen sample faithfully represent the dispersion of the simulation population [6]. Since a result, a sort of pseudo sample is created to satisfy the aforementioned conditions, as pseudorandom or really random selection may be impacted by random fluctuations and cannot accurately represent the pattern of the simulation community [7].

4.2 Bootstrap Methods

Bootstrapping is among the representative resampling techniques [9]. It collects a unique sample from the target population [3]. One of the key concepts of bootstrapping resampling is to consider the observed sample to represent the population, because all accessible information about the base population is included in the example [10]. From this vantage point, researchers can utilize a random selection that takes into account original samples and imitates any implicated test standard deviation [3]. Thus, this method is effective irrespective of the level of understanding well about base distribution. Efron introduced one of the fundamental bootstrapping methods, which recommended applying the variations

ISSN 2616-7433 Vol. 4, Issue 9: 44-48, DOI: 10.25236/FSST.2022.040910

to examine the outlier within the same sample [11]. In generally, the bootstrapping approach significantly depends on recreating the original population's variance [3]. Using different types of time series forecasting to map the aerial trend and the interdependence in bootstrapping datasets is the first stage in reproducing variance [10]. Rising linear and polynomial methods might be used to eliminate trends throughout this procedure [3]. Several experts have also conducted study in this field. For instance, Sherman and Carlstein described the use of diagnostic charts using block histograms to evaluate the success of blocking systems [12].

4.3 Graphical Methods

Graphical statistical approaches are applicable computer statistics techniques [13]. Visualization is the fundamental step in comprehending data [14]. Diverse graphical approaches offer useful tools for not only reflecting the qualities of datasets, but also assisting researchers in identifying the key correlations between variables. Particularly for statistical data assessment, graphical representation is crucial [13]. As indicated by Milici et al., the constant development of advanced equipment and high-speed processors has made graphics increasingly viable [14]. Particularly with the increased resolution, the visuals are rendered more attractively [14]. Therefore, images may disclose even more crucial aspects than other approaches. The objective of the study influences the graphical design style, whilst the number of factors and observations might influence the building of the graphical presentation. Because information cannot be shown readily on a two-dimensional surface, the data presentation in much more than 2 components involves different transformation techniques, such as projections [13]. The most direct effect that computer progress has had on graphical statistical approaches is the simplification and acceleration of visual presentation [14].

4.4 Randomization Methods

The randomized testing is a very efficient technique. As it does not depend on assumptions on possible probability distributions, it has a broad variety of applications. As its core concept, the randomization test compares the observed configuration with all different possibilities [9]. It is important to note that such randomization test may be beneficial without the need for a significant simple test statistic underneath an assumption of interest [4]. Nonetheless, this approach is dependent on the entire data designing phase [15], which necessitates that any random data be respected throughout data collection and sample design. In straightforward uses of a randomization test, it is believed that all outcomes are equally likely to occur under the assumptions 15]. This assumption will be rejected if the observed outcome belongs to a subgroup with a lower probability [9]. Assessing the real - time of two datasets and determining whether they are equivalent is an example of a randomization strategy. Due to the sample observations, the evaluation should be conducted using both regimens [16]. In this instance, either the number of observations or the difference between the means of the samples performed better than the median or mean as a whole [16]. Indeed, any of the aforementioned data might serve as test statistics for the randomization approach.

5. Application Areas and Use Cases

5.1 Financial: Market Risk Assessment

Financial market is a significant application of computer science [17]. In particular, computational statistics plays a crucial role in building an efficient evaluation method for market hazard, such as the calculation of Value at Risk (VaR). Market risk is the economic volatility resulting from an unforeseen change in market prices and inflation [18]. VaR measurement began in the early 1990s. It is a common market risk appraisal approach developed from computational statistics that measures the greatest loss that the present portfolio is expected to experience over a particular time period with a certain amount of certainty [19]. Its two primary features have made it the international standard for risk assessment. VaR gives an initial consistent assessment across various risk variables and situations [18]. Secondly, VaR takes into consideration the connection between many influencing elements and is simple to compute [17]. However, it has several flaws, including non-subadditivity [19]. To address these deficiencies, some investigators, have offered different alternative approaches, such as sufficient risk measurements.

ISSN 2616-7433 Vol. 4, Issue 9: 44-48, DOI: 10.25236/FSST.2022.040910

5.2 Economic: Econometrics Development

The development of computer statistics advanced econometrics greatly by enhancing the effectiveness of data processing and assessment. With the evolution of the internet era, many facets of contemporary human behavior are watched. In a digitalized culture, everyone's actions and behavior cannot escape leaving a digital footprint [20,21]. Through time series, a continual quantity of information was generated in this situation. Econometricians may get these abundant, high-quality, and current data sources at a minimal cost [20]. These data resources may be utilized to investigate and resolve several unresolved economic issues. During this process, the growth of computational statistics gives new analytic models and more precise assessment tools, therefore enhancing studies [20]. Particularly, the fast growth of personal computers is important for processing vast quantities of complicated data [20]. Numerous economic models monitor both apparent and latent variables using statistical computing techniques. For instance, the SV (stochastic volatility) model is applied to calibrate asset return volatility [21,22]. The computational intelligence statistics has made a substantial contribution to the advancement of econometrics.

5.3 Medicine: MRI

Since its invention in the 1970s by Lauterbur and Mansfield, neuroimaging (MRI) technology has had a profound impact on medicine [23]. MRI's principal goal is to detect the function of the brain [24]. There are currently more over 20,000 MRI scans in use for medical reasons across the globe [23]. Radioactive tracers are unnecessary for MRI [24]. Researchers may now, for the first time, investigate brain activity under live settings, which is revolutionary [25]. During the investigation, participants were required to do certain cognitive activities or physical motions. In the meanwhile, magnetic resonance scans of the brain were acquired for the purpose of studying functionality to structural mapping. The study of cognitive development under physiologic and pathologic settings [24], the research of brain damage [23], and so on, have contributed significantly to the advancement of brain functional dynamics. MRI is used outside the medical field. Even geneticists may use this method to examine comparable or dissimilar genes in sophisticated ways. The growth of such a sector necessitates substantial enhancements to computational statistical approaches, as it is accompanied by an increase in the number of data sets and an increase in demand for data gathering and calculating skills.

6. Conclusion

This article concludes with a comprehensive overview of computational statistics. The evolution of statistics, computer technology, and big data might be seen as three crucial forces. In contrast to classical statistics, two characteristics of computational statistics are intense processing and the nature of discovery tools. MCM, randomization, bootstrap, and graphical approach are described as the core methodologies, and their financial, economic, and medical applications are also examined.

References

[1] Büning, H. and Naeve, P. (2019) Computational Statistics. Berlin, Boston: De Gruyter. https://0-doi-org.pugwash.lib.warwick.ac.uk/10.1515/9783110844405
[2] Bekker, A., Chen, D.-G. and Ferreira, J. T. (2020) Computational and methodological statistics and biostatistics : contemporary essays in advancement. Springer (Emerging topics in statistics and biostatistics).
[3] Beserra, A. A. R., Souza, L. C. and Souza, D. F. L. (2020) 'Bootstrap Analysis of Compression Algorithms', IEEE Latin America Transactions, Latin America Transactions, IEEE (Revista IEEE America Latina), IEEE Latin Am. Trans, 18(09), pp. 1639–1645. doi: 10.1109/TLA.2020.9381807.
[4] Gentle, J. E. (2009) Computational statistics. Edited by M. J. Evans. Springer, New York
[5] Härdle, W., Okhrin, O. and Okhrin, Y. (2017) Basic elements of computational statistics. Springer (Statistics and computing).
[6] Barbu, A. G. and Zhu, S. C. (2020) Monte Carlo methods. Springer
[7] Dunn, W. L. and Shultis, J. K. (2012) Exploring Monte Carlo methods. Elsevier/Academic Press.
[8] Kalos, M. H. and Whitlock, P. A. (1986) Monte Carlo methods. Wiley.
[9] Mackinnon, J. G. and Webb, M. D. (2019) Wild bootstrap randomization inference for few treated clusters. Emerald Group Publishing Ltd. (Advances in Econometrics). doi: 10.1108/S0731-905320190000039003.

ISSN 2616-7433 Vol. 4, Issue 9: 44-48, DOI: 10.25236/FSST.2022.040910

[10] Vaish, J., Datta, S. S. and Seethalekshmi, k (2021) 'Short-term Load Forecasting using Bootstrap Aggregation based Ensemble Method', 2021 7th International Conference on Electrical Energy Systems (ICEES), Electrical Energy Systems (ICEES), 2021 7th International Conference on, pp. 245–249. doi: 10.1109/ICEES51510.2021.9383755.

[11] Efron, B. (1987) 'Better Bootstrap Confidence Intervals', Journal of the American Statistical Association, 82(397), pp. 171–185. doi: 10.2307/2289144.

[12] Sherman, M. and Carlstein, E. (1996) 'Replicate Histograms', Journal of the American Statistical Association, 91(434), pp. 566–576. doi: 10.2307/2291653.

[13] Gentle, J.E. (2002) Elements of Computational Statistics. Springer.

[14] Milici, M-R. et al. (2017) 'Graphical Method for Evaluating and Predicting the Human Performance in Real Time', Soft Computing Applications : Proceedings of the 7th International Workshop Soft Computing Applications (SOFA 2016) , Volume 1. (Advances in Intelligent Systems and Computing. 633), p. 327. doi: 10.1007/978-3-319-62521-8_27.

[15] Beerenwinkel, N. and Siebourg, J. (2012) Probability, statistics, and computational science. (Methods in Molecular Biology). doi: 10.1007/978-1-61779-582-4_3.

[16] Lange, K. (2010) Numerical analysis for statisticians. Edited by D. Calitoiu. Springer, New York (Statistics and Computing).

[17] Habart-Corlosquet, M., Janssen, J. and Manca, R. (2013) VaR methodology for non-gaussian finance. ISTE Ltd./John Wiley and Sons Inc. (Focus series in finance, business and management).

[18] Pan,Y. and Yan, F. (2010) 'The risk control of pledge liquidation in logistics finance based on VaR methodology', 2010 International Conference on Logistics Systems and Intelligent Management (ICLSIM), Logistics Systems and Intelligent Management, 2010 International Conference on, 1, pp. 135–138. doi: 10.1109/ICLSIM.2010.5461453.

[19] Xu, C. and Shiina, T. (2018) Risk management in finance and logistics. Springer (Translational systems sciences: volume 14)

[20] Wooldridge, J. M. (2020) Introductory econometrics. a modern approach. Seventh edition.

[21] Baltagi, B. H. (2008) Econometrics. 4th ed. Springer-Verlag Berlin Heidelberg.

[22] Seddighi, H., Lawler, K. A. and Katos, A. V. (2000) Econometrics : a practical approach. Routledge.

[23] Hashemi, R. H., Bradley, W. G. and Lisanti, C. J. (2018) MRI. the basics. 4th ed. Wolters Kluwer Health

[24] Ulmer, S. and Jansen, O. (2020) fMRI : basics and clinical applications. Third edition. Springer (Online access with purchase: Springer).

[25] Dale, B. M., Brown, M. A. and Semelka, R. C. (2015) MRI : basic principles and applications. Fifth edition. Wiley Blackwell