symmetry

MDPI

Editorial

Introduction to the Special Issue in *Symmetry* Titled "Symmetry in Statistics and Data Science"

Christophe Chesneau

Department of Mathematics, LMNO, University of Caen-Normandie, 14032 Caen, France; christophe.chesneau@unicaen.fr

In order to introduce this Special Issue, some motivational facts are given. To begin, the concept of symmetry is a fundamental idea in statistics and data science that underpins many important techniques and methods. Symmetry refers to a property of an object or distribution that remains the same when subjected to a transformation, such as a rotation, translation, or reflection.

Clearly, one of the most important applications of symmetry in statistics is in the study of probability distributions. A distribution is said to be symmetric if it is identical when reflected around a vertical line that passes through its center. Symmetric distributions have many useful properties, such as having the same mean, median, and mode. This makes them particularly useful for statistical inference, where we often need to estimate these quantities from a sample of data.

The concept of symmetry also plays a crucial role in data science and machine learning. In many applications, we are interested in finding patterns or relationships in high-dimensional data sets. One way to accomplish this is by identifying symmetries in the data that can be exploited to reduce the dimensionality or simplify the analysis. For example, in image recognition tasks, we can take advantage of the symmetry properties of images to develop algorithms that are robust to rotations and translations.

Moreover, symmetry has implications for the interpretability of machine learning models. If a model is symmetric with respect to a certain transformation, then it is invariant to that transformation, meaning that the output does not change when the input is transformed. This can make the model more interpretable, as it allows us to focus on the essential features of the input that are invariant under the transformation.

In conclusion, the concept of symmetry is a fundamental idea in statistics and data science that forms the basis of many significant techniques and methods. Understanding symmetry allows us to identify and exploit patterns and relationships in the data, which can lead to more robust and interpretable models. As such, it is a concept that is essential for anyone working in these fields to understand.

This Special Issue was created to collect the most significant modern research dealing more or less directly with this concept. Symmetry can be at the heart of the subject, a property of a major tool, etc. In the end, twelve articles were published in this Special Issue after a thorough peer-review process involving top editors and reviewers. They are authored by respected and prestigious academic researchers from all over the world. Brief summaries of these articles are presented below.

1. The first article, "Computing Expectiles Using k-Nearest Neighbours Approach", by Farooq et al. [1] proposes to investigate the k-nearest neighbors approach, together with the asymmetric least squares loss function, called ex-kNN, for computing expectiles. The ex-kNN is then examined in various configurations. Comparisons are given with the existing method in the literature, demonstrating the importance of the new method. High-dimensional data are considered to illustrate this aspect.
2. The second article, "Bayesian Reference Analysis for the Generalized Normal Linear Regression Model", by Tomazella et al. [2] suggests using the Bayesian reference analysis to estimate the generalized normal linear regression model's parameters. The reference

Citation: Chesneau, C. Introduction to the Special Issue in *Symmetry* Titled "Symmetry in Statistics and Data Science". *Symmetry* **2023**, *15*, 1103. https://doi.org/10.3390/sym15051103

Received: 8 May 2023
Accepted: 16 May 2023
Published: 18 May 2023

prior produced the correct posterior distribution, whereas the Jeffreys prior produced the incorrect one, as the study demonstrates. To support the conclusions, data are analyzed.

3. In the third article entitled "Estimating the Variance of Estimator of the Latent Factor Linear Mixed Model Using Supplemented Expectation-Maximization Algorithm", Angraini et al. [3] introduce the Supplemented EM (SEM) algorithm for the case of fixed variables. The computational aspects of the SEM algorithm have been investigated by means of simulation. Then, real data on political attitudes and behaviors in Flanders (Belgium) are considered.
4. The fourth article, "Spatial-Temporal 3D Residual Correlation Network for Urban Traffic Status Prediction" by Bao et al. [4], innovates by proposing a novel spatial-temporal 3D residual correlation network (ST-3DRCN) for urban traffic status prediction. Diverse data sets are analyzed, showing how the new model outperforms its rivals.
5. In the fifth article entitled "Sec-G Class of Distributions: Properties and Applications", Souza et al. [5] develop a new general class of distributions centered around the functionalities of the secant function, called the Sec-G family. By considering the Kumaraswamy–Weibull distribution as the baseline, the corresponding Sec-G distribution demonstrates a high level of flexibility in practice, making it able to accommodate various kinds of lifetime data.
6. In the sixth article "On Odd Perks-G Class of Distributions: Properties, Regression Model, Discretization, Bayesian and Non-Bayesian Estimation, and Applications", Elbatal et al. [6] propose a modification of the Perks distribution to create the odd Perks-G class of distributions. Its discrete version is also examined. To underline the significance and applicability of the suggested models, real-world data sets are used.
7. In their article entitled "Lagrangian Zero Truncated Poisson Distribution: Properties Regression Model and Applications", Irshad et al. [7] create the Lagrangian zero truncated Poisson distribution (LZTPD), a novel type of discrete Lagrangian distribution. In addition, the Lagrangian zero truncated Poisson regression model (LZTPRM), a novel count regression model, and its parameters are computed using the maximum likelihood estimation approach. Healthcare data are evaluated to show the LZTPRM's superiority, and two real-world data sets are taken into consideration to show the LZTPD's practical applicability.
8. The eighth article, "Statistical Inference of Weighted Exponential Distribution under Joint Progressive Type-II Censoring" by Qiao and Gui [8], contributes to the understanding of the weighted exponential distribution. The authors examine the estimates of parameters of the weighted exponential distribution with the joint progressive type II censoring data. A simulation has been run to evaluate the viability of the aforementioned techniques, and real data analysis is also included as an example.
9. The ninth article, "Analysis of Adaptive Progressive Type-II Hybrid Censored Dagum Data with Applications" by Mohammed et al. [9], seeks to investigate classical and Bayesian estimations of the Dagum distribution based on the adaptive progressive type II hybrid censoring scheme. Two real data sets are also looked at to show how the suggested point and interval estimators can be used.
10. In their article entitled "Robust Procedure for Change-Point Estimation Using Quantile Regression Model with Asymmetric Laplace Distribution", Yang [10] creates a non-iterative sampling process from a Bayesian perspective and offers a robust change-point estimation method based on a quantile regression model with an asymmetric Laplace error distribution. After conducting a simulation study to demonstrate the procedure's effectiveness with encouraging results, real data are studied to demonstrate the algorithm's utility in comparison to the traditional change-point detection approach based on normal regression.
11. The eleventh article, "A New Detection Function Model for Distance Sampling Based on the Burr XII Model" by Alanzi et al. [11], provides a new flexible parametric model based on the functionalities of the Burr XII distribution. The inferential aspect of the model is investigated and tested using two real-world data sets.

12. In the last article, "A Collection of Two-Dimensional Copulas Based on an Original Parametric Ratio Scheme", Chesneau [12] explores new copulas of the ratio type on the theoretical side. Each of them can be the object of a distribution generation, with a new perspective in terms of dependence modeling. Their properties are established, including various copula orderings and correlation measures.

We hope that this Special Issue will encourage future generations to make significant and novel discoveries about the concept of symmetry in statistics and data science.

Thanks to the success of this Special Issue, a second volume will be planned, with the aim of collecting fresh and significant articles on the subject. The best efforts will be made to produce a great volume.

Acknowledgments: We want to express our gratitude to all the authors, assistants, associate editors, and editors for their great work on this Special Issue.

Conflicts of Interest: The author declares that they have no conflict of interest to report regarding the present study.

References

1. Farooq, M.; Sarfraz, S.; Chesneau, C.; Ul Hassan, M.; Raza, M.; Sherwani, R.; Jamal, F. Computing Expectiles Using k-Nearest Neighbours Approach. *Symmetry* **2021**, *13*, 645. [CrossRef]
2. Tomazella, V.; Jesus, S.; Gazon, A.; Louzada, F.; Nadarajah, S.; Nascimento, D.; Rodrigues, F.; Ramos, P. Bayesian Reference Analysis for the Generalized Normal Linear Regression Model. *Symmetry* **2021**, *13*, 856. [CrossRef]
3. Angraini, Y.; Notodiputro, K.; Folmer, H.; Saefuddin, A.; Toharudin, T. Estimating the Variance of Estimator of the Latent Factor Linear Mixed Model Using Supplemented Expectation-Maximization Algorithm. *Symmetry* **2021**, *13*, 1286. [CrossRef]
4. Bao, Y.; Shi, Q.; Shen, Q.; Cao, Y. Spatial-Temporal 3D Residual Correlation Network for Urban Traffic Status Prediction. *Symmetry* **2022**, *14*, 33. [CrossRef]
5. Souza, L.; de Oliveira, W.; de Brito, C.; Chesneau, C.; Fernandes, R.; Ferreira, T. Sec-G Class of Distributions: Properties and Applications. *Symmetry* **2022**, *14*, 299. [CrossRef]
6. Elbatal, I.; Alotaibi, N.; Almetwally, E.; Alyami, S.; Elgarhy, M. On Odd Perks-G Class of Distributions: Properties, Regression Model, Discretization, Bayesian and Non-Bayesian Estimation, and Applications. *Symmetry* **2022**, *14*, 883. [CrossRef]
7. Irshad, M.; Chesneau, C.; Shibu, D.; Monisha, M.; Maya, R. Lagrangian Zero Truncated Poisson Distribution: Properties Regression Model and Applications. *Symmetry* **2022**, *14*, 1775. [CrossRef]
8. Qiao, Y.; Gui, W. Statistical Inference of Weighted Exponential Distribution under Joint Progressive Type-II Censoring. *Symmetry* **2022**, *14*, 2031. [CrossRef]
9. Mohammed, H.; Nassar, M.; Alotaibi, R.; Elshahhat, A. Analysis of Adaptive Progressive Type-II Hybrid Censored Dagum Data with Applications. *Symmetry* **2022**, *14*, 2146. [CrossRef]
10. Yang, F. Robust Procedure for Change-Point Estimation Using Quantile Regression Model with Asymmetric Laplace Distribution. *Symmetry* **2023**, *15*, 447. [CrossRef]
11. Alanzi, A.; Jamal, F.; Tahir, M.; Chesneau, C.; Kanwal, S.; Sami, W. A New Detection Function Model for Distance Sampling Based on the Burr XII Model. *Symmetry* **2023**, *15*, 620. [CrossRef]
12. Chesneau, C. A Collection of Two-Dimensional Copulas Based on an Original Parametric Ratio Scheme. *Symmetry* **2023**, *15*, 977. [CrossRef]