

Comparing Robust Methods: A Review of Techniques and Applications

Checchin Andrea*

Department of Mechanical and Energy Engineering, Beijing University of Technology, Beijing 100124, China

Introduction

Robust statistical methods have gained significant attention in recent years due to their ability to provide reliable results in the presence of outliers and model deviations. This review aims to explore various robust techniques, their theoretical foundations, and practical applications across different domains. By categorizing methods into parametric, non-parametric, and semi-parametric approaches, we will analyze their strengths and limitations. The review concludes with recommendations for future research directions in robust statistics. In statistical analysis, traditional methods often rely on assumptions of normality and homoscedasticity. However, real-world data frequently violate these assumptions, leading to unreliable results. Robust statistical methods have emerged as essential tools to handle these violations effectively. This article reviews various robust techniques, their applications, and their performance relative to classical methods.

Description

Robustness in statistics refers to the quality of a method to remain relatively unaffected by small deviations from model assumptions. A robust estimator is one that provides reliable parameter estimates even when the data includes outliers or is not normally distributed. Influence function measures the effect of a small change in the data on the estimator. A robust estimator has a bounded influence function. Breakdown point is the proportion of contaminated data that can be present before the estimator ceases to provide meaningful results. High breakdown points indicate better robustness. Asymptotic behavior of estimators as the sample size approaches infinity is crucial for evaluating their robustness. M-estimators generalize Maximum Likelihood Estimators (MLEs) and are defined as solutions to the equations derived from the method of moments. They are robust against certain types of data contamination. Commonly used in regression analysis, M-estimators can provide reliable estimates in the presence of outliers. R-estimators focus on rank-based methods that do not rely on the distributional assumptions of the data. They are particularly useful in cases where the underlying distribution is unknown. R-estimators are widely used in non-parametric statistics and are applicable in fields such as biostatistics and economics. The median is the most basic robust estimator of central tendency. It has a breakdown point of 50%, making it highly robust to outliers. Trimmed means involve removing a certain percentage of the lowest and highest values before calculating the mean, while Winsorized means replace extreme values with the nearest values that remain [1].

*Address for Correspondence: Checchin Andrea, Department of Mechanical and Energy Engineering, Beijing University of Technology, Beijing 100124, China, E-mail: andrea@che.cn

Copyright: © 2024 Andrea C. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received: 24 September, 2024, Manuscript No. jbmbs-24-154729; Editor assigned: 26 September, 2024, Pre QC No. P-154729; Reviewed: 10 October, 2024, QC No. Q-154729; Revised: 15 October, 2024, Manuscript No. R-154729; Published: 22 October, 2024, DOI: 10.37421/2155-6180.2024.15.231

GAMs combine parametric and non-parametric elements, allowing for flexibility in modeling complex relationships between variables. They are extensively used in ecological modeling, where the relationships between predictors and responses are often non-linear.

Comparison of robust methods

Strengths: Flexible, can be tailored to specific data characteristics.

Limitations: Sensitive to the choice of tuning parameters.

R-estimators

Strengths: Less sensitive to distributional assumptions, suitable for small samples.

Limitations: May have lower efficiency compared to M-estimators under normal conditions.

Strengths: Do not assume a specific data distribution.

Limitations: Can be less efficient in large samples and may require larger sample sizes for reliable estimates.

Semi-parametric methods

Strengths: Flexibility in modeling complex relationships while retaining some parametric structure.

Limitations: Complexity in implementation and interpretation.

Mean Squared Error (MSE): Measures the average of the squares of the errors—that is, the average squared difference between estimated values and actual values.

Bias: Assesses how far the estimator is from the true parameter value.

Variance: Measures the variability of the estimator [2].

Robust methods are extensively applied in econometrics, particularly in regression analysis where outliers can distort results. Techniques such as robust regression provide reliable estimates of relationships between economic variables. Extending robust methods to multivariate contexts is crucial. Developing robust techniques for multivariate regression, clustering, and principal component analysis will provide deeper insights into complex datasets. In clinical trials, the presence of outliers can skew results. Robust statistical methods, including M-estimators and rank-based methods, are commonly used to analyze clinical data to ensure the validity of conclusions drawn from the studies [3].

In ecological studies, data may often include extreme values due to environmental variability. Robust methods such as GAMs and non-parametric techniques are used to model species distributions and assess the impact of environmental factors. Many robust methods involve complex calculations that can be computationally intensive. Future research should aim at developing faster algorithms and approximations that maintain robustness while improving efficiency. The rise of big data presents unique challenges for traditional statistical methods. Research should explore robust techniques that can efficiently handle large datasets, especially those with high dimensionality and potential outliers. In manufacturing and quality control, robust statistical techniques are used to analyze process data. Techniques such as trimmed means and robust regression are essential for maintaining product quality in the presence of outlier measurements [4].

Many robust techniques involve complex calculations, which can be a barrier to their widespread adoption, particularly in real-time applications. Choosing the appropriate robust method for a given application can be challenging. Developing automated tools for model selection would be a valuable contribution to the field. The integration of robust statistical methods with machine learning techniques is an emerging area of research. Developing robust algorithms that can handle large datasets with outliers will be crucial for future applications. Integrating robust statistical techniques with machine learning algorithms can enhance their performance in handling real-world data. Research should focus on algorithms that can automatically adapt to the presence of outliers and noise. Promoting awareness and understanding of robust statistical methods in academia and industry will encourage their adoption in practice. Educational resources and workshops can play a pivotal role in this endeavor [5].

Conclusion

Robust statistical methods offer valuable tools for analyzing data that deviate from traditional assumptions. By understanding and applying these techniques, researchers can obtain reliable results that are less affected by outliers and model misspecifications. This review highlights the diversity of robust methods, their strengths and limitations, and their applications across various fields. Future research should focus on simplifying these methods, enhancing computational efficiency, and integrating them with modern data analysis techniques.

Acknowledgement

None.

Conflict of Interest

None.

References

1. Souto, Leonardo AV, André Castro, Luiz Marcos Garcia Gonçalves and Tiago P. Nascimento. "Stairs and doors recognition as natural landmarks based on clouds of 3D edge-points from RGB-D sensors for mobile robot localization." *Sensors* 17 (2017): 1824.
2. Yu, Jingrui, ZhenZhen Xiang and Jianbo Su. "Hierarchical multi-level information fusion for robust and consistent visual SLAM." *IEEE Trans Veh Technol* 71 (2021): 250-259.
3. Zhou, Haoyu, Zheng Yao and Mingquan Lu. "Lidar/UWB fusion based SLAM with anti-degeneration capability." *IEEE Trans Veh Technol* 70 (2020): 820-830.
4. Liu, Yongshan, Dehan Kong, Dandan Zhao and Xiang Gong, et al. "A point cloud registration algorithm based on feature extraction and matching." *Math Probl Eng* 2018 (2018): 7352691.
5. Park, Gwangsoo, Byungjin Lee, and Sangkyung Sung. "Integrated pose estimation using 2d lidar and ins based on hybrid scan matching." *Sensors* 21 (2021): 5670.

How to cite this article: Andrea, Checchin. "Comparing Robust Methods: A Review of Techniques and Applications." *J Biom Biosta* 15 (2024): 231.