

Distribution Freedom Analysis of Non-parametric Statistics

Zhihao Zheng

David Game College, UK
2722936505@qq.com

Abstract

This article focuses on the in-depth analysis of the distribution freedom of non-parametric statistics. Firstly, the basic concept of non-parametric statistics was expounded. Then, the connotation and essence of distribution freedom were deeply explored, and its key role in statistical inference was analyzed. Through correlation analysis with other characteristics in non-parametric statistics and comparison with parametric statistical methods, the unique advantages of distribution freedom are highlighted. Meanwhile, this paper explores the value and challenges of distribution freedom in practical applications, aiming to provide theoretical references for a comprehensive and in-depth understanding of the distribution freedom of non-parametric statistics and promote the application and development of non-parametric statistical methods in a wider range of fields.

Keywords: Non-parametric statistics; Distribution freedom; Statistical inference; Parameter statistics

1. Introduction

1.1 Definition and Connotation of Non-parametric Statistics

Nonparametric Statistics is an important branch of mathematical statistics, and its core lies in making statistical inferences without relying on the specific form of the overall distribution [1]. In most statistical problems, the overall distribution form is often unknown or difficult to precisely characterize with a limited number of parameters. At such times, the application of parametric statistical methods is greatly restricted. Non-parametric statistical Methods do not require strict assumptions about the overall Distribution. They only construct statistics based on the information of the sample data itself, such as rank, frequency and other non-parametric features, for inference. Therefore, they are also called distribution-free methods.

For instance, when comparing the mean differences of two independent samples, parametric statistical methods typically assume that both populations follow a normal distribution and have the same variance, and then make inferences through methods such as t-tests. However, when the overall distribution does not conform to the normal distribution or the variances are unknown and unequal, the results of the t-test may be unreliable. At this point, non-parametric statistical methods such as the Wilcoxon rank sum test determine whether there are differences in the distribution positions of two samples by comparing their rank sums, without making any assumptions about the overall distribution, and thus have wider applicability [2].

1.2 The Development History of Non-parametric Statistics

The formation of non-parametric statistics can be traced back to the 1940s. After decades of development, it has become a large-scale, theoretically profound and practically valuable branch of statistics [3]. Early research mainly focused on exploring the non-parametric properties of some simple statistics, such as sign tests and rank tests. With the continuous development of statistical theory and the impetus of practical demands, non-parametric statistical methods have gradually become enriched and improved.

In the middle of the 20th century, a series of important non-parametric test methods were proposed, such as the Kolmogorov-Smirnov test and the Kruskal-Wallis test, etc. [4], which laid the foundation for the application of non-parametric statistics in practical research. In recent years, with the rapid development of computer technology, non-parametric statistical methods have been more widely applied and deeply studied in many fields such as big data analysis, biomedicine, and social sciences. For instance, in biomedical research, non-parametric statistical methods are employed to analyze gene expression data, clinical trial data, etc., providing crucial statistical support for the diagnosis and treatment of diseases [5].

2. Overview of Non-parametric Statistics

2.1 Definition of Non-parametric Statistics

Distribution freedom refers to the fact that non-parametric statistical methods do not rely on the specific form of the overall distribution when making statistical inferences [6]. That is to say, regardless of the distribution that the population follows, non-parametric statistical methods can provide reasonable inference results based on sample data. This characteristic makes non-parametric statistical methods more adaptable and robust when dealing with data of unknown distribution or that does not conform to the assumption of parametric distribution.

The essence of distribution freedom lies in the way non-parametric statistical methods utilize data information. Parameter statistics methods map the information of sample data to the estimation and inference of population parameters by assuming

the form of population distribution. Non-parametric statistical methods, on the other hand, directly start from non-parametric features such as the rank and frequency of the sample data to mine the information contained in the data. Take the rank test as an example. It sorts the sample data by size and assigns the corresponding rank, and then builds the test statistic based on the information of the rank. This processing approach avoids reliance on the overall distribution, making statistical inference more based on the characteristics of the data itself, thereby achieving distribution freedom [7].

2.2 The Development History of Non-parametric Statistics

The idea of non-parametric statistics can be traced back to the late 19th and early 20th centuries. Early research mainly focused on exploring the non-parametric properties of some simple statistics. With the continuous development of statistical theory and the impetus of practical demands, non-parametric statistical methods have gradually become enriched and improved. In the middle of the 20th century, a series of important non-parametric test methods were proposed, such as sign tests and rank tests, which laid the foundation for the application of non-parametric statistics in practical research. In recent years, with the rapid development of computer technology, non-parametric statistical methods have been more widely applied and deeply studied in many fields such as big data analysis, biomedicine, and social sciences [8].

3. The Connotation and Essence of Distribution of Freedom

3.1 Definition of Distribution Freedom

Distribution freedom refers to the fact that non-parametric statistical methods do not rely on the specific form of the overall distribution when making statistical inferences. That is to say, regardless of the distribution that the population follows, non-parametric statistical methods can provide reasonable inference results based on sample data. For instance, in a non-parametric rank test, the distribution of the test statistic is only related to the sample size and is independent of the specific shape of the overall distribution. This characteristic makes non-parametric statistical methods more adaptable and robust when dealing with data of unknown distribution or that does not conform to the assumption of parametric distribution.

3.2 The Essence of Distribution Freedom

The essence of distribution freedom lies in the way non-parametric statistical methods utilize data information. Parameter statistics methods map the information of sample data to the estimation and inference of population parameters by assuming the form of population distribution. Non-parametric statistical methods, on the other hand, directly start from non-parametric features such as the rank and frequency of the sample data to mine the information contained in the data. Take the rank test as an example. It sorts the sample data by size and assigns the corresponding rank, and then builds the test statistic based on the information of the rank. This processing approach avoids reliance on the overall distribution, making statistical inference more based on the characteristics of the data itself, thereby achieving distribution freedom.

4. The Role of Distribution of Freedom in Statistical Inference

4.1 Enhance the Robustness of Inference

In actual research, the overall distribution often fails to precisely meet the distribution form assumed by the parameter statistical method. For instance, the data may have characteristics such as skewness and thick tails, or be affected by outliers. In this case, the inference results of parametric statistical methods may have significant deviations and poor robustness. Non-parametric statistical methods, due to their freedom of distribution and not being restricted by the overall distribution form, can provide relatively reliable inference results in various complex distribution situations, thereby enhancing the robustness of statistical inference.

4.2 Expand the Application Scope of Statistical Methods

The application prerequisite of parametric statistical methods is a clear assumption of the overall distribution, which to a large extent limits their application scope. In many practical problems, the overall distribution is unknown or does not conform to the common form of parameter distribution. At such times, parameter statistics methods cannot be effectively applied. The distribution freedom of non-parametric statistical methods enables them to be applicable to various types of population distributions, whether continuous or discrete, symmetrical or skewed. Non-parametric statistical methods can be used for inference, thus greatly expanding the application scope of statistical methods.

4.3 Simplify the Statistical Inference Process

When making inferences using parametric statistical methods, it is usually necessary to first estimate the parameters of the overall distribution, and then conduct hypothesis testing or construct confidence intervals based on the estimated parameters. This process involves complex parameter estimation and hypothesis testing theories, and the calculation process is also relatively cumbersome. The inference process of non-parametric statistical methods is relatively simple because they do not rely on the specific form of the overall distribution. For instance, non-parametric sign tests only need to focus on the ratio of positive and negative signs in the sample data, without the need for complex parameter estimation, thereby simplifying the statistical inference process.

5. The Correlation between Distribution Freedom and Other Characteristics of Non-parametric Statistics

5.1 Relationship with Rank Statistics

Rank statistics is one of the commonly used tools in non-parametric statistics. It sorts the sample data by size and assigns the corresponding rank. Distribution freedom is closely related to rank statistics. Since rank statistics only focus on the relative order of size of the data and do not involve the specific values of the data or the specific form of the overall distribution, the non-parametric statistical method constructed based on rank statistics has distribution freedom. For instance, the Wilcoxon rank sum test determines whether there are differences in the distribution positions of two samples by comparing their rank sums. The distribution of its test statistic is only related to the sample size and not to the overall distribution.

5.2 Adaptability to Data Types

Non-parametric statistical methods have wide adaptability to different types of data, which complement the freedom of distribution. Whether it is quantitative data or qualitative data, whether it is continuous data or discrete data, non-parametric statistical methods can find appropriate statistics for inference. For instance, for qualitative data, non-parametric methods such as chi-square tests can be used for independence tests; For continuous data, methods such as rank test can be used. The freedom of distribution enables non-parametric statistical methods to break free from the constraints of data distribution forms and better adapt to the characteristics of various data types.

6. Comparison between Distribution Freedom and Parametric Statistical Methods

6.1 Differences in Assumed Conditions

Parametric statistical methods rely on specific assumptions about the overall distribution, such as the normal distribution assumption and the homogeneity of variance assumption, etc. These assumptions are often difficult to fully meet in practical applications. Once the assumptions do not hold true, the inference results of parameter statistics methods may have significant deviations. Non-parametric statistical methods, however, have freedom of distribution and do not require specific assumptions about the overall distribution. Therefore, they are more adaptable to data and can be applied under a wider range of conditions.

6.2 Comparison of Efficiency

When the overall distribution indeed conforms to the form assumed by the parametric statistical method, the parametric statistical method usually has higher efficiency. That is to say, under the same sample size, the parameter statistics method can estimate the population parameters more accurately and conduct hypothesis testing. However, when the overall distribution does not conform to the assumption, the efficiency of parametric statistical methods will drop significantly, and even incorrect conclusions may be drawn. Although non-parametric statistical methods may be slightly less efficient in the case of large samples, they can provide relatively reliable inference results and have better robustness when the overall distribution is unknown or does not conform to the assumptions.

6.3 Different Applicable Scenarios

Parameter statistics methods are applicable to situations where the overall distribution is known or can be reasonably assumed. For instance, in some experimental studies in natural science, through strict control of experimental conditions, it can be considered that the population follows a certain specific parameter distribution. Non-parametric statistical methods are more suitable for situations where the overall distribution is unknown, the data contains outliers or does not conform to the assumption of parametric distribution. For instance, in research fields such as social sciences and biomedicine, the distribution of data is often complex and difficult to describe with a simple parametric distribution. At such times, non-parametric statistical methods have greater advantages.

7. The Value and Challenges of Distribution Freedom in Practical Applications

7.1 Practical Application Value

The freedom of distribution has enabled non-parametric statistical methods to be widely applied in numerous fields. In biomedical research, the distribution of many biological indicators does not follow a normal distribution. The use of non-parametric statistical methods can more accurately analyze drug efficacy, disease influencing factors, etc. In social science research, survey data often present various complex situations, such as small sample sizes and missing values in the data. The distribution freedom of non-parametric statistical methods enables them to better handle these issues and provide reliable statistical support for social science research. In addition, in fields such as quality control and market research, non-parametric statistical methods also play a significant role.

7.2 Challenges Faced

Although the freedom of distribution brings many advantages to non-parametric statistical methods, it also faces some challenges in practical applications. Firstly, the efficiency of non-parametric statistical methods is relatively low in the case of large samples, and a larger sample size may be required to achieve the same inference accuracy as parametric statistical methods. Secondly, the theory of non-parametric statistical methods is relatively complex. For some complex statistical problems, such as multivariate non-parametric statistics and non-parametric regression, their theories and methods are not yet perfect and require further research and development. In addition, in practical applications, how to select an appropriate non-parametric statistical method is also a challenge, which requires comprehensive consideration based on the characteristics of the data and the requirements of the research problem.

6. Conclusion

The distribution freedom of non-parametric statistics is one of its core characteristics, which is of great significance to the theoretical development and practical application of non-parametric statistical methods. Freedom of distribution makes non-parametric statistical methods have higher robustness, a wider range of applications and a simpler inference process in statistical inference.

In practical applications, the freedom of distribution provides an effective tool for solving various complex data analysis problems, especially when the overall distribution is unknown or does not conform to the parameter distribution assumption, non-parametric statistical methods play an irreplaceable role. However, the challenges faced by the freedom of distribution in practical applications include issues such as efficiency, theoretical refinement, and method selection.

Future research should further delve into the theoretical basis of distribution freedom, improve the system of non-parametric statistical methods, enhance the efficiency of non-parametric statistical methods in large sample situations, and at the same time strengthen the guidance and promotion of non-parametric statistical methods in practical applications, so as to promote the application and development of non-parametric statistical methods in a wider range of fields. Provide more powerful statistical support for solving practical problems.

References

- [1] Smeeton, N., Spencer, N. H., & Sprent, P. (2025). *Applied nonparametric statistical methods*. CRC press.
- [2] Dwivedi, A. K., Mallawaarachchi, I., & Alvarado, L. A. (2017). Analysis of small sample size studies using nonparametric bootstrap test with pooled resampling method. *Statistics in medicine*, 36(14), 2187-2205.
- [3] La Rocca, M., Liseo, B., & Salmaso, L. (2020). *Nonparametric Statistics* (Vol. 339, pp. 1-547). Springer Nature.
- [4] Berger, V. W., & Zhou, Y. (2014). Kolmogorov–Smirnov test: Overview. *Wiley statsref: Statistics reference online*.
- [5] Manikandan, S., & Ramachandran, S. S. (2023). Principles and applications of statistics in biomedical research: parametric and nonparametric tests including tests employed for post hoc analysis. In *The Quintessence of Basic and Clinical Research and Scientific Publishing* (pp. 479-499). Singapore: Springer Nature Singapore.
- [6] Lei, L. (2024). CAREER: Distribution-Free and Adaptive Statistical Inference. NSF Award Number 2338464. Directorate for Mathematical and Physical Sciences, 23(2338464), 38464.
- [7] Reddy, T. A., & Henze, G. P. (2023). Parametric and non-parametric regression methods. In *Applied data analysis and modeling for energy engineers and scientists* (pp. 355-407). Cham: Springer International Publishing.
- [8] Kvam, P., Vidakovic, B., & Kim, S. J. (2022). *Nonparametric statistics with applications to science and engineering with R*. John Wiley & Sons.