Artificial Intelligence with Neutrosophic Statistics

Artificial Intelligence with Neutrosophic Statistics:

From Nano to Nature

Edited by

Usama Afzal and Muhammad Aslam

Cambridge Scholars Publishing

Artificial Intelligence with Neutrosophic Statistics: From Nano to Nature

Edited by Usama Afzal and Muhammad Aslam

This book first published 2026

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-0364-6313-7
ISBN (Ebook): 978-1-0364-6314-4

TABLE OF CONTENTS

CHAPTER 1

A CRITICAL EXAMINATION OF UNCERTAINTY IN NATURAL SCIENCE DATA: CHALLENGES AND OPPORTUNITIES

USAMA AFZAL
SCHOOL OF MICROELECTRONICS, TIANJIN UNIVERSITY,
TIANJIN 300072, CHINA

Abbreviations

- CS Classical Statistics
- TM Traditional Methods
- NSs Neutrosophic Sets
- NS Neutrosophic Statistics
- NR Neutrosophic Regression

1.1 Introduction

The contemporary rapid development of science and technologies makes the skill of data analysis and interpretation rather valuable. Data drives decision-making in natural sciences. However, the data we collect is often fraught with uncertainty due to many factors, including measurement error, inherent variability and incomplete information. Addressing this uncertainty is critical to obtain accurate predictions, developing reliable systems and improving our understanding of difficult and complex problems [1, 2].

The data analysis importance cannot be exaggerated. For example in environmental science, understanding automorphic behaviour, environmental pollution and managing natural resources through proper planning can be performed through data analysis [3]. This is very useful to make the right policies to reduce the negative impacts and lead to healthy ecosystems. As in the case of material science, the analysis provides data that result in the discovery of new better materials to enhance technological and industrial advancement. In nanotechnology, accurately interpreting data is critical to develop nanoscale devices and systems that can be able to produce a positive revolution in medicine, electronics and energy production for a reliable environment. Regardless it remains a challenge to analyze complex data sets effectively and properly. One of the main problems is the inherent uncertainty of the data [4, 5].

Such sources of measurement errors might emanate from inadequate instrument calibration and conditions of data collection. For example, environmental sensors may give different readings under serval conditions and factors such as automorphic humidity, automorphic temperature or the presence of any other interfering substance. Inherent changes are another source of uncertainty in data. Such as biological processes and weather patterns in natural systems show stochastic behaviour; this thing makes accurate predictions difficult. In addition, it is worth noting that incomplete data or gaps within the data skew the analysis which leads to wrong conclusions.

Although Classical Statistics (CS) and Traditional Methods (TM) are powerful and well-established, however, there are some limitations due to which TM is not effective in the analysis of uncertain data [6, 7]. These methods are generally premised on different assumptions for example normality, independence and large sample sizes which contradict actual data and calculations. For example, atmospheric data may appear abnormal with heavy tails or skew and observations may be correlated over time or space. Small samples are common in materials science due to the high cost and complexity of the experiment, making traditional statistical situations less efficient. Such weaknesses are attributable to possible lapses in analysis and interpretation, which in essence influence the credibility of the outcome obtained.

The problems have been solved at present with the help of new techniques namely Neutrosophic Sets (NSs) and statistic tools. Neutrosophic Statistics (NS) has been developed by Florentine Smarandache to define flexibility and a complete approach for analyzing indeterminacy and uncertainty [8, 9]. CS only deals with the object as belonging or not belonging to a set, on the other hand, NS consider degrees of membership. To understand this let us take an example. According to classical statistics, a statement is considered true or false at a particular time. Such type of binary perspective does not accept compromise or uncertainty. However, NS offer a more flexible perspective. In this context, a statement can be true and false at the same time based on indeterminacy which allows the degrees of variation [10, 11]. This kind of triad can offer a better characterization of the uncertainty by including inaccurate, inconsistent and incomplete data. On this basis, NS offers a good concept of tools and methods that are oriented especially on dealing with the data under uncertainty. These approaches help researchers and practitioners to overcome the problematic situation of dealing with imprecise information and turn out more accurate information that describes the actual state of data.

Artificial intelligence (AI) has significantly enhanced the capabilities of NS, such as excels data at recognizing particular patterns, machine learning, making decisions from complex data and understanding complex networking. As it can be viewed there can be great benefits from combining AI with NS to gain the potential of both strategies. AI can work with a data set and analyze relationships that can be inconspicuous to humans. At the same time, NS accurately represent uncertainty and ensures the analysis of data. This potentiality facilitates accurate foresight, enhanced decision makings as well as understanding of complex events.

This book is an initial attempt to investigate how NS as an AI tool can address uncertainty issues across different domains. We review the principles and methods of NS, demonstrating their flexibility and effectiveness when working with uncertain data. We will also consider how the neutrosophic method to improve data analysis competencies. With several examples and case studies from the real world, we will show how these progressive methods will be applied to real problems in the fields of environmental science, materials science and nanotechnology. By applying the accurate representation of uncertainty we can expand our capabilities to analyze complex data and gain valuable insights. This book will become a comprehensive guide for researchers, practitioners and students interested, who are working on big data mining and complex data analysis through modern AI methods and approaches. Through this book, the reviewer will explore the boundaries of data science and explore new opportunities for understanding and innovation in a world of fundamental uncertainty.

1.2 Sources of Uncertainty

Fig. 1-1 illustrates the multifaceted nature of uncertainty inherent in natural sciences data.

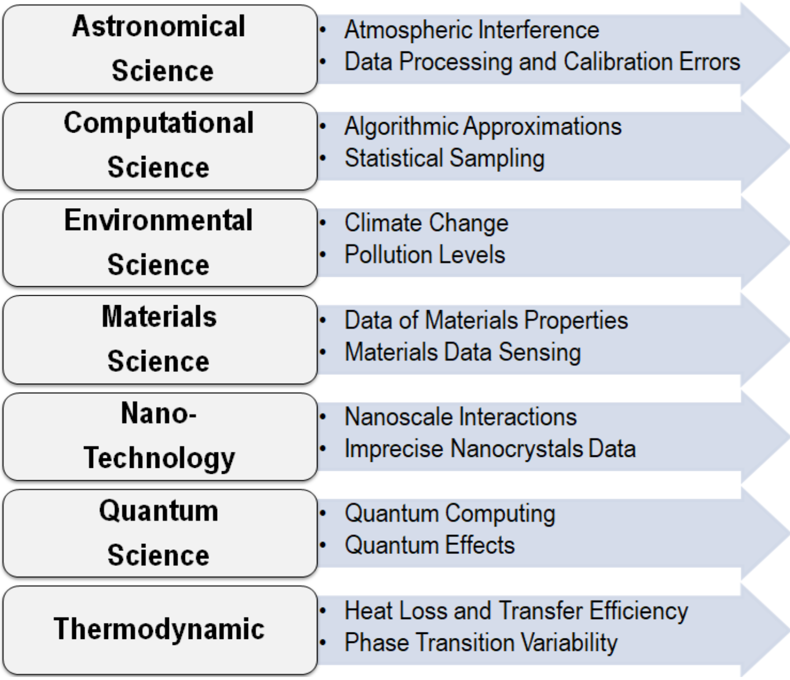

Fig. 1-1 Uncertainty sources

1.2.1 Astronomical & Computational Sciences

In astronomical science, there are numerous sources of uncertainty. Comprehension of heavenly bodies, observational errors result from the instruments themselves, atmospheric disturbances and readout noise, whereas modelling and theoretical errors are relevant to astrophysical phenomena, unprecise physical parameters and assumptions made in modelling or theory. A further source of inconsistency results from using statistical analysis of collected data and from data interpretations which involve statistical errors, systematic disturbances and subjective prejudice. Some limitations relate to cosmic variance, uncertainty principles, distance, and time-scale uncertainties together with degeneracies of model parameters. In particular, the uncertainties are related to such topics as dark matter, dark energy, black holes, star-solar system formation and parameters of the universe. Sources of uncertainty in computational sciences are numerical, modelling, parameter, algorithmic and data uncertainty, which have influences on the accuracy and reliability of computer simulations, such as in climate modelling, fluid dynamics, material science, computational biology, machine learning and quantum computing where lower accuracy and understanding result in higher risks when modelling complex systems. To deal with this problem, computational scientists use techniques such as sensitivity analysis, Bayesian inference, Monte Carlo methods, surrogate modelling, validation and the frameworks of uncertainty quantification. However, for proper data analysis, there is still a need to deploy modern analytical statistical methods that can well analyze data and offer us proper information to make a better decision.

1.2.2 Environmental Sciences

If we discuss environmental science, then one thing that defines the environmental data is very much unpredictable. Variables like changes in biodiversity rates, changes in weather and climatic and pollution factors have numerous causal factors, most of which are interacting and stochastic. Each is a large system with specific functions for accurate measurement and data analysis and likely in forecasting. For example consider weather patterns, which are essentially complex nonlinear systems. The behaviour of the weather is the result of numerous factors while temperature, humidity, wind speed, atmospheric pressure and solar radiation represent the interacting components of it. The relations between these components can be very nonlinear and sometimes produce much-unexpected coupling effects. For example, a minor change in the temperature of the sea surface introduces significant shifts in the weather patterns and leads to El Niño or La Niña events [12]. Such type events, in turn, have large-scale impacts on global climate; affect the rain falling patterns, the agricultural output and even the frequency and the intensity of natural disasters such as hurricanes and floods to understand these phenomena, TM generally cannot analyze this complexity properly and cannot provide precise predictions, leading to substantial uncertainty in weather forecasts. Climate change complicates environmental data. Natural factors, Rainfall intensity and Human factors, It is therefore clear that long-term climate change is due to natural processes and human actions affecting climate models in the long run [13]. Greenhouse gases and emissions, deforestation, urbanization and industrialization have an impact on climate change but their effects are less than those of natural processes which include changes in solar activities, ocean currents and volcanic activities, etc. The encounter of humans and such factors make the system very sensitive and challenging. The relationship between man and the physical environment is very complicated and the system is constantly changing. Future climatic conditions involve developing models with the capacity for incorporating uncertainty in the process and data that are used in forming the model parameters. This makes the climate system hard to understand and as a result, it becomes hard to make accurate long-term climatic predictions since some of the elements interact and give feedback.

Environmental pollution is another field where the acuity of data becomes critically important as uncertainty in data raises many serious issues [14]. The potential source of pollution can be classified into natural, anthropogenic/artificial and exhaust from industries, cars, agriculture/fires in forests and volcanoes. Many factors affect the dispersion and concentration of pollutants in the atmosphere, including climatic conditions, topography and the chemical properties of the pollutants themselves. Such as the surface concentration of ozone, the main reason smog is affected by sunlight, temperature and the presence of elemental chemicals such as nitrogen oxides and volatile organic compounds. Complex interactions between these factors lead to large spatial and temporal fluctuations in pollution levels, which complicate precise air quality monitoring and forecasting. Biodiversity indicators pose a problem for ecological analysis. Biodiversity includes the diversity within a biological ecosystem, species health, genetic diversity and the complexity of ecological interactions [15]. The arrangements of interactions between species and their habitats are ecological networks, sensitive to both abiotic and biotic changes. For example, reductions in the abundance of key species can have major ecosystem-wide impacts, altering species composition, trophic interactions and ecosystem productivity. Collecting and analyzing data on biodiversity requires complex and often large-scale monitoring efforts, which are further complicated by the variability and uncertainty of existing ecosystems.

1.2.3 Material Sciences

Similarly, in material science, data on the properties and behaviour of various materials can affect small changes in composition, processing conditions and environmental effects [16, 17]. Material scientists study the relationship between the structure of materials, processing, properties and characteristics to synthesize new materials with individual properties suitable for concrete applications. However, this task is complex by the fact that minor changes in the composition or processing conditions can lead to large changes in the properties of the material. For example, the mechanical properties of alloys, such as ductility, hardness & strength, are very sensitive to their micro-structural properties, which are affected by factors like alloy composition, thermal processing and deformation processing. Even small changes in the doping of alloying elements or heat treatment process parameters can lead to changes in the microstructure and, in turn, to changes in material properties. Such sensitivity to processing conditions represents a serious problem for material scientists, as it requires accurate control and comprehensive consideration of the factors influencing material behaviour. External environmental factors also play a decisive role in determining the properties of materials. Environmental factors, such as humidity, temperature and exposure to chemicals, can affect the material's performance characteristics and durability. For example, the metals' corrosive resistance and alloys' corrosive resistance is affected by the presence of corrosive agents, such as salt water or acidic environments. Prediction of the long-term properties of materials in such environments requires a detailed understanding of the interactions between materials and environmental factors, as well as degradation and destruction mechanisms.

1.2.4 Nano-Technology

At the nano-level, the problems become even more pronounced, as quantum effects and interactions at the nano-level create levels of complexity that are difficult to accurately capture using TM of measurement and analysis [18, 19]. Nanotechnology includes the handling of matter at the atomic and molecular level, where the behaviour & properties of materials are significantly different from bulk materials. Quantum effects, such as electron-tunnelling, quantum-confinement, and surface-plasmon-resonance, have become noticeable at the nano-level, leading to exclusive electronic, mechanical and optical properties. For example, the electric properties of semiconductor nanoparticles or quantum dots generally depend on their size and shape. Quantum effects produce discrete energy levels that can be tuned, by regulating nanoparticle size which is very useful for biological imaging and light-emitting diodes (LEDs) as well as solar cells. However, accurate characterization and prediction of nanoparticles and quantum dots properties requires accurate control of their structural data and a comprehensive understanding of quantum effects.

Nano-scale interactions also pose serious problems for data analysis [20]. The high surface area-to-volume ratio of nanomaterials means that surface effects, such as adsorption, surface tension and surface reactivity, play a dominant role in determining their properties. For example, the catalytic activity of nanoparticles is influenced by their surface composition and the presence of defects or adsorbed materials on the surface. Conventional measurement methods, developed for porous materials, often lack the precision and sensitivity required to accurately detect these nanoscale interactions.

1.2.5 Quantum Sciences

In quantum science, indeterminacy refers to the fundamental limitation in precisely knowing certain properties of a quantum system, even with unlimited measurement precision. This indeterminacy may be due to numerous factors such as entanglement, fluctuations, interaction of measurements and wave function indefiniteness, interaction of measurements, fluctuations, and entanglement. Indeterminacy brings uncertainty and probability which affects the data accuracy of a quantum system. This work has a profound effect on quantum computing, communication, metrology and field theory, which requires more efficient algorithms, better measurements and methods of statistical analysis. Quantum indeterminacy can be categorized into three types: The decisive uncertainty types are found to be quantum mechanical indeterminacy, epistemic indeterminacy and aleatory indeterminacy. The recognition of the issue of indeterminacy and its prevention is vital for furthering the field of quantum science and creating technologies based on it.

In brief, quantitative assessment and prediction of environmental data, material science data and nanoscale data are associated with problems due to the complexity and uncertainty inherent in these systems. Weather conditions, climate change, pollution levels and biodiversity are affected by many interrelated and random variables, making accurate predictions difficult. Similarly, material properties and behaviour are very sensitive to small changes in composition, processing conditions and external environment. At the nano-level, quantum effects and nanoscale interactions introduce additional complexity that TM of measurement and analysis can hardly capture.

Solving these problems requires innovative methods and structures that can deal with uncertainty and complexity. Neutrosophic Statistics offers a promising approach, providing a flexible and versatile basis for modelling and analyzing uncertain data. By incorporating the principles of neutrosophic clustering into data analysis, researchers will be able to better account for the errors, inconsistencies, and incompleteness that characterize real-world data. This approach can lead to more accurate predictions, better solutions and a deeper understanding of complex phenomena across a wide range of fields.

1.3 Limitations of Traditional Methods

Classical Statistics (CS) and Traditional Methods (TM) have certain limitations that reduce the effectiveness of analyzing imprecise data and obtaining useful results, as illustrated in Fig. 1-2.

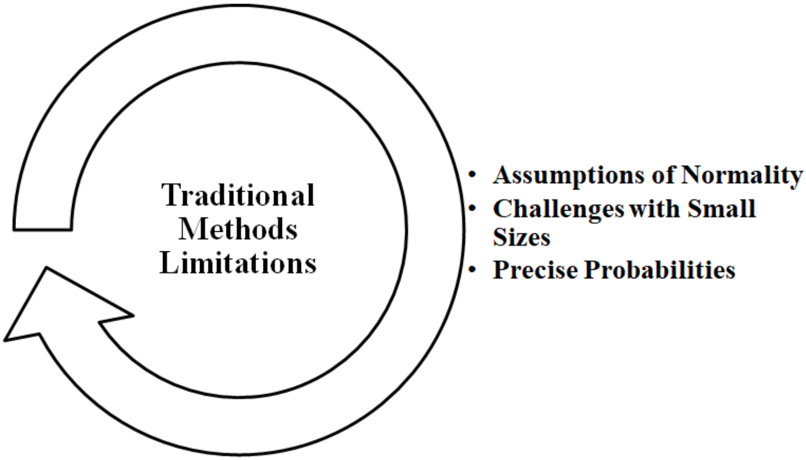

Fig.1-2 Limitations of TM to analyze imprecise data

1.3.1 Assumptions of Normality and Independence

Traditional statistics generally assume that the data follow a normal distribution and that observations are independent of each other [21, 22]. However, such types of assumptions are often violated in real data sets. For example, rainfalling data frequently show a skewed distribution, with small amounts of rain falling occurring frequently and extreme events occurring infrequently. This violates the assumption of normality and can lead us to incorrect predictions and analyses of data. In materials science, the size distribution of defects in manufactured parts can be skewed, with several defects and some large defects. This deviation violates the normality assumption and complicates material properties analysis and failure rate prediction.

There are also problems with establishing independence. Environmental data points are often associated with location and time. For example, concentrations of pollutants, measured in different places in the city, can be correlated due to the influence of prevailing winds. In material science, the properties of adjacent grains in polycrystalline materials are connected by common boundaries and interactions.

1.3.2 Challenges with Small Sample Sizes

In many cases, the collection of a large number of samples is inappropriate or impossible [23, 24]. For example, in nanoscale research, each experiment or measurement can take a long time and is expensive, which limits the amount of samples that can be obtained. In these cases, TM, based on large sample sizes, may not give reliable results, which leads to conclusions with a high degree of uncertainty.

Let us take an example of a study that aims to characterize the properties of a new nano-material. Due to the high cost and complexity of producing the material, only a few samples can be evaluated due to their high cost and complex procedure of material synthesis. In this case, TM may not provide precise estimation for the properties of synthesis material and their variability as there is a high uncertainty in the findings of the study.

1.3.3 Dealing with Precise Probabilities

Traditional statistics deal with exact probabilities [25], but real data often contain inaccuracies and ambiguities. For example, when monitoring the environment, the probability that the level of a particular pollutant exceeds a certain limit may depend on inaccurate measurements and expert conclusions. TM takes this type of uncertainty into account with difficulty, mostly requiring overly simplified assumptions that may lead to misleading outcomes. Consider this scenario: Environmental scientists want to evaluate the risk of a chemical spill in water. Due to incomplete information and differing expert opinions, the probability of the spill going on and its possible effects on water quality are uncertain. TM cannot precisely quantify such types of risks, leading to overestimation or underestimation of possible effects.

1.4 Introduction to Neutrosophic Statistics

Neutrosophic statistics (NS), proposed by Florentine Smarandache, offers a new way to deal with uncertainty in data analysis [8, 9]. The traditional theory of sets is based on binary membership, when an element either belongs to a set or does not belong to a set. This binary approach is often too limited for real-world scenarios where data may be inaccurate, incomplete, or contradictory. NS expands classical set theory, allowing elements to simultaneously have true values, false values, and degrees of indeterminacy. This triangular structure provides a more flexible and comprehensive basis for modelling uncertainty.

1.4.1 General Definition

Let's take a quick look at this strategy as follows:

As we know that X is the neutrosophic value with X_L the large number and X_U the small number providing the ith intervals of the neutrosophic approach as in equation 1.1 [26]:

$$X_i = X_{Li} + X_{Ui}I_i \in (i = 1,2,3, \ldots n_N) \qquad 1.1$$

Here, the neutrosophic variables $X_i \in [X_{Li}, X_{Ui}]$ is composed of two parts; a classical portion with lower values X_{Li} and an indeterminate portion with higher values $X_{Ui}I_i$ and an indeterminacy interval $I_i \in [I_{Li}, I_{Ui}]$ where I_{Li} the lower value of the indeterminacy interval which is always taken as zero as this approach is the generalization of the classical approach and I_{Ui} is the upper value of the indeterminacy interval which can be found through ${(I_{Ui} - I_{Li})}/{I_{Ui}}$.

1.4.2 Understanding of Neutrosophic Statistics

Based on the concept of a neutrosophic set, NS offers a set of tools designed to deal with complex data under indeterminacy and uncertainty. Classical statistics such as regression analysis and hypothesis testing often cannot work properly when dealing with the uncertainty and ambiguity of real-world data. NS allows for more robust and reliable analysis by incorporating the principles of neutrosophic groups into these TM [27]. NS is a branch of Neutrosophic Logic that studies the use of NSs in statistical analysis. In traditional statistics, data points are often considered precise and fixed. However, real data is rarely so clear-cut. Measurement errors, inherent variability, and incomplete information lead to large uncertainties that cannot be adequately addressed by conventional methods [28].

NS solves this problem by representing each data point as a neural set characterized by three parameters: true membership (T), uncertain membership (I), and false membership (F). This tripartite structure allows for a more accurate representation of uncertainty because each data point corresponds to multiple levels of truth, uncertainty, and falsity simultaneously. This approach provides a rich basis for analyzing inaccurate, inconsistent, or incomplete data.

Neutrosophic regression is a prime example of how NS expands the capabilities of TM. Classical regression analysis uses linear or nonlinear equations to model the relationship between the dependent and independent variables. However, these models assume precise and deterministic data points, which is rarely the practice case.

In Neutrosophic regression, the dependent and independent variables appear as NSs. This means that each variable has confidence, uncertainty, and error that reflect the uncertainty of the data. Thus, the NR model accounts for this uncertainty and provides a more accurate and reliable analysis.

1.4.3 Application of Neutrosophic Statistics

Neutrosophic statistics can be applied to a wide range of applications where data uncertainty is a major concern as shown in Fig. 1-3.

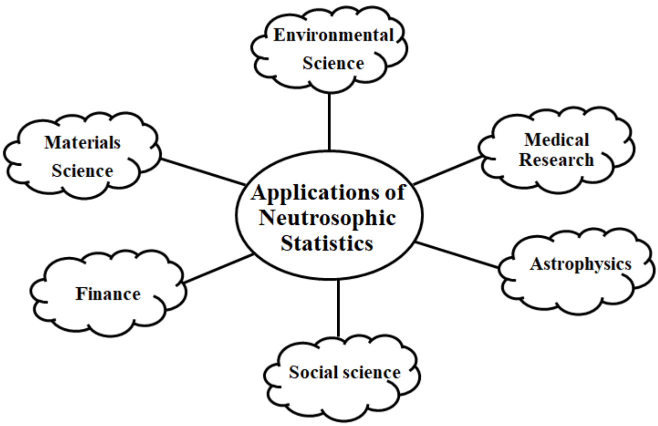

Fig. 1-3 Number of application neutrosophic statistics

In environmental science, data on pollution levels, climate change and biodiversity are often inaccurate due to measurement errors, natural variability and incomplete information. Advanced NS can improve the analysis of these data by providing tools to take these uncertainties into account [29, 30]. For example, when predicting air quality, traditional models may have difficulties adapting to climatic conditions and changes in the concentration of pollutants due to measurement errors. Models of NR can represent concentrations of pollutants and climatic variables in the form of ensembles of the neutrosphere, which allows more accurate and reliable forecasting of air quality by accounting for uncertainty in the data.

In clinical and medical research, clinical trials, and observational studies often deal with uncertain data due to patient heterogeneity, measurement errors, and incomplete observation and records. NS can expand the analysis of these data by providing a framework for interpreting these uncertainties [26, 31]. Such as when we analyze the effectiveness of a new drug, TM may not fully capture variability in patient response. Neutrosophic regression models can represent patient response and other relevant variables as NSs, allowing for more confident conclusions about drug effectiveness while accounting for data uncertainty.

In the financial sector, market data is inherently uncertain due to the impact of many unpredictable factors such as economic conditions, investor behaviour and geopolitical events. NS can improve financial modelling and risk analysis by incorporating these uncertainties into the analysis. Such as, when we are dealing with forecasting the price of shares, traditional models cannot take into account all the uncertainties that affect the market. NR models can represent stock prices and other related variable in the form of NSs combinations and provide more accurate and reliable forecasts, taking into account the uncertainty of the data.

1.5 Benefits of Neutrosophic Statistics

Neutrosophic Statistics is a powerful advance over TM, offers many unique advantages and is particularly suited to the analysis of complex data. These advantages include accuracy, reliability, flexibility and comprehensive analysis. By overcoming the limitations of traditional statistics, NS provides a more accurate and efficient way to deal with data uncertainty.

1.5.1 Enhanced Accuracy

The primary benefit of NS is the ability to improve the accuracy of data analysis. TM often assumes that data points are precise and deterministic, but this is a rare case in real-world scenarios. Measurement error, inherent variability, and incomplete information can introduce significant uncertainty into the data and lead to potential errors in analyses and predictions.

1.5.2 Accounting for Uncertainty

NS tackles this issue by representing each data point as a neutrosophic set, characterized by three parameters: truth-membership (T), indeterminacy-membership (I), and falsity-membership (F). This triadic structure allows for a more nuanced representation of uncertainty, where each data point can simultaneously belong to multiple degrees of truth, indeterminacy, and falsity. By incorporating these degrees, NS provide a more accurate representation of the underlying data.

For example, consider a scenario in environmental science where researchers are trying to predict air quality based on various pollutants. Traditional regression models might struggle to accurately predict air quality due to the variability and uncertainty in pollutant measurements. An NR model, on the other hand, can represent pollutant concentrations as NSs, capturing the uncertainties in the data and leading to more accurate predictions.

1.5.3 Improved Predictions and Decision-Making

By accounting for the uncertainties in the data, NS leads to better predictions and decision-making. In fields such as finance, healthcare, and environmental science, accurate predictions are crucial for making informed decisions. NS provide a more reliable basis for these predictions by considering the uncertainties inherent in the data.

For instance, in finance, predicting stock prices involves significant uncertainty due to the influence of numerous unpredictable factors. Traditional models might fail to account for all these uncertainties, leading to inaccurate predictions. An NR model can represent stock prices and other relevant variables as NSs, providing more accurate and reliable predictions by considering the uncertainties in the data.

1.5.4 Robustness

Another significant advantage of NS is its robustness to data variability and measurement errors. Real-world data is often characterized by variability and errors that can compromise the reliability of TM. NS are designed to handle these challenges, making them suitable for analyzing complex data characterized by uncertainty.

1.5.5 Handling Data Variability

Data variability is a common challenge in many fields. For example, in medical research, patient responses to treatments can vary significantly due to individual differences in genetics, lifestyle, and other factors. TM might not fully capture this variability, leading to potential inaccuracies in the analysis.

NS handle data variability by representing each data point as a neutrosophic set, which can capture the range of possible values for a variable. This allows for a more accurate and robust analysis that accounts for the inherent variability in the data. For instance, when analyzing the effectiveness of a new drug, an NR model can represent patient responses as NSs, capturing the variability in responses and leading to more reliable conclusions about the drug's effectiveness.

1.5.6 Mitigating Measurement Errors

Measurement errors are another common challenge in data analysis. Inaccurate measurements can arise from various sources, including instrument limitations, human error, and environmental factors. TM often assume that measurements are precise, which can lead to inaccuracies in the analysis when measurement errors are present.

NS mitigate the impact of measurement errors by incorporating the degrees of truth, indeterminacy, and falsity into the analysis. This allows the model to account for the uncertainty introduced by measurement errors, leading to a more robust and reliable analysis. For example, in environmental science, measurements of pollutant concentrations can be affected by instrument limitations and environmental conditions. An NR model can represent these measurements as NSs, capturing the uncertainties introduced by measurement errors and providing a more accurate analysis of pollutant levels.

1.5.7 Flexibility

The flexibility of NS is another key benefit that makes it applicable to a wide range of fields and data types. The triadic structure of NSs allows for a flexible representation of data, accommodating different types and degrees of uncertainty.

1.5.8 Accommodating Different Types of Data

NS can be applied to various types of data, including numerical, categorical, and ordinal data. This flexibility makes them suitable for analyzing data from diverse fields, such as finance, healthcare, environmental science, and social sciences.

For example, in social sciences, researchers often deal with subjective data, such as survey responses and opinion polls. TM might struggle to accurately analyze this subjective data due to its inherent uncertainty. NS can represent subjective data as NSs, capturing the uncertainty in the responses and providing a more accurate analysis. This flexibility allows researchers to apply NS to a wide range of data types, enhancing their ability to analyze complex and uncertain data.

1.5.9 Representing Different Degrees of Uncertainty

NS can accommodate different degrees of uncertainty, making them suitable for analyzing data with varying levels of precision and certainty. The triadic structure of NSs allows for a flexible representation of data, capturing the degrees of truth, indeterminacy, and falsity for each data point.

For example, in medical research, data from clinical trials might have varying degrees of certainty due to differences in sample sizes, measurement techniques, and patient responses. TM might not fully capture these differences, leading to potential inaccuracies in the analysis. NS can represent the varying degrees of uncertainty in the data, providing a more accurate and reliable analysis. This flexibility allows researchers to account for the different degrees of uncertainty in the data, enhancing the accuracy and robustness of the analysis.

1.5.10 Comprehensive Analysis

NS provides a comprehensive framework for analyzing complex data, considering the truth, indeterminacy, and falsity of each data point. This leads to a deeper understanding of the data and the relationships between variables.

1.5.11 Multi-Faceted Data Representation

The triadic structure of NSs allows for a multi-faceted representation of data, capturing the degrees of truth, indeterminacy, and falsity for each data point. This comprehensive representation provides a more nuanced and detailed analysis of the data, enhancing the understanding of the underlying phenomena.

For example, in environmental science, researchers might study the relationship between air quality and various pollutants. Traditional regression models might represent pollutant concentrations as precise values, potentially overlooking the uncertainties in the data. An NR model can represent pollutant concentrations as NSs, capturing the degrees of truth, indeterminacy, and falsity for each data point. This comprehensive representation allows for a more detailed analysis of the relationship between air quality and pollutants, leading to a deeper understanding of the underlying factors.

1.5.12 Identifying Hidden Patterns and Relationships

NS can help identify hidden patterns and relationships in complex data by considering the uncertainties and variabilities in the data. This leads to more insightful and reliable conclusions about the underlying phenomena.

For example, in finance, identifying patterns and relationships in stock market data is crucial for making informed investment decisions. TM might overlook the uncertainties and variabilities in the data, potentially leading to inaccurate conclusions. NS can represent stock market data as NSs, capturing the uncertainties and variabilities in the data. This comprehensive analysis allows for the identification of hidden patterns and relationships, providing more insightful and reliable conclusions about stock market trends.

1.5.13 Enhancing Data-Driven Decision-Making

By providing a comprehensive analysis of the data, NS enhances data-driven decision-making. The detailed and nuanced representation of data allows for more informed and reliable decisions, particularly in fields characterized by uncertainty.

For example, in healthcare, making decisions about patient treatment plans requires accurate and reliable analysis of patient data. TM might not fully capture the uncertainties in the data, potentially leading to suboptimal treatment decisions. NS can represent patient data as NSs, capturing the uncertainties in the data and providing a more comprehensive analysis. This enhances the ability to make informed and reliable decisions about patient treatment plans, leading to better healthcare outcomes.

1.6 Summary of Chapter

This chapter introduces Neutrosophic Statistics (NS) as a ground-breaking improvement in data analysis, particularly suitable for treating the complex and uncertain data of practical applications. Traditional Methods (TM) often struggle with the intrinsic uncertainties of data, such as measurement errors, variability, and incomplete information. NS address these limitations by utilizing NSs, which represent data points with three parameters: truth-membership (T), indeterminacy-membership (I), and falsity-membership (F). This structure allows for a nuanced demonstration of uncertainty, leading to improved accuracy and robustness in the analysis of data.

NS also offer noteworthy flexibility and inclusive analysis abilities. They can accommodate numerous sorts and degrees of uncertainty, making them applicable across varied fields such as finance, healthcare and social sciences. By representing particular data like survey responses with the inherent uncertainty captured, NS provides more precise analyses. Additionally, their structure allows for a multi-faceted data representation, which improves the understanding of fundamental phenomena and assists in recognizing hidden patterns and relationships. In general, NS offers a more nuanced and operative method to treat complex data, making it an important tool for researchers and practitioners to aim for precise, consistent and insightful analysis.

References

1. Council, N.R., et al., Completing the forecast: Characterizing and communicating uncertainty for better decisions using weather and climate forecasts. 2006: National Academies Press.
2. Ascough Ii, J., et al., Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. Ecological modelling, 2008. **219**(3-4): p. 383-399.
3. Maxim, L. and J.P. van der Sluijs, Quality in environmental science for policy: Assessing uncertainty as a component of policy analysis. Environmental **science** & policy, 2011. **14**(4): p. 482-492.
4. Wang, Y., Uncertainty in materials modeling, simulation, and development for ICME. Proc. 2015 Materials Science and Technology, 2015.
5. Lookman, T., et al., Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. npj Computational Materials, 2019. **5**(1): p. 21.

6. Hlavacek, I., J. Chleboun, and I. Babuska, *Uncertain input data problems and the worst scenario method*. 2004: Elsevier.
7. Wang, X. and Y. He, *Learning from uncertainty for big data: future analytical challenges and strategies.* IEEE Systems, Man, and Cybernetics Magazine, 2016. **2**(2): p. 26-31.
8. Smarandache, F., A unifying field in logics: neutrosophic logic. Neutrosophy, neutrosophic set, neutrosophic probability: neutrsophic logic. Neutrosophy, neutrosophic set, neutrosophic probability. 2005: Infinite Study.
9. Smarandache, F., *Introduction to Neutrosophic Statistics, Sitech and Education Publisher, Craiova.* Romania-Educational Publisher, Columbus, Ohio, USA, 2014. **123**.
10. Afzal, U., et al., Neutrosophic statistical analysis of resistance depending on the temperature variance of conducting material. Scientific reports, 2021. **11**(1): p. 23939.
11. Afzal, U., Estimate Calculation of Discrete Energy Levels in the Hydrogen Atom Using an Indeterministic Method. High Energy Density Physics, 2025: p. 101192.
12. Collins, M., *El Niño-or La Niña-like climate change?* Climate Dynamics, 2005. **24**: p. 89-104.
13. Randall, D.A., et al., Climate models and their evaluation, in Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC (FAR). 2007, Cambridge University Press. p. 589-662.
14. Reimann, C., et al., Statistical data analysis explained: applied environmental statistics with R. 2011: John Wiley & Sons.
15. Butchart, S.H., et al., *Global biodiversity: indicators of recent declines.* Science, 2010. **328**(5982): p. 1164-1168.
16. Ward, L., et al., A general-purpose machine learning framework for predicting properties of inorganic materials. npj Computational Materials, 2016. **2**(1): p. 1-7.
17. Afzal, U., et al., Chemical Engineering for Advanced Flexible Sensors: Novel Carbon-Metal Oxide Nanocomposites with Superior Multi-Sensing Behavior. Sensors and Actuators B: Chemical, 2025: p. 137449.
18. Di Sia, P., Present and future of nanotechnologies: Peculiarities, phenomenology, theoretical modelling, perspectives. Reviews in Theoretical Science, 2014. **2**(2): p. 146-180.
19. Schommers, W., Basic physics of nanoscience: traditional approaches and new aspects at the ultimate level. 2018: Elsevier.

20. Boulogeorgos, A.A.A., et al., *Machine learning in nano-scale biomedical engineering.* IEEE Transactions on Molecular, Biological and Multi-Scale Communications, 2020. **7**(1): p. 10-39.
21. Gnanadesikan, R., Methods for statistical data analysis of multivariate observations. 2011: John Wiley & Sons.
22. Wetherill, G.B., *Intermediate statistical methods*. 2012: Springer Science & Business Media.
23. Singh, A.S. and M.B. Masuku, *Sampling techniques & determination of sample size in applied statistics research: An overview.* International Journal of economics, commerce and management, 2014. **2**(11): p. 1-22.
24. Beleites, C., et al., *Sample size planning for classification models.* Analytica chimica acta, 2013. **760**: p. 25-33.
25. Rohatgi, V.K. and A.M.E. Saleh, *An introduction to probability and statistics*. 2015: John Wiley & Sons.
26. Afzal, U. and M. Aslam, Use of neutrosophic statistics to analyze the measured data of diabetes, in Cognitive Intelligence with Neutrosophic Statistics in Bioinformatics. 2023, Elsevier. p. 119-135.
27. Firdous, A., & Maryam, H. K. (2025). Neutrosophic Logic Reveals Ontological Indeterminacy in Quantum Systems. Journal of Engineering, Science and Technological Trends, **2**(3).
28. Alsharari, F., et al., An Introduction to Single-Valued Neutrosophic Primal Theory. Symmetry, 2024. **16**(4): p. 402.
29. Ahmad, N., et al., Assessment of soil contamination by heavy metals in the vicinity of the Sahiwal coal power plant, Pakistan, employing an innovative statistical analysis approach. International Journal Of Environmental Analytical Chemistry, 2024: p. 1-14.
30. Ahmad, N., et al., Assessment of heavy metals level in chicken with indeterminate analysis in localities of Lahore, Pakistan. Environmental Monitoring and Assessment, 2024. **196**(5): p. 458.
31. Afzal, U. and M. Aslam, *Neutrosophic statistical analysis of changes in blood pressure, pulse rate and temperature of human body due to COVID-19*, in *Cognitive Intelligence with Neutrosophic Statistics in Bioinformatics*. 2023, Elsevier. p. 153-172.

CHAPTER 2

A COMPREHENSIVE REVIEW OF NEUTROSOPHIC THEORY AND ITS MULTIDISCIPLINARY APPLICATIONS IN NATURAL SCIENCES

SADIQUE AKBAR[1], NAVEED AHMAD[2], MUHAMMAD ASLAM[3], ADEENA FATIMA[4] AND USAMA AFZAL[5]

[1]DEPARTMENT OF PHYSICS, UNIVERSITY OF EDUCATION, LAHORE, PAKISTAN

[2]DEPARTMENT OF PHYSICS, UNIVERSITY OF EDUCATION, LAHORE, PAKISTAN

[3]DEPARTMENT OF STATISTICS, FACULTY OF SCIENCE, KING ABDULAZIZ UNIVERSITY, JEDDAH, SAUDI ARABIA

[4]SCHOOL OF MICROELECTRONICS, TIANJIN UNIVERSITY, TIANJIN 300072, CHINA

[5]SCHOOL OF MICROELECTRONICS, TIANJIN UNIVERSITY, TIANJIN 300072, CHINA

Abbreviations

- NCMs Neutrosophic Cognitive Maps
- NL Neutrosophic Logic
- NM Neutrosophic Measure
- N-MCDM Neutrosophic Multi-Criteria Decision Making
- NSs Neutrosophic Statistics
- SVNS Single-Valued Neutrosophic Set
- SVNSs Single Value Neutrosophic Sets

2.1 Introduction

Lambert (1764) looked into how the contradictory evidence of another impacted the credibility of one witness and the concept of tri-partition/triplet based on truth, falsehood, and indeterminacy emerged. He expanded upon a non-Bayesian method for locating a probabilistic model developed in the 1680s by Hooper: the concept of the combination of witnesses. Lower and higher probabilities were first proposed by Koopman in the 1940s, then by Good and finally, by Dempster in 1967, a formula for integrating arguments was laid down. It was further developed by Shafer in 1976 into the Dempster-Shafer Belief Theory—operations by combining two pieces of evidence derived from two distinct sources. There is a bridge between fuzzy logic and probability through belief functions. A more comprehensive application of mathematical probability that relies on the probabilistic integration of testimony in artificial intelligence [1].

Lambert posits the existence of a probability "a" such that the testimony can be reliable and correct, a probability "b" such that the testimony can be mendacious, and finally, a probability of "1-a-b" that the testimony can be merely negligent. Thus, there are three aspects: correct, mendacious, and negligent, whose sum will be 1. Van Fraassen proposed super-valuation semantics to address the Sorites paradoxes, a concept later explored by Dummett and Fine in 1975. They all tri-partitioned, taking into account an ambiguous predicate that is unclear regarding its edging circumstances. Van expanded the ambiguous base "heap" by confidently categorising elements for which the predicate applies and categorising those for which it does not apply. The boundary of the remaining element was termed the penumbra. A clear frontier between these two extensions is absent for a soritical predicate. Inductive reasoning is invalid; if "SO" is a Sorites conditional statement, the argument "≥ n(San&¥San+1)" is false. Therefore, the conditional base Heap with a positive expansion is true, the Heap with a negative expansion is false, and the Heap with a penumbra is indeterminate. Narinyani employed tripartition in 1980 to define the "indefinite set." Subsequently, Atanassov expanded upon this tripartition in 1982, presenting five broad generalisations of the fuzzy set and examining their characteristics and applications within neural networks in the medical field [1].

The inception of a significant shift across all scientific disciplines occurred when Smarandache established the idea of indeterminacy as a distinct element, which was published in 1998. He defined the neutrosophic set (NS) based on three elements categorised as "Truth (T), Indeterminate (I), and False (F), belonging to [0, 1]. T, I, and F may manifest as intervals and single values. An indeterminate element, also called or Neutral element, as a liberated element separate from "T" and "I", constitutes the primary differentiation between Neutrosophic Theories and traditional or fuzzy extension theories [2]. The indeterminacy is also the primary differentiation between traditional/classical probability and neutrosophic probability [3].

2.1.1 Neutrosophic Theory

Neutrosophic theory refers to the scientific use of neutrosophy in several domains to address issues about indeterminacy [4]. The study of the nature, origin, and extent of neutralities and their connections with other ideational spectra is known as neutrosophy, a relatively young field of philosophy. The neutrosophic theory further generalises fuzzy sets, intuitionistic fuzzy sets, spherical fuzzy sets, picture fuzzy sets, and Pythagorean fuzzy sets [5]. According to this theory, each thing is considered along with its opposite or negation and the broad range of neutralities that exist between them, i.e. entities that support neither nor [4].

2.1.2 Neutrosophic Triplet

The neutrosophic triplets need to be defined here:

Let <D> represent a thought, notion, idea, statement, theory, etc., while <antiD> stands for its polar opposite. There is a neutral region, which is called indeterminacy and represented by <neutD>, between the opposites <D> and <antiD>. The <neutD> is a separate thought, notion, idea, statement, and theory, i.e. the <neutD> is neither <D> nor <antiD>. In some cases, the <neutD> can be a combination of partial <D> and partial <antiD> values [3]. We take into account the logical neutrosophic triplets (<D>, <neutD>, <antiD>) that exist in the real domain and there are many instances of triplets.

Examples of neutrosophic triplets can be seen below:

i) (Won, Tied, Lost)
ii) (Well-wisher, Neutral, Enemy)
iii) (Accurate, Partially accurate, or partially inaccurate, Inaccurate)
iv) (True, Indeterminacy, False), etc.

2.2 Neutrosophic Indeterminacy

Classical Statistics pertains exclusively to deterministic data, probabilistic distributions, and inference methods, whereas Neutrosophic Statistics (NSs) encompasses indeterminate data, characterised by ambiguity, vagueness, partial unknowns, contradictions, and incompleteness along with indeterminate probabilistic distributions and inference methods. This includes the circulations and extrapolations that incorporate varying amounts of indeterminacy, allowing for the use of inexact or uncertain contentions and values, such as charts, schematics, computations, and operations [6, 7]. The precise number of people in the sample or the population is unknown; estimates range from 200 to 250. Some people might only partly fit their mark of belongingness, $T < 1$, to the study population, while others might over-belong their degree of belongingness, $T > 1$. In this case, we have John, who works full-time at the plant, Geffrey, who works part-time, and Mary, who works overtime and belongs 110%. George is sixty-five years old, John is forty, and Mary is twenty. Can you tell me how old the typical employee is at this company?

For classical statistics, when factory affiliation is irrelevant, the average age is 40, calculated as $(40 + 60 + 20) / 3$. The degree of belongingness is taken into account in NS's, which yields: $(40\times1 + 60\times0.5 + 20\times1.1) / (1 + 0.5 + 1.1) = 92 / 2.6 \approx 35.38$.

The average age of 120 divided by 3, which equals 40, is incorrect since Geffrey's 50% effort cannot become equal to Mary's 110% labour, even if in conventional statistics the mark of belongingness is assumed to be $T = 1$ for all employees. Instead of being clear and well-defined like in classical statistics, distribution probability curves can be thick functions or indeterminate functions with approximations, or hazy and contradictory data could characterise them.

For indeterminacy $(I) = 0$, let us denote the neutrosophic components as T, I, and F, where $[0, 1]$ is an interval. Neutrosophic components $(T, 0, F)$ outperform intuitive fuzzy components and fuzzy components with indeterminacy $(I) = 0$ in terms of generality and flexibility. For both the intuitive fuzzy set and the fuzzy set itself, T plus F equals 1.

In the NS, where $0 \leq T + F < 3$, any of the following scenarios is possible:

I. When there is incomplete info, $T + F < 3$, to get all the information data.
II. When there is incomplete info, $T + F = 3$.
III. For info that is paraconsistent or contradictory and originates from separate sources, $T + F > 3$.