Development of Neutrosophic Pareto Distribution for Survival Analysis

Ahmedia Musa M. Ibrahim[1,*]

[1]Finance Department, College of Business Administration in Hawtat Bin Tamim, Prince Sattam bin Abdulaziz University, Hawtat Bin Tamim, Saudi Arabia

Email: am.ibrahim@psau.edu.sa

Abstract

We provide a neutrosophic approach to the Pareto model, which is widely used to model survival data. In this paper, the neutrosophic Pareto model (NPM) is constructed under the framework of neutrosophic statistics, that can manage uncertain nature of data, commonly occur in many real word problems. This formulation generalizes the classical model and is a useful method for dealing with fuzzy or uncertain data typically encountered in many applications in survival data. Using neutrosophic statistical framework, few key mathematic qualities of the proposed model such as its moments, survival function, and hazard rate are presented in the study. These properties are motivated and rigorously established to ensure theoretical soundness of the proposed model. Moreover, the maximum likelihood estimation (MLE) is used to estimate the neutrosophic parameters of the distribution. This approach is essential for deriving accurate parameter estimates from the data available, especially in cases where uncertainty or imprecision is present within the data as it is usually the case for any real-world situation. Based on the simulation experiment, we display the adequate performance of the suggested model. The simulations allow us to evaluate the performance of the routine as well as the stability of the model parameters across different settings. At the end, the real data analysis is conducted to show the applicability of proposed approach. The proposed model processes such a dataset filled with a range of uncertain values and presents its possibilities to be applied for information extraction from real world data sets that are abundant in uncertainty. Our results open a new avenue for neutrosophic statistical model approaches to the analysis of survival data in subsequent studies.

Keywords: Survival model; Estimation; Simulation; Neutrosophic logic; Neutrosophic distribution

1. Introduction

Probability distributions are pertinent in theoretical as well as application based statistics. They explain the probabilities of various results in an experiment or process, enabling researchers and practitioners to make informed choices, forecast future occurrences, and evaluate risks [1]. Knowing the mean, variance, and shape of a probability distribution already provides us with a lot of information about the underlying data or process. Probability distributions find applications in almost every domain such as finance, engineering, healthcare, and quality control, where risk management, process optimization, and decision-making under uncertainty are critical [2].

The Pareto distribution is a power-law probability distribution named after Vilfredo Pareto that describes the distribution of wealth among individuals, but it also manifests in many cases of real world well studied phenomena and such almost always involve skewed data [3]. It is significant because it can be used to represent systems in which a few events are responsible foremost of the overall effect, often referred as the "80/20 rule" [4]. The Pareto distribution is commonly used in the study of survival and healthcare data, where the distribution is applied to

DOI: https://doi.org/10.54216/IJNS.250426

model heavy-tailed events, which may include lifetimes or failure times of some specific medical devices, health conditions, or treatment effects [5]. In healthcare, for example, the distribution can be used to describe the survival times of patients with a chronic disease, where a small percentage of patients may utilize most of the available resources, providing insight for managing healthcare and resources efficiently [6]. In survival analysis, the Pareto distribution can help us in understanding survival data of individual or machine, where the hazard rate (the likelihood of failure) could be decreasing over time such as aging population or machine wear [7]. Additionally, it is used in healthcare in modeling the distribution of medical costs, where a small number of patients might incur an outsized amount of these costs, which helps identify high-risk patients and design cost-efficient interventions [8]. It is widely used to analyze healthcare data involving epidemics where a small number of individuals carry the majority of the infection burden, and in modeling waiting times in healthcare systems or emergency services, assisting in the optimization of healthcare delivery systems [9]. Moreover, in financial modeling, the Pareto distribution is commonly applied in modeling wealth or income distribution, where a relatively small group of people owns a disproportionately large share of total wealth [10]. Then, the system is trained to detect patterns, validating statistically, and modeling things such as insurance risk, customer behavior, ecology, and business management. In summary, the usefulness of the Pareto distribution lies in its capacity to accurately represent uneven distributions, rendering it critical in many fields, especially in the analysis of survival and healthcare data.

Fuzzy theory, which was pioneered by Lotfi Zadeh in the 1960s, is an extension of classical set theory in which an element is permitted to have partial membership in a set rather than being completely in or out of it [11]. Examples demonstrate that this flexibility suggests possible relations of fuzziness in theory that can be helpful about vagueness and uncertainty that are normally available. They use membership functions which map [0, 1] to represent the degree of membership [12-13]. Because of its inherent uncertainty, fuzzy theory has been widely used in related domains including artificial intelligence (AI), decision-making, control systems, and pattern recognition. Fuzzy logic in statistics: Fuzzy theory is an extension to apply fuzzy logic in statistical analysis for uncertain, ambiguous, or incomplete data [14]. For classical approach, the statistics tool such as average, variance, and standard deviation are used to analyze data, while in fuzzy statistics these concepts are generalized which can be used under fuzziness. This idea has been useful in a variety of domains (e.g., medical diagnostics, environmental modeling, economic forecasting), where data may be noisy, subjective, or ambiguous. Fuzzy statistics is a combination of fuzzy logic and statistical techniques that provides a powerful tool for modeling and analyzing uncertainty in complex, real-world systems [15]. The neutrosophic theory, proposed by Florentin Smarandache in the late 20th century, is an extension of classical set theory, which defines a new degree of truth that includes indeterminate truth [16-17]. In neutrosophy, unlike binary logic where an element is true or false, an element can have indeterminate or neutral truth values represented by three components where T = truth, I = indeterminate and F = falsity. This is a more flexible framework to handle vagueness and uncertainty, which allows it to be particularly suited to applications in Artificial Intelligence, Decision Making, and Fuzzy Logic systems, where real-world complexity often do not allow for precise reasoning. Neutrosophic statistics is an extension of classical statistics, which is applied to statistical analysis to deal with indeterminacy, incomplete information, and contradictions in data [18-20]. Traditional statistics assumes well-defined data, but in most real-life scenarios, the data can be ambiguous, uncertain or subject to perception. Neutrosophic statistics extends classical statistical notions (mean, variance, probability distributions, and so on) to accommodate the three components (truth, indeterminate, falsity) and hence provides a more realistic model for the associated phenomena [21]. It allows us to analyze data that does not comply with classical structures and offers strong solutions when working with uncertain or incomplete data. Neutrosophic statistics are applied in many fields such as medical diagnosis, environmental monitoring, social sciences, etc., since there is often inherent uncertainty or ambiguity in data [22-24].

The probabilistic models in neutrosophic distributions extend neutrosophic theory into the domain of probability distributions by introducing truth, indeterminacy, and falsity into the framework of these models. These distributions are used for modelling systems whereby the outcomes cannot only be true or false but can also “not fully determined” [25]. Neutrosophic distributions are used in real-world situations where traditional distributions are unable to accurately model complex uncertain phenomena, e.g., neutrosophic normal distribution, neutrosophic exponential distribution names a few. Models of indeterminate probability distributions based on possibility theory have been proposed and utilized in different fields such as reliability analysis, risk assessment, and decision making where uncertainty and indeterminacy play a crucial role [26]. Neutrosophic theory, statistics, and distributions are strong methods for examining complex, uncertain, and imprecise data, especially by adding truth, indeterminacy, and falsity to statistical models [27-30].

DOI: https://doi.org/10.54216/IJNS.250426

In this study, we investigate the neutrosophical characteristics of the Pareto model and its relation to survival and health care data. In this context, we are using the three parts of the neutrosophic logic (truth, indeterminacy, and falsity) in the Pareto distribution to encapsulate the reality of uncertainty and imprecision that is often involved in healthcare and survival modeling in real life. While standard model may lack the capacity to model ambiguous or missing data. This approach can hinder more flexible modeling of patient survival times, treatment outcomes, and healthcare costs. For more accurate and comprehensive analyses in these realms, the neutrosophic Pareto model represents a reasonable model.

The outline of the paper is as follows: In Section 2, we introduce the neutrosophic generalization of the Pareto model. The mathematical approach to estimate the unknown distributional parameters is elaborated in Section 3. A Monte Carlo simulation to test the theoretical results of the neutrosophic model is introduced in Section 4. The proposed model is applied on real-world data in Section 5. Section 6 concludes with some key findings from the study.

2. Preliminaries

In this section, some statistical properties of the classical Pareto model including its probability density function (PDF), cumulative distribution function (CDF) and related shape parameters are discussed so we can understand the proposed model under neutrosophic framework.

A random variable Z is said to follow Pareto model if it has the following PDF and CDF:

$$g(z,\beta,\alpha) = \frac{\beta\alpha^{\beta}}{z^{\beta+1}} \text{ for } 0 \leq z < \infty;\ \beta,\alpha,\ z > 0 \quad (1)$$

$$G(z,\beta,\alpha) = 1 - \left(\frac{\alpha}{z}\right)^{\beta} \quad (2)$$

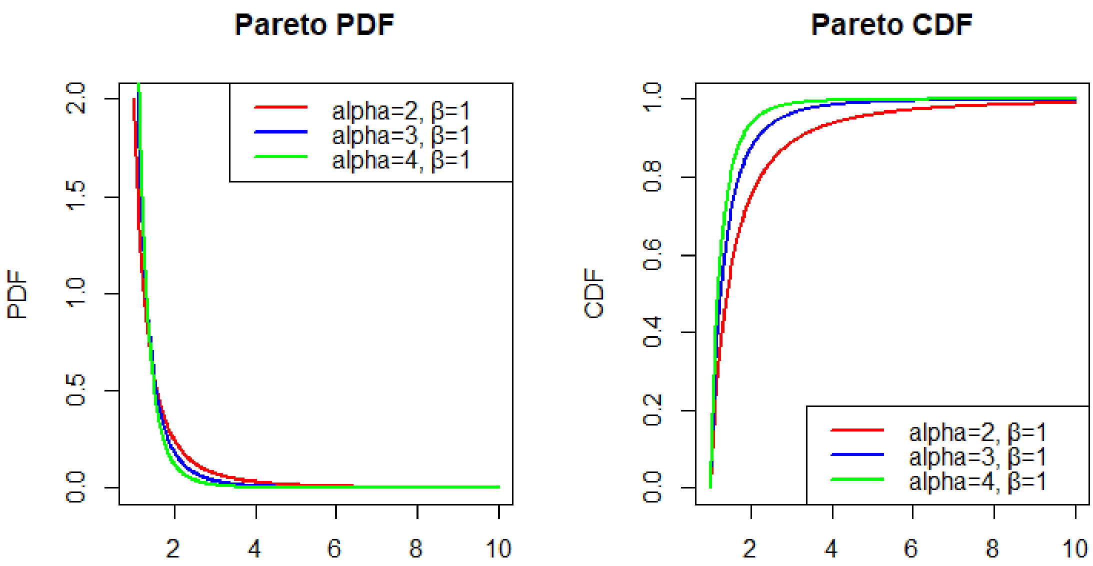

Figure 1. PDF and CDF plot of the Pareto Model with various parameter values

The PDF and CDF of the Pareto distribution are plotted in the first and second subplots respectively as shown in Figure 1 for three sets of parameter α and β. The left panel shows the PDF, illustrating the change of the tail behaviour of the distribution with different α. The right panel shows the corresponding CDF, where the cumulative probabilities can be observed as xx rises. Different colors represent the different curves, where legends providing the values of α and β are included for better understanding.

The hazard function is related concept of any probability distribution with major applications in reliability and survival studies. The hazard function of the Pareto model is defined as:

$$h(z) = \frac{\alpha\beta^{\alpha}}{z^{\alpha+1}} \cdot \frac{1}{1-\left(\frac{\beta}{z}\right)^{\alpha}}, \quad z \geq 0 \quad (3)$$

DOI: https://doi.org/10.54216/IJNS.250426

The hazard function is important in survival analysis because it describes the instantaneous risk that an event will occur at a given time, given that the event has not yet occurred until that time. It assists researchers in determining how probable it is to experience failure, death, or other significant moments, allowing for comparative analyses between groups, evaluations of treatment impacts, and proficient models of time-to-event data for better decision-making.

Reliability function of the Pareto model can also be found by using the expression as given below:

$$R(z) = 1 - F(z) = \left(\frac{\beta}{z}\right)^{\alpha}, \quad z \geq 0 \tag{4}$$

In reliability and survival studies, the reliability function represents the probability that a system or component will perform its intended function for a specified period; hence, it is of utmost importance. This is useful to assess how long product(s) lives, predict when they fail, when to service them, all of which allow for proper functioning of the system and helps maximize their usefulness.

Similarly, other essential properties of the Pareto model can be outlined as:

Mean

$$\mu = \frac{\alpha\beta}{\alpha-1}, \alpha > 1 \tag{5}$$

Variance

$$\sigma^2 = \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)}, \alpha > 2 \tag{6}$$

Skewness Coefficient

$$\gamma_1 = \frac{2(1+\alpha)}{\alpha-3}\sqrt{\frac{\alpha-2}{\alpha}}, \alpha > 3 \tag{7}$$

Kurtosis Coefficient

$$\gamma_2 = \frac{6(\alpha^3+\alpha^2-6\alpha-2)}{\alpha(\alpha-3)(\alpha-4)}, \alpha > 4 \tag{8}$$

These are some basic properties are essential to see for understanding before fitting of Pareto model to real-word scenario. Now in the following section, we will study the structure of Pareto model under neutrosophic framework.

3. Neutrosophic structure

A neutrosophic PDF is the generalization of the classical PDF, taking into consideration the neutrosophic truth (T), indeterminacy (I), and falsity (F) components to capture uncertainty of data and imprecision. In classical statistics, PDF is the function that shows us how likely it is for a random variable to take on certain values, assuming that there is no uncertainty or imprecision in the data. It is well defined through deterministic parameters, i.e. exact probabilities in a fixed framework. On the other hand, the neutrosophic pdf is defined in the context of neutrosophic statistics, where its parameters and the function itself can be intervals or neutrosophic numbers. It allows the model to reflect the uncertainty, imprecision, or incompleteness often seen in real-world datasets.

The PDF of the proposed NMP for uncertain parameters β_n and α_n is given by:

$$h(t, \beta_n, \alpha_n) = \frac{\beta_n \alpha_n^{\beta_n}}{t^{\beta_n+1}} \text{ for } 0 \leq t < \infty;\ \beta_n, \alpha_n, t > 0 \tag{9}$$

Notably, the primary difference resides in the manner uncertainty is treated; where the classical PDF assumes perfect knowledge, the neutrosophic pdf thrives on uncertainty providing an approach to quantify and incorporate it to model real-life data cases. The PDF of the NPM with different values of β_n and fixed $\alpha_n = [1.5, 1.5]$ are shown in Figure 2.

DOI: https://doi.org/10.54216/IJNS.250426

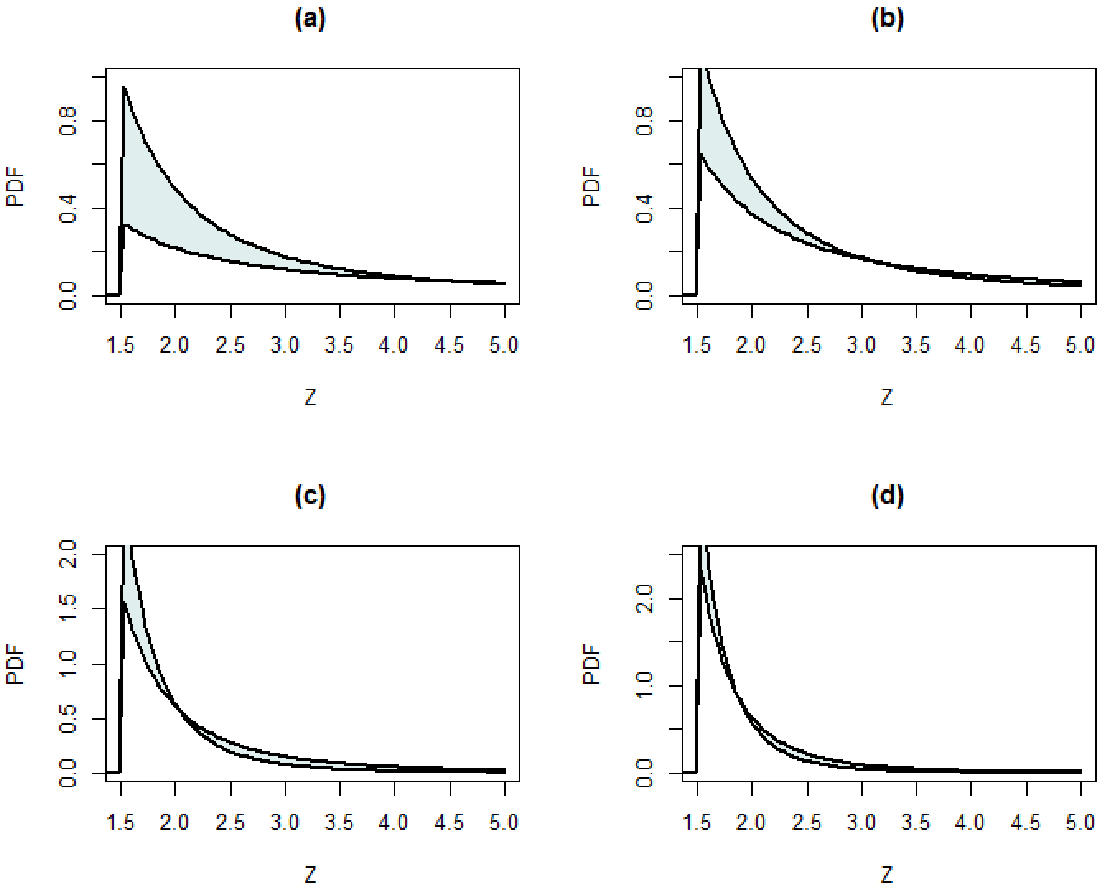

Figure 2. PDF curves of the proposed model with (a) $\beta_n = [0.5,1.5]$ (b)) $\beta_n = [1,1.8]$ (c)) $\beta_n = [2.5,4.2]$ (d)) $\beta_n = [3.5,5.4]$

The influence of unknown neutrosophic parameters on the shape of the Neutrosophic Pareto Distribution is highlighted in figure 1. The figure illustrates the fact that the distribution represents and embraces uncertainty in the form of intervals or ranges for the parameters, reflecting the truth (T), indeterminacy (I), and falsity (F) that characterize the data. In contrast with classical Pareto distribution, where parameters are definite and known, the neutrosophic contribution facilitates a better description of uncertainty and ambiguity. This Norman adaption emphasizes and shows the use of Neutrosophic Pareto Distribution, as a mean of express variables in real life situations, where the exact values of the various parameters are not known, or are not very clearly defined.

The other related function of PDF is the CDF function. The classical CDF is generalized as neutrosophic CDF, which comprises neutrosophic components: truth , indeterminacy and falsity to reflect uncertainty and vagueness in data. It represents the likelihood that a neutrosophic random variable is no greater than a specified value. Neutrosophic CDF has multiple significances: it is not limited to a classical interaction to events as it allows its probabilities and parameters to be given in the form of intervals or neutrosophic numbers. T

$$H(t, \beta_n, \alpha_n) = 1 - \left(\frac{\alpha_n}{t}\right)^{\beta_n} \quad (10)$$

The CDF of the NPM with different values of β_n and fixed $\alpha_n = [1.5, 1.5]$ are shown in Figure 3.

DOI: https://doi.org/10.54216/IJNS.250426

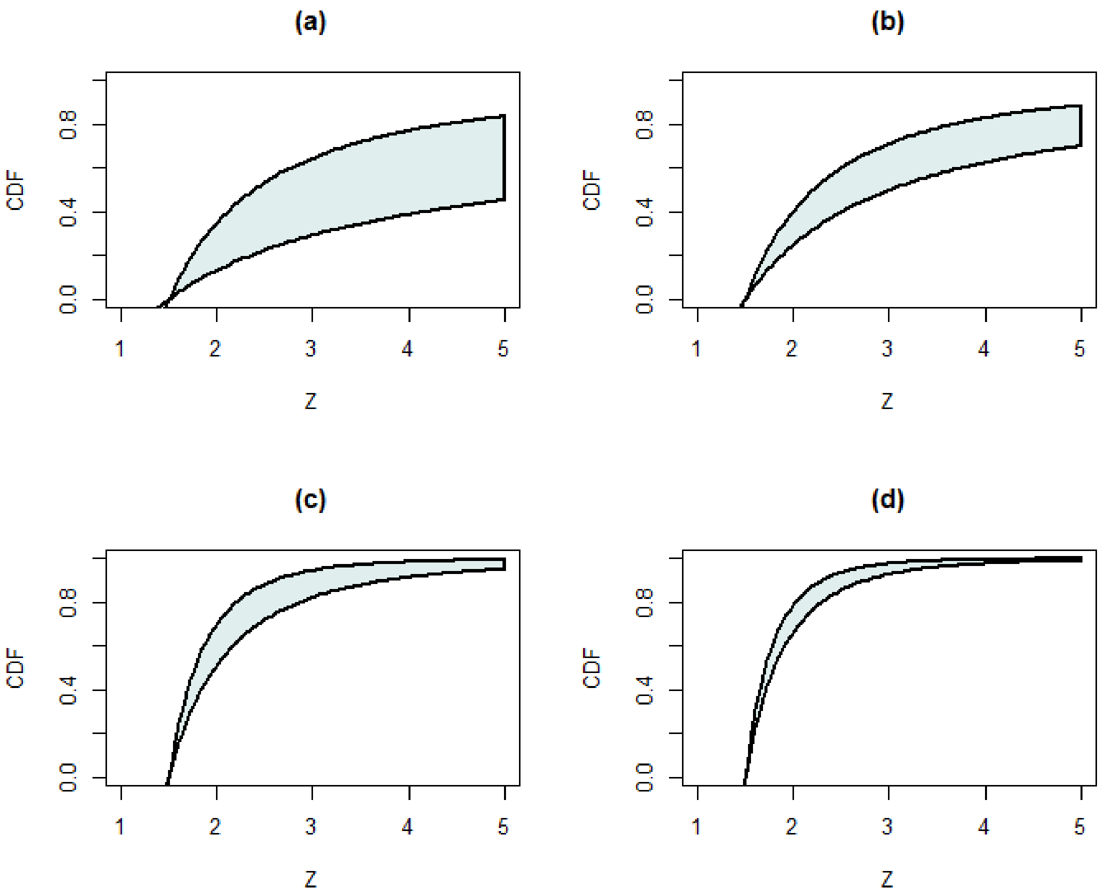

Figure 3. CDF curves of the proposed model with (a) $\beta_n = [0.5,1.5]$ (b)) $\beta_n = [1,1.8]$ (c)) $\beta_n = [2.5,4.2]$ (d)) $\beta_n = [3.5,5.4]$

In Figure 3, we demonstrate the cumulative distribution function (CDF) of the Neutrosophic Pareto Model, illustrating how uncertain parameter values can affect model behavior. To proper identify the uncertainty in parameter estimates, neutrosophic components (truth (T), indeterminacy (I) and falsity (F)) were incorporated into the CDF, we can observe the range or band of CDF instead of a single curve. The complementary cumulative distribution function (CDF) adapts to a broad range of possibilities, showcasing the versatility of the neutrosophic methodology in interpreting data vagueness. The imprecise description of parameters in this problem is inevitable in the real world, so such representation makes the Neutrosophic Pareto Model the best approach to extract the cumulative probabilities than we can do using Survival Data Models, thereby making the Neutrosophic Pareto Model an ideal choice for modeling cumulative probabilities of survival data in a realistic way

The Hazard function of the NPM can be defined as:

$$\vartheta(t) = \frac{\beta_n \alpha_n^{\beta_n}}{t^{\beta_n+1}\left(1-\left(\frac{\alpha_n}{t}\right)^{\beta_n}\right)} \tag{11}$$

The hazard function (also referred to as the failure rate or hazard rate) is a core concept in survival analysis, which describes the instantaneous rate of failure at a given time if the subject has survived until that time. From a mathematical point of view, it is defined by the ratio of the pdf with respect to survival function (1 - CDF). The hazard function is especially useful in the context of time-to-event data in highlighting how the hazard of an event changes over time. This can also reveal qualitative aspects, such as whether the risk levels are increasing or decreasing or remain constant, and is essential in applications such as medicine, engineering, and reliability analysis. Thus, by offering insights into the underlying failure mechanisms, the hazard function assists in devising strategies for enhancing survival rates, optimizing maintenance schedules, and assessing interventions, rendering it an invaluable component of survival analysis.

DOI: https://doi.org/10.54216/IJNS.250426

Reliability function is another important function in the reliability and survival studies. The reliability of the proposed model can be written as:

$$R(t) = 1 - \left(1 - \left(\frac{\alpha_n}{t}\right)^{\beta_n}\right) \quad (12)$$

Reliability function is also known as the survival function in reliability engineering and survival analysis. It is the likelihood of a system, component or individual will perform its intended function without failure for a specified period of time. The reliability function is a rich source of information, which tells us how much longer the system, on average, will stay alive so we can assess the performance over time. It is a widely used approach in product design and testing, maintenance prediction, and risk assessment. Recognizing the most mix-up point with the system, the decision uses dependability capacity to improve framework plans, shape appropriate upkeep plans, and lessen operational hazards.

Other essential properties of the NPM can be established easily under neutrosophic calculus. For example, neutrosophic mean of NPM can be derived as:

$$E(t) = \beta_n \alpha_n^{\beta_n} \int_{\alpha_n}^{t} \frac{t}{t^{\beta_n+1}} dt = \left[\beta_l \alpha_l^{\beta_l} \int_{\alpha_l}^{t} \frac{t}{t^{\beta_l+1}} dt\, , \beta_u \alpha_u^{\beta_u} \int_{\alpha_u}^{t} \frac{t}{t^{\beta_u+1}} dt\right]$$

$$= \left[\frac{\beta_l \alpha_l}{\beta_l - 1}, \frac{\beta_u \alpha_u}{\beta_u - 1}\right] = \frac{\beta_n \alpha_n}{\beta_n - 1} \quad (13)$$

which is the required mean of the NPM.

Likewise, the variance of NPM can be defined as:

$$var(t) = E(t^2) - [E(t)]^2 \quad (14)$$

where

$$E(t^2) = \beta_n \alpha_n^{\beta_n} \int_{\alpha_n}^{t} \frac{t^2}{t^{\beta_n+1}} dt = \left[\beta_l \alpha_l^{\beta_l} \int_{\alpha_l}^{t} \frac{t^2}{t^{\beta_l+1}} dt\, , \beta_u \alpha_u^{\beta_u} \int_{\alpha_u}^{t} \frac{t^2}{t^{\beta_u+1}} dt\right] \left[\frac{\beta_l \alpha_l}{\beta_l - 2}, \frac{\beta_u \alpha_u}{\beta_u - 2}\right] = \frac{\beta_n \alpha_n}{\beta_n - 2} \quad (15)$$

Utilizing Eq (13), Eq (14) and Eq (15) , the following expression can be obtained:

$$\text{var(t)} = \left[\frac{\beta_l \alpha_l^2}{(\beta_l - 1)^2(\beta_l - 2)}, \frac{\beta_u \alpha_u^2}{(\beta_u - 1)^2(\beta_u - 2)}\right] = \frac{\beta_n \alpha_n^2}{(\beta_n - 1)^2(\beta_n - 2)} \quad (16)$$

which is required variance of the NPM.

The rth moment can be established with expression given below:

$$\mu_{rn} = \left[\beta_n \alpha_n^{\beta_n} \int_{\alpha_n}^{t} \frac{t^r}{t^{\beta_n+1}} dt\right] \quad (17)$$

$$= \left[\beta_l \alpha_l^{\beta_l} \int_{\alpha_l}^{t} \frac{t^r}{t^{\beta_l+1}} dt, \beta_u \alpha_u^{\beta_u} \int_{\alpha_u}^{t} \frac{t^r}{t^{\beta_u+1}} dt\right]$$

Further simplification of Eq (17) provided:

$$\mu_{rn} = \left[\frac{\beta_l \alpha_l^r}{(\beta_l - r)}, \frac{\beta_u \alpha_u^r}{(\beta_u - r)}\right] = \frac{\beta_n \alpha_n^r}{(\beta_n - r)} \quad (18)$$

Now putting different values of $r = 1, 2,3,4. ..$,we can obtain different moments about origin. For example first moment can be obtained using $r = 1$ as:

$$\mu'_{1n} = \mu_{1n} = \frac{\beta_n \alpha_n}{(\beta_n - 1)}$$

which is mean of the NPM.

Similarly other origin moments and their relation with moments about mean, we can derive importance properties as:

$$\mu'_{2n} = \mu_{2n} - (\mu_{1n})^2 = \frac{\beta_n \alpha_n^2}{(\beta_n - 1)^2(\beta_n - 2)} \quad (19)$$

DOI: https://doi.org/10.54216/IJNS.250426

$$\mu'_{3n} = \mu_{3n} - 3\mu_{2n}\mu_{1n} + 2(\mu_{1n})^3 = \frac{2\beta_n(\beta_n+1)\alpha_n^3}{(\beta_n-1)^3(\beta_n-2)(\beta_n-3)} \quad (20)$$

$$\mu'_{4n} = \mu_{4N} - 4\mu_{3n}\mu_{1n} + 6\mu_{2n}(\mu_{1n})^2 - 3(\mu_{1n})^4 = \frac{3\beta_n(3\beta_n^3+\beta_n+3)\alpha_n^4}{(\beta_n-1)^4(\beta_n-1)(\beta_n-3)(\beta_n-4)} \quad (21)$$

The shape coefficients can be defined as:

Skewness

$$\alpha_3 = \frac{\mu'_{3n}}{(\mu'_{2n})^{\frac{3}{2}}} \quad (22)$$

where $\mu'_{3n} = \frac{2\beta_n(\beta_n+1)\alpha_n^3}{(\beta_n-1)^3(\beta_n-2)(\beta_n-3)}$ (23)

and $\mu'_{2n} = \frac{\beta_n\alpha_n^2}{(\beta_n-1)^2(\beta_n-2)}$ (24)

Simplification of Eq (23) and Eq (24) yielded:

$$\alpha_3 = \frac{2(\beta_n+1)}{(\beta_n-3)}\sqrt{\frac{\beta_n-2}{\beta_n}} \quad (25)$$

Kurtosis coefficient of the model can be defined as:

$$\alpha_4 = \frac{\mu'_{4n}}{(\mu'_{2n})^2} \quad (26)$$

where $\mu'_{4n} = \frac{3\beta_n(3\beta_n^3+\beta_n+3)\alpha_n^4}{(\beta_n-1)^4(\beta_n-1)(\beta_n-3)(\beta_n-4)}$ (27)

and $\mu'_{2n} = \frac{\beta_n\alpha_n^2}{(\beta_n-1)^2(\beta_n-2)}$ (28)

Putting Eq (27) and Eq (28) in Eq (26) resulted:

$$\alpha_4 = \frac{6(\beta_n^3+\beta_n^2-6\beta_n-2)}{\beta_n(\beta_n-3)(\beta_n-4)} \quad (29)$$

These are some essential properties of the proposed model.

4. Estimation Method

Maximum Likelihood Estimation (MLE) is a statistical method for estimating parameters of a probability distribution, which maximizes the likelihood function. In this approach, the parameter values are chosen that maximize the likelihood of the observed data. MLE is very popular, as it possesses some nice property such as consistency, efficiency and asymptotic normality. It is a key component of many fields; e.g., economics, machine learning, biostatistics, to name a few, that enables a robust parameter estimation of complex models. MLE is a simple yet very general point estimation method that can be used to derive a direct measurement from data for complex models via numerical optimization methods.

The likelihood function of the proposed model with uncertainty parameters is given as:

$$\emptyset(\eta_n, \alpha_n|t) = \prod_{i=1}^{k} h(t_i|\beta_n, \alpha_n) \quad (30)$$

Including the value of PDF in Eq (30) yielded:

$$\emptyset(\eta_n, \alpha_n|t) = \prod_{i=1}^{k}\left[\frac{\beta_n\alpha_n^{\beta_n}}{t^{\beta_n+1}}\right]$$

$$= \beta_n\alpha_n^{\beta_n}\prod_{i=1}^{m}\frac{1}{t^{\beta_n+1}} \quad (31)$$

Taking the log of Eq (31) provided

DOI: https://doi.org/10.54216/IJNS.250426

$l_n(\mathcal{T}_i|\eta_n, \alpha_n) = \log\left[\beta_n \alpha_n^{\beta_n} \prod_{i=1}^{k} \frac{1}{t^{\beta_n+1}}\right]$ (32)

Simplification of Eq (32) yielded:

$$l_n(\mathcal{T}_i|\eta_n, \alpha_n) = k\log(\beta_n) + k - (\beta_n + 1)\sum_{i=1}^{k} \log t_i$$

$$\left[\frac{\delta l_n(t_i, \eta_n)}{\delta\beta_n}, \frac{\delta\omega_n(t_i, \alpha_n)}{\delta\alpha_n}\right] = [0,\ 0]$$

Finally maximization of Eq (32) provided:

$$\widehat{\alpha_n} = \min t_i$$

$$\widehat{\beta_n} = \frac{k}{\sum_{i=1}^{k} \log t_i - k\log(\alpha_n)}$$

Thus, the estimators for unknown itself are imprecise because of uncertainty in the underlying data. Now we can use the Monte Carlo technique to create random numbers which should approximately follow the NPM. In general, the Monte Carlo method includes any technique that makes use of random results in order to solve a problem. The main objective of this work is to validate the theoretical results shown in previous section by simulating random samples from the NPM with a known setting of parameters via the Monte Carlo approach. The inverse CDF method can be used as a direct method for generating random samples from the proposed model. This approach takes advantage of a computer's native pseudo-random number generator to generate random numbers. The inverse CDF of the proposed model can be written as:

$R = \alpha_n[1 - P]^{\frac{-1}{\beta_n}}$ (33)

Here, P indicates random numbers of points taken from the uniform distribution, and R is a desired percentile value of proposed model. Total 40 random samples are generated from the proposed model using the inverse CDF method with parameters setting $\alpha_n = [1.5, 1.5]$ and $\beta_n = [2.5, 3.5]$.For selected values.

These 40 random samples with specific seed setting are given in Table 1.

Table 1: Random sample generated from the proposed model

Random samples				
[1.96, 2.40]	[1.58, 1.64]	[1.76, 1.99]	[2.28, 3.12]	[1.65, 1.78]
[1.57, 1.62]	[1.51, 1.52]	[1.75, 1.97]	[1.62, 1.71]	[1.68, 1.84]
[5.71, 15.56]	[1.62, 1.71]	[1.53, 1.56]	[2.05, 2.60]	[1.97, 2.43]
[1.64, 1.76]	[1.79, 2.05]	[1.55, 1.59]	[1.53, 1.55]	[2.90, 4.76]
[2.35, 3.30]	[2.53, 3.75]	[2.43, 3.48]	[5.75, 15.77]	[1.91, 2.30]
[2.42, 3.47]	[1.75, 1.97]	[1.51, 1.52]	[1.78, 2.03]	[1.54, 1.58]
[2.43, 3.50]	[2.11, 2.72]	[1.90, 2.27]	[1.64, 1.76]	[9.56, 38.39]
[1.56, 1.61]	[1.97, 2.41]	[1.81, 2.09]	[1.73, 1.94]	[2.17, 2.87]

Table 1 shows neutrosophic random samples from the NPM with specific parameters settings and random seed. All the samples are in interval form due to imprecision assumed in the model parameter. If we assume zero indeterminacy in the model parameters, these results would converge to results obtained from the classic model. Likewise, if we randomly select 10000 random samples from the model with already mentioned parameters we

DOI: https://doi.org/10.54216/IJNS.250426

obtain analytical solutions as a function of the baseline parameter values, based on the theoretical results from Section 3. Table 2 lists the estimated values of the different distribution characteristics.

Table 2: Simulated data based statistical properties of the NPM

Basic Properties of NPM	Simulated Results
Mean	[2.10, 3.01]
var	[0.89, 28.64]
Skewness	[9.38, 35.64]
Kurtosis	[205.69, 1733.61]

Table 2 shows the descriptive statistics of the proposed model for the given distributional parameter values. The descriptive metrics for the simulated data are provided as ranges to illustrate uncertainties assumed in the parameter definitions. It is clear from the numerical findings that classical model of the Pareto distribution is not capable of analyzing data involving uncertainty.

5. Conclusion

This study presents the neutrosophic Pareto model (NPM), a new survival model used to tackle uncertainty in the survival data. The proposed model extends the classical Pareto distribution by means of a neutrosophic statistical framework that is used to deal with fuzzy or imprecise data typically arising in practice. Rigorous derivation of key mathematical properties such as moments, survival function and hazard rate ensured theoretical validity of the model. We have used the maximum likelihood estimation (MLE) method to estimate the neutrosophic parameters, showing that with the use of MLE, we can obtain good estimates of parameters even when data is uncertain. The experiments with simulation proved the stability and those resulted performance of the model in different settings, demonstrating the reliability of the model for practical usage. Moreover, the simulated data experimentation demonstrated the feasibility of using the developed model to retrieve useful knowledge from datasets with uncertainty. This indicates the model can be applied to real world data in survival studies.

Acknowledgement: The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2024/01/31702)

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] L. D. Case and T. M. Morgan, "Design of Phase II cancer trials evaluating survival probabilities," *BMC Medical Research Methodology*, vol. 3, pp. 1–12, 2003.

[2] E. T. Lee and J. Wang, *Statistical Methods for Survival Data Analysis*, vol. 476. John Wiley & Sons, 2003.

[3] J. Del Castillo and P. Puig, "Distinguishing between a power law and a Pareto distribution," *Physical Review E*, vol. 107, no. 6, pp. 064113, 2023.

[4] P. Hougaard, "Fundamentals of survival data," *Biometrics*, vol. 55, no. 1, pp. 13–22, 1999.

[5] L. Ndwandwe, J. S. Allison, L. Santana, and I. J. H. Visagie, "Testing for the Pareto type I distribution: A comparative study," *Metron*, vol. 81, no. 2, pp. 215–256, 2023.

[6] R. Chattamvelli and R. Shanmugam, "Pareto distribution," in *Continuous Distributions in Engineering and the Applied Sciences, Part II*. Springer, 2021, pp. 179–188.

[7] Y. Chen, P. Embrechts, and R. Wang, "An unexpected stochastic dominance: Pareto distributions, dependence, and diversification," *Operations Research*, 2024.

[8] K. Jayakumar, B. Krishnan, and G. G. Hamedani, "On a new generalization of Pareto distribution and its applications," *Communications in Statistics-Simulation and Computation*, vol. 49, no. 5, pp. 1264–1284, 2020.

[9] H. Haj Ahmad, E. M. Almetwally, M. Elgarhy, and D. A. Ramadan, "On unit exponential Pareto distribution for modeling the recovery rate of COVID-19," *Processes*, vol. 11, no. 1, pp. 232, 2023.

[10] N. Nanuwong, W. Bodhisuwan, and C. Pudprommarat, "A new mixture Pareto distribution and its application," *Thailand Statistician*, vol. 13, no. 2, pp. 191–207, 2015.

[11] L. A. Zadeh, "Fuzzy logic—A personal perspective," *Fuzzy Sets and Systems*, vol. 281, pp. 4–20, 2015.

DOI: https://doi.org/10.54216/IJNS.250426

[12] L. A. Zadeh, "Fuzzy logic," in *Granular, Fuzzy, and Soft Computing*. 2023, pp. 19–49.

[13] L. A. Zadeh, "Discussion: Probability theory and fuzzy logic are complementary rather than competitive," *Technometrics*, vol. 37, no. 3, pp. 271–276, 1995.

[14] R. Berkachy, *Fuzzy Random Variables and Fuzzy Distributions: The Signed Distance Measure in Fuzzy Statistical Analysis*. Springer, 2021, pp. 87–113.

[15] L. González-De La Fuente, A. Nieto-Reyes, and P. Terán, "Projection depth and Lr-type depths for fuzzy random variables," *Fuzzy Sets and Systems*, vol. 487, pp. 108991, 2024.

[16] U. Rivieccio, "Neutrosophic logics: Prospects and problems," *Fuzzy Sets and Systems*, vol. 159, no. 14, pp. 1860–1868, 2008.

[17] F. Smarandache, "Neutrosophic statistics is an extension of interval statistics, while plithogenic statistics is the most general form of statistics (second version)," *Infinite Study*, vol. 2, 2022.

[18] F. Smarandache, *A Unifying Field in Logics: Neutrosophic Logic. Neutrosophy, Neutrosophic Set, Neutrosophic Probability*. Infinite Study, 2005.

[19] S. Broumi, *Handbook of Research on the Applications of Neutrosophic Sets Theory and Their Extensions in Education*. IGI Global, 2023.

[20] K. J. Silva and R. D. Gupta, "Neutrosophic exponential power distribution: Application to medical survival data," *International Journal of Neutrosophic Analysis and Applications*, vol. 18, no. 1, pp. 45–56, 2021.

[21] A. M. M. Ibrahim, Z. Khan, F. S. Al-Duais, et al., "A new modified logistic distribution: Properties and applications in uncertainty data modeling," *Full Length Article*, vol. 20, no. 2, pp. 27–7, 2023.

[22] M. K. Ullah, Z. H. Khan, and M. S. Noor, "Application of neutrosophic statistics in quality control: A case study on manufacturing data," *International Journal of Statistical Analysis and Applications*, vol. 20, no. 4, pp. 301–310, 2021.

[23] Z. Khan, M. M. A. Almazah, O. Hamood Odhah, and H. M. Alshanbari, "Generalized Pareto model: Properties and applications in neutrosophic data modeling," *Mathematical Problems in Engineering*, vol. 2022, no. 1, pp. 3686968, 2022.

[24] M. Aslam, R. A. R. Bantan, and N. Khan, "Design of a new attribute control chart under neutrosophic statistics," *International Journal of Fuzzy Systems*, vol. 21, pp. 433–440, 2019.

[25] A. A. Alvaracín Jarrín, D. S. Proaño Tamayo, S. A. Montecé Giler, J. C. Arandia Zambrano, and D. M. Macazana, "Neutrosophic statistics applied in social science," *Neutrosophic Sets and Systems*, vol. 44, no. 1, pp. 1, 2021.

[26] U. Afzal, H. Alrweili, N. Ahmad, and M. Aslam, "Neutrosophic statistical analysis of resistance depending on the temperature variance of conducting material," *Scientific Reports*, vol. 11, no. 1, pp. 23939, 2021.

[27] S. Ahmad and T. Qureshi, "Neutrosophic statistical tests for outlier detection with application in health sciences," *Journal of Applied Mathematics and Computational Statistics*, vol. 45, no. 2, pp. 123–132, 2022.

[28] R. A. K. Sherwani, H. Shakeel, M. Saleem, W. B. Awan, M. Aslam, and M. A. Farooq, "A new neutrosophic sign test: An application to COVID-19 data," *PLOS ONE*, vol. 16, no. 8, pp. e0255671, 2021.

[29] G. Işik and İ. Kaya, "Design of single and double acceptance sampling plans based on neutrosophic sets," *Journal of Intelligent & Fuzzy Systems*, vol. 42, no. 4, pp. 3349–3366, 2022.

[30] R. Ahmed, N. Uddin, and M. H. Chowdhury, "A neutrosophic approach to inventory management with shortages under fuzzy demand," *International Journal of Neutrosophic Systems*, vol. 16, no. 3, pp. 37–48, 2021.

DOI: https://doi.org/10.54216/IJNS.250426