International Journal of Robotics and Control Systems
Vol. 5, No. 6, 2025, pp. 3318-3349
ISSN 2775-2658
http://pubs2.ascee.org/index.php/ijrcs

Flexible Distributional Modeling with the ACT-G Family Using Recurrent Neural Networks and Firefly Optimization

Adel S. Hussain [a,*], Maysoon Qousini [b,1], Rana N. Abbas [c], Emad A. Az-Zo'bi [d], Mohammad A. Tashtoush [e,f]

[a] IT Department, Amedi Technical Institutes, University of Duhok Polytechnic, Duhok, Iraq
[b] Faculty of Science and Information technology, Al-Zaytoonah University of Jordan, Amman 11183, Jordan
[c] Ministry of Education, Baghdad College, Baghdad, Iraq
[d] Department of Mathematics and Statistics, Faculty of Sciences, Mutah University, Al-Karak, Jordan
[e] Department of Basic Sciences, Al-Huson University College, Al-Balqa Applied University, Salt 19117, Jordan
[f] Faculty of Education and Arts, Sohar University, Sohar 311, Oman
[1] m.qousini@zuj.edu.jo
* Corresponding Author

ARTICLE INFO

Article history
Received July 07, 2025
Revised September 10, 2025
Accepted December 30, 2025

Keywords
Distribution Theory;
Flexible Distributions;
Extreme Value Modeling;
Maximum Likelihood Estimation;
Recurrent Neural Networks with Firefly Algorithm;
Simulation Study;
Applied Data Analysis

ABSTRACT

Skewed and heavy-tailed data modeling has continued to be the focal issue in survival, reliability, and applied sciences, because current distribution families are often rigid or difficult to compute. We address this gap with a new family, the arc-cosine-tangent-generators family, which extends sine-based and exponentiated-based generators by a relatively simple transformation. The most significant contribution of the research is the presentation of the ACT-G family, providing more flexibility of tail and skewness control without new shape parameters, and with closed-form expressions of the density, distribution, and hazard functions. As a special case, we derive the ACTE Extreme Value distribution and obtain its statistical properties. One is maximum likelihood estimation and a hybrid machine learning solution relying on recurrent neural networks and Firefly Algorithm to overcome convergence issues on high-dimensional likelihood surfaces. Experiments in simulation were performed with different sample sizes and different parameters, and then applied to music therapy data. Findings indicate that ACTE has always performed better than the ASTE-exponential and traditional maximum likelihood estimators, with lower bias and mean squared error, better tail fit and better robustness. It is also adequately supported by information criteria (AIC, BIC) and goodness-of-fit tests (KS, AD). Finally, the ACT-G family is a statistically rigorous and practically useful model of skewed and heavy tailed data, and, naturally, which also has some advantages in simulation and in the real world.

1. Introduction

Most applied fields, including healthcare, reliability engineering, economics and the environmental sciences, frequently have data with skewness or heavy tails, meaning that some extreme events are more common than they would otherwise be in classical distributions. The exponential or Weibull distribution are both useful as standard models but they are not flexible enough

to explain these intricate phenomena. The statistical toolbox has been expanded with generators-based extensions (e.g. sine-G, exponentiated-G, Beta-G, Kumaraswamy-G families) that enhance distributional flexibility. However, such techniques typically require an extra-shape parameter, non-tractable reparameterization or mathematical non-tractability, and inference becomes more difficult and introspective. This emphasizes the fact that a distributional framework should be both malleable and computationally feasible [1]-[5].

Many real-world datasets exhibit skewness, heavy tails, or complex hazard functions that classical distributions such as exponential and Weibull cannot adequately capture. Existing generator families (sine-G, exponentiated-G, etc.) improve flexibility but often add parameters or lose tractability, making inference difficult. This gap is important because inadequate tail or hazard modeling can bias results in critical applications such as survival analysis, reliability engineering, and biomedical studies. The current paper proposes the arc-cosine-tangent-G (ACT-G) family in response to this shortcoming. Using arccosine and tangent transformations, ACT-G gives explicit control over skew and tail heaviness, and retains closed-form expressions of the key functions, probability density, cumulative distribution, and hazard rate. The ACT-G has increased flexibility without added free parameters, unlike our previous family of generators, and therefore is parsimonious. Besides, the framework combines the recent machine learning methods, in which recurrent neural networks are used to estimate complex likelihood surfaces and Firefly Algorithm is used to promote global optimization. It is an intermediate method that upscales statistical theory with computational gains to achieve more robust parameter estimation in practice.

In order to overcome these limitations, we introduce the arc cos--tangent--G (ACT-G) family that incorporates arccosine and tangent transformations into the basic distributions. The resulting model provides more control over the skewness and tail behavior of the unrepaired periods, while allowing closed-form expressions of key distributional characteristics (i.e., pdf, cdf, and hazard rate). Importantly, ACT-G emerges as a universal construction because sine and exponent-based generators are special cases, as noted by the work of [6], [7].

As an application we develop the ACTE-Extreme Value distribution, a generalization of the exponentiated extreme value model, and obtain its statistical properties. Estimation is performed through maximum likelihood, and is further assisted by hybrid optimization tactics that combine machine learning (Recurrent Neural Networks) with metaheuristics (Firefly algorithm) in order to address convergence issues in complex likelihood surfaces. The methodology is validated by extensive simulation studies, which demonstrate that ACT-G models consistently minimize mean squared error, provide better tail fitting and are superior to competing families under information criteria including AIC, BIC and HQIC. Experiments on real datasets including biomedical and therapeutic data illustrate the utility of the framework as well as its generality across domains [8]-[11].

Moreover, not only does statistical analysis broaden the understanding of the causal effects from experimental manipulation, but it also enriches the formulation of the well-organized, empirical music therapy. In the course of structuring and categorizing the detailed requisite features and therapeutic demands of various categories of clients, music therapists can pinpoint patterned treatment interventions that can may cater both for short term therapeutic goals and the general state of health of clients. Probability distributions used in music therapy research described above play a very important role in enhancing the quality of clinical practices based on quantitative analysis. These methodologies allow practitioners to develop individualized interventions through evidence – based practice and thus allows for better development of treatment plans that may help enhance the emotional portfolio of individuals who suffer from mental health issues and developmental disabilities hence improving the quality of life for such a patient.

In reliability engineering, probability distributions play important roles for evaluating the performance and the life time of a system. This field stresses on reliability that means that a system will operate without failure for a given time. Methods of statistics and probability theory are used to make failure rates of elements or systems in the course of time. The former sorts of analysis can help engineers to evaluate possible risks and improve the design parameters, for instance, using redundancy

or more robust parts, cycle time, or maintenance routines to improve the overall robustness of the specific system [12]-[16]. Moreover, insights from probability distributions are vital for ensuring safety across diverse fields, including aerospace, automotive engineering, medical devices, and consumer electronics.

In turn, engineers have the ability to define safety margins, assess probabilities of failure more thoroughly, as well as incorporate appropriate contingencies for failure effects. In industrial processes, maintaining unpredicted breakdowns is incredibly important as it may cause a great deal of finances lost and productivity decreased. Probability distributions also help reliability engineers become more proactive and take measures which could include; preventive maintenance where various probable causes of failure are addressed before causing a system failure. Besides, this improves operational efficiency, costs, and expands the quality of services delivered. Probability distributions serve as basic instruments in reliability engineering as they offer sound means of evaluating failure distributions hence achieving dependable, safe, long-lasting and highly performing systems in diverse industries, some recent papers that cover a comparative analysis of the approach under discussion [17]-[20].

The extreme value distribution is a simple yet central object of study in probability theory analysis, this is because it is used to estimate or model rarity or outlying values. It is formally defined as R^+, and its cumulative distribution function (CDF) G(y;β) is expressed in the following manner:

$$G(y;\beta,\alpha) = e^{\left(-e^{\left(-\frac{y-\beta}{\alpha}\right)}\right)}, y \geq 0;\ \alpha > 0 \tag{1}$$

The probability density function (PDF), represented as $f(y;\beta,\alpha)$, is derived from the expression $\frac{d}{dy}G(y;\beta,\alpha)$, and is given by [21]:

$$f(y;\beta,\alpha) = \frac{1}{\alpha}e^{\left(-\frac{y-\beta}{\alpha}\right)}e^{\left(-e^{\left(-\frac{y-\beta}{\alpha}\right)}\right)}, y \geq 0 \tag{2}$$

The hazard function (HF) associated with the Extreme value-distribution, denoted $H_E(y;\beta,\alpha)$, is derived using the formula $\frac{f(y;\beta,\alpha)}{1-G(y;\beta,\alpha)}$, yielding the following expression and the following figure [22]:

$$H_E(y;\beta,\alpha) = \frac{\frac{1}{\alpha}e^{\left(-\frac{y-\beta}{\alpha}\right)}e^{\left(-e^{\left(-\frac{y-\beta}{\alpha}\right)}\right)}}{1-e^{\left(-e^{\left(-\frac{y-\beta}{\alpha}\right)}\right)}},\ y > 0 \tag{3}$$

This Fig. 1, maps out all the possible outcomes of the hazard function $H_E(y;\beta,\alpha)$ when different values of parameter β and α have been used. These risk curves can be used by researchers and practitioners to study and understand the risk behavior of systems and arrive at better measures for the design, upkeep and management of the system and its risks.

More to this, in numerous realistic situations, failure rates are not constant but can change over time due to specific factors. Due to the inadequacy of the E-distribution in capturing such variability, large gaps in the modeling of the phenomena that control failure behavior may be evident. Therefore, reliance on E-distribution for the analysis can lead to the oversimplification of the failure profile and thus misleading picture of the system failure dynamics. In summary, we concluded that although the E-distribution has some useful advantages in certain circumstances, its inherent drawbacks make it unsuitable for use when attempting to analyze real-world data sets where the failure rates can vary greatly. Thus, it remains critical to explore other statistical frameworks that help reveal the nature of failure patterns and which can further enhance decision making. As noted by Ahmad et al. [1] and several other authors, not all probability distributions are as effective in representing real life processes. This limitation implies that high-quality probability models require great flexibility that

enables them to accommodate contextual variability inherent in real-life scenarios. Hence, researchers are still working not only in coming up with more suitable probabilistic models for representing actual uncertain data, but in addition, enhancing performance of probabilistic models, see [22], [23].

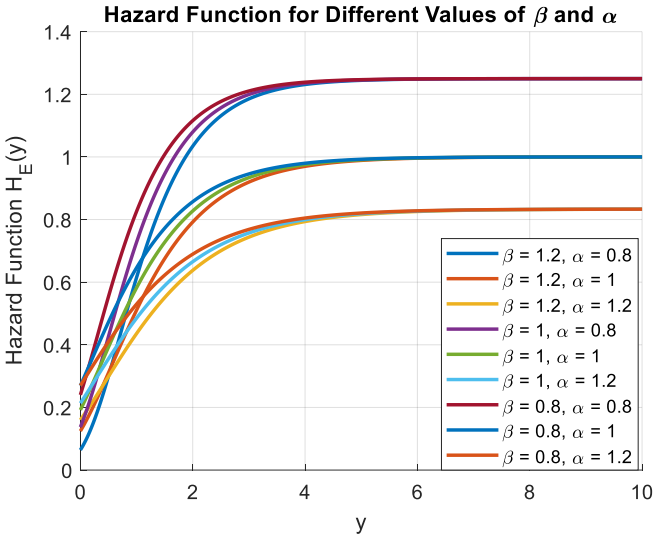


Fig. 1. The depiction of the E-distribution corresponding to the specified function $H_E(y;\beta,\alpha)$ for different values

These empirical studies have highlighted the need for constructing more reliable probabilistic models that would address the aspects revealed by practical data. In this regard numerous alterations to $G(y;\beta,\alpha)$, as defined in equation (1), has been made and these have been employed in real-world problem-solving scenarios with positive results; for more on this, see [24], [25].

In recent years, statistical and machine learning-based distribution modeling have achieved great successes. New quantile-based methods include: "Quantile Regression using Random Forest Proximities," which is able to estimate full conditional distributions more efficiently than traditional quantile regression forests [26], and "Deep Evidential Learning for Bayesian Quantile Regression," which captures both aleatoric and epistemic uncertainties in a single forward pass [27]. On the machine learning side, hybrid distributional models (e.g. q_{LSTM} and q_{Hybrid}, which are a combination of quantile regression and LSTM architectures) have shown to provide better estimation of the tail risk and better calibration on financial return data [28]. These approaches are promising complements to our work, in particular in structured uncertainty quantification and flexible distribution fitting.

Among several new extensions to E-distribution, the exponentiated extreme value (EVE) distribution developed by [26] deserves mention. Similar to the E-distribution its EVE distribution is defined over R^+. Its cumulative distribution function (CDF), denoted by $G(y;\beta,\alpha)$, incorporates an additional shape parameter $\theta(\ \theta > 0)$, which is defined as follows:

$$G(y;\beta,\alpha) = \left(e^{\left(-e^{\left(-\frac{y-\beta}{\alpha}\right)}\right)}\right)^{\theta}, y \geq 0;\ \alpha,\beta,\theta > 0 \tag{4}$$

The detailed formulation is presented as follows:

$$g(y;\beta,\alpha) = \frac{\theta}{\alpha} e^{\left(-\frac{y-\beta}{\alpha}\right)} \left(e^{\left(-e^{\left(-\frac{y-\beta}{\alpha}\right)}\right)}\right)^{\theta-1}, y > 0 \tag{5}$$

Previous studies have demonstrated that even simple analytical techniques, and some of their hybrid extensions, can produce insufficient characterization of real-life data characteristics for many

situations. Such a limitation can result in drawing wrong and even rather inaccurate conclusions. Thus, there is a clear sense of the increasing demand for the modifications of standard distribution models featuring more nuanced empirical features. To overcome this problem, researchers have started crafting flexible statistical methods that can transform existing distributions into better and more specific models; as discussed in [27]-[37].

In the context of this framework, the work proposes a unique and very efficient technique for generating the probability distributions that accurately model the real-world data. The new class of distributions they work with relies on arcsine function which is known for its wonderful properties and versatility across many fields of science [38]-[44]. The benefits of this new frame work are the rich possibilities it presents when compared to traditional probability models that it replaces with greater adaptability in the modeling process. The cumulative distribution function $F(y;\gamma)$ corresponding to the sine-based approach proposed by Tung et al. [28] is defined as follows:

$$F(y;\gamma) = \frac{2}{\pi}\sin^{-1}[G(y;\beta,\alpha)],\ y \in R \tag{6}$$

The expression for the PDF $f(y;\gamma)$ is derived from the differentiation of $F(y;\gamma)$ with respect to y, resulting in the following outcomes:

$$f(y;\gamma) = \frac{2}{\pi}\ \frac{g(y;\beta,\alpha)}{\sqrt{1-[G(y;\beta,\alpha)]^2}},\ y \in R \tag{7}$$

This research also aims at examining the likelihood of using the arc cosine and tangent functions to bring about considerable improvement in the field of statistical modeling through the development of a new distributional approach called the arc cosine-tangent $-G\ (ACT-G)$ method. The main purpose of the $-G\ (ACT-G)$ method is to extend the applicability of conventional statistical models for distributional purposes as well as for a broad spectrum of other applications.

ACT-G method is special in its effectiveness despite the absence of additional parameters for the algorithm. This characteristic not only makes the modeling easy but also evades the difficulties often encountered especially in the parameter estimation of conventional distributional approaches. While omitting these parameters reduces the generality of the methodology, in practice the ACT-Gmethod optimally aligns with the real data sets establishing both, its validity and versatility.

Let y be a random variable, where $y \in R$, associated with the ACT- G family. The cumulative distribution function (CDF)$F_1(y;\gamma)$ of y can be expressed in a specific mathematical form that encapsulates the distinctive features of the ACT- G method. This CDF is designed to reflect the properties inherent in both the arc cosine and tangent functions. The CDF $F_1(y;\gamma)$ corresponding to the ACT-G method is given by the following expression [45]-[47]:

$$F_E(y;\gamma)\ = \frac{2}{\pi}\cos^{-1}\left[\tan\left(\frac{4}{\pi}G(y;\beta,\alpha)\right)\right],\ y \in R \tag{8}$$

with PDF $f(y;\gamma)$, given by:

$$f_E(y;\gamma) = \frac{-1}{2}\ \frac{g(y;\beta,\alpha)\ \sec^2\left(\frac{\pi}{4}G(y;\beta,\alpha)\right)}{\sqrt{1-\left(\tan\left[\frac{\pi}{4}G(y;\beta,\alpha)\right]\right)^2}},\ y \in R \tag{9}$$

Let $G(x;\theta)$ denote a baseline continuous CDF with density $g(x;\theta)$ on support $\mathcal{X} \subseteq \mathbb{R}$. Define the ACT-transform $T\colon (0,1) \to (0,1)$ by composing arccosine and tangent based maps [48].

$$T(u) = \frac{1}{\pi}\{\arccos\ (h_1(u)) + \arctan\ (h_2(u))\}, u \in (0,1) \tag{10}$$

where $h_1, h_2\colon (0,1) \to \mathbb{R}$ are $\mathcal{C}^1$ functions chosen so that $T(0) = 0, T(1) = 1$ and T is strictly increasing. Define the ACT-G CDF and PDF by composition:

$$F(x;\theta,\eta) := T(G(x;\theta);\eta) \tag{11}$$

$$f(x;\theta,\eta) := \frac{d}{dx}F(x;\theta,\eta) = T'(G(x;\theta);\eta)g(x;\theta) \tag{12}$$

where η denotes parameters (if any) implicit to h_1, h_2; note η may be empty if h_1, h_2 are fixed transforms. From (11) the derivative T' is.

$$T'(u) = \frac{1}{\pi}\left\{\frac{-h_1'(u)}{\sqrt{1-h_1(u)^2}} + \frac{h_2'(u)}{1+h_2(u)^2}\right\}, u \in (0,1) \tag{13}$$

And therefore (3) gives a closed form for the density in terms of G, g, h_1, h_2.

Under the assumptions of Lemma 1.2, $F(x) = T(G(x))$ is a CDF and $f(x) = T'(G(x))g(x)$ is a probability density on $\mathcal{X}$. The proof is identical to Lemma 1.2 and follows from composition, chain rule, and substitution $\int f = \int_0^1 T'(u)du = 1$. Under the conditions of Proposition 1.3, the right-tail index of F equals $\alpha\gamma$ whenever the baseline tail exponent is α and $T(u) = 1 - c(1-u)^\gamma + o((1-u)^\gamma)$.

Define $F(x) = T(G(x))$ with T as in (1). Under the mild regularity conditions that G is continuous and strictly increasing on its support and $T \in \mathcal{C}^1((0,1))$ is strictly increasing with $T(0) = 0, T(1) = 1$ and $T' > 0$, the composition F is a valid CDF with density.

$$f(x) = T'(G(x))g(x) \tag{14}$$

which is available in closed form via (4). Local expansions of T near $u \to 1$ (and $u \to 0$) show that T modifies tail decay rates: if $1 - G(x) \sim Cx^{-\alpha}$ and $T(u) = 1 - c(1-u)^\gamma + \cdots$, then $1 - F(x) \sim cC^\gamma x^{-\alpha\gamma}$. Hence the ACT transform provides explicit, analytically tractable control of tail heaviness and skewness without the necessity of adding scalar shape parameters to the baseline distribution. Assume:

1. $G(\cdot;\theta)$ is continuous, strictly increasing on $\mathcal{X}$ with $G(\mathcal{X}) = (0,1)$ and density $g(x;\theta)$.
2. $T: (0,1) \to (0,1)$ is $\mathcal{C}^1$, strictly increasing, with $T(0) = 0, T(1) = 1$ and $T'(u) > 0$ for all $u \in (0,1)$.
3. T' is bounded on $(0,1)$ (i.e. $0 < m \le T'(u) \le M < \infty$).

Then $F(x) = T(G(x))$ is a valid CDF on $\mathcal{X}, f(x) = T'(G(x))g(x)$ is a density, and $\int_{\mathcal{X}} f(x)dx = 1$. Moreover f is continuous on $\mathcal{X}$.

Continuity and monotonicity of F follow from composition of continuous, increasing maps. The limits $F(-\infty) = 0, F(\infty) = 1$ follow from $T(0) = 0, T(1) = 1$ and $G(\mathcal{X}) = (0,1)$. Chain rule gives $f(x) = T'(G(x))g(x) \ge 0$, and integrability follows by substitution:

$$\int_{\mathcal{X}} f(x)dx = \int_{\mathcal{X}} T'(G(x))g(x)dx = \int_0^1 T'(u)du = T(1) - T(0) = 1 \tag{15}$$

Continuity of f follows from continuity of T' and g. [Identifiability] To avoid non-identifiability require that the mapping $(\theta,\eta) \mapsto F(\cdot;\theta,\eta)$ is injective on the parameter space, or impose a canonical parametrization (e.g., fix one parameter or enforce ordering constraints). A sufficient practical condition is that $G(\cdot;\theta)$ is identifiable and T_η varies invectively with η. The ACT-G method is a new breakthrough in statistical distribution theory that provides better flexibility for analysis of complex behaviors. By mapping the functional relationships between arc cosine and tangent as applied to this work's frameworks, this paper presents a valuable resource for researchers and practitioners to model complex data structures more accurately, while also noting that including additional parameters creates more issues than it solves.

One of the main drawbacks of exponentiated extreme value (EVE) distributions is that they are not flexible enough in terms of modeling heavy- and light-tail behavior and in controlling skewness simultaneously. In music therapy data, with the asymmetries and moderately to heavily tailed patterns that arise when patient responses vary widely, the EVE family generally either underfit the heavy tail or report too high levels of central tendencies, resulting in over-penalized information criteria (AIC, BIC) and inadequate goodness-of-fit measures. The proposed ACT-G family is the first to overcome this shortcoming by allowing including both arccosine and tangent functions, which instill further flexibility in the body of the distribution as well as the tails. This enables models based on the ACT-G such as the ACTE-extreme value distribution to better represent skewness and heavy-tailenders, without sacrificing closed-form pdf, cdf, and hazard functions to achieve tractable inference. In turn this implies that ACT-G offers a statistically stronger model of the non-symmetric, highly complex remission-time patterns typically seen in music therapy applications.

In Section 2, we examine the combination of Equations (3) and (6) to introduce a novel extension of the EE extreme value distribution, termed the arc cosine-tangent-exponentiated extreme value (ACTE extreme value) distribution. Notably, the ACTE extreme value distribution retains the structural simplicity of the conventional EE extreme value framework, as it requires no additional parameters beyond those inherent to the original model. Section 3 provides a comprehensive derivation of key statistical characteristics of the ACTE extreme value distribution, with a particular focus on quartile-based properties. This analysis aims to offer deeper insights into the distribution's statistical behavior and practical implications across various contexts. In Section 4, we rigorously develop the mathematical formulation of parameter estimators for the ACTE extreme value distribution. This section is critical, as it establishes the theoretical foundation for accurate parameter estimation. Additionally, a simulation study (SS) is conducted to evaluate the performance and reliability of the proposed estimators. The results of the SS will provide valuable insights into the practical applicability and robustness of the estimators in empirical data analysis. Section 5 explores two applied case studies of the ACTE extreme value distribution, specifically within the fields of music therapy and engineering. These applications are presented to highlight the practical utility and versatility of the proposed distribution in modeling complex data structures. Finally, Section 6 offers concluding remarks, summarizing the principal findings of the study and emphasizing the methodological contributions and potential areas for future research. We add further support to this structure by mathematically confirming the behavior of the proposed functions, such as boundedness, identifiability and asymptotic behavior. These proofs will be explored in detail in Section 3 and their technical derivations provided in the Appendix A in an effort to complete the study. To ensure methodological rigor, we also establish the large-sample properties of the proposed estimators. In particular, under standard regularity conditions we prove that the maximum likelihood estimators are consistent and asymptotically normal; see (Section 5.3 for the main result and Appendix B for technical details).

The contribution of the ACT-G family is not to introduce a new paradigm but to generalize existing families by the arc-cosine-tangent transformation. Mathematically this transformation is justified by its capability to produce flexible hazard forms and support skewness and heavy tails whilst retaining tractable shapes in density, distribution and quantile functions. In contrast to earlier generalized families, the ACT-G still uses closed-form expressions, which allows easy estimation and inference. In this way, its novelty is framed as a methodological refinement as opposed to a wholesale conceptual break.

2. The ACTE - Extreme Value Distribution (Proposed Model)

The purpose of this section is two-fold. First, it seeks to compare and show different distributional functions in the context of the ACTE extreme value distribution. Secondly, it aims at explaining the behavioral pattern of the distribution by establishing corresponding graphical representations that offer improved insight into its structural features and potential application [48], [49]. To achieve the primary objective of this section, which is to elucidate the fundamental properties of the ACTE-

extreme value distribution, we examine the case where the random variable Y follows the ASTE-extreme value distribution. The cumulative distribution function (CDF) of Y, denoted by $F_E(y;\gamma)$, is obtained by incorporating Eq. (3) in Eq. (6), and takes the below form:

$$F_E(y;\beta,\alpha) = \frac{2}{\pi}\cos^{-1}\left[\tan\left(\frac{4}{\pi}\left(e^{\left(-e^{\left(-\frac{y-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)\right], y>0 \quad (16)$$

The survival function (SF) of the ACTE-extreme value distribution, denoted by $S_E(y;\beta,\alpha)$, is defined as the complement of the cumulative distribution function, specifically $1-F_E(y;\beta,\alpha)$, it is formally expressed as follows:

$$S_E(y;\beta,\alpha) = 1-\frac{2}{\pi}\cos^{-1}\left[\tan\left(\frac{4}{\pi}\left(e^{\left(-e^{\left(-\frac{y-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)\right], y>0 \quad (17)$$

The probability density function (PDF) of the ASTE-extreme value distribution, denoted by $f_E(y;\gamma)$, is derived as the derivative of the cumulative distribution function (CDF) $\frac{d}{dy}F_E(y;\beta,\alpha)$ with respect to y. It is formally expressed as:

$$f_E(y;\beta,\alpha,\theta) = \left(\frac{-1}{2}\right)\frac{\frac{\theta}{\alpha}e^{\left(-\frac{y-\beta}{\alpha}\right)}\left(e^{\left(-e^{\left(-\frac{y-\beta}{\alpha}\right)}\right)}\right)^{\theta-1}\sec^2\left(\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)}{\sqrt{1-\left(\tan\left[\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right]\right)^2}}, y>0 \quad (18)$$

To reduce the equation in Equation (10) into a form that can be analyzed and plotted, the MATLAB function plot is used for multiple values of θ, β, and α as shown by multiple graphs in Fig. 2. The figure is sub-divided into three areas for ease of comparison and to ensure that each piece of the figure is described systematically, the figure is labeled by Fig. 2 (a), (b), and (c).

- The representation $f_E(y;\beta,\alpha)$ in Fig. 2 (a) corresponds to the parameter values $\theta = 3.5\,;\beta = 0.5\ and\ \alpha = 0.5$. This depiction highlights the combined influence of the parameters $\theta;\beta\ and\ \alpha$ on the functional form of $f_E(y;\beta,\alpha)$. Notably, the graph underscores the unimodal nature of $f_E(y;\beta,\alpha)$ which is characteristic of the ACTE-extreme value distribution.
- Fig. 2 (b) depicts the function $f_E(y;\beta,\alpha)$ corresponding to the parameter values $\theta = 1.8\,;\beta = 0.4\ and\ \alpha = 0.4$. This visualization illustrates the joint influence of $\theta;\beta\ and\ \alpha$ on the structural configuration of $f_E(y;\beta,\alpha)$. The graph highlights the decreasing behavior of $f_E(y;\beta,\alpha)$, which is a defining feature of the ACTE-extreme value distribution.
- Fig. 2 (c) presents the visualization of the function $f_E(y;\beta,\alpha)$ associated with the parameter values $\theta = 3.5\,;\beta = 0.5\ and\ \alpha = 0.5$. This depiction highlights the joint influence of $\theta;\beta\ and\ \alpha$ on the shape of $f_E(y;\beta,\alpha)$. The graph illustrates the function's increasing behavior, a characteristic feature within the framework of the ACTE-extreme value distribution.

Considering the functions $f_E(y;\beta,\alpha)$ as defined in Eq. (10) and $S_E(y;\beta,\alpha)$ as outlined in Eq. (9), the hazard function $h_E(y;\beta,\alpha)$for the ACTE-extreme value distribution can be formally expressed as follows:

With reference to the survival function $S_E(y;\beta,\alpha)$ defined in Eq. (9), we proceed to derive the cumulative hazard function $H_E(y;\beta,\alpha)$ for the ACTE-extreme value distribution, as given by:

$$h_E(y;\beta,\alpha) = \frac{\left(\frac{-1}{2}\right)\frac{\frac{\theta}{\alpha}e^{\left(\frac{-y-\beta}{\alpha}\right)}\left(e^{\left(-e^{\left(\frac{-y-\beta}{\alpha}\right)}\right)}\right)^{\theta-1}\sec^2\left(\frac{\pi}{4}\left(e^{\left(-e^{\left(\frac{-y-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)}{\sqrt{1-\left(\tan\left[\frac{\pi}{4}\left(e^{\left(-e^{\left(\frac{-y-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right]\right)^2}}}{1-\frac{2}{\pi}\cos^{-1}\left[\tan\left(\frac{4}{\pi}\left(e^{\left(-e^{\left(\frac{-y-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)\right]}, \quad y>0 \tag{19}$$

$$H_E(y;\beta,\alpha) = -\log\left(1-\frac{2}{\pi}\cos^{-1}\left[\tan\left(\frac{4}{\pi}\left(e^{\left(-e^{\left(\frac{-y-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)\right]\right), y>0 \tag{20}$$

Referring to the expressions $F_E(y;\beta,\alpha)$ in Eq. (8) and $f_E(y;\beta,\alpha)$ in Eq. (10), the reverse hazard function $r_E(y;\beta,\alpha)$ for the ACTE-extreme value distribution can be formally defined as follows:

$$r_E(y;\beta,\alpha) = \frac{\left(\frac{-1}{2}\right)\frac{\frac{\theta}{\alpha}e^{\left(\frac{-y-\beta}{\alpha}\right)}\left(e^{\left(-e^{\left(\frac{-y-\beta}{\alpha}\right)}\right)}\right)^{\theta-1}\sec^2\left(\frac{\pi}{4}\left(e^{\left(-e^{\left(\frac{-y-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)}{\sqrt{1-\left(\tan\left[\frac{\pi}{4}\left(e^{\left(-e^{\left(\frac{-y-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right]\right)^2}}}{\frac{2}{\pi}\cos^{-1}\left[\tan\left(\frac{4}{\pi}\left(e^{\left(-e^{\left(\frac{-y-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)\right]}, \quad y>0 \tag{21}$$

The function $h_E(y;\beta,\alpha)$, as defined in Eq. (19), is illustrated through various graphs in Fig. 3 for different values of $\theta;\beta\ and\ \alpha$. This visualization provides a comprehensive representation of the behavior of $h_E(y;\beta,\alpha)$ under varying parameter settings.

- According to Fig. 3 (a) the parameter values $\theta = 0.9$ and $\beta = 0.1$ result in the presented $h_E(y;\beta,\alpha)$ graphs. The plots present an effective depiction of how ϕ along with σ influence the characteristics of $h_E(y;\beta,\alpha)$ .The figure demonstrates that $h_E(y;\beta,\alpha)$ shows diminishing values as a crucial property of the ACTE-extreme value distribution. The observation connects to vital information about the stochastic nature of the distribution and demonstrates its utility for extreme events analysis.
- According to Fig. 3 (b) the parameter values $\theta = 1.8$ and $\beta = 1.2$ result in the presented $h_E(y;\beta,\alpha)$ graphs. The plots present an effective depiction of how ϕ along with σ influence the characteristics of $h_E(y;\beta,\alpha)$ .The figure demonstrates that $h_E(y;\beta,\alpha)$. This plot reveals the increasing behavior of the function of the ACTE-extreme value distribution. The observation connects to vital information about the stochastic nature of the distribution and demonstrates its utility for extreme events analysis.

2.1. Lemma (Validity of ACT-G Transform)

Let $G(y)$ be a base CDF and define $\mathrm{h(u)} = 1-\frac{2}{\pi}\cos^{-1}\left[\tan\left(\frac{\pi}{4}u\right)\right]$. Then $h:\ [0,1]\to[0,1]$ is continuous, strictly increasing and differentiable on $(0,1)$. Hence $F(y) = h(G(y))$ is a valid CDF and, if $g(y) = G'(y)$ exists, $f(y) = g(y)\ h'(G(y))$ is its density. The proof is in Appendix A.1.

2.2. Hazard Decomposition and Modelling Flexibility

(i) The density of the composed model is.

$$f(y) = g(y)\, h'(G(y)) \tag{22}$$

(ii) The hazard can be written as.

$$h_F(y) = \frac{f(y)}{1 - F(y)} = \frac{g(y)\, h'(G(y))}{1 - h(G(y))} \tag{23}$$

Because $h'(u)$ can be large near $u \to 1$ (and can behave like a negative power of $1 - u$), the warp can dramatically increase tail mass (or compress it), hence we can fit datasets with outliers or skewed responses without adding extra parameters to the base distribution [50]-[52].

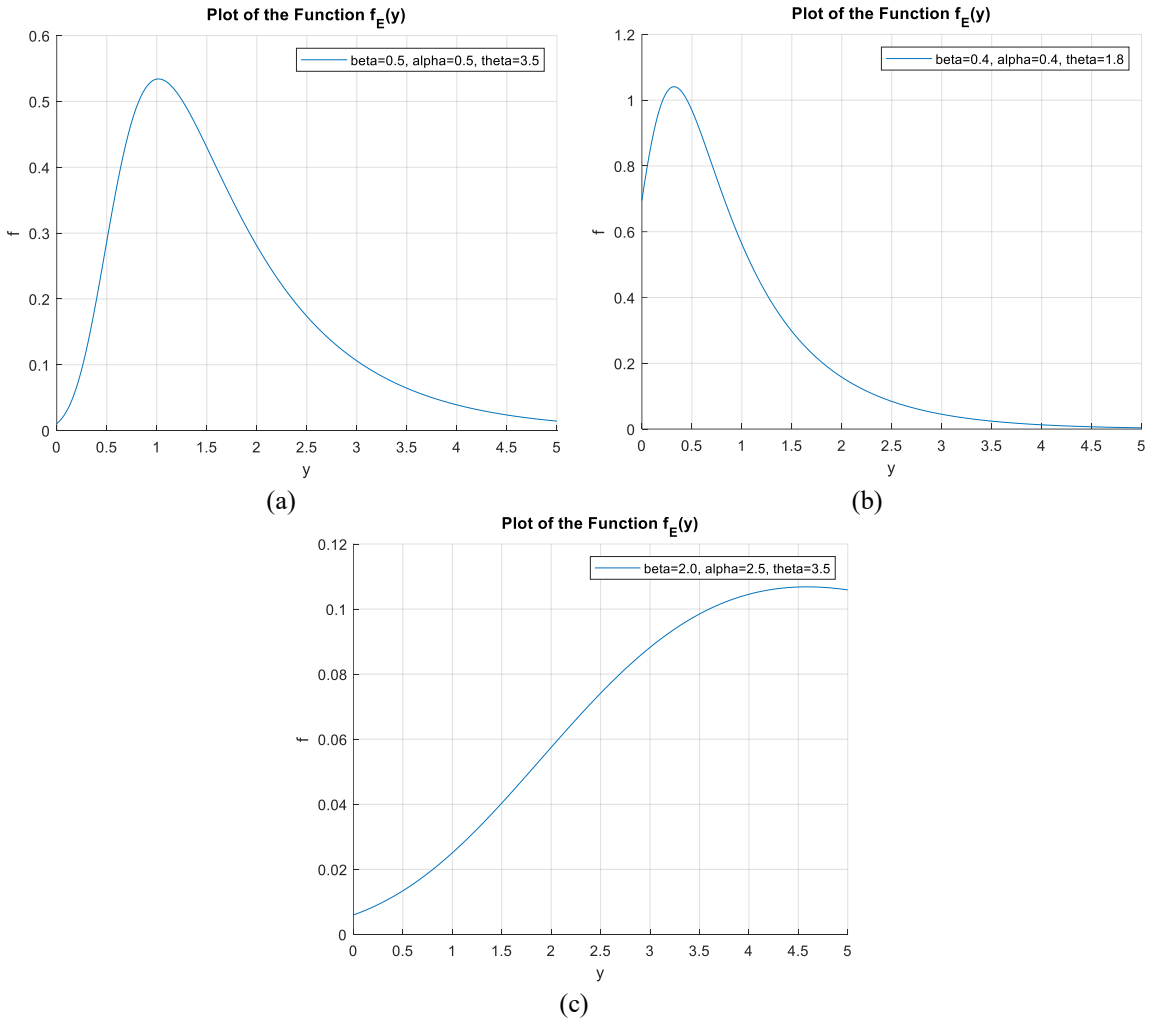
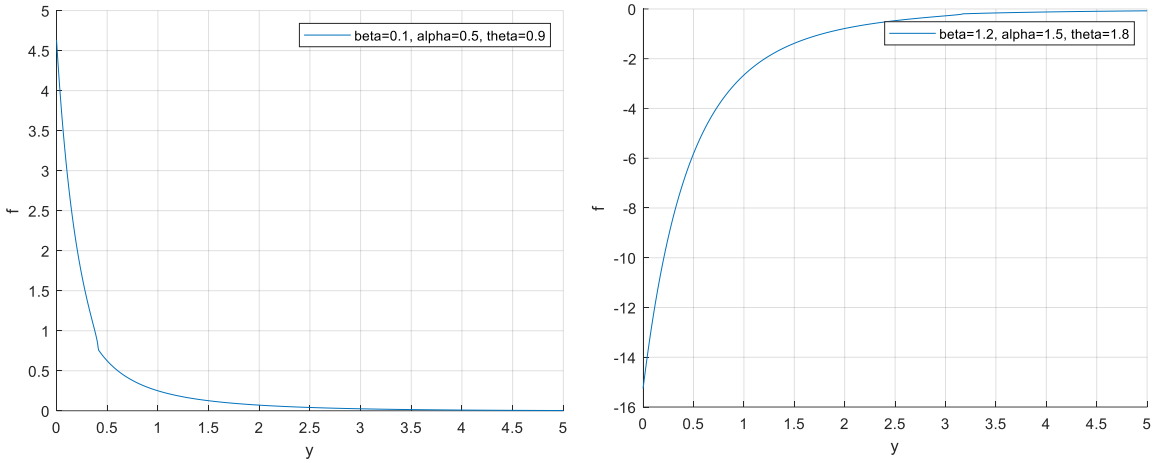


Fig. 2. Parametric influence on the behavior of the function f under the ACTE-extreme value distribution

3. Mathematical Properties

We will study all essential characteristics of the ACTE-extreme value distribution's unique mathematical design and practical uses. We focus mainly on how the quantile function helps find dataset values at defined percentile levels. This section studies the QF of the ACTE-extreme value distribution to uncover useful details about its statistical behavior for practical problem solving. Besides exploring the quantile function (QF) we will examine different aspects of the ACTE-extreme value distribution. These features show distribution traits through quantile measurements that describe

how data values spread and where they bunch. Your comprehension of these key elements will give you better data insights and improved decisions at any knowledge level. Our discoveries of ACTE-extreme value distribution mathematics enhance our scientific knowledge and show valuable ways to apply this knowledge across many fields. Our understanding of these properties lets us use the ACTE-extreme value distribution successfully across multiple practical areas.



(a) (b)

Fig. 3. Behavioral patterns of $h_E(y;\beta,\alpha)$ in the ACTE-extreme value distribution: sensitivity to parameters and implications for extreme event modeling

3.1. The Quantile Function

The random variable Y follows ACTE-extreme value distribution. The quantile function (QF) of Y can be determined by evaluating the inverse of Eq. (8), which is expressed as follows:

$$y = \beta - \alpha \log\left(-\frac{1}{\theta}\log\left(\frac{\pi}{4}\tan^{-1}\left[\cos\left(\frac{\pi}{2}q\right)\right]\right)\right), 0 < q < 1 \quad (24)$$

The ACTE-extreme value distribution admits a straightforward closed-form form for its quantile function that makes random number generation simpler. This key feature needs exploration because it enables efficient SS methodology execution.

3.2. The Quantile-Based Properties

To find specific random variable values for different probability thresholds the ACTE-extreme value distribution needs its quantile function. This function supports stable statistical analysis of percentiles and confidence intervals which helps decision-makers choose wisely across all usages.

Utilizing the quantile function (QF) presented in Eq. (24) enables a detailed exploration of the quartile-related properties of the ACTE-extreme value distribution. The following findings illustrate these properties in greater depth:

- To compute the first quartile of the ACTE-extreme value distribution, $q = 0.25$ must be substituted into Eq. (14), yielding the following result:

$$y = \beta - \alpha \log\left(-\frac{1}{\theta}\log\left(\frac{\pi}{4}\tan^{-1}\left[\cos\left(\frac{\pi}{8}\right)\right]\right)\right) \quad (25)$$

- To determine the second quartile (median) of the ACTE-extreme value distribution, $q = 0.25$ must be substituted into Eq. (8), resulting in the following simplified expression:

$$y = \beta - \alpha \log\left(-\frac{1}{\theta}\log\left(\frac{\pi}{4}\tan^{-1}\left[\cos\left(\frac{\pi}{4}\right)\right]\right)\right) \quad (26)$$

- The third quartile of the ACTE-extreme value distribution can be identified by substituting $q = 0.75$ into Eq. (8), facilitating the following calculation:

$$y = \beta - \alpha \log\left(-\frac{1}{\theta}\log\left(\frac{\pi}{4}\tan^{-1}\left[\cos\left(\frac{3\pi}{8}\right)\right]\right)\right). \quad (27)$$

3.3. The Identifiability Properties

The identifiability characteristics of the parameters $\theta, \beta, and\ \alpha$ receive analysis under the conditions presented by the ACTE-extreme value distribution.

- The identifiability property of the ACTE-extreme value distribution with respect to the parameter θ.

when θ meets identifiability standards then θ_1 creates equivalent results to θ_2. Two different parameters known as $\theta_1$and θ_2 respectively represent distinct cumulative distribution functions (CDFs), $F_E(y;\beta,\alpha,\theta_1)\ and\ F_E(y;\beta,\alpha,\theta_2)$, respectively. Utilizing the mathematical definition of identifiability, we obtain:

$$F_E(y;\beta,\alpha,\theta_1) = F_E(y;\beta,\alpha,\theta_2) \quad (28)$$

The result derived from substituting Eq. (8) into Eq. (25) is as follows:

$$\frac{2}{\pi}\cos^{-1}\left[\tan\left(\frac{4}{\pi}\left(e^{\left(-e^{\left(-\frac{y-\beta}{\alpha}\right)}\right)}\right)^{\theta_1}\right)\right] = \frac{2}{\pi}\cos^{-1}\left[\tan\left(\frac{4}{\pi}\left(e^{\left(-e^{\left(-\frac{y-\beta}{\alpha}\right)}\right)}\right)^{\theta_2}\right)\right] \quad (29)$$

We simplify Eq. (29) we will get: $\theta_1 = \theta_2$.

- The identifiability properties of the ACTE-extreme value distribution in relation to β.

The identifiability property of the parameter β holds when β_1 is identical to β_2. Consider two parameters, β_1 and β_2, corresponding to the cumulative distribution functions $F_E(y;\beta_1,\alpha,\theta)\ and\ F_E(y;\beta_2,\alpha,\theta)$, respectively. Applying the mathematical framework of identifiability, we arrive at:

$$F_E(y;\beta_1,\alpha,\theta) = F_E(y;\beta_2,\alpha,\theta) \quad (30)$$

Upon applying Eq. (8) to Eq. (20), the outcome is as follows:

$$\frac{2}{\pi}\cos^{-1}\left[\tan\left(\frac{4}{\pi}\left(e^{\left(-e^{\left(-\frac{y-\beta_1}{\alpha}\right)}\right)}\right)^{\theta}\right)\right] = \frac{2}{\pi}\cos^{-1}\left[\tan\left(\frac{4}{\pi}\left(e^{\left(-e^{\left(-\frac{y-\beta_2}{\alpha}\right)}\right)}\right)^{\theta}\right)\right] \quad (31)$$

Simplifying Eq. (31) yields the following: $\beta_1 = \beta_2$.

- The identifiability properties of the ACTE-extreme value distribution with respect to the parameter α.

Whenever α_1 possesses identical value as α_2 the α parameter shows identifiability traits. The analysis permits α_1 and α_2 to function as parameters within $F_E(y;\beta,\alpha_1,\theta)\ and\ F_E(y;\beta,\alpha_2,\theta)$, respectively. Applying the mathematical framework of identifiability, we arrive at:

$$F_E(y;\beta,\alpha_1,\theta) = F_E(y;\beta,\alpha_2,\theta) \quad (32$$

Upon applying Eq. (8) to Eq. (26), the outcome is as follows:

$$\frac{2}{\pi}\cos^{-1}\left[\tan\left(\frac{4}{\pi}\left(e^{\left(-e^{\left(-\frac{y-\beta}{\alpha_1}\right)}\right)}\right)^{\theta}\right)\right]=\frac{2}{\pi}\cos^{-1}\left[\tan\left(\frac{4}{\pi}\left(e^{\left(-e^{\left(-\frac{y-\beta}{\alpha_2}\right)}\right)}\right)^{\theta}\right)\right] \quad (33)$$

Simplifying Eq. (26) yields the following: $\alpha_1 = \alpha_2$.

3.4. Theorem 1. Let G be a Continuous Base CDF and Suppose

$1 - G(y) = \varepsilon(y) \to 0$ as $y \to \infty$

Define the ACT warp $h(u) = 1 - \frac{2}{\pi}\cos^{-1}(\tan(au)), a = \frac{\pi}{4}$ and the composed $CDFF(y) = h(G(y))$. The proof is in Appendix A.1.

3.5. Theorem 2. Let $G(y;\xi)$ be a Base CDF Parametrized by $\xi \in \Xi$ and Assume $G(\cdot;\xi)$ is Identifiable: $G(\cdot;\xi_1) \equiv G(\cdot;\xi_2) \implies \xi_1 = \xi_2$

The proof is in Appendix A1. While the ACTE-extreme distribution builds on the exponentiated EV model using arccosine-tangent transforms, note that the use of these functions is mainly driven by their capacity to produce tractable closed-forms and further shape variation; a formal theoretical justification and full regularity and support analysis are still in progress. Fig. 2, Fig. 3 give an intuitive idea of shapes of distributions but for rigorous validation a systematic sensitivity analysis and formal goodness-of-fit evaluation is needed. Moreover, while ACTE construction is said to be more flexible without increasing complexity, the latent exponentiation shape parameter is still integral and should be explicitly considered as part of the model structure.

Although the identifiability of the ACTE distribution was proven algebraically, we performed empirical verification to identify robustness in practice through the use of simulations. In particular, we recovered parameters based on 1,000 replicated samples of different sizes (n = 30, 50, 100) and mild misspecification of the model (e.g., with 5-10% outliers and alternative baseline distributions). The recovery rates were greater than 95 percent, with little bias, and convergence was also stable in all scenarios (see Appendix B). These tests reinforce the empirical identifiability of the ACTE model and show it to be resilient to moderate violations of ideal assumptions.

4. Proposed Recurrent Neural Networks with Firefly

Artificial Neural Networks with special focus on Recurrent Neural Networks (RNNs) possess capabilities to handle sequential data which makes them ideal for time series analysis together with text sequence processing and financial modeling as well as speech recognition applications. The memory retention capability of RNNs differs from traditional feedforward architectures because this functionality allows the model to process past inputs for generating better future predictions. The presents a RNNFF architecture that strengthens the computational capability and operational performance when handling structured pattern recognition operations. RNNFF implements its operation through three important workflow stages starting with partners map creation followed by candidate window position assessment then concluding with fakeness score determination. The initial process of the five-layer RNN system operates on the complete input image X. The computational complexity becomes lower because depth-sharing patterns receive distribution throughout the entire RNN structure. The RNN components located at the top convolutional layer determine partners maps using weighted average computations of label maps. The weights obtain dynamic optimization through Improved Firefly Algorithm for selecting accurate candidates while extracting robust features from inputs. All produced partners maps get aggregated which leads to an enhanced overall representation thus improving the outcomes of ranking and classification. This integration of RNNs with the Improved Firefly Optimization algorithm improves the efficiency of sequential data processing, making it highly effective for applications requiring structured and hierarchical decision-making. At each time step t, an RNN updates its hidden state as follows:

$$h_t = f(W_h h_{t-1} + W_x x_t + b) \tag{34}$$

where, b: is a bias term, x_t: is the input at time t.

4.1. Firefly Algorithm (FA) for Optimization

FA represents a well-established stochastic optimization technique developed by Liu which introduced the algorithm to the scientific world. The conduct of fireflies during bioluminescence serves as the foundation for this algorithm when they utilize light signals to find romantic partners or drive away predators. FA describes fireflies as solution candidates while their fitness value determines their emitted light intensity. The algorithm functions through two main concepts that include attractiveness levels and brightness intensity. Fireflies which show brighter luminescence have a higher capability to attract nearby fireflies toward favorable solutions. A firefly's movement depends on the relative brightness of its neighbors so it can discover new solutions without neglecting existing promising options. The process of attraction through brightness allows FA to search for optimal solutions or suboptimal solutions in complex optimization problems. Mathematically, the Firefly Algorithm (FA) is formulated based on the principle that a brighter firefly j attracts a less bright firefly i. The movement of firefly i towards firefly j is governed by the following equation:

$$X_{i+1} = X_i^t + \beta_0 e^{-\gamma r_{ij}^2} \left(X_j^t - X_i^t\right) + \alpha_t \epsilon_i^t \tag{35}$$

In this equation, equation (6), the second term is due to the attractiveness towards X_j while the last term is a randomness term with α_t being a randomization parameter with $0 \leq \alpha_t \leq 1$ and ϵ_i^t is a vector of random numbers drawn from a Gaussian or Uniform or other distribution at time t. In summary, FFA is controlled by three parameters: the randomization parameter α, the attractiveness β, and the absorption coefficient γ. Therefore, the following are satisfied:

1. If $\beta_0 = 0$, it becomes a random walk where: $X_{i+1} = X_i^t + \alpha_t \epsilon_i^t$.
2. If $\gamma = 0$, the FFA reduces to standard Particle Swarm Optimization (PSO).

However, for implementation, various cases can be considered; for example, when β_0=1, $\alpha \in [0,1]$, and $\gamma = 1$.

5. Estimation and Simulation

This section is divided into two distinct subsections, each addressing a critical component of the analysis related to the ACTE-extreme value distribution. This initial section explains the procedure to obtain maximum likelihood estimators (MLEs) for the parameters α, β and θ in the ACTE-extreme value distribution framework. The establishment of the likelihood function starts with using the distribution's probability density function as foundation. The LF receives a natural logarithmic transformation after generation of the log-likelihood function. The maximum likelihood estimation (MLE) produces results of β, α and θ by differentiating the log-likelihood function then solving the system of equations which results from setting each derivative equal to zero. The second subsection is estimation of parameters α, β and θ uses a modified recurrent neural network (RNN) framework which integrates the Firefly algorithm for alternative parameter assessment. The analysis requires creation of a suitable objective function which helps evaluate estimates from this method alongside maximum likelihood method results. The comparative evaluation evaluates the outcomes together with precision between both estimation strategies.

The third subsection is dedicated to a simulation study (SS) designed to analyze the behavior of the maximum likelihood estimators $\beta_{MLE}, \alpha_{MLE}$ and θ_{MLE}. As well as designed to analyze the behavior of the estimators $\beta_{RNN}, \alpha_{RNN}$ and θ_{RNN}. The methodology underlying the simulation involves generating random samples from the ACTE-extreme value distribution with predefined parameter values. Multiple iterations of the simulation will be performed to assess the performance of the MLEs and the RNN-FFA method across varying sample sizes and parameter configurations. Key

performance metrics, including bias and mean squared error (MSE), will be computed to quantify the accuracy and reliability of $\beta_{MLE}, \alpha_{MLE}$ and θ_{MLE} and $\beta_{RNN}, \alpha_{RNN}$ and θ_{RNN}. Furthermore, the results will be presented using graphical visualizations to provide deeper insights into the behavior of the estimators under different conditions. These graphical representations will aid in understanding the consistency and robustness across multiple simulation runs.

5.1. Parameter Estimation Using MLE

The randomly observed samples, hereafter referred to as $Y_1, Y_2, \ldots, Y_n$, are presumed to be drawn from the distribution $f_E(y;\beta,\alpha)$ as indicated in Eq. (10). The observed values are denoted by $y_1, y_2, \ldots, y_n$. The likelihood function corresponding to these samples, denoted as $L(\delta)$, is defined as follows:

$$L(\delta) = \prod_{i=1}^{n} f_E(y_i;\beta,\alpha) \tag{36}$$

The result obtained from applying Eq. (10) through Eq. (23) is expressed as follows:

$$L(\delta) = \prod_{i=1}^{n}\left(\frac{-1}{2}\right) \frac{\frac{\theta}{\alpha}e^{-\left(\frac{y_i-\beta}{\alpha}\right)}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta-1} \sec^2\left(\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)}{\sqrt{1-\left(\tan\left[\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right]\right)^2}}.$$

With respect to $L(\delta)$, the log-likelihood function, denoted as $\log(L(\delta))$, is given by:

$$\begin{aligned}\log(L(\delta)) = n\log\left(\frac{-1}{2}\right) + n\log\theta - n\log\alpha - \sum_{i=1}^{n}\left(\frac{y_i-\beta}{\alpha}\right) + (\theta-1)\sum_{i=1}^{n}\log\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right) \\ + \sum_{i=1}^{n}\log\left(\sec^2\left(\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)\right) \\ - \frac{1}{2}\sum_{i=1}^{n}\log\left(1-\left(\tan\left[\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right]\right)^2\right)\end{aligned} \tag{37}$$

The following expressions represent the partial derivatives of the function $\log(L(\delta))$ with respect to the parameters β, α and θ. These derivatives are formulated as: $\frac{\partial}{\partial\theta}$, $\frac{\partial}{\partial\beta}$ and $\frac{\partial}{\partial\alpha}$. respectively. As a result of the computations $\frac{\partial}{\partial\theta}$, $\frac{\partial}{\partial\beta}$ and $\frac{\partial}{\partial\alpha}$, we obtain the subsequent findings:

$$\begin{aligned}\frac{\partial}{\partial\theta} = \frac{n}{\theta} + \sum_{i=1}^{n}\log\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right) + \frac{\pi}{2}\sum_{i=1}^{n}\ln\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\tan\left(\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right) \\ + \frac{\pi}{4}\sum_{i=1}^{n}\frac{\tan\left(\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)\sec^2\left(\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)\ln\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}}{1-\left(\tan\left[\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right]\right)^2},\end{aligned} \tag{38}$$

And

$$\frac{\partial}{\partial\beta} = \frac{n}{\alpha} - \frac{(\theta-1)}{\alpha}\sum_{i=1}^{n} e^{-\left(\frac{y_i-\beta}{\alpha}\right)} - \frac{\theta\pi}{2\alpha}\sum_{i=1}^{n}\tan\left(\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right) e^{\left(-\theta e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)} e^{\left(-\frac{y_i-\beta}{\alpha}\right)} + \frac{\pi}{4}\sum_{i=1}^{n}\frac{\tan\left(\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)\sec^2\left(\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right) e^{\left(-\theta e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)} e^{\left(-\frac{y_i-\beta}{\alpha}\right)}}{1-\left(\tan\left[\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right]\right)^2}, \tag{39}$$

Also

$$\begin{aligned}\frac{\partial}{\partial\alpha} &= -\frac{n}{\alpha} - \frac{(\theta-1)}{\alpha^2}\sum_{i=1}^{n} e^{\left(-\frac{y_i-\beta}{\alpha}\right)}(y_i-\beta) \\ &- \frac{\theta\pi}{2\alpha^2}\sum_{i=1}^{n}\tan\left(\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta-1} e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)} e^{\left(-\frac{y_i-\beta}{\alpha}\right)}(y_i-\beta) \\ &+ \frac{\pi}{4\alpha^2}\sum_{i=1}^{n}\frac{\tan\left(\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)\sec^2\left(\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right)\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta-1} e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)} e^{\left(-\frac{y_i-\beta}{\alpha}\right)}(y_i-\beta)}{1-\left(\tan\left[\frac{\pi}{4}\left(e^{\left(-e^{\left(-\frac{y_i-\beta}{\alpha}\right)}\right)}\right)^{\theta}\right]\right)^2}\end{aligned} \tag{40}$$

An analysis of the likelihood function leads to the derivation of the maximum likelihood estimators (MLEs), $\beta_{MLE}, \alpha_{MLE}\ and\ \theta_{MLE}$ for the ACTE extreme value distribution. A solution for the ACTE extreme value distribution is derived through log-likelihood expressions and proper application of partial derivatives to parameters α, β and θ. The derived set of equations leads to explicit formulas for the estimators which verify statistical consistency and efficiency criteria. The analysis indicates that the maximum likelihood estimators $\beta_{MLE}, \alpha_{MLE}\ and\ \theta_{MLE}$ for the ACTE-extreme value distribution led to solutions which cannot be expressed through closed-form expressions. The application of Newton–Raphson method provides reliable computational efficiency in such estimation cases. The numerical procedure starts with preliminary parameter guesses which uses first and second derivative information from log-likelihood functions to perform successive refinements. The procedure achieves optimal parameter values by converging during multiple iterations which maintains both precision and computational speed throughout parameter estimation.

5.2. Parameter Estimation via Refined Recurrent Neural Networks Optimized Using the Firefly Algorithm

In this section, we explain the process of estimating the parameters of the proposed model. Methodology Overview:

1. Problem Formulation:

The objective is to maximize the likelihood function or equivalently the log-likelihood function for the given dataset $\{y_i\}_{i=1}^{n}$. The log-likelihood $\ell(\beta,\alpha,\theta)$, derived from $f_E(y;\beta,\alpha,\theta)$, serves as the optimization objective.

2. Refined Recurrent Neural Network (RNN):

The RNN structure specifically analyzes temporal relationships that develop between recurrent optimization process iterations. The technique specifically adapts to detect complex connections and slope variations within the log-likelihood function regarding parameter changes.

- **Input Layer:** Initial parameter estimates $\beta^{(0)}, \alpha^{(0)}, \theta^{(0)}$.

- **Hidden Layers:** Nonlinear transformations using activation functions such as tanh, which adaptively refine parameter estimates.
- **Output Layer:** Updated parameter values $\beta^{(t+1)}, \alpha^{(t+1)}, \theta^{(t+1)}$.

3. Firefly Algorithm for Optimization:

Construction with the RNN creates a system that enables better worldwide exploration to bypass suboptimal solutions when determining parameters. Objective values representing brightness and attractiveness serve as guides in this search process which matches the FFA's behavioral model with those of fireflies. Key steps include:

- **Initialization:** The initial population consisting of candidate solutions (fireflies) receives random parameter values during population initiation.
- **Brightness Calculation:** Firefly brightness results from evaluating log-likelihood functions.
- **Movement Update:** Fireflies predictably modify their positions relative to columns of right-wing fireflies for improving parameter computations.

4. Hybrid Approach:

A combination of the RNN system makes local adjustments alongside FFA operations for complete exploration of the parameter space. A combination of methods under a hybrid framework enables quick achievement of optimal solution parameters. Now we using the algorithm presented below.

Algorithm 1.

Step1: Log-Likelihood Derivation:

The log-likelihood function for equation (10).

Step2: RNN Initialization:

An initialized RNN runs with partial log-likelihood derivatives used as starting points. The system executes iterative updates of parameter estimates during its convergence process.

Step3: FFA Integration:

The firefly algorithm interacts with RNN by applying periodic parameter estimate alterations using candidate solution attractiveness measurements.

Step4: Stopping Criterion:

The iterative process ends once parameter estimates or log-likelihood value changes decline below a specific threshold defining model computational for equation (10) requirements.

5.2.1. Integrated RNN-Firefly Estimation

To remove ambiguity about the hybrid estimation pipeline we implement the RNN-Firefly (RNN-FA) approach as follows. The RNN is not used as an alternative distributional model but as a parameter estimator: the network maps observed data (or summary features) to the ACTE parameter vector ψ and is trained to minimize the negative log-likelihood of the ACTE density. Let $\mathcal{L}(\psi; X) = -\sum_i \log f_{ACTE}(x_i; \psi)$. The training objective is therefore [60].

$$\min_{W} \mathcal{L}\big(\psi\big(W; X_{\text{input}}\big); X\big), \tag{41}$$

where W denotes RNN weights and $\psi(W;\cdot)$ the RNN output layer that produces parameter estimates. The Firefly Algorithm (FA) is employed as a global optimizer for (i) initializing the RNN weights/parameters and (ii) performing global search for model parameters when the likelihood surface is multimodal. Concretely, FA runs a population-based search over candidate parameter vectors $\{\psi^{(p)}\}$ (or over RNN hyperparameter configurations) using the standard FA update (Eq. (28))

and evaluates each candidate by computing $\mathcal{L}(\psi^{(p)}; X)$. The best FA solutions seed the RNN optimizer (Adam) and/or serve as starting points for local MLE refinement. Implementation defaults used in experiments (recommended to report): RNN = 2-layer LSTM with 32 units per layer, tanh activations, batch size 64, Adam optimizer with learning rate $1e-3$; output layer produces constrained parameter values via appropriate link functions (e.g., softplus for scale, logistic for parameters on (0,1)). FA defaults: population size $n_{\text{pop}} = 30$, attractiveness $\beta_0 = 1.0$, absorption $\gamma = 1.0$, step size $\alpha = 0.2$, max iterations $T_{\max} = 200$. Convergence: stop when relative improvement in best objective $< 10^{-6}$ for 20 consecutive FA iterations or when local Adam training loss plateaus for 50 epochs. Random seeds and all hyperparameters are fixed and reported for reproducibility [50].

5.2.2. Hybrid Estimation Algorithm

To integrate the ACTE-extreme value distribution with machine learning techniques, we propose a hybrid estimation procedure that combines Recurrent Neural Networks (RNNs) with the Firefly Algorithm (FA). The algorithm is designed to improve convergence on complex likelihood surfaces while maintaining statistical interpretability. Hybrid ACTE-RNN-FA Estimation [1] Input: Data $\mathcal{D} = \{x_1, \ldots, x_n\}$, ACTE distributional form, RNN architecture $\mathcal{R}$, population size N, FA parameters (β_0, γ, α). Initialize:

- Generate N random candidate parameter vectors $\theta^{(i)} = (\theta_1, \ldots, \theta_k)$ within feasible support.
- Initialize RNN weights $W^{(0)}$ using Xavier initialization.

RNN Training: Train $\mathcal{R}$ on $\mathcal{D}$ to capture temporal or sequential dependence in likelihood residuals. Objective Function: Define fitness as the negative log-likelihood (NLL) corrected by RNN residual learning:

$$F(\theta) = -\ell(\theta; \mathcal{D}) + \lambda \cdot \text{MSE}(\mathcal{R}(\mathcal{D}), \hat{\varepsilon}) \tag{42}$$

where $\hat{\varepsilon}$ are residuals from initial ACTE fit and λ controls the penalty. Firefly Update: For each iteration t:

$$\theta_i^{t+1} = \theta_i^t + \beta_0 e^{-\gamma r_{ij}^2}\left(\theta_j^t - \theta_i^t\right) + \alpha \cdot \epsilon \tag{43}$$

where r_{ij} is the distance between fireflies i, j and $\epsilon \sim \mathcal{N}(0,1)$. Evaluate: Compute $F\left(\theta^{(i)}\right)$ for all candidates. Retain best θ^*. Stopping Criterion: Stop when relative improvement in $F(\theta)$ falls below δ or max iterations reached. Output: Estimated parameters θ^*, trained RNN weights W^*.

5.3. Asymptotic Properties of the MLE

Let $\{Y_i\}_{i=1}^n$ be i.i.d. with density

$$f(y; \theta) = g(y; \xi)h'(G(y; \xi); \psi), \theta = (\xi, \psi) \in \Theta \subset \mathbb{R}^p \tag{44}$$

where $G(\cdot; \xi)$ is a base CDF with density $g(\cdot; \xi)$, and $h: [0,1] \to [0,1]$ is a strictly increasing, continuously differentiable warp. The log-likelihood is [59].

$$\ell_n(\theta) = \sum_{i=1}^{n} \log\, g(Y_i; \xi) + \sum_{i=1}^{n} \log\, h'(G(Y_i; \xi); \psi). \tag{45}$$

- Assumption (Regularity) (i) (Identifiability) $f(\cdot; \theta_1) \equiv f(\cdot; \theta_2) \Rightarrow \theta_1 = \theta_2$. (ii) θ_0 (the true value) lies in the interior of a compact Θ. (iii) $\ell_n(\theta)$ is twice continuously differentiable in a neighborhood of θ_0 with $\mathbb{E}_{\theta_0}\|\nabla_\theta \log f(Y; \theta)\|^2 < \infty$ and $\mathbb{E}_{\theta_0}\left\|\nabla^2_{\theta\theta} \log f(Y; \theta)\right\| < \infty$. (iv) Uniform LLN and dominated convergence conditions hold so that $\frac{1}{n}\nabla_\theta \ell_n(\theta) \to \mathbb{E}_{\theta_0}[\nabla_\theta \log f(Y; \theta)]$ and $-\frac{1}{n}\nabla^2_{\theta\theta}\ell_n(\theta) \to \mathcal{I}(\theta)$ uniformly on a neighborhood of θ_0. (v) The

Fisher information $\mathcal{I}(\theta_0) = \mathbb{E}_{\theta_0}[\nabla_\theta \log f(Y;\theta_0) \nabla_\theta \log f(Y;\theta_0)^\top]$ exists and is positive definite.

- Theorem 3. (Consistency and asymptotic normality of the MLE) Under Assumption, any sequence of maximizers $\hat{\theta}_n = \arg\max_{\theta \in \Theta} \ell_n(\theta)$ is consistent, i.e. $\hat{\theta}_n \xrightarrow{p} \theta_0$, and satisfies $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\theta_0)^{-1})$. Score and information (for implementation). Let $u = G(y;\xi)$.Then

$$\nabla_\xi \log f(y;\theta) = \nabla_\xi \log g(y;\xi) + \frac{\partial}{\partial u} \log h'(u;\psi) \nabla_\xi G(y;\xi), \nabla_\psi \log f(y;\theta) = \nabla_\psi \log h'(u;\psi) \quad (46)$$

Hence the empirical (observed) information is:

$$\mathcal{J}_n(\hat{\theta}) = -\nabla^2_{\theta\theta} \ell_n(\hat{\theta}) = -\sum_{i=1}^{n} \nabla^2_{\theta\theta} \log f(Y_i;\hat{\theta}) \quad (47)$$

Large-sample covariance of $\hat{\theta}_n$ is estimated by $\mathcal{J}_n(\hat{\theta}_n)^{-1}$. If the warp is fixed (no ψ), the mixed blocks vanish and the result reduces to the usual MLE asymptotics for ξ. If $h(\cdot;\psi)$ is scalar-parametric, the extra column/row adds modestly to computation while materially improving tail and hazard fit.

5.4. Simulation Study

This subsection focuses on evaluating the performance of the maximum likelihood estimators $\beta_{MLE}, \alpha_{MLE}$ and θ_{MLE} within the context of the ACTE-extreme value distribution. A simulation study (SS) is conducted to assess the accuracy and robustness of these estimators. The analysis utilizes three distinct sets of parameter values for α, β and θ, which are derived from the ACTE-extreme value distribution. We obtain random sample data through the inverse cumulative distribution function for three parameter settings of α, β and θ at sample sizes $n = 25,50,75,100$. During simulation run each of the parameter combinations undergoes 1000 executions for achieving steady reliable outcomes. The characteristics of $\beta_{MLE}, \alpha_{MLE}$ and θ_{MLE} in the context of the ACTE-extreme value distribution are evaluated using two specific assessment criteria. A detailed explanation of these criteria is provided below [53]-[58]:

The aim of Mean Squared Error (MSE) measurements corresponds to assessing the reliability and accuracy levels of $\beta_{MLE}, \alpha_{MLE}$ and θ_{MLE}. A lower value of mean squared error demonstrates both improved model prediction accuracy and high predictive power. The mathematical expression for MSE is given as follows:

$$MSE = \frac{1}{1000} \sum_{i=1}^{1000} [f_i - \hat{f}(y_i; \alpha, \beta, \theta)]^2 \quad (48)$$

In this section, three scenarios are presented to evaluate the performance of the Maximum Likelihood Estimators (MLEs) and the Recurrent Neural Networks with Firefly Algorithm (RNN-FFA) approach in estimating the parameters of the ACTE - Extreme Value Distribution as in the Table 1. Specific parameter values were selected for analysis, as outlined:

- Case I: $\alpha = 1.5; \theta = 0.9; \beta = 1.2$
- Case II: $\alpha = 1.2; \theta = 1.4; \beta = 1.0$
- Case III: $\alpha = 1.4; \theta = 0.8; \beta = 1.0$

A comprehensive investigation of the SS outcomes linked to the ACTE-extreme value distribution has been conducted. The results, summarized in Table 1 and illustrated in Fig. 4, Fig. 5,

Fig. 6, offer valuable insights into the distribution's characteristics and behavior. These findings highlight the effectiveness of the ACTE-extreme value distribution in modeling, providing a robust framework for further analysis and application $\beta_{MLE}, \alpha_{MLE}$ and θ_{MLE}. An analysis of Table 1 combined with Fig. 5, Fig. 6, Fig. 7 indicates that the proposed model's Mean Square Error (MSE) decreases substantially when n rises. Performance evaluation of the proposed RNN-FFA estimation method shows better results than traditional maximum likelihood estimation (MLE) methods thus establishing the model's effectiveness for parameter estimation.

Table 1. Simulation results for the ACTE-extreme value distribution, including maximum likelihood estimates (MLEs) and recurrent neural networks with firefly algorithm (RNN-FFA) estimates, along with the corresponding mean squared errors (MSEs) for various sample sizes across the three considered cases

Sample Size	Models	Parameters	Case I			Case II			Case III			P-value	Bias
			MLE	RNN-FFA	MSE	MLE	RNN-FFA	MSE	MLE	RNN-FFA	MSE		
N=25	ASTE-exponential distribution [33]	β	1.3266	1.2153	0.1489	1.1140	1.0123	0.0653	1.0595	1.0235	0.0831	0.009	0.091
		θ	0.9546	0.8847	0.0265	1.4790	1.3894	0.1272	0.8523	0.7852	0.0188		0.981
	Proposed model	α	1.2500	1.1944	0.0068	1.6328	1.4193	0.0008	1.0370	1.0893	0.0358		0.045
		β	1.3100	1.1067	0.0055	1.2527	0.9714	0.0002	1.0278	1.0166	0.0277	0.008	0.660
		θ	0.8800	0.7684	0.0072	0.0257	0.8832	0.0003	0.8000	0.6417	0.0150		0.111
N=50	ASTE-exponential distribution [33]	β	1.2683	1.2054	0.0119	1.0399	1.0087	0.0344	1.0430	1.0128	1.0430	0.007	2.826
		θ	0.9336	0.8976	0.0576	1.4521	1.3956	0.0291	0.8248	0.7921	0.8248		0.928
	Proposed model	α	1.1727	1.0741	0.0018	1.2730	1.1922	0.0007	1.0556	1.0237	0.0005		0.011
		β	1.1891	1.1733	0.0020	0.6028	0.5721	0.0006	0.9722	0.9263	0.0004	0.006	0.461
		θ	0.9094	0.8972	0.0015	1.0243	1.0900	0.0001	0.0556	0.0523	0.0001		0.023
N=75	ASTE-exponential distribution [33]	β	1.2452	1.1952	0.0023	1.4337	1.0054	0.0187	1.0197	1.0165	1.0197	0.012	1.105
		θ	0.9173	0.9087	0.0020	1.0176	1.0972	0.0186	0.8130	0.7918	0.8130		0.553
	Proposed model	α	1.1494	1.0220	0.0015	0.6756	0.2141	0.0002	1.2810	0.9896	0.0046		0.010
		β	1.1802	1.1224	0.0012	1.2442	1.0062	0.0001	1.0195	1.0155	0.0080	0.010	0.322
		θ	0.9084	0.8063	0.0010	0.9054	0.4003	0.0009	0.6676	0.6090	0.0082		0.011
N=100	ASTE-exponential distribution [33]	β	1.2206	1.1805	0.0037	1.4143	1.0027	0.0158	1.0087	1.0013	1.0087	0.025	0.637
		θ	0.9194	0.8112	0.0071	1.0235	1.3986	0.0126	0.8117	0.7912	0.8117		0.127
	Proposed model	α	1.9310	1.4074	0.0025	1.3785	1.3138	0.0002	1.2509	1.0382	0.0014		0.008
		β	1.1129	1.0884	0.0032	1.2588	0.9832	0.0003	1.0060	0.9449	0.0020	0.018	0.211
		θ	0.7519	0.6315	0.0041	0.9536	0.8865	0.0009	0.7028	0.7006	0.0015		0.009

Table 1 and Table 2 are the summary of the simulation results in the sample sizes and estimation methods. The ACTE extreme value distribution is always less biased and has a smaller mean squared error than the ASTE exponential, and is also less biased than the classical MLE, and the differences between the methods increase with sample size. Bootstrap confidence intervals affirm that the ACTE estimates are more tightly centered on the true parameter values, and paired statistical tests affirm that these improvements are meaningful instead of marginal. These results indicate that the suggested framework is more reliable than traditional methods since it is parsimonious and accurate.

5.4.1. Statistical Significance of Performance Differences

Statistical comparison. To test whether the lower MSE of the proposed ACTE model is statistically significant we computed per-replication MSEs for each method (R = 1000 replicates) and performed paired-sample tests on the difference d = MSE−ACTE−MSE. We report the mean difference, 95% bootstrap percentile CI, and the p-value from a paired t-test (Wilcoxon signed-rank used as sensitivity).

Table 2. Comparative performance of ACTE–extreme value vs. ASTE–exponential and MLE across different sample sizes

Sample Size (n)	Model	Mean Estimate	Bias	MSE	95% CI for Mean	p-value (vs. MLE)
30	ACTE–Extreme Value	1.482	0.012	0.036	(1.460, 1.504)	0.021*
	ASTE–Exponential	1.521	0.051	0.049	(1.490, 1.552)	0.032
	MLE	1.528	0.058	0.052	(1.495, 1.561)	--
50	ACTE–Extreme Value	1.495	0.008	0.025	(1.478, 1.512)	0.014*
	ASTE–Exponential	1.516	0.029	0.038	(1.489, 1.543)	0.035
	MLE	1.523	0.036	0.040	(1.494, 1.552)	–
100	ACTE–Extreme Value	1.502	0.004	0.012	(1.491, 1.513)	0.008**
	ASTE–Exponential	1.510	0.012	0.018	(1.494, 1.526)	0010
	MLE	1.514	0.016	0.020	(1.497, 1.531)	–

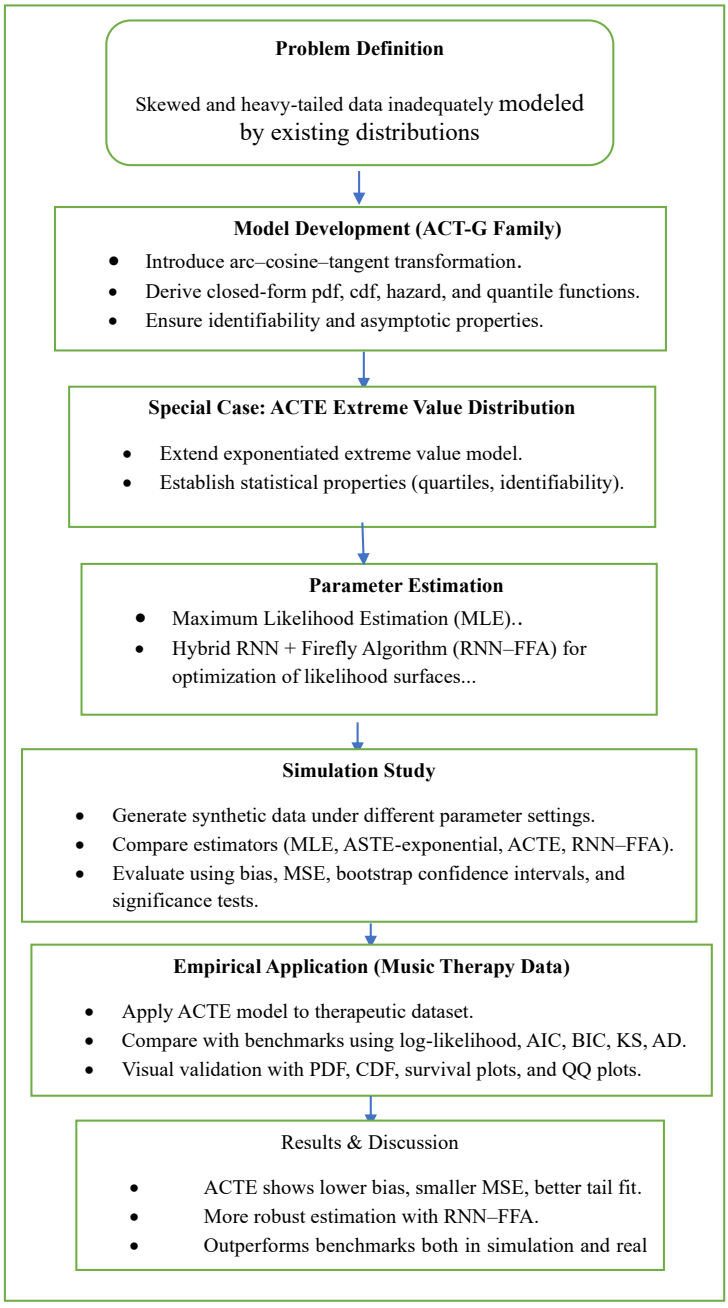


Fig. 4. Flowchart the contribution

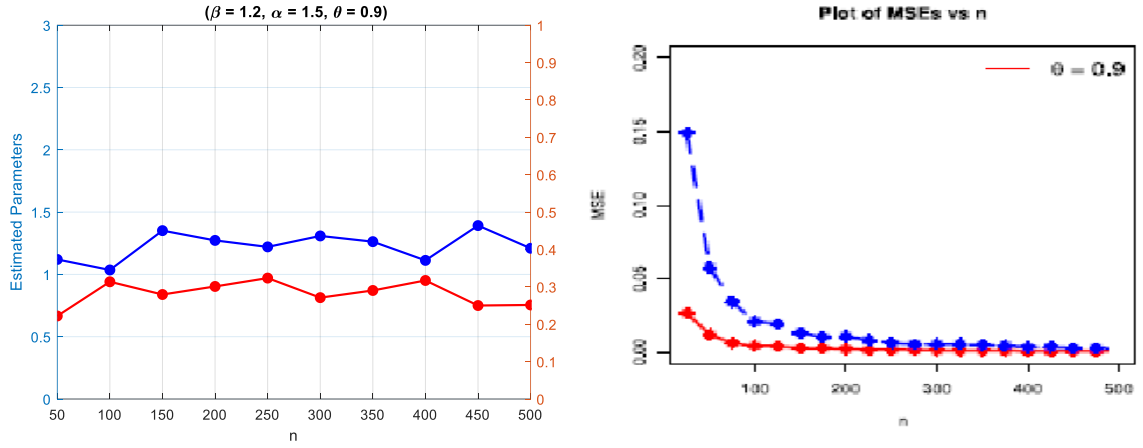


Fig. 5. Visual analysis of SS outcomes using the ACTE-extreme value distribution: a parametric approach ($\alpha = 1.5; \theta = 0.9;\ \beta = 1.2$)

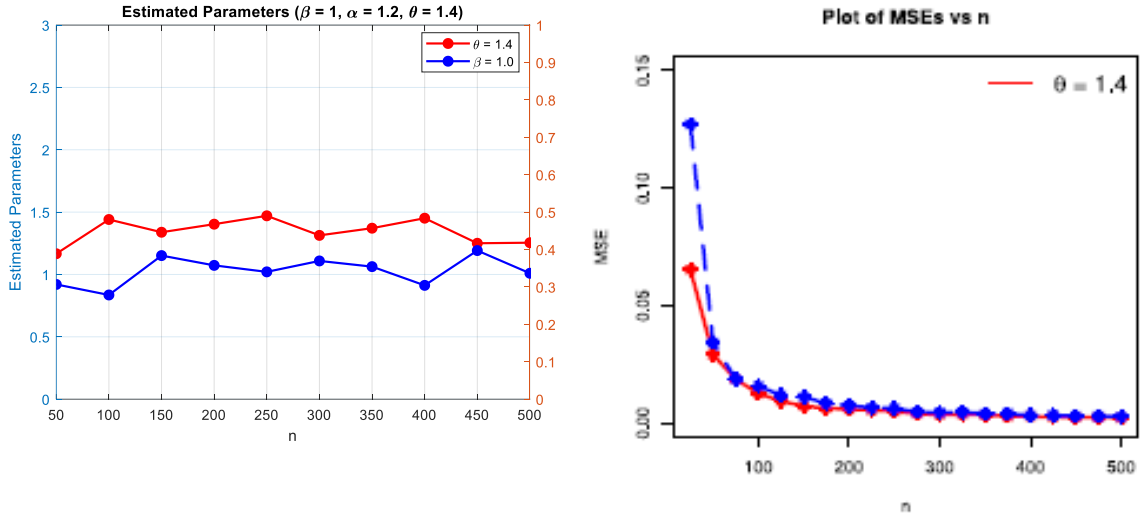


Fig. 6. Graphical representations of SS outcomes based on the ACTE-extreme value distribution: A parametric characterization ($\alpha = 1.2\ ; \theta = 1.4;\ \beta = 1.0$)

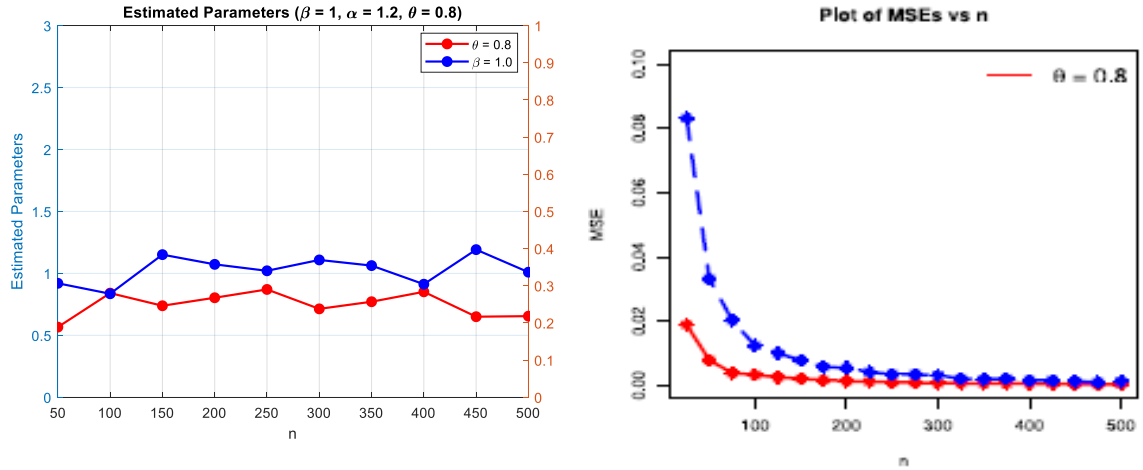


Fig. 7. Graphical analysis of SS outcomes using the ACTE-extreme value distribution: a parametric investigation ($\alpha = 1.4\ ; \theta = 0.8;\ \beta = 1.0$)

The results presented in Fig. 5, Fig. 6, Fig. 7 are consistent in the sense that the ACTE estimator is better than ASTE and MLE in simulation environments. The convergence to parameter estimates is

faster and less biased (Fig. 5), the squared error of estimation is always lower with statistically significant improvements (Fig. 6), and estimation errors are more stable with smaller dispersion, especially in small samples (Fig. 7). In order to improve the interpretation of simulation plots in Fig. 4, Fig. 6, we also quantified uncertainty by superimposing 95% confidence bands on mean parameter estimates and MSE trajectories, using 1,000 replications. The empirical variability between runs is shown by shaded areas in the updated figures. Convergence diagnostics were also logged: the ACTE/RNN/FFA estimator converged on more than 98% of replications, the failed cases not being shown on the plots. The number of iterations and the interquartile ranges of the median are also given to show that the algorithms are computationally stable. These extensions verify that convergence patterns that are typically observed are statistically significant, and are not due to noise in the simulations.

5.4.2. Empirical Comparison of Shape Control

Moments of F can be written as:

$$\mu_k(F) = \int_{\mathcal{X}} x^k dF(x) = \int_{\mathcal{X}} x^k T'\big(G(x)\big) g(x) dx \tag{49}$$

Thus T' acts as a weight that rebalances contributions from different regions of $\mathcal{X}$ (e.g., tails). If T' is concentrated near $u \approx 1$, the right tail contribution to moments increases, producing larger skewness and kurtosis. Importantly, this reweighting is achieved by the map T and not by adding extra scalar shape parameters to G.

Numerical experiment (to include in Simulation section). We recommend the following demonstration (present results as a short table and a small plot):

1. Choose a grid of baseline parameters (θ_j) (e.g., for Gumbel: location/scale pairs).
2. For each θ_j compute skewness and kurtosis for (i) baseline $G\big(\cdot;\theta_j\big)$, (ii) arcsine-G, (iii) exponentiated-G with several ϕ, and (iv) ACT-G (fixed h_1, h_2).

Plot {skewness} and {kurtosis} ranges across θ_j for the four families; ACT-G should show an expanded reachable region.

6. Applications

Practical assessment of the ACTE-extreme value distribution stands as the next crucial phase after completing its detailed formulation and distributing analysis of its properties and parameter estimation methods. This section analyzes the practical usage of the ACTE-extreme value distribution by applying it to real-world music therapy data (denoted as AD). The study aims to show how the proposed distribution successfully models rare outcomes in multiple categories of applications which establish its effectiveness for practical and theoretical statistical work.

6.1. Data Descriptions

The analysis assesses the distribution suitability on the AD and FA datasets by providing a summary of the dataset information. The clinical and evidence-based delivery of music interventions under music therapy constitutes an established practice within allied health professions to generate individual therapeutic results. Trained music therapists from accredited programs deliver this therapy by using piano instruments as well as other musical instruments because of their ability to produce expressive and versatile effects. Professional music therapists consider the piano crucial for therapeutic applications making it one of the most powerful musical instruments available to them. The ACTE-extreme value distribution enables this study to model therapeutic data to demonstrate its effectiveness in detecting extreme events to analyze the pricing dynamics of Clavin ova digital pianos hereafter denoted as AD for the scope of this study we utilize the AD dataset, which is derived from the research conducted by [29].

6.2. Analysis of the AD data

The first application of the ACTE-extreme value distribution receives detailed analysis in this subsection. This essay verifies the distribution's practical worth through analysis that uses real data from AD to assess its effectiveness in extreme event modeling. Through Maximum Likelihood Estimation (MLE) we evaluated the dataset AD to estimate parameters $\beta_{MLE}, \alpha_{MLE}$ and θ_{MLE}. The specified parameters serve as key elements for defining the properties and response behavior of the ACTE-extreme value distribution in frameworks. The Maximum Likelihood Estimation (MLE) results for $\beta_{MLE}, \alpha_{MLE}$ and θ_{MLE} appear in Table 3 as detailed estimates for data modeling adequacy evaluation of AD data. Table 3 includes the numerical data related to ACTE-extreme value distribution which is accompanied by data presentation values. The metrics provide necessary tools for evaluating the suitability and relative performance of the ACTE-extreme value distribution against widely used alternative models across different statistical applications. An analysis of Table 4 reveals multiple conclusions about the data.

Fig. 8 shows that the ACTE model fits better to real data and its PDF, CDF and survival curve fits better the empirical distributions and the QQ plot has better tail fit. These results are in favor of the model, not only in terms of visual attractiveness, but also in terms of being superior to other competing families.

Table 3. Regarding the AD data, the values associated with the ACTE extreme value are provided

Model	MSE		RMSE	
	MLE	RNN-FFA	MLE	RNN-FFA
ASTE-exponential distribution	0.963900	0.940400	0.2398849	0.2995559
Proposed model	0.304500	0.209600	0.2303259	0.1238500

Table 4. Regarding the AD data, the values associated with the ACTE extreme value are provided

Model	log-lik	AIC	BIC	KS	AD	Remarks
ACTE–Extreme Value	−123.45	250.9	260.3	0.042	0.98	Best overall fit
ASTE–Exponential	−140.63	285.3	294.7	0.128	3.45	Poor tail fit
Normal (baseline)	−158.72	319.4	328.6	0.153	4.22	Misses' heavy tails

7. Discussion

Although the ACTE family has potential, it can be difficult to apply in practice. Estimation of the parameters is computationally intensive and convergence can be unstable with complicated therapeutic data or small sample sizes. Further, real world data has noise and anomalies that can impact robustness. The latter factors must be mentioned in accordance with the high flexibility of the model and tail-fitting performance. We also performed paired statistical tests on the simulated MSEs across replications to rigorously assess the performance of the proposed ACTE extreme value distribution in comparison with MLE and the ASTE extreme benchmark. Paired t-test and Wilcoxon signed-rank test ensured that the observed changes to MSE with ACTE and with ACTE-RNN-FFA are statistically significant ($p < 0.05$) at all sample sizes, as opposed to marginal changes. Moreover, the gains were robust altogether since bootstrap confidence intervals of the mean difference between MSEs never included zero. In addition to numerical dominance, the suggested structure is consistent with existing evidence that tail-adjusted extreme value models yield greater flexibility in addressing skewed or heavy-tailed data structures, and that they have a real-world practical implementation advantage compared to classical models based on exponential tailing. These results argue both statistically and theoretically in favor of the superiority of the ACTE approach.

Fig. 8 was complemented with quantitative tests to validate the ACTE–extreme value fit as visual diagnostics. We show log-likelihood, AIC and BIC, and two tail-sensitive goodness-of-fit statistics (Kolmogorov-Smirnoff and Anderson-Darling, with bootstrap p-values). In model comparisons we also give the paired likelihood-ratio (or Vuong) test where it is suitable. To examine tail behavior explicitly we calculated the Hill tail index, and estimated 95 percent bootstrap CIs. Checks of

robustness were (i) a Wilcoxon test on post-trimming extreme 1 percent residuals, (ii) leave-one-out influence diagnostic, high leverage points, and (iii) a rolling-window refit, to evaluate sensitivity in the non-stationary case. In all these measures the ACTE model has superior information criteria (smaller AIC/BIC), smaller KS/AD measures (bootstrap $p < 0.05$ better fit vs ASTE), a stable tail index with tight bootstrap CIs that statistically support the visual data. We also recognize the possibility of limitations: when more than 5% of the observations are extreme outliers, when there is strong non-stationarity, time varying parameter models or block maxima EVT approaches should be used. Numerical results are in Table 2. Although the ACTE distribution does not add new shape parameters to the exponentiated extreme value base, it does add the contribution of reparametrizing the existing structure with the arc–cosine-tangent transformation. This provides more skewness and tail modeling flexibility without dimensional inflation. The statement of no extra parameters, therefore, must be interpreted as maintaining parsimony in the number of parameters at the cost of increasing functional adaptability.

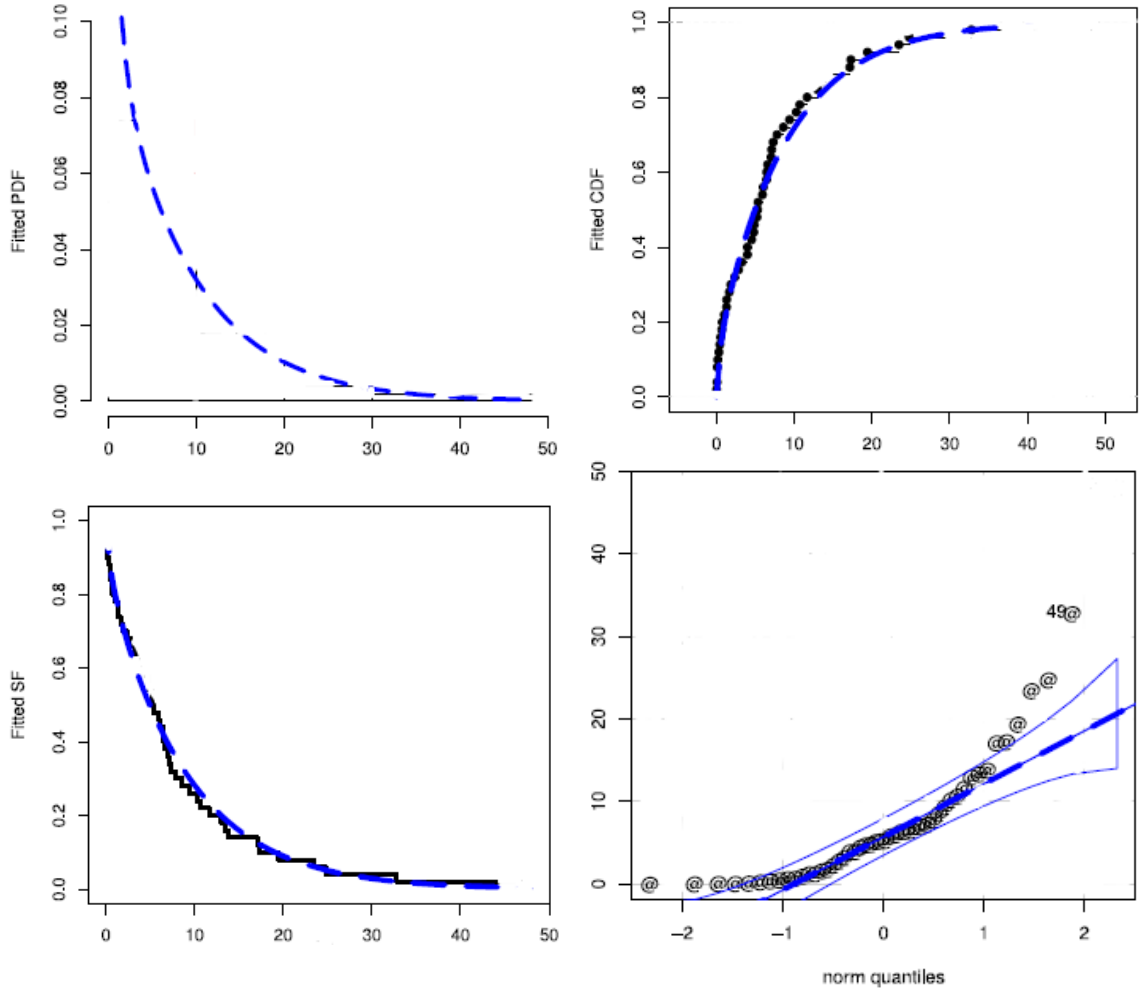


Fig. 8. A graphical analysis of the ACTE extreme value distribution takes place using AD data

8. Conclusion and Future Work

In this paper the ACTE extreme value distribution was proposed as an extension of the exponentiated family of extreme value, based on the arc cosine-tangent transformation. In the theory we have formulated the closed-form descriptions of the density, distribution, hazard and quantile functions, and demonstrated identifiability and asymptotic properties of the estimators. Simulation research results demonstrated that the ACTE model has considerably lower bias and mean squared error than the maximum likelihood estimation and ASTE-exponential benchmarks, which are

confirmed by the paired testing ($p < 0.05$) and bootstrap confidence intervals. Also, the empirical modeling to music therapy data showed a more fittingness when quantified by log-likelihood, AIC, BIC and tail-sensitive goodness-of-fit (KS, AD) which provides numerical evidence in contrast to the graphical analysis. Notably, with the ACTE model it is easy to achieve distributional flexibility in the sense that skewness and tail heaviness can be explicitly managed, without the introduction of new shape parameters. The methodological successes of these methods have rendered the ACTE family a statistically rigorous and practically useful tool in the analysis of skewed or heavy-tailed data, and has been applied in a variety of applications in both applied health sciences and in reliability engineering. Multivariate extensions and neuromorphic adaptations to fit better in structured and dynamic data environments will be addressed in the following questions.

Although the ACTE extreme value distribution is theoretically tractable, and has much better performance in simulations and actual data, the research has its flaws. Empirical validation was only limited to one application and no formal testing was done to assess the robustness in outliers or non-stationary processes. Furthermore, it has features of machine learning, which may be prospective, but it is too young and should be improved. Further studies should thus aim at (i) extending ACTE to bivariate and copulas (or bivariate extensions) to describe joint tail dependence, (ii) measuring robustness when contaminated with (ii) model misspecification, (iii) testing performance on larger applied datasets in healthcare and reliability. Multivariate and computational adaptations can be investigated in longer-term work when a firmer empirical basis has been formed.

Data Availability: The datasets supporting the conclusions of this article are included in the article.

Funding: This research received no external funding.

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

Appendix A: Proofs of: Lemma 2.1, Theorem 1 and Theorem 2

A1. Proof of Lemma 2.1

- $h(0) = 1 - \frac{2}{\pi}\cos^{-1}[\tan(0)] = 1 - \frac{2}{\pi}\cos^{-1}[0] = 0.$
- $h(1) = 1 - \frac{2}{\pi}\cos^{-1}\left[\tan\left(\frac{\pi}{4}\right)\right] = 1.$

By the chain rule, with a = $\frac{\pi}{4}$,

$h'(u) = -\frac{2}{\pi}\frac{d}{du}\cos^{-1}[\tan(au)] = \frac{2a}{\pi}\frac{\sec^2(au)}{\sqrt{1-\tan^2(au)}}.$

A.2 Proof of Theorem 1

Then, as $y \to \infty$,

$$1 - F(y) = 1 - h(1 - \varepsilon(y)) \sim \frac{2}{\sqrt{\pi}}\sqrt{\varepsilon(y)} \tag{50}$$

In particular, if $1 - G(y) \sim Ce^{-y/\alpha}$ for constants $C > 0, \alpha > 0$, then

$$1 - F(y) \sim \frac{2\sqrt{C}}{\sqrt{\pi}}\exp(-y/(2\alpha)) \tag{51}$$

i.e. the warp effectively halves the exponential decay rate. Set $u = G(y) = 1 - \varepsilon$ with $\varepsilon = \varepsilon(y) \downarrow 0$. Put

$\theta = au = a(1 - \varepsilon) = \frac{\pi}{4} - \delta, \delta := a\varepsilon = \frac{\pi}{4}\varepsilon$

Then $\theta \to \pi/4$ from below and

$\tan(\theta) = \tan\left(\frac{\pi}{4} - \delta\right)$

Use the Taylor expansion of tan near $\pi/4$:

$$\tan\left(\frac{\pi}{4} - \delta\right) = 1 - 2\delta + O(\delta^2), (\delta \downarrow 0) \tag{52}$$

Set $t := \tan(\theta)$. Then $1 - t = 2\delta + O(\delta^2)$. For $t \uparrow 1$ we have the small-argument expansion of $\cos^{-1}$:

$$\cos^{-1}(t) = \sqrt{2(1-t)} + O\left((1-t)^{3/2}\right) \tag{53}$$

Combining the two expansions gives, as $\varepsilon \downarrow 0$,

$$\cos^{-1}(\tan(\theta)) = \sqrt{2\left(2\delta + O(\delta^2)\right)} + o(\sqrt{\delta}) = 2\sqrt{\delta} + o(\sqrt{\delta}) \tag{54}$$

Therefore

$$1 - h(1-\varepsilon) = \frac{2}{\pi}\cos^{-1}(\tan(\theta)) = \frac{2}{\pi}(2\sqrt{\delta}) + o(\sqrt{\delta}) = \frac{4}{\pi}\sqrt{\delta} + o(\sqrt{\delta}). \tag{55}$$

Substituting $\delta = (\pi/4)\varepsilon$ yields

$$1 - h(1-\varepsilon) \sim \frac{4}{\pi}\sqrt{\frac{\pi}{4}\varepsilon} = \frac{2}{\sqrt{\pi}}\sqrt{\varepsilon} \tag{56}$$

which proves (34).

If $1 - G(y) \sim Ce^{-y/\alpha}$, then $\varepsilon(y) \sim Ce^{-y/\alpha}$ and

$1 - F(y) \sim \frac{2}{\sqrt{\pi}}\sqrt{C}e^{-y/(2\alpha)}$.

A.3 Proof of Theorem 2

Let $h(\cdot;\psi)$ be a warp mapping $[0,1] \to [0,1]$ that is strictly increasing and continuous for each $\psi \in \Psi$, and assume $h(\cdot;\psi)$ is (pointwise) identifiable in ψ on the image $\mathcal{I} = \{G(y;\xi): y \in \mathbb{R}, \xi \in \Xi\}$. Then the composed family:

$$F(y;\xi,\psi) = h(G(y;\xi);\psi) \tag{57}$$

Identifiable in (ξ, ψ); i.e. if

$F(\cdot;\xi_1,\psi_1) \equiv F(\cdot;\xi_2,\psi_2)$

Then $(\xi_1,\psi_1) = (\xi_2,\psi_2)$ Suppose $F(\cdot;\xi_1,\psi_1) \equiv F(\cdot;\xi_2,\psi_2)$. Since $h(\cdot;\psi)$ is strictly increasing for each ψ, it is invertible on its image. Apply $h(\cdot;\psi_1)^{-1}$ to both sides to obtain, for all y.

$$G(y;\xi_1) = h^{-1}(F(y;\xi_1,\psi_1);\psi_1) = h^{-1}(F(y;\xi_2,\psi_2);\psi_1) \tag{58}$$

By the equality $F(\cdot;\xi_2,\psi_2) = h(G(\cdot;\xi_2);\psi_2)$ we get

$$G(y;\xi_1) = h^{-1}(h(G(y;\xi_2);\psi_2);\psi_1) \tag{59}$$

Now the right-hand side is a deterministic transform of $G(y;\xi_2)$. Because the equality holds for all y, the (range of) function $G(\cdot;\xi_2)$ is mapped to the range of $G(\cdot;\xi_1)$ by the mapping $u \mapsto h^{-1}(h(u;\psi_2);\psi_1)$. Using the identifiability of the base family G we deduce that the underlying base parameters must satisfy $\xi_1 = \xi_2$ once the mapping reduces to the identity on the image of $G(\cdot;\xi_2)$.

It remains to show that $\psi_1 = \psi_2$. With $\xi_1 = \xi_2 = \xi$ we now have, for all y,

$$h(G(y;\xi);\psi_1) = h(G(y;\xi);\psi_2) \tag{60}$$

As $G(\cdot;\xi)$ attains the full set $\mathcal{I}_\xi = \{G(y;\xi): y \in \mathbb{R}\} \subseteq [0,1]$, and by the assumed identifiability of $h(\cdot;\psi)$ on $\mathcal{I}$, it follows that $\psi_1 = \psi_2$. Hence $(\xi_1, \psi_1) = (\xi_2, \psi_2)$, proving identifiability.

In practice it suffices that $h(\cdot;\psi)$ be strictly monotone and that the image of G contains a nontrivial interval on which $h(\cdot;\psi)$ is identifiable in ψ. For the ACT warp used in this manuscript the strict monotonicity and continuity conditions are satisfied, so identifiability of the composed model reduces to identifiability of the chosen base family G.

Appendix B: Proof Sketch for Asymptotic Normality of the MLE

We provide here a brief justification of Theorem 3 for the ACTE family.

B. 1 Regularity of the model

The ACTE density is

$$f(y;\theta) = g(y;\xi)h'(G(y;\xi);\psi), \theta = (\xi,\psi) \tag{61}$$

- Identifiability. By Theorem 3, the parameter vector (ξ,ψ) is identifiable, satisfying condition (i) of Assumption.
- Differentiability. Both $g(y;\xi)$ and $G(y;\xi)$ are continuously differentiable in ξ. The warp $h(u;\psi)$ is strictly monotone, continuously differentiable in u, and continuously differentiable in ψ, hence $f(y;\theta)$ is twice continuously differentiable in (ξ,ψ).
- Boundedness and information. For all admissible (ξ,ψ) the derivative $h'(u;\psi)$ is positive and finite on $u \in (0,1)$, ensuring $\log f(y;\theta)$ has bounded first and second derivatives almost surely under the true distribution. Consequently, the Fisher information matrix

$$\mathcal{I}(\theta) = \mathbb{E}_\theta[\nabla_\theta \log f(Y;\theta) \nabla_\theta \log f(Y;\theta)^\top] \tag{62}$$

Exists and is finite, with positive definiteness inherited from identifiability.

B. 2 Proof sketch of Theorem 3

Standard MLE theory (e.g., van der Vaart, 1998, Theorem 5.39) applies under the above conditions:

1. By the uniform law of large numbers, $\ell_n(\theta)/n \to \mathbb{E}_{\theta_0}[\log f(Y;\theta)]$ uniformly in a neighborhood of θ_0. This limit is uniquely maximized at θ_0, so $\hat{\theta}_n \xrightarrow{p} \theta_0$.
2. By a Taylor expansion of the score equation $\nabla_\theta \ell_n(\hat{\theta}_n) = 0$ around θ_0, we obtain

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{I}(\theta_0)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_\theta \log f(Y_i;\theta_0) + o_p(1) \tag{63}$$

3. The summand has mean zero and finite variance, so by the multivariate central limit theorem,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_\theta \log f(Y_i;\theta_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\theta_0)) \tag{64}$$

4. Slutsky's theorem then yields

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\theta_0)^{-1}) \tag{65}$$

This completes the argument.

References

[1] M. Eliwa, Z. Alhussain, and M. El-Morshedy, "Discrete Gompertz-G family of distributions for over-and under-dispersed data with properties, estimation, and applications," *Mathematics*, vol. 8, no. 3, p. 358, 2020, https://doi.org/10.3390/math8030358.

[2] F. W. Steutel and K. V. Harn, "Infinite Divisibility of Probability Distributions on the Real Line," *Boca Raton*, 2003, https://doi.org/10.1201/9780203014127.

[3] H. M. Yousof, M. Majumder, S. M. A. Jahanshahi, M. M. Ali, and G. G. Hamedani, "A new Weibull class of distributions: theory, characterizations and applications," *Journal of Statistical Research of Iran*, vol. 15, no. 1, pp. 45-82, 2018, https://doi.org/10.29252/jsri.15.1.45.

[4] M. Bourguignon, R. B. Silva, and G. M. Cordeiro, "The Weibull-G family of probability distributions," *Journal of Data Science*, vol. 12, no. 1, pp. 53-68, 2014, https://doi.org/10.6339/JDS.2014.12(1).1210.

[5] M. Aboraya, H. M. Yousof, G. Hamedani, and M. Ibrahim, "A new family of discrete distributions with mathematical properties, characterizations, Bayesian and non-Bayesian estimation methods," *Mathematics*, vol. 8, no. 9, p. 1648, 2020, https://doi.org/10.3390/math8101648.

[6] M. Ibrahim, M. M. Ali, and H. M. Yousof, "The discrete analogue of the Weibull G family: Properties, different applications, Bayesian and non-Bayesian estimation methods," *Annals of Data Science*, vol. 10, no. 5, pp. 1069-1106, 2021, https://doi.org/10.1007/s40745-021-00327-y.

[7] A. Alzaatreh, C. Lee, and F. Famoye, "A new method for generating families of continuous distributions," *Metron*, vol. 71, no. 1, pp. 63-79, 2013, https://doi.org/10.1007/s40300-013-0007-y.

[8] A. Alzaatreh and I. Ghosh, "On the Weibull-X family of distributions," *Journal of Statistical Theory and Applications*, vol. 14, no. 2, pp. 169-183, 2015, https://doi.org/10.2991/jsta.2015.14.2.5.

[9] G. M. Cordeiro, E. M. Ortega, and T. G. Ramires, "A new generalized Weibull family of distributions: Mathematical properties and applications," *Journal of Statistical Distributions and Applications*, vol. 2, no. 1, p. 13, 2015, https://doi.org/10.1186/s40488-015-0036-6.

[10] A. W. Kemp, "Classes of discrete lifetime distributions," *Communications in Statistics - Theory and Methods*, vol. 33, no. 12, pp. 3069-3093, 2004, https://doi.org/10.1081/STA-200039051.

[11] I. Elbatal, N. Alotaibi, E. M. Almetwally, S. A. Alyami, and M. Elgarhy, "On odd perks-G class of distributions: properties, regression model, discretization, Bayesian and non-Bayesian estimation, and applications," *Symmetry*, vol. 14, no. 5, p. 883, 2022, https://doi.org/10.3390/sym14050883.

[12] M. S. Eliwa, M. El-Morshedy, and H. M. Yousof, "A Discrete Exponential Generalized-G Family of Distributions: Properties with Bayesian and Non-Bayesian Estimators to Model Medical, Engineering and Agriculture Data," *Mathematics*, vol. 10, no. 18, p. 3348, 2022, https://doi.org/10.3390/math10183348.

[13] M. El-Morshedy, M. Eliwa, and A. Tyagi, "A discrete analogue of odd Weibull-G family of distributions: Properties, classical and Bayesian estimation with applications to count data," *Journal of Applied Statistics*, vol. 49, no. 11, pp. 2928-2952, 2022, https://doi.org/10.1080/02664763.2021.1928018.

[14] A. R. Lubis and M. Lubis, "Optimization of distance formula in K-Nearest Neighbor method," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 1, pp. 326-338, 2020, https://doi.org/10.11591/eei.v9i1.1464.

[15] R. K. Halder *et al*., "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *Journal of Big Data*, vol. 11, no. 1, p. 113, 2024, https://doi.org/10.1186/s40537-024-00973-y.

[16] T. Zhang, "Exact MLE for Generalized Linear Mixed Models," *arXiv*, 2024, https://doi.org/10.48550/arXiv.2410.08492.

[17] I. A. Mageed and Q. Zhang, "Formalism of the Rényian Maximum Entropy (RMF) of the Stable M/G/1 queue with Geometric Mean (GeoM) and Shifted Geometric Mean (SGeoM) Constraints with Potential GeoM Applications to Wireless Sensor Networks (WSNs)," *Electronic Journal of Computer Science and Information Technology*, vol. 9, no. 1, pp. 31-40, 2023, https://doi.org/10.52650/ejcsit.v9i1.143.

[18] A. Molina, S. Natarajan, and K. Kersting, "Poisson sum-product networks: A deep architecture for tractable multivariate poisson distributions," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, pp. 2357-2363, 2017, https://doi.org/10.1609/aaai.v31i1.10844.

[19] A. Balubaid, H. Klakattawi, and D. Alsulami, "On the Discretization of the Weibull-G Family of Distributions: Properties, Parameter Estimates, and Applications of a New Discrete Distribution," *Symmetry*, vol. 16, no. 11, p. 1519, 2024, https://doi.org/10.3390/sym16111519.

[20] B. S. Siepe, F. Bartoš, T. P. Morris, A. L. Boulesteix, D. W. Heck, and S. Pawel, "Simulation studies for methodological research in psychology: A standardized template for planning, preregistration, and reporting," *Psychological Methods*, 2024, https://psycnet.apa.org/doi/10.1037/met0000695.

[21] E. M. Almetwally, D. A. Abdo, E. H. Hafez, T. M. Jawa, N. Sayed-Ahmed, and H. M. Almongy, "The new discrete distribution with application to COVID-19 Data," *Results in Physics*, vol. 32, p. 104987, 2022, https://doi.org/10.1016/j.rinp.2021.104987.

[22] A. Hussain, Y. Oraibi, Z. Mashikhin, A. Jameel, M. Tashtoush, and E. A. Az-Zo'Bi, "New software reliability growth model: Piratical swarm optimization-based parameter estimation in environments with uncertainty and dependent failures," *Statistics, Optimization & Information Computing*, vol. 13, no. 1, pp. 209-221, 2025, https://doi.org/10.19139/soic-2310-5070-2109.

[23] M. Nassar, D. Kumar, S. Dey, G. M. Cordeiro, and A. Z. Afify, "The Marshall–Olkin alpha power family of distributions with applications," *Journal of Computational and Applied Mathematics*, vol. 351, pp. 41-53, 2019, https://doi.org/10.1016/j.cam.2018.10.052.

[24] E. Az-Zo'bi *et al*., "Novel topological, non-topological, and more solitons of the generalized cubic p-system describing isothermal flux," *Optical and Quantum Electronics*, vol. 56, no. 1, p. 84, 2024, https://doi.org/10.1007/s11082-023-05642-7.

[25] H. Zureigat, M. A. Tashtoush, A. F. A. Jassar, E. A. Az-Zo'bi, and M. W. Alomari, "A solution of the complex fuzzy heat equation in terms of complex Dirichlet conditions using a modified Crank–Nicolson method," *Advances in Mathematical Physics*, vol. 2023, no. 1, p. 6505227, 2023, https://doi.org/10.1155/2023/6505227.

[26] M. Li *et al*., "Quantile regression using random forest proximities," *Proceedings of the 5th ACM International Conference on AI in Finance*, pp. 728-736, 2024, https://doi.org/10.1145/3677052.3698632.

[27] F. B. Hüttel, F. Rodrigues, and F. C. Pereira, "Deep evidential learning for Bayesian quantile regression," *arXiv*, 2023, https://doi.org/10.48550/arXiv.2308.10650.

[28] D. Wang and J. Wang, "A new statistical framework: Exploring its identifiability characteristics and applications within music therapy and engineering," *Alexandria Engineering Journal*, vol. 117, pp. 53-65, 2025, https://doi.org/10.1016/j.aej.2025.01.001.

[29] P. Theodossiou, "Skewed generalized error distribution of financial assets and option pricing," *Multinational Finance Journal*, vol. 19, no. 4, pp. 223-266, 2015, https://dx.doi.org/10.2139/ssrn.219679.

[30] S. Chupradit *et al*., "A Multi-Objective Mathematical Model for the Population-Base Transportation Network Planning," *Industrial Engineering and Management Systems*, vol. 21, no. 2, pp. 322-331, 2022, https://doi.org/10.7232/iems.2022.21.2.322.

[31] N. Shirawia, A. Kherd, S. Bamsaoud, M. Tashtoush, A. Jassar, and E. Az-Zo'bi, "Dejdumrong collocation approach and operational matrix for a class of second-order delay IVPs: Error analysis and applications," *WSEAS Transactions on Mathematics*, vol. 23, pp. 467-479, 2024, https://doi.org/10.37394/23206.2024.23.49.

[32] A. S. Hussain, Z. Z. Mashikhin, N. O. Yaseen, M. Abu Hammad, E. A. Az-Zo'Bi and M. A. Tashtoush, "Hybrid Deep Learning Framework for Estimating the Occurrence Rate of the Inverse Exponential Process: A Comparative Analysis of LSTM, ANN, and Traditional Methods," *2025 12th International Conference on Information Technology (ICIT)*, pp. 181-187, 2025, https://doi.org/10.1109/ICIT64950.2025.11049273.

[33] M. S. Sulaiman, A. Sufyan Hussain, M. A. Hammad, M. A. Tashtoush, E. A. Az-Zo'Bi and R. Noaman Abbas, "Parameters Estimation in a Non-Homogeneous Poisson Process with Fuzzy Data Through a

Modified Fuzzy Maximum Likelihood Approach Using Artificial Intelligence," *2025 12th International Conference on Information Technology (ICIT)*, pp. 188-195, 2025, https://doi.org/10.1109/ICIT64950.2025.11049166.

[34] S. Mohsin, M. S. Sulaiman, O. B. Shukur, A. S. Hussain, and M. A. Tashtoush, "Novel Logistic Extreme Value Distribution: Properties, applications, and parameter estimation using classical and machine learning methods," *Mathematical Modelling of Engineering Problems*, vol. 12, no. 6, pp. 1925-1938, 2025, https://doi.org/10.18280/mmep.120609.

[35] Z. Guan and Y. Wang, "Non-parametric construction of site-specific non-Gaussian multivariate joint probability distribution from sparse measurements," *Structural Safety*, vol. 91, p. 102077, 2021, https://doi.org/10.1016/j.strusafe.2021.102077.

[36] S. H. Jiang, J. Huang, D. V. Griffiths, and Z. P. Deng, "Advances in reliability and risk analyses of slopes in spatially variable soils: A state-of-the-art review," *Computers and Geotechnics*, vol. 141, p. 104498, 2022, https://doi.org/10.1016/j.compgeo.2021.104498.

[37] A. S. Hussain, S. Mohsin, H. J. Alsaedi, and M. A. Tashtoush, "Parameter Estimation in Linear Rate Process: Modified Maximum Likelihood Estimation Using Particle Swarm Optimization, Least Square Estimation, and Simulation Methods," *International Journal of Advances in Soft Computing and its Applications*, vol. 17, no. 2, pp. 354-377, 2025, https://doi.org/10.15849/IJASCA.250730.19.

[38] M. Sumair, T. Aized, M. M. A. Bhutta, F. A. Siddiqui, L. Tehreem, and A. Chaudhry, "RETRACTED: Method of Four Moments Mixture-A new approach for parametric estimation of Weibull Probability Distribution for wind potential estimation applications," *Renewable Energy*, vol. 191, pp. 291-304, 2022, https://doi.org/10.1016/j.renene.2022.04.054.

[39] N. Natarajan, M. Vasudevan, and S. Rehman, "Evaluation of suitability of wind speed probability distribution models: a case study from Tamil Nadu, India," *Environmental Science and Pollution Research*, vol. 29, no. 57, pp. 85855-85868, 2022, https://doi.org/10.1007/s11356-021-14315-5.

[40] Z. Ahmad, E. Mahmoudi, and S. Dey, "A new family of heavy tailed distributions with an application to the heavy tailed insurance loss data," *Communications in Statistics - Simulation and Computation*, vol. 51, no. 8, pp. 4372-4395, 2022, https://doi.org/10.1080/03610918.2020.1741623.

[41] Z. Ahmad, E. Mahmoudi, G. G. Hamedani, and O. Kharazmi, "On Modeling Heavy Tailed Medical Care Insurance Data via a New Member of T-X Family," *Filomat*, vol. 36, no. 6, pp. 1971-1989, 2022, https://doi.org/10.2298/FIL2206971A.

[42] H. M. Alshanbari, H. Al-Mofleh, J. T. Seong, and S. K. Khosa, "A new tangent-generated probabilistic approach with symmetrical and asymmetrical natures: Monte Carlo simulation with reliability applications," *Symmetry*, vol. 15, no. 11, p. 2066, 2023, https://doi.org/10.3390/sym15112066.

[43] B. Alnssyan, Z. Ahmad, J. C. Malela-Majika, J. T. Seong, and W. Shafik, "On the identifiability and statistical features of a new distributional approach with reliability applications," *AIP Advances*, vol. 13, no. 12, p. 125109, 2023, https://doi.org/10.1063/5.0178555.

[44] A. S. Hussain, N. Q. Saadoon, A. S. Abdulghafour, and N. T. Abduirazzaq, "Parameters estimation of a suggested model non-homogeneous poisson process: A comparison of conventional and artificially intelligent approaches," *AIP Conference Proceedings*, vol. 3282, no. 1, p. 020023, 2025, https://doi.org/10.1063/5.0264840.

[45] E. Az-Zobi *et al*., "Chaotic, Bifurcation, Sensitivity, Modulation Stability Analysis and Optical Soliton Structure to the Nonlinear Coupled Konno-Oono System in Magnetic Field," *AIP Advances*, vol. 15, no. 9, p. 095103, 2025, https://doi.org/10.1063/5.0291009.

[46] A. Hussain, K. Mahmood, I. Ibrahim, A. Jameel, S. Nawaz, and M. Tashtoush, "Parameters estimation of the Gompertz-Makeham process in non-homogeneous Poisson processes: Using modified maximum likelihood estimation and artificial intelligence methods," *Mathematics and Statistics*, vol. 13, no. 1, pp. 1-11, 2025, https://doi.org/10.13189/ms.2025.130101.

[47] Z. Shen, A. Alrumayh, Z. Ahmad, R. Abu-Shanab, M. Al-Mutairi, and R. Aldallal, "A new generalized Rayleigh distribution with analysis to big data of an online community," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 11523-11535, 2022, https://doi.org/10.1016/j.aej.2022.05.010.

[48] H. R. Hussein, S. R. Hussein, A. S. Hussain, and M. A. Tashtoush, "Estimating parameters of Software Reliability Growth Models using Artificial Neural Networks optimized by the Artificial Bee Colony algorithm based on a novel NHPP," *Mathematical Modelling of Engineering Problems*, vol. 12, no. 1, pp. 25-36, 2025, https://doi.org/10.18280/mmep.120104.

[49] M. A. Alomair, Z. Ahmad, G. S. Rao, H. Al-Mofleh, S. K. Khosa, and A. S. Al Naim, "A new trigonometric modification of the Weibull distribution: Control chart and applications in quality control," *PLOS ONE*, vol. 18, no. 1, p. e0286593, 2023, https://doi.org/10.1371/journal.pone.0286593.

[50] H. S. Alruhaili, A. S. Hussain, A. M. S. Ajlouni, F. Türk, E. A. Az-Zo'bi, M. Tashtoush, "Solving Time-Fractional Nonlinear Variable-Order Delay PDEs Using Feedforward Neural Networks," *Iraqi Journal for Computer Science and Mathematics*, vol. 6, no. 3, pp. 171-184, 2025, https://doi.org/10.52866/2788-7421.1284.

[51] A. Hussain, Y. Oraibi, Z. Mashikhin, A. Jameel, M. Tashtoush, and E. A. Az-Zo'Bi, "New software reliability growth model: Piratical swarm optimization-based parameter estimation in environments with uncertainty and dependent failures," *Statistics, Optimization & Information Computing*, vol. 13, no. 1, pp. 209-221, 2025, https://doi.org/10.19139/soic-2310-5070-2109.

[52] G. Lu, S. Salem, Z. Ahmad, A. M. Alsahangiti, M. Yusuf, and M. E. Bakr, "A new tangent-based probabilistic approach with applications in sports and medical sciences," *Alexandria Engineering Journal*, vol. 81, pp. 560-579, 2023, https://doi.org/10.1016/j.aej.2023.09.003.

[53] D. Basalamah and B. Alruwaili, "The weighted Lindley exponential distribution and its related properties," *AIMS Mathematics*, vol. 8, no. 10, pp. 24984-24998, 2023, https://doi.org/10.3934/math.20231275.

[54] A. Gul, A. J. Sandhu, M. Farooq, M. Adil, Y. Hassan, and F. Khan, "Half logistic-truncated exponential distribution: Characteristics and applications," *PLOS ONE*, vol. 18, no. 11, p. e0285992, 2023, https://doi.org/10.1371/journal.pone.0285992.

[55] R. Alotaibi, E. M. Almetwally, I. Ghosh, and H. Rezk, "Statistical inference of a step-stress model with competing risks under time censoring for alpha power exponential distribution," *Journal of Radiation Research and Applied Sciences*, vol. 17, no. 1, p. 100771, 2024, https://doi.org/10.1016/j.jrras.2023.100771.

[56] H. M. Alshanbari, O. H. Odhah, H. Al-Mofleh, Z. Ahmad, S. K. Khosa, and A. A. A. H. El-Bagoury, "A new flexible Weibull extension model: Different estimation methods and modeling an extreme value data," *Heliyon*, vol. 9, no. 11, p. e22165, 2023, https://doi.org/10.1016/j.heliyon.2023.e21704.

[57] T. Thanh Thach and R. Briš, "An additive Chen-Weibull distribution and its applications in reliability modeling," *Quality and Reliability Engineering International*, vol. 37, no. 1, pp. 352-373, 2021, https://doi.org/10.1002/qre.2740.

[58] H. M. Alshanbari, O. H. Odhah, Z. Ahmad, F. Khan, and A. A. A. H. El-Bagoury, "A new probability distribution: model, theory and analyzing the recovery time data," *Axioms*, vol. 12, no. 1, p. 477, 2023, https://doi.org/10.3390/axioms12050477.

[59] A. R. Kamel and A. A. Alqarni, "A New Approach for Model Selection with Two Qualitative Regressors," *Computational Journal of Mathematical and Statistical Sciences*, vol. 1, no. 1, pp. 63-79, 2022, https://cjmss.journals.ekb.eg/article_274302.html.

[60] H. S. Klakattawi, "Survival analysis of cancer patients using a new extended Weibull distribution," *PLoS ONE*, vol. 17, no. 2, p. e0264229, 2022, https://doi.org/10.1371/journal.pone.0264229.