Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/140305

From Theory to Practice: The Ristić-Balakrishnan-Harris-G Family of Distributions and Its Applications

Shusen Pu[1,†,∗], *Thatayaone Moakofi*[2,†], *Broderick Oluyede*[2] *and Achraf Cohen*[1]

[1]Department of Mathematics and Statistics, University of West Florida, Pensacola, FL, USA
[2]Department of Mathematics and Statistical Sciences, Botswana International University of Science and Technology, Palapye, Botswana
[†] These authors contributed equally to this work.

Received: 14 Jul. 2024, Revised: 09 Feb. 2025, Accepted: 10 Mar. 2025
Published online: 1 May 2025

Abstract: This research introduces the Ristić-Balakrishnan-Harris-G (RB-Harris-G) family of distributions, a novel framework designed to address critical gaps in survival analysis and reliability modeling. The paper systematically explores the mathematical foundations and statistical properties of this family, emphasizing its capacity to model complex hazard rate behaviors such as bathtub-shaped, increasing, and decreasing failure patterns. Key innovations include a reparameterized quantile function for enhanced simulation capabilities and comprehensive analyses of reliability metrics, Rényi entropy, and order statistics. Parameters are estimated via maximum likelihood estimation, with rigorous simulations confirming estimator consistency. The RB-Harris-Weibull subfamily demonstrates exceptional flexibility, outperforming established models across four diverse datasets (silicon nitride fracture toughness, unemployment insurance, COVID-19 survival, and chemotherapy outcomes) through advanced goodness-of-fit tests. Visual diagnostics, including Kaplan-Meier curves and hazard rate plots, underscore the model's alignment with empirical trends. These advancements provide researchers with robust tools for medical and engineering applications while establishing a foundation for future extensions in time-dependent covariate modeling.

Keywords: Generalized Distributions, Survival Analysis, Hazard Rate Flexibility, Quantile Reparameterization, Monte Carlo Simulations

1 Introduction

Modern statistical challenges demand distributions capable of capturing complex data behaviors inherent in high-dimensional, skewed, or censored datasets. Data distribution and fitting are essential for extracting actionable insights in fields such as reliability engineering and biomedical research. While classical probability models underpin traditional analyses, their rigidity often fails to accommodate non-monotonic hazard rates or heavy-tailed phenomena [1]. This limitation has spurred the development of generalized families, ranging from transformation-based frameworks (e.g., Marshall-Olkin [2]) to exponentiated generators [3]. Recent innovations include the modified Burr III Odds Ratio-G [4], Ristić–Balakrishnan-Topp-Leone-Gompertz-G [5], and alpha power Rayleigh-G [6], among others [7,8,9,10,11,12].

Among these, the Harris-G family [13] stands out for its ability to unify skewness and tail-weight adjustments through its cumulative distribution function (CDF):

$$F_{Harris-G}(x;\nu,\theta,\psi) = 1 - \left[\frac{\theta\bar{B}^{\nu}(x;\Phi)}{1-\bar{\theta}\bar{B}^{\nu}(x;\Phi)}\right]^{\frac{1}{\nu}}, \tag{1}$$

[∗] Corresponding author e-mail: spu@uwf.edu and probability density function (PDF):

$$f_{Harris-G}(x;\nu,\theta,\psi)=\frac{\theta^{\frac{1}{\nu}}g(x;\psi)}{\left(1-\bar{\theta}\bar{B}^{\nu}(x;\Phi)\right)^{1+\frac{1}{\nu}}},\tag{2}$$

where $\nu,\theta>0$, $\bar{\theta}=1-\theta$, and ψ parameterizes the baseline distribution. When $\nu=1$, this reduces to the Marshall-Olkin-G family, illustrating its nested flexibility.

Building on Ristić and Balakrishnan's gamma generator [14], our work extends the Harris-G framework by integrating a shape parameter δ, enabling richer hazard rate dynamics. Prior applications of the gamma generator include heavy-tailed models like the gamma log-logistic Weibull [15] and bathtub hazard formulations such as the gamma exponentiated exponential-Weibull [16] and others [17,18,19]. The RB-Harris-G family synthesizes these advances, supporting reversed-J, symmetric, and skewed densities alongside monotonic/non-monotonic hazard rates.

The necessity of this research stems from critical gaps in contemporary statistical modeling, particularly in survival analysis and reliability engineering. Traditional distributions often impose restrictive assumptions—such as constant or monotonically increasing hazard rates—that poorly align with real-world phenomena. For instance, mechanical systems frequently exhibit "bathtub-shaped" failure patterns (high initial risk, stabilization, then wear-out phases), while medical data, such as post-treatment survival times, may show non-monotonic risks influenced by heterogeneous patient responses. Existing models, including standard Weibull or log-logistic distributions, lack the parametric flexibility to capture these dynamics, leading to biased estimates and unreliable predictions. This limitation becomes acute in high-stakes applications: underestimating early failure risks in engineering components could precipitate catastrophic system failures, while oversimplified medical hazard models might obscure critical intervention windows.

The RB-Harris-G family directly addresses these challenges by unifying adaptable hazard structures with interpretable parameterization. Unlike earlier generalizations that prioritize mathematical tractability over applicability, our framework integrates the gamma generator's shape flexibility [14] with the Harris-G's skewness-tail modulation [13]. This synergy enables explicit modeling of multimodal hazards and heavy-tailed phenomena without sacrificing closed-form quantile expressions—a dual advantage absent in predecessors like the Marshall-Olkin-G or exponentiated Weibull families. Furthermore, the proliferation of complex datasets (e.g., censored biomedical records, multi-modal sensor data) demands distributions that balance computational efficiency with empirical fidelity. By embedding these capabilities, the RB-Harris-G family fills a methodological void, offering practitioners a versatile tool for modern data challenges.

The paper proceeds as follows: in Section 2, we formally introduce the RB-Harris-G family, and Section 3 explores several of its special cases. Section 4 presents the expansions of the probability density function, while Section 5 details its mathematical properties. In Section 6, Monte Carlo simulations are used to validate the estimation methods. Section 7 demonstrates the model's empirical superiority over benchmark alternatives, and Section 8 concludes with a discussion of future research directions.

2 The New Family of Distributions

The development of flexible distribution families hinges on synthesizing existing generators to enhance parametric adaptability. The gamma generator proposed by Ristić and Balakrishnan [14] operationalizes this principle through its CDF:

$$F_{RB}(x;\delta)=1-\frac{1}{\Gamma(\delta)}\int_0^{-\log(G(x))}t^{\delta-1}\mathrm{e}^{-t}\,dt,\quad \delta>0,\tag{3}$$

and corresponding PDF:

$$f_{RB}(x;\delta)=\frac{1}{\Gamma(\delta)}[-\log(G(x))]^{\delta-1}g(x),\quad x\in\mathbb{R},\quad \delta>0.\tag{4}$$

This framework generalizes baseline distributions by introducing a shape parameter δ, which modulates tail behavior and hazard rate flexibility.

To further augment modeling capabilities, we unify the gamma generator with the Harris-G family [13], renowned for its skewness-tail trade-off via parameter θ. The resultant Ristić-Balakrishnan-Harris-G (RB-Harris-G) family synthesizes these advantages, yielding the CDF:

$$F_{RB-Harris-G}(x;\delta,v,\theta,\psi) = 1 - \frac{1}{\Gamma(\delta)}\int_0^{-\log\left(1-\left[\frac{\theta\bar{B}^v(x;\Phi)}{1-\bar{\theta}\bar{B}^v(x;\Phi)}\right]^{1/v}\right)} t^{\delta-1}e^{-t}\,dt$$
$$= 1 - \frac{\gamma\left[\delta, -\log\left(1-\left[\frac{\theta\bar{B}^v(x;\Phi)}{1-\bar{\theta}\bar{B}^v(x;\Phi)}\right]^{1/v}\right)\right]}{\Gamma(\delta)}. \tag{5}$$

Here, $\bar{B}(x;\Phi) = 1 - B(x;\Phi)$ represents the survival function of the baseline distribution $B(x;\Phi)$, while v governs tail heaviness and θ modulates skewness. The associated PDF derives as:

$$f_{RB-Harris-G}(x;\delta,v,\theta,\psi) = \frac{1}{\Gamma(\delta)}\left[-\log\left(1-\left[\frac{\theta\bar{B}^v(x;\Phi)}{1-\bar{\theta}\bar{B}^v(x;\Phi)}\right]^{1/v}\right)\right]^{\delta-1}$$
$$\times \frac{\theta^{1/v}b(x;\Phi)}{\left[1-\bar{\theta}\bar{B}^v(x;\Phi)\right]^{1+1/v}} \tag{6}$$

for $\delta, v, \theta > 0$, where $\bar{\theta} = 1-\theta$. This formulation extends baseline densities $b(x;\Phi)$ through four interpretable parameters, enabling multi-modal and heavy-tailed behaviors.

2.1 Hazard Rate and Quantile Functions

In reliability analysis, the hazard rate function provides critical insights into time-dependent failure risks. For the RB-Harris-G family, this function is analytically tractable:

$$h_{RB-Harris-G}(x;\delta,v,\theta,\psi) = -\log\left(1-\left[\frac{\theta\bar{B}^v(x;\Phi)}{1-\bar{\theta}\bar{B}^v(x;\Phi)}\right]^{1/v}\right)^{\delta-1}\left(\frac{\theta^{1/v}b(x;\Phi)}{[1-\bar{\theta}\bar{B}^v(x;\Phi)]^{1+1/v}}\right)$$
$$\times\left(\gamma\left[\delta, -\log\left(1-\left[\frac{\theta\bar{B}^v(x;\Phi)}{1-\bar{\theta}\bar{B}^v(x;\Phi)}\right]^{1/v}\right)\right]\right)^{-1}, \tag{7}$$

where δ directly influences hazard curvature—lower values induce bathtub shapes, while higher values promote monotonic trends.

Quantile functions are equally vital for simulation and risk quantification. The RB-Harris-G quantile function $Q(p)$ solves:

$$F_{RB-Harris-G}(x;\delta,v,\theta,\psi) = 1 - \frac{\gamma\left[\delta, -\log\left(1-\left[\frac{\theta\bar{B}^v(x;\Phi)}{1-\bar{\theta}\bar{B}^v(x;\Phi)}\right]^{1/v}\right)\right]}{\Gamma(\delta)} = p, \tag{8}$$

yielding the closed-form solution:

$$B(x;\Phi) = 1 - \left[1+\theta\left(1+\left[1-\exp\left(-\gamma^{-1}[(1-p)\Gamma(\delta),\delta]\right)\right]^{-v}\right)\right]^{-1/v} := q.$$

Thus, simulating RB-Harris-G variates reduces to inverting the baseline quantile $B^{-1}(q)$, a computationally efficient process.

3 Selected Sub-Models of the Extended RB-Harris-G Family

In this section, we illustrate how various probability distributions can be derived by specifying a particular baseline distribution function $B(x,\psi)$ within the RB-Harris-G framework. We present three illustrative sub-models that arise naturally from different choices of $B(x,\psi)$. Each sub-model highlights the flexibility of the RB-Harris-G construction and exhibits diverse shapes in both the probability density function (pdf) and the hazard rate function (hrf).

3.1 A Refined Harris-Log-Logistic (RHLLoG) Model

To introduce a Harris-based model connected to the log-logistic family, we start from the standard log-logistic cdf and pdf given by:

$$B(x;c) = 1-\left(1+x^c\right)^{-1}, \qquad b(x;c) = \frac{c\,x^{c-1}}{\left(1+x^c\right)^2},$$

where $c>0$ and $x>0$. By substituting these functions into the RB-Harris-G framework, we obtain the following cdf for the Refined Harris-Log-Logistic (RHLLoG) distribution:

$$F_{\text{RHLLoG}}(x;\delta,v,\theta,c) = 1 - \frac{\gamma\!\left(\delta,\,-\ln\left[1-\left(\frac{\theta(1+x^c)^{-v}}{1-\bar{\theta}\,(1+x^c)^{-v}}\right)^{\frac{1}{v}}\right]\right)}{\Gamma(\delta)}, \tag{9}$$

where $\bar{\theta}=1-\theta$. The corresponding pdf is given by

$$f_{\text{RHLLoG}}(x;\delta,v,\theta,c) = \frac{1}{\Gamma(\delta)}\left(-\ln\left[1-\left(\frac{\theta(1+x^c)^{-v}}{1-\bar{\theta}\,(1+x^c)^{-v}}\right)^{\frac{1}{v}}\right]\right)^{\delta-1} \times \frac{\theta^{\frac{1}{v}}\,c\,x^{c-1}\,(1+x^c)^{-2}}{\left(1-\bar{\theta}\,(1+x^c)^{-v}\right)^{1+\frac{1}{v}}}. \tag{10}$$

All parameters satisfy $\delta,v,\theta,c>0$. Figure 1 illustrates how the RHLLoG distribution can capture numerous shapes in its pdf, including nearly symmetric, left-skewed, and right-skewed forms. The hrf can also take several patterns, such as monotonically increasing or decreasing, bathtub, and inverted bathtub. This variety makes the RHLLoG model highly adaptable to real data scenarios requiring flexible skewness and hazard behavior.

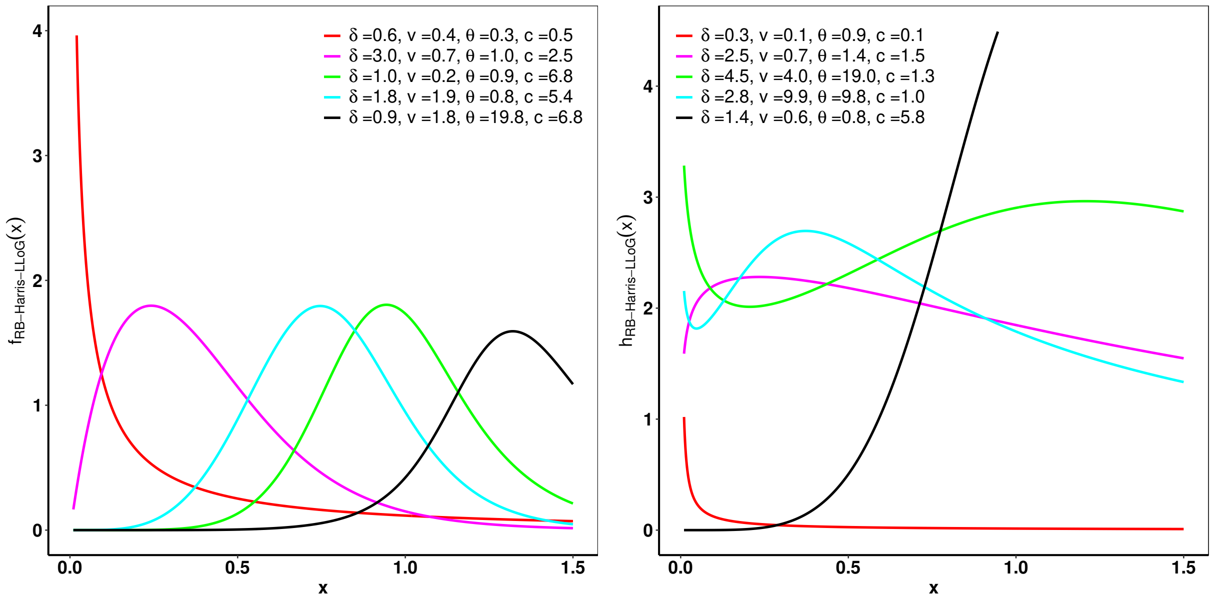

Fig. 1: (A) The pdf of the RHLLoG model for different parameter choices δ, v, θ, and c. (B) The hrf of the RHLLoG model under varying δ, v, θ, and c.

3.2 A Refined Harris-Weibull (RHW) Model

Next, we explore the scenario where the baseline distribution is the Weibull family, characterized by

$$B(x;\lambda) = 1-\exp(-x^{\lambda}), \quad b(x;\lambda) = \lambda x^{\lambda-1}\exp(-x^{\lambda}),$$

for $x > 0$ and $\lambda > 0$. Substituting these into the RB-Harris-G construction yields the cdf of the Refined Harris-Weibull (RHW) distribution:

$$F_{\mathrm{RHW}}(x;\delta,v,\theta,\lambda) = 1 - \frac{\gamma\!\left(\delta, -\ln\left[1-\left(\frac{\theta\exp(-vx^{\lambda})}{1-\bar{\theta}\exp(-vx^{\lambda})}\right)^{\frac{1}{v}}\right]\right)}{\Gamma(\delta)}, \tag{11}$$

and its pdf becomes

$$f_{\mathrm{RHW}}(x;\delta,v,\theta,\lambda) = \frac{1}{\Gamma(\delta)}\left(-\ln\left[1-\left(\frac{\theta\exp(-vx^{\lambda})}{1-\bar{\theta}\exp(-vx^{\lambda})}\right)^{\frac{1}{v}}\right]\right)^{\delta-1} \times \frac{\theta^{\frac{1}{v}}\lambda x^{\lambda-1}\exp(-x^{\lambda})}{\left(1-\bar{\theta}\exp(-vx^{\lambda})\right)^{1+\frac{1}{v}}}. \tag{12}$$

All parameters satisfy $\delta, v, \theta, \lambda > 0$. Figure 2 showcases a selection of pdf and hrf shapes attainable by the RHW model, including unimodal, reverse-J, left-skewed, and right-skewed curves for the pdf. Meanwhile, the hrf can be monotonically increasing or decreasing, or even exhibit bathtub and upside-down bathtub forms. Such flexibility is crucial when modeling lifetimes or reliability data with varying hazard patterns.

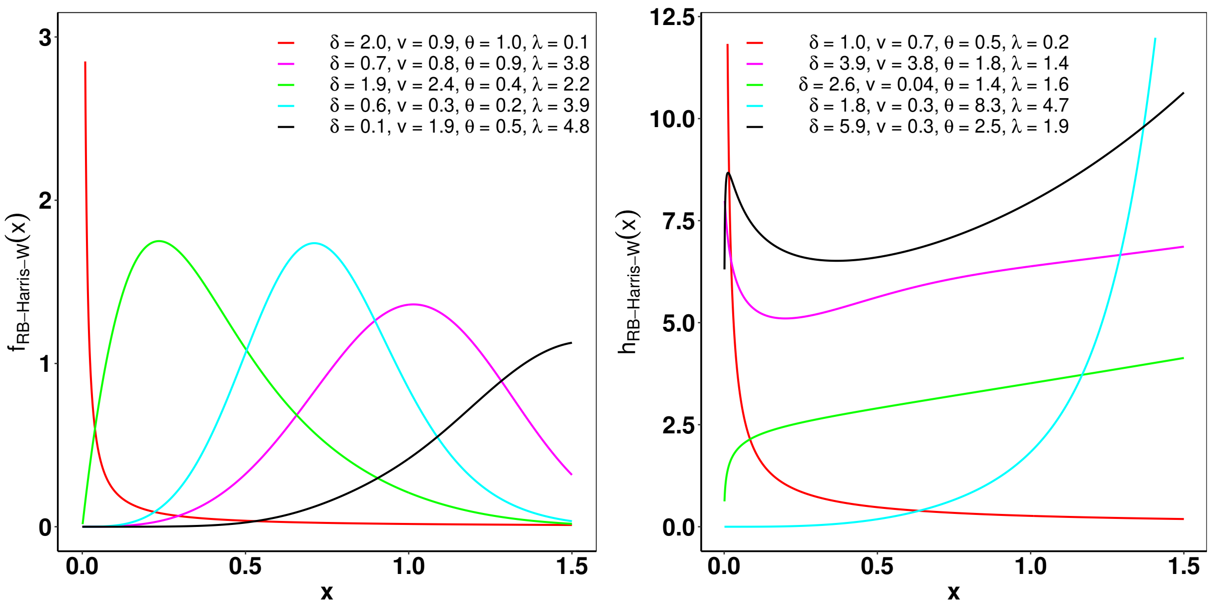

Fig. 2: (A) The pdf of the RHW model for different δ, v, θ, and λ. (B) The hrf of the RHW model for the same parameter sets.

3.3 A Refined Harris-Standard Half Logistic (RH-SHL) Model

Finally, we consider the standard half logistic distribution as the baseline. Its cdf and pdf can be expressed as

$$B(x) = \frac{1-e^{-x}}{1+e^{-x}}, \qquad b(x) = \frac{2e^{-x}}{(1+e^{-x})^2},$$

valid for $x > 0$. Inserting these into the RB-Harris-G structure gives the cdf of the Refined Harris-Standard Half Logistic (RH-SHL) distribution:

$$F_{\text{RH-SHL}}(x;\delta,v,\theta) = 1 - \frac{\gamma\left(\delta, -\ln\left[1-\left(\frac{\theta[1-(1-e^{-x})/(1+e^{-x})]^{v}}{1-\bar{\theta}[1-(1-e^{-x})/(1+e^{-x})]^{v}}\right)^{\frac{1}{v}}\right]\right)}{\Gamma(\delta)}, \tag{13}$$

and its pdf:

$$f_{\text{RH-SHL}}(x;\delta,v,\theta) = \frac{1}{\Gamma(\delta)}\left(-\ln\left[1-\left(\frac{\theta[1-(1-e^{-x})/(1+e^{-x})]^{v}}{1-\bar{\theta}[1-(1-e^{-x})/(1+e^{-x})]^{v}}\right)^{\frac{1}{v}}\right]\right)^{\delta-1}$$

$$\times \frac{\theta^{\frac{1}{v}}\left[2e^{-x}/(1+e^{-x})^{2}\right]}{\left(1-\bar{\theta}\left[1-(1-e^{-x})/(1+e^{-x})\right]^{v}\right)^{1+\frac{1}{v}}}. \tag{14}$$

Here, $\delta, v, \theta > 0$. Figure 3 illustrates how the RH-SHL model can exhibit nearly symmetric, reverse-J, left-skewed, right-skewed, and even bathtub-like shapes in the pdf. The hrf also demonstrates a broad spectrum of behaviors, including monotonically decreasing, monotonically increasing, bathtub, upside-down bathtub, and more complex patterns such as decreasing-increasing-decreasing.

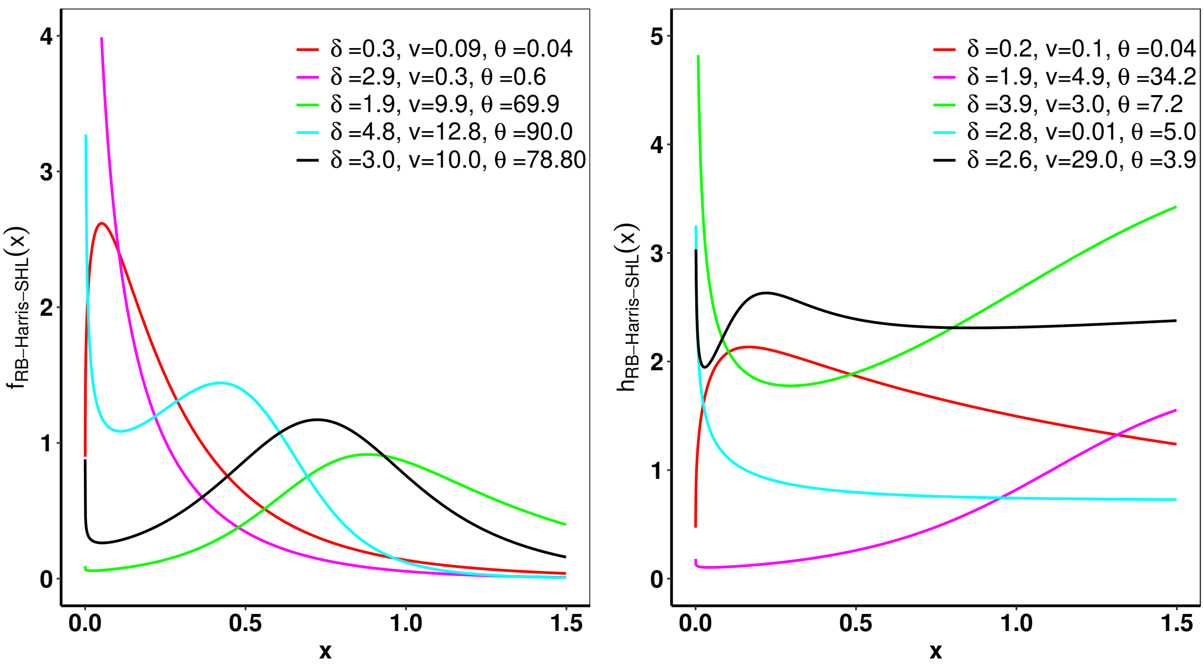

Fig. 3: (A) The pdf of the RH-SHL model for various δ, v, and θ. (B) The hrf of the RH-SHL model for the same parameter settings.

Overall, these three sub-models—RHLLoG, RHW, and RH-SHL—demonstrate the breadth of shapes and hazard patterns attainable within the extended RB-Harris-G family. By choosing an appropriate baseline distribution, researchers and practitioners can adapt the model to a wide range of data types, ensuring both flexibility and interpretability in applied statistical analyses.

4 Series Representation and Properties of the RB-Harris-G Distribution

In this section, we derive an alternative infinite-series formulation for the probability density function of the RB-Harris-G distribution, highlighting its underlying mathematical structure and potential applications in reliability analysis and statistical modeling. Such expansions are particularly useful in analytical derivations, computational implementations, and exploring theoretical properties of generalized probability distributions.

To initiate this expansion, we introduce a transformation variable:

$$z=\left[\frac{\theta\bar{B}^{v}(x;\Phi)}{1-\bar{\theta}\bar{B}^{v}(x;\Phi)}\right]^{\frac{1}{v}}, \tag{15}$$

where parameters and notation are consistent with previously established definitions.

Observing the logarithmic series expansion identity,

$$-\log(1-z)=\sum_{j=0}^{\infty}\frac{z^{j+1}}{j+1}, \tag{16}$$

and applying it to the transformed variable, we obtain:

$$[-\log(1-z)]^{\delta-1}=\left[\sum_{j=0}^{\infty}\frac{z^{j+1}}{j+1}\right]^{\delta-1} \tag{17}$$

$$=z^{\delta-1}\left(1+\sum_{j=1}^{\infty}c_j z^j\right)^{\delta-1}, \tag{18}$$

where we define coefficients $c_j=(j+1)^{-1}$ to simplify notation.

By employing the generalized binomial theorem, the above can be expanded further:

$$[-\log(1-z)]^{\delta-1}=z^{\delta-1}\sum_{m=0}^{\infty}\binom{\delta-1}{m}z^m\left(\sum_{j=0}^{\infty}c_j z^j\right)^m. \tag{19}$$

Expanding the nested power series, we define new coefficients $a_{j,m}$ as:

$$\left(\sum_{j=0}^{\infty}c_j z^j\right)^m=\sum_{j=0}^{\infty}a_{j,m}z^j, \tag{20}$$

with coefficients $a_{j,m}$ computed recursively:

$$a_{0,m}=c_0^m, \tag{21}$$

$$a_{j,m}=\frac{1}{jc_0}\sum_{l=1}^{j}[m(l+1)-j]\,c_l a_{j-l,m}. \tag{22}$$

The details of these recursive computations can be referenced from standard mathematical sources such as [20,21].

Thus, the pdf of the RB-Harris-G distribution originally defined by equation (6) admits a series representation given by:

$$f_{RB-Harris-G}(x;\delta,v,\theta,\psi)=\frac{b(x;\Phi)}{\Gamma(\delta)}\sum_{j,m=0}^{\infty}a_{j,m}\left[\frac{\theta\bar{B}^{v}(x;\Phi)}{1-\bar{\theta}\bar{B}^{v}(x;\Phi)}\right]^{j+m+\delta-1} \tag{23}$$

$$\times\frac{\theta^{1/v}}{[1-\bar{\theta}\bar{B}^{v}(x;\Phi)]^{1+1/v}}. \tag{24}$$

We further simplify the above expression by defining the general weights W_k as follows:

$$W_{k+1}=\frac{1}{\Gamma(\delta)}\sum_{j,m,p=0}^{\infty}\binom{\frac{v(-m-j-\delta)-1}{v}}{p}\binom{v(m+j+\delta+p-1)}{k}a_{j,m} \tag{25}$$

$$\times\frac{\theta^{\frac{v(m+j+\delta-1)+1}{v}}\bar{\theta}^{p}(-1)^{p+k}}{k+1}, \tag{26}$$

yielding a concise expression for the pdf as a mixture of simpler distributions:

$$f_{RB-Harris-G}(x;\delta,\nu,\theta,\psi)=\sum_{k=0}^{\infty}W_{k+1}f_{Exp-G}(x;k+1,\Phi), \tag{27}$$

where $f_{Exp-G}(x;k+1,\Phi)$ is recognized as the exponentiated generalized (Exp-G) pdf characterized by its shape parameter $(k+1)$.

Consequently, the analytical properties and statistical measures of the RB-Harris-G family naturally extend from those known for the Exp-G family, thus providing a robust theoretical foundation and facilitating practical applications in diverse fields such as survival analysis, environmental statistics, and lifetime modeling.

5 Statistical and Analytical Characteristics of the RB-Harris-G Family

Having established the series representation of the RB-Harris-G pdf, we now explore various crucial statistical properties derived directly from this representation. Understanding these characteristics is essential as they offer insight into the behavior of the distribution, facilitate parameter estimation, and support real-world applications such as risk assessment in actuarial science, reliability in engineering systems, and survival analysis in medical research. The series expansion previously introduced allows us to analytically explore various statistical properties in a more tractable form.

5.1 Moments and Generating Functions

Moments are fundamental statistical descriptors that provide insights into the central tendency, dispersion, and shape characteristics of probability distributions. Consider a random variable Z_{k+1} following an Exp-G distribution with shape parameter $(k+1)$. The r^{th} moment of the RB-Harris-G distribution is then given by

$$E(X^r)=\sum_{k=0}^{\infty}\gamma_{k+1}\,E(Z_{k+1}^r), \tag{28}$$

where Z_{k+1} follows an Exp-G distribution characterized by a power parameter $(k+1)$, and the coefficients γ_{k+1} are derived from the pdf series expansion. Furthermore, the moment generating function (MGF), a vital tool for probability theory and statistical inference, can similarly be represented as:

$$M_X(t)=\sum_{k=0}^{\infty}\gamma_{k+1}\,M_{Z_{k+1}}(t), \quad t<1, \tag{29}$$

with $M_{Z_{k+1}}(t)$ denoting the mgf of the Exp-G distribution with power parameter $(k+1)$.

5.2 Conditional and Incomplete Moments

Incomplete and conditional moments are essential for capturing nuanced properties of distributions, particularly useful in reliability analysis, risk management, and economic inequalities measures. The r^{th} incomplete moment, an essential tool in applications involving censored or truncated data, is defined as

$$\varphi_r(z)=\sum_{k=0}^{\infty}\gamma_{k+1}\int_{-\infty}^{z}x^r b_{k+1}(x;\Phi)\,dx, \tag{30}$$

where $b_{k+1}(x;\Phi)$ denotes the Exp-G pdf with shape parameter $(k+1)$.

Moreover, conditional moments, vital in fields like insurance modeling and reliability studies for conditional risk predictions, can similarly be obtained as follows:

$$E(X^r|X\geq a)=\frac{\sum_{k=0}^{\infty}\gamma_{k+1}E(Z_{k+1}^r I_{Z_{k+1}\geq a})}{1-F_{RB-Harris-G}(a;\delta,\nu,\theta,\psi)}, \tag{31}$$

where the incomplete conditional expectation of Z_{k+1} is

$$E(Z_{k+1}^r I_{Z_{k+1}\geq a})=(k+1)\int_{G(z;\psi)}^{1}[Q_G(z;\psi)]^r z^{k+1}dz. \tag{32}$$

The derived incomplete and conditional moments facilitate calculating other important statistical measures like mean deviations and concentration curves, which have broad usage in economics and quality control.

5.3 Entropy Measures

Entropy metrics, specifically the Rényi entropy, have extensive applications in quantifying uncertainty and diversity within statistical and physical systems. The Rényi entropy of the RB-Harris-G distribution, which generalizes the classical Shannon entropy, is given by

$$H_R(\alpha) = \frac{1}{1-\alpha}\log\left[\int_0^\infty (f_{RB-Harris-G}(x))^\alpha\, dx\right], \quad \alpha > 0, \quad \alpha \neq 1. \tag{33}$$

By utilizing the earlier series expansion for the pdf, we rewrite the integrand and simplify to

$$H_R(\alpha) = \frac{1}{1-\alpha}\log\left[\frac{1}{\Gamma^\alpha(\delta)}\sum_{m,i,r=0}^{\infty}\beta_{i,m,r}(\alpha)\int_0^\infty g^\alpha(x;\Phi)B^r(x;\Phi)dx\right], \tag{34}$$

where the coefficients $\beta_{i,m,r}(\alpha)$ are determined through combinatorial relations derived from the original series expansion. This form demonstrates how properties of the RB-Harris-G family directly connect to those of simpler exponentiated distributions.

5.4 Order Statistics and Applications

Order statistics play a critical role in statistics, particularly in predicting extreme values and analyzing reliability of complex systems. Considering independent and identically distributed random variables $X_1, X_2, ..., X_n$ from RB-Harris-G, the pdf of the j^{th} order statistic, denoted by $X_{j:n}$, is given by

$$f_{j:n}(x) = \frac{n!}{(j-1)!(n-j)!}\sum_{p=0}^{n-j}\binom{n-j}{p}(-1)^p\,[F(x)]^{j+p-1}f(x), \tag{35}$$

with $f(x)$ and $F(x)$ as previously defined. Further expansion yields:

$$f_{j:n}(x) = \sum_{k=0}^{\infty}\eta_{k,j}eg_{k+j}(x;\Phi), \tag{36}$$

where $eg_{k+r+1}(x;\Phi)$ represents the Exp-G distribution pdf. This formulation connects order statistics of RB-Harris-G to the Exp-G distribution, enabling deeper theoretical analysis and practical computations.

5.5 Stochastic Comparisons and Their Practical Importance

Understanding the stochastic ordering of random variables is essential for comparing their relative magnitudes and has practical implications in various fields such as economics, reliability engineering, and decision theory. It allows practitioners to assess which of two scenarios or processes presents a higher risk or is more beneficial, depending on the context. In reliability analysis, stochastic orderings help in deciding which component or system exhibits better durability.

Formally, we define stochastic ordering for two random variables U and V with cumulative distribution functions (cdf) $F_U(t)$ and $F_V(t)$, respectively, and survival functions $\bar{F}_U(t) = 1 - F_U(t)$ and $\bar{F}_V(t) = 1 - F_V(t)$. The random variable U is said to be stochastically smaller than V, denoted by $U \leq_{st} V$, if for every real number t,

$$P(U > t) \leq P(V > t).$$

This stochastic order can be characterized further by examining their respective hazard functions, $h_U(t)$ and $h_V(t)$. If $h_U(t) \geq h_V(t)$ for all $t \geq 0$, we say U is smaller than V in hazard rate order, denoted as $U \leq_{hr} V$. Likewise, if the likelihood ratio $\frac{f_U(t)}{f_V(t)}$ is decreasing in t, then we write $U \leq_{lr} V$, indicating a likelihood ratio ordering. These stochastic relationships are hierarchically related as follows:

$$U \leq_{lr} V \Rightarrow U \leq_{hr} V \Rightarrow U \leq_{st} V.$$

Specifically, consider two independent random variables $X_1 \sim RBH_G(x;\delta_1,v,\theta,\psi)$ and $X_2 \sim RB\text{-}Harris\text{-}G(x;\delta_2,v,\theta,\psi)$ with pdfs expressed by equations analogous to those discussed in the previous section. For instance,

$$f_{X_1}(x) = \frac{\theta^{1/v} g(x;\Phi)}{\Gamma(\delta_1)[1-\bar{\theta}\bar{B}^v(x;\Phi)]^{1+\frac{1}{v}}} \left[-\log\left(1-\left[\frac{\theta\bar{B}^v(x;\Phi)}{1-\bar{\theta}\bar{B}^v(x;\Phi)}\right]^{\frac{1}{v}}\right)\right]^{\delta_1-1}, \tag{37}$$

and similarly for X_2. Thus, the ratio of their pdfs becomes a valuable tool for analysis:

$$\frac{f_{X_1}(x)}{f_{X_2}(x)} = \frac{\Gamma(\delta_2)}{\Gamma(\delta_1)} \left[-\log\left(1-\left[\frac{\theta\bar{B}^v(x;\Phi)}{1-\bar{\theta}\bar{B}^v(x;\Phi)}\right]^{\frac{1}{v}}\right)\right]^{\delta_1-\delta_2}. \tag{38}$$

Analyzing its derivative with respect to x gives a condition for the likelihood ratio ordering,

$$\frac{d}{dx}\left(\frac{f_{X_1}(x)}{f_{X_2}(x)}\right) = \frac{\Gamma(\delta_2)}{\Gamma(\delta_1)}(\delta_1-\delta_2)\left[-\log\left(1-y(x)\right)\right]^{\delta_1-\delta_2-1}\frac{dy/dx}{1-y}, \tag{39}$$

where y is defined similarly to previous sections, ensuring consistency in our notation.

Given the condition $\delta_1 < \delta_2$, the derivative is negative, establishing that X_1 is stochastically smaller than X_2 in terms of the likelihood ratio order. Consequently, $X_1 \leq_{lr} X_2$ holds, which in turn implies the hazard rate and general stochastic order. These analytical insights facilitate meaningful comparisons and informed decisions based on stochastic dominance criteria.

5.6 Probability-Weighted Moments (PWMs)

Probability-weighted moments (PWMs) are extensively applied in hydrology, environmental statistics, and economic risk modeling to estimate distribution parameters robustly, especially with smaller samples or censored data. The PWMs of the RB-Harris-G FoD are calculated through:

$$M_{j,l,k} = \sum_{l=0}^{c}\sum_{m,r=0}^{\infty}\tau_{l,m,r}\int_{-\infty}^{\infty} x^a e_{m+r+1}(x;\Phi)dx, \tag{40}$$

where $e_{m+r+1}(x;\Phi)$ denotes an Exp-G distribution with appropriate power parameters. This formulation provides an effective and computationally advantageous method for parameter estimation and practical statistical analyses.

6 Simulation Analysis and Parameter Estimation

Building upon the probability density function expansion introduced previously, this section demonstrates how the derived analytical form aids practical evaluation and estimation of model parameters. Accurate parameter estimation is crucial for practitioners and researchers as it directly influences the efficacy and reliability of subsequent statistical inferences, predictive modeling, and real-world decision-making processes in fields like reliability engineering, finance, and survival analysis. Monte Carlo simulations, a powerful method for evaluating estimator performance, are particularly beneficial in scenarios where analytical expressions are complex or unavailable.

In this context, we carried out an extensive Monte Carlo simulation with $N = 3000$ replications for each of the sample sizes $n = \{25, 50, 100, 200, 400, 800, 1600\}$. Each sample was generated from the RB-Harris-W model with several distinct sets of parameters, aiming to comprehensively assess the behavior and consistency of the maximum likelihood estimates (MLEs). Two fundamental performance metrics are computed: the Average Bias (ABIAS) and the Root Mean Square Error (RMSE), defined respectively as:

$$\text{ABIAS}(\hat{\alpha}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\alpha}_i-\alpha), \quad \text{and} \quad \text{RMSE}(\hat{\alpha}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{\alpha}_i-\alpha)^2}. \tag{41}$$

Tables 1 and 2 summarize the simulation outcomes, showcasing how the estimators behave under different parameter settings and increasing sample sizes. These measures are vital as they indicate estimator bias and precision, guiding the selection and applicability of statistical inference methods. From the provided simulation results, it can be observed that the estimated parameters consistently approach their true values as sample sizes increase, highlighting the unbiasedness and consistency of maximum likelihood estimators (MLEs) in the RB-Harris-G model.

As shown, the estimators exhibit decreasing bias and RMSE as sample sizes increase, validating the consistent property of MLEs. Such empirical validations support the use of RB-Harris-G distributions in practical data analysis tasks, particularly in fields requiring accurate and reliable parameter estimation like medical statistics, quality control, and financial risk management.

Table 1: Monte Carlo Simulation Results

		(0.3, 1.5, 0.3, 0.3)			(1.5, 1.5, 0.3, 0.7)		
Parameter	Sample Size	Mean	RMSE	ABIAS	Mean	RMSE	ABIAS
δ	25	1.1783	2.2125	0.8783	5.6841	9.0801	4.1841
	50	0.7882	1.0010	0.4882	3.3077	4.0001	1.8077
	100	0.5773	0.6097	0.2773	2.6341	2.9459	1.1341
	200	0.4770	0.4310	0.1770	2.1509	1.7875	0.6509
	400	0.3818	0.2571	0.0818	1.8500	1.3611	0.3500
	800	0.3460	0.1584	0.0460	1.6344	0.4849	0.1344
	1600	0.3306	0.1064	0.0306	1.5308	0.2063	0.0308
ν	25	1.3134	0.4814	-0.1865	1.2234	0.5292	-0.2765
	50	1.3367	0.4030	-0.1632	1.2645	0.4418	-0.2354
	100	1.3564	0.3167	-0.1435	1.3556	0.3180	-0.1443
	200	1.3900	0.2484	-0.1099	1.3885	0.2452	-0.1114
	400	1.4271	0.1844	-0.0728	1.4238	0.1811	-0.0761
	800	1.4525	0.1305	-0.0474	1.4556	0.0949	-0.0443
	1600	1.4770	0.0624	-0.0229	1.4760	0.0652	-0.0239
θ	25	1.0213	1.4649	0.7213	0.8995	1.0538	0.5995
	50	0.7935	1.1622	0.4935	0.7212	0.8123	0.4212
	100	0.6241	0.8294	0.3241	0.4995	0.4509	0.1995
	200	0.5089	0.5863	0.2089	0.4205	0.2863	0.1205
	400	0.4086	0.3898	0.1086	0.3637	0.1545	0.0637
	800	0.3654	0.2840	0.0654	0.3368	0.0925	0.0368
	1600	0.3347	0.2017	0.0347	0.3122	0.0397	0.0122
λ	25	0.8613	0.9156	0.5613	2.3795	2.7806	1.6795
	50	0.6266	0.6437	0.3266	1.8134	1.8731	1.1134
	100	0.5251	0.3941	0.2251	1.4993	1.3811	0.7993
	200	0.4205	0.2526	0.1205	1.2158	0.9311	0.5158
	400	0.3607	0.1383	0.0607	1.0054	0.5870	0.3054
	800	0.3244	0.1152	0.0244	0.8613	0.3572	0.1613
	1600	0.3041	0.0385	0.0041	0.7748	0.1945	0.0748

7 Real Data Applications

In this section, we apply the newly introduced RB-Harris-Weibull (RB-Harris-W) distribution to evaluate its flexibility and robustness in modeling practical data across diverse fields. The importance of deriving the probability density function (pdf) expansions in earlier sections lies in their pivotal role in facilitating analytical derivation of various statistical properties, essential for robust parameter estimation and statistical inference. Such expansions help verify properties like stochastic ordering and quantile functions, ensuring the theoretical rigor and applicability of the proposed distribution to real-world data scenarios.

We evaluate the performance of the RB-Harris-W distribution by comparing it with established distributions such as the Kumaraswamy-Weibull (KW) [22], Topp-Leone generated Weibull (TLGW) [23], gamma exponentiated Lindley-log-logistic (GELLoG) [24], gamma log-logistic Weibull (GLLoGW) [15], and alpha power Topp-Leone Weibull (APTLW) distributions [25]. These comparative distributions cover a wide range of shapes and hazard functions, which helps demonstrate the versatility of the RB-Harris-W distribution.

We use several widely recognized criteria for model selection, including the log-likelihood, Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC), and Bayesian Information Criterion (BIC).

Table 2: Monte Carlo Simulation Results

		(0.9, 0.9, 1.0, 0.9)			(0.2, 1.0, 1.0, 0.9)		
Parameter	Sample Size	Mean	RMSE	ABIAS	Mean	RMSE	ABIAS
δ	25	3.7124	4.2351	2.8124	0.8238	1.1511	0.6238
	50	2.4704	2.7542	1.5704	0.4839	0.5771	0.2839
	100	1.8248	1.8078	0.9248	0.3398	0.2575	0.1398
	200	1.4379	1.1475	0.5379	0.2774	0.1757	0.0774
	400	1.1914	0.6963	0.2914	0.2338	0.0779	0.0338
	800	1.0441	0.3871	0.1441	0.2096	0.0333	0.0096
	1600	0.9630	0.1879	0.0630	0.2030	0.0154	0.0030
v	25	2.1797	5.1331	1.2797	3.3844	6.2027	2.3844
	50	1.6300	1.7460	0.7300	2.4607	3.8629	1.4607
	100	1.4355	1.6790	0.5355	1.8605	2.6282	0.8605
	200	1.2548	1.1324	0.3548	1.3084	0.95910	0.3084
	400	1.1935	0.9679	0.2935	1.1266	0.8801	0.1266
	800	1.0978	0.7061	0.1978	1.0507	0.2525	0.0507
	1600	0.9182	0.3791	0.0182	1.0315	0.1512	0.0315
θ	25	4.1539	4.6271	3.1539	1.3292	0.7253	0.3292
	50	2.8901	3.1917	1.8901	1.2541	0.5920	0.2541
	100	2.0467	1.9837	1.0467	1.1995	0.4970	0.1995
	200	1.6452	1.3677	0.6452	1.1263	0.3700	0.1263
	400	1.3544	0.8089	0.3544	1.0871	0.2876	0.0871
	800	1.1842	0.4473	0.1842	1.0464	0.1944	0.0464
	1600	1.0804	0.2195	0.0804	1.0274	0.1058	0.0274
λ	25	1.1034	0.5891	0.2034	0.6802	0.3152	-0.2197
	50	1.0161	0.4362	0.1161	0.7455	0.2400	-0.1544
	100	0.9696	0.3371	0.0696	0.7869	0.1690	-0.1130
	200	0.9435	0.2406	0.0435	0.8318	0.1197	-0.0681
	400	0.9106	0.1585	0.0106	0.8612	0.0762	-0.0387
	800	0.8961	0.1011	-0.0038	0.8853	0.0393	-0.0146
	1600	0.8973	0.0364	-0.0027	0.8965	0.0091	-0.0034

Additionally, we employ goodness-of-fit tests such as Cramér–von Mises (W), Anderson-Darling (A), Kolmogorov-Smirnov (K-S) statistics, and their associated p-values. A distribution with lower values for these statistics and higher K-S p-values generally indicates a better fit.

Table 3: MLEs and Goodness-of-Fit Statistics for Silicon-Nitride Data

	Estimates (SE)				**Statistics**							
Model	δ	v	θ	λ	$-2\log L$	AIC	$AICC$	BIC	W^*	A^*	K-S	p-value
RB-Harris-W	0.4927 (0.1395)	0.2844 (0.0828)	57.4990 (0.0014)	1.6975 (0.1252)	336.6426	344.6426	344.9935	355.7591	0.0419	0.2767	0.0479	0.948
RB-Harris-W(δ, 1, θ, λ)	0.2499 (0.0666)	1 -	284.3733 (134.2492)	1.4985 (0.0831)	371.265	377.2651	377.4738	385.6024	0.591	3.5638	0.1167	0.078
RB-Harris-W(1, v, 1, λ)	1 -	9.9000×10^{-04} (1.9708×10^{-15})	1 -	0.5000 (0.0313)	826.545	830.545	830.6484	836.1032	0.4605	2.766	0.7434	<0.0001
RB-Harris-W(1, 1, θ, λ)	1 -	1 -	0.9174 (0.0877)	0.5219 (0.0327)	840.5279	844.5277	844.6312	850.086	0.4570	2.7459	0.7639	<0.0001
	a	b	λ	c								
KW	0.8261 (0.6252)	0.5374 (7.3712)	0.2326 (0.6203)	5.4234 (1.2277)	337.0526	345.0527	345.4035	356.1692	0.0812	0.4913	0.0685	0.631
	α	θ	λ	β								
TLGW	1.0102×10^{02} (7.1247×10^{-08})	1.2318×10^{-02} (7.8871×10^{-04})	1.5012×10^{-01} (3.6236×10^{-03})	1.6077×10^{01} (1.5918×10^{-06})	342.8601	350.8604	351.2113	361.9769	0.2004	1.1896	0.1062	0.136
	λ	c	α	δ								
GELLoG	3.7014 (0.4972)	1.0472×10^{-07} (0.0129)	0.4907 (0.5142)	15.8740 (2.7646)	353.1245	361.1235	361.4743	372.241	0.3493	2.1263	0.137	0.0220
	c	β	δ	θ								
GLLoGW	10.404 (5.6913×10^{-05})	1.2603 (5.5815×10^{-04})	0.0865 (8.2105×10^{-03})	9.2620×10^{-05} (3.0169×10^{-05})	488.6167	496.6207	496.9715	507.7371	1.9174	10.3571	0.3400	<0.0001
	θ	α	β	λ								
APTLW	0.7681 (0.3931)	60.4077 (0.0032)	3.3023 (0.7334)	0.0059 (0.0081)	337.3089	345.3089	345.6598	356.4254	0.0591	0.3984	0.0567	0.839

7.1 Silicon Nitride Data

This dataset comprises the fracture toughness measurements of silicon nitride expressed in MPa $m^{1/2}$. Silicon nitride is widely used in engineering and materials science due to its exceptional mechanical strength and thermal stability, making its reliable modeling essential for predicting failure rates and ensuring material safety and reliability.

The estimates obtained via maximum likelihood (MLEs), standard errors, and goodness-of-fit metrics are reported in Table 3. The RB-Harris-W distribution demonstrates superior fitting, as evidenced by its lowest information criteria values (e.g., AIC, BIC) and highest K-S p-value. Figures 4 and 5 further confirm the superior visual match between empirical data and RB-Harris-W predictions. The scaled TTT and hazard plots clearly indicate the reliability and robustness of the distribution in capturing complex behaviors of the fracture toughness data, particularly suitable for engineering applications involving reliability and durability analysis.

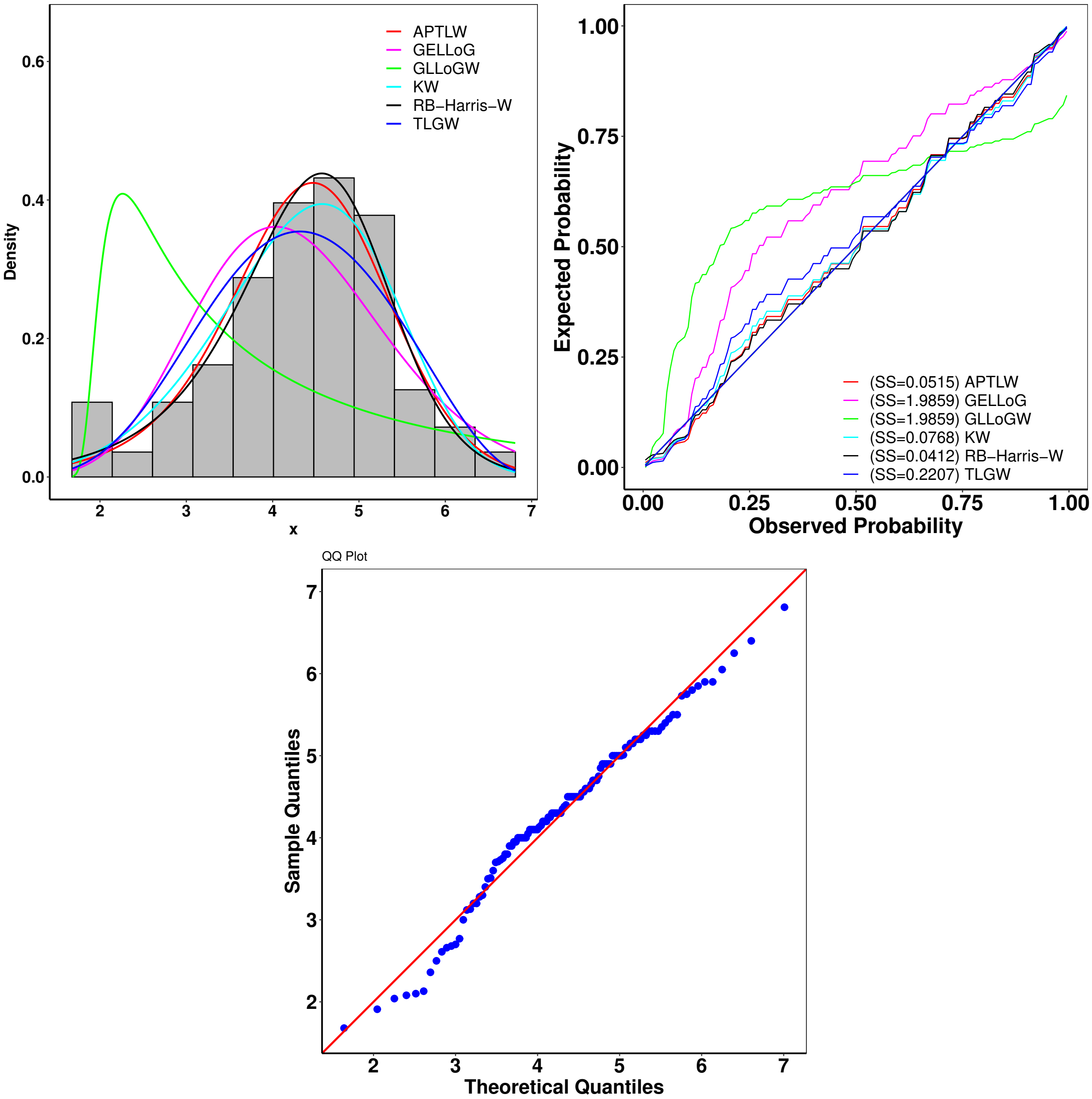

Fig. 4: Fitted density superposed on histogram (left), observed vs. expected probability plot (right), and QQ plot (bottom) for silicon nitride data.

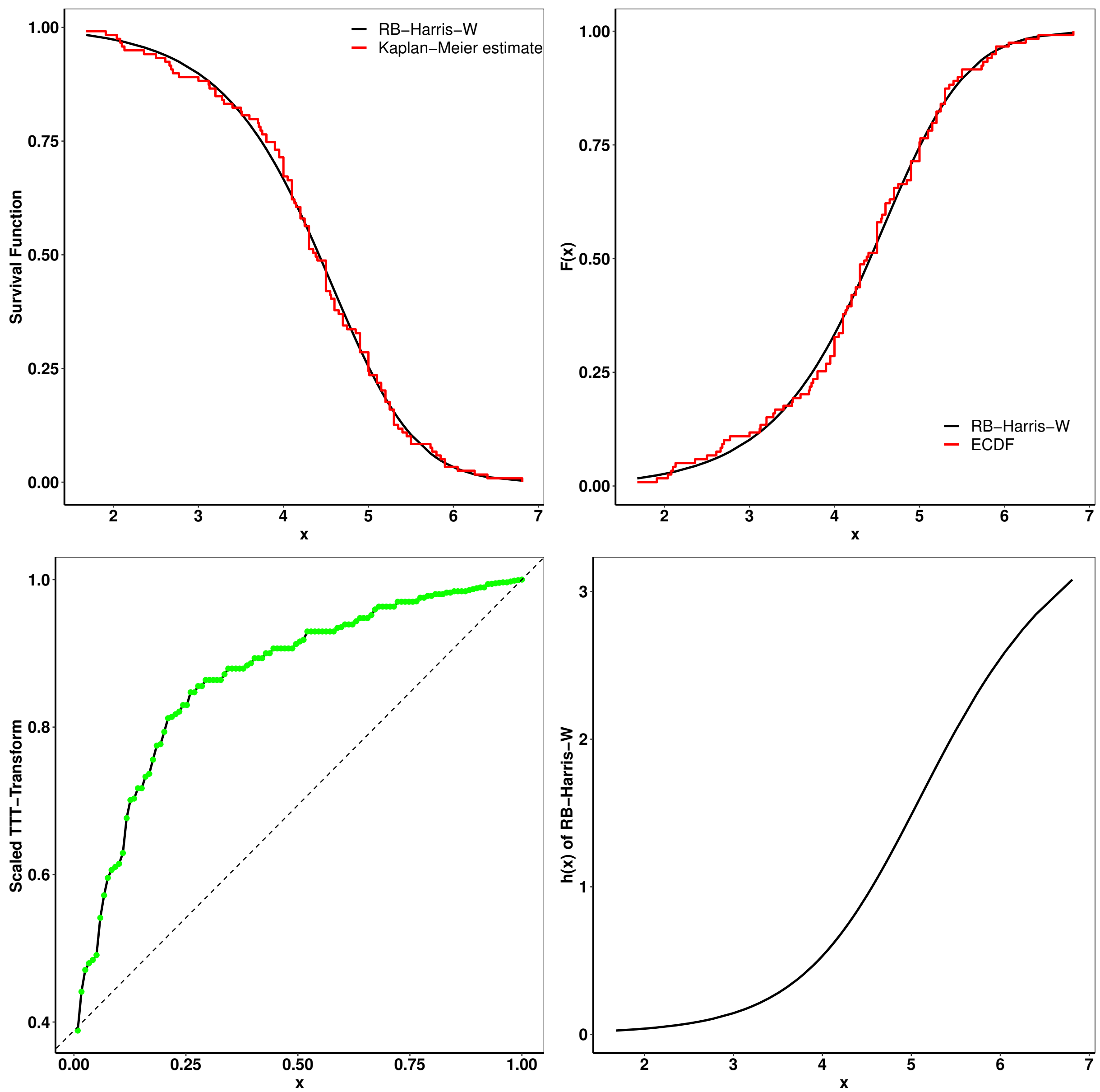

Fig. 5: Kaplan-Meier survival curve, theoretical and empirical cumulative distribution functions, scaled TTT plot, and hazard rate function for the silicon nitride data.

7.2 Insurance Data

The dataset considered here comprises monthly unemployment insurance metrics collected from July 2008 to April 2013. Analyzing such data with flexible distributions is crucial, as insurance claims data often exhibit skewness, kurtosis, and other irregularities challenging classical distributions.

MLE results, standard errors, and various goodness-of-fit statistics are presented in Table 4. Among the tested models, the RB-Harris-W distribution provides the most robust statistical fit, showing the lowest values for criteria such as AIC and BIC and a favorable K-S p-value. The graphical assessments in Figure 6, including PP and QQ plots, indicate that the RB-Harris-W distribution aligns closely with observed data patterns. Furthermore, the hazard rate function and TTT plots (Figure 7) illustrate the suitability of RB-Harris-W for modeling unemployment durations, offering practitioners a valuable tool for insurance data modeling and risk assessment.

Table 4: MLEs and Goodness-of-Fit Statistics for Insurance Data

Model	Estimates (SE)				Statistics							
	δ	v	θ	λ	$-2\log L$	AIC	$AICC$	BIC	W^*	A^*	K-S	p-value
RB-Harris-W	0.9892	0.3582	627.4900	0.6928	495.4847	503.4847	504.2394	511.7265	0.0542	0.3124	0.0802	0.849
	(0.3220)	(0.1542)	(1.8199×10^{-04})	(0.0757)								
RB-Harris-W(δ, 1, θ, λ)	0.9984	1	3.7841×10^{03}	0.5489	499.3748	505.3748	505.8193	511.5562	0.1386	0.7280	0.1000	0.608
	(0.2677)	-	(1.2177×10^{-05})	(0.0172)								
RB-Harris-W(1, v, 1, λ)	1	8.1307×10^{-04}	1	0.2500	786.2729	790.2729	790.491	794.3937	0.3548	1.8761	0.8314	<0.0001
	-	(1.2253×10^{-16})	-	(0.0178)								
RB-Harris-W(1, 1, θ, λ)	1	1	0.9781	0.2085	783.7742	787.7742	787.9924	791.8951	0.3639	1.9265	0.8015	<0.0001
	-	-	(0.1341)	(0.0182)								
	a	b	λ	c								
KW	1.9245×10^{02}	2.8145×10^{03}	1.0681×10^{03}	0.1067	499.3208	507.3205	508.0752	515.5623	0.1665	0.8469	0.1265	0.311
	(6.7663×10^{-08})	(6.0206×10^{-10})	(4.2972×10^{-09})	(4.7349×10^{-04})								
	α	θ	λ	β								
TLGW	$9.4941e\times10^{01}$	1.3352×10^{-02}	1.0763×10^{-02}	9.1307	499.5374	507.5374	508.2921	515.7792	0.1777	0.8992	0.1424	0.19
	(9.7438×10^{-08})	(1.2531×10^{-03})	(6.6421×10^{-04})	(1.8990×10^{-06})								
	λ	c	α	δ								
GELLoG	0.2317	5.2945×10^{-07}	3.9148×10^{-03}	14.8790	498.5636	506.5636	507.3184	514.8054	0.1360	0.6981	0.1230	0.3440
	(0.0761)	(9.5966×10^{-03})	(0.0234)	(9.1622)								
	c	β	δ	θ								
GLLoGW	0.8350	0.1109	4.2111	0.0024	526.2139	534.2139	534.9686	542.4557	0.5782	3.1069	0.2215	0.007
	(0.2185)	(0.1075)	(0.0243)	(0.0007)								
	θ	α	β	λ								
APTLW	0.5283	104.9700	2.1481	1.7417×10^{-04}	496.2395	504.2394	504.9942	512.4812	0.0838	0.4404	0.0966	0.651
	(0.1053)	(1.4404×10^{-04})	(0.0526)	(2.7725×10^{-05})								

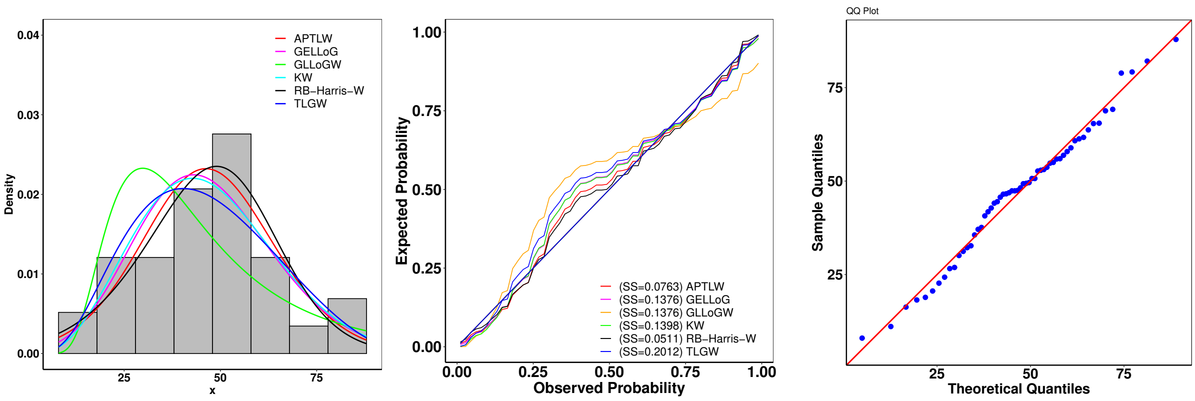

Fig. 6: Fitted densities on histogram (left), observed vs. expected probability plot (right), and QQ plot (bottom) for insurance data.

Table 5: MLEs and Goodness-of-Fit Statistics for COVID-19 Data

Model	Estimates (SE)				Statistics							
	δ	v	θ	λ	$-2\log L$	AIC	$AICC$	BIC	W^*	A^*	K-S	p-value
RB-Harris-W	0.1913	13.5243	115.3497	1.5555	375.1219	383.1219	383.518	393.7757	0.0368	0.2204	0.0543	0.913
	(0.0615)	(6.6047)	(0.2121)	(0.1903)								
RB-Harris-W(δ, 1, θ, λ)	0.2002	1	1.9553	1.5577	378.3722	384.3722	384.6075	392.3625	0.0666	0.4046	0.0754	0.583
	(0.1006)	-	(1.2441)	(0.2563)								
RB-Harris-W(1, v, 1, λ)	1	1.6200×10^{-05}	1	0.5300	591.7618	595.7618	595.8783	601.0886	0.0518	0.2863	0.5844	<0.0001
	-	(7.5720×10^{-14})	-	(0.0384)								
RB-Harris-W(1, 1, θ, λ)	1	1	0.8890	0.6679	584.8981	588.8981	589.0146	594.2249	0.0515	0.2888	0.6091	<0.0001
	-	-	(0.0909)	(0.0430)								
	a	b	λ	c								
KW	23.7652	526.5723	0.1835	2.4539	378.2814	386.2826	386.6786	396.9364	0.0704	0.4322	0.0731	0.621
	(11.2090)	(0.0294)	(0.0406)	(4.8524)								
	α	θ	λ	β								
TLGW	2.2614	1.6008	0.3910	0.9870	376.6661	384.6661	385.0621	395.3198	0.0549	0.3085	0.0693	0.687
	(5.4718)	(4.2322)	(0.3841)	(0.4763)								
	λ	c	α	δ								
GELLoG	0.9081	1.3517	1.4120	3.0399	375.7760	383.776	384.1721	394.4298	0.046	0.2583	0.0624	0.803
	(0.3356)	(1.1318)	(2.9822)	(3.2119)								
	c	β	δ	θ								
GLLoGW	11.9332	0.9698	0.0786	0.4962	417.1421	425.1427	425.5388	435.7965	0.4728	3.0108	0.1892	0.001
	(1.9599)	(0.1178)	(0.0140)	(0.1925)								
	θ	α	β	λ								
APTLW	5.8553	2.7609×10^{-05}	0.5942	0.2816	377.9029	385.903	386.299	396.5567	0.0637	0.3624	0.0673	0.721
	(4.5983×10^{-03})	(2.2908×10^{-04})	(0.0985)	(0.0475)								

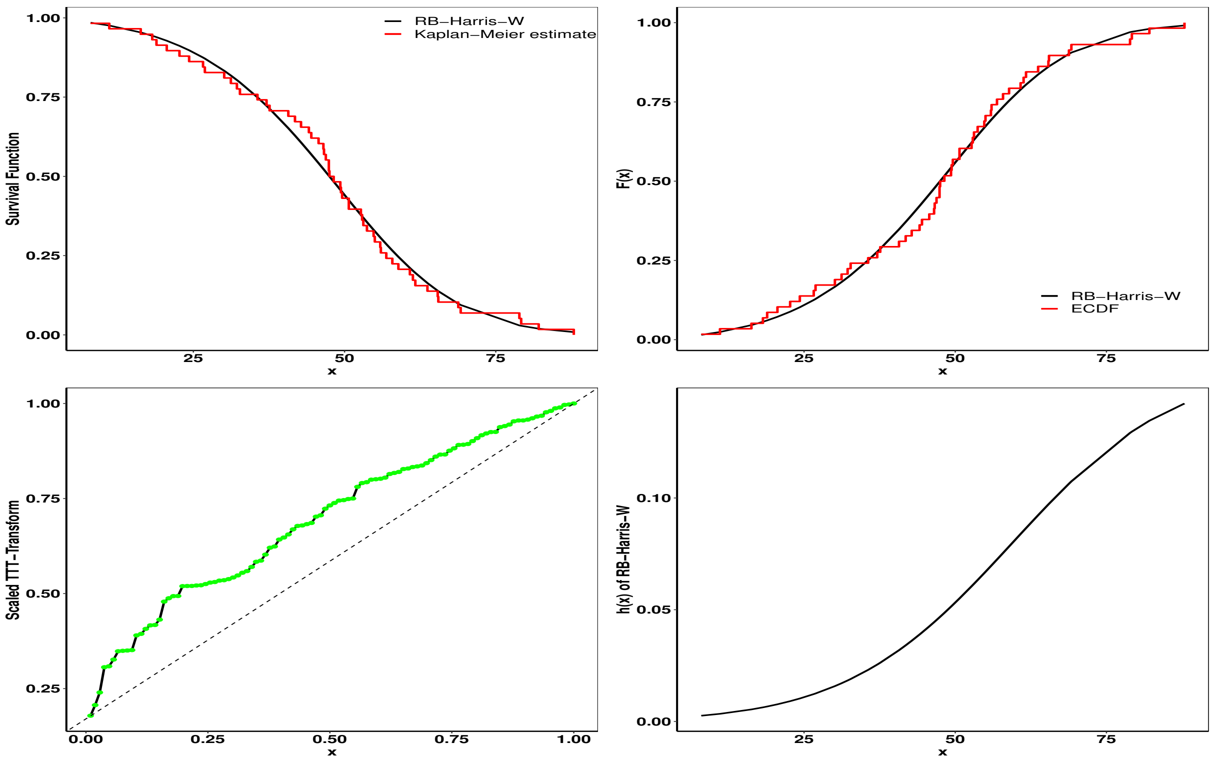

Fig. 7: Kaplan-Meier survival curve, theoretical vs empirical cdfs, TTT, and hazard rate for unemployment insurance data.

7.3 COVID-19 Data

We analyze a dataset containing COVID-19 records from Mexico, gathered during the early pandemic months of March and April 2020. The limited sample size (30 observations) underscores the necessity for robust distributional assumptions capable of accurately representing skewed and potentially heavy-tailed data typical in epidemiological studies.

Table 5 lists the parameter estimates, standard errors, and goodness-of-fit statistics. The RB-Harris-W distribution outperforms comparative distributions significantly, showing minimal information criteria and high p-value metrics, thereby validating its utility for epidemiological applications. Diagnostic plots in Figure 8, including density, QQ, and PP plots, reinforce its excellent fitting capabilities. Such modeling is critical for informed decision-making in public health, enabling accurate estimation and forecasting of disease progression.

7.4 Chemotherapy Data

We analyze survival times for patients undergoing chemotherapy, a context where accurately capturing survival probabilities and hazard rates can directly impact medical decision-making and patient prognosis. Survival data, known for being right-skewed, necessitates distributions with significant flexibility.

The results summarized in Table 6 clearly indicate that the RB-Harris-W distribution provides the best statistical performance among the distributions evaluated. It yields the lowest values of information criteria and excellent goodness-of-fit statistics, accompanied by favorable K-S p-values. Visual analysis through fitted density plots, PP plots, and QQ plots in Figure 9 reinforces the RB-Harris-W distribution's suitability. Moreover, the alignment of TTT and hazard rate functions further validates its robustness (Figure 9). This distribution is especially useful in oncology studies, facilitating precise survival modeling and improving individualized therapeutic interventions based on expected survival duration.

Overall, the results underline the RB-Harris-W distribution's versatility across various data types, showcasing its practical advantage for statistical and reliability modeling.

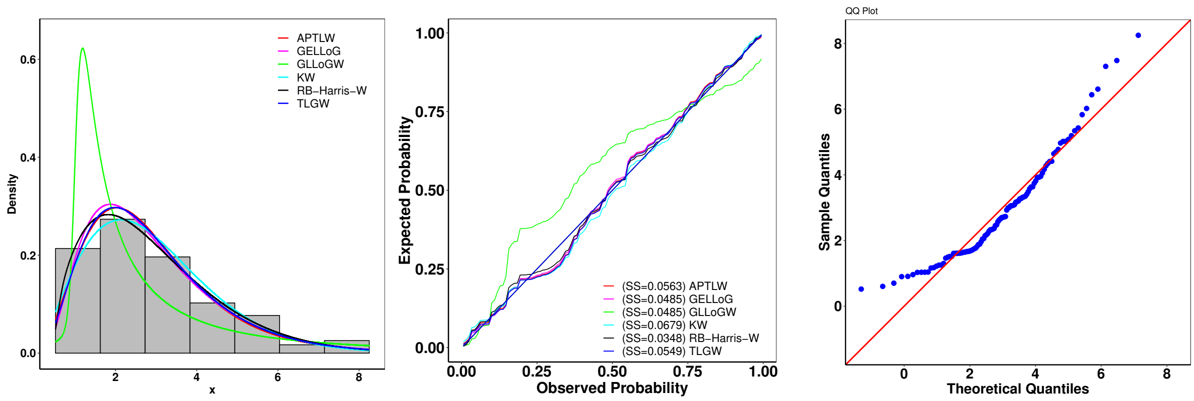

Fig. 8: Density plot over histogram (left), observed vs. expected probability plot (right), and QQ plot (bottom) for COVID-19 data.

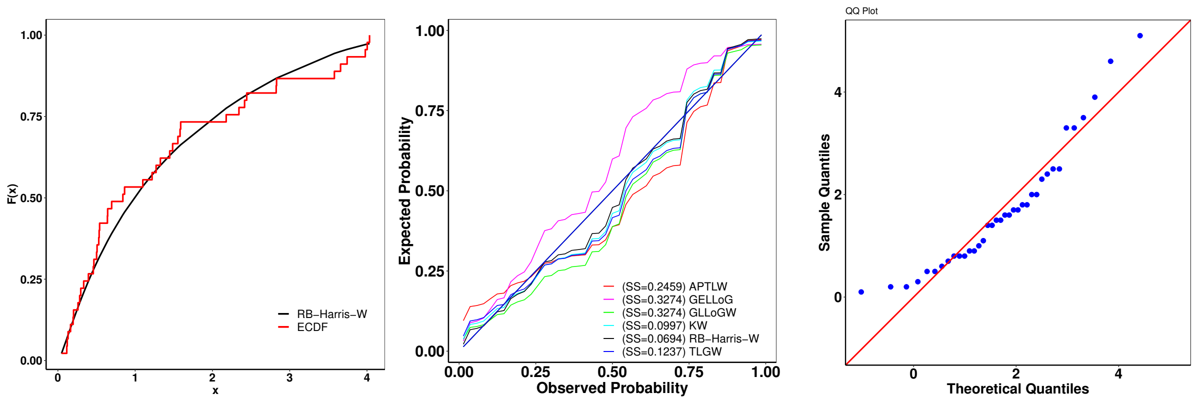

Fig. 9: Density histogram, observed vs. expected probability plots, and QQ plot for chemotherapy data.

Table 6: MLEs and Goodness-of-Fit Statistics for Chemotherapy Data

	Estimates (SE)				Statistics							
Model	δ	v	θ	λ	$-2\log L$	AIC	$AICC$	BIC	W^*	A^*	K-S	p-value
RB-Harris-W	11.6700	0.0622	8.3879×10^{04}	2.6866	110.6116	118.6116	119.6116	125.8382	0.0632	0.4341	0.1025	0.692
	(0.5363)	(0.0542)	(6.6247×10^{-06})	(6.9226)								
RB-Harris-W(δ, 1, θ, λ)	0.4500	1	0.2342	1.0670	117.7912	199.4078	199.9932	204.8278	0.0593	0.4155	0.4027	<0.0001
	(0.1118)	-	(0.1246)	(0.1605)								
RB-Harris-W(1, v, 1, λ)	1	1.0000×10^{-03}	1	0.9360	120.277	124.277	124.5627	127.8903	0.0790	0.5301	0.1443	0.278
	-	(4.1556×10^{-15})	-	(0.0945)								
RB-Harris-W(1, 1, θ, λ)	1	1	0.8392	0.9247	123.1068	127.1068	127.3926	130.7202	0.0739	0.4992	0.1790	0.098
	-	-	(0.1734)	(0.0947)								
	a	b	λ	c								
KW	0.1966	3.1492	4.5772	0.1587	115.5417	123.5417	124.5417	130.7684	0.1007	0.6650	0.1160	0.541
	(0.3522)	(4.2665)	(7.4380)	(0.1314)								
	α	θ	λ	β								
TLGW	0.8514	0.1892	0.2055	4.8706	116.036	124.036	125.036	131.2627	0.1209	0.7923	0.1246	0.451
	(0.7180)	(0.2192)	(0.0298)	(4.8497)								
	λ	c	α	δ								
GELLoG	0.0681	1.6425	1.2348	0.5769	122.8306	130.8306	131.8306	138.0573	0.0825	0.5605	0.1751	0.112
	(0.2585)	(0.2693)	(0.3246)	(0.1356)								
	c	β	δ	θ								
GLLoGW	1.6777	1.5933	0.2500	3.0495	118.5171	126.5171	127.5171	133.7438	0.1156	0.7597	0.1568	0.196
	(0.5807)	(0.2758)	(0.0720)	(1.7102)								
	θ	α	β	λ								
APTLW	0.0743	15.1412	3.6687	0.0062	124.2782	132.2782	133.2782	139.5048	0.2334	1.4805	0.1523	0.223
	(0.0704)	(0.0306)	(2.8463)	(0.0269)								

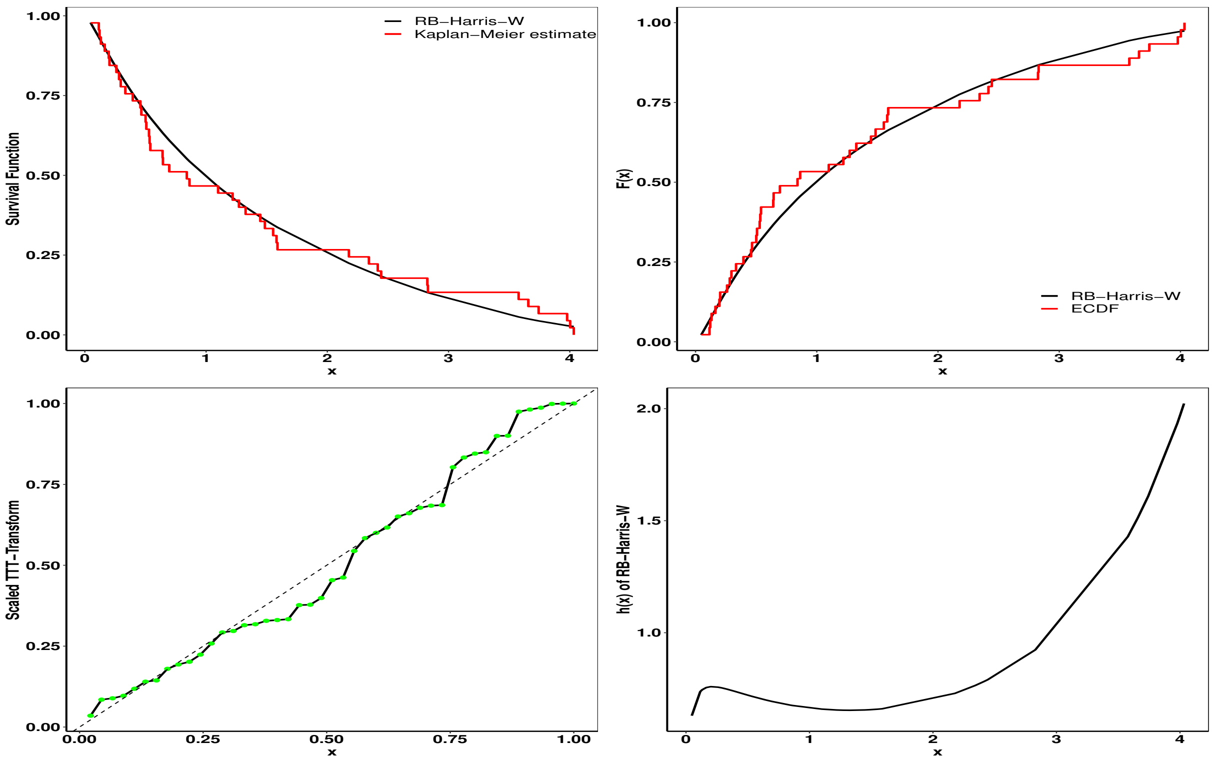

Fig. 10: Survival curve, ECDF, scaled TTT, and hazard function for chemotherapy data.

8 Conclusions

In this paper, we have extensively explored the novel Ristić-Balakrishnan-Harris-G (RB-Harris-G) family of distributions, providing comprehensive insights into its theoretical properties and practical applicability. The RB-Harris-G family of distributions was rigorously formulated through a robust probability density function (pdf) expansion, which laid the essential groundwork for deriving further statistical properties, such as stochastic orderings, moments, and reliability measures. This expansion significantly enhanced our analytical capabilities, allowing a deeper understanding and clearer interpretations of the model behavior.

Three specialized subfamilies—the RB-Harris-log-logistic, RB-Harris-Weibull, and RB-Harris-standard half logistic distributions—were introduced, each demonstrating distinct theoretical attributes suitable for modeling diverse real-world phenomena. Through thorough mathematical treatments, we elucidated their unique traits, providing practitioners with flexible and potent tools to address a wide variety of applied problems.

Parameter estimation was rigorously addressed using the maximum likelihood estimation approach, ensuring precise and reliable estimations. Comprehensive simulation studies confirmed the efficacy of these estimates, demonstrating strong consistency and minimal bias, as evidenced by decreasing RMSEs and average bias with increasing sample sizes.

Applying our model to four diverse real-world datasets—silicon nitride fracture toughness, unemployment insurance metrics, COVID-19 data, and chemotherapy survival times—we demonstrated the versatility and superior performance of the RB-Harris-Weibull distribution. Across these distinct datasets, our proposed distribution consistently exhibited improved fitting capability relative to other established competing models, supported rigorously by numerous goodness-of-fit criteria such as the Akaike Information Criterion, Bayesian Information Criterion, and Kolmogorov-Smirnov statistics. Graphical validation through PP, QQ plots, Kaplan-Meier curves, and scaled TTT transformations further reinforced the practical superiority of the RB-Harris-Weibull distribution.

The significance of the RB-Harris-G family lies not only in its robust flexibility and adaptability to a variety of datasets but also in its ability to model complex hazard structures, which makes it particularly suitable for reliability analysis, survival data modeling, and risk assessment applications. Researchers and practitioners in engineering, actuarial sciences, reliability engineering, and medical survival analysis will find this model invaluable due to its adaptability to skewed, unimodal, or bimodal distributions and its interpretability in terms of hazard and survival analysis.

Future research could beneficially explore further generalizations and hierarchical structures within the RB-Harris framework, investigate additional statistical properties such as conditional moments, or apply Bayesian inference methods to enhance parameter estimation under various prior distributions. In conclusion, the RB-Harris-G family and particularly its Weibull subclass presented here offer robust tools for statistical modeling, with extensive applicability and notable theoretical and practical implications across various scientific disciplines.

Acknowledgments

We gratefully acknowledge Mr. Aaron Stringfellow for his valuable contributions to the initial stages of this research. The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] Shui-Hua Jiang, Xian Liu, Ze Zhou Wang, Dian-Qing Li, and Jinsong Huang. Efficient sampling of the irregular probability distributions of geotechnical parameters for reliability analysis. Structural Safety, **101**:102309, 2023.
[2] Albert W. Marshall and Ingram Olkin. A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. Biometrika, **84**(3):641–652, 1997.
[3] Xinyu Chen, Zhenyu Shi, Yuanqi Xie, Zichen Zhang, Achraf Cohen, and Shusen Pu. Advancing continuous distribution generation: An exponentiated odds ratio generator approach. Entropy, **26**(12):1006, 2024.
[4] Haochong Yang, Mingfang Huang, Xinyu Chen, Ziyan He, and Shusen Pu. Enhanced real-life data modeling with the modified Burr III odds ratio–G distribution. Axioms, **13**(6):401, 2024.
[5] Shusen Pu, Thatayaone Moakofi, and Broderick Oluyede. The Ristić–Balakrishnan–Topp–Leone–Gompertz-G family of distributions with applications. Journal of Statistical Theory and Applications, pp. 1–35, 2023.
[6] Friday Ikechukwu Agu, Joseph Thomas Eghwerido, and Cosmas Kaitani Nziku. The alpha power Rayleigh-G family of distributions. Mathematica Slovaca, **72**(4):1047–1062, 2022.
[7] Broderick Oluyede, Leon Schröder, Sean Fang, Achraf Cohen, Thatayaone Moakofi, Yuhao Zhang, and Shusen Pu. Advancing reliability and medical data analysis through novel statistical distribution exploration. Mathematica Slovaca, **75**(1):225–242, 2025.
[8] Sudakshina Singha Roy, Hannah Knehr, Declan McGurk, Xinyu Chen, Achraf Cohen, and Shusen Pu. The Lomax-exponentiated odds ratio–G distribution and its applications. Mathematics, **12**(10):1578, 2024.
[9] Gauss M. Cordeiro, Abraão D.C. Nascimento, and Jodavid A. Ferreira. The MOGN distribution applied to medical imagery processing. Journal of Statistical Theory and Practice, **17**(1):1–22, 2023.
[10] Ahmed Z. Afify, Shawky Ahmed, and Mazen Nassar. A new inverse Weibull distribution: properties, classical and Bayesian estimation with applications. Kuwait Journal of Science, **48**(3), 2021.
[11] Zubir Shah, Dost Muhammad Khan, Zardad Khan, Muhammad Shafiq, and Jin-Ghoo Choi. A new modified exponent power alpha family of distributions with applications in reliability engineering. Processes, **10**(11):2250, 2022.
[12] Xueyu Wu, Zubair Ahmad, Eslam Hussam, Marwan H. Alhelali, Ramy Aldallal, Muqrin A. Almuqrin, and Fathy H. Riad. A new cosine-Weibull model: Distributional properties with applications to basketball and medical sectors. Alexandria Engineering Journal, **66**:751–767, 2022.
[13] Luis Gustavo Bastos Pinho, Gauss Moutinho Cordeiro, and Juvêncio Santos Nobre. On the Harris-G class of distributions: general results and application. Brazilian Journal of Probability and Statistics, **29**(4):813–832, 2015.
[14] Miroslav M. Ristić and Narayanaswamy Balakrishnan. The gamma-exponentiated exponential distribution. Journal of Statistical Computation and Simulation, **82**(8):1191–1206, 2012.
[15] Susan Foya, Broderick O. Oluyede, Adeniyi F. Fagbamigbe, and Boikanyo Makubate. The gamma log-logistic Weibull distribution: model, properties and application. Electronic Journal of Applied Statistical Analysis, **10**(1):206–241, 2017.
[16] Tibor K. Pogány and Abdus Saboor. The gamma exponentiated exponential–Weibull distribution. Filomat, **30**(12):3159–3170, 2016.
[17] L.G.B. Pinho, G.M. Cordeiro, and J.S. Nobre. The gamma-exponentiated Weibull distribution. Journal of Statistical Theory and Applications, **11**(4):379–395, 2012.
[18] Gauss M. Cordeiro, William D. Aristizábal, Dora M. Suárez, and Sébastien Lozano. The gamma modified Weibull distribution. Chilean Journal of Statistics, **6**(1):37–48, 2015.
[19] Broderick Oluyede, Shusen Pu, Boikanyo Makubate, and Yuqi Qiu. The gamma-Weibull-G family of distributions with applications. Austrian Journal of Statistics, **47**(1):45–76, 2018.
[20] Izrail S. Gradshteyn and Iosif M. Ryzhik. Table of integrals, series, and products. Academic Press, 2014.
[21] Shusen Pu, Broderick O. Oluyede, Yuqi Qiu, and Daniel Linder. A generalized class of exponentiated modified Weibull distribution with applications. Journal of Data Science, **14**(4):585–613, 2016.
[22] Gauss M. Cordeiro, Edwin M.M. Ortega, and Saralees Nadarajah. The Kumaraswamy Weibull distribution with application to failure data. Journal of the Franklin Institute, **347**(8):1399–1429, 2010.

[23] Gokarna R. Aryal, Edwin M. Ortega, G.G. Hamedani, and Haitham M. Yousof. The Topp-Leone generated Weibull distribution: regression model, characterizations and applications. International Journal of Statistics and Probability, **6**(1):126–141, 2017.
[24] Thatayaone Moakofi, Broderick Oluyede, and Boikanyo Makubate. A new gamma generalized Lindley-log-logistic distribution with applications. Afrika Statistika, **15**(4):2451–2481, 2020.
[25] Lazhar Benkhelifa. Alpha power Topp-Leone Weibull distribution: Properties, characterizations, regression modeling and applications. Journal of Statistics and Management Systems, **25**(8):1945–1970, 2022.

Shusen Pu is an Assistant Professor in the Department of Mathematics and Statistics at the University of West Florida (UWF). He earned his Ph.D. in Applied Mathematics from Case Western Reserve University in 2020, specializing in stochastic processes in biological systems. Following his doctoral studies, Dr. Pu served as a postdoctoral researcher, contributing to projects in both the Department of Biomedical Engineering and the Department of Mathematics at Vanderbilt University with a focus on computational neuroscience and data analysis. Dr. Pu's research spans computational neuroscience, mathematical statistics, data analytics, and deep learning. He applies advanced mathematical models and machine learning techniques to uncover the dynamics of neural systems, emphasizing working memory and decision-making processes. His innovative work includes developing novel statistical distributions for applications in medicine, finance, and ecology. At UWF, Dr. Pu leads several research projects and has been recognized for excellence in teaching and undergraduate research mentoring. An active member of the academic community, he frequently organizes conference sessions and presents his findings at major international events, contributing significantly to the advancement of computational biology and data science.

Thatayaone Moakofi obtained his PhD in Statistics from the Botswana International University of Science and Technology (BIUST). His research interests include distribution theory, reliability theory, survival analysis, weighted distributions, and infectious disease modeling. He has published over 23 peer-reviewed research papers and three book chapters in recognized journals. He has over four years of teaching experience. Dr. Moakofi is an active reviewer for reputable journals. His contributions to the field have earned him over 155 citations, an h-index of 7, and an i10-index of 4 according to Google Scholar as of 2025.

Broderick Oluyede is a Full Professor of Mathematics and Statistics, and the former Director of the Statistical Consulting Unit (SCU) in the Department of Mathematical Sciences, Georgia Southern University, Statesboro, Georgia, USA. He is currently a Full Professor of Mathematics and Statistics at the Botswana International University of Science and Technology (BIUST). He has over 30 years of research and teaching experience at Bowling Green State University, Georgia State University, University of Georgia, Oklahoma State University, Georgia Southern University and BIUST. He has authored and co-authored over 200 research papers. His research interests include multivariate statistical analysis, distribution theory, reliability theory, survival analysis, categorical data analysis, biostatistics, order restricted inference, stochastic dependence and weighted distributions. He is a member of editorial boards of several journals.

Achraf Cohen is an Associate Professor in the Department of Mathematics and Statistics at the University of West Florida. His research interests encompass statistical process monitoring, wavelet analysis, anomaly detection, and the optimization of AI models within the contexts of health and education. He has published over 35 peer-reviewed articles on topics such as new probability distributions, AI optimization models, sports medicine, applied statistical modeling, control charts, and functional data classification. Additionally, Dr. Cohen has presented his work at academic conferences in France, the Czech Republic, and the United States. He has also mentored more than 50 undergraduate and graduate students in the CSDA Lab at UWF.