

Modeling Real-life Data Sets with a Novel G Family of Continuous Probability Distributions: Statistical Properties, and Copulas

Murtadha Mansour Abdullah^{1,*} and Afif Masmoudi²



* Corresponding Author

¹University of Wasit. College of Administration and Economics, Iraq; mabdullah@uowasit.edu.iq

²Laboratory of Probability and Statistics, University of Sfax, Tunisia; afif.masmoudi@fss.usf.tn

Abstract

This work presents a novel two-parameter G family of continuous probability distributions with compounded parameters. To determine and examine the pertinent mathematical properties, calculations are performed. In one of the special sections, the standard inverse-Rayleigh baseline model is mathematically and statistically emphasized. We generated a number of bivariate and multivariate distributions using the copula method. These new distributions will aid in the modelling of bivariate and multivariate data. The applicability and flexibility of the new compounded two-parameters-G family are demonstrated through three applications to real-life data sets. These examples demonstrate the applicability of the family.

Key Words: Compound Families; Copulas; Inverse- Rayleigh Distribution; Geometric G Family; Graphical Simulations.

Mathematical Subject Classification: 62E10, 60K10, 60N05

1 Introduction

In the fields of statistics and probability theory, one of the most important classes of distributions is known as the geometric G (GC-G) family of probability distributions. These distributions find use in a wide variety of domains for a variety of purposes. In a series of independent Bernoulli trials, the geometric distribution, which is a member of the family GC-G, is frequently used to describe waiting times or the number of tries until the first success. This is because the geometric distribution is a member of the GC-G family. It has found use in dependability analysis, queueing theory, and a variety of real-life scenarios that include waiting times. The GC-G family of distributions includes models that can accurately simulate difficult-to-predict occurrences. For instance, the negative binomial distribution, which belongs to this family of distributions, can be used to simulate the number of tests that must be conducted in order to see a predetermined number of successes. Because of this, it is useful for modelling unusual events, such as the number of accidents that occur over a specific time or the number of defective items that are produced during a given manufacturing process. Applications in risk analysis and insurance frequently make use of the GC-G distributions, which include the geometric-exponential and geometric-gamma distributions, among others. These distributions can be utilized to represent claim frequencies or inter-arrival durations between claims, both of which are significant components in the process of evaluating risk and calculating insurance rates. Within the realms of reliability and survival research, the GC-G family has a wide variety of potential applications.

Utilizing geometric distribution is one way to model the length of time a system has left before it fails or the amount of time individuals in a population have left to live. This can be done to model the amount of time individuals have left to live. Reliability engineers and medical researchers can make predictions about the reliability and longevity of specific systems or people by analyzing these distributions. Modelling discrete data is made easier with the help of the GC-G family's versatile architecture. Count data and discrete events can both be modelled with different members of this family, such as the geometric distribution, the negative binomial distribution, and the Poisson distribution. Analysis and modelling of discrete phenomena are common applications for these distributions, particularly in the domains of epidemiology, ecology, finance, and telecommunications, to name a few. The GC-G family is frequently utilised for simulation research as well as the creation of random numbers. The process of simulating events using a

geometric distribution or one of its related distributions can be helpful in generating random sequences that replicate real-life conditions. This can be useful for testing theories, evaluating algorithms, and conducting Monte Carlo simulations.

The compound G family is the name that we give to a novel family of probability distributions that we define in this research. This new family of probability distributions is derived from an existing family of probability distributions known as the geometric generated Rayleigh G (GCGR-G) family, which is founded on the generated Rayleigh-G (GR-G) family. This new family of probability distributions is referred to as the GCGR-G family. Let $\omega_{\underline{v}}(\psi)$ and $w_{\underline{v}}(\psi)$ are two functions which denote to the cumulative distribution function (CDF) and its corresponding probability density functions (PDF) with main parameter vector $\underline{v}$. First, we will consider the CDF of the well-known generated-Rayleigh G (GR-G) family class ($W_{\sigma,\underline{v}}(\psi)$) with the following CDF:

$$W_{\sigma,\underline{v}}(\psi) = \exp[-\Delta_{\sigma,\underline{v}}^2(\psi)]|_{\psi \in \mathcal{R}, \sigma > 0}, \quad (1)$$

where $\Delta_{\sigma,\underline{v}}^2(\psi)$ refers to squared odd ratio function which can be explained as

$$\Delta_{\sigma,\underline{v}}(\psi) = \omega_{\underline{v}}^{\sigma}(\psi) / [1 - \omega_{\underline{v}}^{\sigma}(\psi)]|_{\psi \in \mathcal{R}, \sigma > 0},$$

and $\frac{d}{d\psi} W_{\sigma,\underline{v}}(\psi) = w_{\sigma,\underline{v}}(\psi)$ is the PDF corresponding to the CDF in (1). On the other hand, the CDF of the geometric G (GC-G) family is defined as follows: for any random variable (RV) according to any standard baseline model which has the CDF represented by $\omega_{\underline{v}}(\psi)$, then the CDF of the geometric G (GC-G) family is defined as follows:

$$F_{\gamma,\underline{v}}(\psi) = \frac{\gamma \omega_{\underline{v}}(\psi)}{R_{\underline{v}}(\psi) + \gamma \omega_{\underline{v}}(\psi)}|_{\psi \in \mathcal{R}, \gamma > 0}, \quad (2)$$

where $R_{\underline{v}}(\psi) = 1 - \omega_{\underline{v}}(\psi)$ refers to the reliability function (RF) for the base line model. By substituting (1) in (2), we get a new extension of the GC-G class to provide a new flexible compound G family. The new compound G family has a wide physical interpretation as given later in Section 2. The new compound G family is derived via expanding the GC-G family and the GR-G family.

In reality, there are a lot of helpful contributions in the statistical literature on compound G families. Many applied disciplines and fields, including medical, reliability, actuarial sciences, engineering, insurance, econometrics, biology, demography, environmental sciences, etc., employ these contributions in the field of modelling real-life datasets. In this sense, the following examples can be made: Afify et al. (2016a) for the complementary transmuted geometric G family with modelling real-life datasets, Afify et al. (2016a) for the complementary geometric transmuted G family with properties and applications, and Aryal and Yousof (2017) for the exponentiated generalized Poisson family are some examples of the families that have been studied. Cordeiro and de Castro (2011) studied the Kumaraswamy family with some applications. Cordeiro et al (2014).

The new GCGR-G class stands out for having a wide range of applications. Using three examples, we show that the GCGR class provides better matches than many other families. Figures 5, 8, and 11 (1st row right graphs) depict "asymmetric monotonically increasing hazard rate functions (HRF)"; Figures 5 and 8 (2nd row right and left graphs) depict "real-life data with some extreme values," and Figures 5 and Figure 7 (3rd row right and left graphs) depict "real-life data with some extreme values." The new family may be useful in modelling such real-life data. Real-life data without extreme values, as seen in Figure 10's second row right and left graphs, real-life data with symmetric and unimodal nonparametric Kernel densities, as seen in Figures 5, 8, and the first row left graph of Figure 10, real-life data with bimodal and heavy-tailed nonparametric Kernel densities, as seen in Figures 5, and Figure 7, respectively, and real-life data whose nonparametric Kernel density cannot be fit by common models.

2. The new family and its related representations

We build a new two-parameter family of continuous probability distributions called the GCGR-G family by utilizing (1) and (2) to generate the baseline CDF of (2). This family of distributions is known as the GCGR-G family. The CDF of the GCGR-G family can be defined in the following manner:

$$F_{\underline{u}}(\psi) = \frac{\gamma - \gamma \exp[-\Delta_{\sigma,\underline{v}}^2(\psi)]}{1 - \{1 - \exp[-\Delta_{\sigma,\underline{v}}^2(\psi)]\} + \gamma \{1 - \exp[-\Delta_{\sigma,\underline{v}}^2(\psi)]\}}, \quad (3)$$

where $\underline{u} = (\gamma, \sigma, \underline{v})|_{\psi \in \mathcal{R}, \gamma > 0, \sigma > 0}$ refers to the parameter vector. See Aboraya et al. (2020), Ibrahim et al. (2021), Chesneau et al. (2021), Yousof et al. (2021) and Emam et al. (2023b) for additional information on how to

generate a new discrete G family by modeling the count data using the new CDF introduced in (3). The following is the contents of the PDF that match to (3):

$$f_{\underline{u}}(y) = 2\gamma\sigma \frac{\omega_{\underline{u}}(y)\omega_{\underline{u}}^{2\sigma-1}(y)[1 - \omega_{\underline{u}}^{\sigma}(y)]^{-3} \exp[-\Delta_{\sigma,\underline{u}}^2(y)]}{(1 - (1 - \gamma)\{1 - \exp[-\Delta_{\sigma,\underline{u}}^2(y)]\})^2} \Big|_{y \in \mathcal{R}, \gamma > 0, \sigma > 0}. \quad (4)$$

The generalized G (GZX-G family) family of distributions can be used to generate some power series expansions for Equations (3) and (4). The PDF in (4) can be stated as follows by using the generalized binomial expansion and the power series:

$$f_{\underline{u}}(y) = 2\gamma\sigma\omega_{\underline{u}}(y) \sum_{i,j,\tau=0}^{+\infty} \frac{(1-\gamma)^i(-1)^{\tau}}{\tau!(1+i)^{-\tau}} \frac{\omega_{\underline{u}}^{2\sigma(\tau+1)-1}(y)}{[1 - \omega_{\underline{u}}^{\sigma}(y)]^{3+2\tau}} \binom{-2}{i} \binom{i}{j}.$$

Then, using the expansion of Taylor, we have

$$f_{\underline{u}}(y) = \sum_{\tau,w=0}^{+\infty} \Omega_{\tau,w} \pi_{\sigma^*,\underline{u}}(y) \Big|_{\sigma^*=\sigma[2(\tau+1)+w]>0}, \quad (5)$$

where

$$\Omega_{\tau,w} = 2\gamma\sigma \sum_{i=0}^{+\infty} \sum_{j=0}^i \frac{(1-\gamma)^i(-1)^{\tau}(1+i)^{\tau}}{\tau! \sigma^*} \binom{-2}{i} \binom{i}{j} \binom{-3-2\tau}{w}.$$

As a consequence of this, it is possible to infer certain mathematical properties of the GCGR-G family directly from those of the GZX-G family. Equation (5) represents the most important result of this section. In the same way, the CDF of the GCGR-G family can be expressed through the combination of densities from the GZX-G family. By integrating equation (5), we arrive to the same representation for the mixture:

$$F_{\underline{u}}(y) = \sum_{\tau,w=0}^{+\infty} \Omega_{\tau,w} \Pi_{\sigma^*,\underline{u}}(y) \Big|_{\sigma^*>0}, \quad (6)$$

where the function $\Pi_{\sigma^*,\underline{u}}(y)$ refers to the baseline CDF of the GZX-G family with power parameter σ^* .

3. Copulas

In order to represent data with two or more variables, a copula, a fundamental statistical concept, is used. A function that links the marginal distributions of two or more variables to their combined distribution is known as this. It links the marginal distributions of each variable in particular. Copulas have been more well-liked in recent years as a result of their versatility and ability to mimic intricate dependency patterns between variables. They are now better able to replicate complex dependency patterns because to this talent. The significance of copulas and their use in statistical modelling and modelling of bivariate data are demonstrated by the following examples:

1. Copulas are used to show the variables' relationship to one another, which is a key idea in many different fields of data and statistical research. We can describe the joint distribution of all variables while maintaining the marginal distributions of each variable by using copulas. Finding intricate interactions between variables, which cannot be measured using conventional correlation techniques, is made considerably easier as a result.
2. Copulas are used in the realm of finance to emphasise the connection between the returns on different assets. For portfolio optimisation, which tries to create an asset allocation that maximises returns while minimises risk, this is very necessary. It is essential that you have this knowledge at your disposal. When we use copulas as a modelling tool for the dependency between assets, we are able to more accurately assess the risk associated with a portfolio and construct portfolios that are more effective.
3. Copulas can be used in risk management to estimate the likelihood of catastrophic occurrences like market collapses and natural disasters. Copulas are also used to estimate the likelihood of catastrophic catastrophes. Because copulas model the dependence structure that exists between variables, they can give a more precise evaluation of the likelihood that such occurrences will occur. A crucial element of risk management and insurance is this.
4. Copulas can also be used to generate data, which is useful in circumstances when it is challenging or costly to get the required data. By simulating new data sets with the identical dependent structures as the original data, we are able to generate brand-new data for testing and analysis. This provides us with extra data to work with. This will be achieved by using a copula to characterise the structure of the interrelationships between

the variables.

In order to replicate the dependence structure that exists between two or more random variables, the FAGM copula, which is a family of parametric copulas, is utilized. It is an approximation of the copulas of Ali Mikhail Haq, Frank, and Plackett that is of the first order, and it has a simple form that enables explicit calculus and exact results (see Ali et al. (1978) for more information). The FAGM copula is attractive not only due to the ease with which it may show a broad variety of dependency patterns, such as positive, negative, and mixed dependence, but also due to the fact that it can do so. The FAGM copula has been applied in a variety of contexts, such as in the management of financial risk, the underwriting of insurance policies, and the evaluation of environmental risk. For instance, the FAGM copula has been utilized to symbolize the connection between asset values and insurance claims, as well as the connection between environmental variables and other factors. Modelling a wide variety of dependent structures can be accomplished with the flexible assistance of the FAGM copula.

It is simple to implement and has a number of benefits over other families of copulas, including the capacity to show mixed dependency, among other things. In this section, we provide some new bivariate type GCGR-G variations that are based on the FAGM copula that was developed by Morgenstern (1956), Farlie (1960), Gumbel (1958 and 1960), Johnson and Kotz (1975), Balakrishnan and Lai (2009), and Johnson and Kotz (1977). There is also discussion of the "modified FAGM (MFAGM)" copula, the "Clayton copula" copula, the "Archimedean-Ali-Mikhail-Haq" copula, and the "Renyi's entropy" copula (for more information, see Balakrishnan and Lai (2009), Ali et al. (1978), and Pougaza and Djafari (2011)). In addition, the Multivariate GCGR-G type (also known as Mv GCGR-G) is provided (for further information, refer to Balakrishnan and Lai 2009). Nevertheless, in the not-too-distant future, there might be some attempts made to investigate and evaluate these new models.

3.1 BGCGR-G type via FAGM copula

The following are some applications that make use of the FAGM copula:

- I. The FAGM copula may be utilized to provide a representation of the reliance of asset prices. This can be utilized in the process of developing hedging strategies as well as conducting risk assessments on asset portfolios.
- II. A simulation of the interdependence of insurance claims can be carried out with the help of the FAGM copula. This can be utilized to decide pricing and reserving techniques, in addition to evaluating the risk that is posed by an insurance firm.
- III. A simulation of the interdependence of environmental variables can be carried out with the assistance of the FAGM copula. This can be put to use to determine the likelihood of environmental catastrophes and to design strategies for mitigating their effects. Using the joint CDF of the FAGM family we have

$$\mathcal{C}_{\Psi}(\kappa, \kappa) = \kappa\kappa(1 + \Psi\kappa\kappa),$$

where $\kappa \in (0,1)$, $\kappa \in (0,1)$, are two continuous marginal functions and $\Psi \in [-1,1]$ is a dependence parameter. Then, we have

$$\mathcal{C}_{\Psi}(\kappa, 0) = \mathcal{C}_{\Psi}(0, \kappa) = 0|_{(\kappa, \kappa \in (0,1))},$$

which is "grounded minimum condition" and $\mathcal{C}_{\Psi}(\kappa, 1) = \kappa$ and $\mathcal{C}_{\Psi}(1, \kappa) = \kappa$ which is "grounded maximum condition".

In the context of statistical copulas, the grounded minimum condition is an important property that copulas must satisfy to ensure their validity and consistency. A copula is a mathematical function that links marginal distributions to a joint distribution. It plays a crucial role in multivariate statistical analysis, especially in applications like risk management, finance, and insurance, where modeling the dependence structure between variables is crucial. The grounded minimum condition ensures that the copula is well-behaved and interpretable. The fact that the copula attains its minimum value at the lower bounds of its arguments aligns with our intuition about dependence. When all variables are at their minimum value, the joint probability should also be minimized, which is reflected by this condition. The grounded minimum condition is a part of the more general concept of coherence, which imposes certain properties on copulas to ensure their consistency and validity. Coherence is a fundamental requirement for any valid copula function, and the grounded minimum condition contributes to fulfilling this requirement. Then, setting

$$\kappa = 1 - \bar{\kappa}, \bar{\kappa} = 1 - \frac{\gamma_1 - \gamma_1 \exp[-\Delta_{\sigma_1, \underline{\Psi}}^2(y_1)]}{1 - \{1 - \exp[-\Delta_{\sigma_1, \underline{\Psi}}^2(y_1)]\} + \gamma_1 \{1 - \exp[-\Delta_{\sigma_1, \underline{\Psi}}^2(y_1)]\}},$$

and

$$\kappa = 1 - \bar{\kappa}, \bar{\kappa} = 1 - \frac{\gamma_2 - \gamma_2 \exp[-\Delta_{\sigma_2, \underline{y}}^2(\psi_2)]}{1 - \{1 - \exp[-\Delta_{\sigma_2, \underline{y}}^2(\psi_2)]\} + \gamma_2 \{1 - \exp[-\Delta_{\sigma_2, \underline{y}}^2(\psi_2)]\}}.$$

Then, we have

$$F(\psi_1, \psi_2) = (1 + \Psi \overline{\kappa \kappa}) \kappa \kappa.$$

3.2 BGCGR-G type via MFAGM copula

The MFAGM copula is a flexible tool that was produced based on the primary version of the FAGM; it has the ability to be used in the modeling of a large range of different sorts of dependence structures. The MFAGM copula was developed based on the primary version of the FAGM. It is simple to implement, and it offers a number of benefits over other families of copulas, such as the ability to simulate mixed dependency (for more details regarding this topic, refer to Rodriguez-Lallena and Ubeda-Flores (2004)). As a result of this, the FAGM copula is a well-liked choice for a variety of diverse applications due to its many advantageous characteristics. In this section of the article, we are going to look at the MFAGM copula in the following form:

$$\mathcal{C}(\kappa, \kappa, \Psi) = \Psi \overline{\mathcal{J}(\kappa) \mathcal{L}(\kappa)} + \kappa \kappa,$$

where

$$\overline{\mathcal{J}(\kappa)} = \kappa \overline{\mathcal{J}(\kappa)}, \quad \overline{\mathcal{L}(\kappa)} = \kappa \overline{\mathcal{L}(\kappa)}$$

and the two continuous functions $\mathcal{J}(\kappa)$ and $\mathcal{L}(\kappa)$ are two functions on $(0,1)$ with the following condition boundary condition:

$$\mathcal{J}(0) = \mathcal{L}(0) = \mathcal{J}(1) = \mathcal{L}(1) = 0,$$

where

$$A(\kappa; h_1) = \inf \left\{ \frac{\partial}{\partial \kappa} \overline{\mathcal{J}(\kappa)} : h_1(\kappa) \right\} < 0, B(\kappa; h_1) = \sup \left\{ \frac{\partial}{\partial \kappa} \overline{\mathcal{J}(\kappa)} : h_1(\kappa) \right\} < 0,$$

$$A(\kappa; h_2) = \inf \left\{ \frac{\partial}{\partial \kappa} \overline{\mathcal{L}(\kappa)} : h_2(\kappa) \right\} > 0, B(\kappa; h_2) = \sup \left\{ \frac{\partial}{\partial \kappa} \overline{\mathcal{L}(\kappa)} : h_2(\kappa) \right\} > 0.$$

Then,

$$\min(A(\kappa; h_1)B(\kappa; h_1), A(\kappa; h_2)B(\kappa; h_2)) \geq 1,$$

where

$$\frac{\partial}{\partial \kappa} \overline{\mathcal{J}(\kappa)} = \mathcal{J}(\kappa) + \kappa \frac{\partial}{\partial \kappa} \mathcal{J}(\kappa)$$

$$h_1(\kappa) = \left\{ \kappa : \kappa \in (0,1) \mid \frac{\partial}{\partial \kappa} \overline{\mathcal{J}(\kappa)} \text{ exists} \right\},$$

and

$$h_2(\kappa) = \left\{ \kappa : \kappa \in (0,1) \mid \frac{\partial}{\partial \kappa} \overline{\mathcal{L}(\kappa)} \text{ exists} \right\}.$$

A model of the correlation between asset prices can be constructed utilizing the MFAGM copula. The creation of hedging strategies for asset portfolios and risk assessments are also possible using this information. A model of the interdependence of environmental variables can be constructed with the help of the MFAGM copula. This can be used to make predictions about the possibility of environmental disasters and come up with measures for mitigating their effects. It is possible to demonstrate the interconnectedness of insurance claims by utilizing the MFAGM copula. This can be utilized to design pricing and reservation policies, in addition to providing an estimate of the risk level that an insurance company faces. In light of the limitations imposed by the ongoing study, we will conduct the following investigation of four possible kinds of MFAGM copulas.

The Type I BGCGR-G via the MFAGM copula:

Consider the two functional forms $\mathcal{J}(\kappa)$ and $\mathcal{L}(\kappa)$, then the Type I BGCGR-G according the MFAGM copula can be expressed as

$$\mathcal{C}_{\psi}(\kappa, \kappa) = \psi[\widetilde{\mathcal{I}(\kappa)}\widetilde{\mathcal{L}(\kappa)}] + \left(\frac{\gamma_1 - \gamma_1 \exp[-\Delta_{\sigma_1, \underline{y}}^2(y_1)]}{1 - \{1 - \exp[-\Delta_{\sigma_1, \underline{y}}^2(y_1)]\} + \gamma_1 \{1 - \exp[-\Delta_{\sigma_1, \underline{y}}^2(y_1)]\}} \times \frac{\gamma_2 - \gamma_2 \exp[-\Delta_{\sigma_2, \underline{y}}^2(y_2)]}{1 - \{1 - \exp[-\Delta_{\sigma_2, \underline{y}}^2(y_2)]\} + \gamma_2 \{1 - \exp[-\Delta_{\sigma_2, \underline{y}}^2(y_2)]\}} \right),$$

where

$$\widetilde{\mathcal{I}(\kappa)} = \kappa \left\{ 1 - \frac{\gamma_1 - \gamma_1 \exp[-\Delta_{\sigma_1, \underline{y}}^2(y_1)]}{1 - \{1 - \exp[-\Delta_{\sigma_1, \underline{y}}^2(y_1)]\} + \gamma_1 \{1 - \exp[-\Delta_{\sigma_1, \underline{y}}^2(y_1)]\}} \right\} \Big|_{\underline{y}_1 > 0},$$

and

$$\widetilde{\mathcal{L}(\kappa)} = \kappa \left\{ 1 - \frac{\gamma_2 - \gamma_2 \exp[-\Delta_{\sigma_2, \underline{y}}^2(y_2)]}{1 - \{1 - \exp[-\Delta_{\sigma_2, \underline{y}}^2(y_2)]\} + \gamma_2 \{1 - \exp[-\Delta_{\sigma_2, \underline{y}}^2(y_2)]\}} \right\} \Big|_{\underline{y}_2 > 0}.$$

The Type II BGCGR-G via the MFAGM copula:

Consider the two functional forms $\mathcal{I}(\kappa)$ and $\mathcal{L}(\kappa)$, then the Type II BGCGR-G according the MFAGM copula can be expressed as

$$\mathcal{I}(\kappa)|_{(\psi_1 > 0)} = \kappa^{\psi_1} (1 - \kappa)^{1 - \psi_1} \text{ and } \mathcal{L}(\kappa)|_{(\psi_2 > 0)} = \kappa^{\psi_2} (1 - \kappa)^{1 - \psi_2}.$$

Then, the new Type II BGCGR-G version can be derived from

$$\mathcal{C}_{\psi, \psi_1, \psi_2}(\kappa, \kappa) = [1 + \psi \kappa^{\psi_1} \kappa^{\psi_2} (1 - \kappa)^{1 - \psi_1} (1 - \kappa)^{1 - \psi_2}] \kappa \kappa.$$

The Type III BGCGR-G via the MFAGM copula:

According to the Type III of the MFAGM copula, a new two function ($\widetilde{\mathcal{W}(\kappa)}$ and $\widetilde{\mathcal{M}(\kappa)}$) can be considered where

$$\widetilde{\mathcal{W}(\kappa)} = \kappa [\log(1 + \bar{\kappa})] \text{ and } \widetilde{\mathcal{M}(\kappa)} = \kappa [\log(1 + \bar{\kappa})].$$

Then, depending on Ghosh and Ray (2016), one can easily apply the following type

$$\mathcal{C}_{\psi}(\kappa, \kappa) = \kappa \kappa [1 + \psi \widetilde{\mathcal{W}(\kappa)} \widetilde{\mathcal{M}(\kappa)}].$$

The Type IV BGCGR-G via the MFAGM copula:

Following Ghosh and Ray (2016) the bivariate CDF of the type IV BGCGR-G model can be derived from

$$\mathcal{C}(\kappa, \kappa) = \kappa F^{-1}(\kappa) + \kappa F^{-1}(\kappa) - F^{-1}(\kappa) F^{-1}(\kappa),$$

where $F^{-1}(\kappa) = Q(\kappa)$ and $F^{-1}(\kappa) = Q(\kappa)$.

3.3 BGCGR-G type under the Clayton copula

The Clayton copula is a type of copula that can be found in nature. This particular copula permits any non-zero degree of (lower) tail dependence between the different variables to be modelled. In addition to being interchangeable, it possesses an Archimedean copula as one of its properties. The Clayton copula lends itself to a variety of different interpretations, some of which are as follows:

$$\mathcal{C}(\kappa_1, \kappa_2) = (\kappa_1^{-\psi} + \kappa_2^{-\psi} - 1)^{-\frac{1}{\psi}} \Big|_{\psi \in [0, \infty]}.$$

Let us assume that $T \sim \text{GCGR-G}(\delta_1, b_1, \vartheta)$ and $X \sim \text{GCGR-G}(\delta_2, b_2, a)$. Then, setting

$$\kappa_1 = \kappa(t; \underline{y}_1) = \frac{\gamma_1 - \gamma_1 \exp[-\Delta_{\sigma_1, \underline{y}}^2(t)]}{1 - \{1 - \exp[-\Delta_{\sigma_1, \underline{y}}^2(t)]\} + \gamma_1 \{1 - \exp[-\Delta_{\sigma_1, \underline{y}}^2(t)]\}} \Big|_{\underline{y}_1 > 0},$$

and

$$\kappa_2 = \kappa(y; \underline{y}_2) = \frac{\gamma_2 - \gamma_2 \exp[-\Delta_{\sigma_2, \underline{y}}^2(y)]}{1 - \{1 - \exp[-\Delta_{\sigma_2, \underline{y}}^2(y)]\} + \gamma_2 \{1 - \exp[-\Delta_{\sigma_2, \underline{y}}^2(y)]\}} \Big|_{\underline{y}_2 > 0},$$

After conducting an analysis using the Clayton copula, the BGCGR-G type distribution can then be managed in the following manner:

$$F(t, y) = \mathcal{C}(F_{\underline{U}_1}(t), F_{\underline{U}_2}(y)) = \left[\left(\frac{\gamma_1 - \gamma_1 \exp[-\Delta_{\sigma_1, \underline{U}}^2(t)]}{1 - \{1 - \exp[-\Delta_{\sigma_1, \underline{U}}^2(t)]\} + \gamma_1 \{1 - \exp[-\Delta_{\sigma_1, \underline{U}}^2(t)]\}} \right)^{-\psi} + \left(\frac{\gamma_2 - \gamma_2 \exp[-\Delta_{\sigma_2, \underline{U}}^2(y)]}{1 - \{1 - \exp[-\Delta_{\sigma_2, \underline{U}}^2(y)]\} + \gamma_2 \{1 - \exp[-\Delta_{\sigma_2, \underline{U}}^2(y)]\}} \right)^{-\psi} \right]^{-\frac{1}{\psi}}.$$

3.4 BGCGR-G type under the Renyi's entropy

The Renyi's entropy copula allows us to easily build bivariate probability distributions without requiring costly mathematical derivations, making bivariate data mathematics and statistical modeling easy. The Renyi's entropy copula can be used to directly deduce the bivariate version. Renyi's entropy copula introduces no new parameters into the new bivariate model. The Renyi's entropy copula can be found in Pougaza and Djafari (2011). Then, according to the theorem of Pougaza and Djafari (2011), we have:

$$\mathcal{C}(\kappa, \kappa) = y_2 \kappa + y_1 \kappa - y_1 y_2,$$

then, the associated CDF of the BGCGR-G will be

$$\mathcal{C}(y_1, y_2) = R(F_{\underline{P}_1}(y_1), F_{\underline{P}_2}(y_2)) = -y_1 y_2 + y_2 \left\{ \frac{\gamma_1 - \gamma_1 \exp[-\Delta_{\sigma_1, \underline{U}}^2(y_1)]}{1 - \{1 - \exp[-\Delta_{\sigma_1, \underline{U}}^2(y_1)]\} + \gamma_1 \{1 - \exp[-\Delta_{\sigma_1, \underline{U}}^2(y_1)]\}} \right\} + y_1 \left\{ \frac{\gamma_2 - \gamma_2 \exp[-\Delta_{\sigma_2, \underline{U}}^2(y_2)]}{1 - \{1 - \exp[-\Delta_{\sigma_2, \underline{U}}^2(y_2)]\} + \gamma_2 \{1 - \exp[-\Delta_{\sigma_2, \underline{U}}^2(y_2)]\}} \right\}.$$

3.5 BGCGR-G type via the Archimedean Ali-Mikhail-Haq copula

The Archimedean Ali-Mikhail-Haq copula makes it simpler to build bivariate probability distributions without requiring significant mathematical derivations, which makes it easier to mathematically and statistically characterize bivariate data. This is because it simplifies the process of developing bivariate probability distributions. It is possible to generate the bivariate form of the Archimedean Ali-Mikhail-Haq copula (see Ali et al. (1978)) in a straightforward manner by utilizing the two above-mentioned functions. According to the Archimedean Ali-Mikhail-Haq copula, the innovative bivariate model can only add one parameter to the mix. Balakrishnan and Lai (2009) and Ali et al. (1978) are the primary references that are utilized in the construction of the Archimedean Ali-Mikhail-Haq copula. Under more stringent Lipschitz conditions, one may then construct the one-of-a-kind joint CDF of the Archimedean Ali-Mikhail-Haq copula by making use of the fundamental formula that is as follows:

$$\mathcal{C}(\kappa, \kappa) = \frac{1}{1 - \vartheta \overline{\kappa \kappa}} \kappa \kappa |_{\psi \in (-1, 1)}.$$

As a result, the Archimedean Ali-Mikhail-Haq copula's matching joint PDF can be determined as follows:

$$c(\kappa, \kappa) = \frac{1 - \vartheta + 2\vartheta \frac{\overline{\kappa \kappa}}{1 - \vartheta \overline{\kappa \kappa}}}{[1 - \vartheta \overline{\kappa \kappa}]^2} |_{\psi \in (-1, 1)}.$$

Then, for any RVs $T \sim \text{GCGR-G}(\delta_1, b_1, c)$ and $X \sim \text{GCGR-G}(\delta_2, b_2, a)$ we have the following main result

$$\mathcal{C}(\kappa, \kappa) = \frac{\frac{\gamma_2 - \gamma_2 \exp[-\Delta_{\sigma_2, \underline{U}}^2(y)]}{1 - \{1 - \exp[-\Delta_{\sigma_2, \underline{U}}^2(y)]\} + \gamma_2 \{1 - \exp[-\Delta_{\sigma_2, \underline{U}}^2(y)]\}} \times \frac{\gamma_1 - \gamma_1 \exp[-\Delta_{\sigma_1, \underline{U}}^2(t)]}{1 - \{1 - \exp[-\Delta_{\sigma_1, \underline{U}}^2(t)]\} + \gamma_1 \{1 - \exp[-\Delta_{\sigma_1, \underline{U}}^2(t)]\}}}{1 - \vartheta \left(\frac{\gamma_2 - \gamma_2 \exp[-\Delta_{\sigma_2, \underline{U}}^2(y)]}{1 - \{1 - \exp[-\Delta_{\sigma_2, \underline{U}}^2(y)]\} + \gamma_2 \{1 - \exp[-\Delta_{\sigma_2, \underline{U}}^2(y)]\}} \times \frac{\gamma_1 - \gamma_1 \exp[-\Delta_{\sigma_1, \underline{U}}^2(t)]}{1 - \{1 - \exp[-\Delta_{\sigma_1, \underline{U}}^2(t)]\} + \gamma_1 \{1 - \exp[-\Delta_{\sigma_1, \underline{U}}^2(t)]\}} \right)} |_{\psi \in (-1, 1)}.$$

3.6 The MvGCGR-G type

When it comes to the procedure of generating multivariate distributions or families, the Clayton copula is typically considered to be one of the simplest methods that can be used and is widely regarded for this fact. This is due to the fact that Clayton Copula was the one who initially designed the Clayton copula. As a result of this, the Clayton copula should be taken into consideration as a possibility. This is because both its mathematical formulation and its application can be carried out with an amount of work that is manageable. The reason for this is due to the fact that

both can be done with relative ease. In practical applications for multivariate data modeling in engineering and medical journals validity, insurance, and reinsurance, as well as other fields, the multivariate distributions or families that are based on the Clayton copula are adaptive. This is the case because the Clayton copula is a copula that is based on a multivariate distribution. These distributions and families are also flexible enough to be applied in a variety of other settings. This is because the Clayton copula acts as a foundation for the distributions, which are also referred to as families. In accordance with the idea of the Clayton copula, which is summarized as follows, the multivariate GCGR-G version can be formed using the follows:

$$H(\kappa_j) = \left(\sum_{j=1}^w \left\{ \frac{\gamma_j - \gamma_j \exp \left[-\Delta_{\sigma_j, \underline{y}}^2(t) \right]}{1 - \left\{ 1 - \exp \left[-\Delta_{\sigma_j, \underline{y}}^2(t) \right] \right\} + \gamma_j \left\{ 1 - \exp \left[-\Delta_{\sigma_j, \underline{y}}^2(t) \right] \right\}} \right\}^{-\psi} + 1 - w \right)^{-\frac{1}{\psi}}.$$

It is possible that we may devote future independent works to the investigation of a number of these binary and multivariate distributions, along with their application to binary and multivariate data in important journals like engineering, health, insurance, and actuarial sciences. This is as a result of the limitations that the current study imposes.

4. Some properties

It is essential to comprehend these mathematical properties for a number of reasons. The first advantage is that they enable us to construct a variety of statistical measures that can characterize the behavior of a random variable. For instance, the mean and variance can be used to describe the central tendency and variability of a probability distribution. Second, these characteristics enable us to forecast a random variable's future behavior. By comprehending the PMF or PDF, we may forecast the probability of specific events and base our judgements on that knowledge. Creating and evaluating statistical models requires a comprehension of these mathematical features. We build models that precisely mimic the behavior of real-life phenomena using probability distributions with well-defined mathematical features.

4.1 Ordinary moment

Moments are fundamentally important pieces of statistical equipment in a wide variety of fields, including physics, engineering, insurance, economics, and finance. It is a mathematical method that discloses information on the shape, position, and variability of a probability distribution. Specifically, it shows how the shape of the distribution varies. The p^{th} ordinary moment of $\mathcal{Y}$, say $\mu'_{p,y}$, follows from (5) as

$$\mu'_{p,y} = E(\mathcal{Y}^p) = \sum_{\tau, w=0}^{+\infty} \Omega_{\tau, w} E(Z_{\sigma^*}^p).$$

Henceforth, the RV Z_{σ^*} refers to the GZX-G family of probability distributions with power parameter σ^* where

$$E(Z_{\sigma^*}^p) = \sigma^* \int_{-\infty}^{+\infty} y^p \omega_{\underline{y}}(y)^{\sigma^*-1} \omega_{\underline{y}}(y) dy,$$

The quantity $E(Y_{\sigma^*}^p)$ can be numerically evaluated in terms of the baseline quantile function (QF) $Q_{\omega_{\underline{y}}}(u) = \omega_{\underline{y}}^{-1}(u)$ where

$$E(Z_{\sigma^*}^p) = \sigma^* \int_0^1 Q_{\omega_{\underline{y}}}(u)^p u^{\sigma^*-1} du.$$

Moreover, many statistical measures like the variance, skewness, and kurtosis measures can now be calculated via simple relations.

4.2 Moment generating function

The moment generating function, or MGF for short, is significant because it can offer detailed details about the characteristics of a probability distribution. One can precisely pinpoint the distribution's moments by employing the moment generating function. The mean, variance, skewness, and kurtosis are some of these moments. The characteristic function, the cumulant generating function, and the cumulant generating function are all statistical measures that are derived from the moment generating function. The MGF can then be calculated by using the equation below:

$$M_{\mathcal{Y}}(t) = \sum_{\tau, w=0}^{+\infty} \Omega_{\tau, w} M_{\sigma^*}(t),$$

where $M_{\sigma^*}(t)$ is the MGF of the RV Z_{σ^*} . That being said: the $M_{\mathcal{Y}}(t)$ can be easily derived after deriving the $M_{\sigma^*}(t)$ for any baseline model.

4.3 The incomplete moments

Moments and incomplete moments are both significant in terms of the information that can be gleaned about the properties of a probability distribution from the two types of moments. Moments might be whole or they can be missing something. They are helpful in finding, among other statistical measures, the mean, variance, skewness, and kurtosis of a distribution. Additionally, they are important in determining the shape of a distribution. In addition to this, you may make an estimate of the shape of the distribution by using them. The ζ^{th} incomplete moment, say $I_{\zeta, \mathcal{Y}}(t)$, of $\mathcal{Y}$ can be expressed from (5) as:

$$I_{\zeta, \mathcal{Y}}(t) = \int_{-\infty}^t y^{\zeta} f_{\mathcal{Y}}(y) dy = \sum_{\tau, w=0}^{+\infty} \Omega_{\tau, w} I_{-\infty}^t(y^{\zeta}; \sigma^*),$$

where

$$I_{-\infty}^t(y^{\zeta}; \sigma^*) = \int_{-\infty}^t y^{\zeta} \pi_{\sigma^*}(y) dy.$$

4.4 The residual life function and the reversed residual life function

The term "residual life" describes the amount of time that has passed from a particular moment in time or since the system reached a particular age without experiencing a failure. The n^{th} moment of the residual life, say $w_{n, \mathcal{Y}}(t) = E[(\mathcal{Y} - t)^n | \mathcal{Y} > t]$, $n = 1, 2, \dots$. The n^{th} moment of the residual life of $\mathcal{Y}$ is given by

$$w_{n, \mathcal{Y}}(t) = \frac{1}{1 - F_{\mathcal{Y}}(t)} \int_t^{+\infty} (y - t)^n dF(y).$$

Then,

$$w_{n, \mathcal{Y}}(t) = \frac{1}{1 - F_{\mathcal{Y}}(t)} \sum_{\tau, w=0}^{+\infty} \sum_{p=0}^n \Omega_{\tau, w} \binom{n}{p} (-t)^{n-p} I_t^{+\infty}(y^n; \sigma^*),$$

where $I_t^{+\infty}(y^n; \sigma^*) = \int_t^{+\infty} y^n \pi_{\sigma^*}(y) dy$. The events that occur during the residual life provide crucial insights into the functioning and dependability of systems. They are helpful in determining the remaining useful life of an item, organizing maintenance activities, evaluating hazards, establishing warranty durations, and making decisions based on accurate information. When firms take into account these periods in their dependability analyses, they are able to improve the performance of their systems, lower their expenses, and increase their overall operational efficiency. The remaining useful life of a system or component after a specific time point is referred to as the reversed residual life (RSRL), which is an abbreviation. In the process of reliability analysis, the moments of the RSRL play a crucial role, particularly in the context of survival analysis and deterioration modeling.

The n^{th} moment of the RSRL, say $M_{n, \mathcal{Y}}(t) = E[(t - \mathcal{Y})^n | \mathcal{Y} \leq t]$ for $t > 0$ and $n = 1, 2, \dots$, follows as $M_{n, \mathcal{Y}}(t) = \frac{1}{F_{\mathcal{Y}}(t)} \int_0^t (t - y)^n dF(y)$. Therefore, the n^{th} moment of the RSRL of $\mathcal{Y}$ becomes

$$M_n(t) = \frac{1}{F_{\mathcal{Y}}(t)} \sum_{\tau, w=0}^{+\infty} \sum_{p=0}^n \Omega_{\tau, w} (-1)^p \binom{n}{p} t^{n-p} I_0^t(y^n; \sigma^*),$$

where $I_0^t(y^n; \sigma^*) = \int_0^t y^n \pi_{\sigma^*}(y) dy$. The moments of the RSRL can be helpful in determining the pace of improvement in a system's dependability when it is increasing over time, such as when it is undergoing the burn-in phase or when it is improving as a result of learning effects. The evaluation of dependability growth models as well as the identification of prospective trends or anomalies can be made possible through the monitoring of changes in the moments.

4.5 Studying a special model

The inverse-Rayleigh (IR) model is without a doubt among the most important distributions that are utilized in the process of modelling extreme values. Fréchet (1927) is credited with making the first recommendation regarding the IR model. It has a wide range of applications, some of which are accelerated life tests, earthquakes, floods, wind speed, horse racing, showers, lineups at the grocery store, and sea waves (for further information, see Von Mises (1964) and Kotz and Johnson (1992)). One can obtain additional information regarding the IR model by reading the relevant literature. Nadarajah and Kotz (2003), for example, investigated the generalized IR distribution. Please refer to Jahanshahi et al. (2019), Krishna et al. (2013), and Al-Babtain et al. (2020) for additional information regarding pertinent IR extensions. The RV $\mathcal{Y}$ is claimed to have the continuous IR model if its RF is given by

$$R_a(\mathcal{Y}) = 1 - \varpi_{a_1, a_2}(\mathcal{Y}) = \exp \left[- \left(\frac{1}{\mathcal{Y}} a \right)^2 \right] \Big|_{\mathcal{Y} \geq 0},$$

where $a > 0$ is a scale parameter of the base line IR model. The CDF of the geometrically generated Rayleigh Inverse-Rayleigh (GCGR-IR) model can be defined as follows, with reference to equation (3):

$$F_{\underline{\mathcal{U}}}(\mathcal{Y}) = \frac{\gamma - \gamma \exp \left(- \left\{ \exp \left[\sigma \left(\frac{1}{\mathcal{Y}} a \right)^2 \right] - 1 \right\}^{-2} \right)}{1 - (1 - \gamma) \left[1 - \exp \left(- \left\{ \exp \left[\sigma \left(\frac{1}{\mathcal{Y}} a \right)^2 \right] - 1 \right\}^{-2} \right) \right]} \Big|_{\mathcal{Y} \geq 0, \gamma > 0, \sigma > 0},$$

where $\underline{\mathcal{U}} = (\gamma, \sigma, a)$ refers to the parameters vector. For the **GCGR-IR** model, we obtain the following results:

$$\begin{aligned} \mu'_{p, \mathcal{Y}} &= E(\mathcal{Y}^p) = a^p \sum_{\tau, w=0}^{+\infty} \Omega_{\tau, w} \sigma^{* \frac{p}{2}} \Gamma \left(1 - \frac{1}{2} p \right) \Big|_{a_2 > p}, \\ \phi_{1, \mathcal{Y}}(t) &= a \sum_{\tau, w=0}^{+\infty} \Omega_{\tau, w} \sigma^{* \frac{1}{2}} \gamma \left(1 - \frac{1}{2}, \left(\frac{1}{t} a \right)^2 \right) \Big|_{a_2 > 1}, \\ w_{n, \mathcal{Y}}(t) &= \frac{1}{R_{\underline{\mathcal{U}}}(t)} \sum_{\tau, w=0}^{+\infty} \sum_{p=0}^n \Omega_{\tau, w} \binom{n}{p} (-t)^{n-p} a^n \sigma^{* \frac{n}{2}} \Gamma \left(1 - \frac{n}{2}, \left(\frac{1}{t} a \right)^2 \right) \Big|_{2 > n}, \\ M_{n, \mathcal{Y}}(t) &= \frac{1}{F_{\underline{\mathcal{U}}}(t)} \sum_{\tau, w=0}^{+\infty} \sum_{p=0}^n \Omega_{\tau, w} (-1)^p \binom{n}{p} t^{n-p} a^n \sigma^{* \frac{n}{2}} \gamma \left(1 - \frac{n}{2}, \left(\frac{1}{t} a \right)^2 \right) \Big|_{2 > n}, \end{aligned}$$

where

$$\Gamma(w) = \int_0^{+\infty} q^{w-1} \exp(-q) dq,$$

and $\gamma(w, u)$ refers to the lower incomplete gamma function, where

$$\gamma(w, u) = \int_0^u q^{w-1} \exp(-q) dq = \sum_{\kappa=0}^{+\infty} u^{w+\kappa} \frac{(-1)^\kappa}{\kappa! (w + \kappa)}.$$

On the other hand, the function

$$\Gamma(w, u)|_{u>0} = \int_u^{+\infty} q^{w-1} \exp(-q) dq = \Gamma(w, u) = \Gamma(w) - \gamma(w, u),$$

refer to the upper incomplete gamma function.

5. The maximum likelihood (MLE) method

One of the most popular techniques used to estimate unknown parameters in statistical models is the maximum likelihood approach. The likelihood function, which measures the likelihood that the provided data will be observed in line with the predicted model, can be used to determine the parameter values that will maximise this likelihood. It provides a methodical approach to doing so. For the aim of generating parameter estimates in a range of statistical

models, the method known as maximum likelihood estimation, or MLE, is often employed and typically adaptable. Efficiency is one of the ideal characteristics of estimates based on the greatest likelihood. The maximum likelihood technique yields estimate that are asymptotically efficient, which means that they achieve the lowest asymptotic variance among all reliable estimators. This is so that the maximum likelihood technique may make use of the data that is already accessible. To put it another way, compared to estimates produced by other estimators, the maximum likelihood estimates tend to be quite close to the actual values of the parameters being estimated. They also have a comparatively low level of variability.

To obtain the MLE of $\underline{\mathbb{U}}$, the log-likelihood function can be expressed as

$$\begin{aligned} \ell(\underline{\mathbb{U}}) = & n \log(2) + n \log(\gamma) + n \log(\sigma) + \sum_{i=0}^n \log \omega_{\underline{\mathbb{V}}}(\mathcal{Y}_i) + (2\sigma - 1) \sum_{i=0}^n \log \varpi_{\underline{\mathbb{V}}}(\mathcal{Y}_i) \\ & - 3 \sum_{i=0}^n \log[1 - \varpi_{\underline{\mathbb{V}}}^{\sigma}(\mathcal{Y}_i)] - \sum_{i=0}^n \Delta_{\sigma, \underline{\mathbb{V}}}^2(\mathcal{Y}_i) - 2 \sum_{i=0}^n \log T_i(\mathcal{Y}_i), \end{aligned}$$

where

$$T_i(\mathcal{Y}_i) = \left\{ 1 - (1 - \gamma) \left[1 - \exp\left(-\Delta_{\sigma, \underline{\mathbb{V}}}^2(\mathcal{Y})\right) \right] \right\}.$$

The function $\ell(\underline{\mathbb{U}})$ can be numerically maximized either directly by using softwares like the R (using the optim function), the SAS (using the PROC NLMIXED) or the Ox program (by using the sub-routine MaxBFGS) or analytically via solving the nonlinear equations of the likelihood function which obtained by differentiating $\ell(\underline{\mathbb{U}})$. The maximum likelihood method has a close connection to information theory. The likelihood function can be interpreted as a measure of the information contained in the data about the unknown parameters. Maximizing the likelihood is equivalent to maximizing the information extracted from the data. This connection to information theory has led to several developments in statistical theory and provides a theoretical foundation for the maximum likelihood method.

6. Simulation study and assessing the estimation method

In this section, we would like to highlight the significance of simulation by presenting a full simulation study. Additionally, we would want to evaluate the estimators produced by the MLE approach using specific statistical criteria. When evaluating the effectiveness of the MLE technique, simulation studies are an extremely important component. The following is a list of the most essential reasons why simulations are necessary in this setting:

I. We are able to compare the estimated values that were derived by MLE with the actual parameter values in a controlled environment thanks to the power of simulations. We are able to determine the bias of the estimator by performing the simulation process an innumerable number of times. The systematic difference between the values that are estimated and the values that actually exist is referred to as bias. Researchers are able to identify and address any biases that may be present because to the support that simulations provide in understanding how MLE functions in a variety of settings and with varying sample sizes.

II. The term "efficiency" refers to the accuracy and dependability of the estimator in determining the true values of the parameters being estimated. Simulations offer a method for determining whether or not MLE is effective by analyzing the degree to which estimated values vary across a number of different iterations of the simulation. Researchers are able to better comprehend the accuracy of the estimations and evaluate the MLE's performance under a variety of circumstances as a result of this.

III. Consistency is a desired quality in an estimator, which indicates that as the sample size rises, the estimated values should converge to the true parameter values. In other words, consistency means that an estimator should always produce the same results. By producing data sets of varied sizes and determining whether the estimated values approach the true values as the sample size expands, simulations can assist in verifying the reliability of

MLE. This can be done by determining whether the estimated values approach the true values. The consistency of the MLE is a basic characteristic, and simulations are helpful in validating the validity of this property.

IV. Researchers are able to test the robustness of MLE by subjecting it to breaches of the assumptions that underpin the estimation approach by using simulations. Simulations allow for the evaluation of how sensitive MLE is to various kinds of violations by purposefully introducing deviations from the anticipated model circumstances. Analysis of robustness assists in the identification of potential constraints and directs researchers in the selection of acceptable estimating methods or, when necessary, in the modification of the model's underlying assumptions.

V. Simulations offer a foundation upon which MLE can be compared to many other estimating approaches. Researchers are able to evaluate the relative effectiveness of various methods of estimate, such as the method of moments and Bayesian estimation, by simulating data sets and subjecting them to predetermined conditions during the simulation process. When compared to other methods, simulations assist identify the benefits and drawbacks of MLE, which in turn makes it easier to choose the technique that is going to be the most useful for any specific situation.

The following main algorithms are performed to assess the performance of the estimation method under the new compound G family and the IR baseline model:

- 1) Using the inversion method, where $y_u = F^{-1}_{\underline{u}}(y)$, we generate $N=500$ samples of size n from the GCGR-IR distribution;
- 2) Compute the MLEs for each sample of the 500 samples
- 3) Compute the standard errors of estimations (SErEs) of the MLEs for the 1000 samples. The SErEs were computed by the well-known method of inverting our obtained information matrix.
- 4) Compute the biases ($B_{\underline{u}}(n)$) and mean squared errors (MSEs) given for all parameters where $\underline{u} = \gamma, \sigma, \alpha$.
- 5) Repeated steps 1-4 for $n = 50, 60, \dots, 500$ with an initial value for all parameters equal to one, then compute biases, mean squared errors ($MSE_{\underline{u}}(n)$) for γ, σ, α for all $n = 50, 60, \dots, 500$.

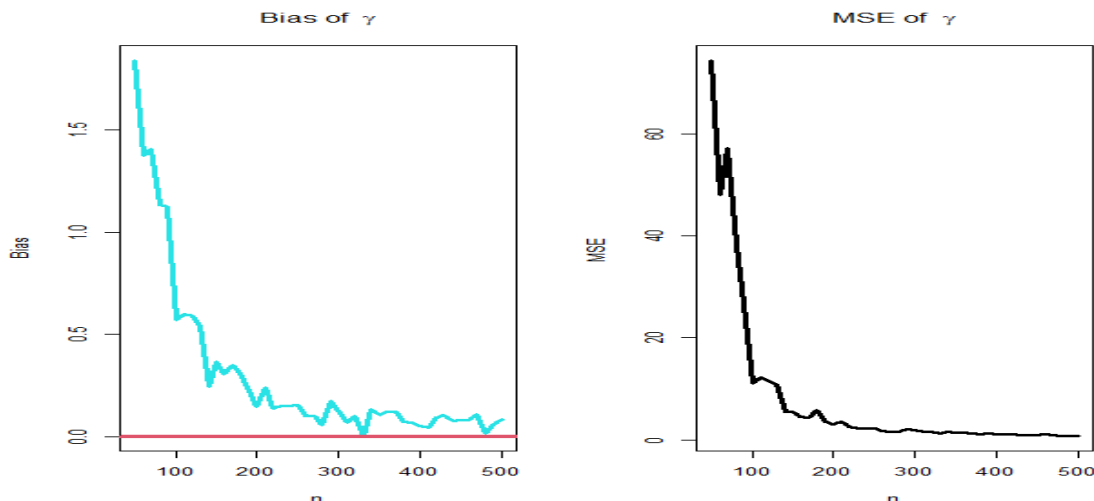


Figure 1: Biases (the left simulation graph) and MSEs (the right simulation graph) for γ and $n = 50, 60, \dots, 500$ for the GCGR-IR model.

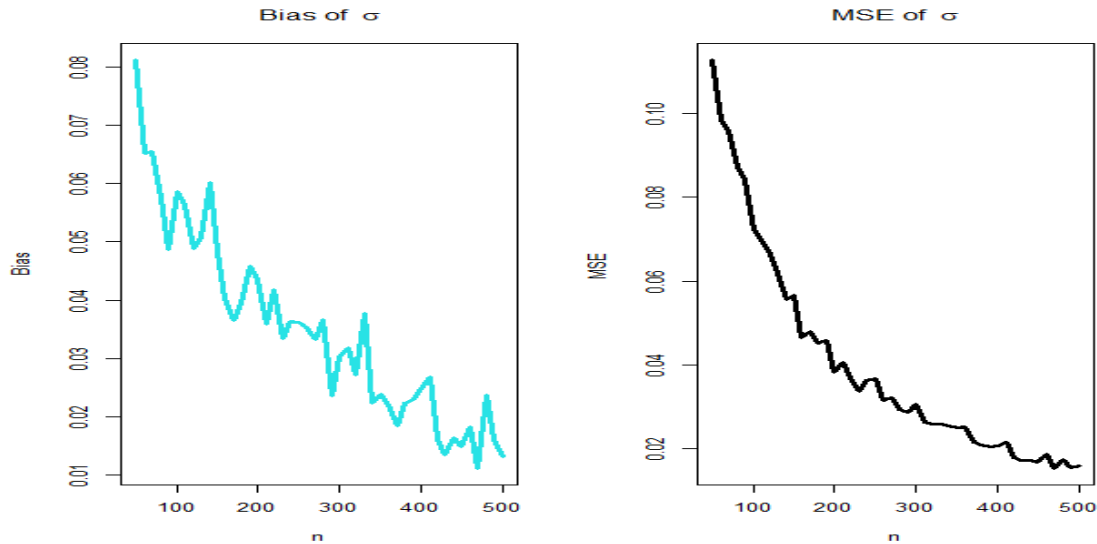


Figure 2: Biases (the left simulation graph) and MSEs (the right simulation graph) for σ and $n = 50, 60, \dots, 500$ for the GCGR-IR model.

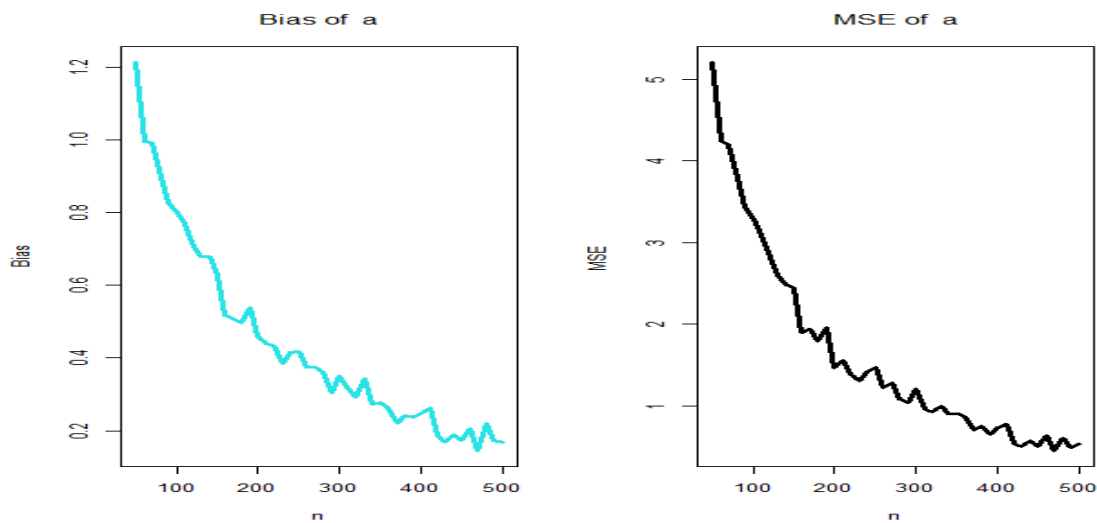


Figure 3: Biases (the left simulation graph) and MSEs (the right simulation graph) for α and $n = 50, 60, \dots, 500$ for the GCGR-IR model.

The graphs that may be found to the left of Figures 1, 2 and 3 each demonstrate the variation of the three types of bias with sample size. The link between the three MSEs and the sample size is depicted in the graphs located on the right-hand side of Figures 1, 2 and 3. Figure 1 shows a broken red line where the biases are when they are all equal to zero. This line represents the biases. As n increases, the biases for each parameter move closer and closer to zero, as can be seen in Figures 1, 2 and 3, and the MSEs for each parameter likewise move closer and closer to zero as n increases.

7. Real data modeling for comparing the competing models

One of the most significant events that takes place within the realm of statistics is something that is referred to as a competition of probability distributions. Researchers are able to make inferences about the features of the population as a result of this activity, which helps to a better knowledge of the process of data gathering and enables them to do so. The selection of the ideal probability distribution, on the other hand, can be a difficult task, especially in circumstances in which there are numerous possible distributions that might be applied to a single dataset. The importance of using actual data models cannot in any way be understated when it comes to the process of comparing various probability distributions applicable to this setting. One of the jobs involved in real-life data modelling is comparing the degree to which various probability distributions are a good fit for the actual datasets that are being modelled. Using this strategy, researchers are able to identify the optimal distribution for a certain dataset and draw inferences about the characteristics of the population. The modelling of real data is vital for a variety of reasons, and each of these reasons is unique in its own way.

Real data modelling is a crucial instrument for comparing and studying a variety of probability distributions because it is based on actual data. Researchers have the option to select the ideal distribution for the data set at hand, arrive at inferences on the parameters of the population, and evaluate the adequacy of the fitted distribution. This opportunity is made available to them thanks to the availability of the data. When researchers employ actual data modelling, they are able to draw more accurate and reliable conclusions regarding the population parameters and the underlying data generating process. This is because actual data modelling takes into account all of the relevant variables. Table 1 reports some competitive models.

Table 1: Some competing models with authors and its corresponding abbreviations.

Model (Author(s)); Abbreviation
Inverse- Rayleigh; RR
Kumaraswamy-Inverse- Rayleigh (Mead and Abd-Eltawab (2014)); KUM-IR
Beta-Inverse- Rayleigh (Barreto-Souza et al. (2011)); Beta-IR
Generalized-inverse- Rayleigh (Nadarajah and Kotz (2003)); GZ-IR
Marshal-Olkin-Inverse- Rayleigh (Krishna et al. (2013)); MRO-IR
McDonald-Inverse- Rayleigh (Shahbaz et al. (2016)); MCD-IR
Transmuted-Inverse- Rayleigh (Mahmoud and Mandouh (2013)); TR-IR
odd log-logistic-Inverse Rayleigh; OLOGL-IR
odd loglogistic generalized-Inverse- Rayleigh; OLLGZ-IR
odd loglogistic Inverse Rayleigh; OLOGL-IR
generalized odd loglogistic- Inverse Rayleigh; GOLOGL-IR

Quantitative analysis, visual inspection, and hybrid methods of both are all viable options for assessing real-world datasets. The "nonparametric Kernel PDF estimation (NK-PDFE)" method for investigating the initial density shape, the "Quantile-Quantile (QN-QN)" plot for investigating the "normality" of the data, the "total time in test (TTT)" plot for investigating the initial shape of the empirical HRFs (for more information, see Aarset (1987)), and the "box plot" for investigating the extremes will all be considered.

7.1 Modeling the reliability stress data

The first raw data set is 100 measurements of carbon fiber breaking stress (in Gba) provided by Nichols and Padgett (2006). Figure 4 displays a box plot (right-hand panel), NK-PDFE plot (left-hand panel), TTT plot (right-hand panel), and QN-QN plot (left-hand panel). According to the data for carbon fiber breaking stress, shown in the first row of the left panel of Figure 4, the distribution is asymmetric bimodal with a right heavy tail. Figure 4's first-row right panel shows that the HRF has been rising consistently over the past few years. The extreme values in this data are readily

apparent in the left and right panels of Figure 4's second row. The third panel in the right-hand row of Figure 4 shows that the current data cannot be characterized by theoretical distributions such the normal, uniform, exponential, logistic, beta, lognormal, and Weibull distributions. Table 2 shows the results of the Kolmogorov-Smirnov test, the Cramér-von Mises statistic, the Anderson-Darling statistic, and the P-value for all of the fitted models. The MLEs and associated SEs are tabulated in Table 3. Table 2 shows that the GCGR-IR model has the most conservative estimates, with $CRVM = 0.061251$, $ANDR = 0.444536$, $KG-SM = 0.057898$, and $P_y = 0.888821$. Because of this, we can choose the GCGR-IR as our preferred model. Figure 5 displays the projected PDF and CDF. The P-P plot and estimated HRF for the current data set are displayed in Figure 6. The revised GCGR-IR model provides satisfactory fits to the experimental functions, as shown in Figures 6 and 7.

Table 2: CRVM, ANDR, KG-SM and P_y for the breaking stress of carbon fibers data.

Criteria→	Goodness of fit criteria			
Distribution↓	CRVM	ANDR	KG-SM	P_y
GCGR-IR	0.061251	0.444536	0.057898	0.888821
OLOGL-IR	0.155353	1.212059	0.655329	< 0.0010
OB-IR	0.066555	0.470644	0.063040	0.822025
MCD-IR	0.133336	1.060777	0.080748	0.533255
OLLGZ-IR	0.120335	0.963936	0.556334	< 0.0010
TR-IR	0.087117	0.620935	0.078216	0.573455
GOLOGL-IR	0.155039	1.212032	0.655233	< 0.0010
GZ-IR	0.109436	0.765675	0.08732	0.428665
RR	0.109044	0.765777	0.0871175	0.428239
KUM-IR	0.081333	0.621768	0.076017	0.611840
MRO-IR	0.088650	0.614260	0.076285	0.516771
Beta-IR	0.081029	0.621314	0.075755	0.614686

Table 3: MLEs and SEs for the breaking stress of carbon fibers data.

Estimates →	Estimates			
Distribution↓	$\hat{\gamma}$	$\hat{\sigma}$	$\hat{\theta}$	$\hat{a}$
GCGR-IR	221.2132 (31.5334)	0.329843 (0.01759)		71.12528 (68.9222)
GCGR-IR	5.195422 (0.00346)	0.599335 (0.03247)		1.077454 (0.04444)
OLLGZ-IR	0.135942 (0.00132)		3.72056 (0.00147)	0.92909 (0.30332)
GOLOGL-IR	0.494683 (0.041436)		0.067494 (0.71959)	1.742613 (9.30073)
OLOGL-IR	0.494589 (0.04132)		0.452429 0.038691	
RR				2.396381 (0.00362)
KUM-IR		0.848944 (16.08346)	1.623929 (0.69791)	1.634104 (9.09417)
GZ-IR		0.939474		1.416949

		(3.54346)		(2.56849)
Beta-IR		0.736632	1.583035	1.668443
		(1.52966)	(0.71324)	(0.76623)
TR-IR	-0.733661			1.265562
	(0.261692)			(0.05799)
MRO-IR	1.2419435	0.003327		6.229612
	(0.118145)	(0.00098)		(1.01341)
MCD-IR	0.850347	44.42347	19.85939	0.203438
	(0.135376)	(25.1072)	(6.70639)	(0.00638)

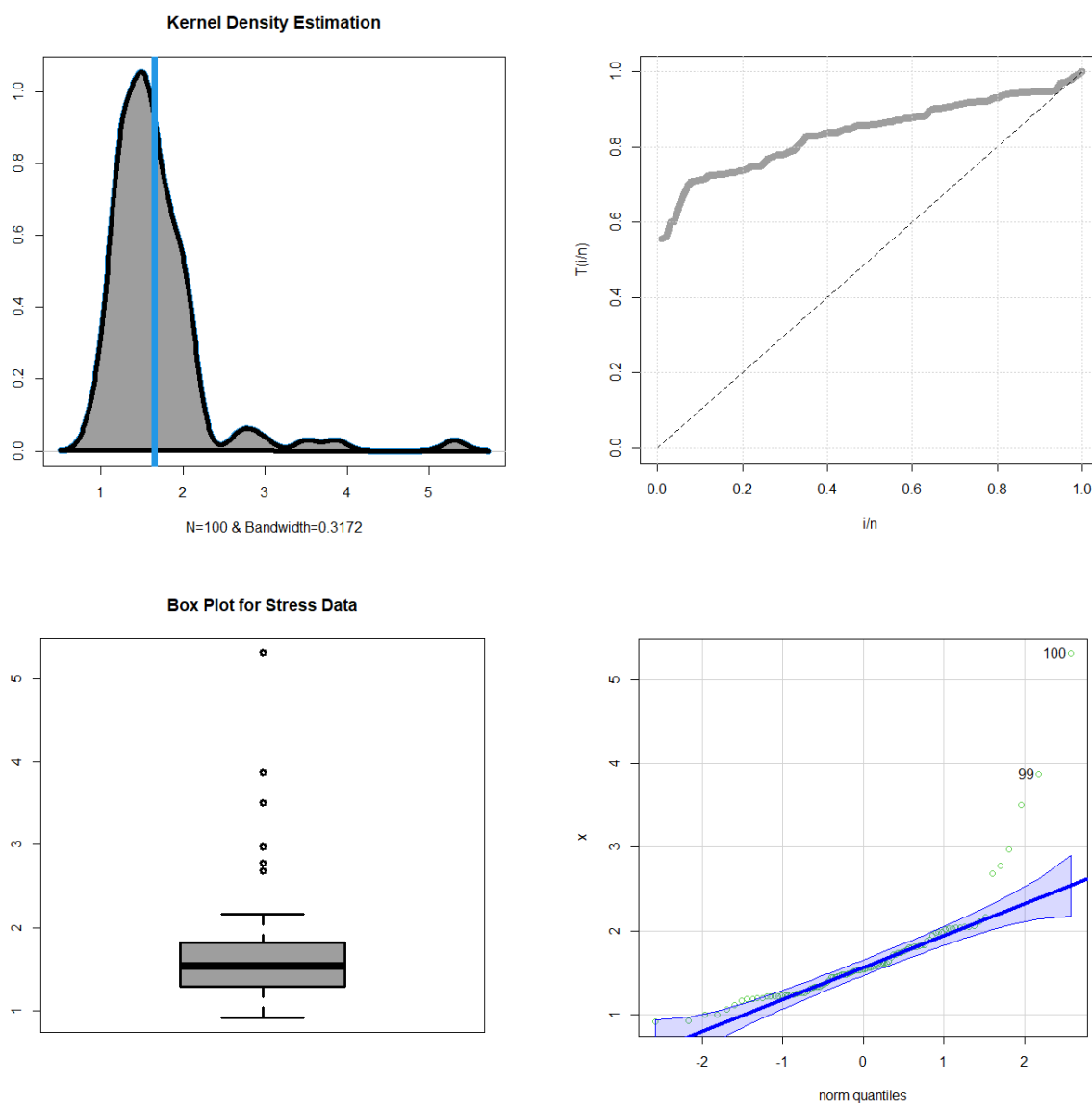


Figure 4: Graphical description for the breaking stress of carbon fibers data.

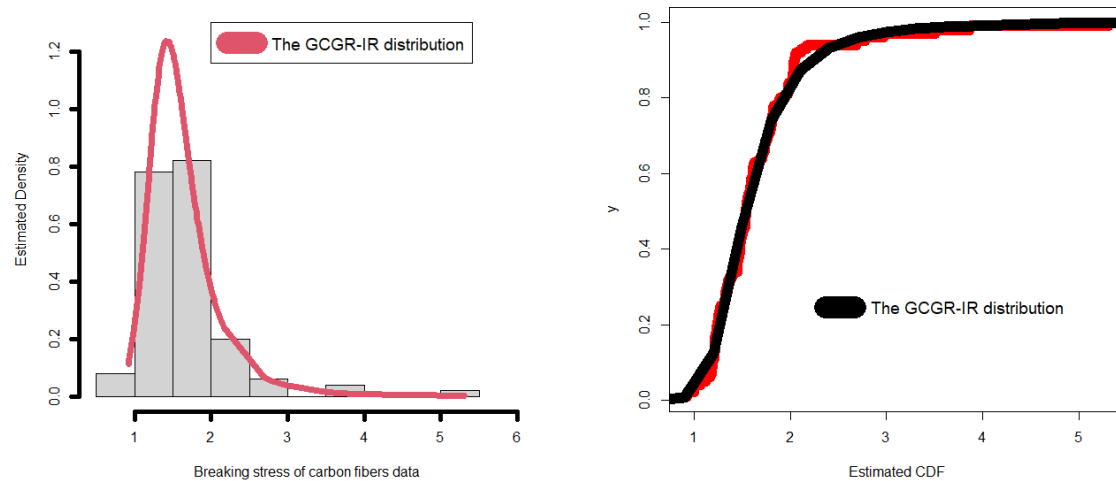


Figure 5: Estimated PDF (left application plot) and estimated CDF (right application plot) for the breaking stress of carbon fibers data.

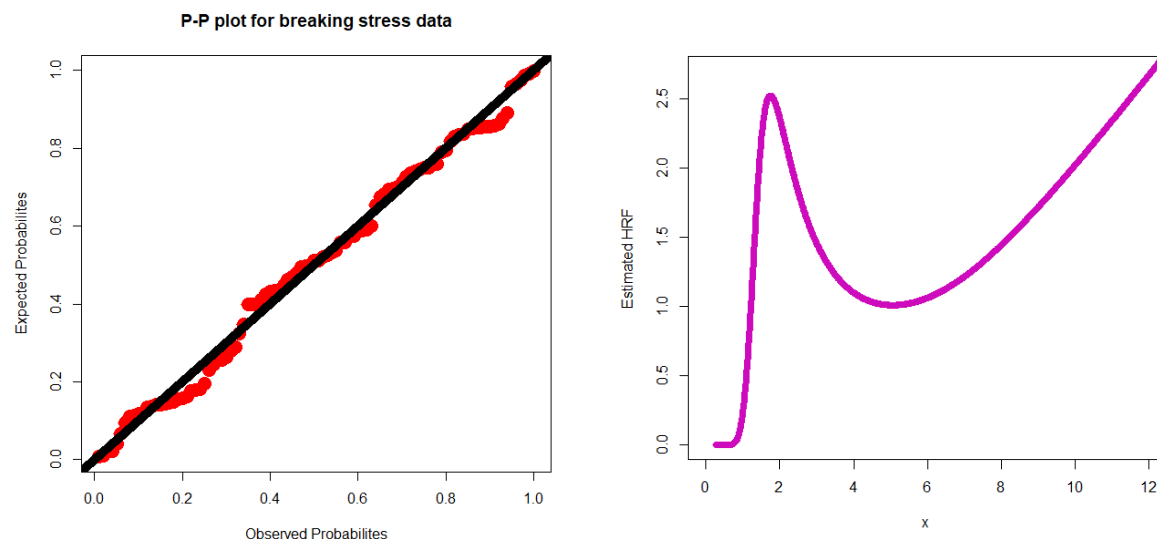


Figure 6: P-P plot (left application plot) and estimated HRF (right application plot) for the breaking stress of carbon fibers data.

7.2 Modeling the reliability glass fibers dataset

The second group of numbers pertains to the glass fiber strengths as described by Smith and Naylor (1987). Box plot, QN-QN plot, NK-PDFE plot, and TTT plot are shown in Figure 7's second row left panel, second row right panel, and first row left panel, respectively. In the first left panel of Figure 7, we can see that the data for glass fibers exhibits an asymmetric bimodal distribution with a right heavy tail. The HRF of glass fibers exhibits a monotonically increasing trend, as shown in the first row, right panel of Figure 7. As may be seen in the panels labeled "second row left" and "second row right" of Figure 7, the statistics for glass fibers include some extreme values. Panel (right third row) of Figure 7 demonstrates that the glass fiber data cannot be explained by the normal, uniform, exponential, logistic, beta, lognormal, or Weibull distributions.

The CRVM, ANDR, KG-SM, and P_v statistics for all fitted models are shown in Table 4. The MLEs and corresponding SEs are shown in Table 5. Table 4 shows that when compared to other models, the CRVM (0.114399), ANDR (0.895333), KG-SM (0.003421), and $P_v = (700635)$ provided by the GCGR-IR model are the lowest. As a result, we can choose the GCGR-IR distribution as our preferred model. Figure 8 displays the projected PDF and CDF. Figure 9 displays the P-P plot and estimated HRF for the glass fiber data. The GCGR-IR model provides a good fit to the empirical functions, as seen in Figures 9 and 10.

Table 4: CRVM, ANDR; KG-SM and P_v for the glass fibers data.

Criteria→ Distribution ↓	Goodness of fit criteria			
	CRVM	ANDR	KG-SM	P_v
GCGR-IR	0.114399	0.895333	0.003421	0. 700635
IR	0.083243	0.695665	0.054245	0. 3422114
OLLGZ-IR	0.104832	0.83665	0.551954	< 0.00100
GOLOGL-IR	0.150665	1.146675	0.679111	< 0.00100
OLOGL-IR	0.151541	1.146119	0.679434	< 0.00100

Table 5: MLEs and SEs for the glass fibers data.

Estimates → Distribution ↓	Estimates			
	$\hat{\gamma}$	$\hat{\sigma}$	$\hat{\theta}$	$\hat{\alpha}$
GCGR-IR	10.34411 (6.22321)	0.0056300 (0.011242)		53.55433 (1.32130)
IR				56.65521 (2.00031)
OLLGZ-IR	0.1449432 (0.01288)1		0.008733 (0.00048)	1.299344 (0.00049)
GOLOGL-IR	0.5025554 (0.052877)		0.071343 (1.13655)	1.7048342 (13.47432)
OLOGL-IR	0.5025122 0.0529461		0.455933 0.048655	

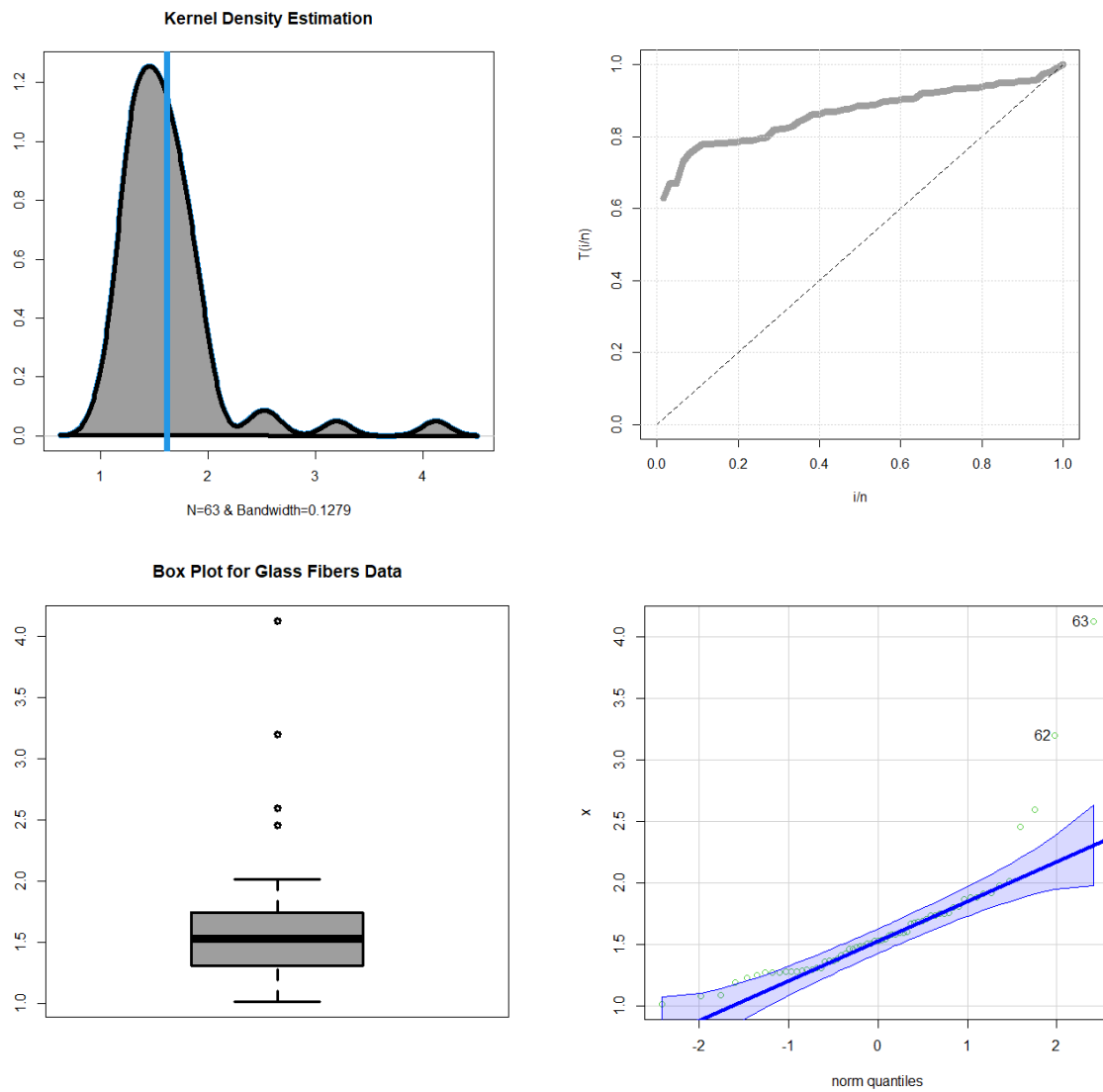


Figure 7: Graphical description for the glass fibers data.

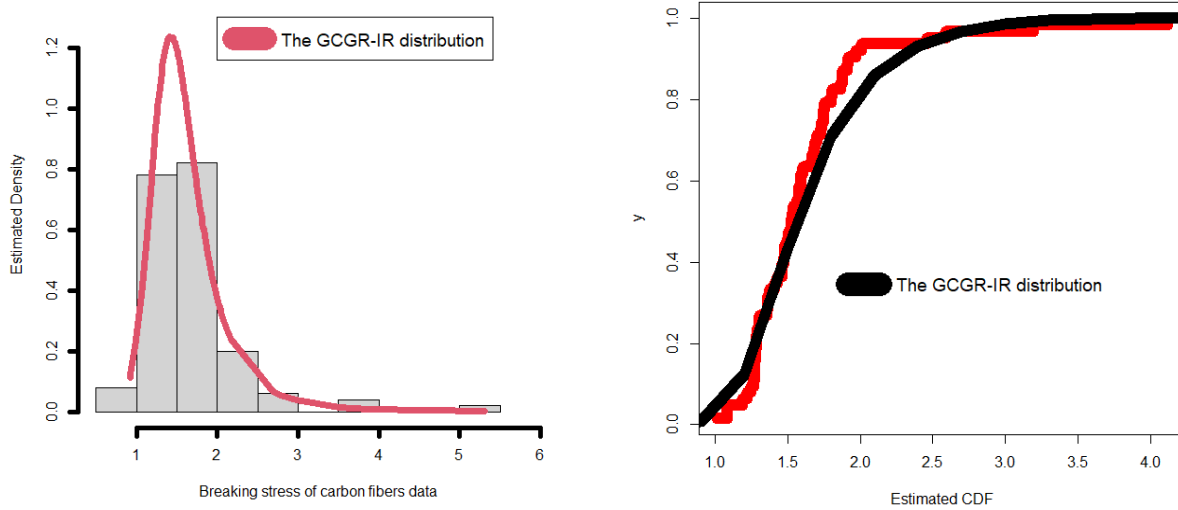


Figure 8: Estimated PDF (left application plot) and CDF (right application plot) for the glass fibers data.

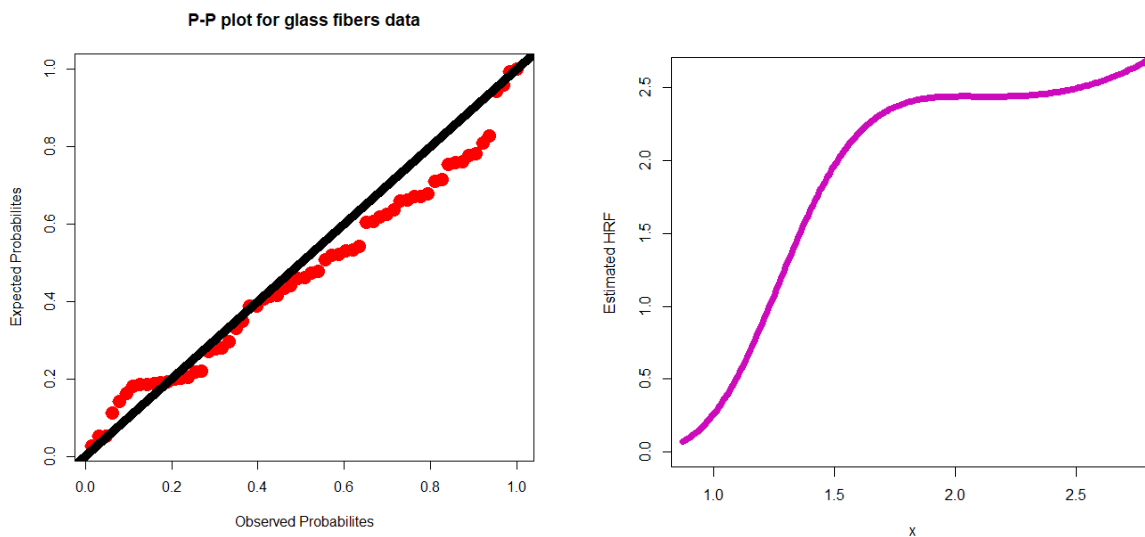


Figure 9: P-P plot (left application plot) and estimated HRF (right application plot) for the glass fibers data.

7.3 Modeling the medical relief times data

The "Wingo data" is the third type of information, and it features the reported alleviation times (in hours) for 50 arthritis patients during a clinical trial. Figure 10 displays a box plot (second panel from the left), NK-PDFE plot (first panel from the left), TTT plot (first panel from the right), and QN-QN plot (second panel from the right). In the first panel on the left of Figure 10, the data for the relief times are presented as symmetric data. The HRF for this data is continuously increasing, as shown in the right panel of the first row in Figure 10. Figure 1's panels for the second row, both left and right, show that the relief times have no outlying values. As shown in the third panel from the right of Figure 10, theoretical distributions including the normal, uniform, exponential, logistic, beta, lognormal, and Weibull cannot account for the relief times.

Table 6 displays the CRVM, ANDR, KG-SM, and P_ values for all of the fitted models. Table 7 details the MLEs and their corresponding SEs. From Table 6, the minimum values for the GCGR-IR model are: CRVM =0.048111, ANDR

=0.401121, KG-SM =0.082471, and P_v =0.900632. Therefore, the GCGR-IR may be selected as the best model. The predicted PDF and CDF are shown in Figure 11. For the data collected during the relief periods, the P-P plot and estimated HRF are shown in Figure 12. Figures 12 and 13 show that the suggested GCGR-IR model provides a good fit to the empirical functions.

Table 6: CRVM, ANDR, KG-SM and P_v for the relief times data.

Criteria→ Distribution ↓	Goodness of fit criteria			
	CRVM	ANDR	KG-SM	P_v
GCGR-IR	0.048111	0.401121	0.082471	0.900632
OB-IR	0.049436	0.42043	0.091266	0.799986
GZ-IR	0.323553	2.030133	0.150688	0.206420
GOLOGL-IR	0.195399	1.349776	0.110001	0.579731
Beta-IR	0.367871	2.513100	0.143344	0.360133
IR	0.323343	2.030323	0.150654	0.206621
OLLGZ-IR	0.157889	1.098854	0.5349543	< 0.0010
TR-IR	0.282090	1.815288	0.137999	0.3045188

Table 7: MLEs and SEs for the relief times data.

Estimates → Distribution ↓	Estimates			
	$\hat{\gamma}$	$\hat{\sigma}$	$\hat{\theta}$	$\hat{\alpha}$
GCGR-IR	0.979543 (0.112198)	0.044489 (0.10154)		31.066812 (21.87953)
OB-IR	10.79053 (0.06675)	6.115531 (4.09556)		0.126864 (0.000235)
GOLOGL-IR	1.96135 (0.23493)	0.111288 (0.00166)		1.412344 (0.005466)
OLLGZ-IR	0.066933 (0.00761)		0.221445 (0.00279)	0.35554 (0.00480)
IR				0.485959 (0.02279)
GZ-IR			0.904841 (18.78439)	0.50151 (3.24447)
Beta-IR		4.0153244 (0.111866)	1.3345431 (0.147803)	2.002210 (0.32101)
TR-IR	-0.581660 (0.2787265)			0.440030 (0.929001)

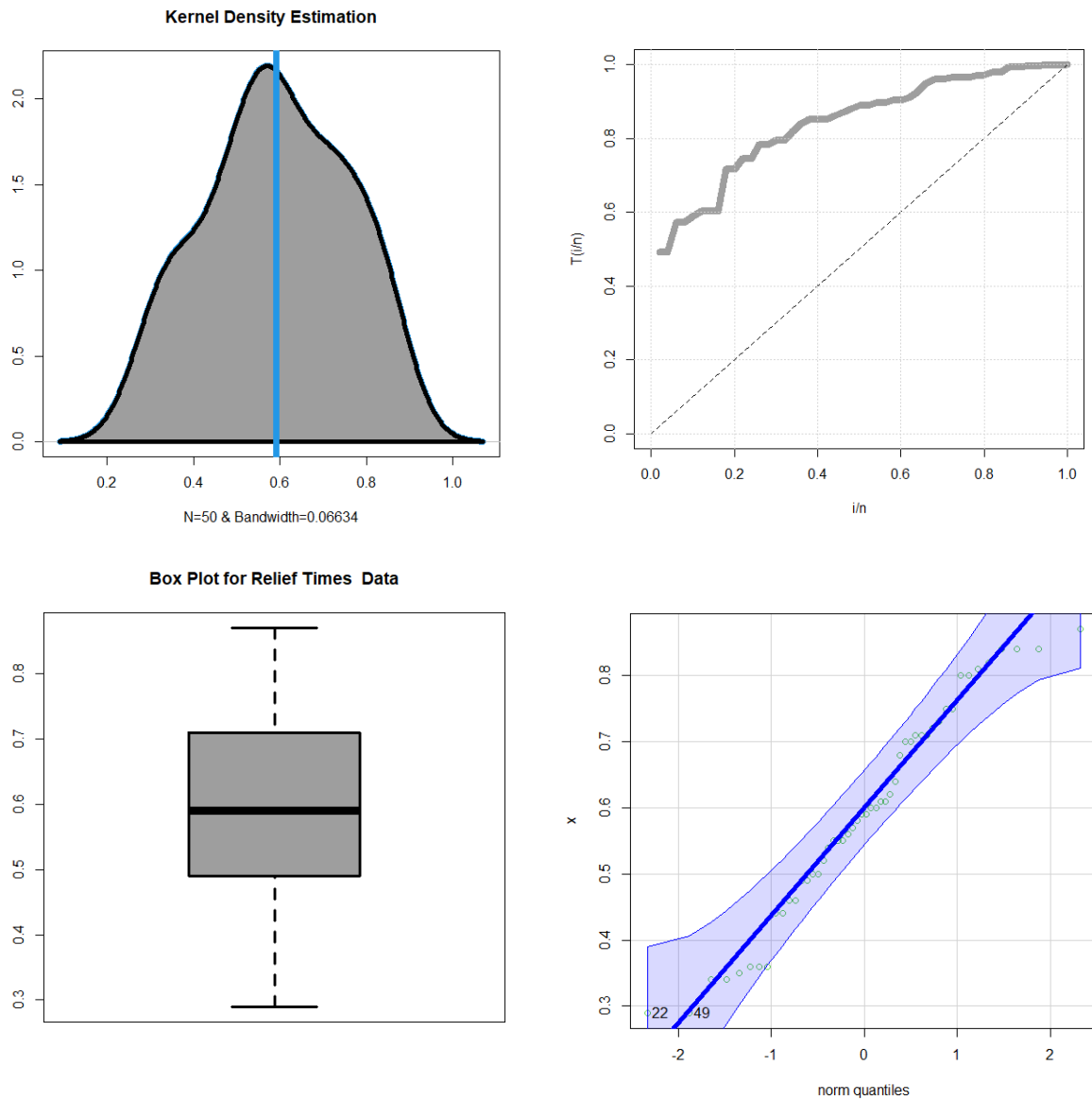


Figure 10: Graphical description for the relief times data.

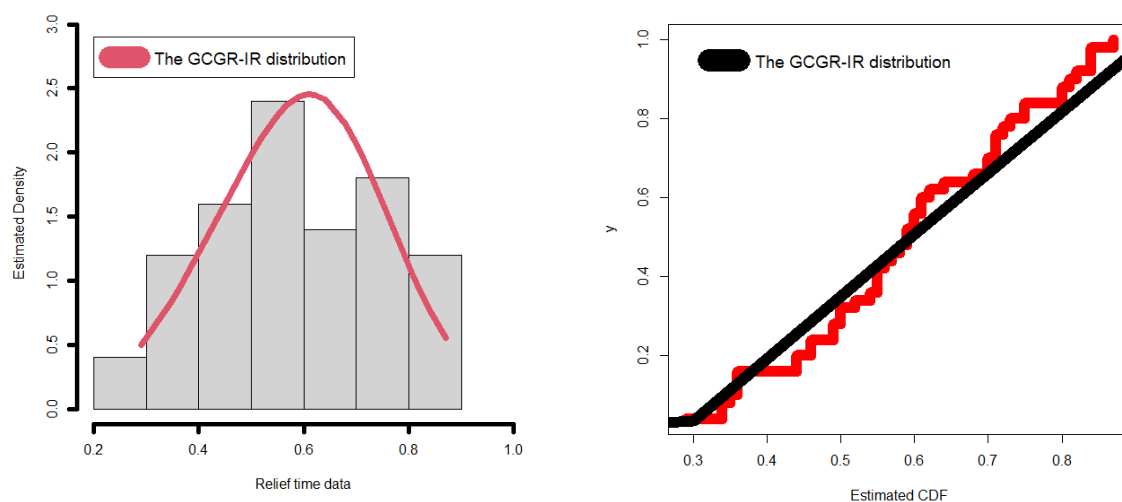


Figure 11: Estimated PDF (left application plot) and estimated CDF (right application plot) for the relief times data.

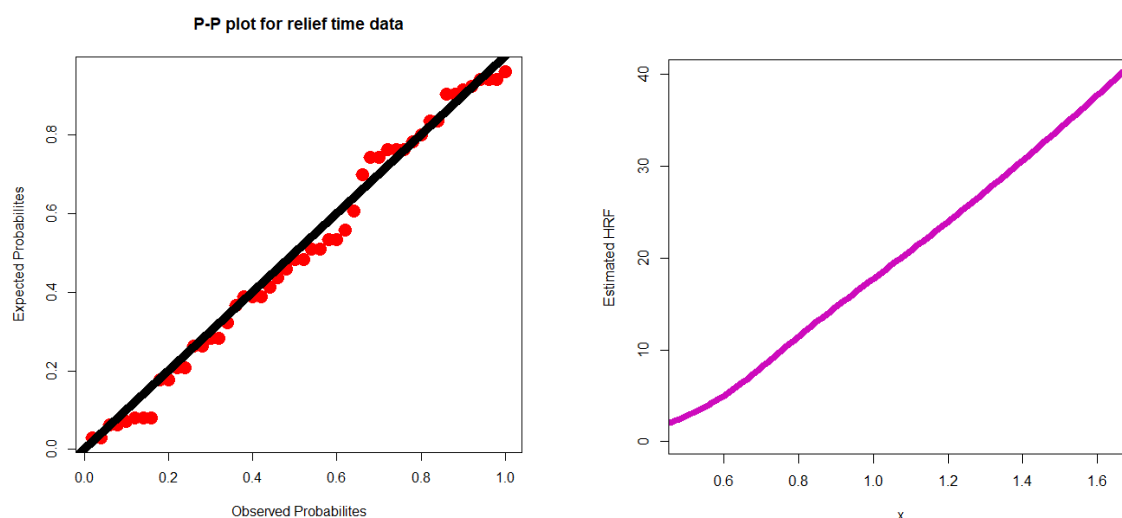


Figure 12: P-P plot (left application plot) and estimated HRF (right application plot) for the relief times data.

8. Conclusions

In this work, a compound G distribution with two parameters that is completely new to the family was introduced as a member of the compound G distributions. It is possible to deduce essential statistical properties, such as the generating function, ordinary moments, and incomplete moments. It is possible to accomplish this. An inquiry has been conducted specifically into the Inverse-Rayleigh model, which serves as the benchmark for comparison with other models. When making an estimation of the characteristics of the new family, the approach that has the best possibility of producing accurate results is the one that is selected. For the purpose of making it easier to compare the many different approaches to estimating, numerical simulations are run through each of them. In order to produce a number of distributions that were bivariate as well as multivariate, we relied on the copula method. These brand-new distributions have the potential to be of great assistance when it comes to modeling data that has both bivariate and multivariate components. In addition to this, three different actual data sets are used in order to assess and compare the various estimating approaches. The flexibility and importance of the purposed family of products are demonstrated by the use of three separate applications to observable data. The following examples demonstrate how the new G family performs better than some of the G families that have been around for a longer period of time:

- I. The Cramér-von Mises statistic, the Anderson-Darling statistic, the Kolmogorov-Smirnov test statistic, and the accompanying p-value all suggest that the new family performs better than the odd Burr G family, the odd log-logistic G family, and the odd log-logistic G family. Additionally, all of these statistics indicate that the new family performs better than the odd Burr G family. Modeling the breaking stress of carbon fibers requires the use of the generalized G family, the odd, transmuted G family, the generalized G family, the McDonald G family, the Marshall-Olkin G family, and the Beta G family. This is according to the Cramér-von Mises statistic, the Anderson-Darling statistic, and the Kolmogorov-Smirnov test.
- II. When it comes to modeling glass fibers, the purposed family performs better than the odd log-logistic G family and the odd log-logistic generalized G family, according to the Cramér-von Mises statistic, the Anderson-Darling statistic, the Kolmogorov-Smirnov test statistic, and its accompanying p-value. This was determined by comparing the purposed family to the Anderson-Darling statistic. The comparison of the purposed family to the other two families allowed for the discovery of this result.
- III. The performance of the new family is superior to that of the odd Burr G family, the generalized odd log-logistic G family, the odd log-logistic generalized G family, the generalized G family, the Beta G family, and the transmuted G family when it comes to modeling the relief timings. The Cramer-von Mises statistic, the Anderson-Darling statistic, the Kolmogorov-Smirnov test statistic, and its accompanying p-value all provide evidence for this assertion.

Based on the above applications we conclude that:

- I. Compound G families of continuous probability distributions allow us to model and describe real-world data accurately. By identifying the distribution that best fits the data, we gain insights into its underlying behavior and characteristics. This can be crucial for understanding the data's central tendency, variability, and potential outliers.
- II. Compound families of probability distributions enable us to perform statistical inference and estimate parameters. This includes tasks like hypothesis testing, confidence intervals, and maximum likelihood estimation. These techniques help us draw conclusions about the population from which the data is sampled.
- III. In many fields like finance, insurance, and engineering, understanding and quantifying risk is essential. Compound G families of continuous probability distributions can be used to model risks and estimate probabilities of extreme events, which are crucial for risk assessment and management.
- IV. Compound G families of continuous probability distributions can be integrated into decision-making processes. They help assess uncertainty and make more informed decisions, particularly in situations where outcomes are uncertain, and risks need to be considered.
- V. Probability distributions are fundamental in forecasting and prediction tasks. By fitting historical data to an appropriate distribution, we can generate probabilistic forecasts for future events or outcomes.
- VI. In quality control and manufacturing, compound continuous probability distributions are used to monitor and control production processes. They help identify deviations and anomalies in the manufacturing process.
- VII. In engineering and reliability studies, compound probability distributions are used to model failure times and lifetimes of products or systems. This information is crucial for design improvements and maintenance planning.

Here are some future research points in this area:

- I. Investigate the extension of the novel geometrically generated Rayleigh family to multivariate distributions. Develop and explore new multivariate distributions with applications in fields like finance, environmental modeling, and engineering.
- II. Apply Bayesian methods for parameter estimation and inference in the context of the new geometric Rayleigh family and copula models. Study the advantages of Bayesian techniques in handling uncertainty.
- III. Explore the capabilities of the proposed models in modeling extreme events, such as natural disasters, financial crises, or rare diseases. Investigate the tail behavior of the distributions and copulas.
- IV. Apply the new models to time series data. Investigate their ability to capture temporal dependencies and trends. Develop time series versions of the proposed distributions and copulas.
- V. Conduct robustness and sensitivity analyses to evaluate how the proposed models perform under different data conditions, parameter settings, and model assumptions.
- VI. Extend the models to incorporate covariates or explanatory variables. Analyze how these factors affect the distributions and dependence structures in real-life data.

- VII.** Investigate the use of the new distributions and copulas for model-based clustering and classification of data. Explore applications in customer segmentation, image analysis, or ecological modeling.
 - VIII.** Study and develop novel dependence measures within the context of copulas. Explore how these measures can provide insights into the relationships between variables in real-life data.
 - IX.** Apply the models to risk assessment and management in areas like insurance, investment, and environmental risk. Evaluate their performance in estimating value at risk (VaR) and conditional value at risk (CVaR).
 - X.** Investigate the use of non-Gaussian copulas (e.g., Archimedean, t-copulas) in combination with the novel geometrically generated Rayleigh family. Assess their performance in modeling data with non-normal dependencies.
 - XI.** Develop user-friendly software tools or R packages that implement the proposed models, making them accessible to a broader audience of researchers and practitioners.
 - XII.** Explore applications of models in emerging fields such as blockchain technology, cybersecurity, and social network analysis, where understanding dependencies and modeling real-life data are of increasing importance.
 - XIII.** Foster collaboration with experts from diverse fields to apply the models to new and unexplored areas, providing insights and solutions to real-world problems.
 - XIV.** Investigate ethical implications related to data modeling and the use of probabilistic models in decision-making, especially in fields like healthcare, finance, and criminal justice.
 - XV.** Develop educational resources and outreach programs to disseminate knowledge about the new models and their applications, fostering a wider understanding of their benefits and limitations.
 - XVI.** A Chi-squared type test for distributional censored and uncensored validity with numerical assessments and real data applications within the context of the new compounded family.
 - XVII.** The geometric generated Rayleigh family for the value at risk analysis with insurance applications (see Ibrahim et al. (2023), Yousof et al. (2023a-f)).
 - XVIII.** The bivariate geometric generated Rayleigh family (Mansour et al. (2020a-f), Elgohari and Yousof (2020a), Shehata and Yousof (2021, 2022), Shehata et al. (2021, 2022) and Elgohari et al. (2020)).
 - XIX.** The geometric generated Rayleigh family for modeling bimodal data.
 - XX.** The geometric generated Rayleigh family for modeling heavy tailed data.
 - XXI.** Characterizing the geometric generated Rayleigh family (see Bhatti et al. (2018, 2019, 2020 and 2022)).
 - XXII.** The geometric generated Rayleigh family for modeling the actuarial data.
 - XXIII.** The geometric generated Rayleigh family under some acceptance sampling plans. (see Ahmed and Yousof (2022), Ahmed et al. (2022a,b) and Tashkandy et al. (2023)).
 - XXIV.** The geometric generated Rayleigh family with censored regression modeling (Altun et al. (2018a-e)).
 - XXV.** The geometric generated Rayleigh family with application in reliability, insurance, economic and medicine (see Elsayed and Yousof (2019a,b), Elsayed and Yousof (2020 and 2021), Korkmaz et al. (2017, 2018, 2019a, 2019b and 2022)).
 - XXVI.** The geometric generated Rayleigh family with right censored Bayesian and Non-Bayesian distributional validation (see Emam et al. (2023b)).
 - XXVII.** Validation of geometric generated Rayleigh family via Nikulin-Rao-Robson goodness-of-fit test under complete and censored applications (Yadav et al. (2020, 2021, 2022)).
 - XXVIII.** Generalized stress-strength and generalized multicomponent stress-strength models under the geometric generated Rayleigh family (see Rasekhi et al. (2022), Saber and Yousof (2022) and Saber et al. (2022a,b,c)).
- These future research points aim to advance the understanding and application of the novel geometrically generated Rayleigh family of distributions and copulas, contributing to a wide range of fields where data modeling is critical.

References

1. Aboraya, M., M. Yousof, H. M., Hamedani, G. G. and Ibrahim, M. (2020). A new family of discrete distributions with mathematical properties, characterizations, Bayesian and non-Bayesian estimation methods. *Mathematics*, 8, 1648.
2. Ahmed, B. and Yousof, H. M. (2022). A New Group Acceptance Sampling Plans based on Percentiles for the Weibull Fréchet Model. *Statistics, Optimization & Information Computing*, forthcoming.
3. Ahmed, B., Ali, M. M. and Yousof, H. M. (2022a). A Novel G Family for Single Acceptance Sampling Plan with Application in Quality and Risk Decisions, *Annals of Data Science*, forthcoming.
4. Ahmed, B., Chesneau, C. Ali, M. M. and Yousof, H. M. (2022b). Amputated Life Testing for Weibull Reciprocal Weibull Percentiles: Single, Double and Multiple Group Sampling Inspection Plans with Applications, *Pakistan Journal of Statistics*

- and Operation Research, 18(4), 995-1013.
5. Aidi, K., Butt, N. S., Ali, M. M., Ibrahim, M., Yousof, H. M. and Shehata, W. A. M. (2021). A Modified Chi-square Type Test Statistic for the Double Burr X Model with Applications to Right Censored Medical and Reliability Data. *Pakistan Journal of Statistics and Operation Research*, 17(3), 615-623.
 6. Afify, A. Z., Alizadeh, M., Yousof, H.M., Aryal, G. and Ahmad, M. (2016a). The transmuted geometric-G family of distributions: theory and applications. *Pakistan Journal of Statistics*, 32, 139-160.
 7. Afify, A. Z., Cordeiro, G.M., Nadarajah, S., Yousof, H.M., Ozel, G., Nofal, Z.M. and Altun, E. (2016b). The complementary geometric transmuted-G family of distributions: model, properties and application. *Hacettepe Journal of Mathematics and Statistics*, 47, 1348-1374.
 8. Al-Babtain, A.A., Elbatal, I. and Yousof, H.M. (2020). A new three parameter Fréchet model with mathematical properties and applications. *Journal of Taibah University for Science*, 14, 265–278.
 9. Altun, E., Yousof, H. M. and Hamedani G. G. (2018a). A new flexible extension of the generalized half-normal lifetime model with characterizations and regression modeling. *Bulletin of Computational Applied Mathematics*, 6(1), 83-115.
 10. Altun, E., Yousof, H. M. and Hamedani, G. G. (2018b). A new generalization of generalized half-normal distribution: properties and regression models. *Journal of Statistical Distributions and Applications*, 5(1), 7.
 11. Altun, E., Yousof, H. M. and Hamedani, G. G. (2018c). A new log-location regression model with influence diagnostics and residual analysis. *Facta Universitatis, Series: Mathematics and Informatics*, 33(3), 417-449.
 12. Altun, E., Yousof, H. M. and Hamedani G. G. (2018d). A Flexible Extension of Generalized Half-Normal Distribution: Characterizations and Regression Models. *International Journal of Applied Mathematics and Statistics*, 57(3), 27-49.
 13. Altun, E., Yousof, H. M., Chakraborty, S. and Handique, L. (2018e). Zografos-Balakrishnan Burr XII distribution: regression modeling and applications. *International Journal of Mathematics and Statistics*, 19(3), 46-70.
 14. Bhatti, F. A., Hamedani, G. G., Yousof, H. M., Ali, A., & Ahmad, M. (2018). On Modified Burr XII-Inverse Exponential Distribution: Properties, Characterizations and Applications. *J Biostat Biom: JBSB-106*. DOI, 10.
 15. Bhatti, F. A., Hamedani, G. G., Korkmaz, M.C., Yousof, H. M., & Ahmad, M. (2020). On the new modified Burr XII distribution: development, properties, characterizations and applications, *Journal of data science*, forthcoming.
 16. Bhatti, F. A., Hamedani, G. G., Korkmaz, M. C., Cordeiro, G. M., Yousof, H. M., & Ahmad, M. (2019). On Burr III Marshal Olkin family: development, properties, characterizations and applications. *Journal of Statistical Distributions and Applications*, 6(1), 1-21.
 17. Bhatti, F. A., Hamedani, G. G., Yousof, H. M., Ali, A., & Ahmad, M. (2020). On Modified Burr XII-Inverse Weibull Distribution: Properties and Applications. *Pakistan Journal of Statistics and Operation Research* 16(4),721-735.
 18. Bhatti, F. A., Hamedani, G. G., Yousof, H. M., & Ali, A. (2022). On the Burr III-Moment Exponential Distribution. *Thailand statistician*, Vol. 20 No. 3, 615-635.
 19. Aryal, G.R. and Yousof, H.M. (2017). The exponentiated generalized-G Poisson family of distributions. *Stochastics and Quality Control*, 32, 7-23.
 20. Chesneau, C., Yousof, H. M., Hamedani, G. and Ibrahim, M. (2022). A New One-parameter Discrete Distribution: The Discrete Inverse Burr Distribution: Characterizations, Properties, Applications, Bayesian and Non-Bayesian Estimations. *Statistics, Optimization & Information Computing*, forthcoming.
 21. Cordeiro, G.M. and de Castro, M. (2011). A new family of generalized distributions. *Journal of statistical computation and simulation*, 81, 883-898.
 22. Cordeiro, G.M., Ortega E.M.M., Nadarajah, S. (2010). The Kumaraswamy Weibull distribution with application to failure data. *Journal of The Franklin Institute*, 347, 1399–1429.
 23. Cordeiro, G.M., Ortega, E. M., Popović, B.V. and Pescim, R.R. (2014). The Lomax generator of distributions: Properties, minification process and regression model. *Applied Mathematics and Computation*, 247, 465-486.
 24. Elgohari, H. and Yousof, H. M. (2020a). A Generalization of Lomax Distribution with Properties, Copula and Real Data Applications. *Pakistan Journal of Statistics and Operation Research*, 16(4), 697-711. <https://doi.org/10.18187/pjsor.v16i4.3260>
 25. Elgohari, H. and Yousof, H. M. (2021b). A New Extreme Value Model with Different Copula, Statistical Properties and Applications. *Pakistan Journal of Statistics and Operation Research*, 17(4), 1015-1035. <https://doi.org/10.18187/pjsor.v17i4.3471>
 26. Elgohari, H. and Yousof, H. M. (2020c). New Extension of Weibull Distribution: Copula, Mathematical Properties and Data Modeling. *Statistics, Optimization & Information Computing*, 8(4), 972-993. <https://doi.org/10.19139/soic-2310-5070-1036>,
 27. Elgohari, H., Ibrahim, M. and Yousof, H. M. (2021). A New Probability Distribution for Modeling Failure and Service Times: Properties, Copulas and Various Estimation Methods. *Statistics, Optimization & Information Computing*, 8(3), 555-586.
 28. Elsayed, H. A., & Yousof, H. M. (2019a). A new Lomax distribution for modeling survival times and taxes revenue data sets. *Journal of Statistics and Applications*, 2(1), 35-58.
 29. Elsayed, H. A. H. and Yousof, H. M. (2019b). The Burr X Nadarajah Haghighi distribution: statistical properties and application to the exceedances of flood peaks data. *Journal of Mathematics and Statistics*, 15, 146-157.
 30. Elsayed, H. A. H., & Yousof, H. M. (2020). The generalized odd generalized exponential Fréchet model: univariate, bivariate and multivariate extensions with properties and applications to the univariate version. *Pakistan Journal of Statistics and Operation Research*, 529-544.
 31. Elsayed, H. A. H. and Yousof, H. M. (2021). Extended Poisson Generalized Burr XII Distribution. *Journal of Applied Probability and Statistics*, 16(1), 01-30.

32. Emam, W.; Tashkandy, Y.; Hamedani, G.G.; Shehab, M.A.; Ibrahim, M.; Yousof, H.M. (2023a). A Novel Discrete Generator with Modeling Engineering, Agricultural and Medical Count and Zero-Inflated Real Data with Bayesian, and Non-Bayesian Inference. *Mathematics* 2023, 11, 1125. <https://doi.org/10.3390/math11051125>
33. Emam, W.; Tashkandy, Y.; Goual, H.; Hamida, T.; Hiba, A.; Ali, M.M.; Yousof, H.M.; Ibrahim, M. (2023b). A New One-Parameter Distribution for Right Censored Bayesian and Non-Bayesian Distributional Validation under Various Estimation Methods. *Mathematics* 2023, 11, 897. <https://doi.org/10.3390/math11040897>
34. Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Pol. Math.*, 6, 93–116.
35. Ibrahim, M.; Emam, W.; Tashkandy, Y.; Ali, M.M.; Yousof, H.M. (2023). Bayesian and Non-Bayesian Risk Analysis and Assessment under Left-Skewed Insurance Data and a Novel Compound Reciprocal Rayleigh Extension. *Mathematics* 2023, 11, 1593. <https://doi.org/10.3390/math11071593>
36. Jahanshahi, S.M.A., Yousof, H.M. and Sharma, V.K. (2019). The Burr X Fréchet Model for Extreme Values: Mathematical Properties, Classical Inference and Bayesian Analysis. *Pakistan Journal of Statistics and Operation Research*, 15, 797.
37. Kim, J.H. and Jeon, Y. (2013). Credibility theory based on trimming. *Insur. Math. Econ.* 53, 36–47.
38. Korkmaz, M. Ç., Altun, E., Chesneau, C. and Yousof, H. M. (2022). On the unit-Chen distribution with associated quantile regression and applications. *Mathematica Slovaca*, 72 (2022), No. 3, 765–786.
39. Korkmaz, M. Ç., Altun, E., Yousof, H. M., Afify, A. Z., & Nadarajah, S. (2018). The Burr X Pareto Distribution: Properties, Applications and VaR Estimation. *Journal of Risk and Financial Management*, 11(1), 1.
40. Korkmaz, M. Ç., Altun, E., Yousof, H. M., & Hamedani, G. G. (2019a). The odd power Lindley generator of probability distributions: properties, characterizations and regression modeling. *International Journal of Statistics and Probability*. ,(2)8 89-70.
41. Korkmaz, M. Ç., Yousof, H. M., & Ali, M. M. (2017). Some theoretical and computational aspects of the odd Lindley Fréchet distribution. *İstatistikçiler Dergisi: İstatistik ve Aktüerya*, 10(2), 129–140.
42. Korkmaz, M. C., Yousof, H. M., Alizadeh, M., & Hamedani, G. G. (2019b). The Topp-Leone generalized odd log-logistic family of distributions: properties, characterizations and applications. *Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics*, 68(2), 1506–1527.
43. Kotz, S. and Shanbhag, D.N. (1980). Some new approach to probability distributions, *Adv. Appl. Probab.* 12, 903–921.
44. Kotz, S.; Johnson, N.L. (1992). *Breakthroughs in Statistics: Foundations and Basic Theory*; Springer-Verlag, Volume 1.
45. Krishna, E.; Jose, K.K., Alice, T. and Risti, M.M. (2013). The Marshall-Olkin Fréchet distribution. *Commun. Stat. Theory Methods*, 42, 4091–4107.
46. Mansour, M. M., Ibrahim, M., Aidi, K., Butt, N. S., Ali, M. M., Yousof, H. M., & Hamed, M. S. (2020a). A New Log-Logistic Lifetime Model with Mathematical Properties, Copula, Modified Goodness-of-Fit Test for Validation and Real Data Modeling. *Mathematics*, 8(9), 1508.
47. Mansour, M. M., Butt, N. S., Ansari, S. I., Yousof, H. M., Ali, M. M., & Ibrahim, M. (2020b). A new exponentiated Weibull distribution's extension: copula, mathematical properties and applications. *Contributions to Mathematics*, 1 (2020) 57–66. DOI: 10.47443/cm.2020.0018
48. Mansour, M., Korkmaz, M. Ç., Ali, M. M., Yousof, H. M., Ansari, S. I., & Ibrahim, M. (2020c). A generalization of the exponentiated Weibull model with properties, Copula and application. *Eurasian Bulletin of Mathematics*, 3(2), 84–102.
49. Mansour, M., Rasekhi, M., Ibrahim, M., Aidi, K., Yousof, H. M., & Elrazik, E. A. (2020d). A New Parametric Life Distribution with Modified Bagdonavičius–Nikulin Goodness-of-Fit Test for Censored Validation, Properties, Applications, and Different Estimation Methods. *Entropy*, 22(5), 592.
50. Mansour, M., Yousof, H. M., Shehata, W. A. M., & Ibrahim, M. (2020e). A new two parameter Burr XII distribution: properties, copula, different estimation methods and modeling acute bone cancer data. *Journal of Nonlinear Science and Applications*, 13(5), 223–238.
51. Mansour, M. M., Butt, N. S., Yousof, H. M., Ansari, S. I., & Ibrahim, M. (2020f). A Generalization of Reciprocal Exponential Model: Clayton Copula, Statistical Properties and Modeling Skewed and Symmetric Real Data Sets. *Pakistan Journal of Statistics and Operation Research*, 16(2), 373–386.
52. Marinho, P. R. D., Silva, R. B., Bourguignon, M., Cordeiro, G. M., & Nadarajah, S. (2019). AdequacyModel: An R package for probability distributions and general-purpose optimization. *PloS one*, 14, e0221487.
53. Nadarajah, S. and Kotz, S. (2003). The exponentiated Fréchet distribution. *Interstate Electronic Journal*, 14, 1–7.
54. Rasekhi, M., Saber, M. M., & Yousof, H. M. (2020). Bayesian and classical inference of reliability in multicomponent stress-strength under the generalized logistic model. *Communications in Statistics-Theory and Methods*, 50(21), 5114–5125.
55. Saber, M. M. and Yousof, H. M. (2022). Bayesian and Classical Inference for Generalized Stress-strength Parameter under Generalized Logistic Distribution, *Statistics, Optimization & Information Computing*, forthcoming.
56. Saber, M. M., Hamedani, G. G., Yousof, H. M. But, N. S., Ahmed, B. and Ibrahim, M. (2022a) A Family of Continuous Probability Distributions: Theory, Characterizations, Properties and Different Copulas, CRC Press, Taylor & Francis Group.
57. Saber, M. M. Marwa M. Mohie El-Din and Yousof, H. M. (2022b). Reliability estimation for the remained stress-strength model under the generalized exponential lifetime distribution, *Journal of Probability and Statistics*, 2021, 1–10.
58. Saber, M. M., Rasekhi, M. and Yousof, H. M. (2022c). Generalized Stress-Strength and Generalized Multicomponent Stress-Strength Models. *Statistics, Optimization & Information Computing*, forthcoming.
59. Shehata, W. A. M. and Yousof, H. M. (2022). A novel two-parameter Nadarajah-Haghighi extension: properties, copulas, modeling real data and different estimation methods. *Statistics, Optimization & Information Computing*, 10(3), 725–749.
60. Shehata, W. A. M. and Yousof, H. M. (2021). The four-parameter exponentiated Weibull model with Copula, properties and

- real data modeling. *Pakistan Journal of Statistics and Operation Research*, 17(3), 649-667.
61. Shehata, W. A. M., Yousof, H. M., & Aboraya, M. (2021). A Novel Generator of Continuous Probability Distributions for the Asymmetric Left-skewed Bimodal Real-life Data with Properties and Copulas. *Pakistan Journal of Statistics and Operation Research*, 17(4), 943-961. <https://doi.org/10.18187/pjsor.v17i4.3903>
 62. Shehata, W. A. M., Butt, N. S., Yousof, H., & Aboraya, M. (2022). A New Lifetime Parametric Model for the Survival and Relief Times with Copulas and Properties. *Pakistan Journal of Statistics and Operation Research*, 18(1), 249-272.
 63. Tashkandy, Y., Emam, W., Ali, M. M., Yousof, H. M., Ahmed, B. (2023). Quality Control Testing with Experimental Practical Illustrations under the Modified Lindley Distribution Using Single, Double, and Multiple Acceptance Sampling Plans. *Mathematics*. 2023; 11(9):2184. <https://doi.org/10.3390/math11092184>
 64. Von Mises, R. (1964). *Selected papers of Richard von Mises*. American Mathematical Society, 1.
 65. Yadav, A. S., Altun, E., & Yousof, H. M. (2021). Burr–Hatke Exponential Distribution: A Decreasing Failure Rate Model, Statistical Inference and Applications. *Annals of Data Science*, 8(2), 241–260.
 66. Yadav, A. S., Goual, H., Alotaibi, R. M., Rezk, H., Ali, M. M., & Yousof, H. M. (2020). Validation of the Topp-Leone-Lomax model via a modified Nikulin-Rao-Robson goodness-of-fit test with different methods of estimation. *Symmetry*, 12(1), 57.
 67. Yadav, A. S., Shukla, S., Goual, H., Saha, M. and Yousof, H. M. (2022). Validation of xgamma exponential model via Nikulin-Rao-Robson goodness-of- fit test under complete and censored sample with different methods of estimation. *Statistics, Optimization & Information Computing*, 10(2), 457-483.
 68. Yousof, H. M., Aidi, K., Hamedani, G. G., & Ibrahim, M. (2023a). Chi-squared type test for distributional censored and uncensored validity with numerical assessments and real data applications. *Japanese Journal of Statistics and Data Science*, 1-30.
 69. Yousof, H. M., Ali, M. M., Aidi, K., Ibrahim, M. (2023b). The modified Bagdonavičius-Nikulin goodness-of-fit test statistic for the right censored distributional validation with applications in medicine and reliability. *Statistics in Transition New Series*, 24(4), forthcoming.
 70. Yousof, H. M., Ansari, S. I., Tashkandy, Y., Emam, W., Ali, M. M., Ibrahim, M., Alkhayyat, S. L. (2023c). Risk Analysis and Estimation of a Bimodal Heavy-Tailed Burr XII Model in Insurance Data: Exploring Multiple Methods and Applications. *Mathematics*. 2023; 11(9):2179. <https://doi.org/10.3390/math11092179>
 71. Yousof, H.M.; Emam, W.; Tashkandy, Y.; Ali, M.M.; Minkah, R.; Ibrahim, M. (2023d). A Novel Model for Quantitative Risk Assessment under Claim-Size Data with Bimodal and Symmetric Data Modeling. *Mathematics* 2023, 11, 1284. <https://doi.org/10.3390/math11061284>
 72. Yousof, H. M., Goual, H., Emam, W., Tashkandy, Y., Alizadeh, M., Ali, M. M., & Ibrahim, M. (2023e). An Alternative Model for Describing the Reliability Data: Applications, Assessment, and Goodness-of-Fit Validation Testing. *Mathematics*, 11(6), 1308.
 73. Yousof, H.M.; Tashkandy, Y.; Emam, W.; Ali, M.M.; Ibrahim, M. (2013f). A New Reciprocal Weibull Extension for Modeling Extreme Values with Risk Analysis under Insurance Data. *Mathematics* 2023, 11, 966. <https://doi.org/10.3390/math11040966>