Austrian Journal of Statistics

AUSTRIAN STATISTICAL SOCIETY

Volume 47, Number 1, 2018

ISSN: 1026597X, Vienna, Austria

Österreichische Zeitschrift für Statistik

ÖSTERREICHISCHE STATISTISCHE GESELLSCHAFT

Austrian Journal of Statistics; Information and Instructions

GENERAL NOTES

The Austrian Journal of Statistics is an open-access journal with a long history and is published approximately quarterly by the Austrian Statistical Society. Its general objective is to promote and extend the use of statistical methods in all kind of theoretical and applied disciplines. Special emphasis is on methods and results in official statistics.

Original papers and review articles in English will be published in the Austrian Journal of Statistics if judged consistently with these general aims. All papers will be refereed. Special topics sections will appear from time to time. Each section will have as a theme a specialized area of statistical application, theory, or methodology. Technical notes or problems for considerations under Shorter Communications are also invited. A special section is reserved for book reviews.

All published manuscripts are available at

http://www.ajs.or.at

(old editions can be found at http://www.stat.tugraz.at/AJS/Editions.html)

Members of the Austrian Statistical Society receive a copy of the Journal free of charge. To apply for a membership, see the website of the Society. Articles will also be made available through the web.

PEER REVIEW PROCESS

All contributions will be anonymously refereed which is also for the authors in order to getting positive feedback and constructive suggestions from other qualified people. Editor and referees must trust that the contribution has not been submitted for publication at the same time at another place. It is fair that the submitting author notifies if an earlier version has already been submitted somewhere before. Manuscripts stay with the publisher and referees. The refereeing and publishing in the Austrian Journal of Statistics is free of charge. The publisher, the Austrian Statistical Society requires a grant of copyright from authors in order to effectively publish and distribute this journal worldwide.

OPEN ACCESS POLICY

This journal provides immediate open access to its content on the principle that making research freely available to the public supports a greater global exchange of knowledge.

ONLINE SUBMISSIONS

Already have a Username/Password for Austrian Journal of Statistics?
Go to http://www.ajs.or.at/index.php/ajs/login

Need a Username/Password?
Go to http://www.ajs.or.at/index.php/ajs/user/register

Registration and login are required to submit items and to check the status of current submissions.

AUTHOR GUIDELINES

The original LaTeX-file guidelinesAJS.zip (available online) should be used as a template for the setting up of a text to be submitted in computer readable form. Other formats are only accepted rarely.

SUBMISSION PREPARATION CHECKLIST

- The submission has not been previously published, nor is it before another journal for consideration (or an explanation has been provided in Comments to the Editor).
- The submission file is preferable in LaTeXfile format provided by the journal.
- All illustrations, figures, and tables are placed within the text at the appropriate points, rather than at the end.
- The text adheres to the stylistic and bibliographic requirements outlined in the Author Guidelines, which is found in About the Journal.

COPYRIGHT NOTICE

Austrian Journal of Statistics

Volume 47, Number 1, 2018

Editor-in-chief: Matthias TEMPL

http://www.ajs.or.at

Published by the AUSTRIAN STATISTICAL SOCIETY

http://www.osg.or.at

Österreichische Zeitschrift für Statistik

Jahrgang 47, Heft 2, 2018

ÖSTERREICHISCHE STATISTISCHE GESELLSCHAFT

Impressum

Published approximately quarterly by the Austrian Statistical Society, C/o Statistik Austria
Guglgasse 13, A–1110 Wien

Contents

Editorial

The Austrian Journal of Statistics is free and open access. Due to this spirit we are moving more and more from hard copies to electronical online versions only. This has several impacts regarding the hard copies volume. Because there is not much demand on the hard copies version, future hard copies of issues will be just a collection of articles without any special formatting on page numbers and table of contents, and also the editorial will be available online only.

This current issue includes five scientific papers for which all information can be accessed online at http://www.ajs.or.at.

Matthias Templ
(Editor-in-Chief)

Institute of Data Analysis and Process Design
Zurich University of Applied Sciences
Rosenstrasse 3, CH–8400 Winterthur,
Switzerland
E-mail: matthias.templ@gmail.com

Winterthur, 29. Januar 2018

Austrian Journal of Statistics
February 2018, Volume 47, 3–31.
http://www.ajs.or.at/
doi:10.17713/ajs.v47i1.689

Linear Association in Compositional Data Analysis

Juan José Egozcue Universitat Politècnica de Catalunya
Vera Pawlowsky-Glahn Universitat de Girona
Gregory B. Gloor University of Western Ontario

Abstract

With compositional data, ordinary covariation indices, designed for real random variables, fail to describe dependence. There is a need for compositional alternatives to covariance and correlation. Based on the Euclidean structure of the simplex, called Aitchison geometry, compositional association is identified to a linear restriction of the sample space when a log-contrast is constant. In order to simplify interpretation, a sparse and simple version of compositional association is defined in terms of balances which are constant across the sample. It is called b-association. This kind of association of compositional variables is extended to association between groups of compositional variables. In practice, exact b-association seldom occurs, and measures of degree of b-association are reviewed based on those previously proposed. Also, some techniques for testing b-association are studied. These techniques are applied to available oral microbiome data to illustrate both their advantages and difficulties. Both testing and measurements of b-association appear to be quite sensitive to heterogeneities in the studied populations and to outliers.

Keywords: Aitchison geometry, balances, CoDa-dendrogram, CoDa-biplot, hypothesis testing, log-ratio, log-contrast, simplex, oral microbiome.

1. Introduction

Measures of statistical dependence—association, concordance, covariation or correlation—have been important since the beginning of modern statistics. The introduction of the correlation coefficient by Galton (for historical details, see Stigler (1989)), nowadays known as the *Pearson correlation coefficient*, was a milestone in the development of multivariate statistics, though its value was discussed in varied frameworks. The controversy on the measure of association proposed by G. U. Yule (Pearson and Heron 1912) revealed both the importance of measures of association and the need of a clear comprehension on the assumptions underlaying the use of correlation-association indices. For instance, the Yule association coefficients, specially developed in the context of two way contingency tables (Yule 1903), were criticized from the point of view of the Pearson correlation coefficient, mainly designed for continuous quantitative variables. Throughout the 20^{th} century, and up to now, alternatives to Pearson correlation have been introduced, such as the Spearman and Kendall correlation coefficients (Spearman 1904; Kendall 1938), or the recently developed distance correlation (Székely, Rizzo, and Barikov 2007, 2009). Also, there were attempts of formalising these concepts and their respective indices (e.g. Schweizer and Wolff (1981); Scarsini (1984)). These indices are mainly used to look for independence of random variables. However, the Pearson correlation coefficient is also commonly involved in modelling (linear) relationships between variables in the omnipresent linear regression. The extreme values (± 1) imply some restriction on the data, i.e. that the sample points are on a straight line or, in general, on a linear manifold of the sample space. This is also clear for distance correlation: a value of 1 implies that the sample is constant, thus lying on a single point of the sample space. Other cases are more involved, as the restrictions are not well defined, and they may depend on the underlying distribution of the data. For instance, both Spearman and Kendall coefficients of ± 1 imply a monotonic functional restriction between the sample points, but the precise relationship requires further exploration. The diversity of alternatives presented arise from the lack of specifications on the hypothesis and applicability of each method.

When multiple variables are at play, correlation coefficients describe bivariate relationships between pairs of variables. Correlation matrices can be difficult to interpret, and data simplifications are often needed to attain further insight. Principal component analysis and subsequent simplifications provide powerful interpretative tools. They consist in looking for linear combinations of the original variables so that the resulting correlations satisfy certain conditions on simplicity, sparsity or geometric orthogonality (Chipman and Gu 2005; Enki, Trendafilov, and Jolliffe 2013). These kind of approaches are of primary importance when dealing with a large number of variables (hundreds or thousands) and dimension reduction is usually required to gain insights into the phenomenon under study.

Compositional data demand careful handling of correlation-association. Pearson (1897) first observed that correlations between ratios with a common denominator were unreliable and coined the term *spurious correlation*. The problem remained at this level, ignored—or even denied (Fisher 1947)—until the sixties. At that time, problems of statistical analysis in Geology (Chayes 1960, 1962, 1971) and in Biology (Mosimann 1962; Connor and Mosimann 1969), motivated research on how to deal with spurious correlation in the analysis of compositional data. However, the discussion remained stuck at how to separate the *true* correlation part from the *spurious* one, something that we now realize makes no sense in view of the sample space of, and the information carried by, compositional data. No discussion on the concept of association was brought up. It was not until the eighties that J. Aitchison situated compositional data analysis on a premise (Aitchison 1982, 1986) by introducing the log-ratio approach, as summarized in Aitchison and Egozcue (2005).

With respect to association between compositional components or parts, Aitchison introduced the variation matrix, the matrix of sample variances of simple log-ratios between parts of a composition. These quantities

$$\mathrm{Var}\left(\ln \frac{x_i}{x_j}\right) \quad , \quad i,j = 1,2,\ldots,D \ ,$$

where x_i, x_j denote parts of a D-part composition, are in fact measures of lack of association between compositional variables. When the variance is null, x_i and x_j are strictly proportional; when it is large, proportionality is lost or it is too noisy to be considered. However, this simple measure lacks a meaningful scale; it lacks statistical techniques for testing; and is missing even a geometrical background that justifies the term *measures of (lack of) association*.

The publication of the Euclidean structure of the simplex as sample space of compositional data (Billheimer, Guttorp, and Fagan 2001; Pawlowsky-Glahn and Egozcue 2001) set the premises for further developments on association of parts of a composition. The goal of the present contribution is to give, within this framework, scale invariant measures of association for compositional parts or groups of parts. The main thesis is that proportionality between the samples of two compositional parts, or groups of parts, is an appropriate criterion of compositional association (Egozcue, Lovell, and Pawlowsky-Glahn 2013; Lovell, Pawlowsky-Glahn, Egozcue, Marguerat, and Bähler 2015).

Section 2 summarizes some main lines of the algebraic-geometric structure of the simplex as sample space of compositional data. Section 3 discusses details on geometric restrictions of the sample space when the association is exact. They inspire the measures of association proposed in Section 4 and testing techniques associated with them (Section 5). Section 6 presents some examples in the context of an *omics*-science using a well-known 16S rRNA tag sequencing data set.

2. Sample space for compositional data

The definition of compositional data has progressively evolved from vectors of positive components adding to a given constant (e.g. Aitchison (1982)), to more recent and general definitions based on equivalence classes (Barceló-Vidal, Martín-Fernández, and Pawlowsky-Glahn 2001). Here, a composition is defined as a D-component vector, all components being strictly positive, where the ratios between different components contain the relevant information. The components do not necessarily add up to a constant. The key concept is the relative character of the information of interest, as multiplication of the composition by any positive constant does not change the information contained in the ratios between components. This was stated as the principle of scale invariance by (Aitchison 1986). A direct consequence is that relevant features (distances, coordinates, sizes, relationships, ...) must be invariant under changes of scale of the compositions. The evolution of the concept of scale invariance lead to the definition of an equivalence relation between vectors of $\mathcal{R}_+^D$ (real vectors with D positive components): two vectors, $\mathbf{x} = (x_1, x_2, \ldots, x_D)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_D)$, are compositionally equivalent, or c-equivalent, if there exists a positive constant α such that

$$x_i = \alpha\, y_i \quad , \quad i = 1, 2, \ldots, D \ .$$

The equivalence classes of this relation are called compositions. This definition was implicit in Aitchison (1986), but recognized explicitly only much later (see e.g. Aitchison (1997); Barceló-Vidal *et al.* (2001); Martín-Fernández, Barceló-Vidal, and Pawlowsky-Glahn (2003); Barceló-Vidal and Martín-Fernández (2016)).

Equivalence classes are usually identified by choosing a representative of each class on which operations and reasoning can then be performed. A representative composition can be selected in many possible ways; one approach is to select the representative belonging to the simplex so that all the components add to a positive constant. In this representation compositions appear expressed in proportions. In some applications, compositions are represented in alternative ways that may not sum to a known constant: for instance, chemical concentrations are frequently expressed in mols per liter; or air pollutant composition is given in μg per m^3. However, the corresponding composition can always be represented on the simplex by a simple change of units. For a discussion of these issues, see Buccianti and Pawlowsky-Glahn (2005). That any composition can be represented on the simplex in a one-to-one way makes the simplex an appropriate representation of the sample space of compositional data. However, the simplex is not the only possible representation; for instance, a positive orthant of a hypersphere can also be used to represent a composition (Aitchison 1986; Wang, Liu, Mok, Fu, and Tse 2007). The advantage of the simplex representation is that it is mathematically easier to handle and has a simpler to interpret algebraic structure.

It is obvious that the sample space of a dataset needs a structure that is appropriate for its proper analysis. In the case of compositional data, the simplex, here denoted as $\mathcal{S}^D$, should be equipped with both internal operations and metrics. For instance, $\mathcal{S}^D$ needs to have an associated σ-field to handle random compositions and allow the assignment of probabilities to events that respects the constant sum (or closed representation) of the data. Moreover, to operate with compositions in some way, the required operation needs to be defined and closed in the sample space. The most common transformation of a composition is perturbation, i.e., componentwise multiplication. For instance, when some market shares, concentrations of chemicals, or abundances of microbial species change, it is said that they increased, e.g. 3%, 1%, or −2%, meaning that previous shares, concentrations or abundances were multiplied by 1.03, 1.01, or 0.98. If the considered compositions are probabilities of a collection of incompatible events, perturbation is identified as Bayes' formula. Even a change of units of only some components of the composition can be viewed as a perturbation.

Perturbation of two compositions $\mathbf{x}, \mathbf{y} \in \mathcal{S}^D$, is defined as

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}(x_1 y_1, x_2 y_2, \ldots, x_D y_D) \ ,$$

where $\mathcal{C}$ is the closure operator which selects the representative on the simplex; that is, $\mathcal{C}$ divides each component by the sum of all of them. Repetition of a perturbation, or multiplication by a real scalar $\alpha \in \mathcal{R}$, called powering, is

$$\alpha \odot \mathbf{x} = \mathcal{C}(x_1^\alpha, x_2^\alpha, \ldots, x_D^\alpha) \ .$$

These two operations in $\mathcal{S}^D$, introduced in Aitchison (1982), configure $\mathcal{S}^D$ as a $(D-1)$-dimensional vector space (Aitchison, Barceló-Vidal, Egozcue, and Pawlowsky-Glahn 2002). This means that it is natural to determine a basis of the space, to choose coordinates, to perform linear combinations or any of all those computations which are allowed and usual in a finite dimensional vector space and do not require a metric.

But compositional analysis also requires distances between compositions and projections of those distances. These requirements complete the structure of the simplex with metric concepts. The Aitchison distance in the simplex (Aitchison, Barceló-Vidal, Martín-Fernández, and Pawlowsky-Glahn 2000) is generated by an inner product given by

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \ln \frac{x_i}{x_j} \ \ln \frac{y_i}{y_j} \ , \tag{1}$$

as $\mathrm{d}_a^2(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x} \ominus \mathbf{y}, \mathbf{x} \ominus \mathbf{y} \rangle_a$, and the corresponding norm $\|\mathbf{x}\|_a^2 = \langle \mathbf{x}, \mathbf{x} \rangle_a$. The subscripts a in the inner product, distance and norm, indicate that these expressions correspond to the Aitchison geometry of the simplex. The symbol $\ominus$ denotes perturbation-difference and can be defined as $\mathbf{x} \ominus \mathbf{y} = \mathbf{x} \oplus ((-1) \odot \mathbf{y})$. The simplex $\mathcal{S}^D$, endowed with perturbation, powering, and with the inner product defined above (Eq. 1) is a $(D-1)$-dimensional Euclidean space (Pawlowsky-Glahn and Egozcue 2001; Billheimer *et al.* 2001). For more details see Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado (2015).

Expressions of metric concepts in the Aitchison geometry can be simplified using the centered log-ratio (clr) transformation of a composition (Aitchison 1982). It is defined as

$$\mathrm{clr}(\mathbf{x}) = \left(\ln \frac{x_1}{\mathrm{g_m}(\mathbf{x})}, \ln \frac{x_2}{\mathrm{g_m}(\mathbf{x})}, \ldots, \ln \frac{x_D}{\mathrm{g_m}(\mathbf{x})} \right) \ ,$$

where $\mathrm{g_m}(\cdot)$ is the geometric mean of the components of its argument. The components of $\mathrm{clr}(\mathbf{x})$ add to zero by construction, that is $\sum_{i=1}^{D} \mathrm{clr}_i(\mathbf{x}) = 0$. The inverse transform is readily identified with $\mathbf{x} = \mathcal{C} \exp(\mathrm{clr}(\mathbf{x}))$, where the function exp operates componentwise on its arguments. The Aitchison inner product (Eq. 1) is then reduced to $\langle \mathbf{x}, \mathbf{y} \rangle_a = \langle \mathrm{clr}(\mathbf{x}), \mathrm{clr}(\mathbf{y}) \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the ordinary inner product in $\mathcal{R}^D$. From this property, the Aitchison distance and Aitchison norm can be expressed in terms of clr values as the ordinary Euclidean distance and norm of the respective clr values. Note that this expression only holds if the whole clr-vector is considered, as it changes when a reduced number of components are considered. It is said that the clr-vector is not subcompositionally coherent, as discussed below.

A crucial point in compositional data analysis is that subsets of compositional components are frequently analyzed, and the results of these analyses are to be taken as representative of what would be found with the complete set of variables. For instance, the chemical analysis of a drinking water is done using the dissolved matter after drying the sample; or a microbiome sample is unlikely to contain proportions of all taxa present in the environment; or an RNA-seq experiment will exclude, by default, ribosomal RNA which forms the majority of the RNA in a cell; or a joint analysis of financial indices does not include all existing indices. As a consequence, a degree of coherence is required when comparing analyses of a composition and a subcomposition, the second being a strict subset of the first (Aitchison 1986; Egozcue 2009; Pawlowsky-Glahn *et al.* 2015). A reasonable required condition is that of subcompositional dominance: the distance between compositions must be greater than, or equal to, the distances between the respective subcompositions. It can be shown that taking a subcomposition in $\mathcal{S}^D$ is equivalent to an orthogonal projection in the Aitchison geometry of the simplex (Egozcue and Pawlowsky-Glahn 2005) and, therefore, within this geometry the required dominance of distances is automatically fulfilled.

In any Euclidean space, Cartesian coordinates and their corresponding orthonormal bases can be built, and the D-part simplex $\mathcal{S}^D$ endowed with the Aitchison geometry is not an exception. There are infinitely many orthonormal bases. Those made of balancing elements are nowadays of particular interest due to their easy interpretation in practical applications. A balancing element (Egozcue, Pawlowsky-Glahn, Mateu-Figueras, and Barceló-Vidal 2003; Egozcue and Pawlowsky-Glahn 2005) is a unitary composition $\mathbf{e}$, which $\mathrm{clr}(\mathbf{e})$ contains coefficients with only two non-null values. Up to permutation of components, the form of the clr of a balancing element is

$$\mathrm{clr}(\mathbf{e}) = (a_+, a_+, \ldots, a_+, a_-, a_-, \ldots, a_-, 0, 0, \ldots, 0) \ ,$$

with n_+ and n_- components with values a_+ and a_-, respectively. As the components of $\mathrm{clr}(\mathbf{e})$ add to zero, $n_+a_+ + n_-a_- = 0$, and we can assume, without loss of generality, that $a_+ > 0$ and $a_- < 0$. Moreover, as $\mathbf{e}$ is unitary, the sum of squares of the components is 1. This implies that

$$a_+ = \frac{1}{n_+}\sqrt{\frac{n_+n_-}{n_+ + n_-}} \quad , \quad a_- = -\frac{1}{n_-}\sqrt{\frac{n_+n_-}{n_+ + n_-}} \ .$$

The orthogonal projection of a composition $\mathbf{x}$ onto $\mathbf{e}$ is $b \odot \mathbf{e}$, where

$$b = \langle \mathbf{x}, \mathbf{e} \rangle_a = \sqrt{\frac{n_+n_-}{n_+ + n_-}} \ \ln \frac{\mathrm{g_m}(\mathbf{x}_+)}{\mathrm{g_m}(\mathbf{x}_-)} \ , \tag{2}$$

and $\mathrm{g_m}(\mathbf{x}_+)$, $\mathrm{g_m}(\mathbf{x}_-)$ are the geometric means of those components of $\mathbf{x}$ which correspond to positions of $\mathbf{e}$ with positive and negative components, respectively. In Equation (2), b is called balance between the groups of parts $\mathbf{x}_+$, $\mathbf{x}_-$.

An orthonormal basis of $\mathcal{S}^D$ made of balancing elements can be build using the sequential binary partition (SBP) procedure (Egozcue and Pawlowsky-Glahn 2005, 2006b, 2011). The assignment of balance-coordinates to a composition $\mathbf{x}$, corresponding to such a basis, is carried out using the isometric log-ratio transformation (ilr)

$$\mathrm{ilr}(\mathbf{x}) = V^\top \mathrm{clr}(\mathbf{x}) \quad , \quad \mathbf{x} = \mathcal{C}\exp(V \ \mathrm{ilr}(\mathbf{x})),$$

where the contrast $(D, D-1)$-matrix V has the clr values of $(D-1)$ orthonormal balancing elements as columns and $\mathrm{clr}(\mathbf{x})$, $\mathrm{ilr}(\mathbf{x})$ are considered as column matrices with D and $D-1$ components, respectively. For further mathematical details see Egozcue, Barceló-Vidal, Martín-Fernández, Jarauta-Bragulat, Díaz-Barrero, and Mateu-Figueras (2011).

3. Linear restrictions as exact association of variables

In multivariate real analysis, a group of g variables, with indices in G, is exactly correlated with another group of h variables, with indices in H, when the sample points are restricted to lay on a hyperplane or straight-line (linear manifold) of the real space. The analytical equation for this restriction contains one or more (affine) linear relations like

$$\sum_{i \in G} \alpha_i x_i = \beta_0 + \sum_{j \in H} \beta_j x_j \ .$$

The particular case in which $g = 1$ and $h = 1$ is readily written as the simple linear model $x_1 = (\beta_0/\alpha_1) + (\beta_1/\alpha_1)x_2$. Exact correlation introduces linear constraints on the data points so that they are restricted to a linear manifold of the whole space. This idea can be translated to a compositional framework. These concepts on correlation can be formulated in two ways: a sample version in which it is assumed that a sample has been observed; and a random variable version in which all the sample space values with their respective probabilities are taken into account. In what follows, the sample version is used in the explanations, but all definitions and properties can be extended to the random variable version. A compositional n-sample is arranged in a (n, D)-matrix $\mathbf{X}$ which rows $\mathbf{x}_i$ are D-part compositions, possibly non-closed. The variables or parts are denoted X_j, $j = 1, 2, \ldots, D$, thus X_j denotes the columns of $\mathbf{X}$, while x_{ij} stands for the i, j-entry of $\mathbf{X}$.

The general expression of a one-dimensional linear restriction in the Aitchison geometry of the simplex (Pawlowsky-Glahn and Egozcue 2001) is that there is a log-contrast which is constant (k) across the sample data-points, that is

$$\sum_{i=1}^{D} \alpha_i \ln X_i = k \quad , \quad \sum_{i=1}^{D} \alpha_i = 0 \ , \tag{3}$$

where the condition on the α values assures that the log-contrast is invariant under scaling of the D-part composition $(X_1, X_2, \ldots, X_D)$. When a number $d < D - 1$ of independent one-dimensional restrictions are satisfied by the sample, the sample-points are confined within an affine subspace of dimension $D - 1 - d$. When Equation (3) holds, it is said that the group of parts with positive α are compositionally associated with the group of parts with negative α, which is called c-association for simplicity.

In compositional data analysis, this kind of general exact c-association, although important for dimension reduction, is not very useful for interpretation. Association, as a concept, tries to facilitate interpretation and this is commonly attained when statements can be easily formulated in terms of the original parts of the composition. Imagine a researcher deals with a composition with some hundreds of parts and he/she attains the result that there is a set of α values (some hundreds of them) satisfying Equation (3). The interpretation of such a log-contrast is cumbersome and the researcher possibly starts trying to find out which α values have large absolute values, identifying in this way which parts play an important role in the log-contrast. This is a typical situation in principal component analysis of high-dimensional multivariate samples or in factor analysis. At least two characteristics may be desirable for interpretation of such log-contrasts: sparsity and simplicity (Chipman and Gu 2005). Sparsity looks for a number of α values, as large as possible, that are equal to zero, so that a reduced number of parts participate in the association. When the number of α-values is restricted, the log-contrast is then simple in a given sense. The problem becomes more complicated when the analyst looks for low variability log-contrasts, since the simple selection of important contributions does not guarantee low variability of the simplified log-contrast.

The complexity of c-association, i.e. of the log-contrast in Equation (3), suggests a more restrictive concept of compositional association fulfilling requirements of sparsity and simplicity. Some log-contrasts are remarkably easier to interpret than the general version in Equation (3). They are the so called balances described in Section 2. When a balance is constant across a sample there is c-association in the sense of Equation (3), since a balance is a log-contrast. That the log contrast is a balance can be shown by defining balance association, or b-association for short.

For a more detailed definition of b-association, let us consider an n-sample D-part composition in $\mathbf{X}$. The parts of $\mathbf{X} \in \mathcal{S}^D$ can be classified in three disjoint groups of parts denoted G, H,

R; when in lower case, these letters denote the number of parts in each group. Without loss of generality, the first group G is placed in the first indices as $G = \{X_1, X_2, \ldots, X_g\}$ of the composition, and the second one, made of the following h parts, is $H = \{X_{g+1}, X_{g+2}, \ldots, X_{g+h}\}$. The remaining parts constitute a third group of parts $R = \{X_{g+h+1}, \ldots, X_D\}$ and contains $r = D - g - h$ parts. Consider the (non-normalized) balance

$$B(G/H) = \ln \frac{\mathrm{g_m}(G)}{\mathrm{g_m}(H)} \ , \tag{4}$$

where $\mathrm{g_m}(\cdot)$ denotes the geometric mean of its arguments. The balance $B(G/H)$ is a log-contrast for which the α-coefficients only take three different values: $1/g$ for group G, $1/h$ for H and 0 for group R, fulfilling simplicity requirements and sparsity when R is large compared to G and H. The groups of parts G and H are b-associated across a given sample, when the balance $B(G/H)$ is constant across that sample. This can be reformulated as G, H are b-associated when $\mathrm{Var}[B(G/H)] = 0$. Here, $\mathrm{Var}[\cdot]$ denotes the sample variance across the sample.

The particular case in which $G = \{X_1\}$ and $H = \{X_2\}$ reduces the balance to a simple log-ratio. This means that b-association in this case is expressed as any of the following equivalent conditions

$$B(\{X_1\}/\{X_2\}) = \ln \frac{X_1}{X_2} = k \quad , \quad \mathrm{Var}[B(\{X_1\}/\{X_2\})] = 0 \ . \tag{5}$$

The b-association of the groups G, H is easily interpreted in terms of proportionality of the geometric means of the parts in each group. When the balance in (4) is constant, then $\ln \mathrm{g_m}(G) = k_1 + \ln \mathrm{g_m}(H)$, for some real constant k_1; taking exponentials it yields $\mathrm{g_m}(G) = k_2\ \mathrm{g_m}(H)$, where $k_2 = \exp(k_1)$ is a positive constant. This motivates the statement that b-association can also be called group-proportionality.

More than one b-association of groups of parts is possible, even involving overlapping groups. Each b-association implies a reduction of one unit in the total dimension $D-1$ of the simplex $\mathcal{S}^D$ whenever the restrictions are linearly independent in the Aitchison geometry of the simplex (Egozcue *et al.* 2011).

3.1. Synthetic example of b-association

In order to illustrate exact b-associations between groups, a synthetic case study of 5-part compositions has been conducted. One hundred compositions, $\mathbf{x}_i$, $i = 1, 2, \ldots, 100$, have been generated as follows:

$$\begin{aligned} \mathbf{x}_i &= (x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}) \\ &= \left(0.25 \cdot i,\ 0.7 \cdot x_{i1},\ (x_{i1})^{1.3},\ (x_{i1})^{-1.3},\ 2(x_{i1} \cdot x_{i4})^{1/2}\right) . \end{aligned}$$

These compositional samples are reduced to proportions taking closure and are plotted in Figure 1. The clr components of these proportions are shown in Figure 2, where the first component $\mathrm{clr}_1(\mathbf{x}_i)$ is taken on the x-axis and each of the other components on the y-axis. In Figure 2 all four dotted lines are straight-lines but only the line $(\mathrm{clr}_1(\mathbf{x}_i), \mathrm{clr}_2(\mathbf{x}_i))$ (blue) has slope equal to 1, as the parts x_{i1}, x_{i2} are proportional. Note that the power relations between x_{i1} and x_{i3} (green) (or x_{i4}, red) appear as straight lines with slopes different from 1. This tells us that X_1 is exactly b-associated with X_2 but not with X_3 and X_4. The variable X_5 also appears as a straight line (non-unitary slope, brown). Therefore, X_5 is not b-associated with X_1. However, X_5 was constructed so that $X_5 = 2\ \mathrm{g_m}(X_1, X_4)$ and $B(\{X_5\}/\{X_1, X_4\}) = \ln 2$. Thus, there is an exact b-association between the groups $G = \{X_5\}$ and $H = \{X_1, X_4\}$, despite the fact that neither X_1 nor X_4 are separately associated with X_5. Consequently, an exact b-association between G and H does not imply b-association between the parts in these two groups. The correlation matrix of the clr-variables has all the entries equal to ± 1 in this case. For this example, the dimension of the sample space is $D - 1 = 4$, but it

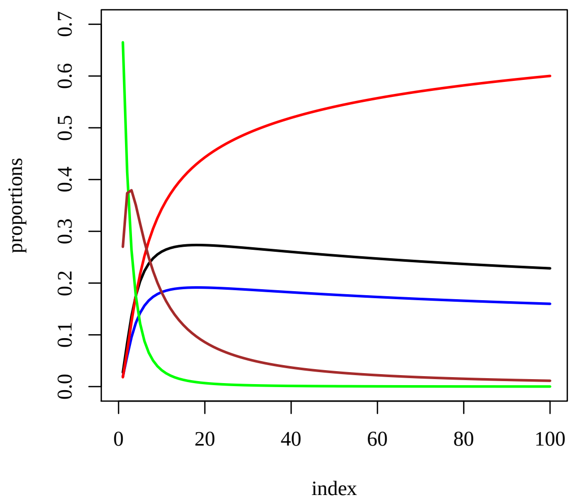

Figure 1: Proportions of one hundred compositions ordered by the subscript i as described in the text. Black, blue, green, red and brown lines correspond to $(x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5})$ respectively, for $i = 1, 2, \ldots, 100$.

was constructed so that two exact b-associations take place. However, there is another exact b-association, namely $B(\{X_1, X_3, X_4\}/\{X_2\})$, which is also constant along the sample, since $x_{i3}x_{i4} = 1$. As each constant log-contrast reduces the dimension of the sample space in one unit, sample compositions are in a 1-dimensional subspace. This is easily shown centering the data matrix and performing the singular value decomposition (SVD): there is only one non-null singular value.

In the previous example, it is surprising that all correlations between clr-components are ± 1, that is, all the points in Figure 2 are aligned, yet only the case with slope equal to 1 is considered as b-association. To clarify this example it is convenient to think on the columns of $\mathbf{X}$, denoted X_j, as compositions that can be represented in $\mathcal{S}^n$. When two of these columns, say X_i, X_j, are proportional, the compositional equivalence indicates that they are equal as compositions, and the Aitchison distance in $\mathcal{S}^n$ is 0. One could also say that both parts are represented in equal proportions along the sample. For instance, this is the case of the first (black) and second (blue) parts of the example shown in Figure 1. This is related to the fact that $\mathrm{Var}(\ln(X_i/X_j)) = 0$, as this sample variance is proportional to $\mathrm{d}_a^2(X_i, X_j)$ (see appendix A). This can be seen in $\mathrm{clr}(\mathbf{X})$, which contains the clr values of the rows $\mathbf{x}_i$, $i = 1, 2, \ldots, n$ (see Figure 2, blue points). This clr-matrix can be centered by columns, which is equivalent to a shift of the origin of the simplex $\mathcal{S}^D$ to the center of the composition. The resulting matrix contains clr values both by rows for compositions in $\mathcal{S}^D$ and by columns for compositions in $\mathcal{S}^n$ and its expression is

$$\mathrm{clr}(\mathbf{X_c}) = \mathrm{clr}(\mathbf{X}) - \frac{1}{n}\mathrm{clr}(\mathbf{X})\mathbf{1}_D\mathbf{1}_n^\top \ ,$$

where $\mathbf{1}_k$ is a k-column vector with unitary entries. The interdistances $\mathrm{d}_a(X_i, X_j)$ can be computed as Euclidean distances between columns of $\mathrm{clr}(\mathbf{X_c})$, as they are not affected by the centering shift. The Euclidean inner product of the columns $\mathrm{clr}(\mathbf{X_c})$ is equal to the sample covariances of $\mathrm{clr}_i(\mathbf{X})$ and $\mathrm{clr}_j(\mathbf{X})$ which are the columns of $\mathrm{clr}(\mathbf{X_c})$ up to an additive constant. Therefore, if the square distance $\mathrm{d}_a^2(X_i, X_j) = 0$, then $\mathrm{Cov}(\mathrm{clr}_i(\mathbf{X}), \mathrm{clr}_j(\mathbf{X})) = 0$. But the reciprocal is not true. Intuitively, in $\mathcal{R}^{n-1}$ the fact that the angle, centered at the origin, between two points is zero does not imply that the distance between these two points is zero; it can be as large as desired. Proportionality of X_i, X_j is equivalent to $\mathrm{d}_a(X_i, X_j) = 0$. However, $\mathrm{Cov}(\mathrm{clr}_i(\mathbf{X}), \mathrm{clr}_j(\mathbf{X})) = 0$ alone is not sufficient for b-association.

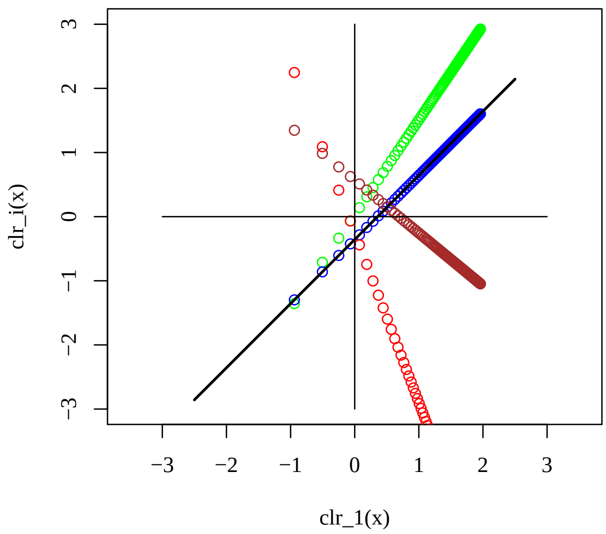

Figure 2: Components of clr of one hundred compositions of proportions shown in Figure 1. Components $\mathrm{clr}_2(\mathbf{x}_i)$ (blue), $\mathrm{clr}_3(\mathbf{x}_i)$ (green), $\mathrm{clr}_4(\mathbf{x}_i)$ (red) and $\mathrm{clr}_5(\mathbf{x}_i)$ (brown) are plotted against $\mathrm{clr}_1(\mathbf{x}_i)$. The black line has slope equal to 1 and passes through the mean value of $(\mathrm{clr}_1(\mathbf{x}_i), \mathrm{clr}_2(\mathbf{x}_i))$

3.2. Hardy-Weinberg law as a case of b-association

The Hardy-Weinberg (HW) law (Hardy 1908; Weinberg 1908) establishes that, under random mating conditions in an isolated population in which the alleles A and B present frequencies f_A, f_B, the genotype frequencies are $f_{AA} = f_A^2$, $f_{BB} = f_B^2$ (homozygote) and $f_{AB} = 2f_A f_B$ (heterozygote). This equilibrium is attained in one generation if there is parental genetic symmetry (Graffelman and Weir 2016). The allelic frequencies are arbitrary across a sample of populations in HW-equilibrium. However, the corresponding genotype frequencies satisfy

$$B(\{AA, BB\}/\{AB\}) = \ln \frac{(f_{AA} \cdot f_{BB})^{1/2}}{f_{AB}} = -\ln 2 \ ,$$

which implies an exact b-association involving the three parts. If balance-coordinates of the simplex $\mathcal{S}^3$ for genotype frequencies are defined as

$$b_1 = \sqrt{\frac{2}{3}} B(\{AA, BB\}/\{AB\}) \quad , \quad b_2 = \sqrt{\frac{1}{2}} B(\{AA\}/\{BB\}) \ ,$$

HW-equilibrium implies b_1 is constant across populations, but b_2 is not constant and can take any real value depending on the allelic frequencies in a population. This shows that if a balance involving several parts of a composition is constant, it does not imply that subsets of the parts are b-associated.

4. Approximate group-proportionality

Exact c-association seldom occurs in practice. A random composition with any exact compositional association between its parts has a degenerate probability distribution corresponding to the restriction imposed in Equation (3). Both for random compositions or sample compositions, interest is focussed in restrictions like those found in Equations (3) or (5) which only hold approximately. As a consequence, a measurement of the degree of association or proportionality is required for a consistent use in compositional analysis.

4.1. Aitchison's first approach

Aitchison (1986) introduced the first measure of (lack of) b-association when proposing the variation matrix as an exploratory tool. The entries of such a matrix are the sample variances $\mathrm{Var}[\ln(X_i/X_j)]$ for $i, j = 1, 2, \ldots, D$. The sum of all these entries, over $2D$, has been identified with the (sample) total variance or metric variance (Pawlowsky-Glahn and Egozcue 2001), that is

$$\mathrm{totVar}[\mathbf{X}] = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \mathrm{Var}\left[\ln \frac{X_i}{X_j}\right] ,$$

taking into account that $\mathrm{Var}[\ln(X_i/X_i)] = 0$. A large value for an entry of the variation matrix indicates a large contribution of the involved two parts to the total variance. Conversely, small values, close to zero, suggest that the two parts are nearly proportional. Therefore, $\mathrm{Var}[\ln(X_i/X_j)]$ is a measure of (lack of) b-association between X_i and X_j. However, the lack of a scale in $\mathrm{Var}[\ln(X_i/X_j)]$ makes its rigorous use difficult. There are different attempts to normalize the entries of a variation matrix.

A first approach reflects the proportion of total variance explained by a single log-ratio; these proportions are $\mathrm{Var}[\ln(X_i/X_j)]/(2D \cdot \mathrm{totVar}[\mathbf{X}])$. Although this normalization is sound, it strongly depends on the number of parts in the composition. A further advance consists of comparing the entries of the variation matrix with those obtained when the total variance is uniformly spread over all entries, a kind of maximal dissociation, which assigns to each non-null entry a value of $2D \cdot \mathrm{totVar}[\mathbf{X}]/(D(D-1))$. In this case, a possible normalization (Egozcue *et al.* 2013; Pawlowsky-Glahn *et al.* 2015) is

$$T_{ij} = \frac{D(D-1)\mathrm{Var}[\ln(X_i/X_j)]}{2D \cdot \mathrm{totVar}[\mathbf{X}]} = \frac{(D-1)\mathrm{Var}[\ln(X_i/X_j)]}{2 \cdot \mathrm{totVar}[\mathbf{X}]} . \tag{6}$$

Then, $T_{ij} < 1$ suggests association between the parts X_i and X_j. But experience points out that only values under 0.2 or even less can be considered candidates of effective b-association. Consequently, values larger then 0.2 can be rejected.

The mentioned lack of scale in the variances of simple log-ratios is moderated with this normalization by introducing a reference composition, $\mathbf{X}$, which includes all the parts involved in the association study. Changes of this reference composition by adding or removing parts will introduce changes in the scale.

4.2. Association between groups of parts

Recently, some additional measures of association between two parts have been introduced. For instance the ϕ-statistic (Lovell *et al.* 2015) and a modification of it (Erb and Notredame 2015). These measures of association are related to the approximate proportionality of single parts, but they can be generalised to association between groups of parts. In general, b-association is based on the proportionality of geometric means of groups of parts, that is $\mathrm{g_m}(G) \simeq k \cdot \mathrm{g_m}(H)$ or, taking logarithms,

$$\ln \mathrm{g_m}(G) \simeq k_1 + \beta_1 \ln \mathrm{g_m}(H) \quad , \quad k_1 = \ln k \quad , \quad \beta_1 = 1 . \tag{7}$$

For measuring b-association, interest is focussed on the slope β_1, which for b-association of G and H should be approximately 1. Equation (7) relates quantities such as $\ln \mathrm{g_m}(G)$ and $\ln \mathrm{g_m}(H)$, which are not scale invariant and, consequently, they are inappropriate for a compositional analysis, since the result would depend on the normalisation of the compositions. There are some options to transform Equation (7) into a scale invariant model preserving the value of β_1. A first choice is to subtract, from each term, $\ln \mathrm{g_m}(G \cup H \cup R)$, that is the logarithm of the geometric mean of all parts in the original composition; in the case that G and H reduce to a single part the equation would be $\mathrm{clr}_1(\mathbf{x}) \simeq k_2 + \beta_1 \mathrm{clr}_2(\mathbf{x})$. A second choice, which will be followed from now on, is to subtract $\ln \mathrm{g_m}(R)$ in Equation (7), thus yielding the linear model

$$B(G/R) = \beta_0 + \beta_1 B(H/R) + \epsilon , \tag{8}$$

where ϵ denotes residuals, and β_1 and the variability of the residuals are to be fitted to the available data. When $\beta_1 \simeq 1$, approximate proportionality holds for small residuals. The size of ϵ is relative to the balances $B(G/R)$ and $B(H/R)$. These balances depend on the original or reference composition which parts are included in the three groups G, H, R. Therefore, the scale of residuals are always relative to the reference composition.

4.3. Measuring b-association through a linear model

The next decision is how to fit the linear model (8) to available data. Both sides of the linear model play a symmetric role and are affected by variability. In such circumstances, ordinary least squares (OLS) regression is not an appropriate method. Following Warton, Wright, Falster, and Westoby (2006), and references therein, regression on the major axis (MA), also known as total least squares, and standardized major axis (SMA) (adopted in Lovell *et al.* (2015)) have been chosen for fitting the model (8). These approaches differ on the residual scores used to minimize their sum of squares. In OLS regression, the residual scores are equal to the ϵ values, that is the distance of data points to the fitted line in the direction of $B(G/R)$ (y-axis). In MA fitting, residual scores are chosen to be orthogonal to the fitted line. The SMA fitting procedure minimizes the sum of triangular areas comprising the data point and the fitted line limited by the projections of the data point on the fitted line following the response and explanatory axes. Figure 3 shows how residual scores are computed in OLS, MA and SMA regression. Note that, in these regression models, both axes have been taken as being orthogonal in the plane of Figure 3, a common practice in regression under the assumption that the sample space is the real space. However, when viewed in the Aitchison geometry of the simplex, these axes correspond to coordinates $B(G/R)$, $B(H/R)$ which are not orthogonal. Fortunately, this does not change the estimated model (i.e. estimation of β_0 and β_1), but it can slightly change the computation of residual scores and their sum of squares. Some tests associated with the sum of squares may be affected, but this details fall out of the scope of this contribution.

Both MA, SMA approaches coincide when the fitted line has slope equal to 1, as demanded for b-association. The estimation of slope in the SMA fit has a simpler expression than in the MA case. The SMA approach provides estimates of the slope slightly biased towards the unitary slope. In general, both approaches may be useful to evaluate b-association.

Our present goal is to discuss measures of (lack of) b-association as related to the model (8). The first one is the ϕ-statistic (Lovell *et al.* 2015) applied to balances appearing in the model given by Equation (8). That is

$$\phi(B(G/R), B(H/R)) = \frac{\mathrm{Var}(B(G/R) - B(H/R))}{\mathrm{Var}(B(G/R))} = 1 + \widehat{\beta_1}^2 - 2\widehat{\beta_1}|\widehat{r_{gh}}| \ , \tag{9}$$

where $\widehat{\beta_1}$ is the SMA-estimate of β_1 and $\widehat{r_{gh}}$ is the estimated correlation coefficient between $B(G/R)$ and $B(H/R)$.

It is obvious that $\phi(B(G/R), B(H/R))$ is equal to 0 when exact proportionality is attained with $\widehat{\beta_1} = 1$ and $\widehat{r_{gh}} = 1$; any departure from these values produces a positive increment of $\phi(B(G/R), B(H/R))$. Therefore, $\phi(B(G/R), B(H/R))$ can be taken as a measure of (lack of) b-association between G and H.

A second option measuring association (Erb and Notredame 2015), can be adapted to b-association between groups as

$$\rho(B(G/R), B(H/R)) = \frac{2\ \mathrm{Cov}(B(G/R), B(H/R))}{\mathrm{Var}(B(G/R)) + \mathrm{Var}(B(H/R))} = \frac{2\widehat{r_{gh}}}{\widehat{\beta_1} + \frac{1}{\widehat{\beta_1}}} \ , \tag{10}$$

where $\widehat{\beta_1}$ and $\widehat{r_{gh}}$ are the estimates appearing in Eq. (9). The statistic ρ is also a measure of b-association and satisfies $-1 \leq \rho(B(G/R), B(H/R)) \leq 1$, the latter value 1 corresponding

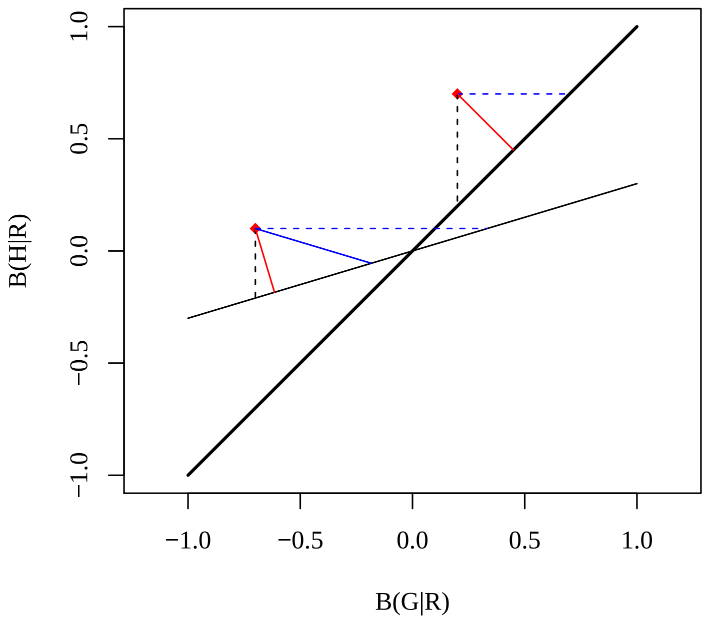

Figure 3: Visualization of residual scores for OLS, MA and SMA. Two artificial fitted lines (black full lines) with slopes 1 (thick line) and 0.3 (thin line). Two data points are shown (red marker): for each of them the OLS residual score (ϵ) is represented by a (vertical) black dashed segment to one of the fitted lines. Red lines are the residual scores used in MA, they are orthogonal to the respective fitted lines. The residual score for SMA for the fitted line with slope 0.3 is the blue segment. The SMA residual score coincides with the MA score (red) when the fitted line has slope equal to 1. The blue segment representing the SMA score is the half-diagonal of the rectangle when the triangle with vertical and horizontal sides is reflected by the fitted line; the square of this half-diagonal (blue) is proportional to the area of the triangle.

to exact b-association. Note again that the presence of the group R in both Eq. (9) and Eq. (10) and recall that these measures are defined with respect to a reference composition.

4.4. A generalisation of the variation matrix

A third possibility of measuring b-association is related to the generalisation of the entries of the variation matrix to the variances of log-ratios between the geometric means of components within each of the groups of parts considered. These variances correspond to the orthogonal projection of $\mathbf{x} \in \mathcal{S}^D$, $D = g + h + r$, such that all variance within the groups G, H, R are filtered out and the between-group variances are retained. The projected composition, up to a closure, is (Egozcue and Pawlowsky-Glahn 2005)

$$\mathbf{x}_b = (\underbrace{\mathrm{g_m}(G), \ldots, \mathrm{g_m}(G)}_{g \text{ components}}, \underbrace{\mathrm{g_m}(H), \ldots, \mathrm{g_m}(H)}_{h \text{ components}}, \underbrace{\mathrm{g_m}(R), \ldots, \mathrm{g_m}(R)}_{r \text{ components}}) \tag{11}$$

Consider the projected variation matrix

$$\mathbf{T}_b = \begin{pmatrix} 0 & gh\mathrm{Var}(B(G/H)) & gr\mathrm{Var}(B(G/R)) \\ hg\mathrm{Var}(B(H/G)) & 0 & hr\mathrm{Var}(B(H/R)) \\ rg\mathrm{Var}(B(R/G)) & rh\mathrm{Var}(B(R/H)) & 0 \end{pmatrix} ,$$

which entries add to $2D$ totVar($\mathbf{x}_b$). Inspired by the normalisation in Eq. (6), several normalisations can be proposed for the entries in $\mathbf{T}_b$. Here, the scaling

$$\frac{\mathbf{T}_b}{12\ D\ \mathrm{totVar}(\mathbf{x}_b)} ,$$

is proposed, and the entry

$$T_{GH} = \frac{gh\mathrm{Var}(B(G/H))}{12(g+h+r)\ \mathrm{totVar}(\mathbf{x}_b)} \ ,$$

is considered as a measure of (lack of) b-association between the groups G and H. Note that the constant 12 is twice the number of non-null entries of $\mathbf{T}_b$.

4.5. Comparison of measures of b-association

To compare the three measures of b-association, $\phi(B(G/R), B(H/R))$, $\rho(B(G/R),B(H/R))$ and T_{GH}, an experiment has been conducted. A set of 100 pairs $(B(G/R), B(H/R))$ has been simulated for different slopes and different orthogonal residuals as follows. First, one bivariate normal 100-sample has been simulated with mean at $(0, 0)$, with null covariance and standard deviations $\sigma_x = 5$ and $\sigma_y = 1$. These data points approximately fill an ellipse with principal axis on the x-axis. The y-axis values are then multiplied by values ranging from 0 to 5, and then rotated to attain different slopes between -3 and 3. This means that, when the standard deviation of the y-axis values of the original data is near to zero and the x-axis is rotated to slope 1, the b-association must be almost perfect. To compute T_{GH}, the number of parts grouped in G, H and R are assumed to be one, $g = h = r = 1$, and the standard deviation of $\ln \mathrm{g_m}(R) = 1$; this is needed to compute $\mathrm{totVar}(\mathbf{x}_b)$.

Figure 4 shows the results of the experiment. In the upper-left panel the values of T_{GH} are shown: exact association corresponds to slope equal 1 and standard deviation equal 0. The smaller values of T_{GH} correspond to better associations, which are illustrated by the white contour lines. The upper-right panel shows the values obtained for $\phi(B(G/R), B(H/R))$. Again the white contours are around the exact association point. Compared to T_{GH}, $\phi(B(G/R), B(H/R))$ has low sensitivity to the increase of standard deviation of orthogonal residuals, but both measures penalize symmetrically around the unitary slope. Contour lines for strong b-associations are very similar in the two cases. The bottom-left panel shows the results for $\rho(B(G/R), B(H/R))$. The main difference with the previous cases is that the values near to 1 correspond to the stronger b-associations, but changing high values into low values, the picture appears to be very similar to the one corresponding to T_{GH} (upper-left panel). For comparison, the correlation $\mathrm{Cor}(B(G/R), B(H/R))$ is presented in the bottom-right panel. Correlations near 1 include strong b-association, but it cannot be used as measure of association, as it is insensitive to changes of positive slopes. This example shows that the three measures of b-association are approximately equivalent except for their scale, and that simple correlation is inappropriate.

Introducing group R in the measures of b-association is important to fix a scale in the measures. However, there is a counterpart: the measures of b-association depend on the reference composition $G \cup H \cup R$ that is considered. Further comments on the subcompositional coherence of b-association are given in the next section on testing b-association.

Several reasons can explain a lack of proportionality even when T_{GH} and $\phi(B(G/R), B(H/R))$ are small, or when $\rho(B(G/R), B(H/R))$ is near to 1. First, these measures are not robust, and outlying samples can cause serious deviation in the results. Second, the presence of different populations in the sample may hide both b-association and the lack of it. The use of robust estimators of the variances in the variation matrix is recommended, but it does not remove the need to carefully inspect for different populations mixed in the sample. These points require further development.

5. Testing b-association

Testing b-association has limitations derived from the fact that the null hypothesis

$$H_0 : \quad B(G/H) = k \ , \tag{12}$$

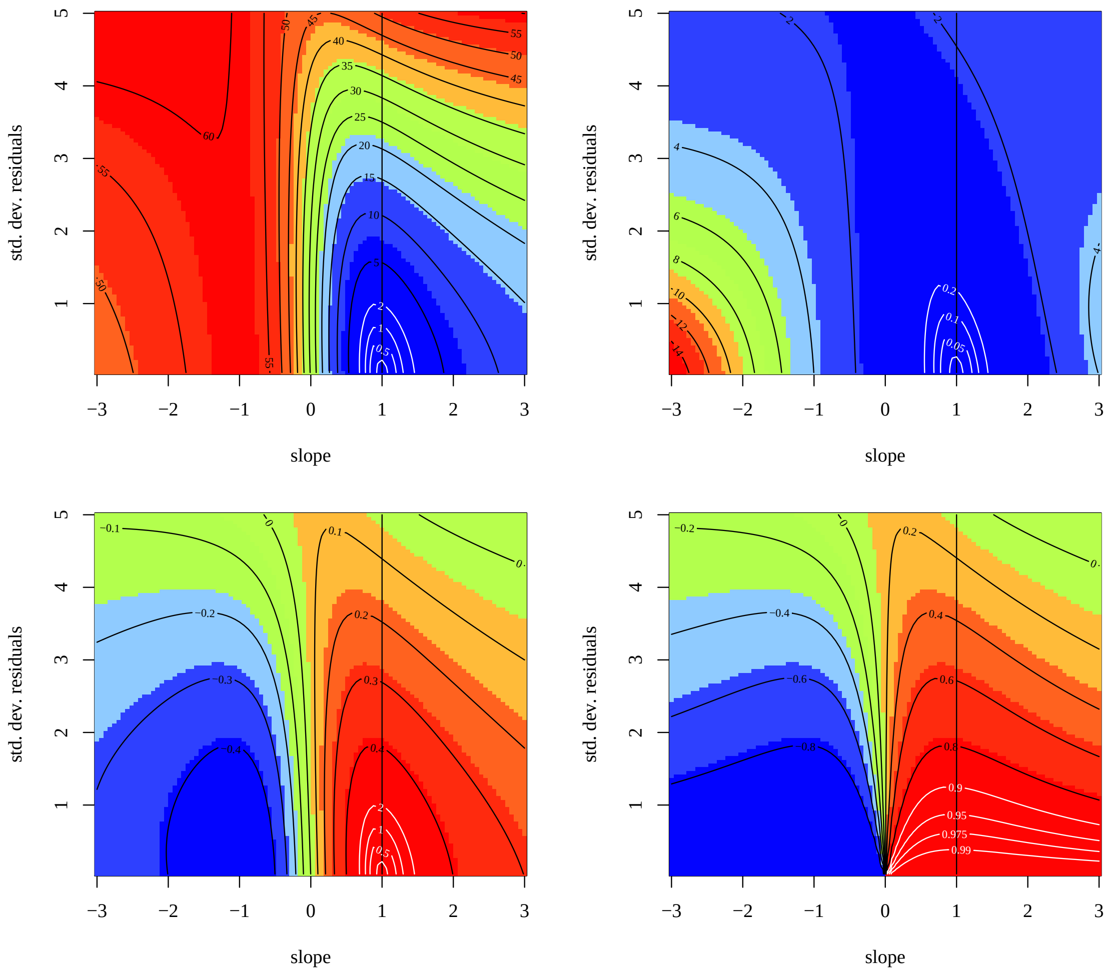

Figure 4: Magnitude of T_{GH} (upper-left), $\phi(B(G/R), B(H/R))$ (upper-right), $\rho(B(G/R), B(H/R))$ (bottom-left), and $\mathrm{Cor}(B(G/R), B(H/R))$ (bottom-right), for changing slope and standard deviation of orthogonal residuals. The scales of the four panels are not comparable as numbers on the isolines point out; only the shape of the isolines is relevant. Red and blue colors correspond to high and low values of the measure respectively. White isolines indicate approximate linear b-association. See text for elaboration.

does not admit variability for natural testing statistics. For instance, if $B(G/H)$ is used as a test statistic, its variance is null under H_0. In these circumstances, any deviation from H_0 leads to an immediate rejection of H_0 at any positive significance level. This situation is not unusual; it appears for example if for two real random variables, one tries to test that the Pearson correlation coefficient is exactly $+1$ or -1. There are some alternatives to test H_0, either modifying H_0, or making some assumption on the source of variability under H_0. Both alternatives are here adopted to propose two tests, one of them changing H_0 to $H_0' : \beta_1 = 1$, that is, testing for unitary slope in the model (8). The other alternative consists of considering a regression model of $B(G/H)$ predicted as a function of all or some balances orthogonal to $B(G/H)$ within a reference composition.

5.1. Test on unitary slope

Consider the linear model of Eq. (8). Exact b-association of the groups of parts G and H, implies H_0 as in Eq. (12). In its place, the hypothesis

$$H_0' : \quad \beta_1 = 1 \ ,$$

is tested. Note that, in case of approximate b-association, H_0' does not imply H_0, since H_0'

may hold even with large residuals. This means that H_0' is weaker than H_0.

Testing H_0' can be performed when model (8) is fitted using either major axis (MA) or standardized major axis approaches (SMA). The estimates of slope β_1 are (Warton *et al.* 2006)

$$\widehat{\beta_{MA}} = \frac{(s_{yy}^2 - s_{xx}^2) + \sqrt{(s_{yy}^2 - s_{xx}^2)^2 + s_{xy}^2}}{2s_{xy}} \ , \quad \widehat{\beta_{SMA}} = \mathrm{sign}(s_{xy}) \frac{s_{yy}}{s_{xx}} \ , \tag{13}$$

where subindices x, y denote the x-axis and the y-axis, which in model (8) are $B(G/R)$ and $B(H/R)$, respectively; the s denotes sample covariance between variables indicated by the subindices. Both estimates are very similar when the slope is near to one. Also, in both MA and SMA, the intercept term is estimated as $\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$, where $\bar{x}$, $\bar{y}$ are the corresponding sample means.

The test statistic for H_0' is based on the correlation between fitted values and residuals when H_0' is assumed to hold, that is $\epsilon' = y - \widehat{\beta_0} - 1 \cdot x$ and the fitted values approach the response as

$$f'_{MA} = 1 \cdot (y - \widehat{\beta_0}) + x \ , \quad f'_{SMA} = (y - \widehat{\beta_0}) + 1 \cdot x \ .$$

For testing H_0', the proposed statistic is

$$F = (n-2) \frac{r^2_{\epsilon' f'}}{1 - r^2_{\epsilon' f'}} \ ,$$

which, under H_0' and independent normality and homoscedasticity of ϵ, has an F-distribution with $(1,\, n-2)$ degrees of freedom.

This test has two additional inconveniences. The first, shared by most F-tests, is that the F-distribution is quite sensitive to departures of normality of residuals; however, even when the normality assumptions do not hold, the F-statistic is still a measure of departures from unit slopes although the obtained p-values using the F-distribution may be biased. The second inconvenience is that this test is actually testing for slope $\beta_1 = \pm 1$, and not simply for $\beta = +1$. When used for a small number of tests, it may be practical to examine by hand whether non rejections of H_0' are due to slopes near to -1. However, this strategy may be very inconvenient when performing some thousands of tests. A solution is to introduce an immediate rejection of H_0' whenever the slope estimate $\widehat{\beta_1}$ (see Eq. 13) is negative. The p-value in this situation is reported as -1.

5.2. Regression test

This testing option requires the establishment of a reference composition including the three groups of parts G, H and R. An appropriate option is to assume that the concatenation (G, H, R) is the reference and this is assumed in the discussion below. However, other subcompositions or orthogonal projections of the composition may be useful references. Consider the linear regression model

$$B(G/H) = \gamma_0 + \gamma_1 B(G, H/R) + \sum_{j=2}^{r} \gamma_j B_j + e \ , \tag{14}$$

where e represents the residuals, $B(G, H/R)$ is the balance of the concatenated G and H over R, and B_j are balances, possibly orthogonal, within the subcomposition R. The standard regression F-test, testing $\gamma_j = 0$ for $j = 1, 2, \ldots, r$, is equivalent to

$$H_0'' : \ B(G/H) = \gamma_0 + e \ ,$$

which is not H_0 in Eq. (12), but allows variability in H_0'' subject to the condition that it does not come from the reference composition. However, H_0'' is acceptable even with large residuals e, if they are not predictable from the reference composition. This means that the test will be more powerful when a large number of terms are considered in the model (14). When R is a large composition containing tens or thousands of parts, this testing procedure may be too strict. This testing procedure was found to be demanding when $r = 5$ or greater (Egozcue *et al.* 2013). Here a reduced power test is proposed for its use in large compositions, both reducing power and computational effort. The approach consists of reducing the model (14) only to the first explanatory variable, the balance $B(G, H/R)$ and its coefficient γ_1, thus reducing the model to $B(G/H) = \gamma_0 + \gamma_1 B(G, H/R) + e$. Note that using this simplified approach, the projected composition in Equation (11) substitutes the original reference composition. The original model in Eq. (14) is only used when a possibly strong association between G and H is detected, and one wants to be more strict at reducing type II errors (false negatives). This simplified test gives p-values equal to the slope test whenever the estimated MA-slope is positive, as both test statistics reduce one to the other. The difference is that the simplified regression test do not distinguish between positive and negative slopes.

As in the case of testing slope (Section 5.1), the standard F-statistic is an acceptable test statistic even if the independent and homoscedastic normality of the residuals e fails. As in a routine exploration of a large composition the F-test will be used without further checking, intermediate p-values obtained (say 0.005 to 0.1) should be examined critically.

6. Example using an 16S rRNA gene profiling case

In the last decade, the interest in the study of microbial communities has grown spectacularly (e.g. Gilbert, Quinn, Debelius, Xu, Morton, Garg, Jansson, Dorrestein, and Knight (2016); Faust and Raes (2012)). In the analysis of a microbiota experiment, e.g. using 16S rRNA gene variable region sequencing techniques, there are several critical points that require further research in order to establish reliable inference procedures (Weiss, Van Treuren, Lozupone, Faust, Friedman, Deng, Xia, Xu, Ursell, Alm, Birmingham, Cram, Fuhrman, Raes, Fengzhu, Zhou, and Knight 2016). Among these points, association or correlation between taxa is an important issue. Most techniques compared in Weiss *et al.* (2016) do not take into account the compositional character of the data, while others are only partially compositional. Only few contributions to the field are consistently compositional, for instance, Fernandes, Reid, Macklaim, McMurrough, Edgell, and Gloor (2014); Gloor, Wu, Pawlowsky-Glahn, and Egozcue (2016); Tsilimigras and Fodor (2016); Silverman, Washburne, Mukherjee, and David (2017). Motivated by this situation, a data set (Human Microbiome Project Consortium 2012) of oral microbiome has been selected to illustrate the compositional techniques here proposed.

The final OTU table and metadata mapping file from the HMQCP v35 dataset were downloaded on Oct 22, 2015 from the base URL
http:\\hmpdacc.org\HMQCP\ site (Gevers, Pop, Schloss, and Huttenhower 2012). Only visit number one for each person's sample was used. The OTUs were aggregated to bacterial genus by name, and there are 229 bacterial genera sampled at 8 different sites in the mouth of $n = 1457$ human individuals, constituting the set of samples. For simplicity, only three of these sites are used: keratinzed gingiva (ak); buccal mucosa (bm); and supragingival plaque –outside plaque– (op). A common feature of 16S rRNA gene sequencing data sets, is that they are sparse, i.e., zero counts are extremely abundant in the data matrix. Log-ratio analysis of compositional data does not allow null proportions as an argument of a logarithm, thus requiring special treatment of such data (Martín-Fernández, Hron, Templ, Filzmoser, and Palarea-Albaladejo 2015a; Gloor, Macklaim, Pawlowsky-Glahn, and Egozcue 2017). The present contribution is not aimed at discussing procedures and methods for treating these zero count data, although Bayesian estimation methods show some promise (Fernandes *et al.* 2014). Therefore, the number of genera used in the examples has been reduced to 12 by removing all genera with a total count across all samples of less than 5000, or that have zero counts in more than 100 samples. Finally, samples were discarded if the 12 genera presented with more than 4 zeros. A non-defined (ND) taxon, initially included in the 12 selected ones, was also removed due to the mixed character of the ND category. The number of genera in

Table 1: Genera used in the present example. Sample size n and the number of zero counts are reported. The three last columns show the center (compositional mean) of the three selected sites attached keratinized gingiva (ak); buccal mucosa (bm); supragingival plaque –over plaque– (op).

ID	genera	n	n zeros	cen-ak	cen-bm	cen-op
1	*Actinomyces*	1436	32	0.0013	0.0140	0.2028
2	*Fusobacterium*	1436	22	0.0033	0.0087	0.0462
3	*Gemella*	1436	24	0.0314	0.0555	0.0057
4	*Granulicatella*	1436	26	0.0107	0.0082	0.0070
5	*Haemophilus*	1436	3	0.2481	0.1393	0.0878
6	*Leptotrichia*	1436	54	0.0006	0.0060	0.0617
7	*Neisseria*	1436	33	0.0044	0.0202	0.1072
8	*Porphyromonas*	1436	47	0.0086	0.0177	0.0280
9	*Prevotella*	1436	10	0.0229	0.0216	0.0423
10	*Streptococcus*	1436	0	0.6364	0.6807	0.3532
11	*Veillonella*	1436	0	0.0325	0.0280	0.0579

the final composition is then $D = 11$.

Table 1 shows the genera which are accounted for, the sample size, which is equal for each site, and the number of zero-counts for each genus. In order to estimate all proportions of taxa, the method GMB (Martín-Fernández *et al.* 2015a) has been used. It consists of estimating proportions as

$$x_{ij} = \frac{n_{ij} + \alpha_{ij}}{n_{i+} + \alpha_{i+}} \quad , \quad \alpha_{ij} = s_i t_{ij} \quad , \quad i = 1, 2, \ldots, n \ , \ j = 1, 2, \ldots, D \ ,$$

where n_{ij} are the counts, possibly zero, for the i-th sample in the j-th taxon, and a subscript $+$ indicates sum in the corresponding dimension. This estimation is a Bayesian point estimation of multinomial probabilities. The values of s_i (strength) and t_{ij} (share) should be assessed from the characteristics of the problem and the available prior knowledge. In this case, following the advice in the previous reference, these values are

$$t_{ij} = \frac{n_{+j} - n_{ij}}{n_{++} - n_{i+}} \quad , \quad s_i = \exp\left(-\frac{1}{D}\sum_{k=1}^{D} \log t_{ik}\right) \ .$$

Table 1 also reports the compositional means for the genera in the three sites ak, bm and op. For each site, they are computed as the compositional average (Pawlowsky-Glahn *et al.* 2015)

$$\widehat{\mathrm{Cen}}[\mathbf{X}] = \frac{1}{n} \odot \bigoplus_{i=1}^{n} \mathbf{x}_i = \mathcal{C}(\mathrm{g_m}(X_1), \mathrm{g_m}(X_2), \ldots, \mathrm{g_m}(X_D)) \ ,$$

which reduces to compute the geometric means of the columns of the data matrix. Observing the values of the proportions in the centres substantial differences between sites are detected. A compositional MANOVA (Martín-Fernández, i Estadella, and Mateu-Figueras 2015b) based on the Aitchison distance d_a could be conducted to find out substantial differences between centres (not presented here).

Principal component analysis of compositional data (Aitchison 1983) and the corresponding biplots (Aitchison and Greenacre 2002) are powerful tools for exploring compositional data. This is also the case in exploring microbiome data (Gloor *et al.* 2016). They are computed by taking clr of the data set, centring it, and then carrying out singular value decomposition. That is

$$\mathrm{clr}(\mathbf{X}) - \mathrm{clr}(\widehat{\mathrm{Cen}}[\mathbf{X}]) = U \ \Lambda \ V^{\top} \ ,$$

where Λ is a diagonal matrix containing the singular values λ_i, $i = 1, 2, \ldots, D$; the (n, D)-matrix U, called the score matrix, contains standardized coordinates of the compositions;

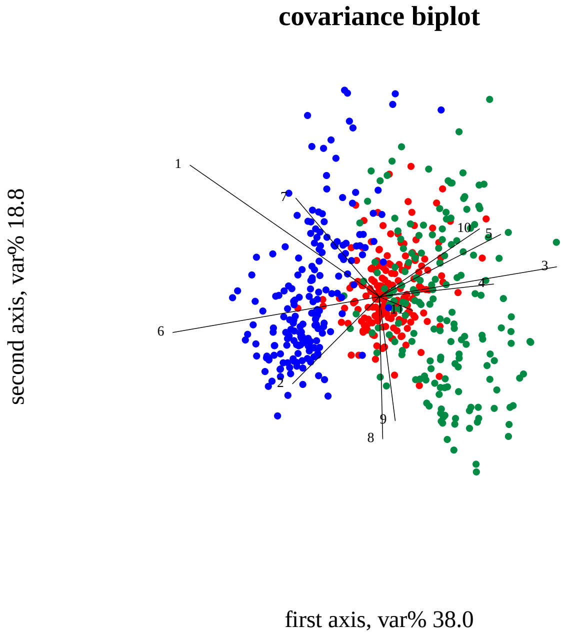

Figure 5: Covariance biplot of the data set projected on the first and second principal axes. Buccal sites: ak, green; bm, red; op, blue. Numbers represent clr-variables of taxa.

and the (D, D)-matrix V (loadings matrix), where the columns are the clr values of the new orthonormal basis of representation. The last singular value is always zero, as the components of the clr values add up to zero and $\mathrm{clr}(\mathbf{X})$ has, at most, rank $D-1$. The matrix $U\Lambda$ (except its last column) contains centered ilr-coordinates with respect to the basis associated with V (except its last column). A biplot consists of a 2-dimensional plot, representing simultaneously either U and $V\Lambda$ (covariance biplot) or $U\Lambda$ and V (form biplot). In a two-dimensional plot of the first two principal components, the proportion of the total variance represented in the plot is $(\lambda_1^2 + \lambda_2^2)/\mathrm{totVar}[\mathbf{X}]$ (for more details see e.g. Pawlowsky-Glahn *et al.* (2015)). Figures 5 and 6 show the covariance and form biplots projected on the first two principal components. The first observation is that the three sites are quite well separated by the first principal component (see this in the form biplot), which is roughly similar to the balance between the groups of parts $\{X_3, X_4, X_5, X_{10}\}$ and $\{X_1, X_6\}$, as revealed by the covariance biplot. The first two principal components explain 56.8% of the total variance and all observations made on the biplots are subject to this limited projection. In a covariance biplot like the one in Figure 5 the line linking two extremes of lines is, up to the projection, proportional to the standard deviation of the log-ratio between the corresponding variables. Therefore, rays where end-points are close suggest pairwise b-association between those variables. In Figure 5 the end-points of variables 8-*Porphyromonas* and 9-*Prevotella* are quite close, thus suggesting b-association. A similar case is visualised for the variables 5-*Haemophilus* and 10-*Streptococcus*. In both cases, the form biplot (Fig. 6) shows that both pairs of end-points are also close; this means that, in the projection, the corresponding unitary vectors are approximately equal, thus supporting the b-association between these couples of taxa. If these b-associations were confirmed, the result would be regarded as strong evidence for compositional association of these genera at the three different sites examined.

Biplots may be useful for a visual detection of pairwise b-associations but they need to be confirmed by examination of the variation matrix and other analyses. Another visual way of detecting possible b-associations, pairwise or involving larger groups, is the CoDa-dendrogram (Egozcue and Pawlowsky-Glahn 2006a) as shown in Figure 7. The tree structure indicates the SBP (sequential binary partition) used for the construction of a basis of balances. The length

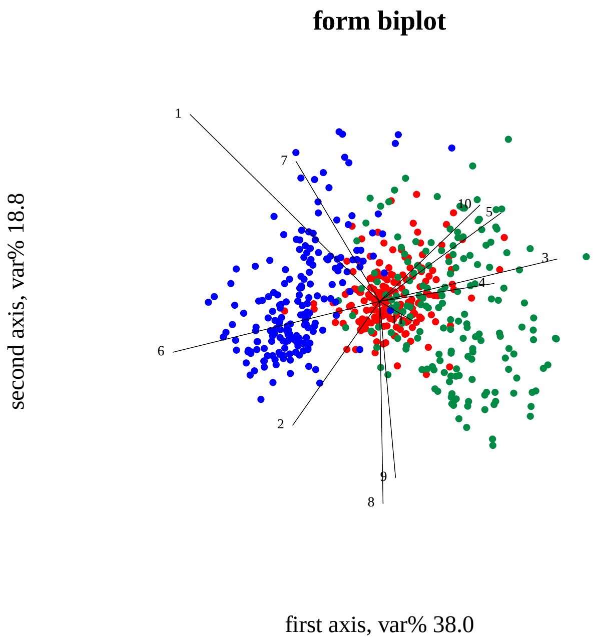

Figure 6: Form biplot of the data set projected on the first and second principal axes. Buccal sites: ak, green; bm, red; op, blue. Numbers represent clr-variables of taxa.

of the vertical bars illustrate the variance of the corresponding balance, and the length of the anchoring branch connecting the nodes is proportional to the mean variance of the balance. Vertical bars in black correspond to the aggregate of all samples at all sites (ak, bm, op), and the colored bars to individual sample groups. A relatively short vertical bar represents stronger evidence of b-association, because it represents a small variance of the balance. The advantage of the CoDa-dendrogram is that it is able to visualize associated groups of parts and not just pairwise b-associations. The counterpart is that only the balances obtained in the SBP are checked. Figure 7 shows the CoDa-dendrogram corresponding to an SBP which has been obtained by hierarchical clustering (Ward's method) of the genera (Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado 2011), using as the distance matrix, the square-roots of the variation matrix of the whole population (sites ak, bm, op), which is actually an Aitchison distance (Appendix A).

In Figure 7, only the balance between 5-*Haemophilus* and 10-*Streptococcus* seems to be approximately constant, that is, it exhibits a small variance of the log-ratio. The b-association seems particularly strong in the case of two individual sample groups (ak-green and bm-red). No b-association of larger groupings of genera is suggested in this dendrogram.

The suggested b-association between 5-*Haemophilus* and 10-*Streptococcus* was examined more closely. Figure 8 shows the fitted line of $B(5\text{-}Haemophilus/R)$ and $B(10\text{-}Streptococcus/R)$ for the three sites ak (left), bm (middle), op (right), where the p-values in the slope test are 0.51, 0.07, 0.0001. That is the b-association is not rejected in the ak samples, is near to rejection in the bm samples and is clearly rejected in op samples.

Similarly, potential b-association between 1-*Actinomyces* and 6-*Leptotrichia* is inspected in Figure 9. In this case, the obtained p-values are 0.82 and 0.11 for samples from ak and bm, while the slope was negative for the op samples thus producing an automatic rejection. In samples from the ak site, the slope test (p-value 0.86) did not reject b-association, but the scatterplot of the points indicates that b-association, if any, is quite noisy and the test is only driven by the slope, which in this case is approximately equal to 1. In these three cases, the obtained p-value is largely influenced by the presence of outliers, thus suggesting the future use of robust estimators for accurate testing.

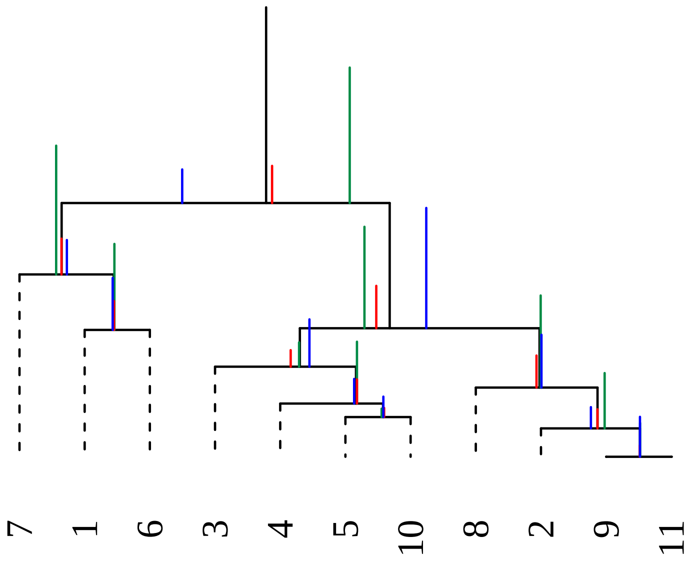

Figure 7: CoDa-dendrogram of approximately principal balances. Buccal sites: ak, green; bm, red; op, blue. Numbers represent clr-variables of taxa.

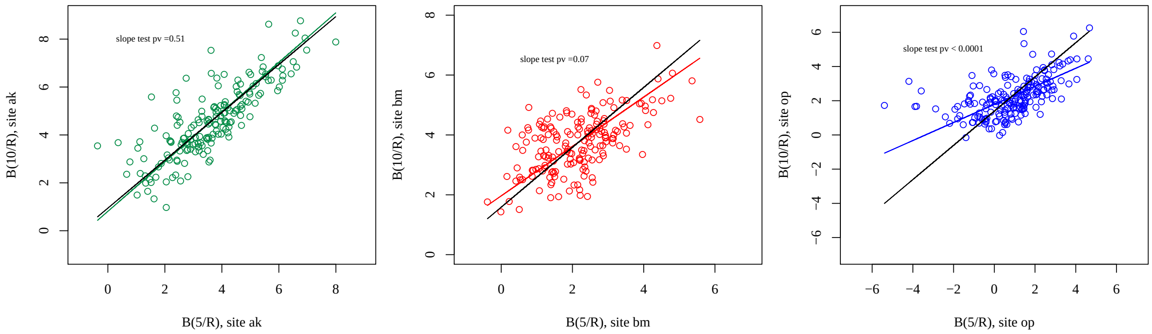

Figure 8: Scatterplot and fitted line of $B(5-Haemophilus/R)$ and $B(10-Streptococcus/R)$ corresponding to buccal sites ak (green), bm (red), op (blue). The null hypothesis corresponds to the unitary slope line (black).

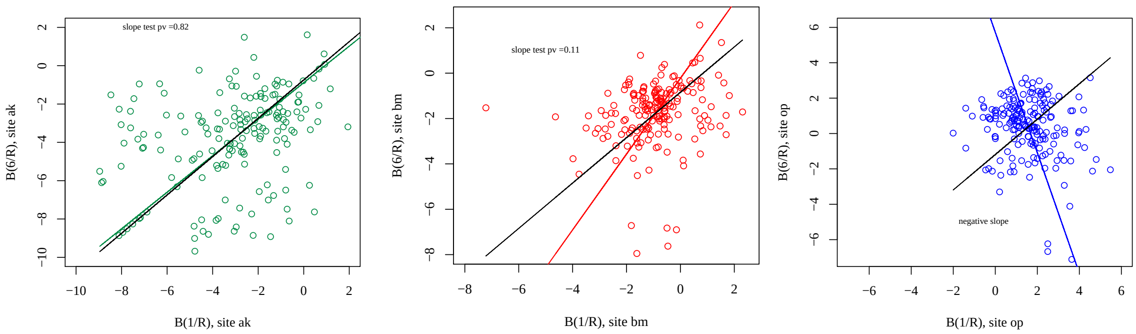

Figure 9: Scatterplot and fitted line of B(1-Actinomyces/R) and B(6-Leptotrichia/R) corresponding to buccal sites ak (green), bm (red), op (blue). The null hypothesis corresponds to the unitary slope line (black).

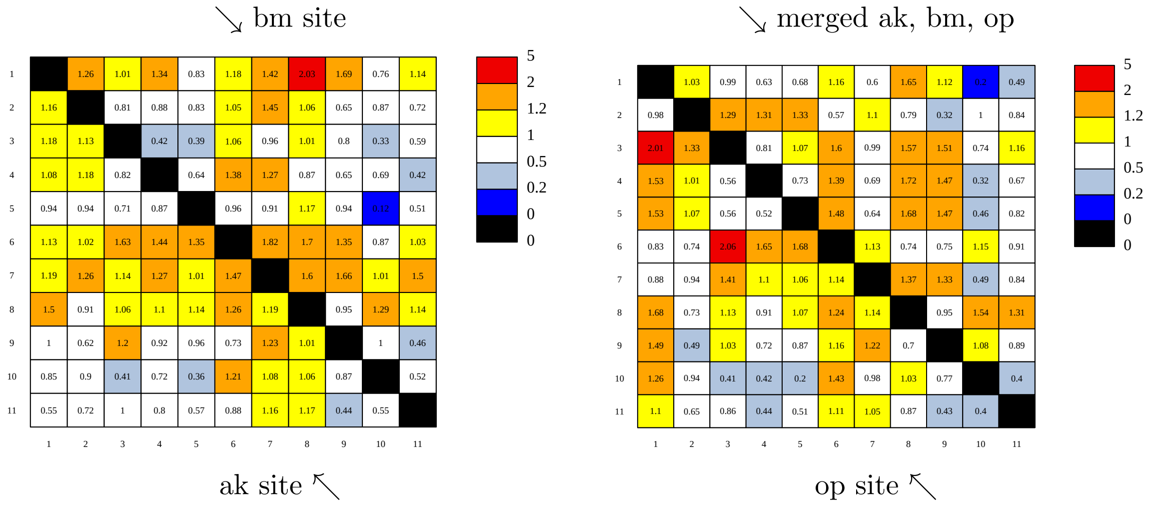

Figure 10: Normalized variation matrices for sites ak (left lower triangle), bm (left upper triangle), op (right lower triangle), for ak, bm, op merged in a single population (right upper triangle)

Figure 10 shows a summary heat map of pairwise b-association for the 11 genera in samples from the three selected sites (ak, bm, op). The top row shows the normalized variation matrices (Eq. 6) for the three sites and of the three sites together (four triangles). The middle and lower rows show the p-values of the slope test and the multiple regression test respectively for the genera at the same sites.

The first observation is that the suggestions of pairwise b-association by the three criteria (normalized variation, and the two p-value based methods) can differ substantially. However, if attention is paid to the strong b-association between 5-*Haemophilus* and 10-*Streptococcus* in bm site shown in Figure 8 (left panel), both the normalized variation matrix (0.12) and the slope test (p-value 0.51) suggest b-association. However, the regression test is clearly significant p-value $< 10^{-4}$, thus rejecting b-association of this pair by this test. In general, we observe that the multiple regression test is too strict, but instead suggests that b-association between larger groups of genera may be present if such groupings are examined. This situation is frequent, and is explored further below. In the lower panel row, there are only few non-significant cells and this number tends to decrease when the considered subcomposition is larger. As a counterpart, a large p-value, say greater than 0.05, in the multiple regression test often implies non-significant slope test and a small value of the normalized variation. This is the case in the joint population (ak, bm, op), for the b-association between 2-*Fusobacterium* and 9-*Prevotella*, which has a 0.05 p-value in the regression test, a 0.99 p-value in the slope test, and a moderate small value of normalized variation (0.49). This may be considered as a surprising case, as fewer b-associations are expected in the joint population of the three cases. However, the total variance is larger in the joint population and all tests and measures must be interpreted relative to the total variance. Together, these results indicate that further refinements in the tests and measures of b-association, and the use of robust statistics, may improve this approach.

As commented above, a rejection of b-association in the multiple regression test may suggest a b-association involving more than two variables. For this example with 11 parts, the number of non-significant cases (0.05 p-value or larger) in the slope test is large. For instance, in the ak site, there are 598 cases of G, H containing 2 parts each which give p-values larger than 0.05. Two of these b-associations, randomly selected, are studied in Figure 12. In the ak site (left panel), the balance of 10-*Streptococcus* and 11-*Veillonella* versus 4-*Granulicatella* and 5-*Haemophilus* is non-significant in the slope-test (p-value 0.409). Although it seems an acceptable b-association, it is quite noisy. The right panel of Figure 12 shows another case of b-association in the site bm (slope-test p-value 0.897). It involves 8-*Porphyromonas*

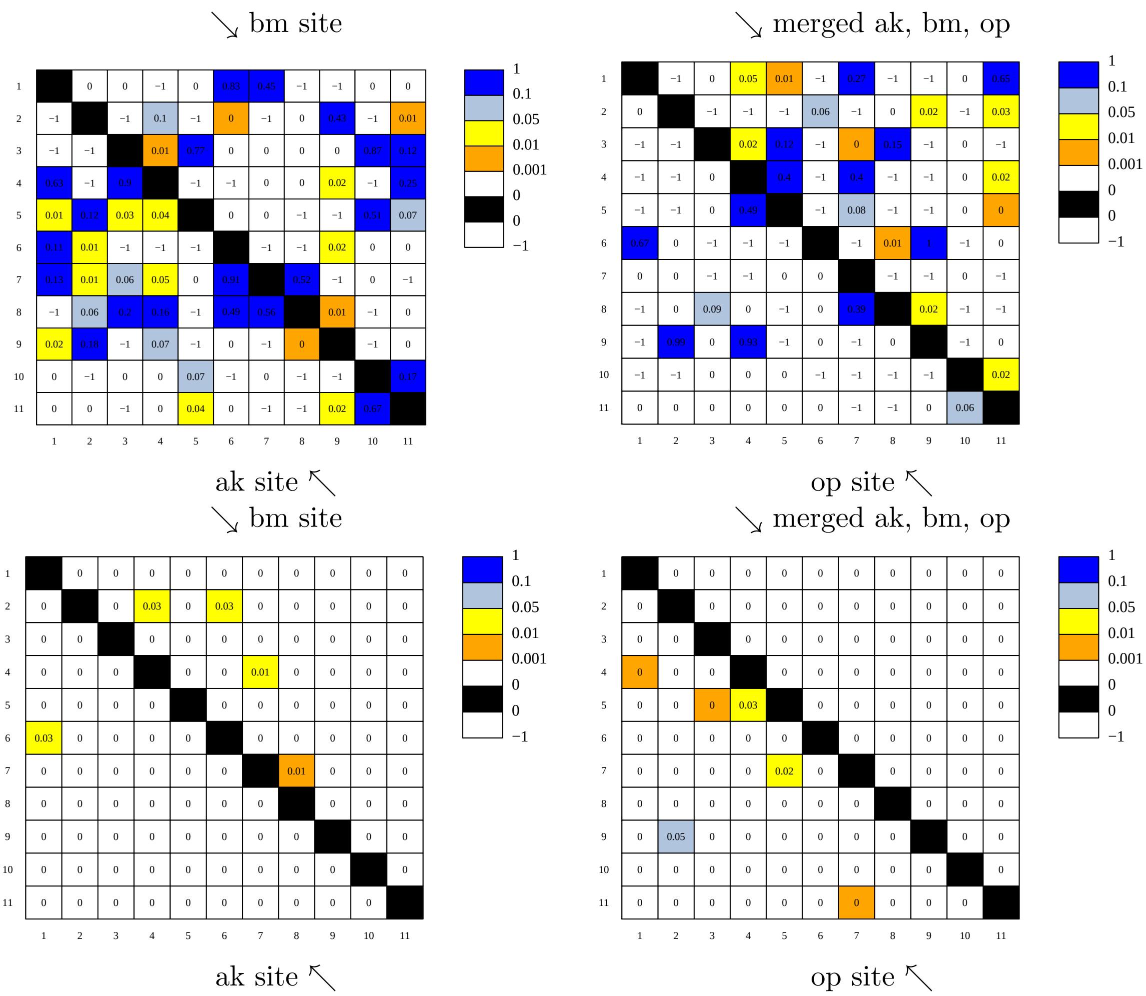

Figure 11: p-values for the slope test of association (panels in the upper row) and for the multiple regression test (panels in the lower row). They correspond to sites ak (left, lower triangles), bm (left, upper triangles), op (right, lower triangles), merged population ak, bm, op (right, upper triangles).

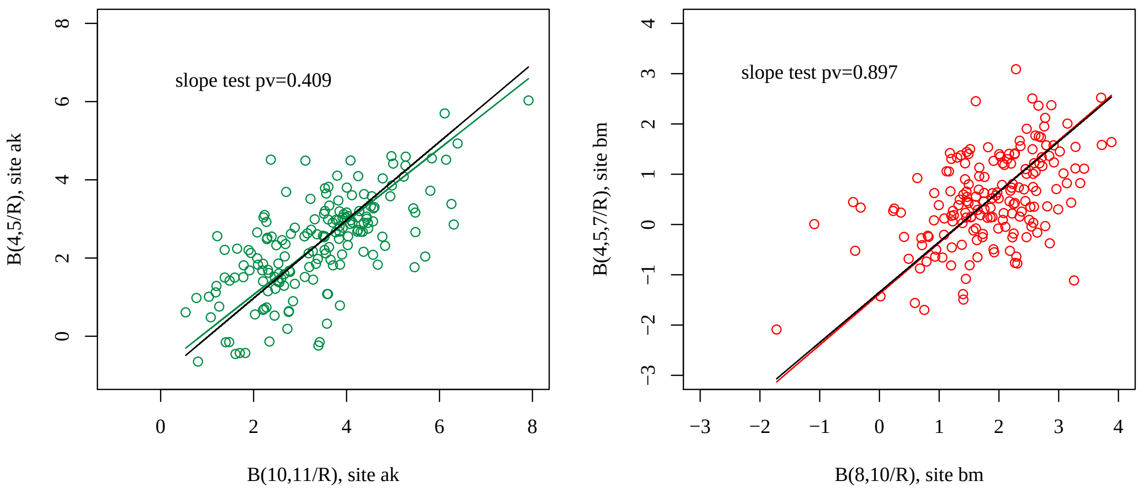

Figure 12: Testing unit slope for group association in two cases for which b-association is not rejected. Black line, unit slope; colored line, MA fitted line. Left panel: site ak, testing constant $B(10, 11/4, 5)$ in the 11-part simplex. Right panel: site bm, testing constant $B(8, 10/4, 5, 7)$

and 10-*Streptococcus* versus 4-*Granulicatella*, 5-*Haemophilus* and 7-*Neisseria*. Again the b-association is noisy. These two examples show that the slope and regression tests are useful to detect approximate unit slope between $B(G/R)$ and $B(H/R)$. However, the both the slope and regression tests are quite insensitive to large variability of residuals around the unit slope line. This is the reason why any of the proposed tests on b-association should be accompanied by some measurement of b-association as they, directly or indirectly, account for the scatter of data points around the unit slope line in plots like those shown in Figures 8, 9 and 12.

7. Conclusions and further research

Linear association of parts in a composition has been examined from a theoretical point of view, beyond the early proposals (Egozcue *et al.* 2013; Lovell *et al.* 2015). Proportionality of two parts in a composition was the idea for introducing b-association which leads to an approximately constant log-ratio. Maintaining this initial idea, the concept is extended to proportionality between geometric means of two groups of parts (b-association) which implies that the corresponding balance is constant. More generally, compositional linear association can be viewed as an approximately constant log-contrast between pairs or groups of parts.

The b-association, defined as the presence of a constant balance, is a simple generalisation of the proportionality between two parts, thus providing a way of thinking on groups of compositional components better than individual ones. That is, it is a way to go from pairwise relationships like that provided by standard pairwise correlations to more complex, but still linear in the simplex, relations involving more than two parts. The counterpart is that procedures to detect such group relations with a reasonable computing effort and their global interpretation is a pending research task.

Although the concept of linear compositional association is quite simple, its measurement and testing is more complex and requires further research. An important point is that the statistics here proposed, both for testing and measuring b-association, depend on the reference composition. Moreover, both measures of b-association and tests are severely affected by outliers and contamination in non-homogeneous populations. This strongly suggests the study of robust statistics both for measuring and testing.

When using b-association in fields like microbiome, or transcriptome, or other 'omic analysis, the presence of zeroes introduces a new source of contamination for detecting b-associations efficiently and accurately. This is especially true when such associations involve more than two parts. It is clear that the number of linear associations (b-associations) is less than predicted by other association criteria such as Pearson's correlation on proportions, or when using Spearman and Kendall correlations which include cases of non-linear associations. Nevertheless, the use of b-association provides a needed sanity-check, since the widely used correlation methods often provide many false positive associations.

Acknowledgements

This work was supported by grants MTM2015-65016-C2-1-R and MTM2015-65016-C2-2-R (MINECO/FEDER) of the Spanish Ministry of Economy and Competitiveness and European Regional Development Fund. The authors thank the constructive comments by the reviewers, which help improving the quality of the manuscript.

References

Aitchison J (1982). "The Statistical Analysis of Compositional Data (with discussion)." *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **44**(2), 139–177.

Aitchison J (1983). "Principal Component Analysis of Compositional Data." *Biometrika*, **70**(1), 57–65.

Aitchison J (1986). *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). (Reprinted in 2003 with additional material by The Blackburn Press), London (UK). ISBN 0-412-28060-4. 416 p.

Aitchison J (1997). "The One-hour Course in Compositional Data Analysis or Compositional Data Analysis Is Simple." In V Pawlowsky-Glahn (ed.), *Proceedings of IAMG'97 - The III Annual Conference of the International Association for Mathematical Geology*, volume I, II and addendum, pp. 3–35. CIMNE, Barcelona, Spain ISBN 978-84-87867-76-7, Barcelona (E). ISBN: 84-87867-97-9.

Aitchison J, Barceló-Vidal C, Egozcue JJ, Pawlowsky-Glahn V (2002). "A Concise Guide for the Algebraic-geometric Structure of the Simplex, the Sample Space for Compositional Data Analysis." In U Bayer, H Burger, W Skala (eds.), *Proceedings of IAMG'02 – The VIII Annual Conference of the International Association for Mathematical Geology*, volume I and II, pp. 387–392. Selbstverlag der Alfred-Wegener-Stiftung, Berlin, 1106 p. ISSN 0946-8978.

Aitchison J, Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V (2000). "Logratio Analysis and Compositional Distance." *Mathematical Geology*, **32**(3), 271–275.

Aitchison J, Egozcue JJ (2005). "Compositional Data Analysis: where Are We and where Should We Be Heading?" *Mathematical Geology*, **37**(7), 829–850.

Aitchison J, Greenacre M (2002). "Biplots for Compositional Data." *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **51**(4), 375–392.

Barceló-Vidal C, Martín-Fernández JA (2016). "The Mathematics of Compositional Analysis." *Austrian Journal of Statistics*, **45**, 57–71. doi:10.17713/ajs.v45i4.142.

Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V (2001). "Mathematical Foundations of Compositional Data Analysis." In G Ross (ed.), *Proceedings of IAMG'01 – The VII Annual Conference of the International Association for Mathematical Geology*, p. 20 p. Cancun (Mex).

Billheimer D, Guttorp P, Fagan W (2001). "Statistical interpretation of Species Composition." *Journal of the American Statistical Association*, **96**(456), 1205–1214.

Buccianti A, Pawlowsky-Glahn V (2005). "New Perspectives on Water Chemistry and Compositional Data Analysis." *Mathematical Geology*, **37**(7), 703–727.

Chayes F (1960). "On Correlation between Variables of Constant Sum." *Journal of Geophysical Research*, **65**(12), 4185–4193.

Chayes F (1962). "Numerical Correlation and Petrographic Variation." *The Journal of Geology*, **70**(4), 440–452.

Chayes F (1971). *Ratio Correlation*. University of Chicago Press, Chicago, IL (USA). 99 p.

Chipman HA, Gu H (2005). "Interpretable Dimension Reduction." *Journal of Applied Statistics*, **32**, 969–987.

Connor RJ, Mosimann JE (1969). "Concepts of Independence for Proportions with a Generalization of the Dirichlet Distribution." *Journal of the American Statistical Association*, **64**(325), 194–206.

Egozcue JJ (2009). "Reply to "On the Harker Variation Diagrams;..." by J. A. Cortés." *Mathematical Geosciences*, **41**(7), 829–834.

Egozcue JJ, Barceló-Vidal C, Martín-Fernández JA, Jarauta-Bragulat E, Díaz-Barrero JL, Mateu-Figueras G (2011). "Elements of Simplicial Linear Algebra and Geometry." In Pawlowsky-Glahn and Buccianti (2011), pp. 141–157. 378 p.

Egozcue JJ, Lovell D, Pawlowsky-Glahn V (2013). "Testing Compositional Association." In PF K Hron, M Templ (eds.), *Proceedings of the 5th Workshop on Compositional Data Analysis –CoDaWork 2013*. ISBN: 978-3-200-03103-6, http://coda.data-analysis.at/.

Egozcue JJ, Pawlowsky-Glahn V (2005). "Groups of Parts and Their Balances in Compositional Data Analysis." *Mathematical Geology*, **37**(7), 795–828.

Egozcue JJ, Pawlowsky-Glahn V (2006a). "Exploring Compositional Data with the CoDa-Dendrogram." In E Pirard, A Dassargues, HB Havenith (eds.), *Proceedings of IAMG'06 – The XI Annual Conference of the International Association for Mathematical Geology*. University of Liège, Belgium, CD-ROM, Liège (B).

Egozcue JJ, Pawlowsky-Glahn V (2006b). "Simplicial geometry for compositional data." In *Compositional Data Analysis in the Geosciences: From Theory to Practice*, volume 264 of *Special Publications*, pp. 145–159. Geological Society, London. ISBN 978-1-86239-205-2.

Egozcue JJ, Pawlowsky-Glahn V (2011). "Basic Concepts and Procedures." In Pawlowsky-Glahn and Buccianti (2011), pp. 12–28. 378 p.

Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C (2003). "Isometric Logratio Transformations for Compositional Data Analysis." *Mathematical Geology*, **35**(3), 279–300.

Enki HA, Trendafilov NT, Jolliffe IT (2013). "A Clustering Approach to Interpretable Principal Components." *Journal of Applied Statistics*, **40**(3), 583–599.

Erb I, Notredame C (2015). "How Should We Measure Proportionality on Relative Gene Expression Data?" *Theory in Biosciences*, **first online**, 1–16. doi:10.1007/s12064-015-0220-8.

Faust K, Raes J (2012). "Microbial Interactions: from Networks to Models." *Nature Reviews*, **August 2012**, 539–550.

Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB (2014). "Unifying the Analysis of High-throughput Sequencing Datasets: Characterizing RNA-seq, 16S rRNA Gene Sequencing and Selective Growth Experiments by Compositional Data Analysis." *Microbiome*, **2**, 15.1–15.13. doi:10.1186/2049-2618-2-15.

Fisher RA (1947). "The Analysis of Covariance Method for the Relation between a Part and the Whole." *Biometrics*, **3**(2), 65–68.

Gevers D, Pop M, Schloss PD, Huttenhower C (2012). "Bioinformatics for the Human Microbiome Project." *PLoS Comput Biol*, **8**(11), e1002779. doi:10.1371/journal.pcbi.1002779.

Gilbert JA, Quinn RA, Debelius J, Xu ZZ, Morton J, Garg N, Jansson JK, Dorrestein PC, Knight R (2016). "Microbiome-wide Association Studies Link Dynamic Microbial Consortia to Disease." *Nature*, **535**, 94–103. doi:10.1038/nature18850.

Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ (2017). "Microbiome Datasets Are Compositional: and This Is Not Optional." *Frontiers Microbiology*. doi:10.3389/fmicb.2017.02224.

Gloor GB, Wu JR, Pawlowsky-Glahn V, Egozcue JJ (2016). "It's All Relative: Analyzing Microbiome Data as Compositions." *Annals of Epidemiology*. doi:10.1016/j.annepidem.2016.03.003.

Graffelman J, Weir BS (2016). "Testing for Hardy-Weinberg Equilibrium at Biallelic Genetic Markers on the X Chromosome Heredity." *Heredity*, **116**(6), 558–568. doi:10.1038/hdy.2016.20.

Hardy GH (1908). "Mendelian Proportions in a Mixed Population." *Science*, **28**, 49–50.

Human Microbiome Project Consortium (2012). "Structure, Function and Diversity of the Healthy Human Microbiome." *Nature*, **486**(7402), 207–14. doi:10.1038/nature11234.

Kendall MG (1938). "A New Measure of Rank Correlation." *Biometrika*, **30**, 81–93.

Lovell D, Pawlowsky-Glahn V, Egozcue JJ, Marguerat S, Bähler J (2015). "Proportionality: A Valid Alternative to Correlation for Relative Data." *PLoS Comput Biol*, **11**(3), e1004075. doi:10.1371/journal.pcbi.1004075.

Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V (2003). "Dealing With Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation." *Mathematical Geology*, **35**(3), 253–278.

Martín-Fernández JA, Hron K, Templ M, Filzmoser P, Palarea-Albaladejo J (2015a). "Bayesian-multiplicative Treatment of Count Zeros in Compositional Data Sets." *Statistical Modelling*, **15**(2), 134–158.

Martín-Fernández JA, i Estadella JD, Mateu-Figueras G (2015b). "On the Interpretation of Differences between Groups for Compositional Data." *Statistics & Operations Research Transactions, SORT*, **39**(2), 231–252.

Martín-Fernández JA, Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2017). "Advances on Principal Balances for Compositional Data."

Mosimann JE (1962). "On the Compound Multinomial Distribution, the Multivariate β-distribution and Correlations among Proportions." *Biometrika*, **49**(1-2), 65–82.

Pawlowsky-Glahn V, Buccianti A (eds.) (2011). *Compositional Data Analysis: Theory and Applications*. John Wiley & Sons. ISBN 978–0–470–71135–4. 378 p.

Pawlowsky-Glahn V, Egozcue JJ (2001). "Geometric Approach to Statistical Analysis on the Simplex." *Stochastic Environmental Research and Risk Assessment (SERRA)*, **15**(5), 384–398.

Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2011). "Principal Balances." In JJ Egozcue, R Tolosana-Delgado, MI Ortego (eds.), *Proceedings of the 4th International Workshop on Compositional Data Analysis (2011)*. CIMNE, Barcelona, Spain ISBN 978-84-87867-76-7.

Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015). *Modeling and Analysis of Compositional Data*. Statistics in practice. John Wiley & Sons, Chichester UK. ISBN 9781118443064. 272 pp.

Pearson K (1897). "Mathematical Contributions to the Theory of Evolution. On a Form of Spurious Correlation which May Arise when Indices Are Used in the Measurement of Organs." *Proceedings of the Royal Society of London*, **LX**, 489–502.

Pearson K, Heron D (1912). "On Theories of Association." *Journal of the Royal Statistical Society*, **LXXV**, 579–652.

Scarsini M (1984). "On Measures of Concordance." *Stochastica*, **VIII**(3), 201–218.

Schweizer B, Wolff EF (1981). "On Nonparametric Measures of Dependence for Random Variables." *The Annals of Statistics*, **9**(4), 879–885.

Silverman JD, Washburne AD, Mukherjee S, David LA (2017). "A Phylogenetic Transform Enhances Analysis of Compositional Microbiota Data." *eLife*, **6**(e21887). doi:10.7554/eLife.21887.

Spearman C (1904). "The Proof and Measurement of Association between Two Things." *American Journal of Psychology*, **15**, 72–101. doi:10.2307/1412159.

Stigler SM (1989). "Francis Galton's Account of the Invention of Correlation." *Statistical Science*, **4**(2), 73–79. URL http://www.jstor.org/stable/2245329.

Székely GJ, Rizzo ML, Barikov NK (2007). "Measuring and Testing Dependence by Correlation of Distances." *The Annals of Statistics*, **35**(6), 2769–2794.

Székely GJ, Rizzo ML, Barikov NK (2009). "Brownian Distance Covariance." *The Annals of Applied Statistics*, **3**(4), 1236–1265.

Tsilimigras MCB, Fodor AA (2016). "Compositional Data Analysis of the Microbiome: Fundamentals, Tools, and Challenges." *Annals Epidemiology*, **26**, 330–335. doi:10.1016/j.annepidem.2016.03.002.

Wang H, Liu Q, Mok HMK, Fu L, Tse WM (2007). "A Hyperspherical Transformation Forecasting Model for Compositional Data." *European Journal of Operational Research*, **179**(2), 459–468.

Warton DI, Wright IJ, Falster DS, Westoby M (2006). "Bivariate Line-fitting Methods for Allometry." *Biol. Rev.*, **81**, 259–291. doi:10.1017/S1464793106007007.

Weinberg W (1908). "Über den Nachweis der Vererbung beim Menschen." *Jahreshefte d. Vereins vaterl. Naturkunde in Württemb.*, **64**, 369–382.

Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, Xia LC, Xu ZZ, Ursell L, Alm EJ, Birmingham A, Cram JA, Fuhrman JA, Raes J, Fengzhu S, Zhou J, Knight R (2016). "Correlation Detection Strategies in Microbial Data Sets Vary Widely in Sensitivity and Precision." *International Society for Microbial Ecology (ISME) Journal*, **10**(7), 1669–1681.

Yule GU (1903). "Notes on the Theory of Association of Attributes in Statistics." *Biometrika*, **2**(2), 121–134.

A. Relation of compositional variation with Aitchison distance

A compositional data matrix $\mathbf{X}$ is considered. The n rows of $\mathbf{X}$ are compositions represented in $\mathcal{S}^D$. Using the notation introduced in Sections 3 and 4, the following theorem holds.

Theorem It holds

$$n \cdot \mathrm{Var}\left(\ln \frac{X_i}{X_j}\right) = \mathrm{d}_a^2(X_i, X_j)\,,$$

where Var is the n-sample variance and $\mathrm{d}_a^2(\cdot,\cdot)$ is the Aitchison distance in $\mathcal{S}^n$. The following proof is similar to that presented in Martín-Fernández, Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado (2017).

Proof: Denoting g_i for the geometric mean of column i, i.e. $g_i = \left(\prod_{k=1}^n x_{ki}\right)^{1/n}$, each entry of the variation matrix can be written as

$$\mathrm{Var}\left(\ln \frac{X_i}{X_j}\right) = \frac{1}{n}\sum_{k=1}^{n}\left(\ln \frac{x_{ki}}{x_{kj}}\right)^2 - \left(\ln \frac{g_i}{g_j}\right)^2.$$

On the other hand, if we take the transpose of the data matrix $\mathbf{X}$, one can consider each part of the composition as an observation of a composition with n parts. These n parts correspond to the n initial samples. The squared Aitchison distance between two of the parts, e.g. X_i and X_j, is then

$$\begin{aligned}
&\mathrm{d}_a^2(X_i, X_j) \\
&= \sum_{k=1}^{n}\left[\ln\frac{x_{ki}}{g_i} - \ln\frac{x_{kj}}{g_j}\right]^2 = \sum_{k=1}^{n}\left[\ln\frac{x_{ki}}{x_{kj}} - \ln\frac{g_i}{g_j}\right]^2 \\
&= \sum_{k=1}^{n}\left(\ln\frac{x_{ki}}{x_{kj}}\right)^2 + \sum_{k=1}^{n}\left(\ln\frac{g_i}{g_j}\right)^2 - 2\sum_{k=1}^{n}\ln\frac{x_{ki}}{x_{kj}}\ln\frac{g_i}{g_j} \\
&= \sum_{k=1}^{n}\left(\ln\frac{x_{ki}}{x_{kj}}\right)^2 + n\left(\ln\frac{g_i}{g_j}\right)^2 - 2n\left(\ln\frac{g_i}{g_j}\right)^2 .
\end{aligned}$$

Thus, it holds that

$$n \cdot \mathrm{Var}\left(\ln\left(\frac{X_i}{X_j}\right)\right) = \mathrm{d}_a^2(X_i, X_j)\,.$$

As a consequence, we can define a *total squared distance* as follows

$$\begin{aligned}
\mathrm{totSDist}[\mathbf{X}] &= \mathrm{totVar}[\mathbf{X}] \\
&= \frac{1}{2D}\sum_{i=1}^{D}\sum_{j=1}^{D}\mathrm{Var}\left[\ln\frac{X_i}{X_j}\right] \qquad \text{(by definition)} \\
&= \frac{1}{2Dn}\sum_{i=1}^{D}\sum_{j=1}^{D}\mathrm{d}_a^2(X_i, X_j)\,.
\end{aligned}$$

Dividing each $\mathrm{d}_a^2(X_i, X_j)$ by $\mathrm{totSDist}[\mathbf{X}]$ we obtain a normalised matrix of inter-distances between compositional parts. In fact,

$$\frac{1}{2Dn}\sum_{i=1}^{D}\sum_{j=1}^{D}\frac{\mathrm{d}_a^2(X_i, X_j)}{\mathrm{totSDist}[\mathbf{X}]} = 1\,.$$

Affiliation:

Juan José Egozcue
Dept. of Civil and Environmental Engineering
Universitat Politècnica de Catalunya
Barcelona, Spain
E-mail: juan.jose.egozcue@upc.edu

Vera Pawlowsky-Glahn
Dept. of Computer Science, Applied Mathematics, and Statistics
Universitat de Girona, Girona, Spain
E-mail: vera.pawlowsky@udg.edu

Gregory B. Gloor
Dept. of Biochemistry
University of Western Ontario
London, Ontario, Canada
E-mail: ggloor@uwo.ca

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 47 *Submitted:* 2017-03-27
February 2018 *Accepted:* 2018-01-02

Austrian Journal of Statistics
February 2018, Volume 47, 33–44.
http://www.ajs.or.at/
doi:10.17713/ajs.v47i1.391

Generalized Point Estimators for Fuzzy Multivariate Data

Owat Sunanta
Vienna University of Technology

Abstract

Data analysis methods are necessary tools in evaluating and for better understanding the information of interest. However, there are limitations in applying standard statistical methods to data analysis in some cases. Data obtained from different sources are often clouded by imprecision and uncertainty, also called fuzziness. To overcome this problem, data analysis has to be adapted and generalized through statistical methods for such fuzzy data to capture the uncertainty. These methods are largely based on the extension principle and/or require other generalized procedures for further calculation of statistics, e.g. the estimation of the unknown statistical parameters. The development of these methods specifically for evaluating univariate data has been flourished. However, to solve complex real-world problems, these methods have to be extended and generalized to handle multivariate fuzzy data. In this research, the methods of generalized point estimators, i.e. sample mean, variance-covariance, and correlation coefficient, are extended for the multivariate case through concepts of fuzzy vector and combined fuzzy sample.

Keywords: fuzzy multivariate data, fuzzy vector, combined fuzzy sample, multivariate statistical analysis.

1. Introduction

Multivariate statistical analysis is concerned with data frequently obtained from empirical measurements on a number of individuals or objects. The sample data or the statistical treatments may be made on a collection of measurements in a manufacturing environment, such as diameter and surface finish of valve seats and those of corresponding nozzle seats in a petroleum system. The measurements made on a single unit (or a specimen) can ideally be assembled into a column vector. The entire vector may be thought of as an observation from a multivariate population or distribution. In statistical terms, when the individual measure is drawn randomly, the resulting vector is considered a random vector with a distribution or probability law describing the population (Anderson 2003). The set of observations on all individuals in a sample constitutes a sample of vectors, and the vectors set side by side make up the matrix of observations. The data to be analyzed are then thought of as displays in a matrix or in several matrices.

Unfortunately, the measurements are often clouded especially by uncertainty and imprecision. The uncertainty of measurement results of continuous quantities differs from probabilistic un-

certainty, however. Individual measurement results also contain a kind of uncertainty, which is called fuzziness. The most suitable mathematical model to describe the fuzziness is by fuzzy numbers and their characterizing functions (Zadeh 1965), (Viertl 2015). The extension principle (Zadeh 1975) in fuzzy set theory has been used in applications which call for an extension of the domain of a relation. In terms of statistical inference for fuzzy data, some literatures present formulating the statistics as by-product of fuzzy random variables Kruse (1984), Kruse and Meyer (1987), Liu and Kao (2002), Nguyen and Wu (2006), Wang (2004). In most cases, the notion of expected values is defined for fuzzy numbers of intervals as in (Dubois and Prade 1987), viewed as random fuzzy sets, which are subsequently used in analyzing fuzzy data. On the other hand, descriptive statistics through point estimators can be formulated differently through embedding the fuzziness through mathematical structures, a focal point of this paper. For univariate data, statistical inference involves combining fuzzy samples and propagation of fuzziness through fuzzy mathematical functions as reported by (Frühwirth-Schnatter 1992), (Viertl 2006), (Viertl 2011).
For analysis of multivariate fuzzy data, displayed in form of matrices of fuzzy numbers, generalized standard statistical inference methods for multivariate fuzzy data have to be considered. Comprehensive literatures in this particular area are limited and/or the proposed definitions are not necessarily concise. In case of multivariate fuzzy data, statistical inferences via point estimators require mathematical statements and involve high-dimension Cartesian products as a result of two-way combining of samples and multiple variables. With combined samples, further succinct mathematical definitions are possible. These concepts will be explained in detail and soon become obvious to the readers in the subsequent sections. For variance/covariance matrices, one of the differences in definitions lies in how the quantification methods of distances between the observed values (samples) and their respective mean are defined, see (Blanco-Fernàndez *et al.* 2014), (Körner 1997), (Kruse 1987), (Voxman 1998), and (Wang 2004) for different definitions, which serve different purposes, i.e. not necessary a generalized method for an $\mathbb{R}^{k \cdot n}$ space. Aside from variance, the measure of association through estimators of the population correlation coefficient is often of interest. In general, for fuzzy data, the correlation coefficient obtained through the Extension Principle is fuzzy as well (Ni and Cheung 2003). Different definitions of fuzzy correlation coefficient have been introduced, some of which result in the possible values of the correlation coefficient within [0,1] (Bustince and Burillo 1995) as opposed to within [-1,1] as defined in (Saneifard and Saneifard 2012) and (Liu and Kao 2002). The latter is the same definition as in standard case and also adopted here.

This paper presents a statistical inference method through mathematical statements and generalized point estimators for fuzzy multivariate data. Section 2 provides the foundation of fuzzy numbers and fuzzy vectors along with a concept of combining fuzzy samples in univariate case. In section 3, fuzzy multivariate data are described. In section 4, statistical functions of fuzzy multivariate data are then presented. The foundations presented in the previous sections are then applied in generalizing the point estimators along with a numerical example as shown in section 5. The paper concludes with final remarks.

2. Fuzzy numbers, fuzzy vectors and combined fuzzy samples

In case of univariate data, to describe observations or measurements of continuous quantities, the definition of general fuzzy numbers is useful.

Definition 1. A general fuzzy number $x^\star$ is defined by its characterizing function $\xi(\cdot)$, which is a real function of one real variable and possesses the following properties:

(1) $\xi : \mathbb{R} \to [0,1]$

(2) The support of $\xi(\cdot)$, denoted by $\text{supp}[\xi(\cdot)]$ and defined by $\text{supp}[\xi(\cdot)] := \{x \in \mathbb{R} : \xi(x) > 0\}$, is a bounded subset of $\mathbb{R}$.

(3) For all $\delta \in (0,1]$, the δ-cut $C_\delta\big[\xi(\cdot)\big]$, defined by
$C_\delta\big[\xi(\cdot)\big] := \{x \in \mathbb{R} : \xi(x) \geq \delta\} = \bigcup_{j=1}^{n_\delta} [a_{\delta,j}; b_{\delta,j}]$,
is non-empty and a finite union of compact intervals.

Along with general fuzzy numbers, how to obtain the characterizing function of a measurement result is also critical. For example, if a measurement $x^\star$ is quantified and represented by an interval $[a,b]$, such an interval can be characterized by its indicator function $\mathbb{1}_{[a,b]}(\cdot)$, i.e. $\xi(\cdot) = \mathbb{1}_{[a,b]}(\cdot)$, which is, in fact, a special kind of characterizing function. On the other hand, if a measurement $x^\star$ is recognized by a function $h(\cdot)$, the characterizing function of $x^\star$ can be derived through a transformation of its inherent function $h(\cdot)$, as following:

$$\xi(x) := \frac{h(x)}{max\{h(x) : x \in \mathbb{R}\}} \quad \forall x \in \mathbb{R}. \tag{1}$$

Further details and examples can be found in Klir and Yuan (1995) and Viertl (2011).

Definition 2. Let $x_1^\star, \ldots, x_n^\star$ be n fuzzy numbers of the observation space $M_x \subseteq \mathbb{R}$ with corresponding characterizing functions $\xi_1(\cdot), \ldots, \xi_n(\cdot)$. To obtain a fuzzy vector $\mathbf{x}^\star$, the fuzzy numbers $x_1^\star, \ldots, x_n^\star$ have to be combined (Viertl and Hareter 2006). Through construction of an n-dimensional vector-characterizing function $\zeta(\cdot, \ldots, \cdot)$ via a triangular norm (t-norm T), the combined fuzzy sample $\mathbf{x}^\star$ forms a fuzzy element $(x_1, \ldots, x_n)^\star$ of the sample space M_x^n, i.e.
$x_i^\star \widehat{=} \xi_i(\cdot), i = 1(1)n \overset{t\text{−norm}\,T}{\Longrightarrow}$ combined fuzzy sample $\mathbf{x}^\star$ and vector-characterizing function $\zeta(\cdot, \ldots, \cdot)$ where $\zeta(x_1, \ldots, x_n) := \kappa_n[\xi_1(x_1), \ldots, \xi_n(x_n)] \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n$
and the combination $\kappa_n = T_n$, which is the n-dimensional extension of the t-norm T by its associativity, i.e. $T_n(y_1, \ldots, y_n) = T\big(y_1, T(\ldots, T(y_{n-1}, y_n) \ldots)\big) \quad \forall (y_1, \ldots, y_n) \in [0,1]^n$.

For statistical and algebraic calculations with fuzzy data, the minimum-t-norm T is optimal (Viertl 2011), i.e.

$$\zeta(x_1, \ldots, x_n) = T_n\big(\xi_1(x_1), \ldots, \xi_n(x_n)\big) = min\{\xi_1(x_1), \ldots, \xi_n(x_n)\} \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Lemma 1. *A fuzzy vector $\boldsymbol{x}^\star$ is obtained via minimum-t-norm when the individual values of the variables x_i are fuzzy numbers $x_i^\star$. Through the minimum-t-norm, the combination of n fuzzy numbers with characterizing functions $\xi_i(\cdot), i = 1(1)n$, a fuzzy vector $\boldsymbol{x}^\star = (x_1, \ldots, x_n)^\star$ is obtained. In this case, the following holds:*

$$C_\delta\big[\zeta(\cdot, \ldots, \cdot)\big] = \times_{i=1}^{n} C_\delta\big[\xi_i(\cdot)\big] \quad \forall \delta \in (0,1]$$

In words, the δ-cuts of the fuzzy vector $\boldsymbol{x}^\star = (x_1, \ldots, x_n)^\star$ are the Cartesian products of the δ-cuts of the fuzzy numbers $x_i^\star, i = 1(1)n$.

Proof.

$$\begin{aligned} C_\delta\big[\zeta(\cdot, \ldots, \cdot)\big] &= \{\mathbf{x} \in \mathbb{R}^n : \zeta(\mathbf{x} \geq \delta\} \\ &= \{\mathbf{x} : min\{\xi_1(x_1), \ldots, \xi_n(x_n)\} \geq \delta\} \\ &= \times_{i=1}^{n} C_\delta\big[\xi_i(\cdot)\big] \end{aligned}$$

□

The concepts of combined fuzzy samples and triangular norms are useful for succinct multivariate statistical analysis of fuzzy data.

3. Fuzzy multivariate data

For multivariate continuous data, i.e. one observation with k variables (dimensions), an idealized measurement results in a k-dimensional real vector $(x_1, \ldots, x_k)$. In reality, there are two possibilities:

First, when the individual values of the variables x_i are fuzzy numbers $x_i^\star$, a vector of fuzzy numbers $(x_1^\star, \ldots, x_k^\star)$ is obtained. In this case, it is necessary to combine the fuzzy numbers $x_1^\star, \dots, x_k^\star$ to obtain a fuzzy element $(x_1, \dots, x_k)^\star$ of the observation space $M_x^k \in \mathbb{R}^k$. This combination is accomplished by a suitable t-norm T and its extension T_k on $[0,1]^k$.

Second, when the vector itself is fuzzy, hence, a fuzzy version of a vector $(x_1, \dots, x_k)^\star$, e.g. fuzzy point in $\mathbb{R}^k$, is obtained. For example, ideally, the position of a ship on a radar screen is a two-dimensional vector (x_1, x_2) or $(x, y) \in \mathbb{R}^2$. In real situations, such position is characterized by a light point on the radar screen, which is not a precise vector, i.e. fuzzy vector.

Definition 3. Using the notation $\mathbf{x} = (x_1, \dots, x_k)$, a k-dimensional fuzzy vector $\mathbf{x}^\star$ is determined by its so-called vector-characterizing function $\zeta(\cdot, \dots, \cdot)$, which is a real function of k real variables $x_1, \dots, x_k$ and possesses the following properties (Viertl and Sunanta 2013):

(1) $\zeta : \mathbb{R}^k \to [0,1]$

(2) The support of $\zeta(\cdot, \dots, \cdot)$ is a bounded set.

(3) For all $\delta \in$ (0,1], the δ-cut $C_\delta\big[\mathbf{x}^\star\big]$, defined by
$$C_\delta\big[\mathbf{x}^\star\big] := \{\mathbf{x} \in \mathbb{R}^k : \zeta(\mathbf{x}) \geq \delta\},$$
is non-empty and a finite union of simply connected and closed sets.

The mathematical formalization of the combined fuzzy samples and that of the fuzzy version of a vector are described here respectively.

If only fuzzy coordinates $x_j^\star, j = 1(1)k$ are available, in order to obtain the vector-characterizing function of the generated fuzzy vector, the characterizing functions of the fuzzy coordinates $x_j^\star$ have to be combined. In this case, the vector-characterizing function of a fuzzy version of a vector $(x_1, \dots, x_k)^\star$ can be derived through the product-t-norm:
$$\zeta(x_1, \dots, x_n) \quad := \quad \textstyle\prod_{j=1}^k \xi_j(x_j) \qquad \forall (x_1, \dots, x_k) \in \mathbb{R}^k$$
For the special case k=2,that is when fuzzy coordinates $x_1^\star$ and $x_2^\star$ of a vector, with characterizing functions $\xi_1(\cdot)$ and $\xi_2(\cdot)$ respectively, are measured, the vector-characterizing function $\zeta(\cdot, \cdot)$ of the fuzzy vector $\mathbf{x}^\star = (x_1, x_2)^\star$ is obtained via the product-t-norm T_{prod}, i.e.
$$\zeta(x_1, x_2) = T_{prod}\big(\xi_1(x_1), \xi_2(x_2)\big) = \xi_1(x_1) \cdot \xi_2(x_2) \qquad \forall (x_1, x_2) \in \mathbb{R}^2$$
The result is a fuzzy vector, denoted as $(x_1, \dots, x_k)^\star$ with a vector-characterizing function $\zeta(x_1, x_2) = \xi_1(x_1) \cdot \xi_2(x_2) \quad \forall (x_1, x_2) \in \mathbb{R}^2$ where $\xi_1(\cdot)$ and $\xi_2(\cdot)$ are the characterizing functions of the fuzzy coordinates $x_1^\star$ and $x_2^\star$ respectively.

In case of a sample of fuzzy observations $\mathbf{x}_i^\star, i = 1(1)n$, with k-dimensional vector-characterizing functions $\zeta_i(x_1, \dots, x_k)$, the generalized minimum rule is applied in oder to obtain the vector-characterizing function $\zeta(x_1, \dots, x_{k \cdot n})$ for the combined fuzzy vector $\mathbf{X}^\star$, which is the combined fuzzy sample. Considering a fuzzy sample $\mathbf{x}_1^\star, \dots, \mathbf{x}_n^\star$ with its corresponding vector-characterizing functions $\zeta_i(\cdot, \dots, \cdot)$, where $\mathbf{x} = (x_1, \dots, x_k) \in \mathbb{R}^k$ and sample space $\mathbb{R}^{k \cdot n}$, the followings can be defined:
Let $\mathbf{x}_i^\star \in \mathcal{F}(\mathbb{R}^k), i = 1(1)n$ and the combined fuzzy sample $\mathbf{X}^\star \in \mathcal{F}(\mathbb{R}^{k \cdot n})$, the vector-characterizing function $\zeta(\cdot, \dots, \cdot)$ of $\mathbf{X}^\star$, which is a function of $k{\cdot}n$ real variables, is obtained via a triangular norm (t-norm T), i.e.
$$\zeta(x_1, \dots, x_{k \cdot n}) = T\big(\zeta_1(x_1, \dots, x_k), \zeta_2(x_{k+1}, \dots, x_{2k}), \dots, \zeta_n(x_{(n-1)k+1}, \dots, x_{k \cdot n})\big).$$
Applying the minimum-t-norm, the vector-characterizing function $\zeta(\cdot, \dots, \cdot)$ of $\mathbf{X}^\star$ is obtained:
$$\zeta(x_1, \dots, x_{k \cdot n}) = \min_{i=1(1)n)} \{\zeta_i(x_{(i-1)k+1}, x_{(i-1)k+2}, \dots, x_{(i-1)k+k})\} \qquad \forall (x_1, \dots, x_{k \cdot n}) \in \mathbb{R}^{k \cdot n}$$
For example, let $k = 2$, i.e. $\mathbf{x}_i^\star = (x_{1 \cdot i}, x_{2 \cdot i})^\star \;\; \widehat{=} \;\; \zeta(\cdot, \cdot), i = 1(1)n$ (i.e. n observations), the combined fuzzy sample $\mathbf{X}^\star$ is obtained by:
$$\mathbf{X}^\star = (x_1, x_2, \dots, x_{1 \cdot n}, x_{2 \cdot n})^\star \quad \widehat{=} \quad \zeta(x_1, x_2, \dots, x_{1 \cdot n}, x_{2 \cdot n}) \text{ where } \zeta : \mathbb{R}^{2 \cdot n} \to [0,1]$$
In this case, the vector-characterizing function $\zeta(\cdot, \dots, \cdot)$ of the combined fuzzy sample is obtained in the following way:

$$\zeta(x_1, x_2, \ldots, x_{1\cdot n}, x_{2\cdot n}) := \min_{i=1(1)n} \{\zeta_i(x_{1\cdot i}, x_{2\cdot i})\}$$
$$= \min\{\zeta_1(x_1, x_2), \ldots, \zeta_n(x_{1\cdot n}, x_{2\cdot n})\} \quad \forall (x_1, x_2, \ldots, x_{1\cdot n}, x_{2\cdot n}) \in \mathbb{R}^{2\cdot n}$$

Lemma 2. *Through the combination of n fuzzy observaions* $\boldsymbol{x}_i^\star, i = 1(1)n$ *of a k-dimensional fuzzy quantity with vector-characterizing functions* $\zeta_i(\cdot)$ *by the minimum-t-norm, n fuzzy k-dimensional vectors are combined into an* $(k{\cdot}n)$*-dimensional fuzzy vector with vector-characterizing function* $\zeta(\cdot, \ldots, \cdot)$*, for which the following property holds:*

$$C_\delta\big[\zeta(\cdot, \ldots, \cdot)\big] = \times_{i=1}^n C_\delta\big[\zeta_i(\cdot)\big] \qquad \forall \delta \in (0, 1]$$

where $\zeta(x_1, \ldots, x_{k\cdot n}) = \min\limits_{i=1(1)n} \{\zeta_1(x_1, \ldots, x_k), \zeta_2(x_{k+1}, \ldots, x_{2k}), \ldots, \zeta_n(x_{(n-1)k+1}, \ldots, x_{k\cdot n})\}$
$\forall (x_1, \ldots, x_{k\cdot n}) \in \mathbb{R}^{k\cdot n}$
The δ*-cuts of the combined fuzzy vector* $\boldsymbol{X}^\star$ *are the Cartesian products of the* δ*-cuts of the fuzzy vectors* $\boldsymbol{x}_i^\star, i = 1(1)n$.

Proof. $C_\delta\big[\zeta(\cdot, \ldots, \cdot)\big] = \{\mathbf{x} \in \mathbb{R}^{k\cdot n} : \zeta(\mathbf{x}) \geq \delta\}$
$= \{\mathbf{x} : \min\{\zeta_1(x_1, \ldots, x_k), \zeta_2(x_{k+1}, \ldots, x_{2k}), \ldots, \zeta_n(x_{(n-1)k+1}, \ldots, x_{k\cdot n})\} \geq \delta\}$,
by $\mathbf{x} \in \times_{i=1}^n C_\delta\big[\zeta_i(\cdot)\big] \iff \zeta_i(\cdot) \geq \delta \quad \forall i = 1(1)n \iff \min\limits_{i=1(1)n} \{\zeta_i(\cdot)\} \geq \delta$, and
it follows $C_\delta\big[\zeta(\cdot, \ldots, \cdot)\big] = \times_{i=1}^n C_\delta\big[\zeta_i(\cdot)\big]$. □

4. Statistical functions of fuzzy multivariate data

Data analysis methods generally use functions $f(x_1, \ldots, x_n), f : \mathbb{R}^n \to$ S, where S is a suitable adopted space, i.e. statistics, to summarize and explain data. In case of non-precise data $x_1^\star, \ldots, x_n^\star$, the function becomes $f(x_1^\star, \ldots, x_k^\star)$. The values $y^\star = f(x_1^\star, \ldots, x_n^\star)$ are non-precise by the propagation of imprecision (fuzziness). The membership function of $y^\star$ is obtained via the extension principle based on combined fuzzy samples and minimum t-norm (see Definition 2). Through the combined fuzzy sample $\mathbf{X}^\star$ (or the fuzzy vector $\mathbf{x}^\star$) with its vector-characterizing function $\zeta(\cdot, \ldots, \cdot)$, the generalization of statistics $f(x_1, \ldots, x_n)$ to handle fuzzy data $x_1^\star, \ldots, x_n^\star$ is possible:

Using the notation $\mathbf{x}^\star = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $\eta(\cdot)$ for the membership function of $y^\star$, the values $\eta(y)$, $\forall y \in$ S, are given by

$$\eta(y) = \left\{ \begin{array}{rl} sup\{\zeta(x_1, \ldots, x_n) : f(x_1, \ldots, x_n) = y\} & \text{if } \exists (x_1, \ldots, x_n) : f(x_1, \ldots, x_n) = y \\ 0 & \text{if } \nexists (x_1, \ldots, x_n) : f(x_1, \ldots, x_n) = y \end{array} \right\}. \quad (1)$$

For a fuzzy vector $\mathbf{x}^\star$ in $\mathbb{R}^n$ with vector-characterizing function $\zeta(\cdot, \ldots, \cdot)$, all δ-cuts $C_\delta\big[\mathbf{x}^\star\big]$ are defined as simply connected and compact subsets of $\mathbb{R}^n$. For continuous functions, the following holds:
Let $x_1^\star, \ldots, x_n^\star$ be fuzzy intervals, i.e. all δ-cuts of $x_i^\star$ are finite closed intervals, and $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function, then $y^\star = f(x_1^\star, \ldots, x_n^\star)$ is a fuzzy interval whose δ-cuts are given by

$$C_\delta\big[y^\star\big] = \big[\min\nolimits_{\mathbf{x}\in C_\delta[\mathbf{x}^\star]} f(\mathbf{x}), \max\nolimits_{\mathbf{x}\in C_\delta[\mathbf{x}^\star]} f(\mathbf{x})\big] \qquad \forall \delta \in (0, 1] \quad (2)$$

Proof. Since $x_i^\star, i = 1(1)n$, are fuzzy intervals, the δ-cuts of the combined fuzzy vector $\mathbf{x}^\star$ are Cartesian products of compact intervals. The continuity of $f(\cdot)$ implies that $f(C_\delta\big[\mathbf{x}^\star\big])$ is compact and simply connected, and, therefore, an interval. By the continuity of $f(\cdot)$, it follows that $f^{-1}(\{y\})$ is closed and, thus, $sup_{\mathbf{x}\in C_\delta[\mathbf{x}^\star]} f(\mathbf{x}) = \max_{\mathbf{x}\in C_\delta[\mathbf{x}^\star]} f(\mathbf{x})$. As a result, $C_\delta[\mathbf{x}^\star]$ is a closed interval. □

In multivariate case, $(n{\cdot}k)$-dimensional fuzzy vector (n samples of k variables) with (combined) vector-characterizing function $\zeta(\cdot,\ldots,\cdot)$, the value $y^{\star\star}$ of the function $f:\mathbb{R}^{k\cdot n}\rightarrow \mathrm{S}$ is defined to have the following membership function:

$$\eta(y)=\left\{\begin{array}{rl} sup\{\zeta(x_1,\ldots,x_{k\cdot n}):f(x_1,\ldots,x_{k\cdot n})=y\} & \text{if } \exists(x_1,\ldots,x_{k\cdot n}):f(x_1,\ldots,x_{k\cdot n})=y \\ 0 & \text{if } \nexists(x_1,\ldots,x_{k\cdot n}):f(x_1,\ldots,x_{k\cdot n})=y \end{array}\right\} \quad (3)$$

For a combined fuzzy vector $\mathbf{X}^\star$ in $\mathbb{R}^{k\cdot n}$ with vector-characterizing function $\zeta(\cdot,\ldots,\cdot)$, the value $y^{\star\star}$, where S=$\mathbb{R}$, all of the δ-cuts $C_\delta[\mathbf{X}^\star]$ are defined as simply connected and compact subsets of $\mathbb{R}^{k\cdot n}$. For continuous functions, the following holds:

$$C_\delta\left[y^{\star\star}\right]=\left[\min_{\mathbf{X}\in C_\delta[\mathbf{X}^\star]} f(\mathbf{X}),\max_{\mathbf{X}\in C_\delta[\mathbf{X}^\star]} f(\mathbf{X})\right] \qquad \forall\delta\in(0,1] \quad (4)$$

Note: As defined in Lemma 2, the δ-cuts of the combined fuzzy sample $\mathbf{X}^\star$ are the Cartesian products of the δ-cuts of the fuzzy vectors $\mathbf{x}_i^\star, i=1(1)n$.

5. Generalized point estimators

The essential characteristics of a univariate distribution are the arithmetic mean, as a measure of location, and the variance, as a measure of dispersion. Similarly, considering a data set, the sample mean and sample variance of the sample are important summary measures. Generalized estimators for fuzzy univariate data are explained in Viertl (2006).

In multivariate case, the means and variances of the separate measurements are of interest. An essential aspect of multivariate analysis is the dependence between the different variables, i.e. the dependence between two variables may involve the covariance between them. In this case, the relevant parameters are in forms of vector and/or matrix, e.g. the mean vector, the variance-covariance matrix, and the correlation coefficient matrix.

For a sample of fuzzy vectors $\mathbf{x}_i^\star, i=1(1)n$ in $\mathbb{R}^k$ with vector-characterizing functions $\zeta_i(\cdot,\ldots,\cdot)$, a point estimate from a statistic (function) $\vartheta(\mathbf{x}_1^\star,\ldots,\mathbf{x}_n^\star)$ becomes a fuzzy element $\hat\theta^\star$ in parameter space Θ. Hence, for a combined fuzzy vector $\mathbf{X}^\star$ in $\mathbb{R}^{k\cdot n}$ with vector-characterizing function $\zeta(\cdot,\ldots,\cdot)$, the generalized point estimator based on a standard estimator $\vartheta:\mathbb{R}^{k\cdot n}\rightarrow\Theta$ yields a fuzzy point estimate $\hat\theta^\star$ whose membership function $\eta(\cdot)$ is defined by its values as follows:

$$\eta(\theta)=\left\{\begin{array}{cl} sup\{\zeta(x_1,\ldots,x_{k\cdot n}):\vartheta(x_1,\ldots,x_{k\cdot n})=\theta\} & \text{if } \exists(x_1,\ldots,x_{k\cdot n}):\vartheta(x_1,\ldots,x_{k\cdot n})=\theta \\ 0 & \text{if } \nexists(x_1,\ldots,x_{k\cdot n}):\vartheta(x_1,\ldots,x_{k\cdot n})=\theta \end{array}\right\} \quad (5)$$

All δ-cuts $C_\delta[\mathbf{X}^\star]$ are assumed to be simply connected and compact subsets of $\mathbb{R}^{k\cdot n}$. For continuous functions $\vartheta(\cdot,\ldots,\cdot)$ and one-dimensional parameter θ, the following holds:

$$C_\delta\left[\hat\theta^\star\right]=\left[\min_{\mathbf{X}\in C_\delta[\mathbf{X}^\star]}\vartheta(\mathbf{X}),\max_{\mathbf{X}\in C_\delta[\mathbf{X}^\star]}\vartheta(\mathbf{X})\right] \qquad \forall\delta\in(0,1] \quad (6)$$

or symbolically $\hat\theta^\star=\vartheta(\mathbf{X}^\star)$

Proof. The δ-cuts of the combined fuzzy sample $\mathbf{X}^\star$ are the Cartesian products of the δ-cuts of the fuzzy vectors $\mathbf{x}_i^\star, i=1(1)n$ (see Lemma 2). The continuity of $\vartheta(\cdot)$ implies that $\vartheta(C_\delta[\mathbf{X}^\star])$ is compact and simply connected, therefore, an interval. By the continuity of $\vartheta(\cdot)$, it also follows that $\vartheta^{-1}(\{\hat\theta\})$ is closed for all $\hat\theta\in\Theta$ and, thus, $sup_{\mathbf{X}\in C_\delta[\mathbf{X}^\star]}\vartheta(\mathbf{X})=\max_{\mathbf{X}\in C_\delta[\mathbf{X}^\star]}\vartheta(\mathbf{X})$. As a result, $C_\delta[\mathbf{X}^\star]$ is a closed interval. □

5.1. Point estimator for mean vector

The arithmetic mean is one of the important parameters that characterize the distribution. In multivariate analysis, the mean vector $\boldsymbol{\mu}$ is of interest, where

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix}.$$

Let $\mathbf{X}^\star$ be a combined fuzzy vector in $\mathbb{R}^{k \cdot n}$ with vector-characterizing function $\zeta(\cdot, \dots, \cdot)$, the δ-cut $C_\delta[\mathbf{X}^\star]$ is defined as simply connected and compact subset of $\mathbb{R}^{k \cdot n}$.

For continuous functions $f : \mathbb{R}^{k \cdot n} \to \mathbb{R}$ with fuzzy point estimate $\hat{\mu}_j^\star$ for the mean, the following holds:

$$C_\delta\left[\hat{\mu}_j^\star\right] = \left[\min_{\mathbf{X} \in C_\delta[\mathbf{X}^\star]} f_{\mu_j}(\mathbf{X}), \max_{\mathbf{X} \in C_\delta[\mathbf{X}^\star]} f_{\mu_j}(\mathbf{X})\right] \qquad \forall \delta \in (0, 1]. \tag{7}$$

Remark 1. Minimum and maximum of $f_{\mu_j}(\mathbf{X})$ are defined as: $\min f_{\mu_j}(\mathbf{X}) = f_{\mu_j}(\underline{\mathbf{X}}) = \frac{1}{n}\sum_{i=1}^n \underline{x}_{ij,\delta}$ and $\max f_{\mu_j}(\mathbf{X}) = f_{\mu_j}(\overline{\mathbf{X}}) = \frac{1}{n}\sum_{i=1}^n \overline{x}_{ij,\delta}$, where $\underline{x}_{ij,\delta}$ and $\overline{x}_{ij,\delta}$ are the lower and upper limits of the δ-cut of a fuzzy interval at a specified δ-level, respectively.

Accordingly, the membership function $\eta(\cdot)$ (as generalized in Equation (5)) of the fuzzy point estimate $\hat{\mu}_j^\star$ can be constructed numerically through lower and upper limits of multiple δ-cuts as well.

5.2. Point estimator for variance-covariance matrix

The sample variance measures how far a set of numbers is spread out. Mathematically, it is simply the average of the squares of the deviations of observations about their sample mean. On the other hand, sample covariance is defined in terms of cross products between two variables (factors) of interest. The sample variance-covariance matrix is an unbiased estimator of the covariance matrix $\boldsymbol{\Sigma}$. For a standard sample $\mathbf{x}^\star = (x_1, \dots, x_{k \cdot n})$, each ij-element in the matrix of sums of squares and cross products of deviations about the sample mean $\bar{\mathbf{x}}$ is generally defined as:

$$\mathbf{A}_{ixj} : a_{ij} = \sum_{i=1}^n (x_{ij} - \bar{\mathbf{x}})(x_{ij} - \bar{\mathbf{x}})'; \text{ where } i = 1(1)n,\ j = 1(1)k, \text{ and } n > k. \tag{8}$$

Remark 2. The constraint $n > k$ is necessary, so that $\mathbf{A}$ is positive definite (Anderson 2003).

The standard point estimator of the variance σ_i^2 is
$\hat{\sigma}_i^2 = (1/(n-1)) \cdot a_{ij} = (1/(n-1)) \sum_{i=1}^n (x_{ij} - \bar{\mathbf{x}}_j)(x_{ij} - \bar{\mathbf{x}}_j)' = (1/(n-1))(\sum_{i=1}^n x_{ij}^2 - n\bar{x}_j^2)$.

Hence, the estimated variance-covariance matrix is derived as:

$$\widehat{\boldsymbol{\Sigma}}_{lxm} : \hat{\sigma}_{lm} = (1/(n-1)) \sum_{i=1}^n (x_{il} - \bar{x}_l)(x_{im} - \bar{x}_m);\ l = 1(1)k,\ m = 1(1)k. \tag{9}$$

Let $\mathbf{X}^\star$ be a combined fuzzy vector in $\mathbb{R}^{k \cdot n}$ with vector-characterizing function $\zeta(\cdot, \dots, \cdot)$, δ-cut $C_\delta[\mathbf{X}^\star]$ is defined as simply connected and compact subset of $\mathbb{R}^{k \cdot n}$. For continuous standard estimation $f : \mathbb{R}^{k \cdot n} \to \mathbb{R}$ and the corresponding fuzzy point estimate $\hat{\sigma}_{lm}^\star$ for the variance-covariance, the following holds:

$$C_\delta\left[\hat{\sigma}_{lm}^\star\right] = \left[\min_{\mathbf{X} \in C_\delta[\mathbf{X}^\star]} f_{\sigma_{lm}}(\mathbf{X}), \max_{\mathbf{X} \in C_\delta[\mathbf{X}^\star]} f_{\sigma_{lm}}(\mathbf{X})\right] \qquad \forall \delta \in (0, 1]. \tag{10}$$

For sample variance-covariance matrix, minimum and maximum of $f_{\sigma_{lm}}(\mathbf{X})$ are defined as:

$$\min f_{\sigma_{lm}}(\mathbf{X}) = \min \hat{\sigma}_{lm}(\mathbf{x}_{min}) \text{ and } \max f_{\sigma_{lm}}(\mathbf{X}) = \max \hat{\sigma}_{lm}(\mathbf{x}_{max}),$$

where $\mathbf{x}_{min}$ and $\mathbf{x}_{max}$ vectors contain the observations that lie closest to and farthest from the corresponding sample means, hence, resulted in minimum and maximum sample variances (covariances) at the specified δ-level respectively.

5.3. Point estimator for correlation matrix

It is frequently informative to separate the information contained in variances from that contained in measures of association and, in particular, the measure of association known as the population correlation coefficient (ρ_{lm}) Wichern and Johnson (2007). Symbolically, the estimated fuzzy sample correlation coefficients are obtained from the following estimator (11), in a similar manner as the standard sample correlation coefficient (r_{lm}), but in forms of $\min r_{lm}$ and $\max r_{lm}$, where l and m represent two distinct variables of interest:

$$f_{\rho_{lm}}(\mathbf{X}) = r_{lm} = \frac{\sum_{i=1}^{n}(x_{il}-\bar{x}_l)(x_{im}-\bar{x}_m)}{\sqrt{\sum_{i=1}^{n}(x_{il}-\bar{x}_l)^2}\sqrt{\sum_{i=1}^{n}(x_{im}-\bar{x}_m)^2}} \tag{11}$$

Applying the δ-cuts $C_\delta[\mathbf{X}^\star]$, the lower and upper boundaries of the estimated sample correlation coefficients are obtained through the following simple linear programs (Liu and Kao 2002):

$$\min r_{lm,\delta} = \left\{ \begin{array}{c} \min \frac{\sum_{i=1}^{n}(x_{il,\delta}-\bar{x}_{l,\delta})(x_{im,\delta}-\bar{x}_{m,\delta})}{\sqrt{\sum_{i=1}^{n}(x_{il,\delta}-\bar{x}_{l,\delta})^2}\sqrt{\sum_{i=1}^{n}(x_{im,\delta}-\bar{x}_{m,\delta})^2}} \\ \text{s.t. } \underline{x}_{il,\delta} \leq x_{il,\delta} \leq \overline{x}_{il,\delta} \quad \forall l \\ \underline{x}_{im,\delta} \leq x_{im,\delta} \leq \overline{x}_{im,\delta} \quad \forall m \\ l = 1(1)k,\ m = 1(1)k \end{array} \right\} \tag{12}$$

$$\max r_{lm,\delta} = \left\{ \begin{array}{c} \max \frac{\sum_{i=1}^{n}(x_{il,\delta}-\bar{x}_{l,\delta})(x_{im,\delta}-\bar{x}_{m,\delta})}{\sqrt{\sum_{i=1}^{n}(x_{il,\delta}-\bar{x}_{l,\delta})^2}\sqrt{\sum_{i=1}^{n}(x_{im,\delta}-\bar{x}_{m,\delta})^2}} \\ \text{s.t. } \underline{x}_{il,\delta} \leq x_{il,\delta} \leq \overline{x}_{il,\delta} \quad \forall l \\ \underline{x}_{im,\delta} \leq x_{im,\delta} \leq \overline{x}_{im,\delta} \quad \forall m \\ l = 1(1)k,\ m = 1(1)k \end{array} \right\} \tag{13}$$

For continuous functions $f : \mathbb{R}^{k \cdot n} \to \mathbb{R}$ with fuzzy point estimate $\hat{\rho}^\star_{lm}$ for the correlation coefficient, where l and m represent two distinct variables of interest, the following holds respectively:

$$C_\delta\left[\hat{\rho}^\star_{lm}\right] = \left[\min_{\mathbf{X}\in C_\delta[\mathbf{X}^\star]} f_{\rho_{lm}}(\mathbf{X}), \max_{\mathbf{X}\in C_\delta[\mathbf{X}^\star]} f_{\rho_{lm}}(\mathbf{X})\right] \qquad \forall \delta \in (0,1] \tag{14}$$

For sample correlation matrix, minimum and maximum of $f_{\rho_{lm}}(\mathbf{X})$ are defined as: $\min f_{\rho_{lm}}(\mathbf{X}) = \min r_{lm}$ and $\max f_{\rho_{lm}}(\mathbf{X}) = \max r_{lm}$. Through the lower and upper boundaries of multiple δ-cuts, i.e. $\min r_{lm,\delta}$ and $\max r_{lm,\delta}$, the membership function $\eta_{lm}(\cdot)$ of the fuzzy point estimate $\hat{\rho}^\star_{lm}$ can be obtained numerically.

Example. As a simplified example, a part of the results of a full-factorial experiment is adapted. Consider results of flat lapping measurements of stainless steel specimens (Table 1), where the process parameters of interest are obtained surface roughness, measured by mean of arithmetic mean Ra (x10^3 mm) $x_1^\star$, and calculated material removal rate MRR (x10^{-3} mm/s) $x_2^\star$. These measurements are naturally uncertain, i.e. considered as fuzzy numbers $x_n^\star = [a,b,c,d]_n$ (intuitively with trapezoidal-shape membership functions) and consequently they are further together considered as vector of fuzzy number $\mathbf{x}_i^\star = [x_1^\star, x_2^\star]$ where $i = 1,2,\ldots,8$ (two variables of eight observations).

Table 1: Results of flat lapping measurements

Expt.	Ra (1000th-mm)				MRR (1000th-mm/s)			
No.	a	b	c	d	a	b	c	d
1	0.001	0.051	0.101	0.151	0.022	0.037	0.057	0.077
2	0.021	0.071	0.121	0.171	0.176	0.191	0.211	0.231
3	0.025	0.075	0.125	0.125	0.008	0.025	0.043	0.065
4	0.032	0.082	0.132	0.182	0.005	0.020	0.040	0.060
5	0.038	0.088	0.132	0.188	0.001	0.016	0.036	0.056
6	0.045	0.095	0.145	0.195	0.013	0.028	0.048	0.068
7	0.052	0.102	0.152	0.202	0.167	0.182	0.202	0.222
8	0.058	0.108	0.158	0.208	0.026	0.041	0.061	0.081

Through the method of combining fuzzy samples described in Section 3, the corresponding combined fuzzy vector $\mathbf{X}^{\star} = (x_{1\cdot i}, x_{2\cdot i})^{\star}$ is obtained along with its vector-characterizing function $\zeta(x_1, x_2, \ldots, x_{1\cdot 8}, x_{2\cdot 8})$. As defined in equation (7), the fuzzy point estimate $\hat{\mu}_j^{\star}$ for the mean can be determined as lower and upper boundaries of the estimated sample mean at different k δ-levels $\left[\overline{\mathbf{X}}_{\delta_k}^{L}, \overline{\mathbf{X}}_{\delta_k}^{U}\right]$ through $\overline{\mathbf{X}}_{\delta_k}^{L} = \min f_{\mu_j}(\mathbf{X}) = f_{\mu_j}(\underline{\mathbf{X}}) = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_{ij,\delta_k}$ and $\overline{\mathbf{X}}_{\delta_k}^{U} = \max f_{\mu_j}(\mathbf{X}) = f_{\mu_j}(\overline{\mathbf{X}}) = \frac{1}{n}\sum_{i=1}^{n} \overline{x}_{ij,\delta_k}$:

$$\overline{\mathbf{X}}_{\delta=0+} = \begin{bmatrix}[0.034, 0.178]\\ [0.052, 0.107]\end{bmatrix}$$

$$\overline{\mathbf{X}}_{\delta=0.25} = \begin{bmatrix}[0.046, 0.167]\\ [0.052, 0.102]\end{bmatrix}$$

$$\overline{\mathbf{X}}_{\delta=0.50} = \begin{bmatrix}[0.059, 0.156]\\ [0.060, 0.097]\end{bmatrix}$$

$$\overline{\mathbf{X}}_{\delta=0.75} = \begin{bmatrix}[0.071, 0.145]\\ [0.063, 0.092]\end{bmatrix}$$

$$\overline{\mathbf{X}}_{\delta=1} = \begin{bmatrix}[0.084, 0.134]\\ [0.067, 0.087]\end{bmatrix}$$

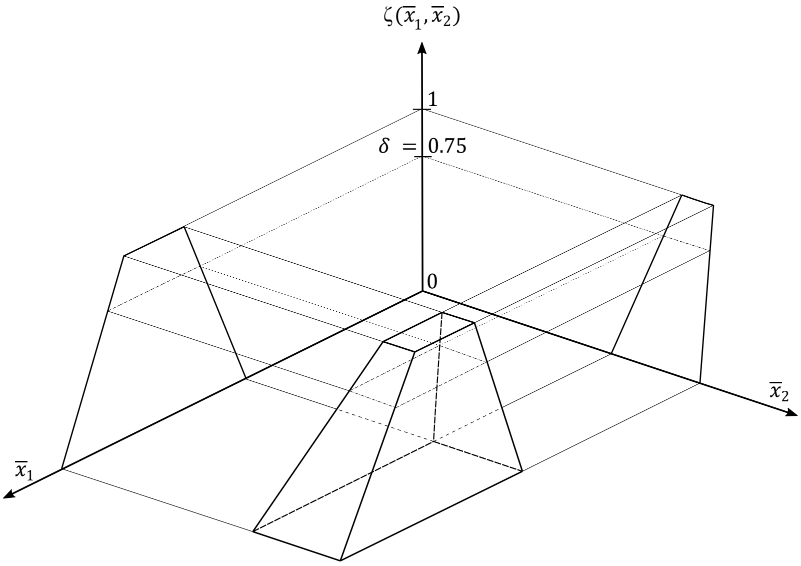

Figure 1: Vector-characterizing function $\zeta(\bar{x}_1, \bar{x}_2)$ of the mean of combined fuzzy vector $\mathbf{X}^{\star}$

Note: At each δ-level, a plane of rectangle is formed, e.g. $\overline{\mathbf{X}}_{\delta=0.75} =$ [0.071,0.145] x [0.063,0.092].

As shown in Figure 1, vector-characterizing function of the mean of the combined fuzzy vector $\mathbf{X}^{\star}$ is of a polygon form. For each δ-cut, a plane representing feasible mean values is obtained as a result.

Next, the search for the minimum and maximum of the variance-covariance continues, i.e. through the feasible regions that are closest/farthest from the means and those with resulting minimum/maximum cross products for the variance-covariance matrix. As defined in Equation (10), the lower and upper boundaries of the δ-cut of the corresponding sample point estimate for fuzzy variance-covariance at different δ-levels are

$$\begin{array}{ccc} & \min \widehat{\boldsymbol{\Sigma}}_{\delta_k} & \max \widehat{\boldsymbol{\Sigma}}_{\delta_k} \\ \widehat{\boldsymbol{\Sigma}}_{\delta=0+} = & \begin{bmatrix} 0.00034 & -0.00018 \\ & 0.00144 \end{bmatrix} & \begin{bmatrix} 0.03095 & 0.00009 \\ & 0.01641 \end{bmatrix} \\ \widehat{\boldsymbol{\Sigma}}_{\delta=0.25} = & \begin{bmatrix} 0.00034 & 0.00003 \\ & 0.00181 \end{bmatrix} & \begin{bmatrix} 0.02228 & 0.00009 \\ & 0.01316 \end{bmatrix} \\ \widehat{\boldsymbol{\Sigma}}_{\delta=0.50} = & \begin{bmatrix} 0.00034 & 0.00004 \\ & 0.00198 \end{bmatrix} & \begin{bmatrix} 0.01503 & 0.00009 \\ & 0.01219 \end{bmatrix} \\ \widehat{\boldsymbol{\Sigma}}_{\delta=0.75} = & \begin{bmatrix} 0.00034 & 0.00005 \\ & 0.00251 \end{bmatrix} & \begin{bmatrix} 0.00922 & 0.00009 \\ & 0.01034 \end{bmatrix} \\ \widehat{\boldsymbol{\Sigma}}_{\delta=1} = & \begin{bmatrix} 0.00022 & -0.00001 \\ & 0.00321 \end{bmatrix} & \begin{bmatrix} 0.00483 & 0.00009 \\ & 0.00866 \end{bmatrix}. \end{array}$$

Note: Since both combined sample itself and the mean are fuzzy, with respect to the shapes of their vector-characterizing functions, for different δ-cuts, the distances between the boundaries of the fuzzy sample and those of the fuzzy mean remain the same, i.e. the boundaries (limits) of the functions move in the same direction with the same distance. Hence, the same value of the min/max variances at different δ-levels of $x_1^\star$ and $x_2^\star$ respectively.
Finally, after applying the δ-cuts as defined in Equation (14), the lower and upper boundaries of the estimated sample correlations $\min r_{lm,\delta}$ and $\max r_{lm,\delta}$ are obtained:

$$\begin{array}{ccc} & \min r_{lm,\delta} & \max r_{lm,\delta} \\ r_{lm,\delta=0+} = & \begin{bmatrix} 1 & -0.00784 \\ & 1 \end{bmatrix} & \begin{bmatrix} 1 & 0.16462 \\ & 1 \end{bmatrix} \\ r_{lm,\delta=0.25} = & \begin{bmatrix} 1 & 0.00152 \\ & 1 \end{bmatrix} & \begin{bmatrix} 1 & 0.11284 \\ & 1 \end{bmatrix} \\ r_{lm,\delta=0.50} = & \begin{bmatrix} 1 & 0.00277 \\ & 1 \end{bmatrix} & \begin{bmatrix} 1 & 0.10828 \\ & 1 \end{bmatrix} \\ r_{lm,\delta=0.75} = & \begin{bmatrix} 1 & 0.00502 \\ & 1 \end{bmatrix} & \begin{bmatrix} 1 & 0.09651 \\ & 1 \end{bmatrix} \\ r_{lm,\delta=0.1} = & \begin{bmatrix} 1 & -0.00168 \\ & 1 \end{bmatrix} & \begin{bmatrix} 1 & 0.10530 \\ & 1 \end{bmatrix} \end{array}$$

The resulting vector-characterizing function is also of a complex form due to fuzzy multiplication and division. As opposed to the maximum covariances, the different in maximum correlations are a result of different minimums of variance products. However, the quantitative values of covariance and correlation follow the same trend in the end. As indicated by the numerical values of correlation coefficient, based on the experimental results, the linear relationship between obtained surface roughness and material removal rate is trivial. In lapping theory, there are many factors that influence the quality of the obtained surface. In other words, a combination of factors, e.g. surface flatness, surface area, abrasive compound, lapping mechanism through rolling, scraping, and embedding, are needed to be considered altogether.

6. Final remarks

The standard statistical inference methods are extended for multivariate fuzzy data. In particular, the concepts of fuzzy vector and fuzzy multivariate data are described. The method of combining n observations of k-dimension is defined and served as focal point of the subsequent definitions. As a result, the combined fuzzy sample along with its corresponding vector-characterizing function is obtained. Accordingly, the generalized methods are then defined for the estimators of sample mean, variance-covariance, and correlation coefficient matrices for multivariate fuzzy data. A numerical example of a two-variable case through δ-cut calculation is also provided. Because of increasing cardinality of the resulting Cartesian products, the higher the dimensions of considering variables are, the more complex the analysis it is. In other words, *nonlinearity* and *nonconvexity* may be expected from multivariate analysis. Hence, the results and interpretation depend rather on the adopted numerical methods and the data themselves as well. The proposed generalized method for multivariate fuzzy data provides a succinct alternative of mathematical definitions for the estimation of different statistics and can be further extended to the generalization of estimation of confidence regions, statistical tests, and Bayes' theorem in a natural way.

References

Anderson TW (2003). *An Introduction to Multivariate Statistical Analysis.* Wiley, NJ.

Blanco-Fernàndez A, *et al.* (2014). "A Distance-based Statistical Analysis of Fuzzy Number-valued Data." *International Journal of Approximate Reasoning*, **55**, 1487–1501.

Bustince H, Burillo P (1995). "Correlation of Interval-valued Intuitionistic Fuzzy Sets." *Fuzzy Sets and Systems*, **74**, 237–244.

Dubois D, Prade H (1987). "The Mean Value of a Fuzzy Number." *Fuzzy Sets and Systems*, **24**, 279–300.

Frühwirth-Schnatter S (1992). "Statistical Inference for Fuzzy Data." *Fuzzy Sets and Systems*, **50**, 143–165.

Klir G, Yuan B (1995). *Fuzzy Sets and Fuzzy Logic – Theory and Applications.* Wiley, NJ.

Körner R (1997). "On the Variance of Fuzzy Random Variables." *Fuzzy Sets and Systems*, **92**, 83–93.

Kruse R (1984). "Statistical Estimation with Linguistic Data." *Information Science*, **33**, 197–207.

Kruse R (1987). "On the Variance of Random Sets." *Journal of Math. Analysis and Applications*, **122**, 469–473.

Kruse R, Meyer KD (1987). *Statistics with Vague Data.* D. Reidel Publishing, Dordrecht.

Liu ST, Kao C (2002). "Fuzzy Measures for Correlation Coefficient of Fuzzy Numbers." *Fuzzy Sets and Systems*, **128**, 267–275.

Nguyen HT, Wu B (2006). "Random and Fuzzy Sets in Coarse Data Analysis." *Computational Statistics and Data Analysis*, **51**(1), 70–85.

Ni Y, Cheung JY (2003). "Correlation Coefficient Estimate for Fuzzy Data." *Intelligent Systems Design and Applications: Advances in Soft Computing*, **23**, 105–114.

Saneifard R, Saneifard R (2012). "Correlation Coefficient between Fuzzy Numbers Based on Central Interval." *Journal of Fuzzy Set Valued Analysis*, **2012**, 1–9.

Viertl R (2006). "Univariate Statistical Analysis with Fuzzy Data." *Journal of Computational Statistics and Data Analysis*, **51**(1), 133–147.

Viertl R (2011). *Statistical Methods for Fuzzy Data.* Wiley, Chichester.

Viertl R (2015). "Measurement of Continuous Quantities and Their Statistical Evaluation." *Austrian Journal of Statistics*, **44**, 25–32.

Viertl R, Hareter D (2006). *Beschreibung und Analyse unscharfer Information.* Springer, Wien.

Viertl R, Sunanta O (2013). "Fuzzy Bayesian Inference." *METRON*, **71**(3), 207–216.

Voxman W (1998). "Some Remarks on Distances between Fuzzy Numbers." *Fuzzy Sets and Systems*, **100**, 353–365.

Wang D (2004). "A Note on Consistency and Unbiasedness of Point Estimation with Fuzzy Data." *Metrika*, **60**, 93–104.

Wichern DW, Johnson RA (2007). *Applied Multivariate Statistical Analysis.* Pearson Prentice Hall, NJ.

Zadeh LA (1965). "Fuzzy Sets." *Information and Control*, **8**(1), 338–353.

Zadeh LA (1975). "The Concept of Linguistic Variable and Its Application to Approximate Reasoning." *Information Science*, **8**, 43–80.

Affiliation:

Owat Sunanta
Institute of Statistics and Mathematical Methods in Economics
Vienna University of Technology
A-1040 Vienna, Austria
Telephone: +43/588/011-0584
E-mail: owat.sunanta@tuwien.ac.at

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 47 *Submitted:* 2016-03-18
February 2018 *Accepted:* 2017-08-09

Austrian Journal of Statistics
February 2018, Volume 47, 45–76.
http://www.ajs.or.at/
doi:10.17713/ajs.v47i1.155

The Gamma-Weibull-G Family of Distributions with Applications

Broderick O. Oluyede
GSU

Shusen Pu
CWRU

Boikanyo Makubate
BIUST

Yuqi Qiu
UCSD

Abstract

Weibull distribution and its extended families has been widely studied in lifetime applications. Based on the Weibull-G family of distributions and the exponentiated Weibull distribution, we study in detail this new class of distributions, namely, Gamma-Weibull-G (GWG) family of distributions. Some special models in the new class are discussed. Statistical properties of the family of distributions, such as expansion of density function, hazard and reverse hazard functions, quantile function, moments, incomplete moments, generating functions, mean deviations, Bonferroni and Lorenz curves and order statistics are presented. We also present Rényi entropy, estimation of parameters by using method of maximum likelihood, asymptotic confidence intervals and applications using real data sets.

Keywords: gamma distribution, Weibull-G distribution, maximum likelihood estimation.

1. Introduction

Numerous classical distributions have been extensively discussed over the past years for modeling data in various areas such as economics, finance, insurance, actuarial, engineering, demography, biological studies, environmental and medical sciences. However, in many of those applied areas, there is more and more demands for extended forms of these distributions. Hence, several methods for generating new families of distributions have been studied in recent years. Researchers often define new families of probability distributions that extend well-known distributions in order to provide great flexibility in modeling data in practice. Recently, (Alzaatreh and Ghosh 2014) introduced and studied the Weibull-X family of distributions, (Alzaatreh, Famoye, and Lee 2014) developed the properties of the gamma-X family of distributions in general and studied the gamma-normal distribution as a special case, and (Nascimento, Bourguignon, Zea, Santos-Neto, Silva, and Cordeiro 2014) introduced the gamma extended Weibull family of distributions. Moreover, (Nadarajah, Cordeiro, and Ortega 2015) provided a comprehensive analysis of general mathematical and statistical properties of Zografos and Balakrishnan-G distributions. See references therein. Most recently, (Tahir, Zubair, Mansoor, Cordeiro, Alizadeh, and Hamedani 2015) introduced a new generator based on the Weibull random variable, namely, the new Weibull-G family.

We introduce a new extended generator, the gamma-Weibull-G family based on their previous work and combined the gamma-generator with the Weibull-G family of distributions which was defined by (Bourguignon, Silva, and Cordeiro 2014). We hope the new family of distributions yields a better fit in certain practical situations. Additionally, we provide a rigorous and comprehensive account of the mathematical properties of the proposed family of distributions.

The results in this paper are organized in the following manner. The new model called gamma Weibull-G (GWG) distribution and its sub-models are given in section 2. In section 3, statistical properties including expansion of the probability density function, hazard rate, reverse hazard rate and quantile functions, moments, conditional moments, moment generating and characteristics functions, mean deviations, Lorenz and Bonferroni curves, order statistics and Rényi entropy are presented. Maximum likelihood estimates of the model parameters and observed information matrix are given in section 4. The special cases of gamma-Weibull-Uniform (GWU) and gamma-Weibull-Weibull (GWW) distributions are presented in details in sections 5 and 6, respectively. A Monte Carlo simulation study to examine the bias and mean square error of the maximum likelihood estimates are presented in section 7. Section 8 contains applications of the new model to real data sets. A short concluding remark is given in section 9.

2. The model

Considering a continuous cumulative distribution function (cdf) $G(x)$ with probability density function (pdf) $g(x)$ and survival function $\bar{G}(x) = 1 - G(x)$, (Bourguignon *et al.* 2014) defined the Weibull-G family of distribution with cdf:

$$\begin{aligned} F_{WG}(x) &= \int_0^{\frac{G(x;\underline{\theta})}{1-G(x;\underline{\theta})}} \alpha\beta t^{\beta-1} e^{-\alpha t^{\beta}} dt \\ &= 1 - \exp\Big\{ -\alpha\Big[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\Big]^{\beta}\Big\}, \quad x, \alpha, \beta > 0, \end{aligned} \tag{1}$$

where $\underline{\theta}$ is a vector of parameters. The pdf of this family of distributions is given by

$$f_{WG}(x) = \alpha\beta g(x;\underline{\theta}) \frac{G(x;\underline{\theta})^{\beta-1}}{\bar{G}(x;\underline{\theta})^{\beta+1}} \exp\Big\{ -\alpha\Big[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\Big]^{\beta}\Big\}. \tag{2}$$

Therefore, given a continuous baseline distribution $G(x)$, one can derive the Weibull-G distribution with two extra parameters α and β from pdf (2). In the rest of this section, we consider the combination of Weibull-G (W-G) distribution with distribution proposed by (Zografos and Balakrishnan 2009) with cdf,

$$F(x;\delta) = \frac{1}{\Gamma(\delta)}\gamma\{\delta, -\log[1-G(x)]\}, \quad x, \delta > 0, \tag{3}$$

where $\gamma(\delta, x) = \int_0^x t^{\delta-1}e^{-t}dt$ is the lower incomplete gamma function and $\Gamma(\cdot)$ is the gamma function. The corresponding pdf is given by

$$f(x;\delta) = \frac{1}{\Gamma(\delta)}\{-\log[1-G(x)]\}^{\delta-1} g(x). \tag{4}$$

By taking $G(x) = F_{WG}(x)$ in (3), we come up with the gamma-Weibull-G family of distributions (GWG), with cdf:

$$F_{GWG}(x) = \frac{1}{\Gamma(\delta)}\gamma\Big\{\delta, \alpha\Big[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\Big]^{\beta}\Big\} \tag{5}$$

and pdf

$$f_{GWG}(x) = \frac{\beta\alpha^{\delta}}{\Gamma(\delta)} g(x;\underline{\theta}) \frac{G(x;\underline{\theta})^{\beta\delta-1}}{\bar{G}(x;\underline{\theta})^{\beta\delta+1}} \exp\Big\{ -\alpha\Big[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\Big]^{\beta}\Big\}. \tag{6}$$

Therefore, for each G distribution, we define the Gamma-Weibull-G (GWG) distribution with three extra parameters α, β and δ by the pdf (6). A random variable X with pdf (6) is denoted by $X \sim GWG(\alpha, \beta, \delta, \underline{\theta})$. The additional parameters induced by the Gamma-Weibull (G-W) generator induces a family of more flexible distributions. An interpretation of the GWG family of distributions can be given as follows (Bourguignon *et al.* 2014) in a similar context. Let Z be a lifetime random variable having a certain continuous G distribution. The odds ratio that an individual (or component) following the lifetime Z will die (failure) at time x is $G(x;\underline{\theta})/\bar{G}(x;\underline{\theta})$. Consider that the variability of this odds of death is represented by the random variable X and assume that it follows the Gamma-Weibull model with scale parameter α, and shape parameters δ and β. We can write

$$Pr(Z < x) = Pr\Big(X \leq \alpha\Big[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\Big]^{\beta}\Big) = F_{GWG}(x;\alpha,\beta,\delta,\underline{\theta}),$$

which is given by (5). To avoid issues or problems with over parametrization and redundancy, we restrict the parameter vector $\underline{\theta}$ to at most a two component vector.

2.1. Sub-models of the GWG distribution

If $\delta = 1$, the GWG distribution corresponds to the Weibull-generator, which was proposed by (Bourguignon *et al.* 2014). When $\beta = 1$, and $\delta = \beta = 1$, GWG distribution reduces to the gamma-exponential-generator, and exponential-generator (Cordeiro, Ortega, and Cunha 2013b), respectively. Moreover, $\beta = 2$, and $\beta = 2, \delta = 1$ lead to the gamma-Rayleigh-generator, and Rayleigh-generator, respectively. In addition, a lot of sub-models can be obtained when we change the parameter vector $\underline{\theta}$ and these special cases as well as the corresponding parameters are listed in Table 1.

Table 1: Distributions and corresponding $G(x;\underline{\theta})/\bar{G}(x;\underline{\theta})$ functions

Distribution	$G(x;\underline{\theta})/\bar{G}(x;\underline{\theta})$	$\Psi = (\alpha,\beta,\delta,\underline{\theta})$	GWG
Uniform	$x/(\theta - x)$	$(\alpha,\beta,\delta,\theta)$	G-W-Uniform
Exponential	$e^{\lambda x} - 1$	$(\alpha,\beta,\delta,\lambda)$	G-W-Exponential
Weibull	$e^{\lambda x^{\gamma}} - 1$	$(\alpha,\beta,\delta,\lambda,\gamma)$	G-W-Weibull
Fréchet	$(e^{\lambda x^{\gamma}} - 1)^{-1}$	$(\alpha,\beta,\delta,\lambda,\gamma)$	G-W-*Fréchet*
Half-logistic	$(e^{x} - 1)/2$	(α,β,δ)	G-W-Half-logistic
Power function	$[(\theta x)^{-k} - 1]^{-1}$	$(\alpha,\beta,\delta,\theta,k)$	G-W-Power function
Pareto	$(x/\theta)^{k} - 1$	$(\alpha,\beta,\delta,\theta,k)$	G-W-Pareto
Burr XII	$[1 - x^{c}]^{k} - 1$	$(\alpha,\beta,\delta,c,k)$	G-W-Burr XII
Log-logistic	$[1 - x^{c}] - 1$	(α,β,δ,c)	G-W-Log-logistic
Lomax	$[1 - x]^{k} - 1$	(α,β,δ,k)	G-W-Lomax
Gumbel	$\{\exp[\exp(-(x-\mu)/\sigma)] - 1\}^{-1}$	$(\alpha,\beta,\delta,\mu,\sigma)$	G-W-Gumbel
Normal	$\Phi((x-\mu)/\sigma)/[1 - \Phi((x-\mu)/\sigma)]$	$(\alpha,\beta,\delta,\mu,\sigma)$	G-W-Normal
Kumaraswamy	$(1 - x^{a})^{-b} - 1$	$(\alpha,\beta,\delta,a,b)$	G-W-Kumaraswamy
Modified Exponential	$e^{\theta x e^{\lambda x}} - 1$	$(\alpha,\beta,\delta,\theta,\lambda)$	G-W-Exponential

3. Statistical properties of the GWG distribution

Some statistical properties, such as expansion of density function, hazard and reverse hazard functions, quantile function, moments, incomplete moments, generating functions, mean deviations, Bonferroni and Lorenz curves and distribution of order statistics of the GWG distribution are presented in this section.

3.1. Expansion of the pdf of GWG distribution

Considering the power series for exponential function, we can obtain the following equation:

$$\exp\left\{-\alpha\Big[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\Big]^{\beta}\right\}=\sum_{i=0}^{\infty}\frac{(-1)^i\alpha^i}{i!}\Big[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\Big]^{i\beta}.$$

Inserting this equation into (6), we obtain

$$f_{GWG}(x) = \frac{\beta\alpha^{\delta}}{\Gamma(\delta)}g(x;\underline{\theta})\frac{G(x;\underline{\theta})^{\beta\delta-1}}{\bar{G}(x;\underline{\theta})^{\beta\delta+1}}\sum_{i=0}^{\infty}\frac{(-1)^i\alpha^i}{i!}\Big[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\Big]^{i\beta} \tag{7}$$

$$= \frac{\beta\alpha^{\delta}}{\Gamma(\delta)}g(x;\underline{\theta})\sum_{i=0}^{\infty}\frac{(-1)^i\alpha^i}{i!}G(x;\underline{\theta})^{\beta\delta+\beta i-1}\bar{G}(x;\underline{\theta})^{-[\beta\delta+\beta i+1]}. \tag{8}$$

Note that

$$\bar{G}(x;\underline{\theta})^{-[\beta\delta+\beta i+1]}=\sum_{j=0}^{\infty}\frac{\Gamma(\beta(\delta+i)+j+1)}{j!\Gamma(\beta(\delta+i)+1)}G(x;\underline{\theta})^j.$$

We obtain

$$f_{GWG}(x) = \frac{\beta\alpha^{\delta}g(x;\underline{\theta})}{\Gamma(\delta)}\sum_{i=0}^{\infty}\frac{(-1)^i\alpha^i}{i!}G(x;\underline{\theta})^{\beta\delta+\beta i-1}\sum_{j=0}^{\infty}\frac{\Gamma(\beta(\delta+i)+j+1)}{j!\Gamma(\beta(\delta+i)+1)}G(x;\underline{\theta})^j$$

$$= \sum_{i,j=0}^{\infty}\omega_{i,j}h_{\beta(\delta+i)+j}(x;\underline{\theta}), \tag{9}$$

where

$$\omega_{i,j}=\frac{(-1)^i\beta\alpha^{\delta+i}\Gamma[\beta(\delta+i)+j+1]}{i!j!\Gamma(\delta)\Gamma[\beta(\delta+i)+1][\beta(\delta+i)+j]}, \tag{10}$$

and

$$h_{\beta(\delta+i)+j}(x;\underline{\theta})=[\beta(\delta+i)+j]g(x;\underline{\theta})G(x;\underline{\theta})^{\beta(\delta+i)+j-1}$$

denote the pdf of the exponentiated generalized (EG) (Cordeiro *et al.* 2013b) distribution with parameter $\beta^*=\beta(\delta+i)+j$. The expansion of the pdf of GWG distribution illustrates that the GWG density function is indeed an infinite linear combination of EG density functions. Thus, a lot of the mathematical and statistical properties of the GWG distribution come directly from those of EG distribution.

3.2. Hazard, reverse hazard and quantile functions

In this section, we provide the hazard, reverse hazard and quantile functions of the GWG distribution.

Hazard and reverse hazard functions

Note that if X is a continuous random variable with cdf $G(x)$, and pdf $g(x)$, then the hazard rate function (hrf), reverse hazard function (rhf) and mean residual life functions are given by $h_G(x)=g(x)/\overline{G}(x)$, $\tau_G(x)=g(x)/G(x)$, and $\delta_G(x)=\int_x^{\infty}\overline{G}(u)du/\overline{G}(x)$ respectively. The functions $h_G(x)$, $\delta_G(x)$, and $\overline{G}(x)$ are equivalent. See (Shaked and Shanthikumar 1994) for additional details. The hazard rate and reverse hazard rate functions of GWG distribution are given by

$$h_{GWG}(x) = \frac{\beta\alpha^{\delta}g(x;\underline{\theta})\frac{G(x;\underline{\theta})^{\beta\delta-1}}{\overline{G}(x;\underline{\theta})^{\beta\delta+1}}\exp\left\{-\alpha\Big[\frac{G(x;\underline{\theta})}{\overline{G}(x;\underline{\theta})}\Big]^{\beta}\right\}}{\Gamma(\delta)-\gamma\Big\{\delta,\alpha\Big[\frac{G(x;\underline{\theta})}{\overline{G}(x;\underline{\theta})}\Big]^{\beta}\Big\}}, \tag{11}$$

and

$$\tau_{GWG}(x) \quad = \quad \frac{\beta\alpha^{\delta} g(x;\underline{\theta}) \frac{G(x;\underline{\theta})^{\beta\delta-1}}{\bar{G}(x;\underline{\theta})^{\beta\delta+1}} \exp\left\{ -\alpha \left[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\right]^{\beta} \right\}}{\gamma\left\{\delta, \alpha\left[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\right]^{\beta}\right\}}, \tag{12}$$

respectively.

Quantile function

The GWG quantile function can be obtained by inverting $F_{GWG}(x) = u$, where $F_{GWG}(x)$ is given by (5). Hence, we obtain the following equation:

$$\left[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\right]^{\beta} = \frac{\gamma^{-1}\Big(\delta, u\Gamma(\delta)\Big)}{\alpha}. \tag{13}$$

Thus, the quantile x_u of the GWG$(\alpha, \beta, \delta, \underline{\theta})$ reduces to the real solution of the following equation:

$$G(x;\underline{\theta}) = \frac{\left[\gamma^{-1}\big(\delta, u\Gamma(\delta)\big)\right]^{\frac{1}{\beta}}}{\left[\gamma^{-1}\big(\delta, u\Gamma(\delta)\big)\right]^{\frac{1}{\beta}} + \alpha^{\frac{1}{\beta}}} := q, \tag{14}$$

that is, the quantile x_u of the GWG$(\alpha, \beta, \delta, \underline{\theta})$ can be derived from the quantile x_q of the baseline distribution with cdf $G(x;\underline{\theta})$ and it is given by

$$x_q = G^{-1}(q), \tag{15}$$

where q is defined in equation (14).

3.3. Moments, incomplete moments, moment generating and characteristic functions

Let $X \sim GWG(\alpha, \beta, \delta, \underline{\theta})$, the s^{th} moment of X can be obtained from (9) as

$$E(X^s) = \sum_{i,j=0}^{\infty} \omega_{i,j} E(Z_{i,j}^s), \tag{16}$$

where $E(Z_{i,j}^s)$ denotes the s^{th} moment of $Z_{i,j}$, which follows the EG distribution with the parameter $\beta^* = \beta(\delta + i) + j$ and $\omega_{i,j}$ is given by equation (10). Similarly, the incomplete moments and the moment generating function (mgf) can be obtained as follows:

$$I_X(y) = \int_0^y x^s f_{GWG}(x)dx = \sum_{i,j=0}^{\infty} \omega_{i,j} I_{i,j}(y),$$

where $I_{i,j}(y) = \int_0^y x^s h_{\beta(\delta+i)+j}(x, \underline{\theta})dx$ and

$$M_X(t) = \sum_{i,j=0}^{\infty} \omega_{i,j} E(e^{tZ_{i,j}}),$$

where $E(e^{tZ_{i,j}})$ is the mgf of the EG distribution with parameter $\beta^* = \beta(\delta + i) + j$ and $\omega_{i,j}$ is given by (10). The characteristic function is given by $\phi(t) = E(e^{itX})$, where $i = \sqrt{-1}$, thus one can obtain

$$\phi(t) = \sum_{i,j=0}^{\infty} \omega_{i,j} \phi_{\beta(\delta+i)+j}(t),$$

with $\phi_{\beta(\delta+i)+j}(t)$ denotes the characteristic function of EG distribution, where parameter $\beta^* = \beta(\delta+i)+j$, and $\omega_{i,j}$ is defined as (10).

Since it is always of interest to know conditional expectations for lifetime models, we present it as follow:

$$E(X^t \mid X > x) = \frac{\Gamma(\delta)}{\Gamma(\delta) - \gamma\left\{\delta, \alpha\left[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\right]^{\beta}\right\}} \times \sum_{i,j=0}^{\infty} \omega_{i,j} I_{i,j}(y),$$

where $I_{i,j}(y) = \int_0^y x^t h_{\beta(\delta+i)+j}(x,\underline{\theta})dx$, and $\omega_{i,j}$ is given by (10).

3.4. Mean deviation, Bonferroni and Lorenz curves

Let $X \sim GWG(\alpha, \beta, \delta, \underline{\theta})$, the mean deviation about the mean and the mean deviation about the median are defined by $\delta_1(X) = \int_0^\infty |x-\mu| f_{GWG}(x)dx$, and $\delta_2(X) = \int_0^\infty |x-M| f_{GWG}(x)dx$, respectively, where $\mu = E(X)$ and $M = Median(X)$ denotes the median. They can be expressed as

$$\delta_1(X) = 2\mu F_{GWG}(\mu) - 2\int_0^{\mu} x f_{GWG}(x)dx \quad \text{and} \quad \delta_2(X) = \mu - 2\int_0^{M} x f_{GWG}(x)dx,$$

where $m(z) = \int_0^z x f_{GWG}(x)dx$ is the first incomplete moment. These quantities have been applied to a wide variety of fields, such as studying of income and property in economics, reliability, demography, insurance and medicine. Bonferroni and Lorenz curves are defined by

$$B(p) = \frac{1}{p\mu}\int_0^q x f(x)dx \quad and \quad L(p) = \frac{1}{\mu}\int_0^q x f(x)dx,$$

respectively, where $\mu = E(X)$ and $q = F_{GWG}^{-1}(p)$ is obtained from equation (14) and (15). Using similar methods in deriving the moments, we can show that

$$B(p) = \frac{1}{p\mu}\int_0^q x f(x)dx = \frac{1}{p\mu}\sum_{i,j=0}^{\infty} \omega_{i,j} I_{\beta(\delta+i)+j}(i,j),$$

and

$$L(p) = \frac{1}{\mu}\int_0^q x f(x)dx = \frac{1}{\mu}\sum_{i,j=0}^{\infty} \omega_{i,j} I_{\beta(\delta+i)+j}(i,j),$$

where $I_{\beta(\delta+i)+j}(i,j) = \int_0^q x f_{\beta(\delta+i)+j}(x,\underline{\theta})dx$, is the first incomplete moment of the EG distribution with parameter $\beta^* = \beta(\delta+i)+j$ and $\omega_{i,j}$ is given by (10).

3.5. Order statistics

In this section, distribution of the i^{th} order statistic for the GWG distribution is presented. Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d) random variables distributed according to (6). The pdf of the i^{th} order statistic, for example $X_{i:n}$, is given by

$$f_{i:n}(x) = \frac{n! f_{GWG}(x)}{(i-1)!(n-i)!}[F_{GWG}(x)]^{i-1}[1 - F_{GWG}(x)]^{n-i}.$$

Using the binomial theorem, the pdf of i^{th} order statistic can be written as

$$f_{i:n}(x) = \frac{n! f_{GWG}(x)}{(i-1)!(n-i)!}\sum_{k=0}^{n-i}\frac{(-1)^k}{[\Gamma(\delta)]^{i+k-1}}\binom{n-i}{k}\left\{\gamma\left(\delta, \alpha\left[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\right]^{\beta}\right)\right\}^{i+k-1}.$$

Note that for the incomplete gamma function $\gamma(x,\delta)$, we apply the following power series (see (Gradshteyn and Ryzhik 2000) for additional details)

$$\gamma(x,\delta) = \sum_{m=0}^{\infty}\frac{(-1)^m x^{m+\delta}}{(m+\delta)m!},$$

to obtain

$$\begin{aligned} f_{i:n}(x) &= \frac{n! f_{GWG}(x)}{(i-1)!(n-i)!} \sum_{k=0}^{n-i} \binom{n-i}{k} \frac{(-1)^k}{[\Gamma(\delta)]^{i+k-1}} \left\{ \alpha \left[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})} \right]^{\beta} \right\}^{\delta(i+k-1)} \\ &\times \left\{ \sum_{m=0}^{\infty} \frac{(-1)^m \left(\alpha \left[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})} \right]^{\beta} \right)^m}{(m+\delta) m!} \right\}^{k+i-1}. \end{aligned}$$

Let $c_m = \frac{(-1)^m}{m!(m+\delta)}$ and using the result on a power series raised to a positive integer, we can write

$$\left\{ \sum_{m=0}^{\infty} c_m \left(\alpha \left[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})} \right]^{\beta} \right)^m \right\}^{k+i-1} = \sum_{m=0}^{\infty} d_{m,k+i-1} \left(\alpha \left[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})} \right]^{\beta} \right)^m,$$

where $d_{0,k+i-1} = c_0^{(k+i-1)}$ and $d_{m,k+i-1} = (mc_0)^{-1} \sum_{l=1}^{m} [(k+i-1)l - m + l] c_l d_{m-l,k+i-1}$. Applying the above equation and replacing $f_{GWG}(x)$ by the right hand side of (6), we obtain

$$\begin{aligned} f_{i:n}(x) &= \frac{n! f_{GWG}(x)}{(i-1)!(n-i)!} \sum_{k=0}^{n-i} \sum_{m=0}^{\infty} \binom{n-i}{k} \frac{(-1)^k d_{m,k+i-1}}{[\Gamma(\delta)]^{i+k-1}} \left\{ \alpha \left[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})} \right]^{\beta} \right\}^{\delta(i+k-1)+m} \\ &= \frac{n!}{(i-1)!(n-i)!} \sum_{k=0}^{n-i} \sum_{m=0}^{\infty} \binom{n-i}{k} \frac{(-1)^k d_{m,k+i-1} \Gamma(\delta i + \delta k + m)}{[\Gamma(\delta)]^{i+k}} \\ &\times \quad f_{GWG}(x; \alpha, \beta, \delta_*). \end{aligned}$$

In other words, the pdf of the i^{th} order statistic is exactly a linear combination of GWG densities with parameters $(\alpha, \beta, \delta_*)$, where $\delta_* = \delta i + \delta k + m$. This is a very useful result since the properties of the order statistics of the GWG distribution can be obtained from those of EG distribution. For instance, we can obtain the moments of the i^{th} order statistic from GWG distribution as:

$$\begin{aligned} E(X_{i:n}^t) &= \sum_{l,j=0}^{\infty} \sum_{k=0}^{n-i} \sum_{m=0}^{\infty} \frac{n! \omega_{i,j}}{(i-1)!(n-i)!} \binom{n-i}{k} \frac{(-1)^k d_{m,k+i-1} \Gamma(\delta i + \delta k + m)}{[\Gamma(\delta)]^{i+k}} \\ &\times E\big(Z_{l,j}(x; \alpha, \beta, \delta_*)^t\big), \end{aligned}$$

where $E(Z_{l,j}(x; \alpha, \beta, \delta_*)^t)$ is the t^{th} moment of EG distribution with parameter $\beta^{**} = \beta(\delta_* + i) + j$, where $\delta_* = \delta i + \delta k + m$ and $\omega_{i,j}$ is given by (10).

3.6. Rényi entropy

In information theory, Rényi entropy generalizes Shannon entropy, Hartley entropy, min-entropy, and collision entropy. Entropies quantify the diversity, uncertainty, or randomness of a system. Rényi entropy is named after Alfréd Rényi (Renyi 1961) and is very important in ecology and statistics as indices of diversity. Rényi entropy is also important in quantum information, where it can be used as a measure of entanglement. As an extension of Shannon entropy, Rényi entropy is a popular measure of entropy, which is defined by

$$I_R(\nu) = (1-\nu)^{-1} \log \left[\int_{-\infty}^{\infty} g^{\nu}(x) dx \right],$$

where $\nu > 0$ and $\nu \neq 1$. We obtain Rényi entropy for the GWG distribution as follows:

$$\begin{aligned} I_R(\nu) &= (1-\nu)^{-1} \log \left[\int_{-\infty}^{\infty} g^{\nu}(x) dx \right] \\ &= (1-\nu)^{-1} \Big\{ \nu \log(\beta) + \delta\nu \log(\alpha) - \nu \log(\Gamma(\delta)) \\ &+ \log \left[\int_{-\infty}^{\infty} g^{\nu}(x;\underline{\theta}) \frac{G(x;\underline{\theta})^{(\beta\delta-1)\nu}}{\bar{G}(x;\underline{\theta})^{(\beta\delta+1)\nu}} \exp\left\{ -\alpha\nu \left[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})} \right]^{\beta} \right\} dx \right] \Big\}. \end{aligned} \tag{17}$$

Using similar expansion for the pdf, that is,

$$\exp\Big\{-\alpha\nu\Big[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\Big]^{\beta}\Big\} = \sum_{i=0}^{\infty}\frac{(-1)^i(\alpha\nu)^i}{i!}\Big[\frac{G(x;\underline{\theta})}{\bar{G}(x;\underline{\theta})}\Big]^{i\beta}, \tag{18}$$

we obtain

$$\begin{aligned} I_R(\nu) &= (1-\nu)^{-1}\Big\{\nu\log(\beta)+\delta\nu\log(\alpha)-\nu\log(\Gamma(\delta)) \\ &\quad +\log\Big[\sum_{i=0}^{\infty}\frac{(-1)^i(\alpha\nu)^i}{i!}\int_{-\infty}^{\infty}g^{\nu}(x;\underline{\theta})\frac{\big[G(x;\underline{\theta})\big]^{(\beta\delta-1)\nu+\beta i}}{\big[\bar{G}(x;\underline{\theta})\big]^{(\beta\delta+1)\nu+\beta i}}dx\Big]\Big\}. \end{aligned} \tag{19}$$

Note that

$$\frac{1}{\bar{G}(x;\underline{\theta})}=\frac{1}{1-G(x;\underline{\theta})}=\sum_{s=0}^{\infty}\big[G(x;\underline{\theta})\big]^{s} \quad and \quad \Big(\sum_{s=0}^{\infty}G(x;\underline{\theta})^{s}\Big)^{m}=\sum_{s=0}^{\infty}b_{s,m}\big[G(x;\underline{\theta})\big]^{s},$$

where

$$b_{s,m}=s^{-1}\sum_{l=1}^{s}\Big[m(l+1)-s\Big]b_{s-l,m}, \tag{20}$$

and $b_{0,m}=1, m=(\beta\delta+1)\nu+\beta i$. Now, we can write Rényi entropy as

$$\begin{aligned} I_R(\nu) &= (1-\nu)^{-1}\Big\{\nu\log(\beta)+\delta\nu\log(\alpha)-\nu\log(\Gamma(\delta))+\log\Big[\sum_{i=0}^{\infty}\sum_{s=0}^{\infty}\frac{(-1)^i(\alpha\nu)^i}{i!} \\ &\times\ b_{s,(\beta\delta+1)\nu+\beta i}\int_{-\infty}^{\infty}g^{\nu}(x;\underline{\theta})\big(G(x;\underline{\theta})\big)^{(\beta\delta-1)\nu+\beta i+s}dx\Big]\Big\} \\ &= (1-\nu)^{-1}\Big\{\nu\log(\beta)+\delta\nu\log(\alpha)-\nu\log(\Gamma(\delta)) \\ &+\ \log\Big[\sum_{i=0}^{\infty}\sum_{s=0}^{\infty}\frac{(-1)^i(\alpha\nu)^i}{i!}b_{s,(\beta\delta+1)\nu+\beta i}\frac{\nu^{\nu}}{[(\beta\delta-1)\nu+\beta i+s+\nu]^{\nu}} \\ &\times\int_{-\infty}^{\infty}\Big[\frac{(\beta\delta-1)\nu+\beta i+s+\nu}{\nu}g(x;\underline{\theta})\big(G(x;\underline{\theta})\big)^{\frac{(\beta\delta-1)\nu+\beta i+s}{\nu}}\Big]^{\nu}dx\Big]\Big\} \\ &= (1-\nu)^{-1}\Big\{\nu\log(\beta)+\delta\nu\log(\alpha)-\nu\log(\Gamma(\delta))+\log\Big[\sum_{i=0}^{\infty}\sum_{s=0}^{\infty}\frac{(-1)^i(\alpha\nu)^i}{i!} \\ &\times\ b_{s,(\beta\delta+1)\nu+\beta i}\frac{\nu^{\nu}}{[(\beta\delta-1)\nu+\beta i+s+\nu]^{\nu}}\times e^{(1-\nu)I_{REG}}\Big]\Big\}, \end{aligned} \tag{21}$$

where I_{REG} denotes the Rényi entropy for the exponentiated generalized distribution (Cordeiro *et al.* 2013b) with parameter $\beta^*=\frac{(\beta\delta-1)\nu+\beta i+s+\nu}{\nu}$ and $b_{s,m}=s^{-1}\sum_{l=1}^{s}\Big[m(l+1)-s\Big]b_{s-l,m}$ with $b_{0,m}=1, m=(\beta\delta+1)\nu+\beta i$. Thus, we can obtain Rényi entropy of the GWG distribution from Rényi entropy of the exponentiated generalized distribution by equation (21).

4. Estimation and observed information matrix

Estimation of the parameters of the GWG distributions can be obtained by the method of maximum likelihood. Let $X\sim GEMW(\alpha,\beta,\delta,\underline{\theta})$ and $\mathbf{\Delta}=(\alpha,\beta,\delta,\underline{\theta})^T$ be the parameter vector. The log-likelihood for $\mathbf{\Delta}$ can be written as

$$\begin{aligned} \ell=\ell(\mathbf{\Delta}) &= n\log(\beta)+n\delta\log(\alpha)-n\log[\Gamma(\delta)]-\alpha\sum_{i=1}^{n}\Big[\frac{G(x_i;\underline{\theta})}{1-G(x_i;\underline{\theta})}\Big]^{\beta}+\sum_{i=1}^{n}\log[g(x_i;\underline{\theta})] \\ &- (\beta\delta+1)\sum_{i=1}^{n}\log[1-G(x_i;\underline{\theta})]+(\beta\delta-1)\sum_{i=1}^{n}\log[G(x_i;\underline{\theta})]. \end{aligned} \tag{22}$$

The first derivative of the log-likelihood function with respect to the parameters $\mathbf{\Delta} = (\alpha, \beta, \delta, \underline{\theta})^T$ are given by

$$\frac{\partial \ell}{\partial \alpha} = \frac{n\delta}{\alpha} - \sum_{i=1}^{n} \left[\frac{G(x_i;\underline{\theta})}{1-G(x_i;\underline{\theta})}\right]^{\beta}, \tag{23}$$

$$\begin{aligned}\frac{\partial \ell}{\partial \beta} &= \frac{n}{\beta} - \alpha \sum_{i=1}^{n} \left[\frac{G(x_i;\underline{\theta})}{1-G(x_i;\underline{\theta})}\right]^{\beta} \log\left[\frac{G(x_i;\underline{\theta})}{1-G(x_i;\underline{\theta})}\right] \\ &\quad -\delta \sum_{i=1}^{n} \log[1-G(x_i;\underline{\theta})] + \delta \sum_{i=1}^{n} \log[G(x_i;\underline{\theta})],\end{aligned} \tag{24}$$

$$\frac{\partial \ell}{\partial \delta} = n\log(\alpha) - \frac{n\Gamma'(\delta)}{\Gamma(\delta)} - \beta \sum_{i=1}^{n} \log[1-G(x_i;\underline{\theta})] + \beta \sum_{i=1}^{n} \log[G(x_i;\underline{\theta})], \tag{25}$$

and

$$\begin{aligned}\frac{\partial \ell}{\partial \underline{\theta}_k} &= \sum_{i=1}^{n} \frac{1}{g(x_i;\underline{\theta})} \frac{\partial g(x_i;\underline{\theta})}{\partial \underline{\theta}_k} + (\beta\delta+1) \sum_{i=1}^{n} \frac{1}{1-G(x_i;\underline{\theta})} \frac{\partial G(x_i;\underline{\theta})}{\partial \underline{\theta}_k} \\ &\quad -\alpha\beta \sum_{i=1}^{n} \frac{G(x_i;\underline{\theta})^{\beta-1}}{\left[1-G(x_i;\underline{\theta})\right]^{\beta+1}} \frac{\partial G(x_i;\underline{\theta})}{\partial \underline{\theta}_k} \\ &\quad +(\beta\delta-1) \sum_{i=1}^{n} \frac{1}{G(x_i;\underline{\theta})} \frac{\partial G(x_i;\underline{\theta})}{\partial \underline{\theta}_k},\end{aligned} \tag{26}$$

where $\underline{\theta}_k$ is the k^{th} element of the vector of parameters $\underline{\theta}$.
The partitioned observed information matrix $\hat{\mathbf{\Delta}} = (\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\underline{\theta}})$ for the GWG distribution is

$$\mathbf{J}(\hat{\mathbf{\Delta}}) = -\begin{bmatrix} \frac{\partial^2 \ell}{\partial \alpha^2} & \frac{\partial^2 \ell}{\partial \alpha \partial \beta} & \frac{\partial^2 \ell}{\partial \alpha \partial \delta} & \frac{\partial^2 \ell}{\partial \alpha \partial \underline{\theta}} \\ \frac{\partial^2 \ell}{\partial \beta \partial \alpha} & \frac{\partial^2 \ell}{\partial \beta^2} & \frac{\partial^2 \ell}{\partial \beta \partial \delta} & \frac{\partial^2 \ell}{\partial \beta \partial \underline{\theta}} \\ \frac{\partial^2 \ell}{\partial \delta \partial \alpha} & \frac{\partial^2 \ell}{\partial \delta \partial \beta} & \frac{\partial^2 \ell}{\partial \delta^2} & \frac{\partial^2 \ell}{\partial \delta \partial \underline{\theta}} \\ \frac{\partial^2 \ell}{\partial \underline{\theta} \partial \alpha} & \frac{\partial^2 \ell}{\partial \underline{\theta} \partial \beta} & \frac{\partial^2 \ell}{\partial \underline{\theta} \partial \delta} & \frac{\partial^2 \ell}{\partial \underline{\theta}^2} \end{bmatrix},$$

which is symmetric matrix. We provide these elements as follows:

$$\frac{\partial^2 \ell}{\partial \alpha^2} = \frac{-n\delta}{\alpha^2}, \quad \frac{\partial^2 \ell}{\partial \alpha \partial \delta} = \frac{n}{\alpha},$$

$$\frac{\partial^2 \ell}{\partial \alpha \partial \beta} = -\sum_{i=1}^{n} \left[\frac{G(x_i;\underline{\theta})}{1-G(x_i;\underline{\theta})}\right]^{\beta} \log\left[\frac{G(x_i;\underline{\theta})}{1-G(x_i;\underline{\theta})}\right],$$

$$\frac{\partial^2 \ell}{\partial \alpha \partial \underline{\theta}_k} = -\beta \sum_{i=1}^{n} \frac{G(x_i;\underline{\theta})^{\beta-1}}{\left[1-G(x_i;\underline{\theta})\right]^{\beta+1}} \frac{\partial G(x_i;\underline{\theta})}{\partial \underline{\theta}_k},$$

$$\frac{\partial^2 \ell}{\partial \beta^2} = \frac{n}{\beta} - \alpha \sum_{i=1}^{n} \left[\frac{G(x_i;\underline{\theta})}{1-G(x_i;\underline{\theta})}\right]^{\beta} \left(\log\left[\frac{G(x_i;\underline{\theta})}{1-G(x_i;\underline{\theta})}\right]\right)^2,$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \delta} = -\sum_{i=1}^{n} \log[1-G(x_i;\underline{\theta})] + \sum_{i=1}^{n} \log[G(x_i;\underline{\theta})],$$

$$\begin{aligned}\frac{\partial^2\ell}{\partial\beta\partial\underline{\theta}_k} &= \delta\sum_{i=1}^{n}\frac{1}{1-G(x_i;\underline{\theta})}\frac{\partial G(x_i;\underline{\theta}_k)}{\partial\underline{\theta}}+\delta\sum_{i=1}^{n}\frac{1}{G(x_i;\underline{\theta})}\frac{\partial G(x_i;\underline{\theta})}{\partial\underline{\theta}_k}\\ &-\alpha\sum_{i=1}^{n}\frac{G(x_i;\underline{\theta})^{\beta-1}}{\big[1-G(x_i;\underline{\theta})\big]^{\beta+1}}\left(1+\beta\log\Big[\frac{G(x_i;\underline{\theta})}{1-G(x_i;\underline{\theta})}\Big]\right)\frac{\partial G(x_i;\underline{\theta})}{\partial\underline{\theta}_k},\end{aligned}$$

$$\frac{\partial^2\ell}{\partial\delta^2} = \frac{n\Big(\big[\Gamma'(\delta)\big]^2-\Gamma(\delta)\Gamma''(\delta)\Big)}{\Gamma^2(\delta)},$$

$$\frac{\partial^2\ell}{\partial\delta\partial\underline{\theta}_k} = \beta\sum_{i=1}^{n}\frac{1}{1-G(x_i;\underline{\theta})}\frac{\partial G(x_i;\underline{\theta})}{\partial\underline{\theta}_k}+\beta\sum_{i=1}^{n}\frac{1}{G(x_i;\underline{\theta})}\frac{\partial G(x_i;\underline{\theta})}{\partial\underline{\theta}_k},$$

and

$$\begin{aligned}\frac{\partial^2\ell}{\partial\underline{\theta}_k\underline{\theta}_j} &= \sum_{i=1}^{n}\left(\frac{-1}{g^2(x_i;\underline{\theta})}\frac{\partial g(x_i;\underline{\theta})}{\partial\underline{\theta}_k}\frac{\partial g(x_i;\underline{\theta})}{\partial\underline{\theta}_j}+\frac{1}{g(x_i;\underline{\theta})}\frac{\partial^2 g(x_i;\underline{\theta})}{\partial\underline{\theta}_k\underline{\theta}_j}\right)+(\beta\delta+1)\\ &\times\sum_{i=1}^{n}\left(\frac{1}{[1-G(x_i;\underline{\theta})]^2}\frac{\partial G(x_i;\underline{\theta})}{\partial\underline{\theta}_k}\frac{\partial G(x_i;\underline{\theta})}{\partial\underline{\theta}_j}+\frac{1}{1-G(x_i;\underline{\theta})}\frac{\partial^2 G(x_i;\underline{\theta})}{\partial\underline{\theta}_k\underline{\theta}_j}\right)\\ &-\alpha\beta\sum_{i=1}^{n}\left(\ell_1\frac{\partial G(x_i;\underline{\theta})}{\partial\underline{\theta}_k}\frac{\partial G(x_i;\underline{\theta})}{\partial\underline{\theta}_j}+\frac{G(x_i;\underline{\theta})^{\beta-1}}{\big[1-G(x_i;\underline{\theta})\big]^{\beta+1}}\frac{\partial^2 G(x_i;\underline{\theta})}{\partial\underline{\theta}_k\underline{\theta}_j}\right)\\ &+(\beta\delta-1)\sum_{i=1}^{n}\left(\frac{1}{G^2(x_i;\underline{\theta})}\frac{\partial G(x_i;\underline{\theta})}{\partial\underline{\theta}_k}\frac{\partial G(x_i;\underline{\theta})}{\partial\underline{\theta}_j}+\frac{1}{G(x_i;\underline{\theta})}\frac{\partial^2 G(x_i;\underline{\theta})}{\partial\underline{\theta}_k\underline{\theta}_j}\right),\end{aligned}$$

where

$$\ell_1=\frac{(\beta-1)G(x_i;\underline{\theta})^{\beta-2}\big[1-G(x_i;\underline{\theta})\big]^{\beta+1}+(\beta+1)G(x_i;\underline{\theta})^{\beta-1}\big[1-G(x_i;\underline{\theta})\big]^{\beta}}{\big[1-G(x_i;\underline{\theta})\big]^{2(\beta+1)}},$$

and $\underline{\theta}_k$ is the k^{th} element of the vector of parameters $\underline{\theta}$.

The asymptotic confidence intervals for the parameters of the GWG distribution are as follows. The expectations in the Fisher Information Matrix (FIM) can be obtained numerically. Let $\hat{\Delta}=(\hat{\alpha},\hat{\beta},\hat{\delta},\hat{\underline{\theta}})$ be the maximum likelihood estimate of $\mathbf{\Delta}=(\alpha,\beta,\delta,\underline{\theta})$. Under the usual regularity conditions and that the parameters are in the interior of the parameter space, but not on the boundary, we have: $\sqrt{n}(\hat{\mathbf{\Delta}}-\mathbf{\Delta})\xrightarrow{d}N_p(\underline{0},I^{-1}(\mathbf{\Delta}))$, where $I(\mathbf{\Delta})$ is the expected Fisher information matrix. The asymptotic behavior is still valid if $I(\mathbf{\Delta})$ is replaced by the observed information matrix evaluated at $\hat{\mathbf{\Delta}}$, that is $J(\hat{\mathbf{\Delta}})$. The multivariate normal distribution $N_p(\underline{0},J(\hat{\mathbf{\Delta}})^{-1})$, where the mean vector $\underline{\mathbf{0}}=(0,0,0,\underline{0})^T$, can be used to construct confidence intervals and confidence regions for the individual model parameters and for the survival and hazard rate functions. That is, the approximate $100(1-\eta)\%$ two-sided confidence intervals for α, β, δ, $\underline{\theta}_k$ are given by:

$$\widehat{\alpha}\pm Z_{\frac{\eta}{2}}\sqrt{\mathbf{I}_{\alpha\alpha}^{-1}(\widehat{\mathbf{\Delta}})},\quad \widehat{\beta}\pm Z_{\frac{\eta}{2}}\sqrt{\mathbf{I}_{\beta\beta}^{-1}(\widehat{\mathbf{\Delta}})},\quad \widehat{\delta}\pm Z_{\frac{\eta}{2}}\sqrt{\mathbf{I}_{\delta\delta}^{-1}(\widehat{\mathbf{\Delta}})},\quad \textit{and}\quad \widehat{\underline{\theta}_k}\pm Z_{\frac{\eta}{2}}\sqrt{\mathbf{I}_{\underline{\theta}_k\underline{\theta}_k}^{-1}(\widehat{\mathbf{\Delta}})},$$

respectively, where $\mathbf{I}_{\alpha\alpha}^{-1}(\widehat{\mathbf{\Delta}}),\mathbf{I}_{\beta\beta}^{-1}(\widehat{\mathbf{\Delta}}),\mathbf{I}_{\delta\delta}^{-1}(\widehat{\mathbf{\Delta}})$, and $\mathbf{I}_{\underline{\theta}_k\underline{\theta}_k}^{-1}(\widehat{\mathbf{\Delta}})$, are the diagonal elements of $\mathbf{I}_n^{-1}(\widehat{\mathbf{\Delta}})=(n\mathbf{I}(\widehat{\mathbf{\Delta}}))^{-1}$, and $Z_{\frac{\eta}{2}}$ is the upper $\frac{\eta}{2}^{th}$ percentile of a standard normal distribution.

5. The gamma-Weibull-uniform distribution

Suppose that the baseline distribution is a uniform distribution on the interval $(0,\theta)$, with $\theta>0$. Then, $g(x;\theta)=1/\theta$ and $G(x;\theta)=x/\theta$, where $0<x<\theta$. The cdf and pdf of the gamma-Weibull-uniform (GWU) distribution are given by

$$F_{GWU}(x) = \frac{1}{\Gamma(\delta)}\gamma\Big[\delta, \alpha\Big(\frac{x}{\theta - x}\Big)^{\beta}\Big], \tag{27}$$

and

$$f_{GWU}(x) = \frac{\beta\theta\alpha^{\delta}}{\Gamma(\delta)}\frac{x^{\beta\delta-1}}{(\theta - x)^{\beta\delta+1}}\exp\Big[-\alpha\Big(\frac{x}{\theta - x}\Big)^{\beta}\Big], \tag{28}$$

respectively, where $0 < x < \theta, \alpha, \beta, \delta > 0$.

5.1. Some sub-models of the GWU distribution

When $\alpha = \beta = \delta = 1$, the GWU distribution reduces to the exponential uniform (EU(θ), EU1) distribution with pdf as follows:

$$f(x) = \frac{\theta}{(\theta - x)^2}\exp\Big(-\frac{x}{\theta - x}\Big).$$

By letting $\alpha = \delta = 1$, we obtain the Weibull uniform (WU(θ, β)) distribution as a special case of the GWU distribution, whose pdf is given by

$$f(x) = \frac{\beta\theta x^{\beta-1}}{(\theta - x)^{\beta+1}}\exp\Big[-\Big(\frac{x}{\theta - x}\Big)^{\beta}\Big].$$

We can also obtain the gamma exponential uniform (GEU(θ, δ), GEU1) distribution by letting $\alpha = \beta = 1$ with pdf given by

$$f(x) = \frac{\theta x^{\delta-1}}{\Gamma(\delta)(\theta - x)^{\delta+1}}\exp\Big(-\frac{x}{\theta - x}\Big).$$

When $\beta = \delta = 1$, the GWU distribution reduces to the exponential uniform (EU(θ, α), EU2) distribution with pdf given by

$$f(x) = \frac{\theta\alpha}{(\theta - x)^2}\exp\Big[-\alpha\Big(\frac{x}{\theta - x}\Big)\Big].$$

By setting $\delta = 1$, we obtain the Weibull uniform distribution, which is also known as the Phani distribution (Phani 1987). The pdf of Phani distribution is given by

$$f(x) = \frac{\beta\theta\alpha x^{\beta-1}}{(\theta - x)^{\beta+1}}\exp\Big[-\alpha\Big(\frac{x}{\theta - x}\Big)^{\beta}\Big].$$

Moreover, when $\alpha = 1$, we can obtain the gamma Weibull uniform (GWU(θ, β, δ), GWU1) distribution as a special case of the GWU distribution, whose pdf is given by

$$f(x) = \frac{\beta\theta}{\Gamma(\delta)}\frac{x^{\beta\delta-1}}{(\theta - x)^{\beta\delta+1}}\exp\Big[-\Big(\frac{x}{\theta - x}\Big)^{\beta}\Big].$$

We can obtain the gamma exponential uniform (GWU(θ, α, δ), GU (Torabi and Montazeri 2012)) distribution by setting $\beta = 1$ and the corresponding pdf is given by

$$f(x) = \frac{\theta\alpha^{\delta}}{\Gamma(\delta)}\frac{x^{\delta-1}}{(\theta - x)^{\delta+1}}\exp\Big[-\alpha\Big(\frac{x}{\theta - x}\Big)\Big].$$

When $\beta = 2$ and $\beta = 2, \delta = 1$, we can obtain gamma Rayleigh uniform (GRU) distribution and Rayleigh uniform (RU) distribution, and the corresponding pdf are given by

$$f(x) = \frac{2\theta\alpha^{\delta}}{\Gamma(\delta)}\frac{x^{2\delta-1}}{(\theta - x)^{2\delta+1}}\exp\Big[-\alpha\Big(\frac{x}{\theta - x}\Big)^{2}\Big],$$

and

$$f(x) \quad = \quad \frac{2\theta\alpha x}{(\theta-x)^3}\exp\Big[-\alpha\Big(\frac{x}{\theta-x}\Big)^2\Big],$$

respectively.

5.2. Expansion of the GWU density

From equation (9), we can obtain

$$f_{GWG}(x) \quad = \quad \sum_{i,j=0}^{\infty}\omega_{i,j}h_{\beta(\delta+i)+j}(x;\theta), \tag{29}$$

where

$$\omega_{i,j} = \frac{(-1)^i\beta\alpha^{\delta+i}\Gamma[\beta(\delta+i)+j+1]}{i!j!\Gamma(\delta)\Gamma[\beta(\delta+i)+1][\beta(\delta+i)+j]}, \tag{30}$$

and

$$h_{\beta(\delta+i)+j}(x;\theta) = \big(\beta(\delta+i)+j\big)\frac{1}{\theta}\Big(\frac{x}{\theta}\Big)^{\beta(\delta+i)+j-1}$$

denotes pdf of the exponentiated uniform (EU) distribution with power $\beta(\delta+i)+j$. Thus, we can obtain statistical properties of the GWU distribution from those of EU distribution.

5.3. Hazard, reverse hazard and quantile functions

The hazard and reverse hazard functions of the GWU distribution are given by

$$h_{GWU}(x) \quad = \quad \frac{\beta\theta\alpha^\delta\frac{x^{\beta\delta-1}}{(\theta-x)^{\beta\delta+1}}\exp\big[-\alpha\big(\frac{x}{\theta-x}\big)^\beta\big]}{\Gamma(\delta)-\gamma\big[\delta,\alpha\big(\frac{x}{\theta-x}\big)^\beta\big]}, \tag{31}$$

and

$$\tau_{GWU}(x) \quad = \quad \frac{\beta\theta\alpha^\delta\frac{x^{\beta\delta-1}}{(\theta-x)^{\beta\delta+1}}\exp\big[-\alpha\big(\frac{x}{\theta-x}\big)^\beta\big]}{\gamma\big[\delta,\alpha\big(\frac{x}{\theta-x}\big)^\beta\big]}, \tag{32}$$

respectively. In addition, we can also obtain the quantile function of GWU distribution from equation (14) and (15) as

$$x_u = q\times\theta = \frac{\theta\Big[\gamma^{-1}\big(\delta,u\Gamma(\delta)\big)\Big]^{\frac{1}{\beta}}}{\Big[\gamma^{-1}\big(\delta,u\Gamma(\delta)\big)\Big]^{\frac{1}{\beta}}+\alpha^{\frac{1}{\beta}}}. \tag{33}$$

Several plots of pdf and hrf of the GWU distribution for selected parameter values are given by Figure 1 and 2. The graphs of the pdf take various types, including unimodal, decreasing, left and right skewed. As we can see from Figure 2, the plot for hrf of the GWU distribution contains decreasing, increasing, bathtub, unimodal and upside down bathtub shapes for the selected values of the GWU parameters.

5.4. Moments, incomplete moments, moment generating and characteristic functions

Let $X\sim GWU(\alpha,\beta,\delta,\theta), 0<x<\theta,\alpha,\beta,\delta>0$, the s^{th} moment of X can be obtained from (29) as

$$E(X^s) = \sum_{i,j=0}^{\infty}\omega_{i,j}E(Z_{i,j}^s),$$

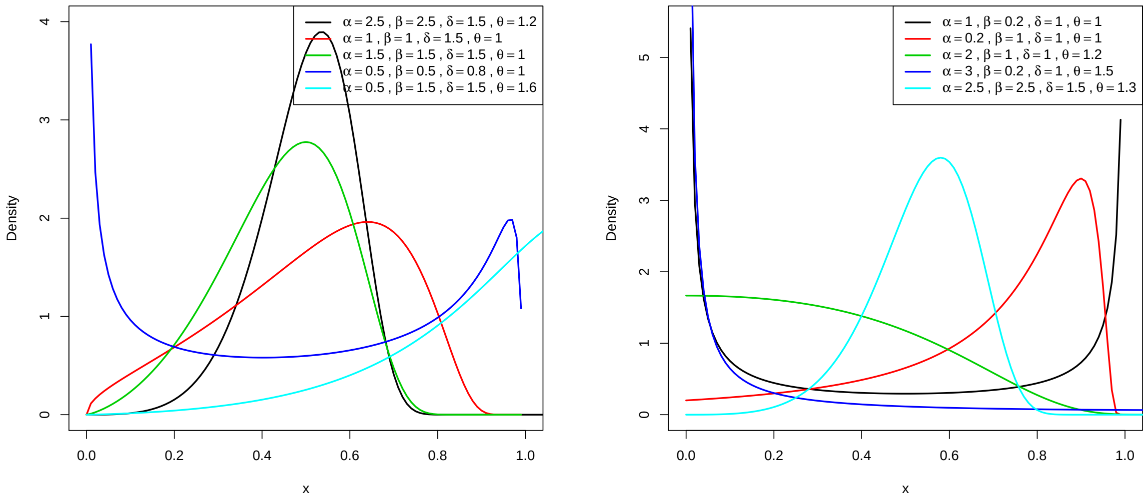

Figure 1: Graphs of GWU pdf with selected parameters

where $E(Z^s_{i,j})$ denotes the s^{th} moment of $Z_{i,j}$, which follows the EU distribution with power $\beta(\delta+i)+j$ and $\omega_{i,j}$ is defined as (30). Similarly, the incomplete moments and mgf are given by

$$I_X(y) = \int_0^y x^s f_{GWG}(x)dx = \sum_{i,j=0}^{\infty} \omega_{i,j} I_{i,j}(y),$$

where $I_{i,j}(y) = \int_0^y x^s h_{\beta(\delta+i)+j}(x,\theta)dx$ and

$$M_X(t) = \sum_{i,j=0}^{\infty} \omega_{i,j} E(e^{tZ_{i,j}}),$$

where $E(e^{tZ_{i,j}})$ is the mgf of the EU distribution with power $\beta^* = \beta(\delta+i)+j$ and $\omega_{i,j}$ is defined as (30). The characteristic function is given by $\phi(t) = E(e^{itX})$, where $i = \sqrt{-1}$. Thus, we have

$$\phi(t) = \sum_{i,j=0}^{\infty} \omega_{i,j} \phi_{\beta(\delta+i)+j}(t),$$

where $\phi_{\beta(\delta+i)+j}(t)$ denotes the characteristic function of the EU distribution with power $\beta(\delta+i)+j$, and $\omega_{i,j}$ is defined as (30).

The first six moments, standard deviation (SD), coefficient of variation (CV), coefficient of skewness (CS), and coefficient of kurtosis (CK) for selected values of the parameters of the *GWU* distribution are listed in Tables 2 below. The variance (σ^2), CV, CS and CK are given by

$$\sigma^2 = \mu_2' - \mu^2, \quad CV = \frac{\sigma}{\mu} = \frac{\sqrt{\mu_2' - \mu^2}}{\mu} = \sqrt{\frac{\mu_2'}{\mu^2} - 1},$$

$$CS = \frac{E\left[(X-\mu)^3\right]}{\left[E(X-\mu)^2\right]^{3/2}} = \frac{\mu_3' - 3\mu\mu_2' + 2\mu^3}{(\mu_2' - \mu^2)^{3/2}},$$

and

$$CK = \frac{E\left[(X-\mu)^4\right]}{\left[E(X-\mu)^2\right]^2} = \frac{\mu_4' - 4\mu\mu_3' + 6\mu^2\mu_2' - 3\mu^4}{(\mu_2' - \mu^2)^2},$$

respectively. The graphs of CS and CK versus δ show the dependence of skewness and kurtosis measures on the shape parameter.

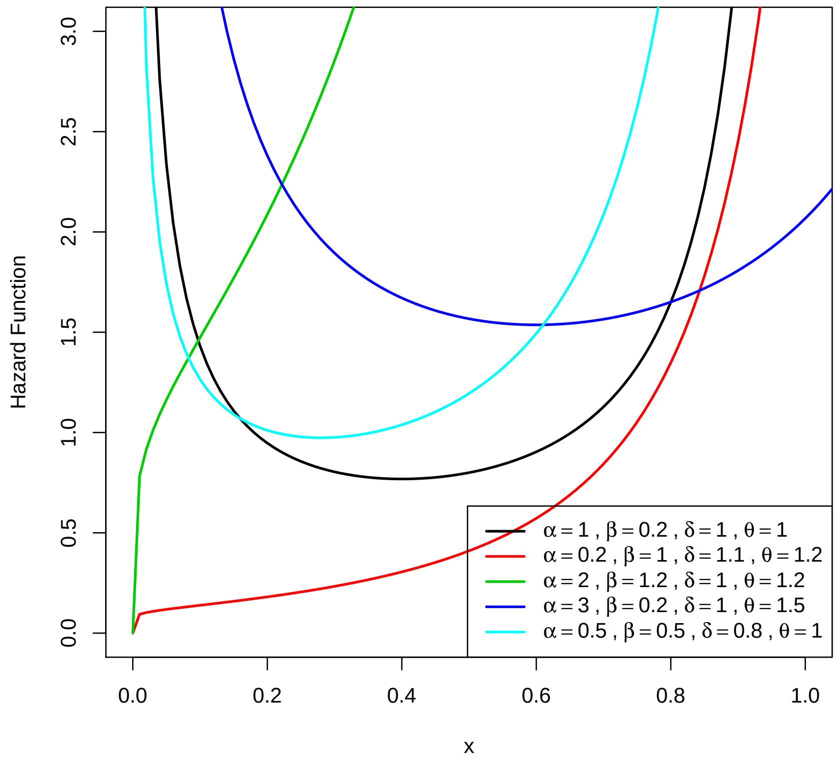

Figure 2: Graphs of GWU hrf with selected parameters

Table 2: GWU moments for selected values

	$(\alpha,\beta,\delta,\theta)$			
$Moments$	(5.0,3.0,2.0,1.0)	(1.0,1.0,1.0,1.0)	(0.5,1.5,1.5,1.6)	(1.0,1.0,1.5,1.5)
$E(X)$	0.4042	0.4037	0.9802	0.7736
$E(X^2)$	0.1673	0.2110	1.0133	0.6865
$E(X^3)$	0.0706	0.1237	1.0852	0.6560
$E(X^4)$	0.0303	0.0778	1.1930	0.6578
$E(X^5)$	0.0132	0.0512	1.3386	0.6831
$E(X^6)$	0.0058	0.0350	1.5273	0.7286
SD	0.0620	0.2191	0.2290	0.2967
CV	0.1534	0.5429	0.2336	0.3835
CS	-0.5254	-0.0171	-0.9034	-0.4321
CK	3.3243	1.9815	3.6357	2.3932

Moreover, the conditional expectations of the GWU distribution can also be obtained from the EU (Cordeiro *et al.* 2013b) distribution with power $\beta(\delta+i)+j$, and we present it as follow:

$$E(X^t \mid X > x) = \frac{\Gamma(\delta)}{\Gamma(\delta) - \gamma\left\{\delta, \alpha\left(\frac{x}{\theta-x}\right)^{\beta}\right\}} \times \sum_{i,j=0}^{\infty} \omega_{i,j} I_{i,j}(y),$$

where $I_{i,j}(y) = \int_0^y x^s h_{\beta(\delta+i)+j}(x,\theta)dx$, and $\omega_{i,j}$ is defined as (30).

5.5. Mean deviation, Bonferroni and Lorenz curves

Let $X \sim GWU(\alpha,\beta,\delta,\theta), 0 < x < \theta, \alpha,\beta,\delta > 0$, $\mu = E(X)$ be the mean and $M =$

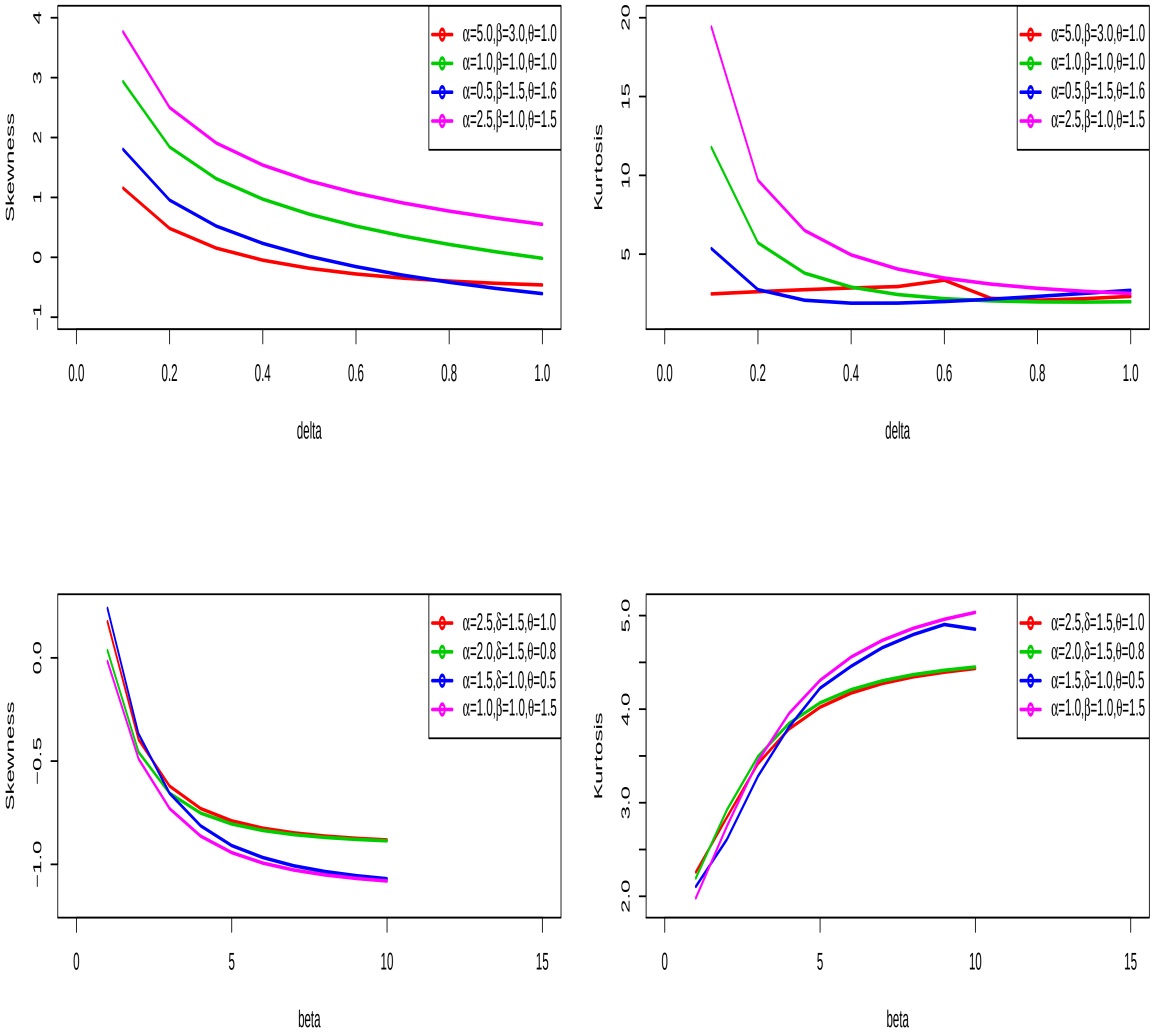

Figure 3: Graphs of GWU skewness and kurtosis with selected parameters

$Median(X)$ the median. Then the mean deviation about the mean and the mean deviation about the median can be obtained from

$$\delta_1(X) = 2\mu F_{GWU}(\mu) - 2\int_0^{\mu} x f_{GWU}(x)dx, \quad \text{and} \quad \delta_2(X) = \mu - 2\int_0^{M} x f_{GWU}(x)dx,$$

respectively, where $m(z) = \int_0^z x f_{GWU}(x)dx$ is the first incomplete moment. Bonferroni and Lorenz curves are given by

$$B(p) = \frac{1}{p\mu}\int_0^q x f(x)dx = \frac{1}{p\mu}\sum_{i,j=0}^{\infty} \omega_{i,j} I_{\beta(\delta+i)+j}(i,j),$$

and

$$L(p) = \frac{1}{\mu}\int_0^q x f(x)dx = \frac{1}{\mu}\sum_{i,j=0}^{\infty} \omega_{i,j} I_{\beta(\delta+i)+j}(i,j),$$

respectively, where $I_{\beta(\delta+i)+j}(i,j) = \int_0^q x f_{\beta(\delta+i)+j}(x,\theta)dx$, is the first incomplete moment of the EU distribution with power $\beta(\delta+i)+j$ and $\omega_{i,j}$ is defined as (30).

5.6. Order statistics

The distribution of the i^{th} order statistic for the GWU distribution is presented in this section. Let $X_1, \ldots, X_n$ be i.i.d random variables distributed according to (29). The pdf of the i^{th} order statistic, for example $X_{i:n}$, is given by

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!}\sum_{q=0}^{n-i}\sum_{m=0}^{\infty}\binom{n-i}{q}\frac{(-1)^q d_{m,q+i-1}\Gamma(\delta i+\delta q+m)}{[\Gamma(\delta)]^{i+k}} f_{GWU}(x;\alpha,\beta,\delta_*),$$

that is, the pdf of the i^{th} order statistic is a linear combination of GWU densities with parameters $(\alpha, \beta, \delta_*, \theta)$, where $\delta_* = \delta i + \delta q + m$. This is a very useful result since properties of the order statistics of the GWU distribution can be obtained from those of EU distribution. For instance, we can obtain

$$\begin{aligned} E(X_{i:n}^t) &= \sum_{l,j=0}^{\infty}\sum_{q=0}^{n-i}\sum_{m=0}^{\infty}\frac{n!\omega_{i,j}}{(i-1)!(n-i)!}\binom{n-i}{q}\frac{(-1)^q d_{m,q+i-1}\Gamma(\delta i+\delta q+m)}{[\Gamma(\delta)]^{i+q}} \\ &\quad E(Z_{l,j}(x;\alpha,\beta,\delta_*)^t), \end{aligned}$$

where $E(Z_{l,j}(x;\alpha,\beta,\delta_*)^t)$ is the t^{th} moment of EU distribution with power $\beta(\delta_*+i)+j$, where $\delta_* = \delta i + \delta q + m$ and $\omega_{i,j}$ is given by equation (30).

5.7. Rényi entropy

By equation (21), Rényi entropy for the GWU distribution is given by

$$\begin{aligned} I_R(\nu) &= (1-\nu)^{-1}\Big\{\nu\log(\beta) + \delta\nu\log(\alpha) - \nu\log(\Gamma(\delta)) + \log\Big[\sum_{i=0}^{\infty}\sum_{s=0}^{\infty}\frac{(-1)^i(\alpha\nu)^i}{i!} \\ &\times\ b_{s,(\beta\delta+1)\nu+\beta i}\frac{\nu^{\nu}}{[(\beta\delta-1)\nu+\beta i+s+\nu]^{\nu}} \times e^{(1-\nu)I_{REU}}\Big]\Big\}, \end{aligned} \tag{34}$$

where I_{REU} denotes the Rényi entropy for the exponentiated uniform (EU) distribution with parameter $\beta^* = \frac{(\beta\delta-1)\nu+\beta i+s+\nu}{\nu}$ and $b_{s,m} = s^{-1}\sum_{l=1}^{s}\Big[m(l+1)-s\Big]b_{s-l,m}$ with $b_{0,m} = 1, m = (\beta\delta+1)\nu+\beta i$.

6. The gamma-Weibull-Weibull distribution

Suppose that the baseline distribution is the Weibull distribution with parameters λ and k. Then, $g(x;\lambda,k) = \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}e^{-(\frac{x}{\lambda})^k}$ and $G(x;\lambda,k) = 1 - e^{-(\frac{x}{\lambda})^k}$, where $x \geq 0$. The cdf and pdf of gamma-Weibull-Weibull (GWW) distribution are given by

$$F_{GWW}(x) = \frac{1}{\Gamma(\delta)}\gamma\Big[\delta, \alpha\big(e^{(\frac{x}{\lambda})^k} - 1\big)^{\beta}\Big],$$

and

$$f_{GWW}(x) \quad = \quad \frac{\beta\alpha^{\delta}}{\Gamma(\delta)} \times \frac{kx^{k-1}e^{(\frac{x}{\lambda})^k}}{\lambda^k}\big[e^{(\frac{x}{\lambda})^k} - 1\big]^{\beta\delta-1}\exp\Big[-\alpha\big(e^{(\frac{x}{\lambda})^k} - 1\big)^{\beta}\Big],$$

respectively, where $x \geq 0, \alpha, \beta, \delta, \lambda, k > 0$.

Graphs of the GWW density functions with several combinations of parameter values are given by Figure 4, which shows different shapes including negative skew, positive skew, monotonically decreasing and almost symmetric bell-shape.

6.1. Some sub-models of the GWW distribution

When $\alpha = \beta = \delta = k = 1$, the GWW distribution reduces to the exponential exponential (EE) distribution with pdf as follows:

$$f(x) \quad = \quad \frac{1}{\lambda} \times \exp\Big(1 + \frac{x}{\lambda} - e^{\frac{x}{\lambda}}\Big).$$

By letting $\alpha = \beta = \delta = 1$, we obtain the exponential power (EP) distribution (Smith and Bain 1975) as a special case of the GWW distribution, whose pdf is given by

$$f(x) \quad = \quad \frac{kx^{k-1}e^{(\frac{x}{\lambda})^k}}{\lambda^k}\exp\Big[1 - e^{(\frac{x}{\lambda})^k}\Big].$$

We can also obtain Chen distribution (Chen 2000) by letting $\beta = \delta = \lambda = 1$ with pdf given by

$$f(x) \quad = \quad \alpha kx^{k-1}e^{x^k}\exp\Big[-\alpha\big(e^{x^k} - 1\big)\Big].$$

When $\beta = \delta = k = 1$, the GWW distribution reduces to Gompertz distribution (Gompertz 1895) with pdf given by

$$f(x) \quad = \quad \frac{\alpha e^{\frac{x}{\lambda}}}{\lambda}\exp\Big[-\alpha\big(e^{\frac{x}{\lambda}} - 1\big)\Big].$$

By setting $\alpha = \beta = 1$, we obtain the gamma exponential power (GEP) distribution. The pdf of GEP distribution is given by

$$f(x) \quad = \quad \frac{1}{\Gamma(\delta)} \times \frac{kx^{k-1}e^{(\frac{x}{\lambda})^k}}{\lambda^k}\big[e^{(\frac{x}{\lambda})^k} - 1\big]^{\delta-1}\exp\Big[1 - e^{(\frac{x}{\lambda})^k}\Big].$$

Moreover, when $\alpha = \delta = 1$, we obtain the exponentiated exponential power (EEP) distribution as a special case of the GWW distribution, whose pdf is given by

$$f(x) \quad = \quad \frac{\beta kx^{k-1}e^{(\frac{x}{\lambda})^k}}{\lambda^k}\big[e^{(\frac{x}{\lambda})^k} - 1\big]^{\beta-1}\exp\Big[-\big(e^{(\frac{x}{\lambda})^k} - 1\big)^{\beta}\Big].$$

We also obtain the gamma exponentiated exponential power (GEEP) distribution by setting $\alpha = 1$ and the corresponding pdf is given by

$$f(x) \quad = \quad \frac{\beta}{\Gamma(\delta)} \times \frac{kx^{k-1}e^{(\frac{x}{\lambda})^k}}{\lambda^k}\big[e^{(\frac{x}{\lambda})^k} - 1\big]^{\beta\delta-1}\exp\Big[-\big(e^{(\frac{x}{\lambda})^k} - 1\big)^{\beta}\Big].$$

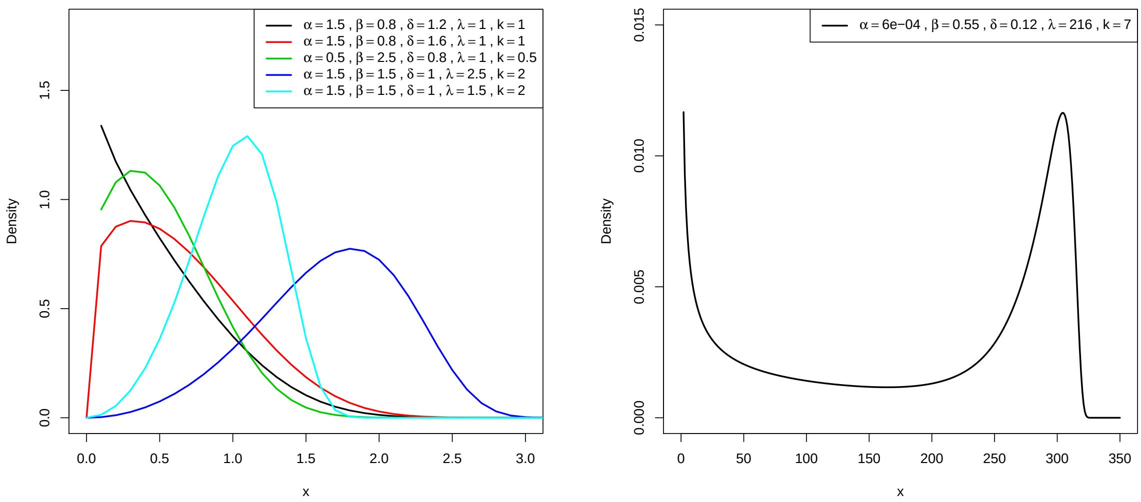

Figure 4: Graphs of GWW pdf with selected parameters. Note that, the focus and originality of this paper is that the GWG family of distributions are capable of generating bathtub-shaped probability density functions (pdf). Because of its importance and in a different scale, we are presenting a bathtub-shaped pdf in a separated plot (right panel) as above.

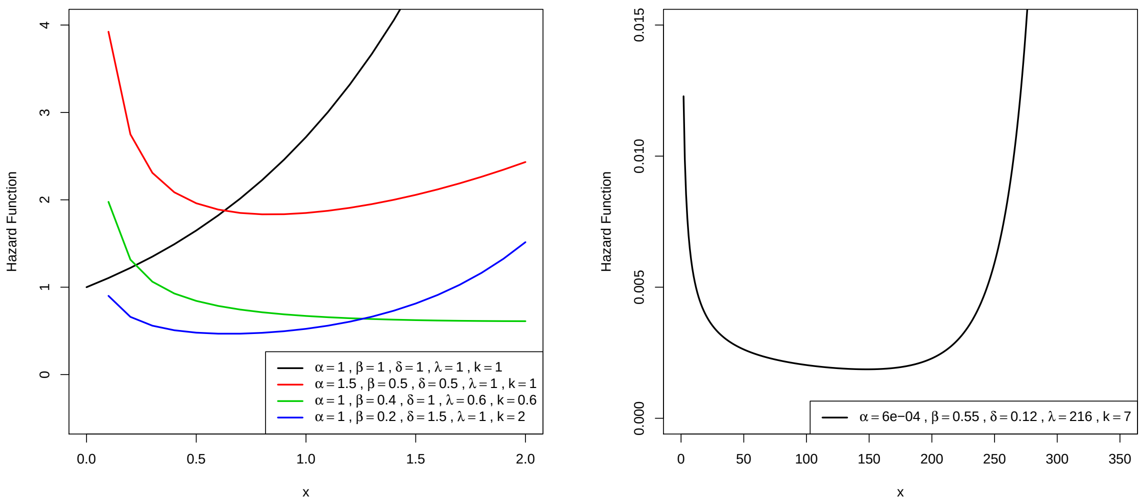

Figure 5: Graphs of GWW hazard rate functions (hrf) with selected parameters. Similar with pdf of GWW, we are presenting a bathtub-shaped hrf separately in the right panel.

6.2. Expansion of the GWW density

From equation (9), we can obtain

$$f_{GWG}(x) = \sum_{i,j=0}^{\infty} \omega_{i,j} h_{\beta(\delta+i)+j}(x;\lambda,k), \tag{35}$$

where

$$\omega_{i,j} = \frac{(-1)^i \beta \alpha^{\delta+i} \Gamma[\beta(\delta+i)+j+1]}{i!j!\Gamma(\delta)\Gamma[\beta(\delta+i)+1][\beta(\delta+i)+j]}, \tag{36}$$

and

$$h_{\beta(\delta+i)+j}(x;\lambda,k) = \big[\beta(\delta+i)+j\big]\frac{k}{\lambda}\Big(\frac{x}{\lambda}\Big)^{k-1} e^{-(\frac{x}{\lambda})^k}\big[1-e^{-(\frac{x}{\lambda})^k}\big]^{\beta(\delta+i)+j-1}$$

denotes pdf of the exponentiated Weibull (EW) (Nassar and Eissa 2003) distribution with power $\beta(\delta+i)+j$. Therefore, we can obtain the statistical properties of the GWW distribution from those of the EW distribution.

6.3. Hazard, reverse hazard and quantile functions

The hazard and reverse hazard functions of the GWW distribution can be obtained as

$$h_{GWW}(x) = \frac{\beta\alpha^{\delta}\frac{kx^{k-1}e^{(\frac{x}{\lambda})^k}}{\lambda^k}\big[e^{(\frac{x}{\lambda})^k}-1\big]^{\beta\delta-1}\exp\Big[-\alpha\big(e^{(\frac{x}{\lambda})^k}-1\big)^{\beta}\Big]}{\Gamma(\delta)-\gamma\Big[\delta,\alpha\big(e^{(\frac{x}{\lambda})^k}-1\big)^{\beta}\Big]},$$

and

$$\tau_{GWW}(x) = \frac{\beta\alpha^{\delta}\frac{kx^{k-1}e^{(\frac{x}{\lambda})^k}}{\lambda^k}\big[e^{(\frac{x}{\lambda})^k}-1\big]^{\beta\delta-1}\exp\Big[-\alpha\big(e^{(\frac{x}{\lambda})^k}-1\big)^{\beta}\Big]}{\gamma\Big[\delta,\alpha\big(e^{(\frac{x}{\lambda})^k}-1\big)^{\beta}\Big]},$$

respectively. The graphs of the hrf for selected parameters are given in Figure 5. The plots show various shapes including monotonically decreasing, monotonically increasing, and bathtub shapes for several combinations of parameters values. This attractive flexibility makes the GWW hazard function be useful and suitable for monotonic and non-monotonic empirical hazard behaviors which are more likely to be encountered in real life situations.

In addition, we can also obtain the quantile function of GWW distribution from equation (14) and (15) as

$$x_u = \lambda\big[-\log(1-q)\big]^{\frac{1}{k}} = \lambda\Bigg[-\log\Bigg(1-\frac{\Big[\gamma^{-1}\big(\delta,u\Gamma(\delta)\big)\Big]^{\frac{1}{\beta}}}{\Big[\gamma^{-1}\big(\delta,u\Gamma(\delta)\big)\Big]^{\frac{1}{\beta}}+\alpha^{\frac{1}{\beta}}}\Bigg)\Bigg]^{\frac{1}{k}}.$$

6.4. Moments, incomplete moments, moment generating and characteristic functions

Let $X \sim GWW(\alpha,\beta,\delta,\lambda,k), x \geq 0, \alpha,\beta,\delta,\lambda,k > 0$, the s^{th} moment of X can be obtained from (35) as

$$E(X^s) = \sum_{i,j=0}^{\infty} \omega_{i,j} E(Z^s_{i,j}),$$

where $E(Z_{i,j}^s)$ denotes the s^{th} moment of $Z_{i,j}$, which follows the EW distribution (Nassar and Eissa 2003), with power $\beta(\delta+i)+j$ and $\omega_{i,j}$ is defined as (36). The s^{th} moment of $Z_{i,j}$ is given by (Pal, Ali, and Woo 2006) and we present it as follows:

$$E(Z_{i,j}^s) = \beta(\delta+i)+j]\lambda^s \Gamma\left(\frac{s}{k}+1\right) \sum_{t=0}^{\beta(\delta+i)+j-1} \binom{\beta(\delta+i)+j-1}{i} (-1)^t (t+1)^{\frac{s}{k}-1},$$

if $(\beta(\delta+i)+j) \in N$,

$$E(Z_{i,j}^s) = [\beta(\delta+i)+j]\lambda^s \Gamma\left(\frac{s}{k}+1\right) \sum_{t=0}^{\beta(\delta+i)+j-1} \frac{P_t^{\beta(\delta+i)+j-1}}{t!} (-1)^t (t+1)^{\frac{s}{k}-1},$$

if $(\beta(\delta+i)+j) \notin N$, for $k = 0, 1, 2, ...$, where

$$P_t^{\alpha} = \alpha(\alpha-1)(\alpha-2)\cdots(\alpha-t+1),\ for\ any\ \alpha \notin N.$$

Similarly, the incomplete moments and mgf can be obtained as

$$I_X(y) = \int_0^y x^s f_{GWW}(x)dx = \sum_{i,j=0}^{\infty} \omega_{i,j} I_{i,j}(y),$$

where $I_{i,j}(y) = \int_0^y x^s h_{\beta(\delta+i)+j}(x, \underline{\theta})dx$ and $M_X(t) = \sum_{i,j=0}^{\infty} \omega_{i,j} E(e^{tZ_{i,j}})$, where $E(e^{tZ_{i,j}})$ is the mgf of the EW (Cordeiro *et al.* 2013b) distribution with power $\beta(\delta+i)+j$ and $\omega_{i,j}$ is defined as (36). The characteristic function is given by $\phi(t) = E(e^{itX})$, where $i = \sqrt{-1}$. Thus, the characteristic function is given by

$$\phi(t) = \sum_{i,j=0}^{\infty} \omega_{i,j} \phi_{\beta(\delta+i)+j}(t),$$

where $\phi_{\beta(\delta+i)+j}(t)$ denotes the characteristic function of the EW distribution with power $\beta(\delta+i)+j$ and $\omega_{i,j}$ is defined as (36).

In addition, the conditional expectations of the GWW distribution can also be obtained from the EW (Nassar and Eissa 2003) distribution as follow:

$$E(X^t \mid X > x) = \frac{\Gamma(\delta)}{\Gamma(\delta) - \gamma\left\{\delta, \alpha\left[e^{(\frac{x}{\lambda})^k} - 1\right]^{\beta}\right\}} \times \sum_{i,j=0}^{\infty} \omega_{i,j} I_{i,j}(y),$$

where $I_{i,j}(y) = \int_0^y x^t h_{\beta(\delta+i)+j}(x, \theta)dx$ is the t^{th} incomplete moment of EW and $\omega_{i,j}$ is defined as (36).

The first six moments, SD, CV, CS, and CK for different selected values of the parameters of the GWW distribution are listed in the Tables 3 below. The graphs of CS and CK versus the shape parameters show the dependence of skewness and kurtosis measures on these parameters, β and δ. Also, plots of CS and CK versus the shape parameter k can be readily obtained.

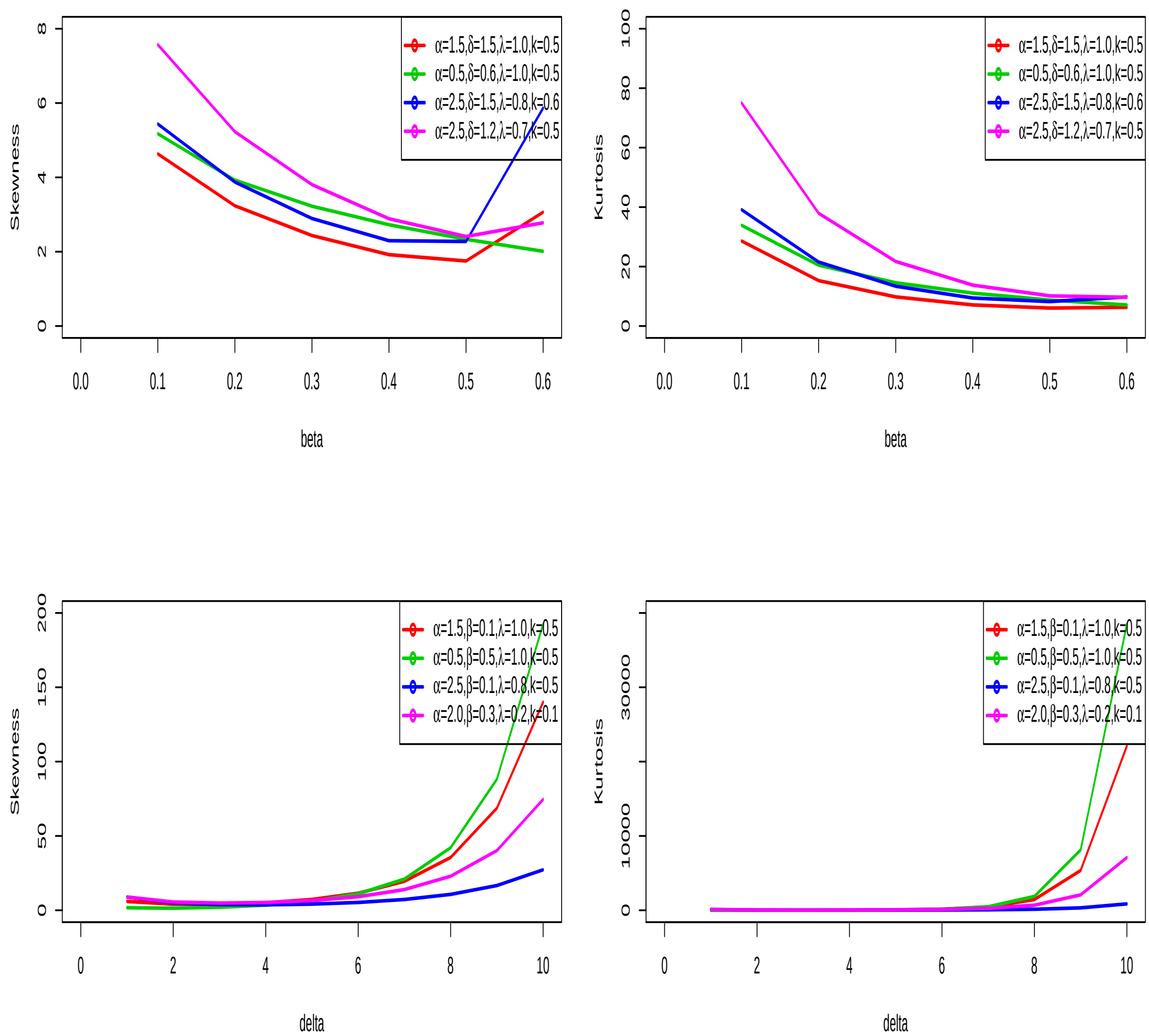

Figure 6: Graphs of GWW skewness and kurtosis with selected parameters

Table 3: GWW moments for selected values

	$(\alpha,\beta,\delta,\lambda,k)$			
Moments	(1.5,0.8,1.5,1.0,0.5)	(0.5,2.5,0.6,1.0, 0.5)	(2.5,1.0,1.5,0.8,0.6)	(0.5,3.0,0.8,0.8, 0.8)
$E(X)$	0.5911	0.4297	0.2391	0.5010
$E(X^2)$	0.8997	0.2921	0.1137	0.2869
$E(X^3)$	2.0769	0.2486	0.0770	0.1792
$E(X^4)$	6.2139	0.2450	0.0658	0.1193
$E(X^5)$	22.3739	0.2691	0.0666	0.0836
$E(X^6)$	92.9078	0.3221	0.0770	0.0610
SD	0.7419	0.3278	0.2377	0.1892
CV	1.2551	0.7628	0.9937	0.3777
CS	2.1910	0.8710	1.6992	-0.0662
CK	9.3210	3.3866	6.6864	2.5279

6.5. Mean deviation, Bonferroni and Lorenz curves

Let $X \sim GWW(\alpha,\beta,\delta,\lambda,k), x \geq 0, \alpha,\beta,\delta,\lambda,k > 0,\ \mu = E(X)$ be the mean and $M = Median(X)$ be the median. Then the mean deviation about the mean and the mean deviation about the median can be obtained as

$$\delta_1(X) = 2\mu F_{GWW}(\mu) - 2\int_0^{\mu} x f_{GWW}(x)dx, \quad \text{and} \quad \delta_2(X) = \mu - 2\int_0^{M} x f_{GWW}(x)dx,$$

respectively, and $m(z) = \int_0^z x f_{GWW}(x)dx$ is the first incomplete moment. Bonferroni and Lorenz curves are given by

$$B(p) = \frac{1}{p\mu}\int_0^q x f(x)dx = \frac{1}{p\mu}\sum_{i,j=0}^{\infty} \omega_{i,j} I_{\beta(\delta+i)+j}(i,j),$$

and

$$L(p) = \frac{1}{\mu}\int_0^q x f(x)dx = \frac{1}{\mu}\sum_{i,j=0}^{\infty} \omega_{i,j} I_{\beta(\delta+i)+j}(i,j),$$

respectively, where $I_{\beta(\delta+i)+j}(i,j) = \int_0^q x f_{\beta(\delta+i)+j}(x;\lambda,k)dx$, is the first incomplete moment of the EW (Nassar and Eissa 2003) distribution with power $\beta(\delta+i)+j$ and $\omega_{i,j}$ is given by (36).

6.6. Order statistics

The distribution of the i^{th} order statistic for the GWW distribution is presented in this section. Let $X_1,\ldots,X_n$ be i.i.d random variables distributed according to (35). The pdf of the i^{th} order statistics, for example $X_{i:n}$, is given by

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!}\sum_{q=0}^{n-i}\sum_{m=0}^{\infty}\binom{n-i}{q}\frac{(-1)^q d_{m,q+i-1}\Gamma(\delta i+\delta q+m)}{[\Gamma(\delta)]^{i+k}} f_{GWW}(x;\alpha,\beta,\delta_*),$$

that is, the pdf of the i^{th} order statistic is a linear combination of GWW densities with parameters $(\alpha,\beta,\delta_*,\lambda,k)$, where $\delta_* = \delta i+\delta q+m$.

Therefore, properties of the order statistics of the GWW distribution can be obtained from those of EW (Nassar and Eissa 2003) distribution. For instance, we can obtain the the t^{th} moment of the i^{th} order statistics and is given by

$$E(X_{i:n}^t) = \sum_{l,j=0}^{\infty}\sum_{q=0}^{n-i}\sum_{m=0}^{\infty}\frac{n!\omega_{i,j}}{(i-1)!(n-i)!}\binom{n-i}{q}\frac{(-1)^q d_{m,q+i-1}\Gamma(\delta i+\delta q+m)}{[\Gamma(\delta)]^{i+q}} \times E(Z_{l,j}(x;\alpha,\beta,\delta_*)^t),$$

where $E(Z_{l,j}(x;\alpha,\beta,\delta_*)^t)$ is the t^{th} moment of EW distribution with power $\beta(\delta_*+i)+j$, where $\delta_* = \delta i+\delta q+m$ and $\omega_{i,j}$ is defined as (36).

6.7. Rényi entropy

By equation (21), we obtain Rényi entropy for the GWW distribution as

$$I_R(\nu) = (1-\nu)^{-1}\Bigg\{\nu\log(\beta)+\delta\nu\log(\alpha)-\nu\log(\Gamma(\delta))+\log\Bigg[\sum_{i=0}^{\infty}\sum_{s=0}^{\infty}\frac{(-1)^i(\alpha\nu)^i}{i!} \times b_{s,(\beta\delta+1)\nu+\beta i}\frac{\nu^{\nu}}{[(\beta\delta-1)\nu+\beta i+s+\nu]^{\nu}}\times e^{(1-\nu)I_{REW}}\Bigg]\Bigg\}, \quad (37)$$

where I_{REW} denotes the Rényi entropy for the exponentiated Weibull distribution (Nassar and Eissa 2003) with parameter $\beta^* = \frac{(\beta\delta-1)\nu+\beta i+s+\nu}{\nu}$ and $b_{s,m} = s^{-1}\sum_{l=1}^{s}\Big[m(l+1)-s\Big]b_{s-l,m}$ with $b_{0,m} = 1, m = (\beta\delta+1)\nu+\beta i$.

7. Monte Carlo simulation study

In this section, a simulation study is conducted to assess the performance and examine the mean estimate, average bias, root mean square error of the maximum likelihood estimators for each parameter for both GWU and GWW distributions. We study the performance of the GWU and GWW distributions by conducting various simulations for different parameter values. The simulation study is repeated for $N = 5000$ times each with sample size $n = 25, 50, 75, 100, 200, 400, 800, 1600, 3200$ and parameter values $I : \alpha = 0.5, \beta = 1.5, \delta = 1.5, \theta = 1.6, II : \alpha = 1, \beta = 1, \delta = 1.5, \theta = 1, III : \alpha = 2.5, \beta = 1.5, \delta = 0.4, \theta = 0.5, IV : \alpha = 1.2, \beta = 1.2, \delta = 0.8, \theta = 1.5$, for the GWU distribution and $I : \alpha = 0.8, \beta = 1.6, \delta = 0.3, \lambda = 0.4, k = 1, II : \alpha = 0.9, \beta = 1.2, \delta = 0.5, \lambda = 0.6, k = 1.2, III : \alpha = 1.0, \beta = 1.0, \delta = 0.8, \lambda = 1.0, k = 1.5, IV : \alpha = 0.5, \beta = 0.8, \delta = 0.5, \lambda = 0.8, k = 0.5$, for the GWW distribution, respectively.

Tables 4, 5 and 6, 7 lists the means MLEs of the model parameters along with the respective average bias and root mean squared errors (RMSE) for the GWU and GWW distributions, respectively. From the results, we can verify that as the sample size n increases, the mean estimates of the parameters tend to be closer to the true parameter values, since RMSEs decay toward zero. The bias and RMSEs are given by:

$$Bias(\hat{\theta}) = \frac{\sum_{i=1}^{n}\hat{\theta}_i}{n} - \theta, \quad \text{and} \quad RMSE(\hat{\theta}) = \sqrt{\frac{\sum_{i=1}^{n}(\hat{\theta}_i-\theta)^2}{n}},$$

respectively.

8. Applications

In this section, we present two examples to illustrate the flexibility of the GWU and GWW distributions and their sub-models for data modeling.

The first data consists of the lifetimes of $n = 50$ devices given by Aarset data (Aarset 1987). It is known to have a bathtub-shaped hazard function, thus been widely studied and is given by: 0.1 0.2 1.0 1.0 1.0 1.0 1.0 2.0 3.0 6.0 7.0 11.0 12.0 18.0 18.0 18.0 18.0 18.0 21.0 32.0 36.0 40.0 45.0 46.0 47.0 50.0 55.0 60.0 63.0 63.0 67.0 67.0 67.0 67.0 72.0 75.0 79.0 82.0 82.0 83.0 84.0 84.0 84.0 85.0 85.0 85.0 85.0 85.0 86.0 86.0.

The second data gives failure and running times of a sample of $n = 30$ devices given by Meeker and Escobar (Meeker and Escobar 1998) and this data also has a bathtub shaped hazard function and is given by: 2 10 13 23 23 28 30 65 80 88 106 143 147 173 181 212 245 247 261 266 275 293 300 300 300 300 300 300 300 300 .

Some descriptive statistics of these two data sets are given in Table 8.

The maximum likelihood estimates (MLEs) of the GWU and GWW parameters α, β, δ, and $\underline{\theta}$ are computed by maximizing the objective function via the subroutine NLMIXED in SAS. The estimated values of the parameters (standard error in parenthesis), -2log-likelihood statistic, Akaike Information Criterion, $AIC = 2p - 2\ln(L)$, Bayesian Information Criterion, $BIC = p\ln(n) - 2\ln(L)$, and Consistent Akaike Information Criterion, $AICC = AIC + 2\frac{p(p+1)}{n-p-1}$, where $L = L(\hat{\boldsymbol{\Delta}})$ is the value of the likelihood function evaluated at the parameter estimates, n is the number of observations, and p is the number of estimated parameters, and Kolmogorov-Smirnov (KS) statistic ($KS = \max_{1\leq i\leq n}\{G_{_{GWG}}(x_i) - \frac{i-1}{n}, \frac{i}{n} - G_{_{GWG}}(x_i)\}$) are presented in Table 9 and 10. In order to compare the models, we use the criteria stated above. Note that for the value of the log-likelihood function at its maximum (ℓ_n), larger value is good and preferred,

Table 4: Monte Carlo simulation results for GWU distribution: mean estimate, average bias, and RMSE

		I			II		
Parameter	n	Mean	Average Bias	RMSE	Mean	Average Bias	RMSE
α	25	1.2113	0.7139	3.1626	1.7216	0.7216	3.5384
	50	1.0893	0.5893	2.3287	1.5490	0.5490	2.6918
	75	0.8740	0.3740	1.7007	1.4273	0.4273	2.2506
	100	0.8345	0.3345	1.4842	1.2994	0.2994	1.8730
	200	0.6818	0.1818	0.9301	1.1211	0.1211	1.3929
	400	0.5626	0.0626	0.5420	1.0271	0.0271	1.1389
	800	0.5132	0.0132	0.3916	0.9001	-0.0999	0.7646
	1600	0.4848	-0.0152	0.2803	0.8124	-0.1876	0.4785
	3200	0.4687	-0.0313	0.2021	0.7909	-0.2091	0.2800
β	25	3.0323	1.5323	2.0056	2.3244	1.3244	1.6974
	50	2.8732	1.3732	2.0006	2.1701	1.1701	1.6716
	75	2.7714	1.2714	1.9362	2.0924	1.0924	1.6628
	100	2.7162	1.2162	1.9034	2.0632	1.0632	1.6511
	200	2.6277	1.1277	1.8255	1.8542	0.8542	1.4436
	400	2.4775	0.9775	1.6645	1.7263	0.7263	1.3110
	800	2.3439	0.8439	1.5235	1.5232	0.5232	1.0124
	1600	2.1536	0.6536	1.2958	1.3518	0.3518	0.6608
	3200	2.0095	0.5095	1.1140	1.2358	0.2358	0.3515
δ	25	1.0949	-0.4051	1.5079	1.0751	-0.4249	1.7584
	50	1.3423	-0.1577	1.7580	1.2335	-0.2665	1.7850
	75	1.3513	-0.1487	1.6424	1.2455	-0.2545	1.6163
	100	1.3641	-0.1340	1.5612	1.1761	-0.3239	1.2890
	200	1.2844	-0.2156	1.1153	1.1413	-0.3587	0.9076
	400	1.2375	-0.2625	0.7970	1.1190	-0.3810	0.7008
	800	1.2370	-0.2630	0.6618	1.1364	-0.3636	0.5612
	1600	1.2675	-0.2325	0.5496	1.1605	-0.3395	0.4537
	3200	1.2930	-0.2070	0.4623	1.1992	-0.3008	0.3670
θ	25	1.8684	0.2684	0.4773	1.2085	0.2085	0.3601
	50	1.8534	0.2534	0.4296	1.1904	0.1904	0.3158
	75	1.8234	0.2234	0.3772	1.1796	0.1796	0.2930
	100	1.8216	0.2216	0.3691	1.1728	0.1728	0.2834
	200	1.8091	0.2091	0.3484	1.1405	0.1405	0.2474
	400	1.7791	0.1791	0.3130	1.1198	0.1198	0.2244
	800	1.7534	0.1534	0.2823	1.0840	0.0840	0.1717
	1600	1.7190	0.1190	0.2419	1.0546	0.0546	0.1102
	3200	1.6922	0.0922	0.2086	1.0350	0.0350	0.0555

Table 5: Monte Carlo simulation results for GWU distribution: mean estimate, average bias, and RMSE

		III			IV		
Parameter	n	Mean	Average Bias	RMSE	Mean	Average Bias	RMSE
α	25	3.7320	1.2320	5.2669	3.1961	1.9961	5.5193
	50	3.8880	1.3880	4.8103	3.0213	1.8213	4.5538
	75	3.6292	1.1292	4.2663	2.6633	1.4633	3.4775
	100	3.6382	1.1382	4.2407	2.5454	1.3454	3.3138
	200	3.3322	0.8322	3.5639	2.1108	0.9108	2.3186
	400	3.2340	0.7340	3.3193	1.7823	0.5823	1.6210
	800	3.0544	0.5544	2.7256	1.5654	0.3654	1.2227
	1600	2.9853	0.4853	2.2392	1.3730	0.1730	0.6680
	3200	2.9198	0.4198	1.8403	1.2753	0.0753	0.3757
β	25	1.5649	0.0649	1.0530	1.7431	0.5431	1.0095
	50	1.3337	-0.1663	0.8907	1.5984	0.3984	1.0152
	75	1.2332	-0.2668	0.7684	1.5318	0.3318	0.9746
	100	1.2347	-0.2653	0.7381	1.4812	0.2812	0.9184
	200	1.2450	-0.2550	0.6390	1.4129	0.2129	0.8179
	400	1.3000	-0.2000	0.5450	1.3378	0.1378	0.6711
	800	1.3560	-0.1440	0.4534	1.2818	0.0818	0.5374
	1600	1.4206	-0.0794	0.3687	1.2569	0.0569	0.3955
	3200	1.4728	-0.0272	0.2983	1.2285	0.0285	0.2700
δ	25	0.8864	0.4864	1.5351	1.1253	0.3253	1.8677
	50	0.9760	0.5760	1.5304	1.2626	0.4626	1.8562
	75	0.8957	0.4957	1.1348	1.2134	0.4134	1.5772
	100	0.8127	0.4127	0.8361	1.1591	0.3591	1.2916
	200	0.6743	0.2743	0.4993	1.0007	0.2007	0.7172
	400	0.5714	0.1714	0.3164	0.9299	0.1299	0.4838
	800	0.5078	0.1078	0.2136	0.8887	0.0887	0.3575
	1600	0.4586	0.0586	0.1388	0.8436	0.0436	0.2551
	3200	0.4279	0.0279	0.0942	0.8238	0.0238	0.1859
θ	25	0.4319	-0.0681	0.2054	1.7041	0.2041	0.6352
	50	0.4428	-0.0572	0.1887	1.6819	0.1819	0.5482
	75	0.4388	-0.0612	0.1788	1.6539	0.1539	0.4784
	100	0.4469	-0.0531	0.1756	1.6482	0.1482	0.4626
	200	0.4528	-0.0472	0.1567	1.6139	0.1139	0.3977
	400	0.4672	-0.0328	0.1413	1.5799	0.0799	0.3236
	800	0.4783	-0.0217	0.1209	1.5487	0.0487	0.2582
	1600	0.4916	-0.0084	0.1037	1.5318	0.0318	0.1835
	3200	0.5012	0.0012	0.0859	1.5150	0.0150	0.1209

Table 6: Monte Carlo simulation results for GWW distribution: mean estimate, average bias, and RMSE

		I			II		
Parameter	n	Mean	Average Bias	RMSE	Mean	Average Bias	RMSE
α	25	1.4111	1.6111	2.0556	1.2597	0.3597	2.0332
	50	1.3657	0.5657	1.9041	1.8731	0.9731	2.3628
	75	1.3151	0.5151	1.7345	1.8999	0.9999	2.2978
	100	1.2305	0.4305	1.6311	1.8312	0.9312	2.1015
	200	1.0764	0.2764	1.3353	1.6426	0.7426	1.7880
	400	0.9667	0.1667	1.0912	1.4320	0.5320	1.4453
	800	0.8323	0.0323	0.8472	1.3082	0.4082	1.2432
	1600	0.8298	0.0298	0.7923	1.1277	0.2277	0.9012
	3200	0.8283	0.0283	0.7736	1.0537	0.1537	0.7300
β	25	1.5221	-0.0779	1.0899	0.9488	-0.2512	1.0695
	50	1.5254	-0.0746	1.2401	1.1865	-0.0135	1.0872
	75	1.5237	-0.0763	1.3291	1.1436	-0.0564	1.0509
	100	1.5279	-0.0721	1.2235	1.1340	-0.0660	1.0689
	200	1.5322	-0.0678	1.2059	1.0849	-0.1151	0.9382
	400	1.5774	-0.0226	1.1630	1.0811	-0.1189	0.8791
	800	1.5734	-0.0266	1.0250	1.0590	-0.1410	0.7351
	1600	1.5838	-0.0162	0.9319	1.0521	-0.1479	0.6227
	3200	1.5438	-0.0562	0.6773	1.0517	-0.1483	0.4862
δ	25	0.7776	0.4776	1.3797	0.8674	0.3674	1.5471
	50	0.9079	0.6079	1.5213	1.4396	0.9396	2.0780
	75	0.9202	0.6202	1.4821	1.4847	0.9847	2.0539
	100	0.8887	0.5887	1.4373	1.4782	0.9782	2.0212
	200	0.8051	0.5051	1.2080	1.4481	0.9481	1.9149
	400	0.7035	0.4035	0.9875	1.3136	0.8136	1.6746
	800	0.5588	0.2588	0.6596	1.2048	0.7048	1.5002
	1600	0.4572	0.1572	0.4272	1.0225	0.5225	1.1389
	3200	0.3928	0.0928	0.2306	0.8521	0.3521	0.7663
λ	25	0.3232	-0.0768	0.2692	0.3881	-0.2119	0.4425
	50	0.3003	-0.0997	0.2543	0.5002	-0.0998	0.3366
	75	0.2979	-0.1021	0.2551	0.5003	-0.0997	0.3436
	100	0.2909	-0.1091	0.2485	0.4953	-0.1047	0.3337
	200	0.2880	-0.1119	0.2517	0.4680	-0.1320	0.3255
	400	0.3032	-0.0968	0.2421	0.4617	-0.1383	0.3143
	800	0.3116	-0.0884	0.2253	0.4649	-0.1351	0.3021
	1600	0.3385	-0.0615	0.2100	0.4704	-0.1296	0.2719
	3200	0.3496	-0.0504	0.1891	0.4952	-0.1048	0.2500
k	25	1.3177	0.3177	0.9108	1.3263	0.1263	1.3839
	50	1.1464	0.1464	0.6429	1.5864	0.3864	1.0178
	75	1.0753	0.0753	0.5442	1.4788	0.2788	0.8604
	100	1.0128	0.0128	0.4760	1.4348	0.2348	0.7814
	200	0.9259	-0.0741	0.3826	1.2766	0.0766	0.5867
	400	0.8912	-0.1088	0.3273	1.1838	-0.0162	0.4632
	800	0.8785	-0.1215	0.2823	1.1273	-0.0727	0.3594
	1600	0.9030	-0.0970	0.2442	1.1089	-0.0911	0.2926
	3200	0.9250	-0.0750	0.1962	1.1245	-0.0755	0.2342

Table 7: Monte Carlo simulation results for GWW distribution: mean estimate, average bias, and RMSE

		III			IV		
Parameter	n	Mean	Average Bias	RMSE	Mean	Average Bias	RMSE
α	25	0.7654	-0.2346	1.7530	1.5137	1.0137	2.2141
	50	2.2911	1.2911	2.7660	1.6133	1.1133	2.1858
	75	2.2837	1.2837	2.5529	1.6818	1.1818	2.1972
	100	2.2296	1.2296	2.4651	1.6428	1.1428	2.1245
	200	1.9894	0.9894	2.0170	1.4936	0.9994	1.8406
	400	1.8281	0.8281	1.7085	1.2771	0.7771	1.4535
	800	1.6887	0.6887	1.4409	1.0323	0.5323	1.0468
	1600	1.6107	0.6107	1.2653	0.7962	0.2962	0.6006
	3200	1.4834	0.4834	1.0017	0.6491	0.1491	0.2891
β	25	0.4491	-0.5509	1.0208	0.8165	0.0165	0.8797
	50	0.9048	-0.0952	0.9894	0.7427	-0.0573	0.8801
	75	0.8385	-0.1615	0.8989	0.6953	-0.1047	0.8374
	100	0.8427	-0.1573	0.8459	0.6918	-0.1082	0.8448
	200	0.8445	-0.1555	0.7775	0.6691	-0.1309	0.7426
	400	0.8337	-0.1663	0.6693	0.6458	-0.1542	0.6458
	800	0.8233	-0.1767	0.5808	0.6562	-0.1438	0.5133
	1600	0.7954	-0.2046	0.5142	0.6921	-0.1079	0.3945
	3200	0.7884	-0.2116	0.4255	0.7244	-0.0756	0.3123
δ	25	0.6440	-0.1560	1.4363	1.4641	0.9641	2.0198
	50	1.9927	1.1927	2.4202	1.6718	1.1718	2.1626
	75	2.0496	1.2496	2.3840	1.7934	1.2934	2.2456
	100	2.0282	1.2282	2.3704	1.7924	1.2924	2.2401
	200	1.9370	1.1370	2.2145	1.7205	1.2205	2.0941
	400	1.9110	1.1110	2.1317	1.5382	1.0382	1.7831
	800	1.8432	1.0432	2.0047	1.2565	0.7565	1.3622
	1600	1.8388	1.0388	1.9322	0.9572	0.4572	0.8627
	3200	1.6646	0.8646	1.6220	0.7516	0.2516	0.4760
λ	25	0.3752	-0.6248	0.8168	0.7368	-0.0632	0.9234
	50	0.8789	-0.1211	0.4086	0.6406	-0.1594	0.8107
	75	0.8689	-0.1311	0.3924	0.5883	-0.2117	0.7975
	100	0.8729	-0.1271	0.3811	0.5814	-0.2186	0.7938
	200	0.8539	-0.1461	0.3682	0.5414	-0.2586	0.7352
	400	0.8330	-0.1670	0.3677	0.5471	-0.2529	0.7351
	800	0.8241	-0.1759	0.3657	0.5901	-0.2099	0.6990
	1600	0.8079	-0.1921	0.3652	0.6731	-0.1269	0.6978
	3200	0.8201	-0.1799	0.3542	0.7212	-0.0788	0.6526
k	25	1.1718	-0.3282	1.9945	0.7013	0.2013	0.4864
	50	2.4561	0.9561	1.8343	0.5924	0.0924	0.3084
	75	2.2968	0.7968	1.5723	0.5536	0.0536	0.2489
	100	2.1698	0.6698	1.3819	0.5310	0.0310	0.2161
	200	1.8856	0.3856	1.0197	0.4885	-0.0115	0.1609
	400	1.6780	0.1780	0.7698	0.4650	-0.0350	0.1248
	800	1.5614	0.0614	0.5897	0.4592	-0.0408	0.1013
	1600	1.4836	-0.0164	0.4575	0.4621	-0.0379	0.0805
	3200	1.4661	-0.0339	0.3724	0.4694	-0.0306	0.0639

Table 8: Descriptive statistics of application data sets

Data	n	Mean	Median	Minimum	Maximum	Variance	SD
Aarset	50	45.686	48.5	0.1	86.0	1078.2	32.8352
Meeker	30	177.03	196.5	2	300	13223	114.9913

Table 9: GWU estimation for Aarset data

Model	α	β	δ	θ	-2 Log Likelihood	AIC	BIC	AICC	KS	SS
EU(θ)	1	1	1	118.59	462.3	464.3	466.2	464.4	0.1962	0.4998
	-	-	-	(5.2776)						
WU(θ, β)	1	0.2745	1	86.1605	423.8	427.8	431.6	428.0	0.2404	1.1142
	-	(0.03494)	-	(0.1979)						
GEU(θ, δ)	1	1	0.6870	128.72	453.3	457.3	461.1	457.5	0.1948	0.4963
	-	-	(0.09336)	(8.8515)						
EU(θ, α)	0.7038	1	1	108.12	461.5	465.5	469.3	465.8	0.2298	0.5739
	(0.2671)	-	-	(9.9097)						
Phani	0.5455	0.3478	1	86.1591	410.3	416.3	422.0	416.8	0.1104	0.0743
	(0.09907)	(0.04773)	-	(0.1684)						
GWU(θ, β, δ)	1	0.2684	1.5248	86.1002	408.8	414.8	420.6	415.4	0.1121	0.0754
	-	(0.03266)	(0.1487)	(0.1220)						
GU(θ, α, δ)	0.02046	1	0.2750	86.7543	415.8	421.8	427.5	422.3	0.1553	0.1952
	(0.01124)	-	(0.05022)	(0.4914)						
GWU	44.0462	0.03578	45.3699	86.0003	398.2	406.3	413.9	407.1	0.0645	0.0336
	(0.00713)	(0.000897)	(0.1289)	(0.0144)						
BW	k	λ	a	b						
	0.9653	2.0997	0.4511	0.04673	477.5	485.5	486.4	493.2	0.1516	0.4494
	(0.02915)	(0.2233)	(0.1527)	(0.006841)						
GMW	β	θ	λ	δ						
	0.003631	0.4785	0.03249	2.7121	453.3	461.3	462.2	468.9	0.1275	0.2723
	(0.01222)	(0.2023)	(0.01365)	(2.2044)						

and for the Kolmogorov-Smirnov test statistic (K-S), smaller value is preferred. The GWU distribution is fitted to the Aarset data (Aarset 1987) set and these fits are compared to the fits using the EU(θ), WU(θ, β), GEU(θ, δ), EU(θ, α), Phani, GWU(θ, β, δ), GEU(θ, α, δ) and GWU distributions. Similarly, the GWW distribution is also fitted to the Meeker and Escobar (Meeker and Escobar 1998) data set and these fits are compared to the fits using the EE, EP, Chen, Gompertz, GEP, EEP and GEEP distributions. The GWW and GWU distributions was compared with the non-nested beta exponentiated Weibull (BEW) (Cordeiro, Gomes, da Silva, and Ortega 2013a), and beta Weibull (BW) distributions as well. The pdf of the BEW distribution is given by

$$g(x) = \frac{\alpha k \lambda^k}{B(a,b)} x^{k-1} e^{(\lambda x)^k} (1 - e^{(\lambda x)^k})^{a\alpha - 1} [1 - (1 - e^{(\lambda x)^k})^{\alpha}]^{b-1}, \quad x > 0. \tag{38}$$

When $\alpha = 1$, we have the BW distribution. We also compared the GWW and GWU distributions with the gamma generalized modified Weibull (GGMW) (Oluyede, Huang, and Yang (2015)) and gamma modified Weibull (GMW) distributions, respectively. The pdf of the GGMW distribution (Oluyede *et al.* 2015) is given by

$$g_{GGMW}(x) \quad = \quad \frac{1}{\Gamma(\delta)} [-\log(1 - e^{-\alpha x - \beta x^{\theta} e^{\lambda x}})]^{\delta - 1} (\alpha + \beta x^{\theta - 1} e^{\lambda x} [\theta + \lambda x]) e^{-\alpha x - \beta x^{\theta} e^{\lambda x}}. \tag{39}$$

When $\alpha = 0$, we have the GMW distribution.

We can use the likelihood ratio (LR) test to compare the fit of the GWU and GWW distribution with their sub-models for a given data set. For example, to test $\lambda = \delta = 1$, for the GWW distribution,the LR statistic is $\omega = 2[\ln(L(\hat{\alpha}, \hat{\beta}, \hat{k}, \hat{\lambda}, \hat{\delta})) - \ln(L(\tilde{\alpha}, \tilde{\beta}, \tilde{k}, 1, 1))]$, where $\hat{\alpha}$, $\hat{\beta}$, $\hat{\lambda}$, $\hat{k}$ and $\hat{\delta}$, are the unrestricted estimates, and $\tilde{\alpha}$, $\tilde{\beta}$, and $\tilde{k}$ are the restricted estimates. The LR test rejects the null hypothesis if $\omega > \chi^2_{\epsilon}$, where χ^2_{ϵ} denote the upper $100\epsilon\%$ point of the χ^2 distribution with 2 degrees of freedom.

We also computed a measure of closeness of each plot to the diagonal line. This measure of closeness (Chambers, Cleveland, Kleiner, and Tukey 1983) is given by the sum of squares (SS)

$$SS = \sum_{j=1}^{n} \left[G_{GWG}(x_{(j)}; \hat{\alpha}, \hat{\beta}, \hat{\delta}, \underline{\hat{\theta}}) - \left(\frac{j - 0.375}{n + 0.25} \right) \right]^2,$$

which is also listed in Table 9 and Table 10.

The 95% confidence intervals for the GWU model (Aarset Data) parameters are: $\alpha \in (44.032, 44.06)$, $\beta \in (0.034022, 0.037538)$, $\delta \in (45.1159, 45.6239)$, $\theta \in (85.972, 86.029)$.

The 95% confidence intervals for the GWW model (Meeker and Escobar Data) parameters are: $\alpha \in (0.0003, 0.0009)$, $\beta \in (0.4986, 0.5874)$, $\delta \in (0.0275, 0.2207)$, $k \in (7.2786, 8.1472)$, $\lambda \in (199.0110, 235.4490)$.

Table 10: GWW estimation for Meeker and Escobar data

Model	α	β	δ	k	λ	-2 Log Likelihood	AIC	BIC	AICC	KS	SS
EE	1	1	1	1	274.87	362.1	364.1	365.6	364.2	0.2294	0.3300
	-	-	-	-	(31.2195)						
EP	1	1	1	1.1028	274.71	361.8	365.8	368.6	366.2	0.2236	0.3206
	-	-	-	(0.1902)	(28.2535)						
Chen	0.005091	1	1	0.3125	1	362.0	366.0	368.8	366.5	0.2059	0.2773
	(0.003423)	-	-	(0.02054)	-						
Gompertz	0.2496	1	1	1	135.09	359.0	363.0	365.8	363.4	0.1889	0.2966
	(0.1795)	-	-	-	(40.6757)						
GEP	1	1	0.1359	6.9362	341.45	350.9	356.9	361.1	357.8	0.2422	0.3943
	-	-	(0.02541)	(0.1927)	(20.2945)						
EEP	1	0.1427	1	7.5321	243.50	341.8	347.8	352.0	348.7	0.2145	0.3033
	-	-	(6.8530)	(0.4399)	(1.4851)						
GEEP	1	13.1977	0.01035	6.9053	324.38	339.8	347.8	353.4	349.4	0.1616	0.1953
	-	(1.6059)	(0.008449)	(0.1905)	(6.5791)						
GWW	0.000610	0.5430	0.1241	7.7129	217.23	326.1	336.1	343.1	338.6	0.1345	0.1404
	(0.000158)	(0.02264)	(0.04927)	(0.2216)	(9.2954)						
BEW	k	λ	α	a	b						
	0.9895	8.0706	0.7834	0.9181	0.0426	371.2	381.2	383.7	388.2	0.1614	0.2904
	(0.03881)	(1.1182)	(0.2485)	(0.5873)	(0.00799)						
GGMW	α	β	θ	λ	δ						
	0.05354	0.004011	0.004549	0.02772	0.06625	345.2	355.2	357.7	362.2	0.1481	0.1885
	(0.00392)	(0.002771)	(0.0893)	(0.002629)	(0.01203)						

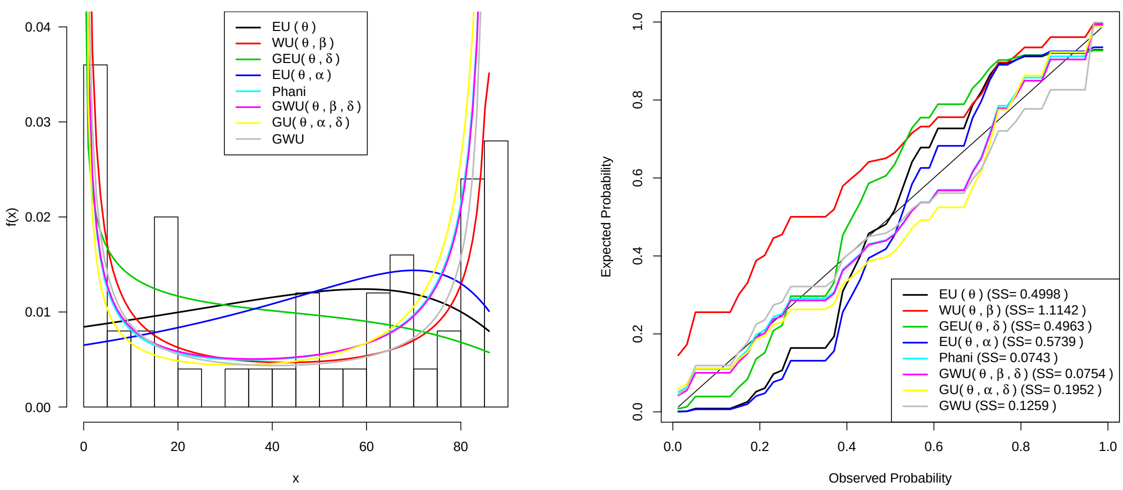

Figure 7: Fitted pdf and observed probabilities of the GWU distribution for Aarset data

Plots of the fitted densities, the histogram of these data and probability plots (Chambers *et al.* 1983) are presented in Figure 7 and Figure 8. For the probability plot, we plotted $G_{GWG}(x_{(j)}; \hat{\alpha}, \hat{\beta}, \hat{\delta}, \underline{\hat{\theta}})$ against $\dfrac{j-0.375}{n+0.25}$, $j = 1, 2, \cdots, n$, where $x_{(j)}$ are the ordered values of the observed data.

For the Aarset data, the LR test statistic of the hypothesis H_0: Phani against H_a: GWU and H_0: GEU(θ, α, δ) against H_a: GWU are $\omega_1 = 12.00$ with p-value$= 5.32 \times 10^{-4}$ and $\omega_2 = 17.5$ with p-value$= 2.87 \times 10^{-5}$. Therefore, we reject H_0 in favor of H_a and conclude that the GWU distribution is significant better than the WU and GEPU distributions. Moreover, the values of the statistics AIC, AICC and BIC show that model GWU is a "better" fit for this data. The value of the goodness-of-fit statistic KS is smallest for the GWU distribution. Also, the value of SS for GWU given by the probability plot is smallest for this model. Consequently, the GWU distribution is a "better"fit when compared to the nested distributions and non-nested four parameter GMW and BW distributions.

Similarly, we can also conduct the LR test for the Meeker and Escobar data for the hypothesis H_0: GEP against H_a: GWW and H_0: GEEP against H_a: GWW. The LR test statistic of these hypothesis are $\omega_3 = 24.8$ with p-value$= 4.12 \times 10^{-6}$ and $\omega_4 = 13.7$ with p-value$= 2.14 \times 10^{-4}$, which implies that we should reject H_0 in favor of H_a and conclude that the GWW distribution is a significant better fit for the Meeker and Escobar data. In addition, the values of the statistics AIC, AICC and BIC clearly show that model GWW is a "better"fit for this data.

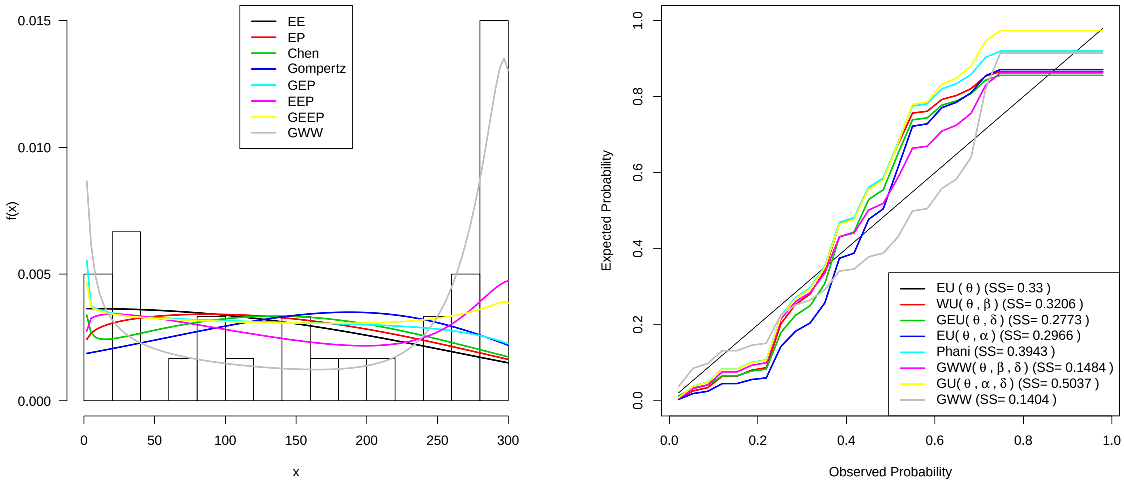

Figure 8: Fitted pdf and observed probabilities of the GWW distribution for Meeker and Escobar data

Furthermore, the value of the statistic KS as well as the value of SS from the probability plot for GWW distribution are the smallest. Clearly the GWW distribution is by far a “better”fit when compared to the nested distributions and the non-nested GGMW and BEW distributions.

9. Concluding remarks

In this paper, the gamma-Weibull-G family of distributions was introduced. This paper also contains the statistical properties of the GWG family such as expansion of density function, hazard and reverse hazard functions, quantile function, moments, incomplete moments, generating functions, mean deviations, Bonferroni and Lorenz curves, order statistics and maximum likelihood estimation for parameters as well as the observed information matrix of the GWG family of distributions. We also discussed two special cases of the GWG family of distributions, namely the gamma-Weibull-uniform and gamma-Weibull-Weibull distributions in detail along with applications of these two special cases to real life data.

Acknowledgements
The authors would like to thank the editor and the referees for carefully reading the paper and for their valuable comments, which greatly improved the presentation in this paper.

References

Aarset M (1987). “How to Identify Bathtub Hazard Rate.” *IEEE Transactions on Reliability*, **36**(1), 106–108.

Alzaatreh A, Famoye F, Lee C (2014). “The Gamma-Normal Distribution: Properties and Applications.” *Computational Statistics and Data Analysis*, **69**, 67–80.

Alzaatreh A, Ghosh I (2014). “On the Weibull-X Family of Distributions.” *Journal of Statistical Theory and Applications*, **14**(2), 169–183.

Bourguignon M, Silva R, Cordeiro G (2014). “The Weibull-G Family of Probability Distributions.” *Journal of Data Science*, (12), 53–68.

Chambers JM, Cleveland WS, Kleiner B, Tukey P (1983). *Graphical Methods for Data Analysis.* Wadsworth.

Chen Z (2000). "A New Two Parameter Lifetime Distribution with Bathtub Shape of Increasing Failure Rate Function." *Statistics and Probability Letters*, **49**, 155–161.

Cordeiro G, Gomes A, da Silva C, Ortega M (2013a). "The Beta Exponentiated Weibull Distribution." *Journal of Statistical Computation and Simulations*, **38**(1), 114–138.

Cordeiro GM, Ortega E, Cunha D (2013b). "The Exponentiated Generalized Class of Distributions." *Journal of Data Science*, **11**(11), 1–27.

Gompertz B (1895). "On The Nature of the Function Expressive of The Law of Human Mortality and The New Model of Determining The Value of Life Contingencies." *Philosophical Transactions of the Royal Society of London*, **115**, 513–585.

Gradshteyn IS, Ryzhik IM (2000). *Tables of Integrals, Series and Products.* Academic Press, San Diego.

Meeker W, Escobar L (1998). *Statistical Methods for Reliability Data.* John Wiley.

Nadarajah S, Cordeiro G, Ortega EMM (2015). "The Zografos and Balakhrishnan-G Family of Distributions: Mathematical Properties and Applications." *Communications in Statistics-Theory and Methods*, **44**, 186–215.

Nascimento ADC, Bourguignon M, Zea LM, Santos-Neto M, Silva RB, Cordeiro GM (2014). "The Gamma Extended Weibull Family of Distributions." *Journal of Statistical Theory and Applications*, **13**, 1–16.

Nassar MM, Eissa F (2003). "On the Exponentiated Weibull Distribution." *Communications in Statistics-Theory and Methods*, **32**, 1317–1336.

Oluyede B, Huang S, Yang T (2015). "A New Class of Generalized Modified Weibull Distribution with Applications." *Austrian Journal of Statistics*, **44**(3), 45–68.

Pal M, Ali MM, Woo J (2006). "Exponentiated Weibull Distribution." *Statistica, anno LXVI*, **2**.

Phani KK (1987). "A New Modified Weibull Distribution Function." *Journal of American Ceramic Society*, **70**, 182–184.

Renyi A (1961). "On Measure of Entropy and Information." *Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 547–561.

Shaked M, Shanthikumar J (1994). *Stochastic Orders and Their Applications.* Academic Press.

Smith RM, Bain LJ (1975). "An Exponential Power Life-testing Distribution." *Communications in Statistics-Theory and Methods*, **4**, 469–481.

Tahir MH, Zubair M, Mansoor M, Cordeiro GM, Alizadeh M, Hamedani GG (2015). "A New Weibull-G Family of Distributions." *Hacettepe University Bulletin of Natural Sciences and Engineering, Series B: Mathematics and Statistics*, **45**(57).

Torabi H, Montazeri HN (2012). "The Gamma-uniform Distribution and It's Applications." *Kybernetika*, **48**, 16–30.

Zografos K, Balakrishnan N (2009). "On Families of Beta- and Generalized Gamma-generated Distribution and Associated Inference." *Stat. Method*, **6**, 344–362.

Affiliation:

Broderick O. Oluyede
Department of Mathematical Sciences
Georgia Southern University (GSU)
Statesboro, GA 30460
E-mail: boluyede@georgiasouthern.edu

Shusen Pu
Department of Mathematics and Statistics
Case Western Reserve University (CWRU)
Cleveland, OH

Boikanyo Makubate
Department of Mathematics and Statistical Sciences
Botswana International University of Science and Technology (BIUST)
Palapye, BW

Yuqi Qiu
Department of Biostatistics
University of California San Diego (UCSD)
San Diego, CA

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 47 *Submitted:* 2015-12-09
February 2018 *Accepted:* 2017-02-07

Austrian Journal of Statistics
February 2018, Volume 47, 77–94.
http://www.ajs.or.at/
doi:10.17713/ajs.v47i1.578

Bayesian Estimation for Inverse Weibull Distribution under Progressive Type-II Censored Data with Beta-Binomial Removals

Pradeep K. Vishwakarma, Arun Kaushik*, Aakriti Pandey
Umesh Singh, Sanjay K. Singh
Department of Statistics, Institute of Science,
Banaras Hindu University, India-221 005

Abstract

This paper deals with the estimation procedure for inverse Weibull distribution under progressive type-II censored samples when removals follow Beta-binomial probability law. To estimate the unknown parameters, the maximum likelihood and Bayes estimators are obtained under progressive censoring scheme mentioned above. Bayes estimates are obtained using Markov chain Monte Carlo (MCMC) technique considering square error loss function and compared with the corresponding MLE's. Further, the expected total time on test is obtained under considered censoring scheme. Finally, a real data set has been analysed to check the validity of the study.

Keywords: inverse Weibull distribution, progressive type-II censoring, beta-binomial removals, maximum likelihood estimates Bayes estimates, MCMC technique.

1. Introduction

In the modern lifetime scenario, it is often difficult or time taking to observe all the units put on experiment due to some controlled or uncontrolled reasons, eg. time and cost constraints, accidental damage, disaster etc. The observations come from this type of situation are called censored observation. The type-I and type-II are the two very common censoring schemes which is widely used in the fields of survival and reliability studies. In type-I censoring schemes, the experimental time is fixed, say (T_0) but the number of observed failure is a random variable while in type-II censoring schemes, number of observed failure is fixed, say (m) but the experimental time is a random. Unfortunately, none of these censoring schemes have discussed the importance of removals of the live units from the test at any time before the completion of the experiment. Further, it is also possible that some units are intentionally or unintentionally removed from the experiment while they are still alive, the data arise from such type of phenomena call it as censored data. In survival/ reliability studies, we usually deals with these censored data. Therefore, there is a need of more appropriate and flexible sampling procedure for life-testing experiments. To attain this, a new censoring schemes is introduced besides the above two schemes, namely progressively type-I and type-II censoring schemes which facilitates the removals of the units during the experiment.

Here, in this paper, we emphasize on estimation procedure under progressive type-II censoring scheme, which is developed by Cohen (1963). The detailed description of the considered scheme are as follows: Suppose, we have n experimental units are put on test at time (T_0) and going to observe m failure units/items during the experiment. The experiment proceeds in such a way that when first failure x_1 observed, R_1 of the surviving units are randomly selected from remaining $(n-1)$ surviving units and then removed i.e, we get R_1 removals from the experiment and immediate after the second failure x_2 is obtained, again R_2 of the surviving units are randomly selected from remaining $(n-R_1-2)$ surviving units and removed i.e, R_2 removals obtained. This procedures continues until the m^{th} failures. Then, at this instance, the experiment terminates and remaining $R_m = n - R_1 - R_2 - R_3 - \cdots - R_{m-1} - m$ surviving units are randomly removed from the experiment. If theses removals $R_1 = R_2 = R_3 = \ldots R_{m-1} = R_m = 0$, then $m = n$, which correspond to complete sample situation and if $R_1 = R_2 = R_3 = \cdots = R_{m-1} = 0$, then $R_m = n - m$, which is simply conventional type-II censoring. Thus the progressive type-II censoring scheme is the generalization of type-II censoring schemes. This type of schemes generally obtained in medical/engineering fields. For example, Consider a medical experiment with n cancer patients but after the death of the first patient, some patients leave the experiment and go for treatment to other medical institution. Similarly, after the second death a few more leave and so on. Finally, the doctor stops taking observation as soon as the predetermined number of deaths (say, m) are recorded.

Statistical inferences based on estimation of parameters for different lifetime models under progressive type-II censoring scheme have been studied by Cohen (1963), Childs and Balakrishnan (2000), Balakrishnan and Sandhu (1995) and cited authors therein. Balakrishnan and Aggarwala (2000) is recommended to the readers for more detail. It may be noted here that, in this censoring scheme, the number of removals $R_1, R_2, R_3, \ldots$, at each stage are pre-fixed. However, in some practical situations, these removals may occur at random e.g. in the previous example, the number of patients leaves the hospital at each stage is random and can not be pre-determined. Utilizing this concept, Tse, Yang, and Yuen (2000), Wu and Chang (2003) and Yuen and Tse (1996) have considered that the number of units removed at each stage follows some specific distribution with certain probability for progressively censored samples. After that, several papers have been published on the estimation of the model parameters for various lifetime distributions under this procedure, see Singh, Singh, and Sharma (2014); Kaushik, Singh, and Singh (2017); Singh, Singh, and Kumar (2013b) and references cited therein.

It is assumed that the probability of removals remains same for all surviving units as well as it remains same at all stages in the case of progressive type-II censored sample with binomial removals. But this assumption seems to be too restrictive and unrealistic to be true in practical situations. For example, in the case of clinical study, if the deaths are recorded in the early stages of the test then definitely the probability of a removals will be high in the beginning and may decrease as the time passes. On other hand, if all the patients in the study are surviving for a longer period i.e. even the first death takes place after a long time, the chance of a drop-out of patients will be relatively small in the beginning and may increase at later stage. This confirms that, the probability of a drop-out at each stage of the experiment may not be remain constant through out the entire experiment. However, this drop out rate can not be observed although it effects the number of drop outs. Hence, one must think that the number of removals is random in nature and it follows a binomial distribution at each stage of removal with random probability of removal following some probability distribution. Due to flexible nature of the beta distribution, capable of having wide range of shapes, we have modelled the uncertainty about in the probability of a removal at various stages of the experiment as random realization of beta variables. Compounding the distribution of number of removals with the probability of removals, results into the distribution of R_i to be beta-binomial and thus, the scheme named as progressive type-II censoring scheme with beta-binomial removals, which is abbreviated as *PT-II CBBR*. In the present piece of work, we have considered that the lifetime of the experimental units follow inverse Weibull (IW) distribution. The probability density function of inverse Weibull distribution is given as,

$$f(x) = \alpha\lambda x^{-\alpha-1}e^{-\lambda x^{-\alpha}}, \quad x > 0, \alpha, \lambda > 0. \tag{1}$$

The corresponding cdf and hazard function are given by

$$F(x) = e^{-\lambda x^{-\alpha}}, \quad x > 0, \alpha, \lambda > 0 \tag{2}$$

and

$$h(t) = \frac{\alpha\lambda t^{-\alpha-1}}{e^{\lambda t^{-\alpha}} - 1}, \tag{3}$$

respectively, where $\alpha > 0$ and $\lambda > 0$ are the shape and scale parameters, respectively. The plots of pdf and hazard function for different values of shape parameters are given in Figure 1. As we can see that it is heavy tail distribution and as $\alpha \to \infty$, the tail probability decreases. For $0 < \alpha \leq 1$, the mean does not exist and for $1 < \alpha \leq 2$, the mean exists but the variance does not exist.

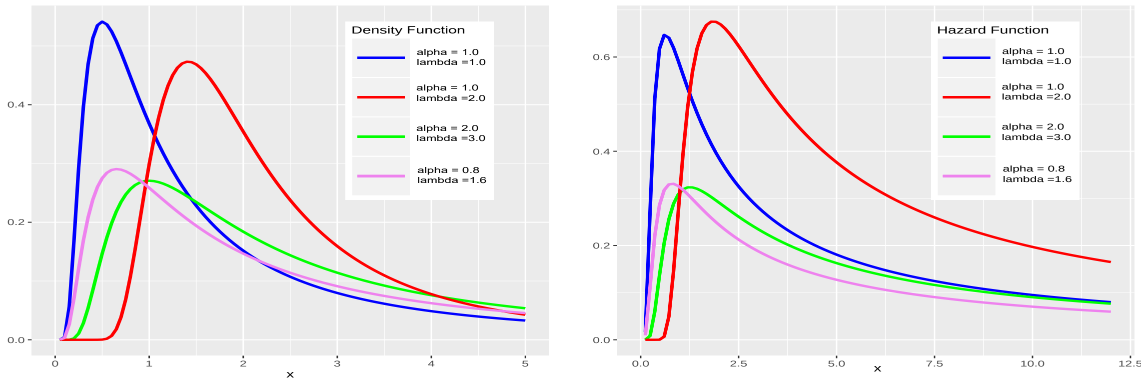

Figure 1: Probability density function and hazard rate for different values of shape and scale parameters

The inverse Weibull distribution is more useful in those situations where data indicates the non-monotone hazard rate characteristics. There are many real life examples where data don't shows the monotone hazard rate. For example, Langlands, Pocock, Kerr, and Gore (1997) have studied breast cancer data and observed that the mortality increases initially, reaches to a peak after some time and then declines slowly i.e., associated hazard rate is modified bathtub or particularly uni-modal. Such types of data can be modelled through the inverse Weibull (IW) distribution. Also, Nelson (1982) showed that, this distribution play an important role in many applications including the dynamic components of diesel engines the time to the breakdown of an insulating fluid subject to the action of a constant tension. Calabria and Pulcini (1990) provide an interpretation of inverse Weibull distribution in the context of the load-strength relationship for a component. Recently, Maswadah (2003) has fitted the inverse Weibull distribution to the flood data reported by Dumonceaux and Antle (1973)

In survival studies, the inverse Weibull distribution has been considered by several authors. Khan, Pasha, and Pasha (2008) have discussed the classical statistical properties of IW distribution. Kim, Lee, and Kang (2014) have derived the non-informative prior for the parameters of IW distribution. Noor and Aslam (2013) have proposed the Bayes estimators of the parameters of inverse Weibull mixture distribution using type-I censored samples. Recently, Singh, Singh, and Kumar (2013a) have discussed the classical as well as Bayesian estimation procedures for unknown parameters of inverse Weibull distribution under conventional type-I and type-II censoring schemes. Thus, there is a need to developed a estimation procedure (classical as well as Bayesian) for IW distribution under more realistic and advanced censoring scheme such as progressive censoring schemes, where removals have also some probability distributions.

The main objective of this chapter is to provide the maximum likelihood estimates and Bayes estimates of unknown parameters under progressive type-II censoring scheme, where removals follow the beta-binomial probability law. The Bayes estimates are obtained using informative and non-informative prior under squared error loss function. It is noted here that the estimates obtained are not in explicit forms and they can be analysed by some suitable numerical integration technique. Therefore, we use Newton-Raphson method to find MLEs. MCMC technique has been used to solve the integration involve in posterior distribution. Also, we compare the MLEs with corresponding Bayes estimates of the unknown parameters by Monte-Carlo simulations.

The rest of the chapter is organized as follows: the maximum likelihood estimators (MLEs) of the parameter are obtained under the progressive type-II censored data with beta-binomial removals in section 2. In section 3, we have obtained Bayes estimators for unknown parameters of the IW distribution under progressive type-II censoring scheme with beta-binomial removals. The risk of estimates has been obtained. The comparison of MLE and correspond Bayes estimator under squared error loss function in term of their risks have been studied in section 4. A real data study to illustrate the application of the results in section 5. Finally, conclusions are presented in section 6.

2. The likelihood function

The likelihood function under progressive type-II censoring with pre-determined number of removals $\mathbf{R} = (R_1 = r_1, R_2 = r_2, \cdots, R_m = r_m)$ is given by

$$L(\underset{\sim}{\boldsymbol{x}}|\alpha, \lambda) = C \prod_{i=1}^{m} f(x_i, \alpha) \{1 - F(x_i, \alpha, \lambda)\}^{r_i}, \tag{4}$$

where, $C = n(n - m - r_1)...(n - m - \sum_{i=1}^{m} r_i + 1)$. Substituting (1) and (2) into (4), we get

$$L(\alpha, \lambda; \underset{\sim}{\boldsymbol{x}}|R = r) = C\alpha^m \lambda^m e^{-\lambda \sum_{i=1}^{m} x_i^{-\alpha}} \prod_{i=1}^{m} \left[x_i^{-\alpha-1} \left(1 - e^{-\lambda x_i^{-\alpha}}\right)^{r_i} \right]. \tag{5}$$

It has been discussed above that in the progressive censoring scheme if the experimental units are removed from the test with given probability, say p, then following Tse *et al.* (2000), the number of removals at i^{th} stage is given by,

$$Pr(R_i = r_i|p) = \binom{n - m - \sum_{j=1}^{i-1} r_j}{r_i} p^{r_i} (1 - p)^{n-m-\sum_{j=1}^{i} r_j}; i = 1, 2, ..., m - 1 \tag{6}$$

Further, it is assumed in the considered form of the censoring scheme that the probability of removals is not fixed over whole of the experimentation period and assumed to be a random variable following the probability density function

$$g(p|\xi, \zeta) = \frac{1}{B(\xi, \zeta)} p^{\xi-1} (1 - p)^{\zeta-1}; \xi > 0, \zeta > 0, 0 < p < 1. \tag{7}$$

Thus, the unconditional distribution of $R_i's$ can be derived as follows

$$Pr(R_i = r_i, \xi, \zeta) = \int_0^1 Pr(R_i = r_i|p)g(p|\xi, \zeta)dp \tag{8}$$

Substituting (6) and (7) in (8), we get,

$$Pr(R_i = r_i, \xi, \zeta) = \frac{1}{B(\xi,\zeta)} \binom{n-m-\sum\limits_{j=1}^{i-1} r_j}{r_i} \int_0^1 p^{r_i+\xi-1}(1-p)^{n-m-\sum\limits_{j=1}^{i} r_j+\zeta-1} dp$$

After simplification, which can be written as

$$Pr(R_i = r_i, \xi, \zeta) = \binom{n-m-\sum\limits_{j=1}^{i-1} r_j}{r_i} \frac{B\left(\xi + r_i,\ \zeta + n - m - \sum\limits_{j=1}^{i} r_j\right)}{B(\xi,\zeta)} \tag{9}$$

where, $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \xi, \zeta > 0, r_i = 0, ..., n-m-\sum\limits_{j=1}^{i-1} r_j; i = 1, ..., (m-1).$

It is the probability mass function (pmf) of beta-binomial distribution and it is denoted by $BB(n', \xi, \zeta)$.

Thus, the joint probability of $R_1 = r_1, R_2 = r_2, ..., R_m = r_m$ is given by

$$\begin{aligned} P(\mathbf{R}{=}\mathbf{r}, \xi, \zeta) = P[R_1 = r_1] \times P[R_2 = r_2|R_1 = r_1] \times \cdots \times \\ P[R_{m-1} = r_{m-1}|R_{m-2} = r_{m-2}, ..., R_1 = r_1] \end{aligned} \tag{10}$$

Now, we further assume that $R_i's$ is independent of $x_i's$ for all i. Then, the full likelihood function takes the following form

$$\begin{aligned} L(\underset{\sim}{x}, \Theta, \mathbf{R}, \xi, \zeta) &= L(\underset{\sim}{x}, \Theta|\mathbf{R})P(\mathbf{R}{=}\mathbf{r}, \xi, \zeta) \\ &= C\alpha^m \lambda^m e^{-\lambda \sum_{i=1}^m x_i^{-\alpha}} \prod_{i=1}^m \left[x_i^{-\alpha-1}\left(1 - e^{-\lambda x_i^{-\alpha}}\right)^{r_i}\right] \\ &\prod_{i=1}^{m-1} \binom{n-m-\sum\limits_{j=1}^{i-1} r_j}{r_i} \frac{B\left(\xi + r_i,\ \zeta + n - m - \sum\limits_{j=1}^{i} r_j\right)}{B(\xi,\zeta)}. \end{aligned} \tag{11}$$

Maximum likelihood estimation for α and λ

The likelihood function for progressively type-II censored sample with beta-binomial removals from (11) can be re-written as,

$$L(\underset{\sim}{\boldsymbol{x}}, \alpha, \lambda, R, \xi, \zeta) \propto L_1(\alpha, \lambda|R)L_2(\xi, \zeta|R), \tag{12}$$

where

$$L_1(\alpha, \lambda|R) = \alpha^m \lambda^m e^{-\lambda \sum_{i=1}^m x_i^{-\alpha}} \prod_{i=1}^m \left[x_i^{-\alpha-1}\left(1 - e^{-\lambda x_i^{-\alpha}}\right)^{r_i}\right] \tag{13}$$

and

$$L_2(\xi, \zeta|R) = \prod_{i=1}^{m-1} \binom{n-m-\sum\limits_{j=1}^{i-1} r_j}{r_i} \frac{B\left(\xi + r_i,\ \zeta + n - m - \sum\limits_{j=1}^{i} r_j\right)}{B(\xi,\zeta)}, \tag{14}$$

It may be noted here that $L_1(\alpha, \lambda|R)$ involves the parameters of the considered model but it is independent of the parameters of the distributions considered for removal probability. On the other hand $L_2(\xi, \zeta|R)$ is free from the parameters of the considered model and includes only the parameters of the distribution of the probability of removals. Therefore for finding the MLE of α and λ, we must use $L_1(\alpha, \lambda|R)$ only. On maximizing the likelihood function given by $L_1(\alpha, \lambda|R)$, we obtain the maximum likelihood estimates for model parameters α and λ. Hence, the corresponding log-likelihood function of (13) is given by,

$$\ln L_1 = \ln C + m\ln\alpha + m\ln\lambda - \lambda\sum_{i=1}^{m} x_i^{-\alpha} - (\alpha+1)\sum_{i=1}^{m}\ln x_i + \sum_{i=1}^{m} r_i \ln(1 - e^{-\lambda x_i^{-\lambda}}) \tag{15}$$

Then, the MLEs $\hat{\alpha}$ and $\hat{\lambda}$ of α and λ, respectively can be obtained as the simultaneous solution of the following two non-linear equations:

$$\frac{m}{\alpha} - \lambda\sum_{i=1}^{m}(x_i^{-\alpha}\ln x_i) - \sum_{i=1}^{m}\ln x_i + \lambda\sum_{i=1}^{m}\left[\frac{r_i x_i^{-\alpha} e^{-\lambda x_i^{-\alpha}}\ln x_i}{(1 - e^{-\lambda x_i^{-\alpha}})}\right] \tag{16}$$

and

$$\frac{m}{\lambda} - \sum_{i=1}^{m} x_i^{-\alpha} + \sum_{i=1}^{m}\left[\frac{r_i x_i^{-\alpha} e^{-\lambda x_i^{-\alpha}}}{(1 - e^{-\lambda x_i^{-\alpha}})}\right]. \tag{17}$$

As the above equations cannot be evaluated analytically, one can use numerical technique such as Newton-Raphson method to solve these and find the MLEs. While attempting to obtain interval estimates, we note that, the exact distribution of MLEs are not easy to obtain, we suggest to use the concept of large sample theory to obtain the confidence intervals based on $I(\alpha, \lambda)$, the Fisher‘s information matrix, which can be estimated by

$$I\left(\hat{\alpha}, \hat{\lambda}\right) = \begin{bmatrix} -\frac{d\ln L}{d\alpha^2} & -\frac{d\ln L}{d\lambda d\alpha} \\ -\frac{d\ln L}{d\alpha d\lambda} & -\frac{d\ln L}{d\lambda^2} \end{bmatrix}_{(\hat{\alpha}, \hat{\lambda})} \tag{18}$$

The diagonal elements of $I^{-1}\left(\hat{\alpha}, \hat{\lambda}\right)$ provides the asymptotic variances for the parameters α and λ, respectively. Thus, two-sided $100(1-\gamma)\%$ normal approximation confidence interval of α and λ can be obtained as $\left\{\hat{\alpha} \mp Z_{\gamma/2}\sqrt{var(\hat{\alpha})}\right\}$ and $\left\{\hat{\lambda} \mp Z_{\gamma/2}\sqrt{var(\hat{\lambda})}\right\}$, respectively.

Maximum likelihood estimation for ξ and ζ:

From equation (14), the joint probability of **R=r** is given by

$$P(\mathbf{R{=}r}, \xi, \zeta) = C^{\star}\frac{1}{B(\xi,\zeta)^{m-1}}\prod_{i=1}^{m-1} B(\xi + r_i,\ n - m - \sum_{j=1}^{i} r_j + \zeta) \tag{19}$$

Where, $C^{\star} = \frac{(n-m)!}{\prod_{i=1}^{m-1} r_i!(n-m-\sum_{j=1}^{m-1} r_j)!}$.

Taking logarithm on both sides of equation (19), we get

$$\begin{aligned} LogP = {} & \ln(C^{\star}) - (m-1)[\ln\Gamma(\xi) + \ln\Gamma(\zeta) - \ln\Gamma(\xi+\zeta)] + \sum_{i=1}^{m-1}\ln\Gamma(\xi + r_i) \\ & + \sum_{i=1}^{m-1}\ln\Gamma(n - m - \sum_{j=1}^{i} r_j + \zeta) - \sum_{i=1}^{m-1}\ln\Gamma(n - m - \sum_{j=1}^{i-1} r_j + \xi + \zeta) \end{aligned} \tag{20}$$

The MLEs $\hat{\xi}$ and $\hat{\zeta}$ of ξ and ζ, respectively can be obtained as the simultaneous solution of the following two normal non-linear equations:

$$-\frac{(m-1)}{\Gamma(\xi)}D_\xi(\xi)+\frac{(m-1)}{\Gamma(\xi+\zeta)}D_\xi(\xi+\zeta)+\sum_{i=1}^{m-1}\frac{D_\xi(\xi+r_i)}{\Gamma(\xi+r_i)}-\sum_{i=1}^{m-1}\frac{D_\xi(n-m-\sum\limits_{j=1}^{i-1}r_j+\xi+\zeta)}{\Gamma(n-m-\sum\limits_{j=1}^{i-1}r_j+\xi+\zeta)}=0 \tag{21}$$

$$\begin{aligned}-\frac{(m-1)}{\Gamma(\zeta)}D_\zeta(\zeta)+\frac{(m-1)}{\Gamma(\xi+\zeta)}D_\zeta(\xi+\zeta)+\sum_{i=1}^{m-1}\frac{D_\zeta(n-m-\sum\limits_{j=1}^{i}r_j+\zeta)}{\Gamma(n-m-\sum\limits_{j=1}^{i}r_j+\zeta)}\\ -\sum_{i=1}^{m-1}\frac{D_\zeta(n-m-\sum\limits_{j=1}^{i-1}r_j+\xi+\zeta)}{\Gamma(n-m-\sum\limits_{j=1}^{i-1}r_j+\xi+\zeta)}=0\end{aligned} \tag{22}$$

Where $D_a(\phi(a))=\frac{d}{da}\Gamma(\phi(a))$ is dia-gamma function. The equation's are not solvable in nice closed form therefore, we suggest the use of iterative methods.

3. Bayes estimation

In the Bayesian paradigm, we need to assume the prior distributions for unknown model parameters. The prior probability density for the parameters α and λ are assumed to be of the following forms

$$g_1(\alpha)=\frac{\nu_1^{\mu_1}}{\Gamma\mu_1}e^{-\nu_1\alpha}\alpha^{\mu_1-1};\ \alpha>0,\ \mu_1>0,\ \nu_1\ >0 \tag{23}$$

$$g_2(\lambda)=\frac{\lambda_2^{\mu_2}}{\Gamma\mu_2}e^{-\nu_2\lambda}\lambda^{\mu_2-1};\ \lambda>0,\ \mu_2>0,\ \nu_2\ >0, \tag{24}$$

where, $\mu_1,\ \nu_1,\ \mu_2$ and ν_2 are the hyper-parameters. The joint prior pdf of α and λ may be obtain as,

$$g(\alpha,\lambda)=g_1(\alpha)\,g_2(\lambda)\quad;\quad\alpha>0,\quad\lambda>0 \tag{25}$$

Thus, the joint posterior of α and λ is given by

$$\pi(\alpha,\lambda|\underset{\sim}{\boldsymbol{x}})=\frac{J}{\iint_0^\infty Jd\alpha d\lambda} \tag{26}$$

where,

$$J=J(\alpha,\lambda)\ =\ \alpha^{m+\mu_1-1}\lambda^{m+\mu_2-1}e^{-\left(\nu_1\alpha+\nu_2\lambda+\lambda\sum_{i=1}^m x_i^{-\alpha}\right)}\prod_{i=1}^m\left[x_i^{-\alpha-1}\left(1-e^{-\lambda x_i^{-\alpha}}\right)^{r_i}\right]$$

Let $h(\cdot)$ be a function of α and λ. Then, the Bayes estimator of $h(\cdot)$ under the squared error loss function is given by,

$$\begin{aligned}\hat{h}_B(\alpha,\lambda)&=E_\pi(h(\alpha,\lambda))\\&=\frac{\iint_0^\infty h(\alpha,\lambda)\,J\,d\alpha\,d\lambda}{\iint_0^\infty J\,d\alpha\,d\lambda}\end{aligned} \tag{27}$$

It is clear from the expression (26) that there is no closed form for the estimators, so we suggest using an MCMC procedure to compute the Bayes estimates, see (Smith and Roberts 1993; Brooks 1998; Hastings 1970) for more detail. After getting MCMC samples from the posterior distribution, we can find the Bayes estimate for the parameters in the following way

$$[E(\Theta|data)] = \left[\frac{1}{N-N_0}\sum_{i=N_0+1}^{N}\Theta_i\right],$$

where N_0 is burn-in period of the Markov chain and $\Theta_i = [\alpha_i, \lambda_i]'$. For computation of the highest posterior density (HPD) interval of Θ, order the MCMC sample of Θ as $\Theta_{(1)}$, $\Theta_{(2)}$, $\Theta_{(3)}, \cdots, \Theta_{(N)}$. Then construct all the 100(1-γ)% credible intervals of Θ say $(\Theta_{(1)}, \Theta_{(N\lfloor 1-\gamma\rfloor+1)})$, $(\Theta_{(2)}, \Theta_{(N\lfloor 1-\gamma\rfloor+2)}) \cdots$,$(\Theta_{(\lfloor N\gamma\rfloor)}, \Theta_{(N)})$. Finally, the HPD credible interval of α and β is that interval which has the shortest length.

In order to obtain the MCMC samples from the joint posterior density of α and λ, we use the Metropolis-Hastings (M-H) algorithm. We consider a bivariate normal distribution as the proposal density i.e. $N_2(\mu, \Sigma)$ where Σ is the variance-covariance matrix. It may be noted here that if we generate observations from the bivariate normal distribution, we may get negative values also, which are not possible as the parameters under consideration are positive valued. Therefore, we take the absolute value of the generated observations. Following this, the Metropolis-Hastings algorithm associated with the target density $\pi(\cdot)$ and the proposal density $N_2(\mu, \Sigma)$ produces a Markov chain Θ^i through the following steps.

① Set initial values $\Theta_0 = [\alpha_0, \lambda_0]'$.

② Generate new candidate parameter values $\Theta_* = [\alpha_*, \lambda_*]'$ from $N_2(\mu, \Sigma)$.

③ Calculate the ratio

$$\rho\left(\Theta_*, \Theta_{i-1}\right) = \min\left\{\tfrac{\pi(\Theta_*)}{\pi(\Theta_{i-1})}, 1\right\}.$$

④ Draw u from uniform(0,1).

$$\begin{cases}\text{Accept } \Theta_* \text{ as } \Theta_i \text{ if } u < \rho\left(\Theta_*, \Theta_{i-1}\right). \\ \text{If } \Theta_* \text{ is not accepted, then } \Theta_i = \Theta_{i-1}.\end{cases}$$

In using the above algorithm, the problem arises as to how to choose the initial guess. Here, we propose the use of the MLEs of (α, λ), obtained by using the method described in section 2, as initial values for the MCMC process. The choice of covariance matrix Σ is also an important issue; see Natzoufras (2009) and Kaushik *et al.* (2017) for details. One choice for Σ would be the asymptotic variance-covariance matrix $I^{-1}(\hat{\alpha}, \hat{\lambda})$. While generating M-H samples by taking $\Sigma = I^{-1}(\hat{\alpha}, \hat{\lambda})$, we noted that the acceptance rate for such a choice of Σ is about 15%. By acceptance rate, we mean the proportion of times a new set of values is generated at the iteration stages. It is well known that if the acceptance rate is low, a good strategy is to run a small pilot run using a diagonal Σ as a rough estimate of the correlation structure for the target posterior distribution and then re-run the algorithm using the corresponding estimated variance-covariance matrix; for more details see Gelmen, Carlin, Stern, and Rubin (1995). Therefore, we have also used the latter described strategy for the calculations in the following sections.

4. Expected total time on test

In practice, it is also desired to have an idea of the duration of a life test since the experiment termination time is directly associated with the cost of the experiment. For progressive type-II censoring scheme, the termination time is given by the expectation of the m^{th} order statistics. From Balakrishnan and Aggarwala (2000), the conditional expectation of X_m given $\mathbf{R} = (R_1 = r_1, R_2 = r_2, ..., R_m = r_m)$ can be defined as

$$E\left(X_m \middle| \mathbf{R}\right) = C(r) \sum_{l_1=1}^{r_1} ... \sum_{l_m=1}^{r_m} (-1)^{\sum_{i=1}^{m} l_i} \frac{\binom{r_1}{l_1} ... \binom{r_m}{l_m}}{\prod_{i=1}^{m-1} h(l_i)} \int_0^{\infty} x f(x) F^{h(l_m)-1}(x) dx \tag{28}$$

Where, $h(l_i) = \sum_{j=1}^{i} l_j + i$. Putting (1), (2) in (28) and then simplifying, we have

$$E\left(X_m \middle| \mathbf{R}\right) = C(r)\alpha\lambda \sum_{l_1=1}^{r_1} ... \sum_{l_m=1}^{r_m} (-1)^{\sum_{i=1}^{m} l_i} \frac{\binom{r_1}{l_1} ... \binom{r_m}{l_m}}{\prod_{i=1}^{m-1} h(l_i)} \int_0^{\infty} x^{-\alpha} e^{-\lambda h(l_m) x^{-\alpha}} dx \tag{29}$$

The expected termination time PT-II CBBR is evaluated by taking the expectation on both sides of (29) with respect to the $\mathbf{R}$. That is,

$$\begin{aligned} E(X_m) &= E_R[E\left(X_m \middle| \mathbf{R}\right)] \\ &= \sum_{r_1=0}^{g(r_1)} \sum_{r_2=0}^{g(r_2)} ... \sum_{r_{m-1}=0}^{g(r_{m-1})} P(\mathbf{R}{=}\mathbf{r}, \xi, \zeta) E\left(X_m \middle| \mathbf{R}\right) \end{aligned} \tag{30}$$

Where, $g(r_i) = n - m - r_1 - ... - r_{i-1}$.

5. Simulation study

In this section, we have compared the performances of the various estimators on the basis of their mean square errors (MSEs). It may be mentioned here that the exact expression for the mean square errors cannot be obtained, because the estimators are not in explicit form. Therefore, MSEs are estimated on the basis of a Monte-Carlo simulation study of 2000 samples. For this purpose, we generated a specified number of observations from the distribution given in equation (1) for fixed values of the parameters under the specified censoring schemes and calculated different estimates of α and λ following the procedure described in the previous sections. This process was repeated 2000 times to obtain the simulated biases and MSEs. We have computed the MLEs by using the Newton-Raphson algorithm and Bayes estimates using MCMC method. The ML and Bayes estimates of (α, λ) are denoted as $(\alpha_{ML}, \lambda_{ML})$ and (α_B, λ_B), respectively. It is noted that Newton-Raphson algorithm has a convergence rate of 90%-95%. We have reported the results omitting the cases where algorithm do not converge. To simulate a progressive type-II censored sample from the considered distribution, we have used the algorithm given by Balakrishnan and Cramer (2014).

It may be noted here that the MSE of these estimators will depend on the sample size n, m, values of α, λ and hyper-parameters μ_1, μ_2, ν_1 and ν_2. We considered a number of values for the sample size n; namely n = 10, 20, 30, 40 and 50 and m is taken as 50%, 60%, 70%, 80%, 90% and 100% of the n. For an informative prior, the hyper parameters are chosen on the basis of the information possessed by the experimenter, denoted as Bayes1. Also, we have considered the choice of hyper-parameters as $\mu_1 = \mu_2 = \nu_1 = \nu_2 = 0$ which reduces the prior to a non-informative prior, denoted as Bayes2. In most of the cases, the experimenter can have a notion of what are the expected value of the parameter and can always associate a degree of belief to this value. In other words, the experimenter can specify the prior mean and prior variance for the parameters. The prior mean reflects the experimenter's belief about the parameter in the form of its expected value and the prior variance reflects his confidence in this expected value. Keeping this point in mind, we have chosen the hyper-parameters in such a way that the prior mean is equal to the true value of the parameter and the belief in the prior mean is either strong or weak, i.e. the prior variance is small or large, respectively; for details see Singh *et al.* (2013a). The MSE's of the estimates of parameters with corresponding confidence interval have been calculated and the results are summarized in Tables 1, 2 and (3). From Tables 1-3, we have observed that the both MLEs and Bayes estimates provides more accurate estimates with increasing sample information m although, the Bayes estimates of the model parameters are more nearer to the true values of the model parameters. The HPD intervals have shorter width than the asymptotic confidence intervals.

Table 1 provides the MSE of estimates of the parameters for $\alpha = 2$, $\lambda = 3$, $\xi = 0.1$, $\zeta = 3$ and hyper-parameters $\mu_1 = 4, \mu_2 = 6, \nu_1 = 2, \nu_2 = 3$. It can be seen from the Table that in general, the MSE's decrease as n increases in all the considered cases. It can also be seen that the MSE of the MLE is more than that of the corresponding Bayes estimate in all cases but the difference between the MSEs of the Bayes and ML estimates decreases for increases in the value of n. It is also noted here that MSEs decreases as m increases for fix value of n. Similar trend found in the MSE of the Bayes estimates and Bayes estimates with informative prior having least MSE. The 95% asymptotic interval estimates are wider than HPD intervals. In Table 2, we have shown the effect of variation of ξ and ζ. Here, we noticed that increment in the values of ξ reflect the negative effect on the performance of all considered point and interval estimates, however as ζ increases the estimators performance becomes better. Table 3 provide the effect of magnitude of parameters α and λ on the considered estimators. As α increases or λ increases, in both of the situation the estimated MSEs of all the estimators increases, along with the width of the asymptotic and HPD interval estimators are increases.

For investigating the expected total test time (TTT), we have considered the different combination of the model parameters which are given below:

$$\begin{array}{llll}
\alpha = 2, & \lambda = 3, & \xi(= 0.5, 1.0), & \zeta\{= (15, 9, 3, 1, 0.5, 0.25), (15, 9, 3, 1, 0.5, 0.25)\}; \\
\alpha = 0.8, & \lambda = 1.6, & \zeta(= 0.5, 1.0), & \xi\{= (15, 9, 3, 1, 0.5, 0.25), (15, 9, 3, 1, 0.5, 0.25)\}; \\
\alpha = 2, & \lambda = 3, & \xi(= 1, 3), & \zeta\{= (15, 9, 3, 1, 0.5, 0.25), (15, 9, 3, 1, 0.5, 0.25)\}; \\
\alpha = 0.8, & \lambda = 1.6, & \zeta(= 1, 3), & \xi\{= (15, 9, 3, 1, 0.5, 0.25), (15, 9, 3, 1, 0.5, 0.25)\}.
\end{array}$$

For each combination, we have taken $n = (10, 20, 30, 40$ and $50)$ and for each n, m is chosen so that the sample contains the 100%, 90%, $\cdots$, 50% units of the available sample units, respectively. All results are summarised in Tables 4 - 7. From this, we have observed that the expected TTT is an increasing function of n and m as it is expected. It is interesting to know that the expected TTT decreases as ζ increases while ξ is fixed and on other hand expected TTT increases as ξ increases for given fixed value of ζ. It can also be observed that the expected TTT increases as α increases and decreases as λ decreases.

6. Real data illustration

In this section, we illustrate our proposed methodology with the four real examples. The first data set considered by us, represents the times between successive failures of air conditioning equipment in a *Boeing 720* airplane, reported by Proschan (1963):

75	57	48	29	502	12	70	21	29
386	59	27	153	26	326			

Second data set used by Bhaumik, Kapur, and Gibbons (2009), is vinyl chloride data obtained from clean upgradient monitoring wells in mg/litre:

5.1	1.2	1.3	0.6	0.5	2.4	0.5	1.1	8.0
0.8	0.4	0.6	0.9	0.4	2.0	0.5	5.3	3.2
2.7	2.9	2.5	2.3	1.0	0.2	0.1	0.1	1.8
0.9	2	4	6.8	1.2	0.4	0.2		

Table 1: MLEs and Bayes estimators (MSE in brackets) with corresponding asymptotic and HPD intervals for the paramaters for fixed values of $\alpha = 2$, $\lambda = 3$, $\xi = 0.1$, $\zeta = 3$

n	m	MLEs	Asymptotic CI	Bayes1	HPD CI	Bayes2	HPD CI
20	10	2.0951(0.1175)	(1.8206,2.2811)	1.9836(0.0775)	(1.9036,2.1983)	2.0675(0.1037)	(1.8501,2.2509)
		3.2994(0.6705)	(1.9215,4.5489)	3.1328(0.4308)	(2.4173,4.0567)	3.2707(0.5914)	(2.0943,4.3746)
	12	2.0848(0.1144)	(1.8265,2.2749)	1.9932(0.0744)	(1.9094,2.1930)	2.0778(0.0994)	(1.8582,2.2426)
		3.2889(0.6389)	(1.9833,4.4865)	3.1224(0.4144)	(2.4485,4.0252)	3.2612(0.5559)	(2.1631,4.3060)
	14	2.0812(0.1119)	(1.8312,2.2699)	2.0008(0.0720)	(1.9135,2.1886)	2.0595(0.0993)	(1.8587,2.2430)
		3.2751(0.6153)	(2.0296,4.4408)	3.2048(0.3954)	(2.4850,3.9887)	3.2219(0.5419)	(2.1901,4.2790)
	16	2.0708(0.1121)	(1.8305,2.2704)	2.0102(0.0719)	(1.9137,2.1876)	2.0693(0.0956)	(1.8662,2.2356)
		3.2651(0.6127)	(2.0354,4.4355)	3.1950(0.3909)	(2.4936,3.9802)	3.2118(0.5220)	(2.2286,4.2410)
	18	2.0607(0.1086)	(1.8376,2.2634)	2.0002(0.0682)	(1.9206,2.1814)	2.0594(0.0927)	(1.8718,2.2297)
		3.2544(0.5962)	(2.0669,4.4038)	3.1853(0.3744)	(2.5243,3.9491)	3.2016(0.5056)	(2.2600,4.2089)
	20	2.0511(0.1068)	(1.8413,2.2599)	1.9902(0.0660)	(1.9252,2.1763)	2.0496(0.0908)	(1.8759,2.2261)
		3.2365(0.5746)	(2.1097,4.3606)	3.1670(0.3573)	(2.5569,3.9164)	3.1834(0.4851)	(2.2996,4.1694)
30	15	2.0449(0.1048)	(1.8223,2.2327)	1.9333(0.0645)	(1.9047,2.1502)	2.0177(0.0882)	(1.8576,2.1980)
		3.2279(0.5487)	(2.0261,4.1739)	3.1693(0.3361)	(2.4624,3.7405)	3.0871(0.4616)	(2.2103,3.9909)
	18	2.0383(0.0722)	(1.8859,2.1692)	1.9867(0.0564)	(1.9204,2.1346)	1.9778(0.0675)	(1.8978,2.1577)
		3.1577(0.3589)	(2.3980,3.8031)	3.0992(0.2815)	(2.5665,3.6360)	3.0170(0.3369)	(2.4516,3.7513)
	21	2.0311(0.0703)	(1.8896,2.1650)	1.9866(0.0552)	(1.9223,2.1327)	2.0099(0.0655)	(1.9009,2.1539)
		3.1553(0.3159)	(2.4820,3.7185)	3.0782(0.2449)	(2.6362,3.5664)	3.0716(0.2931)	(2.5354,3.6670)
	24	2.0479(0.0689)	(1.8931,2.1621)	1.9996(0.0524)	(1.9279,2.1268)	1.9861(0.0638)	(1.9045,2.1503)
		3.1520(0.3054)	(2.5027,3.6987)	3.0577(0.1660)	(2.7858,3.4167)	3.1264(0.2480)	(2.6230,3.5796)
	27	2.0378(0.0657)	(1.8986,2.1560)	2.0010(0.0502)	(1.9325,2.1234)	1.9755(0.0599)	(1.9119,2.1429)
		3.1219(0.2948)	(2.5240,3.6780)	3.0438(0.2225)	(2.6781,3.5244)	3.0960(0.2701)	(2.5800,3.6225)
	30	2.0281(0.0557)	(1.9186,2.1361)	2.0009(0.0400)	(1.9518,2.1037)	1.9655(0.0502)	(1.9302,2.1242)
		3.1018(0.2495)	(2.6121,3.5893)	3.0240(0.1768)	(2.7653,3.4372)	3.0764(0.2256)	(2.6664,3.5360)
50	25	2.0460(0.0619)	(1.8894,2.1314)	2.0122(0.0466)	(1.9218,2.0986)	2.0136(0.0577)	(1.8994,2.1216)
		3.1277(0.4023)	(2.2880,3.8648)	3.0696(0.3005)	(2.5047,3.6442)	3.0975(0.3718)	(2.3588,3.7921)
	30	2.0360(0.0604)	(1.8915,2.1280)	2.0016(0.0448)	(1.9248,2.0952)	2.0035(0.0559)	(1.9020,2.1175)
		3.1158(0.3608)	(2.3686,3.7836)	3.0409(0.2657)	(2.5714,3.5785)	3.1153(0.3314)	(2.4362,3.7136)
	35	2.0336(0.0609)	(1.8908,2.1288)	2.0102(0.0447)	(1.9250,2.0948)	2.0168(0.0557)	(1.9033,2.1173)
		3.1127(0.3005)	(2.4868,3.6650)	3.0420(0.2221)	(2.6537,3.4954)	3.1108(0.2756)	(2.5444,3.6067)
	40	2.0306(0.0485)	(1.9150,2.1050)	2.0181(0.0329)	(1.9479,2.0727)	2.0301(0.0434)	(1.9267,2.0936)
		3.0915(0.2328)	(2.6194,3.5321)	3.0407(0.1552)	(2.7809,3.3697)	3.0830(0.2064)	(2.6776,3.4732)
	45	2.0210(0.0442)	(1.9241,2.0960)	2.0078(0.0282)	(1.9564,2.0632)	2.0208(0.0389)	(1.9352,2.0847)
		3.0819(0.2003)	(2.6834,3.4689)	3.0304(0.1261)	(2.8363,3.3148)	3.0727(0.1747)	(2.7390,3.4128)
	50	2.0109(0.0408)	(1.9306,2.0901)	1.9976(0.0247)	(1.9637,2.0572)	2.0105(0.0357)	(1.9420,2.0789)
		3.0765(0.1744)	(2.7337,3.4175)	3.0254(0.1052)	(2.8760,3.2750)	3.0672(0.1508)	(2.7848,3.3663)

Table 2: MLEs and Bayes estimators (MSE in brackets) with corresponding asymptotic and HPD intervals for the paramaters for fixed value of $\alpha = 2$, $\lambda = 3$, $n = 50$, $n = 35$

ξ	ζ	MLE	Asymptotic CI	Bayes1	HPD CI	Bayes2	HPD CI
0.1	0.1	2.0135(0.0607)	(1.9315,2.2814)	2.0339(0.0448)	(1.9657,2.1985)	2.0777(0.0555)	(1.9431,2.2518)
		3.0823(0.2975)	(2.6525,4.5494)	3.0120(0.2199)	(2.8180,4.0571)	3.0804(0.2737)	(2.7078,4.3751)
	0.5	2.0117(0.0606)	(1.9315,2.2752)	2.0320(0.0451)	(1.9645,2.1939)	2.0758(0.0551)	(1.9447,2.2434)
		3.0793(0.2981)	(2.6513,4.4870)	3.0092(0.2205)	(2.8173,4.0254)	3.0769(0.2728)	(2.7092,4.3062)
	3	2.0094(0.0608)	(1.9313,2.2701)	2.0302(0.0451)	(1.9645,2.1890)	2.0738(0.0552)	(1.9436,2.2435)
		3.0759(0.2979)	(2.6519,4.4413)	3.0060(0.2195)	(2.8189,3.9889)	3.0738(0.2732)	(2.7085,4.2791)
	15	2.0076(0.0603)	(1.9323,2.2706)	2.0279(0.0443)	(1.9663,2.1885)	2.0718(0.0557)	(1.9426,2.2356)
		3.0726(0.2972)	(2.6538,4.4362)	3.0029(0.2196)	(2.8189,3.9802)	3.0707(0.2722)	(2.7106,4.2418)
0.5	0.1	2.0296(0.0615)	(1.9305,2.2643)	2.0502(0.0453)	(1.9646,2.1821)	2.0131(0.0557)	(1.9428,2.2297)
		3.1066(0.3000)	(2.6480,4.4044)	3.0364(0.2226)	(2.8130,3.9495)	3.1052(0.2753)	(2.7048,4.2092)
	0.5	2.0278(0.0613)	(1.9307,2.2605)	2.0489(0.0448)	(1.9654,2.1767)	2.0113(0.0560)	(1.9427,2.2266)
		3.1042(0.3005)	(2.6469,4.3610)	3.0332(0.2216)	(2.8150,3.9164)	3.1022(0.2752)	(2.7053,4.1696)
	3	2.0260(0.0612)	(1.9073,2.2332)	2.0464(0.0455)	(1.9412,2.1508)	2.0908(0.0555)	(1.9209,2.1987)
		3.1005(0.2997)	(2.5140,4.1747)	3.0308(0.2212)	(2.6807,3.7407)	3.0986(0.2747)	(2.5712,3.9915)
	15	2.0244(0.0616)	(1.9074,2.1698)	2.0446(0.0448)	(1.9427,2.1348)	2.0885(0.0562)	(1.9193,2.1582)
		3.0975(0.2993)	(2.5150,3.8031)	3.0273(0.2215)	(2.6801,3.6365)	3.0954(0.2743)	(2.5716,3.7516)
3	0.1	2.0220(0.0609)	(1.9079,2.1653)	2.0420(0.0449)	(1.9421,2.1336)	2.0862(0.0557)	(1.9199,2.1546)
		3.0946(0.2994)	(2.5148,3.7186)	3.0243(0.2212)	(2.6810,3.5674)	3.0926(0.2742)	(2.5722,3.6672)
	0.5	2.0196(0.0614)	(1.9069,2.1629)	2.0401(0.0451)	(1.9419,2.1269)	2.0847(0.0556)	(1.9207,2.1506)
		3.0913(0.2992)	(2.5147,3.6990)	3.0211(0.2209)	(2.6816,3.4168)	3.0897(0.2741)	(2.5722,3.5800)
	3	2.0177(0.0609)	(1.9082,2.1563)	2.0379(0.0450)	(1.9420,2.1237)	2.0817(0.0561)	(1.9192,2.1432)
		3.0886(0.2988)	(2.5160,3.6782)	3.0184(0.2208)	(2.6822,3.5244)	3.0869(0.2743)	(2.5719,3.6233)
	15	2.0155(0.0612)	(1.9079,2.1363)	2.0359(0.0448)	(1.9425,2.1038)	2.0804(0.0554)	(1.9206,2.1245)
		3.0856(0.2984)	(2.5160,3.5897)	3.0155(0.2208)	(2.6816,3.4380)	3.0837(0.2736)	(2.5736,3.5367)
15	0.1	2.1158(0.0640)	(1.8848,2.1323)	2.1226(0.0472)	(1.9202,2.0996)	2.0991(0.0588)	(1.8970,2.1219)
		3.2392(0.3131)	(2.4625,3.8649)	3.1651(0.2320)	(2.6354,3.6446)	3.2371(0.2872)	(2.5214,3.7930)
	0.5	2.0744(0.0627)	(1.8872,2.1283)	2.0803(0.0459)	(1.9231,2.0958)	2.0579(0.0577)	(1.8991,2.1184)
		3.1754(0.3074)	(2.4739,3.7840)	3.1029(0.2265)	(2.6458,3.5791)	3.1732(0.2816)	(2.5331,3.7142)
	3	2.0340(0.0615)	(1.8900,2.1289)	2.0702(0.0455)	(1.9237,2.0953)	2.0176(0.0564)	(1.9020,2.1183)
		3.1131(0.3007)	(2.4862,3.6656)	3.0427(0.2222)	(2.6543,3.4964)	3.1112(0.2764)	(2.5425,3.6071)
	15	2.0320(0.0618)	(1.8888,2.1057)	2.0687(0.0457)	(1.9233,2.0737)	2.0157(0.0560)	(1.9022,2.0942)
		3.1103(0.3003)	(2.4878,3.5327)	3.0394(0.2220)	(2.6539,3.3698)	3.1083(0.2761)	(2.5430,3.4740)

Table 3: Average ML estimators and Bayes Estimators (Average MSE in brackets) with corresponding asymptotic/HPD Confidence interval for fixed value of $\xi = 0.1$, $\zeta = 3$, n=50 and m=35

α	λ	Par	ML		Bayes1		Bayes2	
			Estimates(MSE)	Asymptotic CI	Estimates(MSE)	HPD CI	Estimates(MSE)	HPD CI
0.8	0.5	α	0.8025(0.0280)	(0.7472,0.8558)	0.8015(0.0124)	(0.7770,0.8244)	0.8016(0.0223)	(0.7575,0.8443)
		λ	0.5293(0.1661)	(0.2081,0.8495)	0.5240(0.0875)	(0.3546,0.6937)	0.5295(0.1413)	(0.2561,0.8021)
	1.6	α	0.8051(0.0309)	(0.7452,0.8648)	0.8018(0.0152)	(0.7723,0.8314)	0.8043(0.0253)	(0.7546,0.8534)
		λ	1.6290(0.1781)	(1.2849,1.9731)	1.6247(0.0996)	(1.4310,1.8164)	1.6297(0.1534)	(1.3328,1.9251)
	2	α	0.8066(0.0340)	(0.7401,0.8714)	0.8030(0.0182)	(0.7661,0.8375)	0.8056(0.0283)	(0.7494,0.8596)
		λ	2.0291(0.1901)	(1.6619,2.3953)	2.0248(0.1117)	(1.8080,2.2393)	2.0298(0.1652)	(1.7096,2.3488)
	3	α	0.8098(0.0369)	(0.7377,0.8809)	0.8037(0.0212)	(0.7619,0.8441)	0.8080(0.0314)	(0.7467,0.8679)
		λ	3.0295(0.2019)	(2.6385,3.4186)	3.0249(0.1236)	(2.7854,3.2630)	3.0297(0.1773)	(2.6868,3.3717)
1	0.5	α	1.0022(0.0400)	(0.9236,1.0792)	1.0009(0.0243)	(0.9531,1.0475)	1.0019(0.0345)	(0.9340,1.0686)
		λ	0.5363(0.2139)	(0.1205,0.9510)	0.5306(0.1357)	(0.2668,0.7936)	0.5366(0.1892)	(0.1689,0.9042)
	1.6	α	1.0051(0.0431)	(0.9215,1.0881)	1.0017(0.0273)	(0.9480,1.0550)	1.0040(0.0375)	(0.9311,1.0767)
		λ	1.6366(0.2261)	(1.1977,2.0748)	1.6302(0.1476)	(1.3435,1.9163)	1.6368(0.2013)	(1.2458,2.0268)
	2	α	1.0065(0.0460)	(0.9164,1.0954)	1.0023(0.0303)	(0.9434,1.0613)	1.0057(0.0404)	(0.9264,1.0833)
		λ	2.0369(0.2380)	(1.5743,2.4977)	2.0302(0.1597)	(1.7203,2.3400)	2.0367(0.2133)	(1.6228,2.4502)
	3	α	1.0092(0.0488)	(0.9138,1.1040)	1.0035(0.0333)	(0.9378,1.0680)	1.0075(0.0435)	(0.9233,1.0916)
		λ	3.0367(0.2500)	(2.5505,3.5216)	3.0300(0.1717)	(2.6969,3.3631)	3.0371(0.2253)	(2.5993,3.4734)
2	0.5	α	2.0021(0.0521)	(1.9001,2.1032)	2.0014(0.0362)	(1.9294,2.0710)	2.0018(0.0464)	(1.9109,2.0923)
		λ	0.5721(0.2619)	(0.0608,1.0832)	0.5608(0.1836)	(0.2015,0.9181)	0.5738(0.2374)	(0.1096,1.0364)
	1.6	α	2.0056(0.0549)	(1.8972,2.1122)	2.0017(0.0393)	(1.9245,2.0786)	2.0040(0.0493)	(1.9073,2.1005)
		λ	1.6725(0.2739)	(1.1372,2.2061)	1.6604(0.1955)	(1.2782,2.0422)	1.6735(0.2494)	(1.1867,2.1595)
	2	α	2.0069(0.0580)	(1.8926,2.1192)	2.0021(0.0422)	(1.9198,2.0846)	2.0058(0.0524)	(1.9028,2.1071)
		λ	2.0730(0.2861)	(1.5136,2.6302)	2.0603(0.2076)	(1.6550,2.4649)	2.0735(0.2613)	(1.5637,2.5829)
	3	α	2.0091(0.0610)	(1.8896,2.1286)	2.0031(0.0452)	(1.9145,2.0918)	2.0076(0.0554)	(1.8987,2.1156)
		λ	3.0721(0.2980)	(2.4911,3.6531)	3.0602(0.2196)	(2.6316,3.4883)	3.0733(0.2732)	(2.5395,3.6060)
3	0.5	α	3.0022(0.0539)	(2.8950,3.1078)	3.0005(0.0383)	(2.9255,3.0759)	3.0017(0.0483)	(2.9057,3.0966)
		λ	0.6080(0.2919)	(0.0354,1.1810)	0.5907(0.2137)	(0.1711,1.0093)	0.6103(0.2674)	(0.0855,1.1339)
	1.6	α	3.0050(0.0570)	(2.8930,3.1169)	3.0018(0.0413)	(2.9201,3.0825)	3.0046(0.0513)	(2.9034,3.1050)
		λ	1.7087(0.3041)	(1.1117,2.3043)	1.6909(0.2257)	(1.2478,2.1322)	1.7099(0.2793)	(1.1621,2.2576)
	2	α	3.0062(0.0600)	(2.8880,3.1234)	3.0029(0.0443)	(2.9149,3.0888)	3.0051(0.0544)	(2.8977,3.1114)
		λ	2.1089(0.3160)	(1.4880,2.7277)	2.0900(0.2377)	(1.6238,2.5564)	2.1101(0.2913)	(1.5384,2.6805)
	3	α	3.0100(0.0630)	(2.8855,3.1327)	3.0032(0.0473)	(2.9106,3.0961)	3.0079(0.0574)	(2.8948,3.1196)
		λ	3.1090(0.3280)	(2.4653,3.7512)	3.0904(0.2497)	(2.6010,3.5794)	3.1098(0.3034)	(2.5150,3.7038)

Table 4: Expected total time on test $E[x_m]$ for $\alpha = 2$, $\lambda = 3$

n	m		$\xi = 0.5$							$\xi = 1$						
		$\zeta \longrightarrow$	15	9	3	1	0.5	0.25	0.1	15	9	3	1	0.5	0.25	0.1
10	5		4.14	4.29	5.34	9.92	16.31	20.21	22.37	4.37	4.76	7.18	15.24	19.69	21.95	23.14
10	6		5.62	5.87	8.07	16.24	23.22	26.31	28.49	5.93	6.81	11.72	22.61	25.94	27.96	29.27
10	7		7.93	8.58	12.78	25.73	30.81	32.61	34.54	8.85	10.42	19.4	30.96	32.9	34.98	34.97
10	8		12.04	13.46	19.66	32.86	37.81	38.81	39.06	13.48	16.35	28.34	37.7	39.62	39.61	40.66
10	9		21.43	22.97	31.68	42.81	45.37	46.36	47.26	25.23	29.19	39.97	45.63	45.62	47.16	47.15
10	10		53.76	53.76	53.76	53.76	53.76	53.76	53.76	51.84	51.84	51.84	51.84	51.84	51.84	51.84
20	10		4.68	5.14	9.52	32.75	45.34	46.49	51.25	5.51	6.8	20.63	45.17	47.17	48.22	49.46
20	12		6.64	7.66	17.31	49.51	58.42	58.41	62.69	8.21	11.03	34.93	55.08	60.21	60.2	60.19
20	14		9.97	11.89	30.98	63.35	68.04	70.19	73.42	13.16	19.86	55.02	70.98	73.88	73.87	73.86
20	16		16.68	20.77	49.51	79.26	79.25	81.41	83.03	23.14	35.77	70.44	82.13	82.12	82.11	82.66
20	18		33.79	42.75	72.53	92.3	93.73	93.72	94.31	46.4	60.01	89.78	94.42	94.41	97.04	97.03
20	20		106.53	106.53	106.53	106.53	106.53	106.53	106.53	105.95	105.95	105.95	105.95	105.95	105.95	105.95
30	15		5.19	6.12	17.91	62.48	71.96	75.99	75.98	6.68	9.99	45.41	72.23	75.31	76.48	76.47
30	18		7.57	9.65	35.8	85.2	87.13	91.4	91.39	10.9	18.38	71.34	89.97	89.96	94.09	94.08
30	21		12.05	16.76	64.7	99.88	103.86	103.97	106.34	18.95	35.51	96.32	107.6	112.06	112.57	112.56
30	24		21.1	31.33	95.65	121.96	124.63	124.62	127.81	37.2	65.97	119.8	122.34	123.86	128.91	128.9
30	27		46.95	67.99	132.31	138	142.09	142.08	142.07	77.66	108.58	139.44	141.42	141.41	141.4	141.54
30	30		155.65	155.65	155.65	155.65	155.65	155.65	155.65	160.83	160.83	160.83	160.83	160.83	160.83	160.83
40	20		5.7	7.39	34.6	91.22	100.19	103.79	103.78	8.26	15.2	80.56	98.5	99.27	102.84	104.36
40	24		8.77	12.76	66.61	112.59	119.96	119.95	121.01	14.97	31.98	111.78	120.89	123.22	123.21	126.38
40	28		14.6	22.18	98.89	144.27	144.25	144.24	150.73	27.93	64.99	140.41	140.4	151.7	151.69	151.68
40	32		26.37	46.63	143.81	166.69	167.99	167.98	167.97	56.84	107.87	158.83	160.86	166.28	166.27	166.26
40	36		62.58	97.06	183.99	191.68	191.66	197.22	197.21	112.85	159.9	179.14	189.35	189.34	189.33	194.26
40	40		219.46	219.46	219.46	219.46	219.46	219.46	219.46	212.09	212.09	212.09	212.09	212.09	212.09	212.09
50	25		6.3	9.16	56.96	121.19	132.62	132.61	132.6	10.46	24.17	112.63	132.04	132.03	132.02	132.01
50	30		10.14	16.82	103.61	144.1	155.32	155.46	156.27	20.67	53.48	145.24	149.07	153.81	153.8	153.79
50	35		17.41	32.14	152.99	179.2	179.19	180.86	180.85	43.59	103.26	175.07	175.06	182.02	182.01	185.33
50	40		33.32	66.71	190.01	200.61	209.79	209.78	213.51	84.51	156.67	199.53	203.22	210.74	210.73	210.72
50	45		79.76	137.76	250.87	250.86	250.85	250.84	250.83	164.01	217.16	222.59	238.58	238.57	238.56	238.55
50	50		259.99	259.99	259.99	259.99	259.99	259.99	259.99	261.62	261.62	261.62	261.62	261.62	261.62	261.62

Table 5: Expected total time on test $E[x_m]$ for $\alpha = 0.8$, $\lambda = 1.6$

n	m		$\xi = 0.5$							$\xi = 1$						
		$\zeta \longrightarrow$	15	9	3	1	0.5	0.25	0.1	15	9	3	1	0.5	0.25	0.1
10	5		2.65	2.8	3.68	8.31	15.77	20.44	23.39	2.87	3.23	5.46	14.03	19.95	22.79	24.04
10	6		3.95	4.22	6.19	15.62	25.25	29.11	30.93	4.43	5.11	10.08	23.86	28.04	30.14	31.5
10	7		6.12	6.76	10.93	24.95	35.22	36.5	40.08	6.98	8.43	18.77	33.89	39.76	39.75	39.73
10	8		10.53	11.94	19.74	37.02	47.53	47.52	50.13	12.79	16.14	30.15	47.71	47.7	49.23	49.22
10	9		20.71	24.22	36.93	54.77	55.73	57.23	57.22	25.28	31.11	47.8	57.36	57.57	57.56	57.55
10	10		68.86	68.86	68.86	68.86	68.86	68.86	68.86	64.88	64.88	64.88	64.88	64.88	64.88	64.88
20	10		3.18	3.56	8.07	35.95	53.75	65.49	65.48	3.83	5.05	21.64	54	57.63	60.44	60.43
20	12		4.82	5.88	17.11	59.37	77.21	77.2	82.36	6.34	9.4	40.99	76.88	78.28	80.41	82.24
20	14		8.24	10.61	35.08	83.64	92.71	101.15	101.14	11.51	18.83	71.06	94.38	95.5	99.96	104.02
20	16		15.52	20.03	64.41	108.47	113.31	120.45	121.73	23.36	37.1	101.87	118.23	118.22	119.19	119.18
20	18		36.47	50.27	111.99	133.38	133.37	133.36	140.27	54.09	80.43	135.15	139.35	144.84	144.83	144.82
20	20		163.32	163.32	163.32	163.32	163.32	163.32	163.32	163.6	163.6	163.6	163.6	163.6	163.6	163.6
30	15		3.60	4.44	18.59	85.87	95.28	103.16	109.53	4.96	8.27	57.2	98	112.49	112.48	112.47
30	18		5.94	7.91	43.66	121.49	136.58	136.57	137.03	9.19	17.79	100.58	128.76	136.28	136.26	136.25
30	21		10.23	15.57	87.09	159.28	171.41	171.4	171.38	18.05	42.45	148.58	162.97	162.96	167.86	167.85
30	24		20.87	34.92	142.66	194.16	198.59	203.7	203.69	44.03	89.39	180.83	199.15	199.13	199.12	201.29
30	27		55.82	89.45	198.19	228.69	231.09	231.08	233.05	105.82	164.53	226.63	228.4	228.39	228.38	233.93
30	30		276.39	276.39	276.39	276.39	276.39	276.39	276.39	270.2	270.2	270.2	270.2	270.2	270.2	270.2
40	20		4.06	5.54	39.16	134.45	153.45	153.44	157.5	6.48	14.14	109.25	151.35	151.34	154.3	154.29
40	24		7.01	10.66	90.32	180.53	189.86	197.83	197.81	13.58	35.69	172.15	189.22	193.74	193.73	198.72
40	28		13.11	23.44	155.95	235.69	242.85	242.84	249.26	31.45	88.41	223.43	244.67	244.66	244.65	245.11
40	32		27.54	56.58	224.23	284.21	284.2	284.19	284.18	79.07	164.19	272.27	289.49	289.47	289.46	289.45
40	36		80.51	146.86	314.85	349.19	349.18	349.17	349.16	185.82	275.36	314.88	326.04	333.13	333.12	333.11
40	40		388.78	388.78	388.78	388.78	388.78	388.78	388.78	373.99	373.99	373.99	373.99	373.99	373.99	373.99
50	25		4.56	7.07	76.98	190.46	202.82	202.8	213.33	8.59	23.95	167.62	197.49	200.65	202.65	210.87
50	30		8.23	15.38	157.51	236.9	250.73	250.72	257.53	20.79	66.23	239.33	262.75	262.74	262.73	262.72
50	35		16.44	36.33	255.71	309.86	314.72	323.41	323.39	51.49	153.8	313.71	336.49	336.48	336.47	336.45
50	40		38.52	91.82	358.63	363.76	363.79	376.93	384.22	131.05	289.07	368.52	388.16	388.15	394.25	394.24
50	45		116.4	215.53	423.94	428.49	459.53	459.52	460.2	285.76	386.26	434.95	441.18	441.17	457.89	457.88
50	50		513.71	513.71	513.71	513.71	513.71	513.71	513.71	510.02	510.02	510.02	510.02	510.02	510.02	510.02

Table 6: Expected total time on test $E[x_m]$ for $\alpha = 2$, $\lambda = 3$

n	m		$\zeta = 1$							$\zeta = 3$						
		$\xi \longrightarrow$	15	9	3	1	0.5	0.25	0.1	15	9	3	1	0.5	0.25	0.1
10	5		21.91	21.9	21.43	15.65	10.06	6.73	4.86	21.76	20.88	14.38	7.2	5.3	4.57	4.14
10	6		29.11	28.31	28.3	22.19	16.7	10.23	6.84	28.51	27.56	22.8	11.93	8.06	6.34	5.58
10	7		34.12	34.11	34.1	29.77	24.41	16.55	10.43	35.43	35.42	29.62	19.9	12.62	9.49	7.94
10	8		41.69	39.29	39.28	38.23	33.2	25.43	17.02	40.73	40.67	36.24	29.11	19.71	14.76	12.41
10	9		45.56	45.55	45.26	44.77	43.84	36.89	26.91	48.05	46.16	46.15	39.59	32.54	25.33	21.6
10	10		51.63	51.63	51.63	51.63	51.63	51.63	51.63	53.76	53.76	53.76	53.76	53.76	53.76	53.76
20	10		50.87	47.78	46.85	41.38	32.14	17.56	7.41	49.82	47.49	43.43	20.99	9.74	6.04	4.73
20	12		61.76	60.92	60.43	57	49.38	30.82	12.12	58.44	58.43	56.38	35.49	17.27	9.54	6.7
20	14		72.45	71.1	71.06	69.4	62.26	45.47	20.6	72.19	69.6	68.66	54.37	30.51	15.63	10.11
20	16		86.63	83.73	81.35	79.76	76.66	64.67	35.24	81.19	81.18	80.85	72.99	49.68	28.45	16.83
20	18		91.39	91.38	91.37	91.36	88.21	83.67	61.2	93.35	93.34	93.33	89.93	76.03	51.75	32.85
20	20		106	106	106	106	106	106	106	106.53	106.53	106.53	106.53	106.53	106.53	106.53
30	15		74.15	74.14	74.13	71.72	63.56	40.46	11.89	73.93	72	70.99	44.72	19.23	8.15	5.34
30	18		94.78	93.12	92.22	85.39	81.98	63.06	22.87	93.29	87.15	87.03	72.3	36.57	14.6	7.97
30	21		107.11	107.1	107.09	102.96	102.14	91.79	41.92	107.27	105.67	105.66	97.75	62.05	26.45	12.57
30	24		125.81	125.8	124.16	124.15	123.61	110.85	67.62	126.44	122.25	121.1	119.69	98.26	50.02	22.35
30	27		139.48	139.47	135.26	135.25	135.24	130.27	107.42	140.74	139.38	137.96	134.97	130.13	91.88	46.61
30	30		160.75	160.75	160.75	160.75	160.75	160.75	160.75	155.65	155.65	155.65	155.65	155.65	155.65	155.65
40	20		101.18	101.17	101.16	101.15	93.1	68.03	20.42	99.31	99.3	96.99	80.61	34.33	11.36	5.85
40	24		120.61	120.6	120.59	120.58	111.11	97.54	39.57	119.48	119.47	119.46	111.93	65.8	21.95	8.95
40	28		141.73	141.72	141.71	141.7	141.69	130.92	69.29	151.41	147.29	140.09	140.08	103.44	44.54	15.22
40	32		163.59	163.58	163.57	163.56	163.55	163.54	109.28	165.94	165.92	161.08	160.71	142.77	81.55	28.77
40	36		193.74	189.75	189.46	183.34	183.32	182.53	154.89	191.66	188.91	187.71	184.1	171.09	136.58	65.21
40	40		212.51	212.51	212.51	212.51	212.51	212.51	212.51	219.46	219.46	219.46	219.46	219.46	219.46	219.46
50	25		126.6	126.59	126.58	123.55	120.89	99.43	33.45	126.17	124.56	124.55	107.83	59.12	16.51	6.76
50	30		151.1	151.09	151.07	151.06	150.38	135.61	62.74	157.09	153.19	153.18	145.56	106.78	35.79	10.61
50	35		175.91	175.9	175.88	175.87	173.35	173.34	104.7	173.69	173.68	173.67	173.66	148.35	68.61	18.67
50	40		210.41	208.61	208.6	208.59	206.75	204.72	154.5	204.31	204.3	204.29	204.27	187.87	123.06	37.49
50	45		246.15	231.31	231.3	228.72	228.71	228.7	210.21	231.15	227.27	227.26	227.25	227.24	192.28	83.63
50	50		261.62	261.62	261.62	261.62	261.62	261.62	261.62	259.99	259.99	259.99	259.99	259.99	259.99	259.99

Table 7: Expected total time on test $E[x_m]$ for $\alpha = 0.8$, $\lambda = 1.6$

n	m		$\zeta = 1$							$\zeta = 3$						
		$\xi \longrightarrow$	15	9	3	1	0.5	0.25	0.1	15	9	3	1	0.5	0.25	0.1
10	5		24.7	23.62	21.48	13.71	8.66	4.75	3.29	22.91	20.83	13.28	5.41	3.73	3.01	2.7
10	6		31.27	31.26	29.61	23.74	16.01	8.72	5.13	30.97	30.75	24.28	10.12	6.14	4.63	4.04
10	7		41.86	39.94	38.62	34.36	25.15	15.53	8.46	39.52	37.68	35.21	18.76	11.2	7.63	6.22
10	8		51.42	49.26	46.73	45.44	37.16	26.57	15.47	47.81	47.77	43.26	32.27	20.15	13.6	10.69
10	9		57.87	57.86	57.12	57.11	51.1	42.5	28.4	56.95	55.14	55.13	49.11	36.01	28.37	21.13
10	10		64.88	64.88	64.88	64.88	64.88	64.88	64.88	68.86	68.86	68.86	68.86	68.86	68.86	68.86
20	10		60.41	59.28	59.27	51.92	37.2	17.87	5.76	62.01	60.89	52.86	21.5	7.57	4.45	3.16
20	12		76.9	75.75	75.74	73.44	59.84	33.88	10.47	80.34	79.99	71.72	41.51	16.71	7.98	4.99
20	14		97.02	97.01	97	94.97	88.74	60.68	21.57	93.85	93.84	93.06	70.53	33.6	14.83	8.48
20	16		123.42	121.32	120.46	109.59	109.58	89.61	42.2	119.63	119.63	119.63	119.63	119.63	119.63	119.63
20	18		133.67	133.66	133.65	128.87	128.86	120.47	81.96	142.22	137.83	131.7	131.69	102.76	69.65	36.89
20	20		163.6	163.6	163.6	163.6	163.6	163.6	163.6	163.32	163.32	163.32	163.32	163.32	163.32	163.32
30	15		106.97	106.96	104.49	98.39	81.8	44.28	10.52	103.23	103.22	94.44	59.09	17.27	6.36	3.69
30	18		137.99	137.98	135.06	135.05	119.34	80.9	23.7	131.01	131	129.83	101.29	42.07	12.85	6.01
30	21		165.34	165.33	164.46	164.45	151.73	131.8	48.86	162.89	162.88	161.04	135.98	88.6	28.74	10.71
30	24		192.69	192.1	192.09	192.08	192.07	182.78	92.67	199.23	198.09	190.15	190.14	148.14	62.01	22.25
30	27		240.11	233.36	221.5	221.49	221.48	215.97	160.35	231.29	231.28	229.57	228.22	190.41	136.69	58.26
30	30		270.2	270.2	270.2	270.2	270.2	270.2	270.2	276.39	276.39	276.39	276.39	276.39	276.39	276.39
40	20		157.2	151.26	151.25	144.24	134.25	95.36	21.22	149.95	149.94	148.7	112.44	39.91	9.78	4.17
40	24		192.91	192.9	185.49	185.48	185.46	151.89	49.92	185.39	185.38	185.37	162.71	91.31	21.69	7.36
40	28		233.05	233.04	231.92	228.38	228.37	199.66	100.75	233.47	224.61	224.6	216.69	152.02	56.29	13.82
40	32		287.34	287.33	287.32	284.45	263.35	263.33	159.76	287.7	287.68	268.38	268.37	247.82	121.28	30.94
40	36		322.06	322.05	315.43	315.42	314.51	314.5	270.89	327.21	327.2	324.56	324.55	317.18	236.51	87.07
40	40		373.99	373.99	373.99	373.99	373.99	373.99	373.99	388.78	388.78	388.78	388.78	388.78	388.78	388.78
50	25		210.96	206.74	195.87	195.86	193.44	152.43	38.63	214.3	197.84	192.46	166.54	81.04	15.86	4.88
50	30		266.94	262.3	251.68	244.63	241.88	225.92	91.04	249.75	249.74	249.72	227.43	147.81	40.85	8.77
50	35		306.25	296.78	296.77	296.76	296.75	292.66	164.97	318.49	318.48	308.98	306.7	256.79	96.77	18.26
50	40		388.53	374.21	374.08	362.64	361.01	357.37	262.62	381.86	380.28	362.88	362.87	338.05	208.41	42.96
50	45		425.17	425.16	425.15	425.14	416.05	416.03	377.57	441.93	438.81	438.8	425.8	410.41	349.73	123.17
50	50		510.02	510.02	510.02	510.02	510.02	510.02	510.02	513.71	513.71	513.71	513.71	513.71	513.71	513.71

The third data set represents the lifetime‘s data relating to relief times (in minutes) of 20 patients receiving an analgesic and reported by Gross and Clark (1975).

1.1	1.4	1.3	1.7	1.9	1.8	1.6	2.2	1.7
2.7	4.1	1.8	1.5	1.2	1.4	3.0	1.7	2.3
1.6	2.0							

Fourth data set reported by Efron (1988) represent the survival times of a group of patients suffering from Head and Neck cancer disease and treated using a combination of radiotherapy and chemotherapy (RT+CT).

12.20	23.56	23.74	25.87	31.98	37.0	41.35	47.38	55.46
58.36	63.47	68.46	78.26	74.47	81.43	84.00	92.00	94.00
110.0	112.0	119.0	127.0	130.0	133.0	140.0	146	155.0
159.0	173.0	179.0	194.0	195.0	209.0	249.0	281.0	319.0
339.0	432.0	469.0	519.0	633.0	725.0	817.0	1776	

The MLEs for the unknowns are calculated for all above data sets based on complete sample and reported in Table 8, using the procedure explained in section 2. In this Table, $\hat{\lambda}_{ML}$ and $\hat{\alpha}_{ML}$ represent the maximum likelihood estimates for the parameters α and λ, respectively. The quantity reported in brackets is the standard deviation (sd) computed based on the square root of inverse of estimated Fisher information matrix as given in equation (18). Here, we also compute the K-S statistic and corresponding p-value for the purpose of goodness-of-fit. The quantity log of likelihood, AIC and BIC are also presented.

For the illustration of our methodology, we have generated censored data for a prefixed m, ξ and ζ. It may be worthwhile to mention here that the number of drop-outs are random and we are generating the progressive type-II censored data from the complete sample data, therefore, we can study the average performance of the estimators. For this purpose, we generated 2000 censored data sets for given m and accordingly the ξ's and ζ's from the considered complete data set. The m is chosen 80% of the complete sample size and $\xi = 3, \zeta = 0.1$ are considered. The average ML estimates, Bayes estimates with corresponding average MSE and confidence interval are reported in Table 9.

Table 8: MLEs with other statistic for considered real data sets

data	$\hat{\alpha}_{ML}(sd)$	$\hat{\lambda}_{ML}(sd)$	KS-Statistic	p-value	log-lik	AIC	BIC
Boing AC	1.1458(0.2357)	65.3018(53.3062)	0.1480	0.8970	-84.0772	172.1544	173.5705
Chloride	0.8803(0.1093)	0.6539(0.1347)	0.1134	0.7740	-58.6266	121.2532	124.3059
Relief Time	4.0173(0.6972)	6.0221(1.9636)	0.1019	0.9850	-15.4087	34.8174	36.8089
RT+CT	1.0134(0.1119)	80.7880(36.9616)	0.0926	0.8110	-279.5701	563.1403	566.7086

Table 9: Average estimates with corresponding average MSE and Assymptotic/HPD confidence Interval for fixed $\xi = 3$, $\zeta = 0.1$ and $m = 0.80 \times$ sample size, for considered real data sets

	Data	α		λ	
Boing AC	ML	1.1248(0.2650)	(0.6117, 1.6413)	65.8131(55.1362)	(-42.2472, 173.8737)
	Bayes	1.0250(0.2260)	(0.6846, 1.6378)	65.3132(53.1317)	(-36.7806, 171.4009)
Chloride	ML	0.8995(0.1385)	(0.6342, 1.1639)	0.6331(0.1643)	(0.3170, 0.9491)
	Bayes	0.8498(0.1249)	(0.6541, 1.1325)	0.1339(0.1569)	(0.3501, 0.9172)
Relief Time	ML	3.9963(0.7268)	(2.5818, 5.4195)	5.7018(1.9933)	(1.8030, 9.6023)
	Bayes	4.0066(0.7254)	(2.6356, 5.4123)	5.8021(1.9809)	(1.9252, 9.0770)
RTCT	ML	0.9928(0.1415)	(0.7258, 1.2645)	79.7676(37.8907)	(5.5043, 154.0242)
	Bayes	1.0932(0.1294)	(0.7915, 1.1693)	81.2678(35.1421)	(5.6798, 150.0634)

7. Conclusion

In this chapter, we have developed a sampling procedure for life-testing experiment called as progressive Type-II censoring scheme with beta-binomial removals (PT-II CBBR) which covers the uncertainty of the real phenomenon of life-testing procedure. The Bayesian procedure provides the more accurate and precise estimates of the parameters even if we consider the vague prior. Finally, we can conclude that the discussed methodology provides the more flexible procedure for life-testing experiment and can be recommended for their use in medical, engineering and in other areas where such type of life-tests are needed.

References

Balakrishnan N, Aggarwala R (2000). *Progressive Censoring: Theory, Methods, And Applications*. Birkhauser, Boston.

Balakrishnan N, Cramer E (2014). *The Art of Progressive Censoring: Applications to Reliability and Quality*. Statistics for Industry and Technology. Springer New York. ISBN 9780817648060.

Balakrishnan N, Sandhu RA (1995). "A Simple Simulational Algorithm for Generating Progressive Type-II Censored Samples." *American Statistics*, **49**(2), 229–230.

Bhaumik D, Kapur K, Gibbons R (2009). "Testing Parameters of a Gamma Distribution for Small Samples." *Technometrics*, **51**, 326–334.

Brooks S (1998). "Markov Chain Monte Carlo Method and Its Application." *Journal of Royal Statistical Society, Series. D*, **47**, 69–100.

Calabria R, Pulcini G (1990). "On the Maximum Likelihood and Least-squares Estimation in the Inverse Weibull Distributions." *Statistica Applicata*, **2**(1), 53–66.

Childs A, Balakrishnan N (2000). "Conditional Inferences Procedure for the Laplace Distributions when the Observed Samples Are Progressively Censored." *Metrika*, **52**, 253–265.

Cohen AC (1963). "Progressively Censored Samples in Life Testing." *Technometries*, **5**, 327–329.

Dumonceaux R, Antle CE (1973). "Discrimination between the Log-normal and the Weibull Distributions." *Technometrics*, **15**(4), 923–926.

Efron B (1988). "Logistic Regression, Survival Analysis and the Kaplan-Meier Curve." *Journal of the American Statistical Association*, **83**, 414–425.

Gelmen A, Carlin J, Stern H, Rubin D (1995). *Bayesian Data Analysis*. Text in Statistical Science, Chapman & Hall, London.

Gross AJ, Clark V (1975). *Survival Distributions: Reliability Applications in the Biomedical Sciences*. John Wiley & Sons.

Hastings W (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrika*, **55**, 97–109.

Kaushik A, Singh U, Singh SK (2017). "Bayesian Inference for the Parameters of Weibull Distribution Under Progressive Type-I Interval Censored Data with Beta-Binomial Removals." *Communications in Statistics-Simulation and Computation*, **46**(4), 3140–3158.

Khan M, Pasha G, Pasha A (2008). "Theoretical Analysis of Inverse Weibull Distribution." *WSEAS Transictions on Mathematics*, **7**.

Kim DH, Lee WD, Kang SG (2014). "Non-informative Priors for Inverse Weibull distribution." *Journal of Statistical Computation and Simulation*, **84**(5), 1039–1054.

Langlands AO, Pocock SJ, Kerr GR, Gore SM (1997). "Long-term Survival of Patients with Breast Cancer: A Study of the Curability of the Disease." *British Medical Journal*, **2**, 1247–1251.

Maswadah M (2003). "Conditional Confidence Interval Estimation for the Inverse Weibull Distribution Based on Censored Generalized Order Statistics." *Journal of Statistical Computation and Simulation*, **73**(12), 887–898.

Natzoufras I (2009). *Bayesian Modeling Using WinBugs*. Series B, Wiley, New York.

Nelson WB (1982). *Applied Life Data Analysis*. John Wiley & Sons.

Noor F, Aslam M (2013). "Bayesian Inference of the Inverse Weibull Mixture Distribution Using Type-I Censoring." *Journal of Applied Statistics*. Doi:10.1080/02664763.2013.780157.

Proschan F (1963). "Theoretical Explanation of Observed Decreasing Failure Rate." *Technometrics*, **5**, 375–383.

Singh S, Singh U, Sharma V (2014). "Expected Total Test Time and Bayesian Estimation for Generalized Lindley Distribution under Progressively Type-II Censored Sample where Removals Follow the Beta-binomial Probability Law." *Applied Mathematics and Computation*, **222**, 402–419.

Singh SK, Singh U, Kumar D (2013a). "Bayesian Estimation of Parameters of Inverse Weibull Distribution." *Journal of Applied Statistics*. Doi:10.1080/02664763.2013.789492.

Singh SK, Singh U, Kumar M (2013b). "Estimation of Parameters of Exponentiated Pareto Distribution for Progressive Type-II Censored Data with Binomial Random Removals Scheme." *Electronic Journal of Applied Statistics*, **06**, 130–148.

Smith A, Roberts G (1993). "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods." *Journal of Royal Statistical Society, Series B*, **55**, 3–23.

Tse SK, Yang C, Yuen H (2000). "Statistical Analysis of Weibull Distributed Lifetime Data under Type II Progressive Censoring with Binomial Removals." *Journal of Applied Statistics*, **27**(8), 1033–1043.

Wu SJ, Chang CT (2003). "Inference in the Pareto Distribution Based on Progressive Type -II Censoring with Random Removals." *Journal of Applied Statistics*, **30**, 163–172.

Yuen HK, Tse SK (1996). "Parameters Estimation for Weibull Distribution under Progressive Censoring with Random Removal." *Journal Statistical Computation and Simulation*, **55**, 55–71.

Affiliation:

Arun Kaushik
Department of Statistics, Institute of Science
Banaras Hindu University
Varanasi, India-221005
E-mail: arundevkauhsik@gmail.com
URL: http://arun-kaushik.github.io/

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 47 *Submitted:* 2016-07-09
February 2018 *Accepted:* 2016-09-07

Austrian Journal of Statistics
February 2018, Volume 47, 95–104.
http://www.ajs.or.at/
doi:10.17713/ajs.v47i1.579

Moments Inequalities for NBRUL Distributions with Hypotheses Testing Applications

M. A. W. Mahmoud
Al-Azhar University

R. M. EL-Sagheer
Al-Azhar University

W. B. H. Etman
Al-Azhar University

Abstract

In this paper, moment inequalities for the new better than renwal used in Laplace transform order (NBRUL) class of ageing distributions are derived. This inequalities demonstrate that if the mean life is finite, then all higher order moments exist. A new test for exponentiality versus NBRUL can be constructed using thes inequalities. Pitman's asymptotic efficiencies and critical values of the proposed test are calculated and tabulated. The powers of this test are estimated for some famously alternatives distributions in reliability such as Linear failure rate, Weibull and gamma distributions. Finally, examples in different areas are used as a practical applications of the proposed test.

Keywords: classes of life distributions, NBRUL, moments inequalities, life testing.

1. Introduction

Classes of life distributions are defined to classify the life distributions according to their aging properties. The definitions of these classes helped statisticians to define the test statistics. The test statistics are defined based on definition of the classes. The main aim of constructing new tests is to gain higher efficiencies. Many authors proposed tests for exponentiality versus some classes of life distributions based on the moment inequalities. Testing exponentiality against IFR, NBU and NBUE based on moment inequalities have been studied by (Ahmad 2001); (Ahmad and Mugdadi 2004) constructed the tests of NBUC, IFRA and DMRL depends on the moment inequalities, while testing NRBU and RNBU based on moment inequalities have been studied by (Mahmoud, EL-arishy, and Diab 2003). Using the moment inequalities of the class NBUL, (Mahmoud, Diab, and Kayid 2009) constructed a test statistic for testing exponentiality versus this class.

In this paper our theme formulates a new test statistic for testing exponentiality against NBRUL class based on the moment inequalities and discuss this test. The main classes of life distributions which have been introduced in the literature are based on new better than used (NBU), new better than used failure rate (NBUFR), new better than average failure rate (NBAFR), new better than used renewal failure rate (NBURFR), new better than used average renewal failure rate (NBARFR), new better than renewal used (NBRU) and exponential better than used in Laplace transform order (EBUL). Testing exponentiality against some classes of life distributions has been introduced by many researchers from many points of views. For more details one can refer to (Bryson and Siddiqui 1969), (Deshpande, Kochar, and Singh 1986), (Abouammoh and Ahmed 1988, 1992), (Abouammoh, Abdulghani, and

Qamber 1994), (Mahmoud, Moshref, and Mansour 2015), (Kumazawa 1983), (Ahmad 1994, 2001). (Abouammoh and Newby 1989), (Mahmoud and Abdul Alim 2002, 2003, 2008), (Ahmad, Alwasel, and Mugdadi 2001), (Abu-Youssef 2009), (Ismail and Abu-Youssef 2012) , (Mahmoud and Rady 2013). Recently (Atallah, Mahmoud, and Al-Zahrani 2014) developed a new method for testing exponentiality which is more general and flexible than goodness approach.

The rest of this paper can be organized as follows, Section 2 gives a brief knowledge about renewal classes. In Section 3 moment inequalities for the NBRUL class are developed. In Section 4, Testing exponentiality against NBRUL is proposed based on moment inequalities. In Section 5, Pitman's asymptotic efficiency (PAE) of the test for several common alternatives will be considered. In Section 6, Monte Carlo null distribution critical points from the null distribution for sample size $n = 5(5)35;39;40(5)50$. In section 7, The power estimate for the test are calculated. Finally, the application of the proposed test for real data sets are discussed in Section 8.

2. Renewal classes

Let T be a random variable represents life time of a device (system or component) with a continuous life distribution $F(t)$. Upon arising the failure of the device, it can be substituted by a sequence of mutually independent devices which are identically distributed with the same life distribution $F(t)$. The following stationary renewal distribution constitutes the remaining life distribution of the device under operation at time t.

$$W_F(t) = \mu_F^{-1} \int_0^t \overline{F}(u)\,du, \quad t \geq 0,$$

where $\mu_F = \mu = \int_0^\infty \overline{F}(u)\,du$.

It is easy to show that

$$\overline{W_F}(t) = \mu_F^{-1} \int_t^\infty \overline{F}(u)\,du, \quad t \geq 0.$$

For extra details, see (Barlow and Proschan 1981), (Abouammoh and Ahmed 1988, 1992). Now we need to present the definitions of the NBRU (NWRU) and NBRUL (NWRUL) classes of life distributions.

Definition 2.1. (Abouammoh *et al.* 1994) If X is a random variable with survival function $\bar{F}(x)$, then X is said to have new better (worse) than renewal used property, denoted by NBRU (NWRU), if

$$\overline{W_F}(x|t) \leq (\geq) \overline{F}(x|0), \quad x \geq 0, t \geq 0,$$

or

$$\overline{W_F}(x+t) \leq (\geq) \overline{W_F}(t)\,\overline{F}(x), \quad x \geq 0, t \geq 0.$$

Depending on the definition (2.1), (Mahmoud, EL-Sagheer, and Etman 2016) defined a new class which is called new better (worse) than renewal used in Laplace transform order NBRUL (NWRUL) as follows

Definition 2.2. X is said to be NBRUL (NWRUL) if

$$\int_0^\infty e^{-sx}\overline{W_F}(x+t)\,dx \leq (\geq) \overline{W_F}(t) \int_0^\infty e^{-sx}\overline{F}(x)\,dx, \quad x,t,s \geq 0.$$

It is obvious that NBRU⇒NBRUL⇒NBRUE.

3. Moments inequalities

In this section, the moment inequalities for NBRUL class are established.

Theorem 3.1. *Let F be NBRUL life distribution such that all moments exist and finite then for integers $r \geq 0$ and $s \geq 0$. Then*

$$\begin{aligned}\frac{\mu_{(r+2)}}{s(r+1)(r+2)}[1-\zeta(s)] \quad &\geq \quad \frac{-(-1)^r r!}{s^{r+2}}[\mu_F - \frac{1}{s}(1-\zeta(s))] \\ &+\frac{r!}{s^{r+1}}\sum_{i=0}^{r}(-1)^i \frac{s^{r-i}}{(r-i+2)!}\mu_{(r-i+2)}, \end{aligned} \tag{1}$$

where $\mu_{(r)} = E(X^r), \zeta(s) = Ee^{-sX}$.

Proof. Since F is NBRUL, then

$$\int_0^\infty e^{-sx}\overline{W_F}(x+t)\,dx \leq \overline{W_F}(t)\int_0^\infty e^{-sx}\overline{F}(x)\,dx, \quad x,t \geq 0. \tag{2}$$

Making use of (2), yields

$$\int_0^\infty t^r \int_0^\infty e^{-sx}\overline{W_F}(x+t)\,dxdt \leq \int_0^\infty t^r\overline{W_F}(t)\int_0^\infty e^{-sx}\overline{F}(x)dxdt. \tag{3}$$

The left hand side of (3) can be written as

$$\int_0^\infty t^r\overline{W_F}(t)\int_0^\infty e^{-sx}\overline{F}(x)dxdt = E\int_0^\infty t^r\overline{W_F}(t)\int_0^\infty e^{-sx}I(X>x)dxdt,$$

where

$$I(X>x) = \begin{cases} 0 & if\ x \geq X, \\ 1 & if\ x < X. \end{cases}$$

After some calculations, the left hand side of (3) is given by

$$\frac{\mu_F^{-1}\mu_{(r+2)}}{s(r+1)(r+2)}(1-\zeta(s)). \tag{4}$$

Also, the right hand side of (3) can be put in the following form

$$\int_0^\infty t^r\overline{W_F}(t)\int_0^\infty e^{-sx}\overline{F}(x)dxdt = \int_0^\infty e^{-sv}\overline{W_F}(v)\int_0^v u^r e^{su}dudv. \tag{5}$$

After some calculations (5) can be rewritten as

$$\begin{aligned}\int_0^\infty t^r\overline{W_F}(t)\int_0^\infty e^{-sx}\overline{F}(x)dxdt \quad &= \quad \frac{r!}{s^{r+1}}\mu_F^{-1}\sum_{i=0}^{r}(-1)^i\frac{s^{r-i}}{(r-i+2)!}\mu_{(r-i+2)} \\ &-\frac{(-1)^r r!}{s^{r+2}}\mu_F^{-1}\left[\mu_F - \frac{1}{s}(1-\zeta(s))\right]. \end{aligned} \tag{6}$$

From (4) and (6), Eq. (1) can be proved. □

Remark. For $r=1$, Eq.(1) will be reduced to

$$\frac{\mu_3}{6s}[1-\zeta(s)] \leq \frac{1}{s^3}\left[\mu - \frac{1}{s}(1-\zeta(s))\right] + \frac{1}{s^2}\left[\frac{s}{6}\mu_{(3)} - \frac{1}{2}\mu_{(2)}\right], \tag{7}$$

where $\mu_{(r)} = \int_0^\infty x^r dF(x)$.

4. Testing against NBRUL alternatives

Using Inequality (7) we can test the null hypothesis $H_0 : F$ is exponential aganist $H_1 : F$ is NBRUL and not exponential. $\delta_1(s)$ has been used as follows

$$\delta^{(1)}(s) = \frac{1}{2s^2}\mu_{(2)} - \frac{1}{6s}\mu_{(3)}\zeta(s) - \frac{1}{s^4}\zeta(s) - \frac{1}{s^3}\mu + \frac{1}{s^4}. \tag{8}$$

Note that under $H_0, \delta^{(1)}(s) = 0$, while under $H_1, \delta^{(1)}(s) > 0$.
Let $X_1, X_2, X_3, \ldots\ldots\ldots, X_n$ be a random sample from a distribution F. The empirical estimate $\delta_n^{(1)}(s)$ of $\delta^{(1)}(s)$ can be obtained as

$$\delta_n^{(1)}(s) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} [\frac{1}{2s^2} X_i^2 - \frac{1}{6s} X_i^3 e^{-sX_j} - \frac{1}{s^4} e^{-sX_i} - \frac{1}{s^3} X_i + \frac{1}{s^4}].$$

To make the test invarient, let $\Delta_n^{(1)}(s) = \frac{\delta_n^{(1)}(s)}{\overline{X}^4}$, where $\overline{X}$ is the sample mean. Then

$$\Delta_n^{(1)}(s) = \frac{1}{n^2 \overline{X}^4} \sum_{i=1}^{n} \sum_{j=1}^{n} [\frac{1}{2s^2} X_i^2 - \frac{1}{6s} X_i^3 e^{-sX_j} - \frac{1}{s^4} e^{-sX_i} - \frac{1}{s^3} X_i + \frac{1}{s^4}]. \tag{9}$$

One can note that $\delta^{(1)}(s)$ is an unbiased estimator of $\delta_n^{(1)}(s)$.
Now, set

$$\phi_s(X_i, X_j) = \frac{1}{2s^2} X_i^2 - \frac{1}{6s} X_i^3 e^{-sX_j} - \frac{1}{s^4} e^{-sX_i} - \frac{1}{s^3} X_i + \frac{1}{s^4}, \tag{10}$$

and define the symmetric Kernel

$$\psi_s(X_i, X_j) = \frac{1}{2!} \sum_R \phi_s(X_i, X_j),$$

where the summation is over all arrangements of X_i, X_j. Then $\Delta_n^{(1)}(s)$ in (9) is equivalent to the U_n-statistic given by

$$U_n = \frac{1}{\binom{n}{2}} \sum_{i<j} \psi_s(X_i, X_j). \tag{11}$$

The asymptotic normality of $\Delta_n^{(1)}(s)$ can be summarized in the following theorem.

Theorem 4.1. *(i) As $n \to \infty$, $\sqrt{n}(\Delta_n^{(1)}(s) - \Delta^{(1)}(s))$ is asymptotically normal with mean 0 and variance $\sigma^2(s)$, where*

$$\begin{aligned} \sigma^2(s) &= Var\{\frac{1}{2s^2} X^2 - \frac{1}{6s} X^3 \zeta(s) - \frac{1}{s^4} e^{-sx} - \frac{1}{s^3} X + \frac{1}{2s^2} \mu_{(2)} \\ &\quad - \frac{1}{6s} e^{-sx} \mu_{(3)} - \frac{1}{s^4} \zeta(s) - \frac{1}{s^3} \mu + \frac{2}{s^4}\}. \end{aligned} \tag{12}$$

(ii) Under H_0, the variance $\sigma_0^2(s)$ is

$$\sigma_0^2(s) = \frac{19 + 14s + s^2}{(1+s)^4 (1+2s)}. \tag{13}$$

Proof. Using standard U-statistic theory, (Lee 1989),

$$\sigma^2(s) = Var\{E[\phi_s(X_1, X_2) \mid X_1] + E[\phi_s(X_1, X_2) \mid X_2]\}.$$

Recall the definition of $\phi_s(X_i, X_j)$ in (10), thus it is easy to show that

$$E(\phi_s(X_1, X_2) \mid X_1) = \frac{1}{2s^2} X^2 - \frac{1}{6s} X^3 \int_0^\infty e^{-sx} dF(x) - \frac{1}{s^4} e^{-sx} - \frac{1}{s^3} X + \frac{1}{s^4},$$

and

$$\begin{aligned} E(\phi_s(X_1, X_2) \quad | \quad X_2) &= \frac{1}{2s^2} \int_0^\infty x^2 dF(x) - \frac{1}{6s} e^{-sx} \int_0^\infty x^3 dF(x) - \frac{1}{s^4} \int_0^\infty e^{-sx} dF(x) \\ &\quad - \frac{1}{s^3} \int_0^\infty x dF(x) + \frac{1}{s^4}, \end{aligned}$$

therefore,

$$\begin{aligned}\sigma^2(s) &= Var\{\frac{1}{2s^2}X^2-\frac{1}{6s}X^3\int_0^\infty e^{-sx}dF(x)-\frac{1}{s^4}e^{-sx}-\frac{1}{s^3}X+\frac{1}{2s^2}\int_0^\infty x^2dF(x)\\ &\quad -\frac{1}{6s}e^{-sx}\int_0^\infty x^3dF(x)-\frac{1}{s^4}\int_0^\infty e^{-sx}dF(x)-\frac{1}{s^3}\int_0^\infty xdF(x)+\frac{2}{s^4}\}.\end{aligned}$$

Under H_0

$$\sigma_0^2(s)=\frac{19+14s+s^2}{(1+s)^4(1+2s)}.$$

□

5. The Pitman asymptotic efficiency

To judge on the quality of this procedure, Pitman asymptotic efficiencies (PAEs) are computed and compared with some other tests for the following alternative distributions:

(i) The Weibull distribution: $\bar{F}_1(x)=e^{-x^{\theta}}, x\geq 0,\theta\geq 1$.

(ii) The linear failure rate distribution (LFR): $\bar{F}_2(x)=e^{-x-\frac{\theta}{2}x^2}, x\geq 0,\theta\geq 0$.

(iii) The Makeham distribution: $\bar{F}_3(x)=e^{-x-\theta(x+e^{-x}-1)}, x\geq 0,\theta\geq 0$.

Note that For $\theta=1,\bar{F}_1(x)$ reduces to exponential distribution while for $\theta=0,\bar{F}_2(x)$and $\bar{F}_3(x)$ reduces to exponential distribution. The PAE is defined by:

$$PAE(\Delta_n^{(1)}(s))=\frac{1}{\sigma_\circ(s)}\left|\frac{d}{d\theta}\delta_\theta^{(1)}(s)\right|_{\theta\to\theta_\circ}. \tag{14}$$

At $s=5$,

$$\delta_\theta^{(1)}(s)=\frac{1}{2s^2}\mu_{\theta(2)}-\frac{1}{6s}\mu_{\theta(3)}\zeta_\theta(s)-\frac{1}{s^4}\zeta_\theta(s)-\frac{1}{s^3}\mu_\theta+\frac{1}{s^4},$$

where

$$\begin{aligned}\mu_\theta &= \int_0^\infty \overline{F}_\theta(u)du, \mu_{\theta(2)}=2\int_0^\infty u\overline{F}_\theta(u)du, \mu_{\theta(3)}=3\int_0^\infty u^2\overline{F}_\theta(u)du,\\ \zeta_\theta(s) &= E_\theta(e^{-su})=\int_0^\infty e^{-su}dF_\theta(u)=-\int_0^\infty e^{-su}d\overline{F}_\theta(u).\end{aligned}$$

Hence,

$$\frac{d}{d\theta}\delta_\theta^{(1)}(s)=\frac{1}{2s^2}\grave{\mu}_{\theta(2)}-\frac{1}{6s}(\mu_{\theta(3)}\grave{\zeta}_\theta(s)+\grave{\mu}_{\theta(3)}\zeta_\theta(s))-\frac{1}{s^4}\grave{\zeta}_\theta(s)-\frac{1}{s^3}\grave{\mu}_\theta,$$

where

$$\begin{aligned}\grave{} &= \frac{d}{d\theta}, \grave{\mu}_\theta=\int_0^\infty \grave{\overline{F}}_\theta(u)du, \grave{\mu}_{\theta(2)}=2\int_0^\infty u\grave{\overline{F}}_\theta(u)du,\\ \grave{\mu}_{\theta(3)} &= 3\int_0^\infty u^2\grave{\overline{F}}_\theta(u)du, \grave{\zeta}_\theta(s)=-\int_0^\infty e^{-su}d\grave{\overline{F}}_\theta(u).\end{aligned}$$

Upon using the definition of the PAE in (14), we obtain

$$PAE(\delta^{(1)})=\frac{1}{\sigma_0}\mid\frac{1}{2s^2}\grave{\mu}_{\theta(2)}-\frac{1}{6s}(\mu_{\theta(3)}\grave{\zeta}_\theta(s)+\grave{\mu}_{\theta(3)}\zeta_\theta(s))-\frac{1}{s^4}\grave{\zeta}_\theta(s)-\frac{1}{s^3}\grave{\mu}_\theta\mid_{\theta\to\theta_0}.$$

Table 1: Comparison between the PAEs of our test and some other tests

Test	Weibull	LFR	Makeham
(Kango 1993)	0.132	0.433	0.144
(Mugdadi and Ahmad 2005)	0.170	0.408	0.039
(Abdel Aziz 2007)	0.223	0.535	0.184
(Mahmoud and Abdul Alim 2002, 2003, 2008)	0.050	0.217	0.144
Our test $\Delta_n^{(1)}(5)$	1.046	0.932	0.233

Table 2: Critical Values of the statistic $\Delta_n^{(1)}(5)$

n	90%	95%	99%
5	0.039527	0.051659	0.081158
10	0.024073	0.029370	0.041747
15	0.019395	0.023351	0.032516
20	0.016674	0.019865	0.026928
25	0.015041	0.017706	0.023554
30	0.013927	0.016372	0.021441
35	0.013179	0.015419	0.019853
39	0.012648	0.014794	0.018750
40	0.012483	0.014519	0.018562
45	0.011991	0.013944	0.018084
50	0.011490	0.013433	0.017040

When $s = 5$; this leads to

$$PAE\ [\Delta_n^{(1)}(5), Weibull] = 1.04561,\ PAE[\Delta_n^{(1)}(5), LFR] = 0.931891 \text{ and}$$

$$PAE\ [\Delta_n^{(1)}(5), Makeham] = 0.232973,\ \text{where } \sigma_0(5) = 0.0894239.$$

From Table 1, it is obvious that $\Delta_n^{(1)}(5)$ is better than the other tests based on the PAEs.

6. Monte Carlo null distribution critical points

In this section the Monte Carlo null distribution critical points of $\Delta_n^{(1)}(5)$ are simulated based on 10000 generated samples of size $n = 5(5)35, 39, 40(5)50$. From the standard exponential distribution by using Mathematica 8 program. Table 2 gives the upper percentile points of statistic $\Delta_n^{(1)}(5)$ for different confidence levels $90\%, 95\%$ and 99%.

From Table 2, it is obvious that the critical values are decreasing as the samples size increasing and they are increasing as the confidence levels increasing.

7. Power estimates of the test $\Delta_n^{(1)}(5)$

In this section the power of our test $\Delta_n^{(1)}(5)$ will be estimated at $(1-\alpha)\%$ confidence level, $\alpha = 0.05$ with suitable parameters values of θ at $n = 10, 20$ and 30 with respect to three alternatives Linear failure rate (LFR) , Weibull and Gamma distributions based on 10000 samples.

Table 3 shows that the power estimates of our test $\Delta_n^{(1)}(5)$ are good power for all alternatives and increases when the value of the parameter θ and the sample sizes increasing.

Table 3: The Power Estimates of $\Delta_n^{(1)}(5)$

n	θ	LFR	Weibull	Gamma
10	2	0.6674	0.9978	0.9922
	3	0.8501	1.0000	0.9991
	4	0.9324	1.0000	1.0000
20	2	0.9360	1.0000	0.9888
	3	0.9816	1.0000	0.9988
	4	0.9911	1.0000	0.9998
30	2	0.9828	1.0000	0.9861
	3	0.9944	1.0000	0.9992
	4	0.9983	1.0000	1.0000

8. Applications to real data

In this section, we apply our test to some real data-sets at 95% confidence level.

1- Consider the data in (Al-Gashgari, Shawky, and Mahmoud 2016) which represent 39 liver cancers patients taken from Elminia cancer center Ministry of Health – Egypt, which entered in (1999). The ordered life times (in days)

10	14	14	14	14	14	15	17	18	20
20	20	20	20	23	23	24	26	30	30
31	40	49	51	52	60	61	67	71	74
75	87	96	105	107	107	107	116	150	

In this case, $\Delta_n^{(1)}(5) = 0.0000132958$ which is less than the corresponding critical value in Table 2, then we reject H_1 which states that the data set have NBRUL property.

2- Consider the real data-set given in (Grubbs 1971) and have been used in (Shapiro 1995). This data set gives the times between arrivals of 25 customers at a facility.

1.80	2.89	2.93	3.03	3.15	3.43	3.48	3.57	3.85	3.92
3.98	4.06	4.11	4.13	4.16	4.23	4.34	4.37	4.53	4.62
4.65	4.84	4.91	4.99	5.17					

Since $\Delta_n^{(1)}(5) = 0.00119843$ and this value less than the corresponding critical value in Table 2. Then we conclude that this data set have the exponential property.

3- Consider the data in (Abouammoh *et al.* 1994). These data represent 40 patients suffering from blood cancer from one of the Ministry of Health Hospital in Saudi Arabia and the ordered life times (in days):

115	181	255	418	441	461	516	739	743	789
807	865	924	983	1024	1062	1063	1169	1191	1222
1222	1251	1277	1290	1357	1369	1408	1455	1478	1549
1578	1578	1599	1603	1604	1696	1735	1799	1815	1852

Since $\Delta_n^{(1)}(5) = 1.81685 \times 10^{-8}$ and this value less than the corresponding critical value in Table 2. Then we conclude that this data set have the exponential property.

9. Conclusion

The NBRUL class of life distributions is considered. The moments inequalities are derived. A new test statistics for exponentiality versus NBRUL class is proposed based on the moment inequalities. Quality criteria of the test is shown by the famous criterion which is Pitman asymptotic efficiency. The upper percentiles and the power of the proposed test are calculated and tabulated. Our test is applied to some real data to show the usefulness of the test.

References

Abdel Aziz A (2007). "On Testing Exponentiality Against RNBRUE Alternatives." *Applied Mathematical Sciences*, **35**, 1725–1736.

Abouammoh A, Abdulghani S, Qamber I (1994). "On Partial Orderings and Testing of New Better Than Renewal Used Classes." *Reliability Engneering and System Safety*, **43**, 37–41.

Abouammoh A, Ahmed A (1988). "The New Better Than Used Class of Life Distribution." *Advances in Applied Probability*, **20**, 237–240.

Abouammoh A, Ahmed A (1992). "On Renewal Failure Rate Classes of Life Distributions." *Statistics and Probability Letters*, **14**, 211–217.

Abouammoh A, Newby J (1989). "On Partial Ordering the Tests of Generalized New Better Than Used Class of Life Distribution." *Reliability Engneering and System Safety*, **25**, 207–217.

Abu-Youssef S (2009). "A Goodness of Fit Approach to Monotone Variance Residual Life Class of Life Distributions." *Applied Mathematical Sciences*, **15**, 715–724.

Ahmad I (1994). "Class of Statistics Useful in Testing Increasing Failure Rate Average and New Better than Used Life Distributions." *Journal of Statistical Plannng and Inference*, **41**, 141–149.

Ahmad I (2001). "Moments Inequalities of Aging Families of Distributions with Hypothesis Testing Applications." *Journal of Statistical Plannng and Inference*, **92**, 121–132.

Ahmad I, Alwasel I, Mugdadi A (2001). "A Goodness of Fit Approach to Major Life Testing Problems." *International Journal of Reliability Application*, **2**, 81–97.

Ahmad I, Mugdadi A (2004). "Further Moments Inequalities of Life Distributions with Hypothesis Testing Applications: the IFRA, NBUC, DMRL Classes." *Journal of Statistical Plannng and Inference*, **120**, 1–12.

Al-Gashgari F, Shawky A, Mahmoud M (2016). "A Nonparametric Test for Testing Exponentiality against NBUCA Class of Life Distributions Based on Laplace Transform." *Quality and Reliability Engineering International*, **32**, 29–36.

Atallah M, Mahmoud M, Al-Zahrani B (2014). "A New Test for Exponentiality versus NBUmgf Life Distributions Based on Laplace Transform." *Quality and Reliability Engineering International*, **30**, 1353–1359.

Barlow R, Proschan F (1981). *Statistical Theory of Reliability and Life Testing*. To Begin with Silver Spring, M D.

Bryson M, Siddiqui M (1969). "Some Criteria for Ageing." *Journal of the American Statistical Association*, **64**, 1472–1483.

Deshpande J, Kochar S, Singh H (1986). "Aspects of Positive Ageing." *Journal of Applied Probability*, **23**, 748–758.

Grubbs F (1971). "Fiducial Bounds on Reliability for the Two Parameter Negative Exponential Distribution." *Technometrics*, **13**, 873–876.

Ismail A, Abu-Youssef S (2012). "A Goodness of Fit Approach to the Class of Life Distribution with Unknown Age." *Quality and Reliability Engineering International*, **25**, 761–766.

Kango A (1993). "Testing for New is Better than Used." *Communications in Statistics-Theory and Methods*, **12**, 311–321.

Kumazawa Y (1983). "A Class of Tests Statistics for Testing whether New Is Better than Used." *Communications in Statistics-Theory and Methods*, **12**, 311–321.

Lee A (1989). *U-Statistics*. Marcel Dekker, New York.

Mahmoud M, Abdul Alim N (2002). "On Testing Exponentiality Against NBARFR Life Distributions." *Statistica*, **4**, 619–631.

Mahmoud M, Abdul Alim N (2003). "On Testing Exponentiality against NBURFR Class of Life Distributions." *International Jornal of Reliability and Applications*, **4**, 57–69.

Mahmoud M, Abdul Alim N (2008). "A Goodness of Fit Approach to for Testing NBUFR (NWUFR) and NBAFR (NWAFR) Properties." *International Jornal of Reliability and Applications*, **9**, 125–140.

Mahmoud M, Diab L, Kayid M (2009). "Moments Inequalities for NBUL Distributions with Hypotheses Testing Applications." *Contemporary Engineering Sciences*, **7**, 319–332.

Mahmoud M, EL-arishy S, Diab L (2003). "Moments Inequalities for Testing New Renewal Better than Used and Renewal New Better than Used Classes." *International Journal of Reliability and Applications*, **4**, 97–123.

Mahmoud M, EL-Sagheer R, Etman W (2016). "Testing Exponentiality aganist New Better than Renewal Used in Laplace Transform Order." *Journal of Statistics Applications & Probability*, **5**(2), 279–285.

Mahmoud M, Moshref M, Mansour M (2015). "A New Class of Life Distributions with Hypotheses Testing Applications." *Journal of Applied Probability and Statistics*, **1**, 35–47.

Mahmoud M, Rady D (2013). "Testing Exponentiality Against HNBUET Class Based on Goodness of Fit Approach." *Al-Azhar Bulletin of Science*, **24**(1), 81–92.

Mugdadi A, Ahmad I (2005). "Moment Inequalities Derived from Comparing Life with Its Equilibrium Form." *Journal of Statistical Plannng and Inference*, **134**, 303–317.

Shapiro S (1995). *Goodness of Fit Tests in The Exponential Distribution Theory Methods and Applications*. Balakrishnan N, Basu AP. Editors, Gordon and Breach, Amsterdam.

Affiliation:

Mohamed A. W. Mahmoud
Mathematics Department, Faculty of Science,
AL-Azhar University, Nasr City (11884),
Cairo, Egypt.
E-mail: mawmahmoud11@azhar.edu.eg

R. M. EL-Sagheer
Mathematics Department, Faculty of Science,
AL-Azhar University, Nasr City (11884),
Cairo, Egypt.
E-mail: rashadmath@azhar.edu.eg

W. B. H. Etman
Mathematics Department, Faculty of Science,
AL-Azhar University, Nasr City (11884),
Cairo, Egypt.
E-mail: w.h.etman2012@gmail.com

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 47 *Submitted:* 2016-07-09
February 2018 *Accepted:* 2016-08-19

ISSN 1026–597X

Contents