Austrian Journal of Statistics

AUSTRIAN STATISTICAL SOCIETY

Volume 46, Number 1, 2017

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

The author(s) retain any copyright on the submitted material. The contributors grant the journal the right to publish, distribute, index, archive and publicly display the article (and the abstract) in printed, electronic or any other form.

Manuscripts should be unpublished and not be under consideration for publication elsewhere. By submitting an article, the author(s) certify that the article is their original work, that they have the right to submit the article for publication, and that they can grant the above license.

Austrian Journal of Statistics

Volume 46, Number 1, 2017

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 46, Heft 1, 2017

ÖSTERREICHISCHE STATISTISCHE GESELLSCHAFT

Impressum

Printed by Statistics Austria, A-1110 Vienna

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

Contents

Editorial

This volume include five scientific papers and one book review.

The first contribution deals with multivariate multi-sample rank tests, the second contribution with tests for ranking data. Both works do not only show new theory for the proposed tests, but they also include interesting applications.

The last three contributions introduce a new distributions useful in particular areas such as reliability, life-time analysis, finance and assurance.

Walter Krämer was recently awarded with the Bruckmann-Price of the Austrian Statistical Society. His new book "Statistik für alle - Die 101 wichtigsten Begriffe anschaulich erklärt" is reviewed by Andreas Quatember, who was itself awarded with the Bruckmann-Price. Worth reading!

Matthias Templ
(Editor-in-Chief)

Zurich University of Applied Sciences
Rosenstrasse 3
CH–8400 Winterthur, Switzerland
E-mail: matthias.templ@gmail.com

Winterthur, 5. Januar 2017

Austrian Journal of Statistics
February 2017, Volume 46, 3–13.
http://www.ajs.or.at/
doi:10.17713/ajs.v46i1.130

The Multivariate Multisample Nonparametric Rank Statistics for the Location Alternatives

Hidetoshi Murakami
Tokyo University of Science

Abstract

Multisample testing problems are among the most important topics in nonparametric statistics. Various nonparametric tests have been proposed for multisample testing problems involving location parameters, and the analysis of multivariate data is important in many scientific fields. One type of multivariate multisample testing problem based on Jurečková-Kalina-type rank of distance is discussed in this paper. A multivariate Kruskal-Wallis-type statistic is proposed for testing the location parameter with both equal and unequal sample sizes. Simulations are used to compare the power of proposed nonparametric statistics with the Wilks' λ, the Pillai's trace and the Lawley-Hotelling trace for various population distributions.

Keywords: Jurečková-Kalina-type ranks of distances, multivariate multisample rank test, power comparison.

1. Introduction

Testing hypotheses is one of the most important challenges in nonparametric statistics. Various nonparametric tests have been proposed for one-sample, two-sample and multisample testing problems involving the location, scale, location-scale and other parameters. Recent progress in computerized measurement technology has permitted the accumulation of multivariate data, increasing the importance of multivariate data in many scientific fields. When we consider testing a multivariate multisample hypothesis, one of the most important statistical procedures, we naturally consider vector-valued observations. If only a marginal study of each component of these vectors is carried out, then outliers, strongly influential points and useful relationships among variables may not be detected. Thus, a multivariate examination of the data is necessary. However, in many applications, the underlying distribution is not adequately understood to assume normality or any other specific distribution, and the nonparametric test statistic must be used. Because it is important to determine how to represent ranks for multivariate data in nonparametric statistics, various researchers have proposed the distances of observation for the rank tests. Jurečková and Kalina (2012) proposed a rank test based on observation distances for two-sample problems with a discussion about the unbiasedness of test statistics under the alternatives hypothesis.

Recently, Murakami (2015a) applied the Jurečková-Kalina rank of distance to the Ansari-

Bradley, Lepage and Baumgartner statistics. In addition, Murakami (2015b) considered the use of Jurečková-Kalina-type rank of distance with the Wilcoxon-type statistic. We extend this concept of rank of distance to a multisample setting. In Section 2, we introduce multivariate multisample nonparametric statistics based on Jurečková-Kalina-type rank of distance. We consider the Kruskal-Wallis test (Gibbons and Chakraborti 2010), the multisample median test (Hájek et al. 1999), the multisample Lepage-type test (Rublík 2007), the Wilks' λ (Rencher 1998), the Pillai's trace (Rencher 1998) and the Lawley-Hotelling trace (Rencher 1998) in this paper. In addition, we propose another type of multivariate Kruskal-Wallis test. In Section 3, we compare the powers of the proposed test with the multivariate multisample parametric and nonparametric tests for various distributions by using simulation studies. The simulations include 100,000 Monte Carlo replications. Conclusions are stated in Section 4.

2. Multivariate multisample nonparametric statistics

In this section, we introduce the multivariate multisample nonparametric statistics for the vector-valued observations. MANOVA is one of the most important types of statistical procedures in many scientific fields, especially in biometry. However, in many applications, the underlying distribution is not adequately understood to assume normality or some other specific distribution. Additionally, if we carry out only a marginal component of the vector-valued observation, we may not detect outliers, strongly influential points and useful relationships among variables. Then, we require to determine how to represent ranks for the vector-valued observation.

Let $\{\boldsymbol{x}_{ij}; i = 1,\dots,k,\ j = 1,\dots,n_i\}$ be k independent samples from p-variate populations having continuous unknown distribution functions $F_i^{(p)}$. Under these circumstances, we are interested in the following hypothesis:

$$\begin{aligned} H_0 &: \ F_1^{(p)} = F_2^{(p)} = \cdots = F_i^{(p)} \\ H_1 &: \ \text{not } H_0. \end{aligned}$$

To test this hypothesis, we utilize the multivariate multisample nonparametric statistics. For multivariate data in nonparametric statistics, it is important to determine how to represent a rank of the vectored-value observation. Jurečková and Kalina (2012) proposed a distance of observation for $k = 2$, and the proposed rank of distance was found to be invariant for a shifted location parameter. To introduce their rank of distance, let

$$\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_{n_1+n_2}) = (\boldsymbol{x}_{11},\dots,\boldsymbol{x}_{1n_1},\boldsymbol{x}_{21},\dots,\boldsymbol{x}_{2n_2})$$

denote the pooled sample. For every fixed j and under fixed $\boldsymbol{x}_{1j}$, $1 \le j \le n_1$, they considered the distances $\{\ell^*_{jt} = L(\boldsymbol{x}_{1j},\boldsymbol{\zeta}_t); t = 1,\dots,n_1+n_2, j \ne t\}$, where $L(\cdot,\cdot)$ denotes Euclidean distance. Then, conditionally given $\boldsymbol{x}_{1j}$, the vector $\{\ell^*_{jt} = L(\boldsymbol{x}_{1j},\boldsymbol{\zeta}_t); t = 1,\dots,n_1, j \ne t\}$ is a random sample from a first population $G_1^{(p)}$, while $\{\ell^*_{jt} = L(\boldsymbol{x}_{1j},\boldsymbol{\zeta}_t); t = n_1+1,\dots,n_1+n_2\}$ is a random sample from a second population $G_2^{(p)}$. Jurečková and Kalina (2012) decided to work with the ranks of ℓ^*_{jt}, $t = 1,\dots,n_1+n_2$, $j \ne t$.

Herein, we extend this concept of rank of distance to a multisample setting. Let

$$\boldsymbol{Z} = (\boldsymbol{Z}_1,\dots,\boldsymbol{Z}_N) = (\boldsymbol{x}_{11},\dots,\boldsymbol{x}_{1n_1},\boldsymbol{x}_{21},\dots,\boldsymbol{x}_{2n_2},\dots,\boldsymbol{x}_{k1},\dots,\boldsymbol{x}_{kn_k})$$

denote the pooled sample with $N = n_1 + \cdots + n_k$. We consider Jurečková-Kalina-type of distances such that $\{\ell_{st} = L(\boldsymbol{Z}_s,\boldsymbol{Z}_t); s,t = 1,\dots,N, s \ne t\}$ for every fixed s. Conditionally given as $\boldsymbol{x}_{ij}$, the vector $\{\ell_{u(i,j)v(i)};\ u(i,j) = \sum_{q=1}^{i-1} n_q + j, v(i) = \sum_{q=1}^{i-1} n_q + r,\ r = 1,\dots,n_i,\ r \ne j\}$ is then a random sample based on the distribution function $F_i^{(p)}(z|\boldsymbol{x}_{ij}) = G_i^{(p)}$, where we define $n_q = 0$ for $i = 1$. Assuming that the distribution functions $G_i^{(p)}$ are continuous, the rank of ℓ_{st} is denoted as

$$R_{st} = (R_{s1},\dots,R_{s,t-1},R_{s,t+1},\dots,R_{sN}),$$

where $s, t = 1, \ldots, N$ and $t \neq s$ for fixed s. We then consider the following rank statistics:

- The Kruskal-Wallis statistic T_1 (Gibbons and Chakraborti 2010):

$$T_1 := T_s^{(p)} = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left(W_{si} - \frac{N+1}{2} \right)^2,$$

  where

$$W_{si} = \frac{1}{n_i} \sum_{j=1}^{n_i} R_{su(i,j)}.$$

  Exact probabilities for the Kruskal-Wallis statistic are listed in Gibbons and Chakraborti (2010) for small sample sizes. The limiting distribution for the Kruskal-Wallis statistic is the chi-square distribution with $k-1$ degrees of freedom.

- The multisample median statistic T_2 (Hájek et al. 1999):

$$T_2 := T_s^{(p)} = 4 \sum_{i=1}^{k} \frac{1}{n_i} \left(A_{si} - \frac{n_i}{2} \right)^2,$$

  where

$$A_{si} = \sum_{j=1}^{n_i} \frac{1}{2} \left\{ \operatorname{sign} \left(\sum_{j=1}^{n_i} R_{su(i,j)} - \frac{N+1}{2} \right) + 1 \right\}.$$

  Exact probabilities for the multisample median statistic are listed in Jorn and Klotz (2002) for small sample sizes. The limiting distribution for the multisample median statistic is the chi-square distribution with $k-1$ degrees of freedom.

- The multisample Lepage-type statistic T_3 (Rublík 2007):

$$T_3 := T_s^{(p)} = T_1 + T_4,$$

  where

$$T_4 := T_s^{(p)} = \frac{180}{N(N+1)(N^2-4)} \sum_{i=1}^{k} n_i \left(M_{si} - \frac{N^2-1}{12} \right)^2$$

  and

$$M_{si} = \frac{1}{n_i} \sum_{j=1}^{n_i} \left(R_{su(i,j)} - \frac{N+1}{2} \right)^2.$$

  Exact probabilities for the multisample Lepage-type statistic are listed in Murakami (2008) for small sample sizes. The limiting distribution for the multisample Lepage-type statistic is the chi-square distribution with $2(k-1)$ degrees of freedom.

Note that in a one-dimensional setting, the multisample median test uses less information than the Kruskal-Wallis test does, and may therefore be less powerful. The asymptotic relative efficiency of the multisample median test is $2/3$ with respect to the Kruskal-Wallis test for a normal distribution (e.g. Gibbons and Chakraborti 2010). The multisample version of the Lepage statistic is preferable for location, scale and location-scale parameters. However, Rublík (2007) showed that the multisample version of a combination of the Kruskal-Wallis and multisample Mood statistics is more efficient than the multisample Lepage statistic for shifted location, scale and location-scale parameters with various distributions.

The statistic $T_s^{(p)}$ is equally distributed for $s = 1, \ldots, N$ under the null hypothesis. Randomization of $T_1^{(p)}, \ldots, T_N^{(p)}$ maintains the simple structure of the test. Thus, we obtain

$$\Pr(T^{(p)} = T_s^{(p)}) = \frac{1}{N}, \quad s = 1, \ldots, N, \tag{1}$$

where the randomization in (1) is independent of the observations. For any C,

$$\Pr(T^{(p)} > C) = \frac{1}{N} \sum_{s=1}^{N} \Pr(T_s^{(p)} > C),$$

and the statistic rejects H_0 if $T^{(p)} > C$.

Herein we suggest another multivariate Kruskal-Wallis-type test, namely $V^{(p)}$, as follows:

$$V^{(p)} = \max_{1 \le s \le N} T_1.$$

3. Simulation study

We employed R software to investigate the behavior of the T_1, T_2 and T_3 statistics in simulation studies. Additionally, we used the Wilks' λ, namely W_λ, the Pillai's trace, specifically PT, and the Lawley-Hotelling trace, specifically LH, as a classical MANOVA test (Rencher 1998). The simulations included 100,000 replications, and the significance level was 5%. To compare the power of the classical MANOVA test and tests based on the multivariate nonparametric statistics, we carried out a simulation study of different populations with various distributions. In this paper, we have focused on the cases $(n_1, n_2, n_3) = (5, 5, 5)$, $(15, 10, 5)$ and $(20, 20, 20)$ for $p = 2$ and 3 and the following distributions:

- $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$: the multivariate normal distribution.
- $t(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \delta_i)$: the multivariate t distribution with δ degrees of freedom.
- $LN(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$: the multivariate lognormal distribution.

To generate random numbers, we used the packages "mvrnorm," "rmt," and "rlnorm.rplus" for the multivariate normal, multivariate t and multivariate lognormal distributions, respectively. We define a p-dimensional matrix as follows:

$$I^{(2)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \rho_2^{(2)} = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}, \quad \rho_3^{(2)} = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix},$$

$$I^{(3)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \rho_2^{(3)} = \begin{pmatrix} 1 & 0.4 & 0.2 \\ 0.4 & 1 & 0.3 \\ 0.2 & 0.3 & 1 \end{pmatrix}, \quad \rho_3^{(3)} = \begin{pmatrix} 1 & 0.5 & 0.3 \\ 0.5 & 1 & 0.4 \\ 0.3 & 0.4 & 1 \end{pmatrix}.$$

In this paper, we assume $\boldsymbol{\mu}_1 = \mathbf{0}$ and $\boldsymbol{\Sigma}_1 = I^{(p)}$, and we consider the following cases for the multivariate normal, multivariate t and multivariate lognormal distributions.

Case 1	
$\boldsymbol{\mu}_2 = \mathbf{0}$	$\boldsymbol{\mu}_3 = \mathbf{0}$
$\boldsymbol{\Sigma}_2 = I^{(p)}$	$\boldsymbol{\Sigma}_3 = I^{(p)}$

Case 2	
$\boldsymbol{\mu}_2 = \mathbf{0}$	$\boldsymbol{\mu}_3 = \mathbf{0}$
$\boldsymbol{\Sigma}_2 = \rho_2^{(p)}$	$\boldsymbol{\Sigma}_3 = \rho_3^{(p)}$

Case 3	
$\boldsymbol{\mu}_2 = \mathbf{1.0}$	$\boldsymbol{\mu}_3 = \mathbf{1.0}$
$\boldsymbol{\Sigma}_2 = I^{(p)}$	$\boldsymbol{\Sigma}_3 = I^{(p)}$

Case 4	
$\boldsymbol{\mu}_2 = \mathbf{1.0}$	$\boldsymbol{\mu}_3 = \mathbf{1.0}$
$\boldsymbol{\Sigma}_2 = \rho_2^{(p)}$	$\boldsymbol{\Sigma}_3 = \rho_3^{(p)}$

Case 5	
$\boldsymbol{\mu}_2 = \mathbf{1.0}$	$\boldsymbol{\mu}_3 = \mathbf{2.0}$
$\boldsymbol{\Sigma}_2 = I^{(p)}$	$\boldsymbol{\Sigma}_3 = I^{(p)}$

Case 6	
$\boldsymbol{\mu}_2 = \mathbf{1.0}$	$\boldsymbol{\mu}_3 = \mathbf{2.0}$
$\boldsymbol{\Sigma}_2 = \rho_2^{(p)}$	$\boldsymbol{\Sigma}_3 = \rho_3^{(p)}$

In the case of $(n_1, n_2, n_3) = (5, 5, 5)$, we used the exact critical value of the T_1, T_2 and T_3 statistics by Gibbons and Chakraborti (2010), Jorn and Klotz (2002) and Murakami (2008), respectively. Since it is difficult to evaluate the exact critical value of the statistic for the large sample sizes, we estimated the critical value via a permutation approach for $(n_1, n_2, n_3) = (15, 10, 5)$ and $(20, 20, 20)$. Additionally, we apply the following method to the $V^{(p)}$ statistic. Our method for estimating the critical value is as follows:

1. Construct a dataset $\boldsymbol{Z}$ by generating N integers from 1 to N (without ties) for each dimension.
2. Calculate the statistics T_1, T_2, T_3 and $V^{(p)}$.
3. Construct a permutation dataset $\boldsymbol{Z}^*$.
4. Calculate the T_1, T_2, T_3 and $V^{(p)}$ statistics from the dataset $\boldsymbol{Z}^*$.
5. Independently repeat steps 3 and 4 B times.
6. Sort the statistics $T_{m(1)}, \ldots, T_{m(B)}$, $m = 1, 2, 3$ and $V^{(p)}_{(1)}, \ldots, V^{(p)}_{(B)}$.

$T_{m(CV)}$ and $V^{(p)}_{(CV)}$ are then the estimated critical value of the statistics, where $CV = B \times \alpha\%$. We simulated $B = 100,000$ replications in this study.

Table 1 lists the simulation results for the multivariate normal distribution.

Table 1 shows that the classical MANOVA tests were more powerful than the multivariate multisample nonparametric statistics. Compared with nonparametric statistics, the proposed statistic was more efficient than the randomized nonparametric statistics were. Therefore, the $V^{(p)}$ statistic was more effective than the other nonparametric statistics for parameters associated with the multivariate normal distribution.

For a non-normal distribution, we used the multivariate t distribution with 2 degrees of freedom, and the results are listed in Table 2.

Table 2 shows that the classical MANOVA tests did not maintain 5% significance levels (not conservative) under the null hypothesis for unequal sample sizes. The non-conservative test is meaningless for testing the hypothesis. Moreover, the suggested statistic was more powerful than the parametric and nonparametric statistics. Therefore, the $V^{(p)}$ statistic was more effective than the other statistics were for parameters associated with the multivariate t distribution.

We used the multivariate lognormal distribution to simulate an asymmetrical distribution; the results are listed in Table 3.

The results presented in Table 3 reveal the following facts: The classical MANOVA tests did not maintain 5% significance levels (not conservative) under the null hypothesis for unequal

Table 1: Simulated power for the multivariate normal distributions

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
			Case of $n_1 = n_2 = n_3 = 5$ for $p = 2$			
W_λ	0.050	0.052	0.363	0.306	0.845	0.767
PT	0.050	0.053	0.352	0.294	0.820	0.740
LH	0.050	0.052	0.366	0.311	0.853	0.778
T_1	0.049	0.049	0.176	0.178	0.458	0.440
T_2	0.039	0.039	0.120	0.128	0.325	0.310
T_3	0.050	0.051	0.133	0.134	0.354	0.338
$V^{(2)}$	0.048	0.049	0.217	0.215	0.640	0.582
			Case of $n_1 = 15$, $n_2 = 10$ and $n_3 = 5$ for $p = 2$			
W_λ	0.047	0.047	0.814	0.771	0.994	0.986
PT	0.047	0.048	0.815	0.772	0.994	0.986
LH	0.048	0.049	0.820	0.778	0.994	0.987
T_1	0.050	0.051	0.390	0.387	0.668	0.649
T_2	0.040	0.041	0.283	0.278	0.525	0.501
T_3	0.050	0.052	0.307	0.307	0.604	0.591
$V^{(2)}$	0.050	0.051	0.700	0.668	0.964	0.938
			Case of $n_1 = n_2 = n_3 = 20$ for $p = 2$			
W_λ	0.049	0.050	0.988	0.973	1.000	1.000
PT	0.049	0.050	0.987	0.972	1.000	1.000
LH	0.049	0.050	0.988	0.973	1.000	1.000
T_1	0.050	0.051	0.552	0.564	0.876	0.866
T_2	0.049	0.051	0.460	0.472	0.808	0.794
T_3	0.050	0.054	0.494	0.509	0.835	0.831
$V^{(2)}$	0.051	0.051	0.955	0.940	1.000	1.000

sample sizes. The $V^{(p)}$ statistic was the most powerful statistic for the shifted location parameters when the sample sizes were equal and unequal. Therefore, the $V^{(p)}$ statistic was more effective than the other parametric and nonparametric statistics for parameters associated with the multivariate lognormal distribution.

4. Concluding remarks

In this paper, we considered multivariate multisample nonparametric statistics by applying Jurečková-Kalina-type rank of distance. Simulation studies showed that the multivariate Kruskal-Wallis-type statistic, named $V^{(p)}$, was more powerful than the Kruskal-Wallis, multivariate multisample median and Lepage-type statistics for shifted location parameters under the multivariate normal, t and lognormal distributions. Additionally, the proposed statistic was more efficient than the classical MANOVA test for equal and unequal sample sizes with non-normal distributions. As ties occur frequently in practice, in future research we should investigate the powers of multivariate multisample nonparametric statistics under multivariate discrete distributions.

Acknowledgements

The author would like to thank the editor and the referees for their valuable comments and suggestions. The author appreciates that this research is supported by the Grant-in-Aid for Young Scientists B of JSPS, KAKENHI Number 26730025.

Table 1: Continued for the multivariate normal distributions

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
	Case of $n_1 = n_2 = n_3 = 5$ for $p = 3$					
W_λ	0.049	0.053	0.347	0.228	0.849	0.671
PT	0.049	0.053	0.338	0.228	0.778	0.608
LH	0.049	0.053	0.416	0.287	0.906	0.755
T_1	0.049	0.049	0.201	0.208	0.536	0.505
T_2	0.039	0.039	0.136	0.156	0.399	0.375
T_3	0.050	0.051	0.151	0.152	0.428	0.392
$V^{(3)}$	0.050	0.050	0.282	0.276	0.803	0.700
	Case of $n_1 = 15$, $n_2 = 10$ and $n_3 = 5$ for $p = 3$					
W_λ	0.050	0.047	0.908	0.821	0.999	0.994
PT	0.050	0.047	0.907	0.821	0.999	0.993
LH	0.050	0.048	0.909	0.823	0.999	0.994
T_1	0.050	0.052	0.434	0.428	0.732	0.705
T_2	0.040	0.042	0.324	0.316	0.592	0.554
T_3	0.050	0.057	0.350	0.353	0.673	0.660
$V^{(3)}$	0.049	0.050	0.835	0.770	0.995	0.975
	Case of $n_1 = n_2 = n_3 = 20$ for $p = 3$					
W_λ	0.050	0.052	0.999	0.985	1.000	1.000
PT	0.050	0.052	0.999	0.983	1.000	1.000
LH	0.050	0.052	0.999	0.986	1.000	1.000
T_1	0.049	0.053	0.585	0.600	0.917	0.907
T_2	0.049	0.053	0.498	0.519	0.857	0.844
T_3	0.050	0.061	0.529	0.547	0.882	0.880
$V^{(3)}$	0.050	0.051	0.992	0.981	1.000	1.000

References

Gibbons JD, Chakraborti S (2010). *Nonparametric Statistical Inference, 5th Edition*. CRC Press, New York.

Hájek J, Sidák Z, Sen PK (1999). *Theory of Rank Tests, 2nd Edition*. Academic Press, San Diego.

Jorn H, Klotz J (2002). "Exact Distribution of the K-Sample Mood and Brown Median Test." *Journal of Nonparametric Statistics*, **14**, 249–257.

Jurečková J, Kalina J (2012). "Nonparametric Multivariate Rank Tests and Their Unbiasedness." *Bernoulli*, **18**, 229–251.

Murakami H (2008). "Table for Exact Critical Values of Multisample Lepage Type Statistics when $k = 3$." *Preprint Series of CHUO MATH*, **80**, 1–6.

Murakami H (2015a). "A Randomized Baumgartner Statistic for Multivariate Two-Sample Testing Hypothesis." *Journal of Statistical Computation and Simulation*, **85**, 189–201.

Murakami H (2015b). "Power Comparison of Multivariate Wilcoxon-Type Tests Based on the Jurečková-Kalina's Ranks of Distances." *Communications in Statistics – Simulation and Computation*, **44**, 2176–2194.

Rencher AC (1998). *Multivariate Statistical Inference and Applications*. Wiley, Toront.

Rublík F (2007). "On the Asymptotic Efficiency of the Multisample Location-Scale Rank Tests and Their Adjustment for Ties." *Kybernetika*, **43**, 279–306.

Table 2: Simulated power for the multivariate t distribution

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Case of $n_1 = n_2 = n_3 = 5$ for $p = 2$						
W_λ	0.031	0.031	0.153	0.129	0.409	0.350
PT	0.032	0.033	0.153	0.128	0.400	0.341
LH	0.031	0.031	0.152	0.129	0.413	0.352
T_1	0.048	0.049	0.110	0.113	0.239	0.236
T_2	0.038	0.039	0.087	0.092	0.198	0.193
T_3	0.050	0.051	0.097	0.097	0.204	0.199
$V^{(2)}$	0.046	0.047	0.137	0.138	0.364	0.347
Case of $n_1 = 15$, $n_2 = 10$ and $n_3 = 5$ for $p = 2$						
W_λ	0.076	0.074	0.344	0.314	0.594	0.556
PT	0.077	0.074	0.345	0.315	0.595	0.556
LH	0.079	0.077	0.351	0.321	0.600	0.562
T_1	0.050	0.050	0.223	0.224	0.402	0.393
T_2	0.040	0.040	0.181	0.180	0.349	0.335
T_3	0.050	0.051	0.200	0.198	0.362	0.354
$V^{(2)}$	0.047	0.048	0.429	0.417	0.727	0.698
Case of $n_1 = n_2 = n_3 = 20$ for $p = 2$						
W_λ	0.034	0.034	0.470	0.406	0.854	0.805
PT	0.034	0.035	0.469	0.405	0.853	0.804
LH	0.034	0.034	0.472	0.408	0.855	0.806
T_1	0.049	0.050	0.360	0.376	0.695	0.695
T_2	0.049	0.050	0.333	0.353	0.674	0.673
T_3	0.049	0.053	0.340	0.348	0.676	0.669
$V^{(2)}$	0.050	0.050	0.758	0.730	0.996	0.992

Affiliation:

Hidetoshi Murakami
Department of Mathematical Information Science
Tokyo University of Science
1-3 Kagurazaka, Shinjyuku-ku, Tokyo, Japan E-mail: murakami@gug.math.chuo-u.ac.jp

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 46 *Submitted:* 2015-10-02
February 2017 *Accepted:* 2016-10-24

Table 2: Continued the multivariate t distribution

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
	Case of $n_1 = n_2 = n_3 = 5$ for $p = 3$					
W_λ	0.027	0.028	0.149	0.097	0.435	0.304
PT	0.031	0.032	0.160	0.107	0.417	0.301
LH	0.030	0.031	0.184	0.125	0.499	0.364
T_1	0.048	0.049	0.114	0.121	0.257	0.258
T_2	0.038	0.039	0.090	0.103	0.225	0.223
T_3	0.050	0.051	0.103	0.103	0.226	0.218
$V^{(3)}$	0.048	0.047	0.150	0.155	0.414	0.394
	Case of $n_1 = 15$, $n_2 = 10$ and $n_3 = 5$ for $p = 3$					
W_λ	0.082	0.078	0.446	0.370	0.728	0.638
PT	0.082	0.079	0.446	0.370	0.728	0.638
LH	0.083	0.079	0.448	0.372	0.729	0.640
T_1	0.050	0.052	0.229	0.238	0.421	0.413
T_2	0.041	0.041	0.179	0.187	0.373	0.353
T_3	0.050	0.054	0.222	0.225	0.396	0.386
$V^{(3)}$	0.041	0.044	0.480	0.459	0.786	0.741
	Case of $n_1 = n_2 = n_3 = 20$ for $p = 3$					
W_λ	0.034	0.034	0.603	0.449	0.931	0.858
PT	0.035	0.035	0.599	0.444	0.930	0.855
LH	0.034	0.034	0.607	0.454	0.932	0.860
T_1	0.049	0.051	0.364	0.400	0.715	0.723
T_2	0.049	0.050	0.332	0.371	0.694	0.701
T_3	0.050	0.056	0.370	0.390	0.716	0.713
$V^{(3)}$	0.043	0.044	0.832	0.801	0.999	0.996

Table 3: Simulated power for the multivariate lognormal distribution

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
			Case of $n_1 = n_2 = n_3 = 5$ for $p = 2$			
W_λ	0.031	0.031	0.159	0.125	0.522	0.400
PT	0.034	0.033	0.169	0.134	0.491	0.378
LH	0.031	0.031	0.153	0.120	0.535	0.410
T_1	0.048	0.049	0.160	0.146	0.443	0.399
T_2	0.038	0.040	0.115	0.110	0.345	0.301
T_3	0.050	0.051	0.143	0.128	0.445	0.390
$V^{(2)}$	0.044	0.045	0.188	0.170	0.580	0.504
			Case of $n_1 = 15$, $n_2 = 10$ and $n_3 = 5$ for $p = 2$			
W_λ	0.080	0.075	0.773	0.727	0.989	0.979
PT	0.081	0.075	0.773	0.728	0.989	0.979
LH	0.083	0.077	0.778	0.733	0.989	0.980
T_1	0.045	0.046	0.401	0.368	0.698	0.656
T_2	0.040	0.042	0.323	0.284	0.573	0.523
T_3	0.050	0.051	0.461	0.407	0.850	0.797
$V^{(2)}$	0.045	0.047	0.624	0.553	0.937	0.886
			Case of $n_1 = n_2 = n_3 = 20$ for $p = 2$			
W_λ	0.035	0.036	0.812	0.713	0.999	0.994
PT	0.035	0.037	0.811	0.711	0.999	0.993
LH	0.035	0.036	0.814	0.715	0.999	0.994
T_1	0.048	0.051	0.509	0.502	0.820	0.809
T_2	0.046	0.049	0.495	0.466	0.805	0.783
T_3	0.045	0.050	0.631	0.578	0.987	0.976
$V^{(2)}$	0.050	0.053	0.913	0.853	1.000	1.000

Table 3: Continued for the multivariate lognormal distribution

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
			Case of $n_1 = n_2 = n_3 = 5$ for $p = 3$			
W_λ	0.028	0.028	0.158	0.091	0.538	0.323
PT	0.031	0.031	0.194	0.115	0.469	0.288
LH	0.033	0.033	0.187	0.113	0.633	0.405
T_1	0.049	0.050	0.182	0.152	0.509	0.436
T_2	0.038	0.040	0.125	0.115	0.413	0.336
T_3	0.050	0.053	0.167	0.133	0.528	0.430
$V^{(2)}$	0.046	0.048	0.223	0.180	0.695	0.562
			Case of $n_1 = 15$, $n_2 = 10$ and $n_3 = 5$ for $p = 3$			
W_λ	0.088	0.080	0.879	0.772	0.999	0.990
PT	0.088	0.080	0.879	0.772	0.999	0.990
LH	0.088	0.080	0.880	0.773	0.999	0.990
T_1	0.050	0.053	0.471	0.402	0.761	0.708
T_2	0.040	0.044	0.382	0.299	0.642	0.563
T_3	0.050	0.055	0.548	0.438	0.909	0.844
$V^{(3)}$	0.041	0.047	0.721	0.570	0.977	0.910
			Case of $n_1 = n_2 = n_3 = 20$ for $p = 3$			
W_λ	0.036	0.038	0.919	0.733	1.000	0.996
PT	0.036	0.038	0.916	0.727	1.000	0.996
LH	0.036	0.038	0.922	0.738	1.000	0.997
T_1	0.048	0.057	0.530	0.521	0.846	0.833
T_2	0.046	0.055	0.535	0.489	0.839	0.812
T_3	0.045	0.057	0.718	0.618	0.995	0.988
$V^{(3)}$	0.043	0.049	0.958	0.823	1.000	1.000

Austrian Journal of Statistics
February 2017, Volume 46, 15–32.
http://www.ajs.or.at/
doi:10.17713/ajs.v46i1.133

The Analysis of Ranking Data Using Score Functions and Penalized Likelihood

Mayer Alvo
University of Ottawa

Hang Xu
University of Hong Kong

Abstract

In this paper, we consider different score functions in order summarize certain characteristics for one and two sample ranking data sets. Our approach is flexible and is based on embedding the nonparametric problem in a parametric framework. We make use of the von Mises-Fisher distribution to approximate the normalizing constant in our model. In order to gain further insight in the data, we make use of penalized likelihood to narrow down the number of items where the rankers differ. We applied our method on various real life data sets and we conclude that our methodology is consistent with the data.

Keywords: rankings, score function, Rao score test, Spearman, Kendall, penalized likelihood, von Mises distribution.

1. Introduction

Ranking data occur quite frequently in practice. There are examples in sport competitions (Deng, Han, Li, and Liu 2014), in the selection of candidates in political elections (Croon 1989; Lee and Philip 2012), in the arrangement of web-pages when using search engines (Aslam and Montague 2001) and in flagging disease related gene in bioinformatics (DeConde, Hawley, Falcon, Clegg, Knudsen, and Etzioni 2006) just to name a few. In all cases it is of interest to present summaries of the data. Some of the methods include the calculation of the average ranks. Borda counts (Baba 1986) usually sum up the rank as each item's score and present transformations such as ordering or averaging. There are also more sophisticated methods which involve building models to explain the ranking data. For example, Kemeny rankings (Mallows 1957; Kemeny and Snell 1962) are based on fitting a distance-based model with Kendall distance and using the modal rank as a summary analogous to the mean for numerical data. Deng *et al.* (2014) suggested a Bayesian approach to aggregate the observed rankings whereas Dwork, Kumar, Naor, and Sivakumar (2001) suggested a model using a Markov chains approach to have a summary of the word association or the web-page for example.

Here we first introduce a parametric model for the one sample case which incorporates an arbitrary score function. Such score functions include both the Spearman and Kendal scores. The former focuses attention on a single item at a time whereas the latter allows for comparisons between pairs of items. In the era of big data, the number of items being ranked is usually very large. For this reason, we may apply penalized likelihood methods to focus on those items that are particularly preferred by the rankers. The parametric model can be extended to the two sample case whereby we now compare groups of judges. Bootstrap methods can be particularly useful for constructing a confidence interval for the parameters in the models.

The article is organized as follows. In Section 2 we formally introduce the parametric model and indicate how to calculate the maximum likelihood estimates (MLE). We also obtain the estimates using penalized likelihood (PMLE) . In Section 3, we introduce the Spearman score function, explain its significance and then apply it to three real data sets. In Section 4 we consider the Kendall score function, explain its significance and use it on the same data sets. In Section 5 we consider the two sample problem and illustrate our results on two real life data sets. We briefly summarize our results in Section 6.

2. The probability models

Neyman (1937) first proposed smooth tests for testing the null hypothesis that data come from a uniform distribution on the interval $(0,1)$. The smooth alternative density proposed by Neyman was given by

$$p(y,\theta) = exp\left\{\sum_{i=1}^{k}\theta_i h_i(y) - K(\theta)\right\}$$

where $\theta^{'} = (\theta_1, \theta_2, ..., \theta_k)$ is a set of unknown parameters, $K(\theta)$ is the normalizing constant and $h_i(y)$ are the Legendre orthonormal polynomials with respect to the uniform on $(0,1)$. Motivated by this model, we can define a similar probability function for ranking data. Assume there are n judges each of whom ranks t items. Let $\{\omega_j\}$ be the set of $t!$ possible rankings and define the probability that $X(\omega_j) = x_j$ by

$$\pi_j\left(\theta\right) = \exp\left\{\theta' x_j - K\left(\theta\right)\right\}\frac{1}{t!},\; j = 1, ..., t!$$

where $\theta^{'} = (\theta_1, ..., \theta_k)$ is a k-dimensional vector of parameters, $K\left(\theta\right)$ is a normalizing constant. $X(\omega_j)$ is a k-dimensional vector transformed from the original ranking ω_j. The transformation rule $X\left(\omega_j\right) = x_j$ is defined by certain scoring function. The likelihood function is obtained from the multinomial distribution and is proportional to

$$L\left(\theta\right) \sim \prod \left[\pi_j\left(\theta\right)\right]^{n_j}$$

where n_j is the number of judges choosing ω_j . Let the total sample size $n = \sum_{j=1}^{t!} n_j$. The log-likelihood function is then given by

$$l(\theta) \sim \sum_{j=1}^{t!} n_j\left[\theta' x_j - K(\theta)\right]$$

As mentioned in the introduction, when the number of items being ranked is large, the dimension $t!$ becomes very large. Penalized likelihood is a parametric method which helps to narrow down the number of items to look at. The main idea is to add a penalty term in the log likelihood function. In this paper, we consider l_2 norm penalized terms and minimize this function in order to obtain the Maximum Penalized likelihood (MPL) estimates of the vector parameter θ:

$$\Lambda(\theta, c) = -\theta'\left[\sum_{j=1}^{t!} n_j x_j\right] + nK(\theta) + \lambda(\sum_{i=1}^{t}\theta_i^2 - c)$$

for some prescribed values of the constant c. When t is large (say $t \geq 10$), the exact computation of the normalizing constant $K(\theta)$ involves a summation of $t!$ items. McCullagh (1993) suggested using the normalizing constant from the von Mises-Fisher distribution. Following on this suggestion, we approximate $K(\theta)$ with

$$K(\theta) \approx \frac{1}{t!}(2\pi)^{\frac{t}{2}} I_{\frac{t}{2}-1}(\|\theta\|)\, \|\theta\|^{-\frac{t}{2}+1}$$

where $\|\theta\|$ is the norm of θ and $I_v(z)$ is the modified Bessel function of the first kind given by

$$I_v(z) = \sum_{k=0}^{\infty} \frac{1}{\Gamma(k+1)\Gamma(v+k+1)} \left(\frac{z}{2}\right)^{2k+\nu}$$

We may now find approximately the Maximum Penalized likelihood estimation for θ after ensuring that $\|x\| = 1$ in our model. There are several ways to minimize the target function $\Lambda(\theta, c)$. In this paper, we used three methods: the BFGS Quasi-Newton method (Broyden 1970; Fletcher 1970; Goldfarb 1970; Shanno 1970),the Trust-Region-Reflective method (Coleman and Li 1994, 1996) and the Interior Point Algorithm (Byrd, Hribar, and Nocedal 1999; Byrd, Gilbert, and Nocedal 2000; Waltz, Morales, Nocedal, and Orban 2006). As well, we imposed the constraint: $\theta \geq 0$ on θ in order to compare the results when there is no constraint. Differentiating with respect to θ we have

$$\frac{\partial \Lambda(\theta, c)}{\partial \theta} = -\sum_{j=1}^{t!} n_j x_j + 2\lambda\theta + \frac{n\theta}{\|\theta\|} \times \frac{\partial K(\theta)}{\partial \|\theta\|}$$

where

$$\frac{\partial K(\theta)}{\partial \|\theta\|} = \frac{1}{t!}(2\pi)^{\frac{t}{2}} \left[\|\theta\|^{1-\frac{t}{2}} \left(I_{\frac{t}{2}}(\|\theta\|) - \frac{1-\frac{t}{2}}{\|\theta\|} I_{\frac{t}{2}-1}(\|\theta\|) \right) + (1-\frac{t}{2})\, \|\theta\|^{-\frac{t}{2}}\, I_{\frac{t}{2}-1}(\|\theta\|) \right]$$

The critical points of the minimization occur at saddle points, rather than at local maxima (or minima). For the Quasi-Newton method, the target function is the magnitude of the gradient which is the square root of the sum of the squares of the partial derivatives instead of the $\Lambda(\theta, c)$ to solve the problem. In the applications that follow, the results from the three different methods yield the same solutions and the algorithms implemented in MATLAB converge very fast.

Following the estimation of θ, we proceeded to apply the basic bootstrap method in order to assess the distribution of θ. The basic idea of the bootstrap is to sample n rankings with replacement from the data. Then we find the MLE of each bootstrap sample. Repeating this procedure several times say 10^4, leads to a distribution of θ. We can draw useful inference from the distribution θ and determine whether or not the $\theta_j' s$ are significantly different from zero. Two sided confidence intervals can also be calculated. We illustrate this on real life data.

3. Using the Spearman score function

3.1. The Spearman score function and its meaning

In this section we restrict our attention to the Spearman score function defined as:

$$X(\omega_j) = \left(\omega_j(1) - \frac{t+1}{2}, ..., \omega_j(t) - \frac{t+1}{2}\right), \; j = 1, ..., t!$$

where $\omega_j(i)$ is the rank given to Item i.

Let T_S be the $t \times t!$ matrix of possible values of X: $T_S = (X(\omega_j))$. Taking $t = 3$ as an example, let ω_j be the possible rankings:

$$\omega_1 = \begin{pmatrix}1\\2\\3\end{pmatrix}, \omega_2 = \begin{pmatrix}1\\3\\2\end{pmatrix}, \omega_3 = \begin{pmatrix}2\\1\\3\end{pmatrix}, \omega_4 = \begin{pmatrix}2\\3\\1\end{pmatrix}, \omega_5 = \begin{pmatrix}3\\1\\2\end{pmatrix}, \omega_6 = \begin{pmatrix}3\\2\\1\end{pmatrix}$$

Then the matrix T_S becomes:

$$T_S = \begin{bmatrix} -1 & -1 & 0 & 0 & 1 & 1 \\ 0 & 1 & -1 & 1 & -1 & 0 \\ 1 & 0 & 1 & -1 & 0 & -1 \end{bmatrix}$$

Let $\pi_j = n_j/n$ and let π be the column vector of π_j. The first part of our likelihood function $\Lambda(\theta, c)$ becomes:

$$-n\theta' T_s \pi = -\frac{n}{\|x\|} \times \begin{bmatrix}\theta_1 & \theta_2 & \theta_3\end{bmatrix} \begin{bmatrix} \pi_5 + \pi_6 - \pi_1 - \pi_2 \\ \pi_2 + \pi_4 - \pi_3 - \pi_5 \\ \pi_1 + \pi_3 - \pi_4 - \pi_6 \end{bmatrix}$$

We can notice that for θ_1,

$$\pi_5 + \pi_6 = \Pr(\mathit{giving\, rank}\; 3\; \mathit{to\, Item}\; 1)$$

and

$$\pi_1 + \pi_2 = \Pr(\mathit{giving\, rank}\; 1\; \mathit{to\, Item}\; 1).$$

So here θ_1 weights the difference in probability between giving the top rank and the lowest rank to Item 1 ($\Pr(\mathit{giving\,rank}\,3\,\mathit{to\,Item}\,1) - \Pr(\mathit{giving\,rank}\,1\,\mathit{to\,Item}\,1)$). There is a similar interpretation for θ_2 and θ_3. In general, the components of the matrix T_S actually focus on a special characteristic of the data, namely on the difference in weighted average of the rankings to the t objects. The weights here are $i - \frac{t+1}{2}$, $i = 1, ..., t$. The θ_i's here represent the coefficients attributed to each item.

Similarly, we can also compute the matrix of possible scores for $t = 4$. It can be seen that the first row element for $T_S\pi$ is $-1.5\Pr(\mathit{giving\ rank}\ 1\ \mathit{to\ Item}\ 1) - 0.5\Pr(\mathit{giving\ rank}\ 2) + 0.5\Pr(\mathit{giving\ rank}\ 3) + 1.5\Pr(\mathit{giving\ rank}\ 4)$. This is the weighted average of probability giving the high rank compare with the low rank to Item 1. Notice that when t is odd, the weight of the middle item is 0 which means that our comparison is symmetric and balanced.

3.2. Application to real data

Sutton data

For the first example, we consider the Sutton data ($t = 3$) analyzed by C. Sutton in her 1976 thesis on leisure preferences and attitudes on retirement of the elderly for 14 white and 13 black females in the age group 70- 79 years. Each individual was asked: with which sex do you wish to spend your leisure? Each female was asked to rank the three responses: male(s), female(s) or both, assigning rank 1 for the most desired and 3 for the least desired. The first item in the ranking corresponds to "male", the second to "female" and the third to "both". To illustrate the approach in the one sample case, we combined the data from the two groups as in Table 1.

We applied our penalized likelihood in this situation and the results are shown in Table 2.

To better illustrate, we rearrange our result (unconstrained θ, c=1) and data in Table 3. It can be seen that θ_1 is the largest coefficient and Item 1 "male" shows the greatest difference between the number of judges choosing rank 1 or rank 3 which means that the judges dislike

Table 1: Sutton data on leisure preferences

Rankings	(123)	(132)	(213)	(231)	(312)	(321)
Frequencies	1	1	1	5	7	12

Table 2: Maximum penalized likelihood estimation for Sutton's data

Choice of c	$\theta \geq 0$	θ_1	θ_2	θ_3	$\Lambda(\theta, c)$
c=0.5	No	0.53	-0.06	-0.47	50.00
	Yes	0.71	0.00	0.00	52.88
c=1	No	0.75	-0.09	-0.66	50.36
	Yes	1.00	0.00	0.00	54.44
c=2	No	1.06	-0.12	-0.93	54.62
	Yes	1.41	0.00	0.00	60.38
c=10	No	2.36	-0.28	-2.08	159.96
	Yes	3.16	0.00	0.00	172.84
MLE	No	0.60	-0.07	-0.53	49.90
	Yes	0.61	0.00	0.00	52.79

spending leisure with male the most. For Item 3 "both", the greater value of negative θ_3 means judges prefer to spend leisure with both sex the most. θ_2 is close to zero and we deduce the judges show no strong preference on Female. This is consistent with the hypothesis that θ close to zero means randomness. To conclude, the results also show that θ_i weights the difference in probability giving the top rank and the lower rank to Item i. Negative θ_i means the judges prefer Item i more and positive θ_i means the judges are more likely to give a lower rank to Item i.

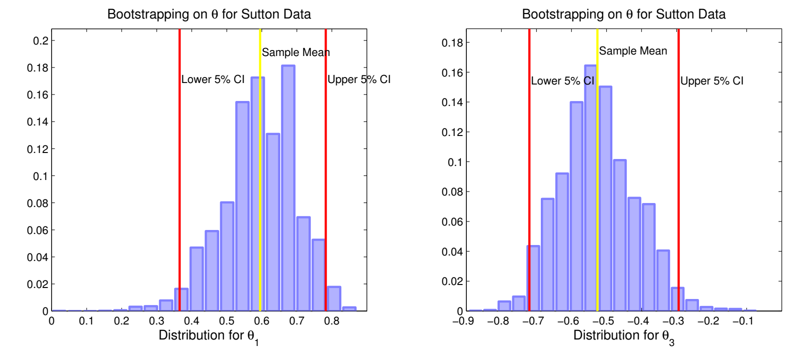
Applying the bootstrap method on the Sutton data we plot the distribution of θ in Figure 1. The bootstrap sample size is 10^4 in this case. For H_0: $\theta_i = 0$, we can see that θ_1 and θ_3 are significantly different from 0 and θ_2 is not significantly different from zero. We can also see that the distribution of θ_1 and θ_2 is not completely bell shaped. This is mainly because the sample size of the data is small. Using a traditional t-test method may be misleading in this case.

Song data

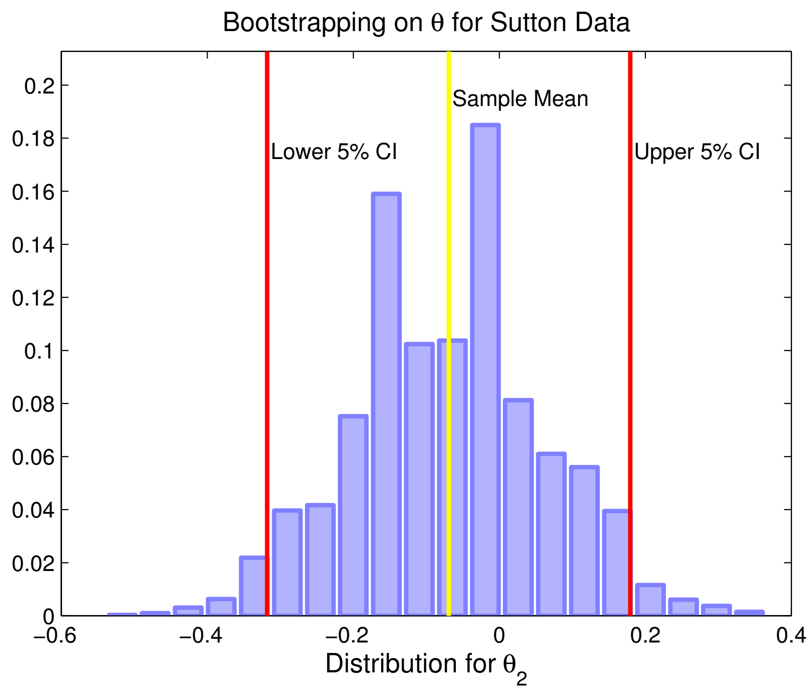
Our second example is the Song data (t=5) from Critchlow, Fligner, and Verducci (1991). Ninety-eight students were asked to rank 5 words, (1) score, (2) instrument, (3) solo, (4) benediction and (5) suit, according to the association with the word "song". Critchlow *et al.* (1991) reported that the average ranks for words (1) to (5) are 2.72, 2.27, 1.60, 3.71 and 4.69 respectively. However, the available data given in Critchlow *et al.* (1991) is in grouped format and the ranking of 15 students are unknown and hence discarded, resulting in 83 rankings, as shown in Table 4.

Table 3: The Sutton data and the estimation of θ

Items	Number of judges	Action	Difference	θ	value
Item 1 (Male):	2	choose to rank 1	-17	θ_1	**0.75**
	19	choose to rank 3			
Item 2 (Female):	8	choose to rank 1	2	θ_2	-0.09
	6	choose to rank 3			
Item 3 (Both):	17	choose to rank 1	15	θ_3	**-0.66**
	2	choose to rank 3			



(a) Significantly different from zero



(b) Not significantly different from zero

Figure 1: The distribution of θ for Sutton data by bootstrap method

Table 4: Song data set

Rankings	Observed frequency
(32145)	19
(23145)	10
(13245)	9
(42135)	8
(12345)	7
(31245)	6
(32154)	6
(52134)	5
(21345)	4
(24135)	3
(41235)	2
(43125)	2
(52143)	2
others	0

Table 5: Maximum penalized likelihood estimation for Song's data

Choice of c	$\theta \geq 0$?	θ_1	θ_2	θ_3	θ_4	θ_5	$\Lambda(\theta, c)$
c=0.5	No	-0.07	-0.22	-0.41	0.21	0.48	-30.15
	Yes	0.00	0.00	0.00	0.29	0.65	-17.81
c=1	No	-0.10	-0.31	-0.58	0.30	0.69	-49.60
	Yes	0.00	0.00	0.00	0.41	0.91	-32.15
c=2	No	-0.15	-0.44	-0.81	0.43	0.97	-76.45
	Yes	0.00	0.00	0.00	0.57	1.29	-51.76
c=10	No	-0.33	-0.98	-1.82	0.96	2.17	-176.13
	Yes	0.00	0.00	0.00	1.28	2.89	-120.94
MLE	No	-0.49	-1.46	-2.73	1.44	3.25	-220.60
	Yes	0.00	0.00	0.00	1.76	3.97	-141.12

Table 6: The Song data and the estimation of θ

Items	number of judges	Action	Average ranks	θ	value
Item 1 (Score):	16	choose to rank 1	2.72	θ_1	-0.10
	7	choose to rank 5			
Item 2 (instrument):	10	choose to rank 1	2.27	θ_2	-0.31
	0	choose to rank 5			
Item 3 (solo):	55	choose to rank 1	1.60	θ_3	**-0.58**
	0	choose to rank 5			
Item 4 (benediction):	0	choose to rank 1	3.71	θ_4	0.30
	6	choose to rank 5			
Item 5 (suit):	0	choose to rank 1	4.69	θ_5	**0.69**
	70	choose to rank 5			

We applied the penalized likelihood in this situation and the results are shown in Table 5. We also rearranged the results (unconstrained θ, c=1) and data in Table 6. Note that for convenience we only show the number of judges who rank the top and the lowest but ranking 2 or 4 also is involved in determining the value of θ. From the results, we can see the value of θ successfully captures the properties of the data. θ_5 is the largest positive value and most of the judges think word "suit" is not related to the word "song". θ_3 is the largest negative value and 55 of the 83 judges think that word "solo" is the closest to the word "song". From the results, we see that θ_i successfully weights the difference in probability giving the upper rank and the lower rank to Item i.

We also applied our bootstrap method on the Song data and plotted the distribution of θ in Figure 2. The bootstrap sample size is 10^4 in this case. For H_0: $\theta_i = 0$, we can see that all the θ's are significantly different from zero. As well, their distributions of θ are more bell shaped in view of the larger sample size.

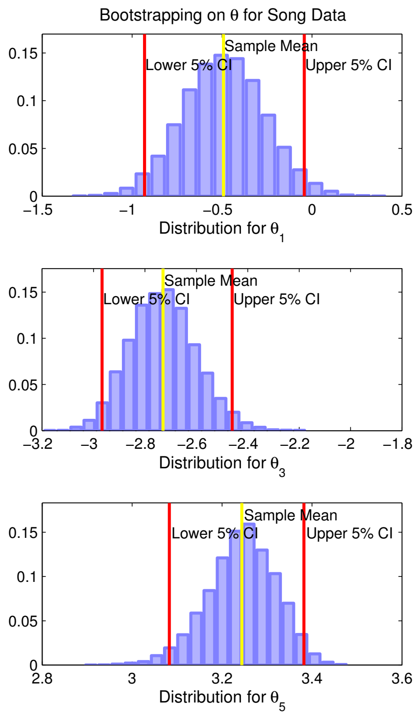


Figure 2: The distribution of θ for Song data by bootstrap method

Goldberg data

Our final example is due to Goldberg (1976) data (t=10). In the data, 143 graduates were asked to rank 10 occupations according to the degree of social prestige. These 10 occupations are: (i) Faculty member in an academic institution (Fac), (ii) Mechanical engineer (ME), (iii) Operation researcher (OR), (iv) Technician (Tech), (v) Section supervisor in a factory (Sup), (vi) Owner of a company employing more than 100 workers (Own), (vii) Factory foreman (For), (viii) Industrial engineer (IE), (ix) Manager of a production department employing more than 100 workers (Mgr) and (x) Applied scientist (Sci). The data are given in Cohen and Mallows (1980) and have been analyzed by many researchers.

Feigin and Cohen (1978) analyzed the Goldberg data and found three outliers due to the fact that the corresponding graduates wrongly presented rankings in reverse order. After reversing these 3 rankings, the average ranks received by the 10 occupations are 8.57, 4.90, 6.29, 1.90, 4.34, 8.13, 1.47, 6.27, 5.29, 7.85, with the convention that bigger rank means more prestige. Then the preference of graduates is in the order: Fac $\succ$ Own $\succ$Sci $\succ$ OR $\succ$ IE $\succ$ Mgr $\succ$ ME $\succ$ Sup $\succ$ Tech $\succ$ For.

We applied our penalized likelihood method and the results are shown in Table 7.

We also rearranged our results (unconstrained θ, c=1) and data in Table 8. For convenience we only show the number of judges who rank the top and the lowest but giving other rankings also involves in determining the value of θ. We can see that the results are consistent with the average ranks and our preference result from θ is the also consistent with the results from Feigin and Cohen (1978). For Item 7 "factory foreman", 93 of 143 graduates give rank 1

Table 7: Maximum penalized likelihood estimation for Goldberg's data

Choice of c	$\theta \geq 0$	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	$\Lambda(\theta, c)$
c=0.5	No	0.30	-0.06	0.08	-0.35	-0.11	0.25	-0.39	0.07	-0.02	0.23	-81.89
	Yes	0.45	0.00	0.12	0.00	0.00	0.39	0.00	0.11	0.00	0.35	-53.50
c=1	No	0.42	-0.08	0.11	-0.49	-0.16	0.36	-0.55	0.11	-0.03	0.32	-115.80
	Yes	0.64	0.00	0.16	0.00	0.00	0.55	0.00	0.16	0.00	0.49	-75.66
c=2	No	0.59	-0.11	0.15	-0.69	-0.22	0.51	-0.78	0.15	-0.04	0.45	-163.77
	Yes	0.90	0.00	0.23	0.00	0.00	0.77	0.00	0.23	0.00	0.69	-107.01
c=10	No	1.32	-0.26	0.34	-1.55	-0.50	1.13	-1.73	0.33	-0.09	1.01	-366.21
	Yes	2.02	0.00	0.52	0.00	0.00	1.73	0.00	0.51	0.00	1.54	-239.27
MLE	No	8.74	-1.70	2.24	-10.25	-3.31	7.48	-11.46	2.20	-0.61	6.67	-2277.02
	Yes	13.03	0.00	3.34	0.00	0.00	11.17	0.00	3.28	0.00	9.95	-1447.59

which means most of them think factory foreman has the lowest social prestige and our θ_7 is the lowest (that bigger rank means more prestige). For t=10, it is almost impossible to calculate the exact value of $K(\theta)$. Our results show that the von Mises-Fisher distribution approximation of $K(\theta)$ actually works well and can be easily used when t is large.

Table 8: The Goldberg data and the estimation of θ

Items	number of judges	Action	Average ranks	θ value
Item 1 (Fac)	0	choose to rank 1	**8.57**	**0.42**
	49	choose to rank 10		
Item 2 (ME)	0	choose to rank 1	4.90	-0.08
	0	choose to rank 10		
Item 3 (OR)	3	choose to rank 1	6.29	0.11
	1	choose to rank 10		
Item 4 (Tech)	45	choose to rank 1	1.90	-0.49
	0	choose to rank 10		
Item 5 (Sup)	0	choose to rank 1	4.34	-0.16
	0	choose to rank 10		
Item 6 (Own)	0	choose to rank 1	8.13	0.36
	54	choose to rank 10		
Item 7 (For)	93	choose to rank 1	**1.47**	**-0.55**
	0	choose to rank 10		
Item 8 (IE)	0	choose to rank 1	6.27	0.11
	1	choose to rank 10		
Item 9 (Mgr)	1	choose to rank 1	5.29	-0.03
	3	choose to rank 10		
Item 10 (Sci)	1	choose to rank 1	7.85	0.32
	35	choose to rank 10		

The bootstrap distribution of θ is exhibited in Figure 3. The bootstrap sample size is 10^4 in this case. For H_0: $\theta_i = 0$, we see that all the θ_i except θ_9 are significantly different from zero. We can also see that the distribution of the θ's are all bell shaped because the sample size is large.

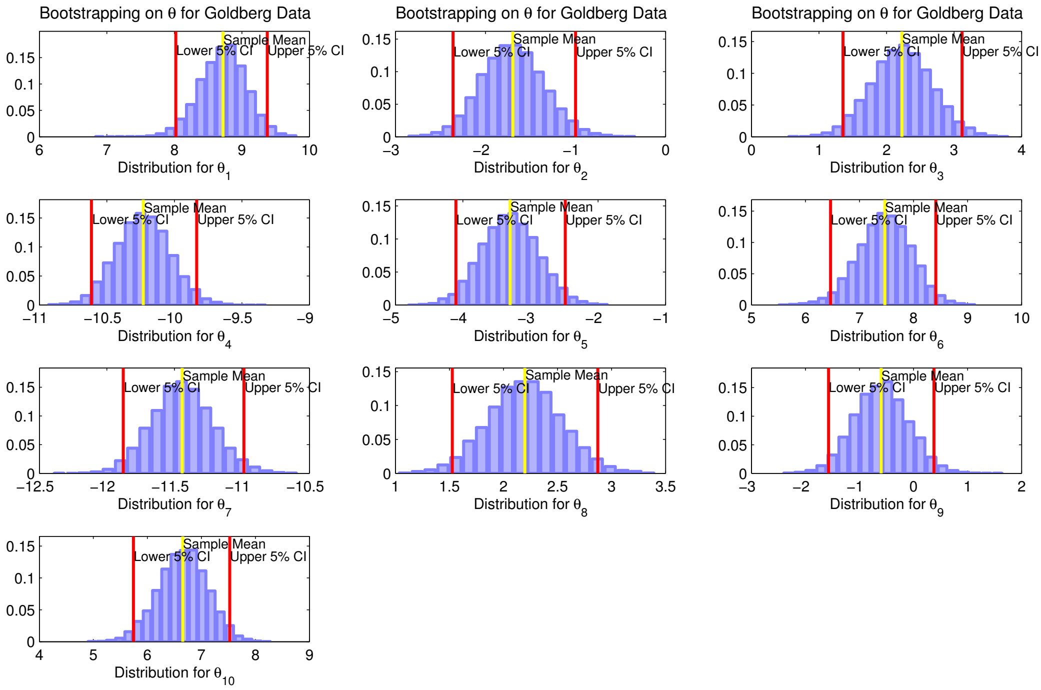


Figure 3: The distribution of θ for Goldberg data by bootstrap method

4. Pair comparison using the Kendall score function

4.1. Kendall score function and its meaning

In this section, we re-consider the previous data sets through the lens of the Kendall score statistic. Specifically, for the Kendall score function, the $X(\omega_j)$ vector takes values$\left(t_K\left(\omega_j\right)\right)_q$where the q^{th} element is the pair comparison between jth rank given to item m and n denoted by

$$\left(t_K\left(\omega_j\right)\right)_q = sgn\left[\omega_j\left(m\right) - \omega_j\left(n\right)\right]$$

for $q = (n-1)\left(t - \frac{n}{2}\right) + (m-n), 1 \leq n < m \leq t$. The matrix of possible values of X becomes

$$T_K = (t_K(\omega_1), ..., t_K(\omega_{t!}))'$$

which is of dimension $\binom{t}{2} \times t!$. And θ_q from 1 to $\binom{t}{2}$ weights the pair m and n. In the case of the Kendall score function, the $\theta's$ focus on the comparison between pairs of items ranked by the judges. As an example, consider once again the case $t = 3$. Then,

$$X(\omega) = \begin{pmatrix} sgn(\omega(2) - \omega(1)) \\ sgn(\omega(3) - \omega(1)) \\ sgn(\omega(3) - \omega(2)) \end{pmatrix}$$

where θ_1 weights the comparison between Items 1 and 2, θ_2 weights Items 1, 3 and θ_3 Items

2, 3. As well,

$$T_K = \begin{pmatrix} 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix}$$

When $\theta_q < 0$ (for say Item i and Item j), it means that the judges prefer Item j over Item i ($\mu(i) > \mu(j)$). When θ_q is close to zero, the judges have no special preference between this pair. In the next section we apply the Kendall score function to the previous data sets.

4.2. Application to real data

Sutton data

For the first example, we consider the Sutton data ($t = 3$). We applied our penalized likelihood with Kendall score function in this situation and the results are shown in Table 9.

Table 9: MPLE using Kendall score function for Sutton data

Pair Compare			choice of c				
Item i	Item j	θ	c=0.5	c=1	c=2	c=10	MLE
1	2	θ_1	-0.35	-0.49	-0.70	-1.56	-0.60
1	3	θ_2	-0.56	-0.80	-1.13	-2.53	-0.97
2	3	θ_3	-0.24	-0.34	-0.48	-1.08	-0.41
$\Lambda(\theta, c)$			42.79	40.17	40.20	127.76	39.59

Table 10: Pair comparison from the Sutton data and the estimation of θ

Item i	Item j	number of judges	Pair comparison	θ
1	2	7	more prefer 1	-0.49
		20	more prefer 2	
1	3	3	more prefer 1	-0.80
		24	more prefer 3	
2	3	9	more prefer 2	-0.34
		18	more prefer 3	

We rearrange the Sutton data focusing on pair comparison and our results (c=1) in Table 10. First, from our estimated θ, we can find that all $\theta_i's$ are negative. This is consistent with our interpretation of θ. The judges strongly prefer Males to Both and Males to Females. They least prefer Females to Both. We can conclude that our $\theta's$ well represent the paired comparisons among the judges.

Song data

As to the Song data (t=5) the results are shown in Table 11.

Since the Song data is about how much 5 words are close to the word “song”. We can summarize the results in Table 12. We note that θ_7, θ_8 and θ_9 all have the same value 0.29. For these, all of the judges think Item i is closer to Item j. Once again, we conclude that θ well represents the paired preferences among judges.

Goldberg data

For the Goldberg (1976) data (t=10) we note that a large rank means more prestige, so our interpretation is reversed. We applied penalized likelihood with the Kendall score function in this situation and part of the results are shown in Table 13.

Table 11: MPLE using Kendall score function for Song data

Pair Compare			choice of c				
Item i	Item j	θ	c=0.5	c=1	c=2	c=10	MLE
1	2	θ_1	-0.09	-0.12	-0.17	-0.39	-0.76
1	3	θ_2	-0.15	-0.21	-0.30	-0.67	-1.30
1	4	θ_3	0.16	0.22	0.31	0.70	1.36
1	5	θ_4	0.24	0.34	0.48	1.07	2.09
2	3	θ_5	-0.16	-0.22	-0.31	-0.70	-1.36
2	4	θ_6	0.25	0.36	0.51	1.13	2.21
2	5	θ_7	0.29	0.41	0.57	1.28	2.51
3	4	θ_8	0.29	0.41	0.57	1.28	2.51
3	5	θ_9	0.29	0.41	0.57	1.28	2.51
4	5	θ_{10}	0.23	0.33	0.46	1.04	2.03
$\Lambda(\theta, c)$			-125.39	-184.29	-266.92	-602.04	-973.61

Table 12: Interpretation of θ in song data. A$\succ$B means A is preferred to B, that is A is closer to the word "Song".

Behavior of θ	Interpretation of θ
$\theta_1, \theta_2 < 0$	score$\prec$instrument, solo
$\theta_3 > 0$	score$\succ$benediction, suit
$\theta_5 < 0$	instrument$\prec$solo
$\theta_6, \theta_7 > 0$	instrument$\succ$benediction, suit
$\theta_8, \theta_9 > 0$	solo$\succ$benediction, suit
$\theta_{10} > 0$	benediction$\succ$suit

Table 13: MPLE using Kendall score function for Goldberg data

Pair Compare			choice of c				
Item i	Item j	θ	c=0.5	c=1	c=2	c=10	MLE
1	2	θ_1	-0.13	-0.18	-0.25	-0.56	-4.09
1	3	θ_2	-0.09	-0.13	-0.18	-0.40	-2.89
1	4	θ_3	-0.15	-0.21	-0.29	-0.65	-4.76
1	5	θ_4	-0.12	-0.17	-0.25	-0.55	-4.02
1	6	θ_5	-0.02	-0.02	-0.03	-0.07	-0.50
1	7	θ_6	-0.15	-0.21	-0.29	-0.65	-4.76
1	8	θ_7	-0.11	-0.15	-0.21	-0.48	-3.49
1	9	θ_8	-0.12	-0.17	-0.24	-0.53	-3.89
1	10	θ_9	-0.03	-0.04	-0.05	-0.11	-0.83
2	3	θ_{10}	0.06	0.09	0.12	0.28	2.03
2	4	θ_{11}	-0.14	-0.20	-0.28	-0.63	-4.62
...	...	...	...	...	...	...	...
9	10	θ_{45}	0.09	0.13	0.19	0.42	3.09
$\Lambda(\theta, c)$			-490.79	-694.06	-981.55	-2194.80	-15177.07

Table 14: Interpretation of θ in Goldberg data. A$\succ$B means A has more prestige than B. Underlined θ_i in left column means that its absolute value is bigger than 0.1. The underlined occupation in right column means that the preference is very strong for this comparison.

Behavior of θ	Interpretation of θ
$\theta_1, \underline{\theta_2}, \theta_3, \underline{\theta_4}, \underline{\theta_5}, \theta_6, \underline{\theta_7}, \underline{\theta_8}, \underline{\theta_9} < 0$	Fac$\succ$$\underline{\text{ME}}$, OR, $\underline{\text{Tech}}$, $\underline{\text{Sup}}$, Own, $\underline{\text{For}}$, $\underline{\text{IE}}$, $\underline{\text{Mgr}}$, Sci
$\underline{\theta_{11}}, \theta_{12}, \underline{\theta_{14}} < 0$	ME$\succ$$\underline{\text{Tech}}$, Sup, $\underline{\text{For}}$
$\theta_{10}, \theta_{13}, \underline{\theta_{15}}, \theta_{16}, \theta_{17} > 0$	ME$\prec$OR, Own, $\underline{\text{IE}}$, Mgr, Sci
$\underline{\theta_{18}}, \theta_{19}, \underline{\theta_{21}}, \theta_{22}, \theta_{23} < 0$	OR$\succ$$\underline{\text{Tech}}$, Sup, $\underline{\text{For}}$, IE, Mgr
$\theta_{20}, \theta_{24} > 0$	OR$\prec$Own, Sci
$\theta_{27} < 0$	Tech$\succ$For
$\underline{\theta_{25}}, \underline{\theta_{26}}, \underline{\theta_{28}}, \underline{\theta_{29}}, \underline{\theta_{30}} > 0$	Tech$\succ$$\underline{\text{Sup}}$, $\underline{\text{Own}}$, $\underline{\text{IE}}$, $\underline{\text{Mgr}}$, $\underline{\text{Sci}}$
$\underline{\theta_{32}} < 0$	Sup$\succ$For
$\underline{\theta_{31}}, \theta_{33}, \theta_{34}, \theta_{35} > 0$	Sup$\prec$$\underline{\text{Own}}$, IE, Mgr, Sci
$\underline{\theta_{36}}, \theta_{37}, \underline{\theta_{38}}, \theta_{39} < 0$	Own$\succ$$\underline{\text{For}}$, IE, $\underline{\text{Mgr}}$, Sci
$\underline{\theta_{40}}, \underline{\theta_{41}}, \underline{\theta_{42}} > 0$	For$\prec$$\underline{\text{IE}}$, $\underline{\text{Mgr}}$, $\underline{\text{Sci}}$
$\theta_{43} < 0$	IE $\succ$Mgr
$\theta_{42} > 0$	IE$\prec$Sci
$\theta_{45} > 0$	Mgr$\prec$Sci

The Goldberg data compares occupations with more social prestige. We can also interpret the estimated θ to get the pair comparison results. We arrange the behavior and interpretation of θ in Table 14. In the table, the underlined θ_i in the left column means that its absolute value is larger than 0.1. The underlined occupation in the column on the right means that the preference is very strong. From the interpretation of θ, we find that the pair comparisons make sense. To conclude, the Kendall score function provides a nice way to look at the data by pair preference.

5. Extension to the two-sample ranking problem

5.1. Extension to two-sample ranking problems and its meaning

We may extend the approach to the two sample case. Let $X_1, .X_2$ be two independent random variables under our model whose distributions are given respectively by:

$$\pi_j(\theta_l) = \exp\left\{\theta_l' x_j - K(\theta_l)\right\} p_{lj},\ j = 1, ..., t!,\ l = 1, 2$$

where $\theta_l = (\theta_{l1}, ..., \theta_{lt})'$ represents the vector of parameters for population l and x_j is as in the one-sample case, a t-dimensional vector of scores. We shall use the Spearman scores here throughout the two-sample case. Set $\gamma = \theta_1 - \theta_2$ and write

$$\theta_l \quad = \quad \mu + b_l \gamma$$

for $l = 1, 2$ where

$$\mu = \frac{n_1\theta_1 + n_2\theta_2}{n_1 + n_2},\ b_1 = \frac{n_{2,}}{n_1 + n_2},\ b_2 = -\frac{n_1}{n_1 + n_2}.$$

Suppose that the observed vector of frequencies for the l^{th} population is

$$n_l = (n_{l1}, ..., n_{lt!})'$$

The logarithm of the likelihood L as a function of (μ, γ) is proportional to

$$\log L(\mu, \gamma) \quad \sim \quad \sum_{l=1}^{2} \sum_{j=1}^{t!} n_{lj} \left\{(\mu + b_l \gamma)' x_j - K(\theta_l)\right\}$$

We may now consider penalized likelihood to determine significant components of γ which most separate the populations. Hence, we consider minimizing with respect to the parameters μ and γ the function:

$$\Lambda(\mu,\gamma) = -\sum_{l=1}^{2}(\mu + b_l\gamma)\sum_{j=1}^{t!} n_{lj}x_{lj} + \sum_{l=1}^{2} n_l K(\mu + b_l\gamma) + \lambda(\sum_{i=1}^{t}\gamma_i^2 - c)$$

for some prescribed values of the constant c and λ. We may continue to use the normalizing constant from the von Mises-Fisher distribution to approximate $K(\theta)$. Differentiating we get

$$\frac{\partial \Lambda(\mu,\gamma)}{\partial \gamma} = -\sum_{l=1}^{2} b_l \sum_{j=1}^{t!} n_{lj}x_{lj} + 2\lambda\gamma + \sum_{l=1}^{2} n_l b_l \times \frac{\partial K(\theta_l)}{\partial \theta_l}$$

$$\frac{\partial \Lambda(\mu,\gamma)}{\partial \mu} = -\sum_{l=1}^{2}\sum_{j=1}^{t!} n_{lj}x_{lj} + \sum_{l=1}^{2} n_l \times \frac{\partial K(\theta_l)}{\partial \theta_l}$$

where

$$\frac{\partial K(\theta)}{\partial \theta} = \frac{1}{t!}\frac{\theta}{\|\theta\|}(2\pi)^{\frac{t}{2}}\left[\|\theta\|^{1-\frac{t}{2}}\left(I_{\frac{t}{2}}(\|\theta\|) - \frac{1-\frac{t}{2}}{\|\theta\|} I_{\frac{t}{2}-1}(\|\theta\|)\right) + (1-\frac{t}{2})\,\|\theta\|^{-\frac{t}{2}}\, I_{\frac{t}{2}-1}(\|\theta\|)\right]$$

Here γ_i shows the difference between the two groups with respect to their preference on Item i. A negative value of γ_i means that group 1 shows more preference for Item i compared to population 2. A positive value of γ_i means that group 2 shows more preference on Item i compared to population 1. For γ_i close to zero, there is no difference between the two groups on that item. As we shall see, this interpretation is consistent with the results in the real data applications. From the definition, we know that μ is the common part of θ_1 and θ_2. More specifically, μ is the weight average of θ_1 and θ_2 taking into account the sample sizes of the populations.

5.2. Application to real data

Two-sample Sutton data

For the first example, we consider the Sutton data ($t = 3$) found in Table 15.

Table 15: Sutton data on leisure preferences (two-sample problem)

rankings	(123)	(132)	(213)	(231)	(312)	(321)
Frequencies for white females	0	0	1	0	7	6
Frequencies for black females	1	1	0	5	0	6

We applied penalized likelihood and the results are shown in Table 16.

Table 16: Maximum penalized likelihood estimation for the two-sample Sutton data

Choice of c	γ_1	γ_2	γ_3	μ_1	μ_2	μ_3	$\Lambda(\mu,\gamma)$
c=0.5	0.34	-0.57	0.24	0.59	-0.07	-0.52	46.88
c=1	0.48	-0.81	0.34	0.58	-0.06	-0.52	46.38
c=2	0.67	-1.15	0.48	0.57	-0.06	-0.51	46.46
c=10	1.50	-2.57	1.07	0.47	-0.04	-0.43	58.73
MLE	0.56	-0.95	0.40	0.58	-0.06	-0.52	46.30

Table 17: Two-Sample Sutton data and the estimation of μ, γ

Item:	No. white female	Diff.. for white	No. black female	Diff. for black	Sum	Action	γ	μ
Male	0	-13	2	-4	2	give rank 1	0.48	0.58
	13		6		19	give rank 3		
Female	8	8	0	-6	8	give rank 1	-0.81	-0.06
	0		6		6	give rank 3		
Both	6	5	11	10	17	give rank 1	0.34	-0.52
	1		1		2	give rank 3		

Table 18: Average rank for the game data

Average Rank	Xbox	PlayStation	PSPortable	GameCube	GameBoy	PC
Frequent player	2.59	2.59	3.94	4.69	5.04	2.14
Seldom player	2.86	3.17	3.64	3.88	4.24	3.21

We rearranged one of our estimation results (c=1) and the original data in Table 17. First, it is easy to see that μ is just like the θ's in the one-sample problem. For example, μ_3 is the smallest value and the whole population prefers Item "both" best. μ_3 is largest and the whole population mostly dislikes Item "male". This is not surprising since we know that μ is the common part of θ_1 and θ_2. For the parameter γ, we first consider Item "female". We see that white females prefer to spend leisure time with females (8 assign rank 1) whereas black females do not (6 give rank 3). We find that γ_2 is negative and is largest in absolute value. There is a significant difference between the opinions with respect to Item 2 "female". For Item "male" and "both", we find black females prefer them more than white females. To conclude, the results are consistent with the interpretation of μ and γ.

Game data

For the second example, we consider the Game data ($t = 6$) (Fok, Paap, and Van Dijk 2012). In this data, 91 Dutch students were asked to consider buying a new platform to play computer games. They had to rank 6 different platforms suitable to play computer games. The 6 platforms are the X-box (360), the PlayStation (2 or 3), the Gamecube (or Wii), the PlayStation Portable, the Gameboy or regular PC. In addition, we know the average number of hours that each student spends on gaming each week. We separate the students into two groups: frequent player (49 students) and seldom player (42 students). We classify a student as a frequent player if the average number of hours that he spends on gaming each week is larger than two. Otherwise he is classified as a seldom player

The game data is too large to exhibit here. Instead. we present the average ranks for the two groups in Table 18.

We applied penalized likelihood in this situation and the results are shown in Table 19.

To better illustrate the estimation results and the data, we calculate the weighted average of the observed probability of giving the high rank compared with the low rank for each item. The weight for each rank is $i - \frac{t+1}{2}$, $i = 1, ..., t$ (-2.5,-1.5,-0.5,0.5,1.5,2.5) here. The observations giving ranks close to the top and the bottom will get higher weight. A negative weighted average means preference in this case. We summary the results in table 20. First, for the parameter μ, there is a strong relationship with the weighted average of the total population. For the total population, GameBoy receives the most low ranks and μ_5 is the largest, which is consistent with our interpretation of μ. Then, it can be found that the results for γ are consistent with the trend of the difference of weighted average between two samples. For example, for the Item "PC", frequent players exhibits the largest difference of

Table 19: Maximum penalized likelihood estimation for the game data

Choice of c		Xbox	PlayStation	PSPortable	GameCube	GameBoy	PC	$\Lambda(\mu,\gamma)$
c=0.5	γ	0.00	-0.23	0.12	0.34	0.26	-0.50	-228.28
	μ	-2.66	-2.16	1.01	2.74	3.94	-2.88	
c=1	γ	0.09	-0.31	0.17	0.48	0.31	-0.74	-229.52
	μ	-2.67	-2.15	1.01	2.72	3.94	-2.84	
c=2	γ	0.27	-0.41	0.24	0.66	0.32	-1.08	-230.74
	μ	-2.70	-2.13	0.99	2.69	3.93	-2.78	
c=10	γ	1.24	-0.73	0.47	1.32	0.14	-2.44	-231.87
	μ	-2.84	-2.04	0.94	2.53	3.91	-2.50	
MLE	γ	0.96	-0.64	0.40	1.14	0.20	-2.06	-232.06
	μ	-2.81	-2.07	0.96	2.58	3.92	-2.58	

Table 20: Weighted average of observed probability and the estimation for the game data

	Weighted average of observed probability giving the high rank compare to the low rank				parameter	
Item:	Frequent players	Seldom players	Difference	Total populations	γ	μ
Xbox	-0.91	-0.64	-0.27	-0.79	0.09	-2.67
PlayStation	-0.91	-0.33	-0.57	-0.64	-0.31	-2.15
PSPortable	0.44	0.14	0.30	0.30	0.17	1.01
GameCube	1.19	0.38	**0.81**	0.82	**0.48**	2.72
GameBoy	1.54	0.74	0.80	**1.17**	0.31	**3.94**
PC	-1.36	-0.29	**-1.07**	-0.86	**-0.74**	-2.84

opinion with seldom players and the γ_5 has the largest absolute value among γ. To conclude, our application on Game data also shows that our methodology is consistent with the data.

6. Conclusion

In this paper, we considered both the Spearman and Kendall score functions in the one or two sample problems as a way to summarize ranking data. A parametric model was formulated and then we applied penalized likelihood on the original parametric model to narrow down the items being ranked. We used the von Mises-Fisher distribution to approximate the normalizing constant and then determined the MLE estimates in several examples. This estimation procedure is fast and simple. In all cases, the estimation is shown to be consistent with the data. Our methodology was applied on various popular ranking data sets.

Acknowledgments

This research was supported by a discovery grant from the Natural Sciences and Engineering Research Council of Canada OGP0009068. As well, the research of Hang Xu and Mayer Alvo was also supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No.17303515).

References

Aslam Ja, Montague M (2001). "Models for Metasearch." in *Proceedings of the 24th Annual*

International Acm Sigir Conference on Research and Development in Information Retrieval, Pp. 276–284. Acm.

Baba Y (1986). “Graphical Analysis of Rank Data.” *Behaviormetrika*, **19**, 1–15.

Broyden CG (1970). “The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations.” *IMA Journal of Applied Mathematics*, **6**(1), 76–90.

Byrd RH, Gilbert JC, Nocedal J (2000). “A Trust Region Method Based on Interior Point Techniques For Nonlinear Programming.” *Mathematical Programming*, **89**(1), 149–185.

Byrd RH, Hribar ME, Nocedal J (1999). “An Interior Point Algorithm for Large-scale Nonlinear Programming.” *SIAM Journal on Optimization*, **9**(4), 877–900.

Cohen A, Mallows C (1980). “Analysis of Ranking Data.” *Technical memorandum*, AT&T Bell Laboratories, Murray Hill, N.J.

Coleman T, Li Y (1994). “On the Convergence of Reflective Newton Methods for Large-scale Nonlinear Minimization Subject to Bounds vol. 67.” *Ithaca, NY, USA: Cornell University.*

Coleman TF, Li Y (1996). “An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds.” *SIAM Journal on Optimization*, **6**(2), 418–445.

Critchlow DE, Fligner MA, Verducci JS (1991). “Probability Models on Rankings.” *Journal of Mathematical Psychology*, **35**(3), 294–318.

Croon MA (1989). “Latent Class Models for the Analysis of Rankings.” *Advances in Psychology*, **60**, 99–121.

DeConde RP, Hawley S, Falcon S, Clegg N, Knudsen B, Etzioni R (2006). “Combining Results of Microarray Experiments: a Rank Aggregation Approach.” *Statistical Applications in Genetics and Molecular Biology*, **5**(1).

Deng K, Han S, Li KJ, Liu JS (2014). “Bayesian Aggregation of Order-based Rank Data.” *Journal of the American Statistical Association*, **109**(507), 1023–1039.

Dwork C, Kumar R, Naor M, Sivakumar D (2001). “Rank Aggregation Methods for the Web.” In *Proceedings of the 10th International Conference on World Wide Web*, pp. 613–622. ACM.

Feigin PD, Cohen A (1978). “On a Model for Concordance Between Judges.” *Journal of the Royal Statistical Society Series B*, **40**, 203–213.

Fletcher R (1970). “A New Approach to Variable Metric Algorithms.” *The Computer Journal*, **13**(3), 317–322.

Fok D, Paap R, Van Dijk B (2012). “A Rank-Ordered Logit Model With Unobserved Heterogeneity In Ranking Capabilities.” *Journal of Applied Econometrics*, **27**(5), 831–846.

Goldfarb D (1970). “A Family of Variable-metric Methods Derived by Variational Means.” *Mathematics of Computation*, **24**(109), 23–26.

Kemeny JG, Snell JL (1962). *Mathematical Models in the Social Sciences*, volume 9. Ginn New York.

Lee PH, Philip L (2012). “Mixtures of Weighted Distance-based Models for Ranking Data With Applications in Political Studies.” *Computational Statistics & Data Analysis*, **56**(8), 2486–2500.

Mallows CL (1957). “Non-null Ranking Models. I.” *Biometrika*, **44**, 114–130.

McCullagh P (1993). "Models on Spheres and Models for Permutations." In MA Fligner, JS Verducci (eds.), *Probability Models and Statistical Analyses for Ranking Data*, pp. 278–283. Springer-Verlag.

Neyman J (1937). "Smooth Test for Goodness of Fit." *Skandinavisk Aktuarietidskrift*, **20**, 149–199.

Shanno DF (1970). "Conditioning of Quasi-newton Methods for Function Minimization." *Mathematics of Computation*, **24**(111), 647–656.

Waltz RA, Morales JL, Nocedal J, Orban D (2006). "An Interior Algorithm for Nonlinear Optimization That Combines Line Search and Trust Region Steps." *Mathematical Programming*, **107**(3), 391–408.

Affiliation:

Mayer Alvo
Department of Mathematics and Statistics
University of Ottawa
585 King Edward Ave, Ottawa, ON K1N 6N5, Canada
E-mail: malvo@uOttawa.ca
URL: http://science.uottawa.ca/mathstat/en/people/alvo-mayer

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 46 *Submitted:* 2015-10-08
February 2017 *Accepted:* 2016-01-19

Austrian Journal of Statistics
February 2017, Volume 46, 33–39.
http://www.ajs.or.at/
doi:10.17713/ajs.v46i1.139

A New Extended Burr XII Distribution

Indranil Ghosh*
University of North Carolina Wilmington

Marcelo Bourguignon
Universidade do Rio Grande do Norte

Abstract

In this paper, we propose a new lifetime distribution, namely the extended Burr XII distribution (using the technique as mentioned in Cordeiro et al. (2015)). We derive some basic properties of the new distribution and provide a Monte Carlo simulation study to evaluate the maximum likelihood estimates of model parameters. For illustrative purposes, two real life data sets have been considered as an application of the proposed model.

Keywords: Burr XII distribution, estimation, half logistic family of distributions.

1. Introduction

The armory of statistical distributions is truly illimitable. New distributions are being unearthed literally on a weekly basis elicited by either theoretical considerations or by pressing practical applications or both. However, in various spheres of applied areas, for example, survival data analysis, finance and risk modeling, insurance, modeling rare events and biology, it has become imperative to develop an extended class of classical distributions to enhance its flexibility in modeling real life data that appears to have high degree of skewness and kurtosis. In the last few years, new classes of distributions were defined by extending the Weibull distribution to cope with bathtub shaped failure rates. Mudholkar and Srivastava (1993) and Mudholkar et al. (1996) pioneered and studied the exponentiated Weibull distribution to analyze bathtub failure data. A good review of some of these extended models is presented in Pham and Lai (2007).

In this paper, we consider the technique of mixing two absolutely continuous distributions. We consider here the particular case of the general Type I half logistic family of distributions, studied by Cordeiro et al. (2015). The cumulative distribution function (cdf) and the probability density function (pdf) is given by

$$F(x;\lambda,\xi) = \int_0^{-\log[1-G(x;\xi)]} \frac{2\,\lambda\,\mathrm{e}^{-\lambda t}}{(1+\mathrm{e}^{-\lambda t})^2}dt = \frac{1-[1-G(x;\xi)]^{\lambda}}{1+[1-G(x;\xi)]^{\lambda}},$$

where $G(x;\xi)$ is the baseline cdf depending on a parameter vector ξ and $\lambda > 0$ is an additional shape parameter. Also, the corresponding pdf will be

$$f(x;\lambda,\xi) = \frac{2\,\lambda\,g(x;\xi)\,[1-G(x;\xi)]^{\lambda-1}}{\left\{1+[1-G(x;\xi)]^{\lambda}\right\}^2}.$$

Next, if we consider the baseline distribution as $G(x) = 1-(1+x^c)^{-1}$ (Burr XII distribution), then, the pdf and the cdf for a extended Burr XII distribution with parameters c and λ are respectively,

$$f(x;\lambda,c) = \frac{2\,c\,\lambda\,x^{c-1}(1+x^c)^{\lambda-1}}{\left[1+(1+x^c)^{\lambda}\right]^2}, \quad x>0, \quad c>0, \tag{1}$$

and

$$F(x;\lambda,c) = 1-\frac{2}{1+(1+x^c)^{\lambda}}. \tag{2}$$

If a random variable X has the density as in (1), then we say that X follows Extended Burr XII (henceforth EBXII in short) distribution with parameters c and λ. One significant usefulness of Equation (1) is its ability to analyze skewed data that can not be properly fitted by its parent distributions. Furthermore, it permits greater flexibility in its tails and can be widely applied in many areas of reliability and biology. The proposed distribution has only two parameters and the density and the cumulative distribution functions of the new distribution are simple, i.e., the EBXII distribution is in fact very tractable. We mention here an important fact that our derived model, although bears the same name as in Shao et al. (2004) and Mudholkar et al. (1996), who alternatively called this distribution as generalized Weibull, is completely different from all of them in terms of the genesis of the model and subsequently its stochastic properties.

The rest of the paper is organized as follows. In Section 2, we discuss briefly some properties of the extended Burr XII (henceforth EBXII) distribution. These include reliability parameter and order statistics. In Section 3, we discuss the likelihood inference for the EBXII and conduct a simulation study for specific choices of the model parameters. In Section 4, we illustrate the applicability of the EBXII with two data sets and compared with the rival Burr XII distribution. Finally, in Section 5, we provide some concluding remarks.

2. Properties of the extended Burr XII distribution

In this section, we consider some properties of the EBXII distribution, such as the reliability parameter and order statistics. Hereafter, the random variable X following Equation (2) with parameters c and λ is denoted by $X \sim \mathrm{EBXII}(c,\lambda)$.

The reliability parameter R is defined as $R = P(X > Y)$, where X and Y are independent random variables. If X and Y are two continuous and independent random variables with the cdfs $F_1(x)$ and $F_2(y)$ and their pdfs $f_1(x)$ and $f_2(y)$ respectively. Then the reliability parameter R can be written as

$$R = P(X > Y) = \int_{-\infty}^{\infty} F_2(t) f_1(t)dt.$$

Theorem 1. Suppose that $X \sim EBXII(c,\lambda_1)$ and $Y \sim EBXII(c,\lambda_2)$, and they are independent. Then

$$P(X>Y) = 1-4B\left(\frac{\lambda_2}{\lambda_1}+1, 1-\frac{\lambda_2}{\lambda_1}\right).$$

Proof: From (1) and (2), we have

$$\begin{aligned} P(X>Y) &= \int_0^{\infty}\left[1-\frac{2}{1+(1+t^c)^{\lambda_2}}\right]\left\{\frac{2\,c\,\lambda\,t^{c-1}(1+t^c)^{\lambda_1-1}}{\left[1+(1+t^c)^{\lambda_1}\right]^2}\right\}dt \\ &= 1-4\int_1^{\infty} u^{-2}(u-1)^{\frac{\lambda_2}{\lambda_1}}du, \quad \text{on substitution} \quad u = 1+(1+t^c)^{\lambda_1} \\ &= 1-4B\left(1+\frac{\lambda_2}{\lambda_1}, 1-\frac{\lambda_2}{\lambda_1}\right). \end{aligned}$$

Now, we will consider the expression for the general r-th order statistic and the large sample distribution of the sample minimum and the sample maximum when a random sample of size n are drawn from the EBXII distribution. The density function of the r-th order statistic, $X_{r:n}$, for a random sample of size n drawn from (1) is given by

$$f_{X_{r:n}}(x) \quad = \quad \frac{f(x)}{B(r,n-r+1)} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} [1-F(x)]^{n-r+j}\, I(0<x<\infty).$$

Thus the pdf of $X_{r:n}$ can be alternatively written as

$$f_{r:n}(x) = \frac{1}{B(r,n-r+1)} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \frac{2^{n-r+j+1} c\lambda x^{c-1}(1+x^c)^{\lambda-1}}{[1+(1+x^c)^\lambda]^{2+n-r+j}} I(0<x<\infty).$$

In order to derive the asymptotic distribution of the sample minima $X_{1:n}$, we consider Theorem 8.3.6 of Arnold et al. (2008). Observe that, since $F^{-1}(0)=0$, it follows from the theorem that the asymptotic distribution of the sample minima $X_{1:n}$ is not of Fréchet type. The asymptotic distribution of $X_{1:n}$ will be of Weibull type with parameter $\delta > 0$ if

$$\lim_{\varepsilon \to 0_+} \frac{F(\varepsilon x)}{F(\varepsilon)} = x^\delta, \quad \text{for all} \quad x>0.$$

By using L'Hôpital's rule, it follows that

$$\lim_{\varepsilon \to 0_+} \frac{F(\varepsilon x)}{F(\varepsilon)} = x \lim_{\varepsilon \to 0_+} \frac{f(\varepsilon x)}{f(\varepsilon)} = x \lim_{\varepsilon \to 0_+} \frac{\frac{2c\lambda(\varepsilon x)^{c-1}(1+(\varepsilon x)^c)^{\lambda-1}}{\left(1+(1+(\varepsilon x)^c)^\lambda\right)^2}}{\frac{2c\lambda \varepsilon^{c-1}(1+x^c)^{\lambda-1}}{\left(1+(1+\varepsilon^c)^\lambda\right)^2}} = x^c \lim_{\varepsilon \to 0_+} \frac{\frac{(1+(\varepsilon x)^c)^{\lambda-1}}{\left(1+(1+(\varepsilon x)^c)^\lambda\right)^2}}{\frac{(1+\varepsilon^c)^{\lambda-1}}{\left(1+(1+\varepsilon^c)^\lambda\right)^2}}.$$

Hence,

$$\lim_{\varepsilon \to 0_+} \frac{F(\varepsilon x)}{F(\varepsilon)} = x^c.$$

Hence we obtain that the asymptotic distribution of the sample minima $X_{1:n}$ is of the Weibull type with shape parameter c. The asymptotic distribution of the sample maxima $X_{n:n}$, can be viewed as $F_n(x)$, where $F_n(x) = 1 - F_1(-x)$, where $F_1(.)$ is the cdf of $X_{1:n}$.

3. Maximum likelihood estimation

In this section, we address the parameter estimation of the EBXII(c,λ) under the classical set up. Let x_1, x_2, $\cdots$, x_n be a random sample of size n drawn from the density in (1). The log-likelihood function is given by

$$\ell = n\log(2\,c\,\lambda) + (c-1)\sum_{i=1}^n \log(x_i) + (\lambda-1)\sum_{i=1}^n \log\left(1+x_i^c\right) - 2\sum_{i=1}^n \log\left[1+(1+x_i^c)^\lambda\right]. \tag{3}$$

The derivatives of (3) with respect to c and λ are given by

$$\frac{\partial}{\partial c}\ell = \frac{n}{c} + \sum_{i=1}^n \log(x_i) + \sum_{i=1}^n x_i^c\,(1+x_i^c)^{-1}\log(x_i)\left\{(\lambda-1) - 2\,\lambda\,(1+x_i^c)^\lambda\,[1+(1+x_i^c)^\lambda]^{-1}\right\}. \tag{4}$$

$$\frac{\partial}{\partial \lambda}\ell = \frac{n}{\lambda} + \sum_{i=1}^n \log(1+x_i^c) - 2\sum_{i=1}^n [1+(1+x_i^c)^\lambda]^{-1}\,(1+x_i^c)^\lambda \log(1+x_i^c). \tag{5}$$

The MLE $\hat{c}$ and $\hat{\lambda}$ are obtained by setting (4) and (5) to zero and solving them simultaneously.

Next, a small Monte Carlo simulation experiment is conducted to evaluate the maximum likelihood estimation of the EBXII distribution parameters. We set the sample size at $n = 50, 100, 200, 400$ and 800. The Monte Carlo simulation experiments are performed using the R programming language; see http://www.r-project.org. All results were obtained from 10,000 Monte Carlo replications and the simulations were carried out using the R programming language; see http://www.r-project.org.

Table 1 reports the empirical mean and the mean squared error (in parentheses) of the corresponding estimator. From this table, note that, as the sample size increases, the empirical bias and mean squared error decrease in all the cases analyzed, as expected.

Table 1: Empirical means and mean squared errors (in parentheses).

c	λ	$\widehat{c}$	$\widehat{\lambda}$
		$n = 50$	
0.5	0.5	0.5495 (0.0465)	0.4911 (0.0204)
1.5	0.5	1.6357 (0.2732)	0.4839 (0.0448)
0.5	1.5	0.5179 (0.0058)	1.5051 (0.0469)
1.5	1.5	1.5442 (0.0583)	1.5067 (0.0557)
		$n = 100$	
0.5	0.5	0.5188 (0.0082)	0.4968 (0.0096)
1.5	0.5	1.5609 (0.0824)	0.4937 (0.0181)
0.5	1.5	0.5075 (0.0026)	1.5021 (0.0234)
1.5	1.5	1.5210 (0.0264)	1.5017 (0.0269)
		$n = 200$	
0.5	0.5	0.5096 (0.0035)	0.4985 (0.0036)
1.5	0.5	1.5308 (0.0359)	0.4931 (0.0173)
0.5	1.5	0.5041 (0.0013)	1.4999 (0.0117)
1.5	1.5	1.5101 (0.0107)	1.5030 (0.0110)
		$n = 400$	
0.5	0.5	0.5049 (0.0016)	0.4989 (0.0018)
1.5	0.5	1.5145 (0.0162)	0.4976 (0.0058)
0.5	1.5	0.5020 (0.0006)	1.5007 (0.0054)
1.5	1.5	1.5066 (0.0054)	1.5001 (0.0054)
		$n = 800$	
0.5	0.5	0.5023 (0.0008)	0.4995 (0.0009)
1.5	0.5	1.5066 (0.0076)	0.4993 (0.0023)
0.5	1.5	0.5010 (0.0003)	1.5003 (0.0027)
1.5	1.5	1.5037 (0.0027)	1.4996 (0.0027)

4. Applications

For illustrative purposes, we consider two data sets and compare with the Burr XII (BXII) distribution. For each data set, we estimate the unknown parameters of each distribution by the maximum-likelihood method (as discussed in Section 3) and all the computations were done using the subroutine NLMixed of the SAS software. We obtain the values of the Akaike information criterion (AIC), Bayesian information criterion (BIC) and consistent Akaike information criterion (CAIC). First, we describe the three data sets:

• Data set I: The first example consist of thirty successive values of March precipitation (in inches) in Minneapolis/St Paul (Hinkley 1977). The data are: 0.77, 1.74, 0.81, 1.2, 1.95, 1.2, 0.47, 1.43, 3.37, 2.2, 3, 3.09, 1.51, 2.1, 0.52, 1.62, 1.31, 0.32, 0.59, 0.81, 2.81, 1.87, 1.18, 1.35, 4.75, 2.48, 0.96, 1.89, 0.9, 2.05.

• Data set II: In the second data set, we consider vinyl chloride data obtained from clean upgradient monitoring wells in mg/L; this data set was used for Bhaumik et al. (2009). The data are: 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, 2.0, 4.0, 6.8, 1.2, 0.4, 0.2.

Tables 2 and 3 provides some descriptive statistics and the MLEs (with corresponding standard errors in parentheses) of these two data sets. Since the values of the AIC, BIC and CAIC are smaller for the EBXII distribution compared with those values of the BXII model, this new distribution seems to be a very competitive model for these data.

Table 2: Descriptive statistics for the two data sets.

	n	Min.	Q_1	Q_2	Mean	Q_3	Max.	Var.
Set I	30	0.320	0.915	1.470	1.675	2.088	4.750	1.0012
Set II	34	0.100	0.500	1.150	1.879	2.475	8.0	3.8126

Table 3: MLEs of the model parameters with corresponding SE's (given in parentheses) for the two data sets and the corresponding AIC, CAIC and BIC statistics.

Set I	$\widehat{c}$	$\widehat{\lambda}$	$\widehat{k}$	AIC	BIC	CAIC
EBXII	2.8689	0.8811	–	82.4	85.2	82.8
	(0.5977)	(0.1979)	(–)			
BXII	3.2555	–	0.5770	84.5	87.3	85.0
	(0.6455)	(–)	(0.1371)			
Set II	$\widehat{c}$	$\widehat{\lambda}$	$\widehat{k}$	AIC	BIC	CAIC
EBXII	1.3457	1.3837	–	115.4	118.5	115.8
	(0.2405)	(0.2405)	(–)			
BXII	1.5621	–	0.9305	116.2	119.2	116.5
	(0.2479)	(–)	(0.1791)			

Plots of the pdf of the EBXII and BXII fitted models to these data are displayed in Figure 1. They indicate that the EBXII distribution is superior to the BXII distribution in terms of model fitting. Based on these plots, we conclude that the EBXII distribution provides a better fit to these data than the BXII model.

5. Concluding remarks

There has been a growing interest among statisticians and applied researchers in developing flexible lifetime models for the betterment of modeling survival data. In this paper, we introduce a two parameter extended Burr XII distribution which is obtained by considering a Burr XII distribution as the baseline cdf in the Cordeiro et al. (2015) model. We study some of its statistical and mathematical properties. Parameter estimation is approached by maximum likelihood. The usefulness of the new distribution is illustrated in an analysis of two real data sets. We hope that the proposed extended model will invite wider applications in survival analysis.

Acknowledgments

The authors are grateful to the editor and the two referees for their helpful comments and suggestions which has significantly improved this paper.

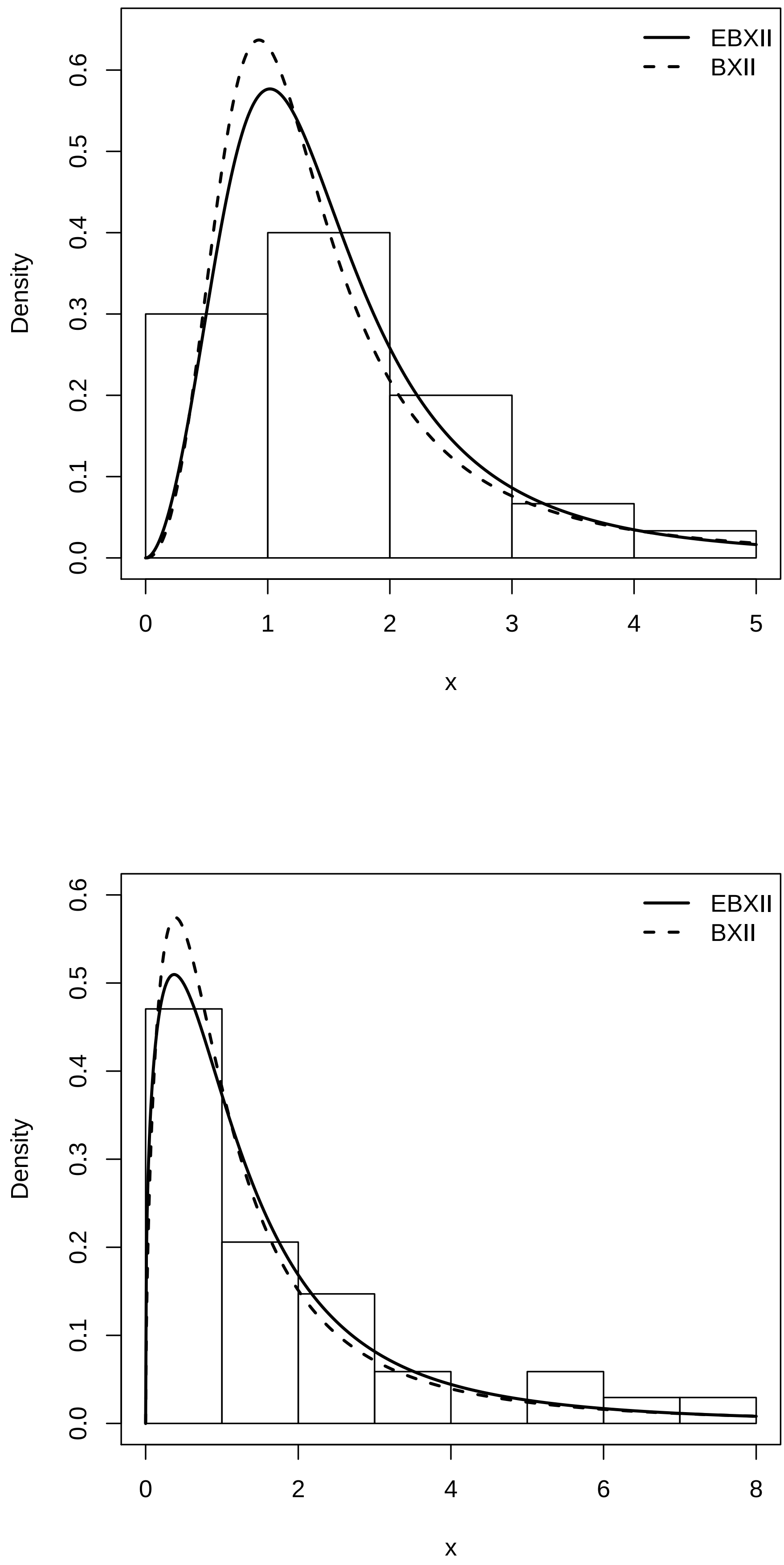


Figure 1: Plots of the fitted densities of the EBXII and BXII distributions for the first and second data sets.

References

Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (2008). *A First Course in Order Statistics*, Classic Edition. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Cordeiro, G.M., Alizadeh, M. and Marinho, P.D. (2015). The Type I Half-Logistic Family of Distributions, *Journal of Statistical Computation and Simulation*, **86**, 707–728.

Bhaumik, D.K., Kapur, K., Gibbons, R.D. (2009). Testing Parameters of a Gamma Distribution for Small Samples. *Technometrics*, **51**, 326–334.

Hinkley, D. (1977). On Quick Choice of Power Transformation. *Applied Statistics*, **26**, 67–69.

Mudholkar, G.S. and Srivastava, D.K. (1993). Exponentiated Weibull Family Analyzing Bathtub Failure-Rate Data. *IEEE Transactions and Reliability*, **42**, 299–302.

Mudholkar, G.S., Srivastava, D. K. and Kollia, G.D. (1996). A Generalization of the Weibull Distribution With Application to the Analysis of Survival Data. *Journal of the American Statistical Association*, **91**, 1575–1583.

Pham, H. and Lai, C.D. (2007). On Recent Generalizations of the Weibull Distribution. *IEEE Transactions on Reliability*, **56**, 454–458.

Shao, Q.X., Wong, H., Xia, J. and I, W.C. (2004). Models for Extreme Using the Extended Three-Parameter Burr XII System With Application to Flood Frequency Analysis. *Hydrological Sciences Journal*, **49**, 685–702.

Affiliation:

* Corresponding Author Indranil Ghosh
Department of Mathematics & Statistics
University of North Carolina Wilmington
Wilmington, NC, USA
E-mail: ghoshi@uncw.edu

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 46 *Submitted:* 2015-10-20
February 2017 *Accepted:* 2016-04-13

Austrian Journal of Statistics
February 2017, Volume 46, 41–63.
http://www.ajs.or.at/
doi:10.17713/ajs.v46i1.144

The Beta Exponential Fréchet Distribution with Applications

M. E. Mead
Zagazig University

Ahmed Z. Afify
Benha University

G. G. Hamedani
Marquette University

Indranil Ghosh
UNCW

Abstract

We define and study a new generalization of the Fréchet distribution called the beta exponential Fréchet distribution. The new model includes thirty two special models. Some of its mathematical properties, including explicit expressions for the ordinary and incomplete moments, quantile and generating functions, mean residual life, mean inactivity time, order statistics and entropies are derived. The method of maximum likelihood is proposed to estimate the model parameters. A small simulation study is also reported. Two real data sets are applied to illustrate the flexibility of the proposed model compared with some nested and non-nested models.

Keywords: generating function, maximum likelihood, entropy, Fréchet distribution, beta exponential–G family.

1. Introduction

In the past few decades, many generators have been proposed by extending some useful statistical distributions. Such generated families of distributions have been extensively used for modeling and analyzing lifetime data in many applied sciences such as reliability, engineering, actuarial sciences, demography, economics, hydrology, biological studies, insurance, medicine and finance, among others. However, there still remain many real world phenomena involving data, which do not follow any of the classical statistical distributions.

The Fréchet distribution was proposed to model extreme events such as annually maximum one-day rainfalls and river discharges by Fréchet (1924). This distribution has found wide application in extreme value theory. Further details about the Fréchet distribution can be found in Kotz and Nadarajah (2000).

Some extensions of the Fréchet distribution are available in the literature, such as the exponentiated Fréchet (EFr) (Nadarajah and Kotz, 2003), beta Fréchet (BFr) (Nadarajah and Gupta, 2004 and Barreto-Souza et al., 2011), transmuted Fréchet (TFr) (Mahmoud and Mandouh, 2013), gamma extended Fréchet (GEFr) (Silva et al., 2013), Marshall-Olkin Fréchet (Krishna et al., 2013), transmuted exponentiated Fréchet (TEFr) (Elbatal et al., 2014), Kumaraswamy Fréchet (Kw-Fr) (Mead and Abd-Eltawab, 2014), transmuted Marshall-Olkin

Fréchet (TMOFr) (Afify et al., 2015a) and Weibull Fréchet (WFr) (Afify et al., 2016).

The cumulative distribution function (cdf) and probability density function (pdf) of the Fréchet (Fr) distribution are, respectively, given by (for $x > 0$)

$$G_{\mathrm{Fr}}(x;\theta,\beta) = e^{-\left(\frac{\theta}{x}\right)^{\beta}} \text{ and } g_{\mathrm{Fr}}(x;\theta,\beta) = \beta\theta^{\beta}x^{-\beta-1}e^{-\left(\frac{\theta}{x}\right)^{\beta}}, \tag{1}$$

where $\theta > 0$ is a scale parameter and $\beta > 0$ is a shape parameter, respectively.

The aim of this paper is to provide another extension of the Fréchet model using the Beta exponential-G (BEx-G) family of distributions proposed by Alzaatreh et al. (2013). So, we propose the new *beta exponential Fréchet* (BExFr for short) distribution by adding three extra shape parameters to the Fréchet distribution. The objective of this work is to study some mathematical properties of the five-parameter BExFr model with the hope that it will attract wider applications in reliability, engineering and other areas of research.

For an arbitrary baseline cdf $G(x)$, Alzaatreh et al. (2013) defined the BEx-G family of distributions by the cdf and pdf

$$F\left(x;a,b,\lambda\right) = \frac{1}{B\left(a,b\right)}B\left(1-\left(1-G(x)\right)^{\lambda};a,b\right) \tag{2}$$

and

$$f(x;a,b,\lambda) = \frac{\lambda g\left(x\right)}{B(a,b)}\left[1-G\left(x\right)\right]^{\lambda b-1}\left\{1-\left[1-G\left(x\right)\right]^{\lambda}\right\}^{a-1}, \tag{3}$$

respectively, where $g(x) = dG(x)/dx$ and a, b and λ are three extra positive shape parameters, $B\left(z;a,b\right) = \int_0^z t^{a-1}\left(1-t\right)^{b-1}dt$ is the incomplete beta function, $B\left(a,b\right) = \Gamma\left(a\right)\Gamma\left(b\right)/\Gamma\left(a+b\right)$ and $\Gamma\left(.\right)$ is the gamma function. Clearly, when $a = b = \lambda = 1$, we obtain the baseline distribution. If X is a random variable with pdf (3), we write $X \sim$ BEx-G(a,b,λ). An attractive feature of this model is that these parameters can afford greater control over the weights in both tails and in its center.

Next, we consider the Fr model in order to define the new distribution by taking $G(x)$ in (2) to be the cdf in (1) of the Fr distribution. Then, the cdf, say $F(x) = F(x;a,b,\lambda,\beta,\theta)$, of the BExFr distribution (for $x > 0$) reduces to

$$F(x) = \frac{1}{B\left(a,b\right)}B\left(1-\left(1-e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)^{\lambda};a,b\right). \tag{4}$$

The corresponding pdf follows by inserting (1) in equation (3)

$$f\left(x\right) = \frac{\lambda\beta\theta^{\beta}}{B(a,b)}x^{-\beta-1}e^{-\left(\frac{\theta}{x}\right)^{\beta}}\left[1-e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right]^{\lambda b-1}\left\{1-\left[1-e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right]^{\lambda}\right\}^{a-1}. \tag{5}$$

Henceforth, $X \sim$BExFr$(a,b,\lambda,\beta,\theta)$ denotes a random variable having density function (5).

The survival function (sf) and hrf of X are, respectively, given by

$$S(x) = 1-\frac{1}{B(a,b)}B\left(1-\left(1-e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)^{\lambda};a,b\right)$$

and

$$h(x) = \frac{\lambda\beta\theta^{\beta}x^{-\beta-1}e^{-\left(\frac{\theta}{x}\right)^{\beta}}\left[1-e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right]^{\lambda b-1}\left\{1-\left[1-e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right]^{\lambda}\right\}^{a-1}}{\left\{B(a,b)-B\left(1-\left(1-e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)^{\lambda};a,b\right)\right\}}.$$

The BExFr distribution appears to have the ability to model failure rate models which are quite common in reliability and biological studies. Furthermore, a possible application of the

BExFr distribution could be in modeling ordinal data as a latent response models. Ordinal data are commonly used in many areas of application, some examples being the quality of an item or service or performance (poor, fair, good, very good or excellent), seriousness of a defect (minor, major, critical), taste of food (too mild, just right, too spicy) and extent of agreement (strongly disagree, disagree, neutral, agree, strongly agree). Many techniques are available for analyzing stochastic shifts in ordinal data; for a review see Agresti (1984). However, serious difficulties arise when inferences are desired on both location and dispersion effects; see Nair (1986) and Hamada and Wu (1990) and the accompanying discussions. The main cause of difficulty in separating the location effects from dispersion effects when the data are ordinal is that the number of categories is usually small (between 3 and 10). Therefore, when the location parameter is pushed to the limit (either too high or too low), most of the data fall in the extreme category giving a false impression of reduced variance. A common approach to the analysis of ordinal data is to assume a continuous latent response distribution that is observed through windows of ordered intervals with fixed, but unknown cutpoints. This approach is implicit in the proportional odds model (McCullagh, 1980), which can be derived from an underlying logistic response distribution and in other generalized linear models (McCullagh and Nelder, 1989). These models are typically based on the assumption of a symmetric continuous latent response having an infinite domain.

To resolve these difficulties, we propose the BExFr distribution as a model for the latent variable. The following two properties of the beta distribution make it especially suitable for modeling ordinal data:

- The BExFr distribution has an infinite domain.
- The BExFr distribution can flexibly model a wide variety of shapes including a bell-shape (symmetric or skewed), U-shape and J-shape.

The BExFr distribution is a very flexible model having several special cases. It contains 32 sub-models listed in Table 1. The BExFr includes some important sub-models, namely: the beta exponential inverse Rayleigh (BExIR), beta exponential inverse exponential (BExIEx), beta Fréchet (BFr), beta inverse Rayleigh (BIR), beta inverse exponential (BIEx), exponentiated exponential Fréchet (EExFr), exponentiated exponential inverse Rayleigh (EExIR), exponentiated exponential inverse exponential (EExIEx), beta exponential generalized inverse Weibull (BExGIW) and beta exponential generalized inverse Rayleigh (BExGIR) distributions. Figure 1 displays some plots of the BExFr density for some values of the parameters a, b, λ, β and θ. Further, plots of the hrf of the new distribution are shown in Figure 2.

We provide a comprehensive description of some mathematical properties of the BExFr distribution. The paper is outlined as follows. In Section 2, we derive useful representations for the pdf and cdf of the BExFr. Some mathematical properties including the quantile function (qf), ordinary and incomplete moments, moment generating function (mgf), Rényi, Shannon and q-entropies, mean residual life (MRL) and mean inactivity time (MIT) are discussed in Section 3. In section 4, we consider order statistics for a random sample of size n drawn from the BExFr distribution. Certain characterizations are presented in Section 5. In Section 6, we obtain the maximum likelihood estimates (MLEs) of the model parameters. Section 7 deals with a small simulation study. In Section 8, the potentiality of the new model is illustrated by means of two applications to two real data sets. Finally, in Section 9, we provide some concluding remarks.

2. Mixture representation

In this section, we derive mixture representations for the pdf and cdf of X. In order to obtain

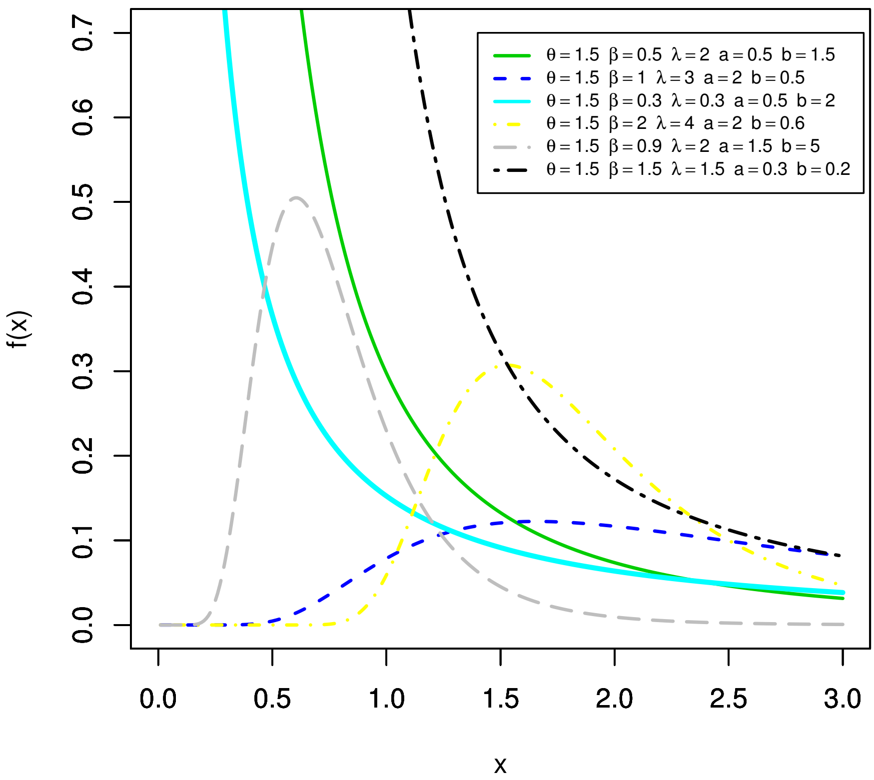


Figure 1: The plots of BExFr density function.

a simple form for the BExFr pdf, we expand (5) using the power series

$$(1-z)^{b-1} = \sum_{j=0}^{\infty} \frac{(-1)^j \,\Gamma(b)}{j!\,\Gamma(b-j)} z^j, \quad |z| < 1, \ b > 0. \tag{6}$$

Using expansion (6) in equation (5) and after some algebra, the pdf of X can be written as

$$f(x) = \frac{\lambda\beta\theta^{\beta}}{B(a,b)} x^{-\beta-1}\, e^{-\left(\frac{\theta}{x}\right)^{\beta}} \sum_{j=0}^{\infty} \frac{(-1)^j \ \Gamma(a)}{j!\,\Gamma(a-j)} \underbrace{\left[1-e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right]^{\lambda(b+j)-1}}_{A}.$$

By applying (6) in the quantity A, the last equation becomes

$$f(x) = \sum_{k=0}^{\infty} \upsilon_k \beta\,(k+1)\,\theta^{\beta} x^{-\beta-1} e^{-(k+1)\left(\frac{\theta}{x}\right)^{\beta}}, \tag{7}$$

$$\upsilon_k = \sum_{j=0}^{\infty} \frac{(-1)^{j+k} \ \lambda\Gamma(a)\,\Gamma(\lambda(b+j))}{j! \ (k+1)\,B(a,b)\Gamma(a-j)\,\Gamma(\lambda(b+j)-k)}.$$

Equation (7) can be rewritten as

$$f(x) = \sum_{k=0}^{\infty} \upsilon_k \, h_{k+1}(x), \tag{8}$$

where $h_{k+1}(x)$ is the Fréchet pdf with shape parameter β and scale parameter $\theta\,(k+1)^{1/\beta}$.

Reduced Model	Parameters					Author
	θ	β	a	b	λ	
BExIR	–	2	–	–	–	New
BExIEx	–	1	–	–	–	New
BFr	–	–	–	–	1	Nadarajah and Gupta (2004)
BIR	–	2	–	–	1	–
BIEx	–	1	–	–	1	–
EExFr	–	–	1	–	–	New
EExIR	–	2	1	–	–	New
EExIEx	–	1	1	–	–	New
EFr	–	–	1	–	1	Nadarajah and Kotz (2003)
EIR	–	2	1	–	1	–
EIEx	–	1	1	–	1	–
ExFr	–	–	1	1	–	New
ExIR	–	2	1	1	–	New
ExIEx	–	1	1	1	–	New
Fr	–	–	1	1	1	Fréchet (1924)
IR	–	2	1	1	1	Trayer (1964)
IEx	–	1	1	1	1	Keller and Kamath (1982)
BExGIW	$qc^{1/\beta}$	–	–	–	–	New
BExGIR	$qc^{1/2}$	2	–	–	–	New
BExGIEx	qc	1	–	–	–	New
BGIW	$qc^{1/\beta}$	–	–	–	1	Baharith et al. (2014)
BGIR	$qc^{1/2}$	2	–	–	1	–
BGIEx	qc	1	–	–	1	–
EExGIW	$qc^{1/\beta}$	–	1	–	–	New
EExGIR	$qc^{1/2}$	2	1	–	–	New
EExGIEx	qc	1	1	–	–	New
EGIW	$qc^{1/\beta}$	–	1	–	1	–
EGIR	$qc^{1/2}$	2	1	–	1	–
EGIEx	qc	1	1	–	1	–
GIW	$qc^{1/\beta}$	–	1	1	1	de Gusmao et al. (2011)
GIR	$qc^{1/2}$	2	1	1	1	–
GIEx	qc	1	1	1	1	–

Table 1: Sub-models of the BExFr distribution.

Equation (8) reveals that the BExFr density function can be expressed as a mixture of Fréchet densities. So, several of its structural properties can be derived from those of the Fréchet distribution.

By integrating (8), we obtain

$$F\left(x\right)=\sum_{k=0}^{\infty}\upsilon_k\, H_{k+1}\left(x\right),$$

where $H_{k+1}\left(x\right)$ is the cdf of the Fréchet model with shape parameter β and scale parameter $\theta\left(k+1\right)^{1/\beta}$.

3. Mathematical properties

Established algebraic expansions to determine some structural quantities of the BExFr distribution can be more efficient than computing those directly by numerical integration of its density function.

Let X and Y be two random variables. X is said to be stochastically greater than or equal to Y, denoted by $X \underset{st}{\geqslant} Y$ if $P(X > x) \geq P(Y > x)$ for all x in the support set of X.

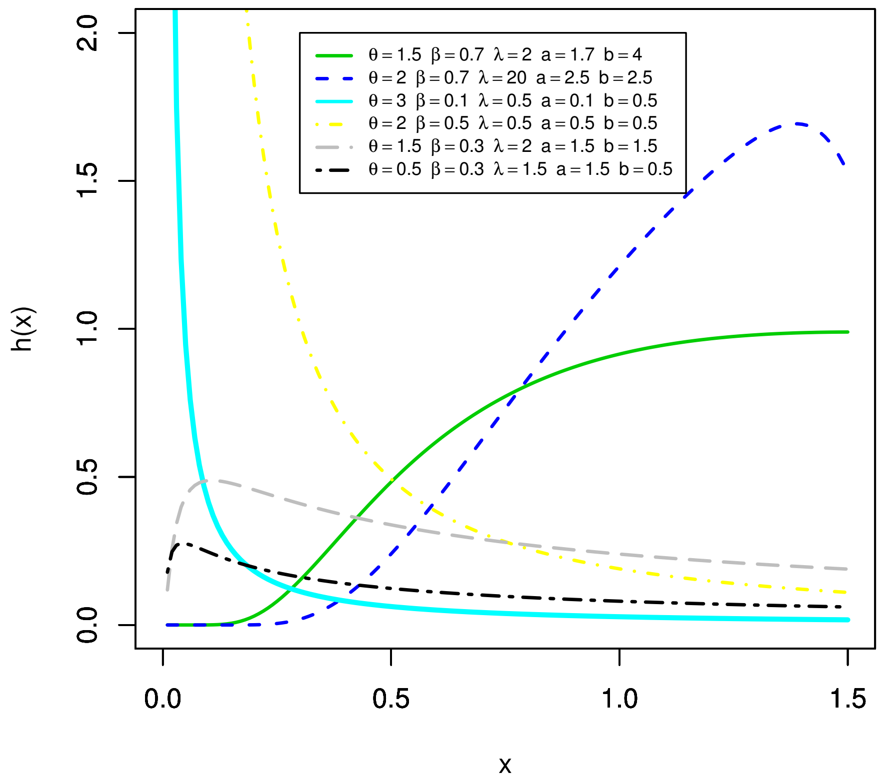


Figure 2: The hrf plots for the BExFr model.

Theorem 1. Suppose $X \sim \text{BExFr}(a, b, \lambda_1, \beta_1, \theta_1)$ and $Y \sim \text{BExFr}(a, b, \lambda_2, \beta_2, \theta_2)$. If $\beta_1 < \beta_2$, $\theta_1 > \theta_2$ and $\lambda_1 < \lambda_2$ Then $X \underset{st}{\geqslant} Y$, for integer values of β_1 and β_2.

Proof. At first, we consider the following:

$$I_x(a,b) = \frac{\int_0^x u^{a-1}(1-u)^{b-1}du}{B(a,b)}.$$

Next, note that the incomplete beta function $Beta_x(a, b)$ is an increasing function of x for fixed a and b. For any real number $x \in (0, \infty)$, $\beta_1 < \beta_2$, $\theta_1 > \theta_2$ and $\lambda_1 < \lambda_2$, we have

$$\left[1 - e^{-\left(\frac{\theta_1}{x}\right)^{\beta_1}}\right]^{\lambda_1} > \left[1 - e^{-\left(\frac{\theta_2}{x}\right)^{\beta_2}}\right]^{\lambda_2}$$

This implies that $I_{\left[1-e^{-\left(\frac{\theta_1}{x}\right)^{\beta_1}}\right]^{\lambda_1}}(a,b) \leq I_{\left[1-e^{-\left(\frac{\theta_2}{x}\right)^{\beta_2}}\right]^{\lambda_2}}(a,b)$. Equivalently, it implies that $P(X > x) \geq P(Y > x)$ and this completes the proof. □

Note: For fractional choices of λ_1, λ_2, β_1, β_2, the reverse of the theorem will be observed.

Corollary 1. From Theorem 1, we conjecture the following:

- For increasing θ, and β and λ decreasing, the hrf will exhibit DFR.

- For decreasing θ, and β and λ increasing, the hrf will exhibit IFR.

3.1. Quantile function

Let $Q_{a,b}\,(u)$ be the beta qf with parameters a and b. The qf of the BExFr distribution, say $x = Q(u)$, is given by

$$Q(u) = \theta\left\{-\ln\left[1-\left[1-Q_{a,b}\,(u)\right]^{1/\lambda}\right]\right\}^{-1/\beta},\ 0<u<1.$$

This scheme is useful to generate BFr random variates because of the existence of fast generators for beta random variables in most statistical packages, i.e. if V is a beta random variable with parameters a and b, then

$$X = \theta\left\{-\ln\left[1-(1-V)^{1/\lambda}\right]\right\}^{-1/\beta}$$

follows the BExFr distribution.

3.2. Moments

Henceforth, let Z be a random variable having the Fréchet distribution (1) with parameters θ and β. For $r<\beta$, the rth ordinary and incomplete moments of Z are, respectively, given by

$$\mu'_r = \theta^r\,\Gamma(1-r/\beta) \qquad \text{and} \qquad \varphi_r(t) = \theta^r\gamma(1-r/\beta,(\theta/t)^\beta)\,,$$

where $\gamma\,(s,t) = \int_0^t x^{s-1}\,e^{-x}dx$ is the lower incomplete gamma function.
Then, the rth moment of X, say μ'_r, can be expressed as

$$\mu'_r = \theta^r\sum_{k=0}^{\infty}\upsilon_k\ (k+1)^{r/\beta}\ \Gamma(1-r/\beta). \tag{9}$$

Setting $r=1$ in (9), we have the mean of X.
Using the relation between the central and non-central moments, we obtain the nth central moment of X, say μ_n, as follows

$$\mu_n = \theta^r\sum_{r=0}^{n}\sum_{k=0}^{\infty}\binom{n}{r}\upsilon_k\ \left(-\mu'_1\right)^{n-r}(k+1)^{r/\beta}\ \Gamma(1-r/\beta).$$

The skewness and kurtosis measures can be determined from the central moments using well-known relationships.

3.3. Moment generating function

First, we provide the generating function of the Fréchet model as discussed by Afify et al. (2016). Setting $y = x^{-1}$, we can write the mgf of Z as

$$M(t;\theta,\beta) = \beta\theta^\beta\int_0^\infty e^{\frac{t}{y}}\,y^{\beta-1}\,e^{-(\theta y)^\beta}dy.$$

After expanding $e^{\frac{t}{y}}$, we can write

$$\begin{aligned} M(t;\theta,\beta) &= \beta\theta^\beta\int_0^\infty\sum_{m=0}^{\infty}\frac{t^m}{m!}\,y^{\beta-m-1}\,e^{-(\theta y)^\beta}dy \\ &= \sum_{m=0}^{\infty}\frac{\theta^m\,t^m}{m!}\,\Gamma\left(\frac{\beta-m}{\beta}\right), \end{aligned}$$

where the gamma function is well-defined for any non-integer β.

Consider the Wright generalized hypergeometric function defined by

$$ {}_p\Psi_q\left[\begin{array}{c}(\gamma_1, A_1), \ldots, (\gamma_p, A_p)\\ (\beta_1, B_1), \ldots, (\beta_q, B_q)\end{array}; x\right] = \sum_{n=0}^{\infty} \frac{\prod_{j=1}^{p} \Gamma\left(\gamma_j + A_j\, n\right)}{\prod_{j=1}^{q} \Gamma\left(\beta_j + B_j\, n\right)} \frac{x^n}{n!}. $$

Then, we can write $M(t;\theta,\beta)$ as

$$ M(t;\theta,\beta) = {}_1\Psi_0\left[\begin{array}{c}(1, -\beta^{-1})\\ -\end{array}; \theta\, t\right]. $$

Combining expressions (8) and the last equation, we obtain the mgf of X, say $M(t)$, as

$$ M(t) = \sum_{k=0}^{\infty} \upsilon_k\, {}_1\Psi_0\left[\begin{array}{c}(1, -\beta^{-1})\\ -\end{array}; \theta(k+1)^{1/\beta}\, t\right]. $$

3.4. Incomplete moments

The nth incomplete moment, say $\vartheta_n(t)$ of the BExFr model is given by $\vartheta_n(t) = \int_0^t x^n\, f(x)dx$. From equation (8), we can write

$$ \vartheta_n(t) = \sum_{k=0}^{\infty} \upsilon_k \int_0^t x^n\, h_{k+1}(x). $$

Using the lower incomplete gamma function, we obtain (for $n < \beta$)

$$ \vartheta_n(t) = \sum_{k=0}^{\infty} \upsilon_k\, \theta^n\, (k+1)^{n/\beta} \gamma\left(1 - \frac{n}{\beta}, (k+1)\left(\frac{\theta}{t}\right)^{\beta}\right). \tag{10} $$

The important application of the first incomplete moment is related to the Bonferroni and Lorenz curves. These curves are very useful in economics, reliability, demography, insurance and medicine.

Further, the amount of scatter in a population is evidently measured to some extent by the totality of deviations from the mean and median. The mean deviations about the mean and about the median of X can be expressed as $\delta_\mu\left(X\right) = \int_0^\infty |X - \mu_1'|\, f(x)dx = 2\mu_1' F(\mu_1') - 2\vartheta_1(\mu_1')$ and $\delta_M(X) = \int_0^\infty |X - M|\, f(x)dx = \mu_1' - 2\vartheta_1(M)$, respectively, where $\mu_1' = E(X)$ comes from (11), $F(\mu_1')$ is simply calculated from (5), $\vartheta_1(\mu_1')$ is the first incomplete moment and M is the median of X.

3.5. Mean residual life and mean inactivity time

The MRL has many applications in biomedical sciences, life insurance, maintenance and product quality control, economics and social studies, demography and product technology (see Lai and Xie, 2006). Guess and Proschan (1988) gave an extensive coverage of possible applications of the mean residual life. The MRL (or the life expectancy at age t) represents the expected additional life length for a unit, which is alive at age t.

The MRL is given by

$$ m_X\left(t\right) = E\left(X - t \mid X > t\right), t > 0. $$

Then, the MRL of X can be obtained as

$$ m_X\left(t\right) = \left[1 - \vartheta_1\left(t\right)\right]/R(t) - t, \tag{11} $$

where $\vartheta_1(t)$ is the first incomplete moment of X and by setting $n = 1$ in equation (10), we obtain

$$\vartheta_1(t) = \sum_{k=0}^{\infty} \upsilon_k\, \theta\, (k+1)^{1/\beta} \gamma\left(1 - \frac{1}{\beta}, (k+1)\left(\frac{\theta}{t}\right)^{\beta}\right). \tag{12}$$

By substituting (12) in equation (11), we obtain

$$m_X(t) = \frac{\theta}{R(t)} \sum_{k=0}^{\infty} \upsilon_k\ (k+1)^{1/\beta} \gamma\left(1 - \frac{1}{\beta}, (k+1)\left(\frac{\theta}{t}\right)^{\beta}\right) - t.$$

The MIT represents the waiting time elapsed since the failure of an item on condition that this failure had occurred in $(0, t)$. The MIT of X is defined (for $t > 0$) by $M_X(t) = E(t - X \mid X \leq t)$.

The MIT of X is given by

$$M_X(t) = t - [\varphi_1(t) / F(t)]. \tag{13}$$

By inserting (12) in equation (13), we obtain the MIT of X as

$$M_X(t) = t - \frac{\theta}{F(t)} \sum_{k=0}^{\infty} \upsilon_k\ (k+1)^{1/\beta} \gamma\left(1 - \frac{1}{\beta}, (k+1)\left(\frac{\theta}{t}\right)^{\beta}\right).$$

3.6. Entropies

The Rényi entropy of a random variable X represents a measure of variation of the uncertainty. The Rényi entropy is defined by

$$I_q(x) = \frac{1}{1-q} \log \int_{-\infty}^{\infty} f^q(x)\, dx,\ \ q > 0 \text{ and } q \neq 1.$$

From equation (5), we can write

$$\begin{aligned} f^q(x) &= \left(\frac{\lambda\beta\theta^{\beta}}{B(a,b)}\right)^{q} x^{-q(\beta+1)} e^{-q\left(\frac{\theta}{x}\right)^{\beta}} \left[1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right]^{q(\lambda b - 1)} \\ &\quad \times \left\{1 - \left[1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right]^{\lambda}\right\}^{q(a-1)}. \end{aligned}$$

Applying the power series (6) to the last equation and after some simplifications, we can write

$$f^q(x) = \beta^q \theta^{q\beta} \sum_{k=0}^{\infty} \omega_k\ x^{-q(\beta+1)} e^{-(k+q)\left(\frac{\theta}{x}\right)^{\beta}},$$

where

$$\omega_k = \sum_{j=0}^{\infty} \frac{(-1)^{j+k}\, \lambda^q \Gamma(q(\lambda b - 1) + 1)\, \Gamma(\lambda(bq + j) - q + 1)}{j!\, k!\, [B(a,b)]^q\, \Gamma(q(\lambda b - 1) - j + 1)\, \Gamma(\lambda(bq + j) - q - k + 1)}.$$

Then, the Rényi entropy of X is given by

$$I_q(x) = \frac{1}{1-q} \log \left[\beta^q \theta^{q\beta} \sum_{k=0}^{\infty} \omega_k \underbrace{\int_0^{\infty} x^{-q(\beta+1)} e^{-(k+q)\left(\frac{\theta}{x}\right)^{\beta}} dx}_{I}\right].$$

Then,

$$I = \frac{\theta^{1-q(\beta+1)}}{\beta}\ (k+q)^{-s/\beta}\, \Gamma\left(\frac{s}{\beta}\right),$$

where $s = q(\beta + 1) - 1$.

Now, we can write the Rényi entropy of X as

$$I_q(x) = \frac{1}{1-q} \log \left\{ \sum_{k=0}^{\infty} \omega_k \left(\frac{\beta}{\theta}\right)^{q-1} (k+q)^{-s/\beta} \Gamma\left(\frac{s}{\beta}\right) \right\}. \tag{14}$$

The q-entropy, say $H_q(x)$, is defined by

$$H_q(x) = \frac{1}{q-1} \log\left[1 - I_q(x)\right],$$

where

$$I_q(x) = \int_{\Re} f^q(x)\, dx,\ q > 0 \quad \text{and} \quad q \neq 1.$$

From equation (14), we obtain

$$H_q(f) = \frac{1}{q-1} \log \left\{ 1 - \sum_{k=0}^{\infty} \omega_k \left(\frac{\beta}{\theta}\right)^{q-1} (k+q)^{-s/\beta} \Gamma\left(\frac{s}{\beta}\right) \right\}.$$

The Shannon entropy, say E_{sh}, of a random variable X is defined by

$$E_{sh} = E\left\{-\left[\log f(x)\right]\right\}.$$

It is a special case of the Rényi entropy when $q \uparrow 1$. So, based on equation (8), we can write

$$E_{sh} = -\left\{ \log \left[\sum_{k=0}^{\infty} \upsilon_k\, E\left(Y_{k+1}\right) \right] \right\},$$

where $Y_{k+1} \sim \mathrm{Fr}(\theta(k+1)^{1/\beta}, \beta)$. But $E\left(Y_{k+1}\right) = \theta\ (k+1)^{1/\beta}\, \Gamma(1 - 1/\beta)$, so the E_{sh} of X is given by

$$E_{sh} = -\left\{ \log \left[\sum_{k=0}^{\infty} \upsilon_k\, \theta\ (k+1)^{1/\beta}\, \Gamma(1 - \frac{1}{\beta}) \right] \right\}.$$

4. Order statistics

In this section we consider the expression for the general r-th order statistic and the large sample distribution of the sample minimum and the sample maximum when a random sample of size n are drawn from the BExFr$(a, b, \lambda, \beta, \theta)$ distribution. The density function of the rth order statistic, $X_{r:n}$, for a random sample of size n drawn from (5), is given by

$$f_{X_{r:n}}(x) = \frac{1}{B(r, n-r+1)} (F(x))^{r-1} (1 - F(x))^{n-r} f(x).$$

Then the rth order statistic of X is given by

$$\begin{aligned} f_{X_{r:n}}(x) &= \frac{1}{B(r, n-r+1)} f(x) \sum_{j=0}^{n-r} (-1)^j \binom{n-r}{j} I(0 < x < 1) \\ &\quad \times \left(\frac{B\left(1 - \left(1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)^{\lambda}; a, b\right)}{B(a, b)} \right)^{r-1+j}. \end{aligned}$$

Using the series expression for the incomplete beta function:

$$I_x(a,b,) = \frac{B(x,a,b)}{B(a,b)} = \sum_{k=a}^{a+b-1} x^k (1-x)^{a+b-1-k},$$

the pdf of $X_{r:n}$ can be written as

$$\begin{aligned}
f_{r:n}(x) &= \frac{f(x)}{B(r,n-r+1)} \sum_{j=0}^{n-r} \sum_{k=a}^{a+b-1} (-1)^j \binom{n-r}{j} \\
&\times \left\{ \sum_{k=a}^{a+b-1} \left[1 - \left(1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)^{\lambda} \right]^k \left(1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)^{\lambda(a+b-1-k)} \right\}^{r-1+j} \\
&= \frac{f(x)}{B(r,n-r+1)} \sum_{j=0}^{n-r} \sum_{k_1=a}^{a+b-1} \cdots \sum_{k_{r-1+j}=a}^{a+b-1} (-1)^{j+s_k} \binom{n-r}{j} \\
&\times \frac{Beta(s_k+a,(r-1+j)(a+2b-1)-s_k)}{Beta(a,b)} p_k \\
&\times f\left(x|s_k+a,(r-1+j)(a+2b-1)-s_k),\lambda,\beta,\theta\right), \qquad (15)
\end{aligned}$$

where $s_k = \sum_{i=1}^{r-1+j} k_i$ and $p_k = \prod_{i=1}^{r-1+j} \binom{a+b-1}{k_i}$.

Also, $f\left(x|s_k+a,(r-1+j)(a+2b-1)-s_k),\lambda,\beta,\theta\right)$ is the density of a BExFr distribution with parameters $s_k+a,(r-1+j)(a+2b-1)-s_k)$, λ, β, θ respectively.

From (15), it is interesting to note that the pdf of the rth order statistic $X_{r:n}$ can be expressed as an finite sums of the BExFr pdf 's. However, if a and b are not integers, then the sums will terminate at ∞. Note that, using moments expression, one can easily get an expression for the general mth order moment for the order statistics.

5. Characterizations

This section deals with various characterizations of BExFr distribution. These characterizations are based on: (i) a simple relationship between two truncated moments; (ii) the hazard function. It should be mentioned that for characterization (i) the cdf need no have a closed form. We believe, due to the nature of the cdf of BExFr, there may not be other possible characterizations than the ones presented in this section.

5.1. Characterizations based on two truncated moments

In this subsection we present characterizations of BExFr distribution in terms of a simple relationship between two truncated moments. Our first characterization employs a theorem due to Glänzel (1987), see Theorem 1 below. Note that the result holds also when the interval H is not closed. It should also be mentioned that this characterization is stable in the sense of weak convergence.

Theorem 2. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a given probability space and let $H=[d,e]$ be an interval for some $d<e$ $(d=-\infty,\ e=\infty$ might as well be allowed). Let $X:\Omega \to H$ be a continuous random variable with the distribution function F and let g and h be two real functions defined on H such that

$$\mathbf{E}\left[g(X) \mid X \geq x\right] = \mathbf{E}\left[h(X) \mid X \geq x\right] \eta(x), \quad x \in H,$$

is defined with some real function η. Assume that $g,\ h \in C^1(H)$, $\eta \in C^2(H)$ and F is twice continuously differentiable and strictly monotone function on the set H. Finally, assume that the equation $h\eta = g$ has no real solution in the interior of H. Then F is uniquely determined by the functions g, h and η, particularly

$$F(x) = \int_a^x C \left| \frac{\eta'(u)}{\eta(u)h(u) - g(u)} \right| \exp(-s(u))\ du\ ,$$

where the function s is a solution of the differential equation $s' = \frac{\eta'\, h}{\eta\, h\, -\, g}$ and C is the normalization constant, such that $\int_H dF = 1$.

Proposition 1. Let $X : \Omega \to (0, \infty)$ be a continuous random variable and let

$$h(x) = \left[1 - \left(1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)^{\lambda}\right]^{1-a} \text{ and } g(x) = h(x)\left(1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)\ for x > 0.$$

The random variable X belongs to BExFr family (5) if and only if the function η defined in Theorem 2 has the form

$$\eta(x) = \frac{\lambda b}{\lambda b + 1}\left(1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right), \quad x > 0.$$

Proof. Let X be a random variable with density (5), then

$$(1 - F(x))\, E\left[h(x)\ \mid X \geq x\right] = \frac{1}{bB(a,b)}\left(1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)^{\lambda b},\ x > 0$$

and

$$(1 - F(x))\, E\left[g(x)\ \mid X \geq x\right] = \frac{\lambda}{(\lambda b + 1)\, B(a,b)}\left(1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)^{\lambda b + 1},\ x > 0,$$

and finally

$$\eta(x)\, h(x) - g(x) = h(x)\frac{1}{bB(a,b)}\left(1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)\left\{-\frac{1}{\lambda b + 1}\right\} < 0 \text{ for } x > 0.$$

Conversely, if η is given as above, then

$$s'(x) = \frac{\eta'(x)\, h(x)}{\eta(x)\, h(x) - g(x)} = \frac{\lambda b \beta x^{-(\beta+1)} e^{-\left(\frac{\theta}{x}\right)^{\beta}}}{1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}}, \quad x > 0,$$

and hence

$$s(x) = -\ln\left\{\left(1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)^{\lambda b}\right\}, \quad x > 0.$$

Now, in view of Theorem 2, X has density (5).

Corollary 2. Let $X : \Omega \to (0, \infty)$ be a continuous random variable and let $h(x)$ be as in Proposition 1. The pdf of X is (5) if and only if there exist functions g and η defined in Theorem 2 satisfying the differential equation

$$\frac{\eta'(x)\, h(x)}{\eta(x)\, h(x) - g(x)} = \frac{\lambda b \beta x^{-(\beta+1)} e^{-\left(\frac{\theta}{x}\right)^{\beta}}}{1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}}, \quad x > 0. \tag{16}$$

The general solution of the differential equation in Corollary 2 is

$$\eta(x) = \left[1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right]^{-\lambda b} \left[- \int \frac{\lambda b \beta e^{-\left(\frac{\theta}{x}\right)^{\beta}} \left[1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right]^{\lambda b - 1}}{h(x)\, x^{\beta+1}} g(x)\, dx + D \right],$$

where D is a constant. Note that a set of functions satisfying the differential equation (16) is given in Proposition 1 with $D = 0$. Clearly, there are other triplets (h, g, η) satisfying the conditions of Theorem 2.

5.2. Characterization based on hazard function

It is known that the hazard function, h_F, of a twice differentiable distribution function, F, satisfies the first order differential equation

$$\frac{f'(x)}{f(x)} = \frac{h_F'(x)}{h_F(x)} - h_F(x). \tag{17}$$

For many univariate continuous distributions, this is the only characterization available in terms of the hazard function. The following characterization establish a non-trivial characterization for BExFr distribution in terms of the hazard function when $a = 1$, which is not of the trivial form given in (17).

Proposition 2. Let $X : \Omega \to (0, \infty)$ be a continuous random variable. Then for $a = 1$, the pdf of X is (5) if and only if its hazard function $h_F(x)$ satisfies the differential equation

$$\begin{aligned} h_F'(x) + (\beta+1)\, x^{-1} h_F(x) &= \lambda b \beta^2 \theta^{2\beta} x^{-2(\beta+1)} e^{-\left(\frac{\theta}{x}\right)^{\beta}} \left(1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)^{-2} \\ &= \lambda b \beta \theta^{\beta} x^{-(\beta+1)} \frac{d}{dx} \left\{ \left(1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)^{-1} \right\}. \end{aligned} \tag{2}$$

Proof. If X has pdf (5), then clearly (18) holds. Now, if (18) holds, then

$$\frac{d}{dx}\left\{x^{\beta+1} h_F(x)\right\} = \lambda b \beta \theta^{\beta} \frac{d}{dx} \left\{ \left(1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)^{-1} \right\},$$

or, equivalently,

$$h_F(x) = \frac{\lambda b \beta \theta^{\beta} x^{-(\beta+1)} e^{-\left(\frac{\theta}{x}\right)^{\beta}}}{1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}}.$$

Integrating the above equation from 0 to x, we obtain

$$1 - F(x) = \left(1 - e^{-\left(\frac{\theta}{x}\right)^{\beta}}\right)^{\lambda b}, \; x \geq 0.$$

6. Maximum likelihood estimation

In this section, we consider the estimation of the parameters of the BExFr model by the maximum likelihood. Consider the random sample $X_1, \ldots, X_n$ of size n from this distribution.

The log-likelihood function for the parameter vector $\varphi = (a, b, \lambda, \beta, \theta)^\intercal$, say $\ell(\varphi)$, is given by

$$\begin{aligned}\ell(\varphi) &= n\left[\log\lambda + \log\beta + \beta\log\theta - \log B(a,b)\right] - (\beta+1)\sum_{i=1}^{n}\log\left(x_i\right) \\ &\quad -\sum_{i=1}^{n}\left(\frac{\theta}{x_i}\right)^{\beta} + (\lambda b - 1)\sum_{i=1}^{n}\log\left(s_i\right) + (a-1)\sum_{i=1}^{n}\log\left(1 - s_i^{\lambda}\right),\end{aligned}$$

where $s_i = 1 - e^{-\left(\frac{\theta}{x_i}\right)^{\beta}}$.

This equation can be maximized either directly by using the `R` (`optim` function), `MATH-CAD` program, `SAS` (`PROC NLMIXED`) or by solving the nonlinear equations obtained by differentiating the log-likelihood. Therefore, the score vector is $\mathbf{U}(\varphi) = \frac{\partial\ell}{\partial\varphi} = \left(\frac{\partial\ell}{\partial a}, \frac{\partial\ell}{\partial b}, \frac{\partial\ell}{\partial\lambda}, \frac{\partial\ell}{\partial\beta}, \frac{\partial\ell}{\partial\theta}\right)^\intercal$.

$$\frac{\partial\ell}{\partial a} = n\left[\psi\left(a+b\right) - \psi\left(a\right)\right] + \sum_{i=1}^{n}\log\left(1 - s_i^{\lambda}\right),$$

$$\frac{\partial\ell}{\partial b} = n\left[\psi\left(a+b\right) - \psi\left(b\right)\right] + \lambda\sum_{i=1}^{n}\log\left(s_i\right),$$

$$\frac{\partial\ell}{\partial\lambda} = \frac{n}{\lambda} + b\sum_{i=1}^{n}\log\left(s_i\right) - (a-1)\sum_{i=1}^{n}\frac{s_i^{\lambda}\log\left(s_i\right)}{1 - s_i^{\lambda}}$$

$$\begin{aligned}\frac{\partial\ell}{\partial\beta} &= \frac{n}{\beta} + n\log\theta - \sum_{i=1}^{n}\log\left(x_i\right) - \sum_{i=1}^{n}\left(\frac{\theta}{x_i}\right)^{\beta}\log\left(\frac{\theta}{x_i}\right) \\ &\quad + (\lambda b - 1)\sum_{i=1}^{n}\frac{z_i}{s_i} - \lambda\left(a-1\right)\sum_{i=1}^{n}\frac{z_i s_i^{\lambda-1}}{1 - s_i^{\lambda}}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial\ell}{\partial\theta} &= \frac{n\beta}{\theta} - \frac{\beta}{\theta}\sum_{i=1}^{n}\left(\frac{\theta}{x_i}\right)^{\beta} - \frac{\beta\lambda\left(a-1\right)}{\theta}\sum_{i=1}^{n}\frac{\left(\frac{\theta}{x_i}\right)^{\beta}s_i^{\lambda-1}}{1 - s_i^{\lambda}} \\ &\quad + \frac{\beta\left(\lambda b - 1\right)}{\theta}\sum_{i=1}^{n}\frac{1}{s_i}\left(\frac{\theta}{x_i}\right)^{\beta}e^{-\left(\frac{\theta}{x_i}\right)^{\beta}},\end{aligned}$$

where $z_i = \left(\frac{\theta}{x_i}\right)^{\beta}e^{-\left(\frac{\theta}{x_i}\right)^{\beta}}\log\left(\frac{\theta}{x_i}\right)$ and $\psi\left(.\right)$ is the digamma function which is the derivative of $\log\Gamma\left(.\right)$, where $\Gamma\left(.\right)$ is the gamma function.

We can obtain the estimates of the unknown parameters by setting the score vector to zero, $\mathbf{U}(\widehat{\varphi}) = 0$. Solving these equations simultaneously yields the MLEs $\widehat{\varphi} = (\widehat{a}, \widehat{b}, \widehat{\lambda}, \widehat{\beta}, \widehat{\theta})^\intercal$ of $\varphi = (a, b, \lambda, \beta, \theta)^\intercal$. These equations cannot be solved analytically and statistical software can be used to solve them numerically by means of iterative techniques such as the Newton-Raphson algorithm. For the new distribution all the second-order derivatives exist.

For interval estimation of the model parameters, we require the 5×5 observed information matrix $J\left(\varphi\right) = \left\{J_{rs}\right\}$ (for $r, s = a, b, \lambda, \beta, \theta$). Under standard regularity conditions, the multivariate normal $N_5(0, J(\widehat{\varphi})^{-1})$ distribution can be used to construct approximate confidence intervals for the model parameters. Here, $J(\widehat{\varphi})$ is the total observed information matrix evaluated at $\widehat{\varphi}$. Based on this multivariate normal approximation, the approximate $100(1-\phi)\%$ confidence intervals for a, b, λ, β and θ can be determined by the usual way.

7. Simulation

In this section, we consider the maximum likelihood estimation of parameters for the two models derived in the preceding sections. The maximum likelihood estimators can be obtained by direct maximization of the likelihood functions given earlier. Here, we maximized the log-likelihood function using `SAS PROC NLMIXED`. For each maximization, the `SAS PROC NLMIXED` function was executed for a wide range of initial values, and the maximum likelihood estimates were determined as the ones that corresponds to the largest of the maxima.

To illustrate the feasibility of the suggested estimation strategy, a small simulation study was undertaken. The simulation study was carried out for one representative set of parameters $(\lambda, \beta, \theta, a, b) = (1.6, 2.3, 1.2, 1.8, 0.9)$ and the process was repeated 30000 times. Three different sample sizes $n = 50$, 100 and 200 were considered. The bias (actual-estimate) and the standard deviation of the parameter estimates for the maximum likelihood estimates were determined from this simulation study and are presented in Table 2.

Table 2. Bias and standard deviation of the parameter estimates.

Parameter	Sample size (n=50)	Sample size (n=100)	Sample size (n=200)
λ	0.1108(0.5382)	0.0614(0.2345)	0.0437(0.1139)
β	0.1678(0.4628)	-0.1321(0.1894)	0.0672(0.0933)
θ	0.0268(0.4321)	0.1483(0.2467)	0.0946(0.1264)
a	0.0825(0.0667)	0.0779(0.0627)	0.0621(0.0358)
b	0.127(0.2368)	0.0651(0.0789)	0.0317(0.0223)

The figures in Table 2 indicate that the estimates are quite stable and, more important, are close to the true values for the these sample sizes. Furthermore, as the sample size increases, the SEs decreases as expected.

8. Applications

In this section, we provide two applications to two real data sets to prove the importance and flexibility of the BExFr distribution. The first real data set represents the survival times, in weeks, of 33 patients suffering from acute Myelogeneous Leukaemia. These data have been analyzed by Feigl and Zelen (1965). The data are: 65, 156, 100, 134, 16, 108, 121, 4, 39, 143, 56, 26, 22, 1, 1, 5, 65, 56, 65, 17, 7, 16, 22, 3, 4, 2, 3, 8, 4, 3, 30, 4, 43. The second data set, strength data, which were originally reported by Badar and Priest (1982) and it represents the strength measured in GPA for single carbon fibers and impregnated 1000-carbon fiber tows. Single fibers were tested under tension at gauge lengths of 10 mm with sample size (n = 63). This data set consists of observations: 1.901, 2.132, 2.203, 2.228, 2.257, 2.350, 2.361, 2.396, 2.397, 2.445, 2.454, 2.474, 2.518, 2.522, 2.525, 2.532, 2.575, 2.614, 2.616, 2.618, 2.624, 2.659, 2.675, 2.738, 2.740, 2.856, 2.917, 2.928, 2.937, 2.937, 2.977, 2.996, 3.030, 3.125, 3.139, 3.145, 3.220, 3.223, 3.235, 3.243, 3.264, 3.272, 3.294, 3.332, 3.346, 3.377, 3.408, 3.435, 3.493, 3.501, 3.537, 3.554, 3.562, 3.628, 3.852, 3.871, 3.886, 3.971, 4.024, 4.027, 4.225, 4.395, 5.020. These data have been used by Afify et al. (2015b) and Afify et al. (2015c) to fit the exponentiated transmuted generalized Rayleigh and transmuted Weibull Lomax distributions, respectively.

We shall compare the fit of the proposed BExFr distribution (and its sub-models namely: BFr, EFr and Fr distributions) with several other competitive models namely: the generalized inverse gamma (Mead, 2015), McDonald Lomax (McL) (Lemonte and Cordeiro, 2013), gamma Lomax (GL) (Cordeiro et al., 2015) and Zografos-Balakrishnan log-logistic (ZBLL) (Zografos and Balakrishnan, 2009) models with corresponding densities (for $x > 0$):

GIG: $f(x;\theta,\beta,a,b,\lambda) = \frac{b\theta^{ab}}{\Gamma_\lambda(a,\beta)} x^{-(ab+1)} \left[\left(\frac{\theta}{x}\right)^b + \beta\right]^{-\lambda} \exp\left[-\left(\frac{\theta}{x}\right)^b\right]$;

McL: $f(x;\theta,\beta,a,b,\lambda) = \frac{\theta\lambda}{\beta B\left(\frac{a}{\lambda},b\right)} \left(1+\frac{x}{\beta}\right)^{-(\theta+1)} \left[1-\left(1+\frac{x}{\beta}\right)^{-\theta}\right]^{a-1}$

$\times \left\{1-\left[1-\left(1+\frac{x}{\beta}\right)^{-\theta}\right]^{\lambda}\right\}^{b-1}$;

GL: $f(x;\theta,\beta,a)=\frac{\theta\beta^{\theta}}{\Gamma(a)\Gamma(\beta+x)^{\theta+1}}\left[-\theta\ln\left[\left(\frac{\beta}{\beta+x}\right)\right]\right]^{a-1}$;

ZBLL: $f(x;\theta,\beta,a)=\frac{\beta\theta^{-\beta}}{\Gamma(a)}x^{\beta-1}\left[1+\left(\frac{x}{\theta}\right)^{\beta}\right]^{-2}\left\{\ln\left[1+\left(\frac{x}{\theta}\right)^{\beta}\right]\right\}^{a-1}$;

where the parameters of the above densities are all positive real numbers, $\Gamma(a)$ is the gamma function and $\Gamma_{\lambda}(a,\beta)$ is the generalized gamma function (Kobayashi, 1991) defined by

$$\Gamma_{\lambda}(a,\beta)=\int_{0}^{\infty}y^{a-1}(\beta+y)^{-\lambda}\exp(-y)\,dy.$$

In order to compare the models, we consider some goodness-of-fit measures including $-2\widehat{\ell}$, where $\widehat{\ell}$ is the maximized loglikelihood, Anderson-Darling (A^{*}) and Cramér-von Mises (W^{*}) statistics (full details can be found in Chen and Balakrishnan, 1995). In general, the model with minimum values for these statistics could be chosen as the best model to fit the data.

Tables 4 and 5 list the MLEs of the model parameters, their corresponding standard errors (given in parentheses) and the values of these statistics ($-2\widehat{\ell}$, A^{*} and W^{*}) for the fitted models to both data sets.

Tables 3 and 4 compare the BExFr model with the BFr, EFr, Fr, GIG, McL, ZBLL and GL distributions. It is noted, from Tables 4 and 5, that the BExFr distribution gives the lowest values for the $-2\widehat{\ell}$, A^{*} and W^{*} statistics among all fitted models. Thus, the BExFr distribution could be chosen as the best models. These results are obtained using the `MAT-HCAD PROGRAM`.

Table 3. MLEs, their corresponding standard errors and the statistics $-2\widehat{\ell}$, W^{*} and A^{*} for the first data set.

Model	Estimates					$-2\widehat{\ell}$	W^{*}	A^{*}
	$\widehat{\theta}$	$\widehat{\beta}$	$\widehat{a}$	$\widehat{b}$	$\widehat{\lambda}$			
BExFr	29.5877	0.1107	21.0415	19.7308	1.7253	**308.117**	**0.11211**	**0.70495**
	(201.769)	(0.146)	(64.505)	(62.632)	(4.578)			
BFr	6.89494	0.13852	16.8911	24.04154		308.213	0.11407	0.71477
	(60.760)	(0.138)	(43.704)	(47.769)				
EFr	13.3423	0.5732			1.4581	310.894	0.15728	0.95136
	(14.377)	(0.206)			(1.007)			
Fr	7.8651	0.6944				311.997	0.17306	1.0475
	(2.091)	(0.091)						
GIG	3.1481	0.89959	8.42866	0.29097	18.95355	308.233	0.11333	0.7127
	(45.860)	(4.01682)	(10.3956)	(0.033)	(17.1406)			
McL	0.57665	8.46832	1.62102	1.72961	0.46481	310.958	0.13714	0.84928
	(1.067)	(28.781)	(2.762)	(2.593)	(3.251)			
ZBLL	7.45579	1.01264	1.4687			318.823	0.156487	0.95522
	(0.248)	(0.138)	(0.179)					
GL	1.1115	3.52517	2.1469			311.037	0.1419	0.86637
	(0.271)	(6.722)	(1.6968)					

Figures 3 and 4 display the estimated pdf's and cdf's of the BExFr distribution for the first real data set, where figures 5 and 6 represent the same for the second data set respectively. It is shown from these figures that the BExFr provides a close fit to these data sets.

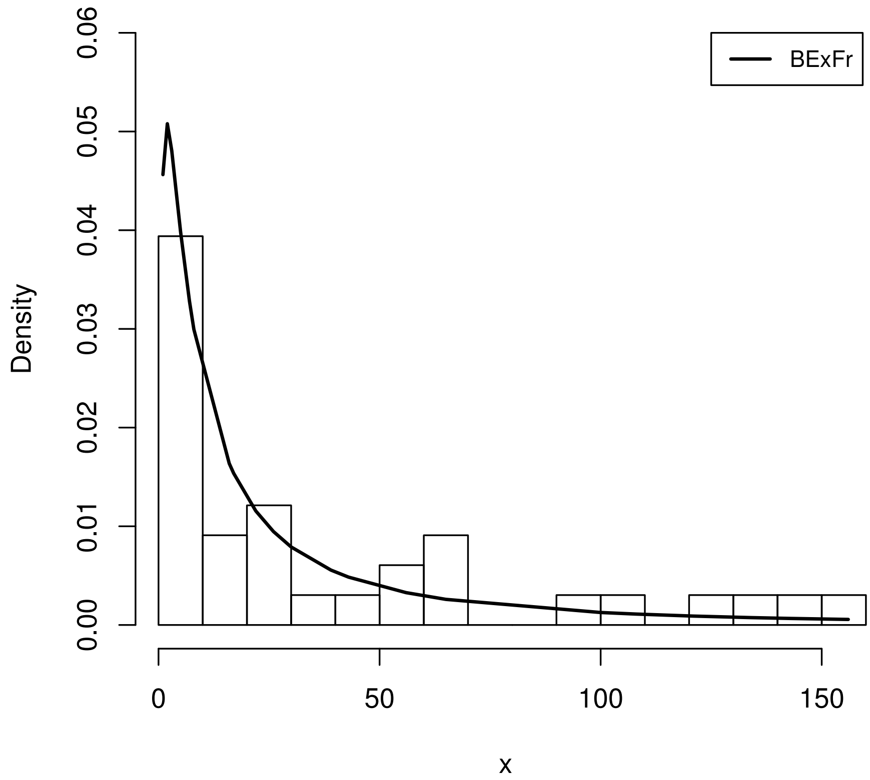


Figure 3: The fitted BExFr density for the first data set

Table 4. MLEs, their corresponding standard errors and the statistics $-2\widehat{\ell}$, W^* and A^* the second data set

Model	Estimates					$-2\widehat{\ell}$	W^*	A^*
	$\widehat{\theta}$	$\widehat{\beta}$	$\widehat{a}$	$\widehat{b}$	$\widehat{\lambda}$			
BExFr	12.6463	0.4852	25.7682	14.2299	7.2493	**112.704**	**0.05574**	**0.31204**
	(75.689)	(0.978)	(71.831)	(67.197)	(28.051)			
BFr	25.03468	0.55403	7.91533	185.80664		113.065	0.0558	0.32555
	(21.8041)	(0.255)	(6.395)	(155.015)				
EFr	4.2957	2.3636			7.0322	112.701	0.05895	0.31988
	(1.613)	(1.028)			(8.513)			
Fr	2.7214	5.4338				117.804	0.12884	0.69597
	(0.067)	(0.508)						
GIG	1.7724	0.26685	3.12731	3.72201	8.12837	113.065	0.06669	0.3536
	(0.5801)	(1.18644)	(3.3986)	(2.082)	(6.41594)			
McL	1.89877	3.68277	37.12441	26.14064	2.85382	113.018	0.05712	0.32925
	(14.485)	(10.640)	(37.783)	(236.492)	(5.007)			
ZBLL	2.7035	8.16714	1.45705			140.08	0.10167	0.53337
	(0.0033)	(0.823)	(0.13)					
GL	67.73728	2.76071	50.15703			112.922	0.05965	0.33634
	(29.078)	(4.312)	(36.5936)					

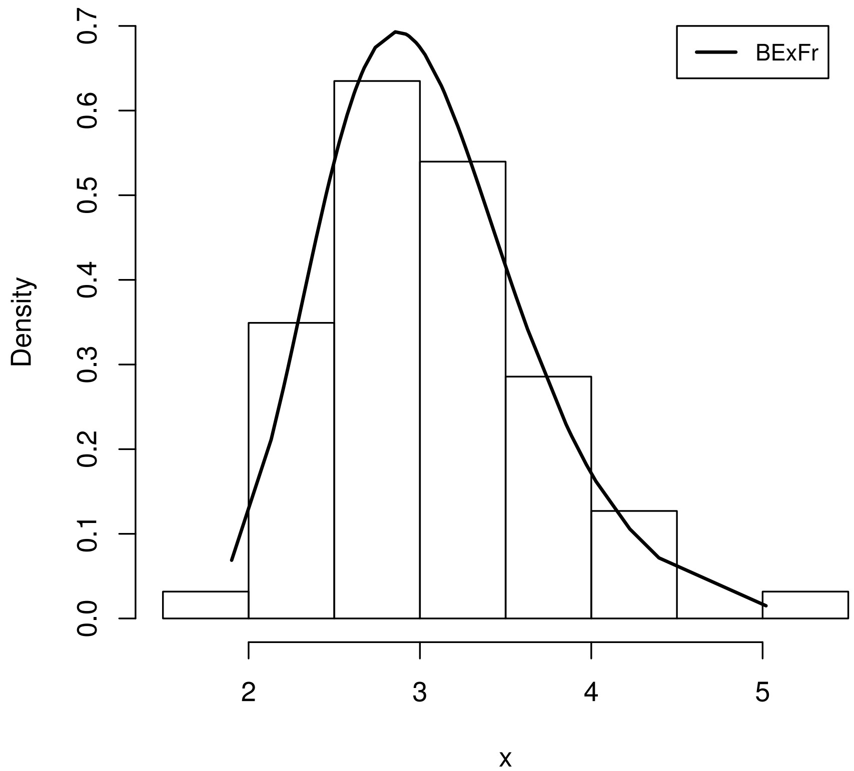


Figure 4: The fitted BExFr density for the second data set.

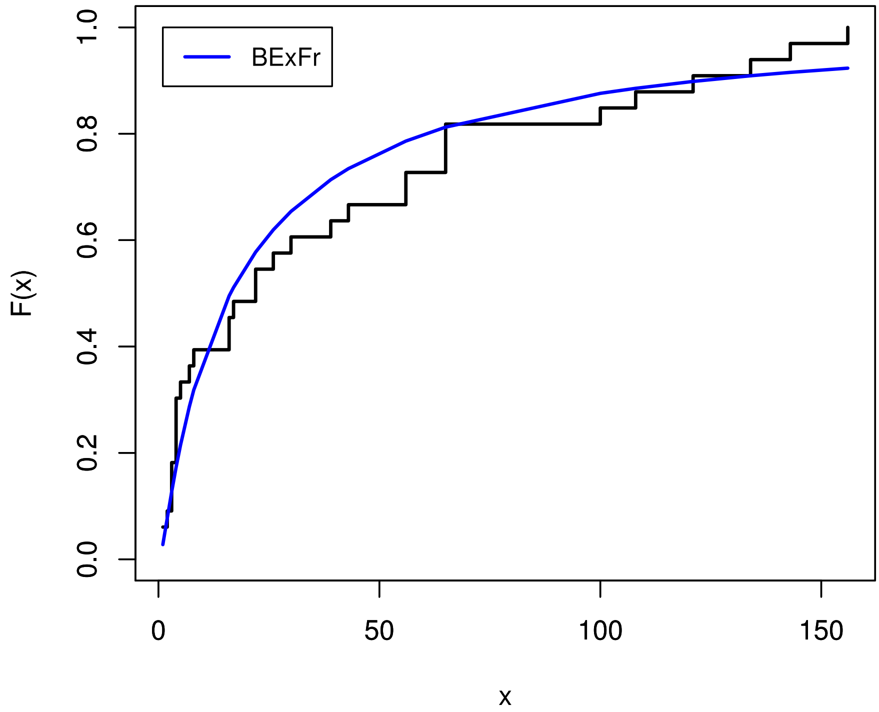


Figure 5: The estimated cdf of the BExFr model for the first data set.

9. Concluding remarks

In this paper, we propose a new five-parameter model, called the *beta exponential Fréchet* (BExFr) distribution, which extends the Fréchet distribution. In fact, the BExFr distribution is motivated by the wide use of the Fréchet distribution in extreme value theory and also for the fact that the generalization provides more flexibility to analyze real data. The BExFr density function can be expressed as a mixture of Fréchet densities. We derive explicit expressions for the ordinary and incomplete moments, moment generating function, entropies, mean residual life and mean inactivity time. We discuss the maximum likelihood estimation of the model parameters. Two applications illustrate that the proposed model provides consistently better fit than the other competitive models.

Estimation of the model parameters under the bayesian paradigm is currently underway and will be reported in a separate article elsewhere. However, we must make a note of the fact under the Bayesian setting, a non informative prior approach is essentially maximum likelihood estimation under the classical approach. In the absence of an appropriate conjugate prior, the choice of prior will be a challenging in such a setting.

As a future work we will consider the following:

- Bivariate and multivariate extension of the BExFr distribution. In particular with the copula based construction method, trivariate reduction etc.

- Comparison of the derived models with the available popular bivariate beta- G and bivariate exponential type models.

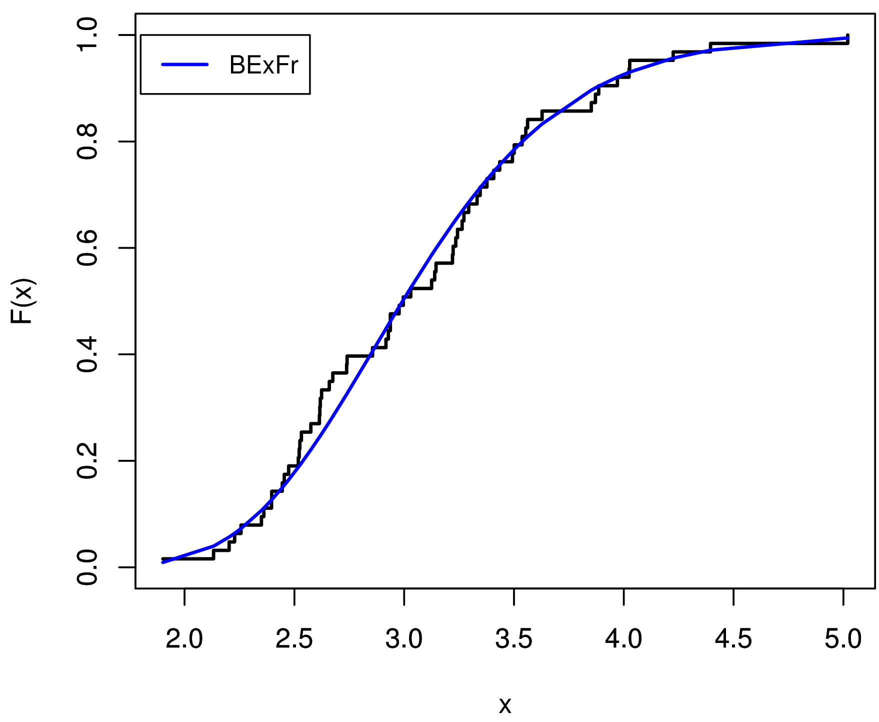


Figure 6: The estimated cdf of the BExFr model for the second data set.

Acknowledgment

The authors would like to thank the editor and anonymous referees for their valuable comments and suggestions that improved the quality of the paper.

References

Agresti, A. (1984). *Analysis of Ordinal Categorical Data.* New York: John Wiley.

Afify, A.Z., Hamedani, G.G., Ghosh, I. and Mead M.E. (2015a). The Transmuted Marshall-Olkin Fréchet Distribution: Properties and Applications. *International Journal of Statistics and Probability*, **4**, 132–148.

Afify, A.Z., Nofal, Z.M. and Ebraheim, A.N. (2015a). Exponentiated Transmuted Generalized Rayleigh Distribution: A New Four Parameter Rayleigh Distribution. *Pakistan Journal of Statistics and Operation Research*, **11**, 115–134.

Afify, A.Z., Nofal, Z.M., Yousof, H.M., El Gebaly, Y.M. and Butt, N.S. (2015b). The Transmuted Weibull Lomax Distribution: Properties and Application. *Pakistan Journal of Statistics and Operation Research*, **11**, 135–152.

Afify, A.Z., Yousof, H.M., Nofal, Z.M. and Cordeiro, G.M. (2015). The Weibull Fréchet Distribution and Its Applications. *Journal of Applied Statistics (Forthcoming.)*

Alzaatreh, A., Famoye, F. and Lee, C. (2013). A New Method for Generating Families of Continuous Distributions. *Metron*, **71**, 63–79.

Arnold, B.C. and Laguna, L. (1976). A Stochastic Mechanism leading to Asymptotically Paretian Distributions. *Proceedings of the American Statistical Association. Business and Economic Statistics section.*

Badar, M.G. and Priest, A.M. (1982). Statistical Aspects of Fiber and Bundle Strength in Hybrid Composites. In: Hayashi, T., Kawata, K., Umekawa, S. (Eds.). *Progress in Science and Engineering Composites.* ICCM-IV, Tokyo, 1129–1136.

Baharith, L.A., Mousa, S.A., Atallah, M.A. and Elgayar, S.H. (2014). The Beta Generalized Inverse Weibull Distribution. *British Journal of Mathematics & Computer Science*, **4**, 252–270.

Barreto-Souza, W.M., Cordeiro, G.M. and Simas, A.B. (2011). Some Results for Beta Fréchet Distribution. *Communications in Statistics-Theory and Methods*, **40**, 798–811.

Chen, G. and Balakrishnan, N. (1995). A General Purpose Approximate Goodness-of-fit Test. *Journal of Quality Technology*, **27**, 154–161.

Cordeiro, G.M., Ortega, E.M.M. and Popovic, B.V. (2015). The Gamma-Lomax Distribution. *Journal of Statistical computation and Simulation*, **85**, 305–319.

De Gusmao F.R.S., Ortega E.M.M., and Cordeiro G.M. (2011). The Generalized Inverse Weibull Distribution. *Statistical Papers*, **52**, 591–619.

Elbatal, I., Asha, G. and Raja, V. (2014). Transmuted Exponentiated Fréchet Distribution: Properties and Applications. *Journal of Statistics Applications & Probability*, **3**, 379–394.

Feigl, P. and Zelen, M. (1965). Estimation of Exponential Probabilities with Concomitant Information. *Biometrics*, **21**, 826–838.

Fréchet, M. (1924). Sur la Loi des Erreurs d'Observation. *Bulletin de la Société Mathématique de Moscou*, **33**, 5–8.

Glänzel, W. (1987). A Characterization Theorem Based on Truncated Moments and Its Application to Some Distribution Families. *Mathematical Statistics and Probability Theory (Bad Tatzmannsdorf, 1986)*, **B**, Reidel, Dordrecht, 75–84.

Guess, F. and Proschan, F. (1988). Mean Residual Life, Theory and Applications. In: Krishnaiah, P.R., Rao, C.R. *Handbook of Statistics, Reliability and Quality Control*, **7**, 215–224.

Hamada, M. and Wu C.F.J. (1990). A Critical look at Accululation Analysis and Related Methods (with discussion). *Technometrics*, **28**, 302-306.

Keller, A.Z. and Kamath, A.R. (1982). Reliability Analysis of CNC Machine Tools. *Reliability Engineering*, **3**, 449–473.

Kobayashi, K . (1991). On Generalized Gamma Functions Occurring in Diffraction Theory. *Journal Physical Society of Japan*, **60**, 1501–1512.

Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications.* Imperial College Press, London.

Krishna, E., Jose, K.K., Alice, T. and Ristic, M.M. (2013). The Marshall-Olkin Fréchet Distribution. *Communications in Statistics-Theory and Methods*, **42**, 4091–4107.

Lai, C.D., and Xie, M. (2006). *Stochastic Ageing and Dependence for Reliability.* Springer Science and Business Media.

Lemonte, A.J. and Cordeiro, G.M. (2013). An Extended Lomax Distribution. *Statistics*, **47**, 800–816.

Mahmoud, M.R. and Mandouh, R.M. (2013). On the Transmuted Fréchet Distribution. *Journal of Applied Sciences Research*, **9**, 5553–5561.

McCullagh, P. (1980). Regression Models for Ordinal Data (with discussion). *Journal of Royal Statistical Society,* Series B, **42**, 109–142.

Mead, M.E. (2015). Generalized Inverse Gamma Distribution and Its Application in Reliability. *Communications in Statistics - Theory and Methods*, **44**, 1426–1435.

Mead, M.E. and Abd-Eltawab A.R. (2014). A Note on Kumaraswamy-Fréchet Distribution. *Australian Journal of Basic and Applied Sciences*, **8**, 294–300.

Nadarajah, S. and Gupta, A.K. (2004). The Beta Fréchet Distribution. *Far East Journal of Theoretical Statistics*, **14**, 15–24.

Nadarajah, S. and Kotz, S. (2003). The Exponentiated Fréchet Distribution. *InterStat.* Available online at http://interstat.statjournals.net/YEAR/2003/abstracts/0312001.pdf.

Taguchi G. (1986). *Introduction to Quality Engineering.* White Plains, NY: Kraus International Publications.

Treyer, V.N. (1964). Doklady Acad, Nauk, Belorus, U.S.S.R.

Zografos, K. and Balakrishnan, N. (2009). On Families of Beta- and Generalized Gamma-generated Distributions and Associated Inference. *Statistical Methodology*, **6**, 344–362.

Affiliation:

M. E. Mead
Department of Statistics, Mathematics and Insurance
Zagazig University, Egypt
E-mail: mead9990@gmail.com

Ahmed Z. Afify
Department of Statistics, Mathematics and Insurance
Benha University, Egypt
E-mail: ahmed.afify@fcom.bu.edu.eg

G. G. Hamedani
Department of Mathematics, Statistics and Computer Science
Marquette University, USA
E-mail: gholamhoss.hamedani@marquette.edu

Indranil Ghosh
Department of Mathematics and Statistics
University of North Carolina Wilmington, USA
E-mail: ghoshi@uncw.edu

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 46 *Submitted:* 2015-11-01
February 2017 *Accepted:* 2016-02-16

Austrian Journal of Statistics
February 2017, Volume 46, 65–87.
http://www.ajs.or.at/
doi:10.17713/ajs.v46i1.222

The Odd Lindley-G Family of Distributions

Frank Gomes-Silva
Federal R. Univ. of Pernambuco
Ana Percontini
State Univ. of F. de Santana
Edleide de Brito
Federal Univ. of Bahia
Manoel W. Ramos
Federal Institute of Paraíba
Ronaldo Venâncio
Recife Military School
Gauss Cordeiro
Federal Univ. of Pernambuco

Abstract

We propose a new generator of continuous distributions with one extra positive parameter called the odd Lindley-G family. Some special cases are presented. The new density function can be expressed as a linear combination of exponentiated densities based on the same baseline distribution. Various structural properties of the new family, which hold for any baseline model, are derived including explicit expressions for the quantile function, ordinary and incomplete moments, generating function, Rényi entropy, reliability, order statistics and their moments and k upper record values. We provide a Monte Carlo simulation study to evaluate the maximum likelihood estimates. We discuss estimation of the model parameters by maximum likelihood and provide an application to a real data set.

Keywords: estimation, Lindley distribution, generating function, odd Lindley-G distribution; k upper record values, moment, Monte Carlo simulation.

1. Introduction

In recent years, several ways of generating new distributions from classic ones has attracted theoretical and applied statisticians due to their flexible properties. Many classical distributions have been extensively used over the past decades for modeling data in several areas. In fact, for instance, Johnson, Kotz, and Balakrishnan (1994, 1995) presented a comprehensive discussion on hundreds of continuous univariate distributions. However, in many applied areas including, but not limited to lifetime analysis, finance and insurance, there is a clear need for extended forms of these distributions, which are more flexible for fitting specific real world scenarios. Consequently, recent developments focus on definition of the new families of distributions that extend well-known distributions and at the same time provide great flexibility in modelling data in practice. Some well-established generators and other recently proposed are the Marshall-Olkin generated family (MO-G) by Marshall and Olkin (1997), beta-G by Eugene, Lee, and Famoye (2002), Kumaraswamy-G (Kw-G for short) by Cordeiro and de Castro (2011), McDonald-G (Mc-G) by Alexander, Cordeiro, Ortega, and Sarabia (2012), gamma-G by Zografos and Balakrishnan (2009), transformed-transformer (T-X) by Alzaatreh, Lee, and Famoye (2013), exponentiated T-X by Alzaghal, Felix, and Carl (2013), Weibull-G by Bourguignon, Silva, and Cordeiro (2014), exponentiated half-logistic family by Cordeiro, Alizadeh, and Ortega (2014), logistic-X by Tahir, Cordeiro, Alzaatreh, Mansoor, and Zubair (2016a), a new Weibull-G by Tahir, Zubair, Mansoor, Cordeiro, Alizadeh, and Hamedani (2016b) and Kumaraswamy odd log-logistic-G by Alizadeh, Emadi, Doostparast, Cordeiro, Ortega, and Pescim (2015). The Lindley distribution was originally proposed by Lindley (1958) as a counterexample of fiducial statistics. Ghitany, Atieh, and Nadarajah (2008) showed through a numerical example that the hazard function of the Lindley distribution does not exhibit a constant hazard rate, indicating its flexibility over the exponential distribution. It has recently received considerable attention as an appropriate model to analyze lifetime data especially in applications modeling stress-strength reliability; see, for example, Ghitany *et al.* (2008), Zakerzadeh and Dolati (2009), Mazucheli and Achcar (2011), Gupta and Singh (2012), Warahena-Liyanage and Pararai (2014). Several other authors including Sankaran (1970), Nadarajah and Tahmasbi (2011) and Asgharzedah, Bakouch, and Esmaeli (2013) developed some structural properties of various generalized Lindley distributions. Nonetheless, there are situations in which the Lindley distribution and all of its generalizations may not be suitable from a theoretical or an applied point of view.

Let $r(t)$ be the probability density function (pdf) of a random variable $T \in [a, b]$ for $-\infty \leq a < b < \infty$ and let $W[G(x)]$ be a function of the cumulative distribution function (cdf) of a random variable X such that $W[G(x)]$ satisfies the following conditions:

$$\begin{cases} (i) & W\left[G(x)\right] \in [a, b], \\ (ii) & W\left[G(x)\right] \text{ is differentiable and monotonically non-decreasing, and} \\ (iii) & W\left[G(x)\right] \rightarrow a \;\text{ as }\; x \rightarrow -\infty \text{ and } W\left[G(x)\right] \rightarrow b \;\text{ as }\; x \rightarrow \infty. \end{cases} \tag{1}$$

Alzaatreh *et al.* (2013) defined the T-X family of distributions by

$$F(x) = \int_a^{W[G(x)]} r(t)\, dt, \tag{2}$$

where $W[G(x)]$ satisfies the conditions (1). The pdf corresponding to (2) is given by

$$f(x) = \left\{ \frac{d}{dx} W[G(x)] \right\} r\left\{ W[G(x)] \right\}. \tag{3}$$

In this paper, we propose a new wider class of continuous distributions called the *Odd Lindley-G* ("OL-G" for short) family by taking $W[G(x)] = \frac{G(x;\boldsymbol{\xi})}{1-G(x;\boldsymbol{\xi})}$ and $r(t) = \frac{a^2}{1+a}\,(1+t)\,\mathrm{e}^{-a\,t}$, $t > 0$, $a > 0$, where $G(x;\boldsymbol{\xi})$ is a baseline cdf, which depends on a parameter vector $\boldsymbol{\xi}$ and $\overline{G}(x;\boldsymbol{\xi}) = 1 - G(x;\boldsymbol{\xi})$ is the baseline survival function. Its cdf is given by

$$\begin{aligned} F(x; a, \boldsymbol{\xi}) &= \int_0^{\frac{G(x;\boldsymbol{\xi})}{1-G(x;\boldsymbol{\xi})}} \frac{a^2}{1+a}\,(1+t)\,\mathrm{e}^{-a\,t}\, dt \\ &= 1 - \frac{a + \overline{G}(x;\boldsymbol{\xi})}{(1+a)\,\overline{G}(x;\boldsymbol{\xi})} \exp\left\{ -a\,\frac{G(x,\boldsymbol{\xi})}{\overline{G}(x;\boldsymbol{\xi})} \right\}. \end{aligned} \tag{4}$$

For each baseline G, the OL-G family of distributions is defined by the cdf (4). Equation (4) is a wider family of continuous distributions. Further, we can omit sometimes the dependence on the vector $\boldsymbol{\xi}$ of the parameters and write simply $G(x) = G(x;\boldsymbol{\xi})$.

The corresponding density function to (4) is given by

$$f(x; a, \boldsymbol{\xi}) = \frac{a^2}{(1+a)}\, \frac{g(x,\boldsymbol{\xi})}{\overline{G}(x;\boldsymbol{\xi})^3} \exp\left\{ -a\,\frac{G(x,\boldsymbol{\xi})}{\overline{G}(x;\boldsymbol{\xi})} \right\}, \tag{5}$$

where $g(x;\boldsymbol{\xi})$ is the baseline pdf. Equation (5) is most tractable when the cdf $G(x)$ and the pdf $g(x)$ have simple analytic expressions. Hereafter, a random variable X with density function (5) is denoted by $X \sim$ OL-G$(a, \boldsymbol{\xi})$. Table 1 lists $G(x;\boldsymbol{\xi})/\overline{G}(x;\boldsymbol{\xi})$ and the corresponding parameters for some special distributions.

Most of the distributions lack physical motivation for modeling lifetime data. We now provide, in a similar context, a physical interpretation for the proposed family inspired in Cooray (2006). Let Y be a lifetime random variable having a certain continuous G distribution. The odds ratio that an individual (or component) following the lifetime Y will die (failure) at time x is $G(x;\boldsymbol{\xi})/\overline{G}(x;\boldsymbol{\xi})$. Consider that the variability of this odds of death is represented by the random variable X and assume that it follows the Lindley model with scale a. We can write

$$\Pr(Y \leq x) = \Pr\left(X \leq \frac{G(x;\boldsymbol{\xi})}{\overline{G}(x;\boldsymbol{\xi})}\right) = F(x;a,\boldsymbol{\xi}),$$

which is given by (4). The rest of the paper is organized as follows. In Section 2, we present

Table 1: Distributions and corresponding $G(x;\boldsymbol{\xi})/\overline{G}(x;\boldsymbol{\xi})$ functions

Distribution	$G(x;\boldsymbol{\xi})/\overline{G}(x;\boldsymbol{\xi})$	$\boldsymbol{\xi}$
Uniform $(0 < x < \theta)$	$x/(\theta - x)$	θ
Exponential $(x > 0)$	$\mathrm{e}^{\lambda x} - 1$	λ
Weibull $(x > 0)$	$\mathrm{e}^{\lambda x^{\gamma}} - 1$	(λ, γ)
Fréchet $(x > 0)$	$(\mathrm{e}^{\lambda x^{\gamma}} - 1)^{-1}$	(λ, γ)
Half-logistic $(x > 0)$	$(\mathrm{e}^{x} - 1)/2$	$\varnothing$
Power function $(0 < x < 1/\theta)$	$[(\theta\, x)^{-k} - 1]^{-1}$	(θ, k)
Pareto $(x \geq \theta)$	$(x/\theta)^{k} - 1$	(θ, k)
Burr XII $(x > 0)$	$[1 + (x/s)^{c}]^{k} - 1$	(s, k, c)
Log-logistic $(x > 0)$	$[1 + (x/s)^{c}] - 1$	(s, c)
Lomax $(x > 0)$	$[1 + (x/s)]^{k} - 1$	(s, k)
Gumbel $(-\infty < x < \infty)$	$\{\exp[\exp(-(x-\mu)/\sigma)] - 1\}^{-1}$	(μ, σ)
Kumaraswamy $(0 < x < 1)$	$(1 - x^{\alpha})^{-\beta} - 1$	(α, β)
Normal $(-\infty < x < \infty)$	$\Phi((x-\mu)/\sigma)/(1 - \Phi((x-\mu)/\sigma))$	(μ, σ)

four special models of the new family. A range of mathematical properties of Equation (5) is derived in Section 3 including a useful expansion for the pdf and explicit expressions for the moments and generating function. General expressions for the Rényi entropy, reliability, order statistics and k upper record values are discussed in Section 4. Estimation of the model parameters by maximum likelihood is performed in Section 5. In Section 6, we conduct a simulation study for specific choices of a model parameter. An application to a real data set illustrates the performance of the new family in Section 6. The paper is concluded in Section 7.

2. Four special models of the OL-G family

In this section, we present four special models of the OL-G family.

2.1. Odd Lindley Weibull (OLW)

The Weibull cdf with parameters $\alpha > 0$ and $\lambda > 0$ is $G(x) = 1 - \mathrm{e}^{-(\lambda x)^{\alpha}}$ (for $x > 0$). The cdf of a random variable X having the OLW distribution, say $X\sim$ OLW(a, α, λ), is given by

$$F_{\mathrm{OLW}}(x) = (1+a)^{-1}\exp\{-a[a + \mathrm{e}^{(\lambda x)^{\alpha}}]\}\left\{(1+a+a^2)\exp[a\mathrm{e}^{(\lambda x)^{\alpha}}] - \mathrm{e}^{a(1+a)}\left[1 + a\,\mathrm{e}^{(\lambda x)^{\alpha}}\right]\right\},$$

and the associated density function reduces to

$$f_{\mathrm{OLW}}(x) = a^2\,(1+a)^{-1}\,\alpha\,\lambda^{\alpha}\,x^{\alpha-1}\,\mathrm{e}^{2(\lambda x)^{\alpha}}\exp\{-a[\mathrm{e}^{(\lambda x)^{\alpha}} - 1]\}. \qquad (6)$$

The hazard rate function (hrf) corresponding to (6) is given by

$$\begin{aligned}\tau_{\mathrm{OLW}}(x) &= a^2\,\alpha\,\lambda^\alpha\,x^{\alpha-1}\,\mathrm{e}^{2(\lambda x)^\alpha}\exp\{-a[\mathrm{e}^{(\lambda x)^\alpha}-1]\}\Big\{1+a-\exp\{-a[a+\mathrm{e}^{(\lambda x)^\alpha}]\}\\ &\times\Big[(1+a+a^2)\exp[a\mathrm{e}^{(\lambda x)^\alpha}]-\mathrm{e}^{a(1+a)}\left\{1+a\,\mathrm{e}^{(\lambda x)^\alpha}\right\}\Big]\Big\}^{-1}.\end{aligned}$$

Plots of the density and hrf of the OLW distribution for some parameter values are displayed in Figure 1.

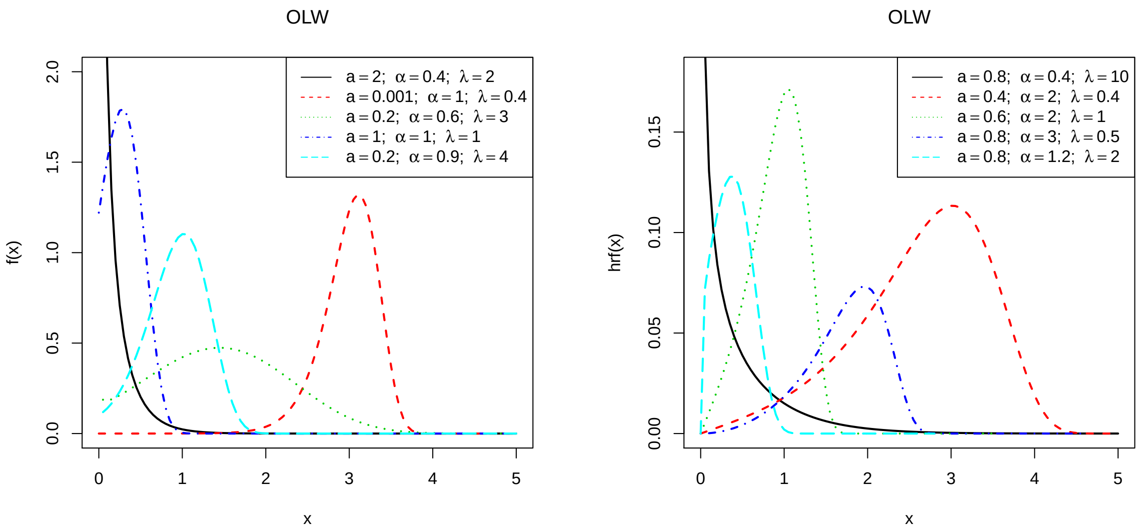


Figure 1: Plots of the OLW density and hrf functions for some parameter values.

2.2. Odd Lindley Kumaraswamy (OLKw)

The Kumaraswamy cumulative distribution (for $x \in [0,1]$) is $G(x) = 1-(1-x^\alpha)^\beta$, where the parameters are $\alpha > 0$ and $\beta > 0$. The OLKw cumulative distribution is given by

$$F_{\mathrm{OLKw}}(x) = (1+a)^{-1}\left\{(1+a+a^2)\,\mathrm{e}^{-a^2}+\exp[a-a(1-x^\alpha)^{-\beta}]\,[-1-a(1-x^\alpha)^{-\beta}]\right\}$$

and the associated density function reduces to

$$f_{\mathrm{OLKw}}(x) = a^2\,(1+a)^{-1}\,\alpha\,\beta\,x^{\alpha-1}\,(1-x^\alpha)^{-2\beta-1}\exp\{-a[(1-x^\alpha)^{-\beta}-1]\}. \qquad (7)$$

The hrf corresponding to (7) is given by

$$\begin{aligned}\tau_{\mathrm{OLKw}}(x) &= \alpha\,\beta\,x^{\alpha-1}\,(1-x^\alpha)^{-2\beta-1}\exp\{-a[(1-x^\alpha)^{-\beta}-1]\}\\ &\times\left\{1+a-\left[(1+a+a^2)\,\mathrm{e}^{-a^2}+\exp[a-a(1-x^\alpha)^{-\beta}]\,\{-1-a(1-x^\alpha)^{-\beta}\}\right]\right\}^{-1}.\end{aligned}$$

Plots of the density and hrf of the OLKw distribution for some parameter values are displayed in Figure 2.

2.3. Odd Lindley half-logistic (OLHL)

The half-logistic cumulative distribution (for $x > 0$) is given by $G(x) = \frac{1-\mathrm{e}^{-x}}{1+\mathrm{e}^{-x}}$. The OLHL cumulative distribution becomes

$$\begin{aligned}F_{\mathrm{OLHL}}(x) &= [2(1+a)]^{-1}\exp\left[\frac{a}{2}(1-2a-\mathrm{e}^{x})\right]\\ &\times\left\{-(2+a)\,\mathrm{e}^{a^2}+2(1+a+a^2)\exp\left[\frac{a}{2}(-1+\mathrm{e}^{x})\right]-a\,\mathrm{e}^{a^2+x}\right\}\end{aligned}$$

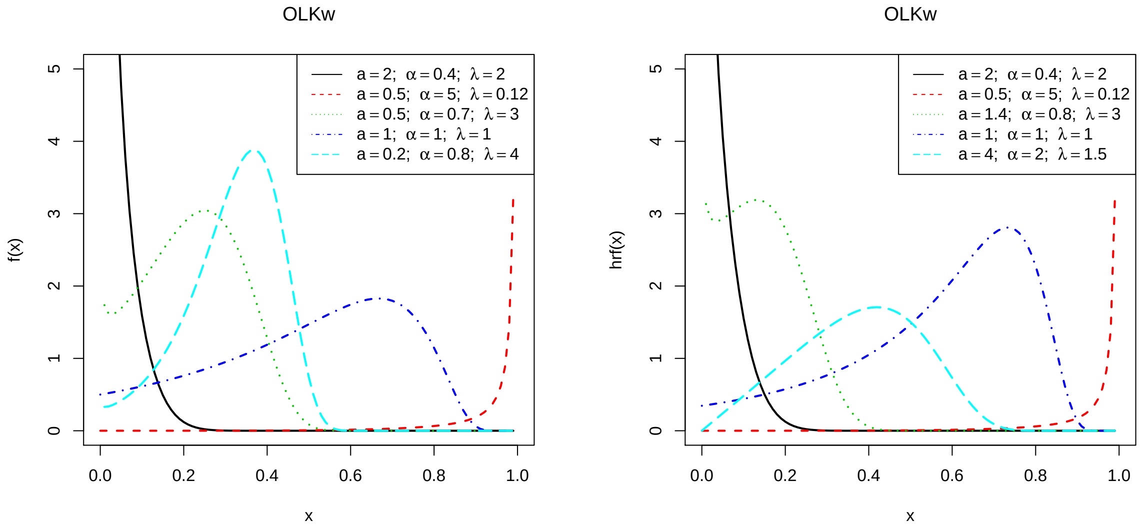


Figure 2: Plots of the OLKw density and hrf functions for some parameter values.

and the associated density function reduces to

$$f_{\text{OLHL}}(x) \quad = \quad a^2\,[4(1+a)]^{-1}\,(1+\mathrm{e}^x)\exp\left[\frac{a}{2}(1-\mathrm{e}^x)+x\right]. \tag{8}$$

The corresponding hrf to (8) is given by

$$\begin{aligned}\tau_{\text{OLHL}}(x) \quad &= \quad \frac{a^2}{2}\,(1+\mathrm{e}^x)\exp\left[\frac{a}{2}(1-\mathrm{e}^x)+x\right]\left\{2(1+a)-\exp\left[\frac{a}{2}(1-2a-\mathrm{e}^x)\right]\right.\\ &\times \quad \left.\left[-(2+a)\,\mathrm{e}^{a^2}+2(1+a+a^2)\exp\left[\frac{a}{2}(-1+\mathrm{e}^x)\right]-a\,\mathrm{e}^{a^2+x}\right]\right\}^{-1}.\end{aligned}$$

Plots of the pdf and hrf of the OLHL distribution for some parameter values are displayed in Figure 3.

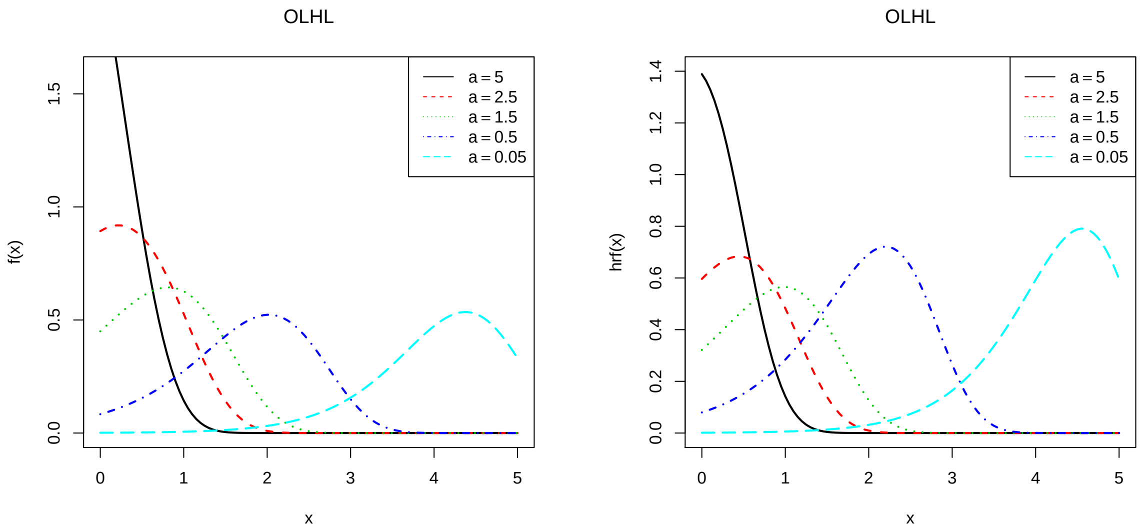


Figure 3: Plots of the OLHL density and hrf functions for some parameter values.

2.4. Odd Lindley Burr XII (OLBXII)

Zimmer, Keats, and Wang (1998) introduced the three parameter Burr XII (BXII) distribution with cdf and pdf (for $x > 0$): $G(x; s, k, c) = 1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}$ and $g(x; s, k, c) = c\,k\,s^{-c}\,x^{c-1}\left[1 + \left(\frac{x}{s}\right)^c\right]^{-k-1}$, respectively, where $k > 0$ and $c > 0$ are shape parameters and $s > 0$ is a scale parameter. The OLBXII cumulative distribution becomes

$$\begin{aligned} F_{\mathrm{OLBXII}}(x) &= (1+a)^{-1} \exp\{a + [1 + (\frac{x}{s})^c]^k\} \\ &\times \left\{(1 + a + a^2) \exp\{a[1 + (\frac{x}{s})^c]^k\} - e^{a(1+a)} \left(1 + a[1 + (\frac{x}{s})^c]^k\right)\right\} \end{aligned}$$

and the associated density function reduces to

$$f_{\mathrm{OLBXII}}(x) = a^2\,c\,k\,s^{-c}\,(1+a)^{-1}\,x^{c-1}\left[1 + \left(\frac{x}{s}\right)^c\right]^{2k-1} \exp\left[a\left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^k\right\}\right]. \quad (9)$$

The corresponding hrf is given by

$$\begin{aligned} \tau_{\mathrm{OLBXII}}(x) &= a^2\,c\,k\,s^{-c}\,x^{c-1}\left[1 + \left(\frac{x}{s}\right)^c\right]^{2k-1} \exp\left[a\left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^k\right\}\right] \\ &\times \left\{1 + a - \exp\{a + [1 + (\frac{x}{s})^c]^k\}\right. \\ &\times \left.\left[(1 + a + a^2) \exp\{a[1 + (\frac{x}{s})^c]^k\} - e^{a(1+a)} \left(1 + a[1 + (\frac{x}{s})^c]^k\right)\right]\right\}^{-1}. \end{aligned}$$

Plots of the density and hrf of the OLBXII distribution for some parameter values are displayed in Figure 4.

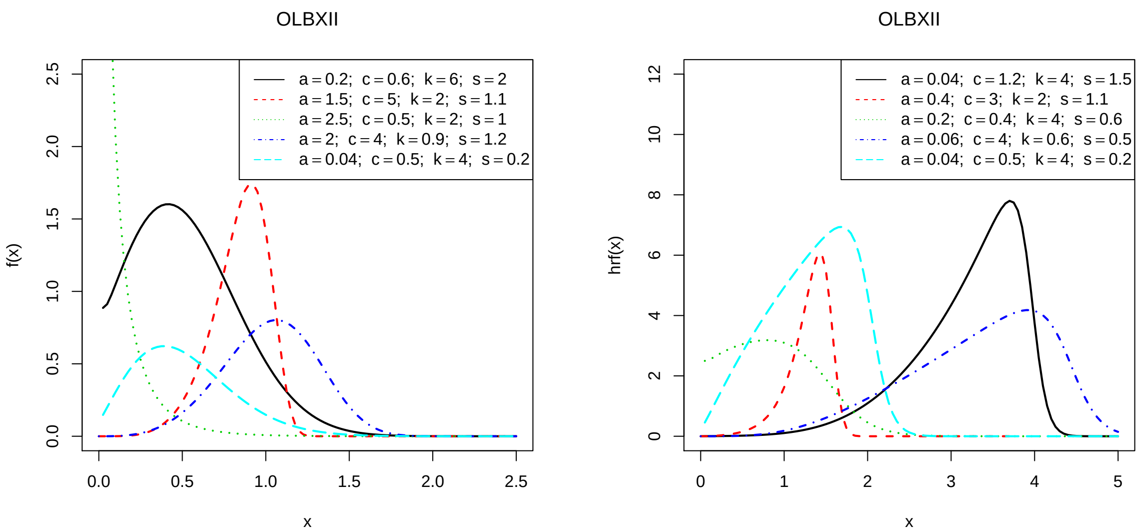


Figure 4: Plots of the OLBXII density and hrf functions for some parameter values.

3. Main properties

3.1. Survival and hazard

The corresponding survival function to (4) is given by

$$S(x; a, \boldsymbol{\xi}) = 1 - F(x; a, \boldsymbol{\xi}) = \frac{a + \overline{G}(x; \boldsymbol{\xi})}{(1+a)\,\overline{G}(x; \boldsymbol{\xi})} \exp\left\{-a\,\frac{G(x, \boldsymbol{\xi})}{\overline{G}(x; \boldsymbol{\xi})}\right\}. \quad (10)$$

The hrf of X becomes

$$\tau(x;a,\boldsymbol{\xi}) = \frac{a^2\, g(x,\boldsymbol{\xi})}{\overline{G}(x;\boldsymbol{\xi})^2\,[a+\overline{G}(x;\boldsymbol{\xi})]} = \frac{a^2}{\overline{G}(x;\boldsymbol{\xi})\,[a+\overline{G}(x;\boldsymbol{\xi})]}\,\tau(x;\boldsymbol{\xi}), \tag{11}$$

where $\tau(x;\boldsymbol{\xi}) = g(x;\boldsymbol{\xi})/\overline{G}(x;\boldsymbol{\xi})$. The multiplying quantity $a^2/\{\overline{G}(x;\boldsymbol{\xi})\,[a+\overline{G}(x;\boldsymbol{\xi})]\}$ works as a corrected factor for the baseline hrf. Equation (4) can deal with general situations in modeling survival data with various shapes of the hrf.

3.2. Quantile functions

Let X be an arbitrary random variable with cdf $F(x) = \Pr(X \le x)$, where $x \in \mathbb{R}$. For any $u \in (0,1)$, the uth quantile function (qf) $Q(u)$ of X is the solution of

$$F(Q(u)) = u, \tag{12}$$

for $Q(u) > 0$.

For any fixed $a > 0$, from Equation (4), we obtain

$$-\frac{1+a-G(Q(u))}{1-G(Q(u))}\,\mathrm{e}^{-\frac{aG(Q(u))}{1-G(Q(u))}} = (1+a)(u-1).$$

Multiplying both sides of this equation by $\mathrm{e}^{-(1+a)}$ gives

$$-\frac{1+a-G(Q(u))}{1-G(Q(u))}\,\mathrm{e}^{-\frac{1+a-G(Q(u))}{1-G(Q(u))}} = (1+a)(u-1)\,\mathrm{e}^{-(1+a)}.$$

In the above equation, we note that $-\frac{1+a-G(Q(u))}{1-G(Q(u))}$ is the Lambert W function of the real argument $(1+a)(u-1)\mathrm{e}^{-(1+a)}$. The Lambert W function is defined by

$$W(x)\,\mathrm{e}^{W(x)} = x.$$

The Lambert function has two real branches with a branching point located at $(-\mathrm{e}^{-1},1)$. The lower branch, $W_{-1}(x)$, is defined in the interval $[-\mathrm{e}^{-1},1]$ and has a negative singularity for $x \to 0^-$. The upper branch, $W_0(x)$, is defined for $x \in [-\mathrm{e}^{-1},\infty]$.

Then, we have

$$W\left((1+a)(u-1)\mathrm{e}^{-(1+a)}\right) \;=\; -\left(1+\frac{a}{1-G(Q(u))}\right). \tag{13}$$

Clearly, for any $a > 0$ and $u \in (0,1)$, we have $\left(1+\frac{a}{1-G(Q(u))}\right) > 1$ and then $(1+a)(u-1)\mathrm{e}^{-(1+a)} < 0$. Therefore, considering the lower branch of the Lambert W function, we can write (13) as

$$W_{-1}\left((1+a)(u-1)\mathrm{e}^{-(1+a)}\right) \;=\; -\left(1+\frac{a}{1-G(Q(u))}\right).$$

Hence, the qf of X is given by

$$Q(u) = G^{-1}\Big\{1+a\left[1+W_{-1}\left((1+a)(u-1)\mathrm{e}^{-(1+a)}\right)\right]^{-1}\Big\}. \tag{14}$$

3.3. Shapes of the OL-G family

The shapes of the density and hazard rate functions can also be described analytically. The critical points of the OL-G density function are the roots of the equation:

$$\frac{g'(x)}{g(x)} + 3\,\frac{g(x)}{\overline{G}(x)} - a\,\frac{g(x)}{\overline{G}(x)^2} = 0. \tag{15}$$

The critical points of $h(x)$ are obtained from the following equation

$$\frac{g'(x)}{g(x)} + \frac{g(x)}{a+\overline{G}(x)} + 2\,\frac{g(x)}{\overline{G}(x)^2} = 0. \tag{16}$$

By using most computer algebra systems, we can examine Equations (15) and (16) to determine the local maximums and minimums and inflexion points.

3.4. Useful expansions

Several structural properties of the extended distributions may be easily explored using mixture forms of exponentiated-G ("Exp-G") distributions. In this section, we obtain expansions for $f(x)$ and $F(x)$. First, we define the Exp-G distribution for an arbitrary parent distribution $G(x)$, say $W \sim$ Exp-G(c), if W has cdf and pdf given by

$$H_c(x;\boldsymbol{\xi}) = G(x;\boldsymbol{\xi})^c \qquad \text{and} \qquad h_c(x;\boldsymbol{\xi}) = c\,g(x;\boldsymbol{\xi})\,G(x;\boldsymbol{\xi})^{c-1},$$

respectively. Next, we obtain an expansion for $f(x)$. Using the power series for the exponential function, we have

$$\exp\left\{-a\left[\frac{G(x;\boldsymbol{\xi})}{\overline{G}(x;\boldsymbol{\xi})}\right]\right\} = \sum_{k=0}^{\infty}\frac{(-1)^k\,a^k}{k!}\left[\frac{G(x;\boldsymbol{\xi})}{\overline{G}(x;\boldsymbol{\xi})}\right]^k.$$

Inserting this expansion in Equation (5), we have

$$f(x;a,\boldsymbol{\xi}) = \frac{a^2}{1+a}\,g(x,\boldsymbol{\xi})\sum_{k=0}^{\infty}\frac{(-1)^k a^k}{k!}\,\frac{G(x;\boldsymbol{\xi})^k}{\overline{G}(x;\boldsymbol{\xi})^{k+3}}. \tag{17}$$

By using the generalized binomial expansion, we can write

$$\overline{G}(x;\boldsymbol{\xi})^{-(k+3)} = \sum_{i=0}^{\infty}\frac{\Gamma(i+k+3)}{i!\,\Gamma(k+3)}\,G(x;\boldsymbol{\xi})^i. \tag{18}$$

Inserting (18) in (17), the OL-G density function can be expressed as an infinite mixture of Exp-G density functions

$$f(x;a,\boldsymbol{\xi}) = \sum_{i,\,k=0}^{\infty}\gamma_{i,k}\,h_{i+k+1}(x;\boldsymbol{\xi}), \tag{19}$$

where

$$\gamma_{i,k} = \frac{(-1)^k\,a^{2+k}\,\Gamma(i+k+3)}{(a+1)(i+k+1)\,i!\,k!\,\Gamma(k+3)}.$$

The cdf of X can be given by integrating (19) as

$$F(x;a,\boldsymbol{\xi}) = \sum_{i,\,k=0}^{\infty}\gamma_{i,k}\,H_{i+k+1}(x;\boldsymbol{\xi}). \tag{20}$$

The properties of Exp-G distributions have been studied by many authors in recent years, see Mudholkar and Srivastava (1993) and Mudholkar, Srivastava, and Freimer (1995) for exponentiated Weibull, Gupta, Gupta, and Gupta (1998) for exponentiated Pareto, Gupta and Kundu (1999) for exponentiated exponential, Nadarajah (2005) for exponentiated Gumbel, Shirke and Kakade (2006) for exponentiated log-normal and Nadarajah and Gupta (2007) for exponentiated gamma distributions.

Thus, some structural properties of the new family such as the ordinary and incomplete moments and generating function can be determined from well-established properties of the Exp-G distributions.

Note that Equations (19) and (20) are the main results of this section.

3.5. Moments

From now on, we assume that $Y_{i+k} \sim \text{Exp-G}(i+k+1)$. Many of the important features and characteristics of a distribution can be obtained using ordinary moments. A first formula for the nth moment of X can be obtained from (19) as

$$\mu'_n = E(X^n) = \sum_{i,\,k=0}^{\infty} \gamma_{i,k}\, E(Y_{i+k}^n). \tag{21}$$

Closed-form expressions for moments of several Exp-G distributions are given by Nadarajah and Kotz (2006) that can be used to obtain OL-G moments. For instance, the moments of the OLW model (for $n > -\alpha$) discussed in Section 2 can be derived from closed-forms moments of the exponetiated Weibull given by Nadarajah and Kotz (2006). In this case, we obtain

$$\mu'_n = \lambda^{-n}\,\Gamma(n/\alpha+1) \sum_{i,k=0}^{\infty} (i+k+1)\,\gamma_{i,k} \sum_{j=0}^{\infty} \frac{(-i-k)_j}{j!\,(j+1)^{n/\alpha+1}}. \tag{22}$$

A second alternative formula for μ'_n can be obtained from (21) in terms of the baseline quantile function $Q_G(u)$. We obtain

$$\mu'_n = \sum_{i,k=0}^{\infty} (i+k+1)\,\gamma_{i,k}\,\tau_{n,i+k}, \tag{23}$$

where the integral depends on the baseline qf

$$\tau_{n,j} = \int_0^1 Q_G(u)^n\, u^j\, du. \tag{24}$$

Note that the ordinary moments of several OL-G distributions can be determined directly from Equations (23) and (24). We now provide the PWMs (Probability Weighted Moments) for one distribution discussed in Section 2. Cordeiro and Nadarajah (2011) determined $\tau_{r,s}$ for some well-known distribution such as normal, beta, gamma, and Weibull distributions, which can be applied to obtain raw moments of the corresponding OL-G distributions.

For instance, in OL-N distribution discussed in Section 2, the quantities $\tau_{r,s}$ can be expressed in terms of the Lauricella functions of type A, Exton (1978) and Trott (2006) defined by

$$F_A^{(n)}(a; b_1,\ldots,b_n; c_1,\ldots,c_n; x_1,\ldots,x_n) = \\ \sum_{m1=0}^{\infty} \cdots \sum_{m_n=0}^{\infty} \frac{(a)_{m_1+\ldots+m_n}(b_1)_{m_1}\ldots(b_n)_{m_n}}{(c_1)_{m_1}\ldots(c_n)_{m_n}}\, \frac{x_1^{m_1}\ldots x_n^{m_n}}{m_1!\ldots m_n!},$$

where $(a)_i = a(a+1)\ldots(a+i-1)$ is the ascending factorial given by (with the convention that $(a)_0 = 1$).

In fact, Cordeiro and Nadarajah (2011) determined $\tau_{r,s}$ for the standard normal distribution as

$$\tau_{r,s} = 2^{r/2}\,\pi^{-(s+1/2)} \sum_{\substack{l=0\\(r+s-l)\,\text{even}}}^{s} \binom{s}{l} 2^{-l}\,\pi^{l}\,\Gamma\left(\frac{r+s-l+1}{2}\right) \times$$
$$F_A^{(s-l)}\left(\frac{r+s-l+1}{2}; \frac{1}{2},\ldots,\frac{1}{2}; \frac{3}{2},\ldots,\frac{3}{2}; -1,\ldots,-1\right).$$

This equation holds when $r+s-l$ is even and it vanishes when $r+s-l$ is odd. So, any OLN moment can be expressed as an infinite weighted linear combination of Lauricella functions of type A.

Further, the central moments (μ_r) and cumulants (κ_r) of X can be determined from the ordinary moments using the recurrence equations

$$\mu_r = \sum_{k=0}^{r}(-1)^k \binom{r}{k} \mu_1'^{k}\,\mu_{r-k}' \qquad \text{and} \qquad \kappa_r = \mu_r' - \sum_{k=1}^{r-1}\binom{r-1}{k-1}\kappa_k\,\mu_{r-k}',$$

respectively, where $\kappa_1 = \mu_1'$. Then, $\kappa_2 = \mu_2' - \mu_1'^2$, $\kappa_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$, $\kappa_4 = \mu_4' - 4\mu_3'\mu_1' - 3\mu_2'^2 + 12\mu_2'\mu_1'^2 - 6\mu_1'^4$, etc. The skewness $\rho_1 = \kappa_3/\kappa_2^{3/2}$ and kurtosis $\rho_2 = \kappa_4/\kappa_2^2$ can be obtained from the second, third and fourth cumulants.

The incomplete moments play an important role for measuring inequality. For example, the main application of the first incomplete moment refers to the Lorenz and Bonferroni curves.

The nth incomplete moment of X is calculated as

$$m_n(y) = E\left(X^n | X < y\right) = \sum_{i,k=0}^{\infty} (i+k+1)\,\gamma_{i,k} \int_0^{G(y)} Q_G(u)^n u^{i+k} du.$$

The last integral can be evaluated for most baseline G distributions.

3.6. Generating function

Here, we provide two formulae for the mgf $M(t) = M(t; a, \boldsymbol{\xi}) = E[\exp(t\,X)]$ of X. The first one comes from (19) as

$$M(t) = \sum_{i,k=0}^{\infty} \gamma_{i,k}\, M_{i+k}(t), \tag{25}$$

where $M_{i+k}(t)$ is the mgf of Y_{i+k}. Hence, $M(t)$ can be determined from the generating function of the Exp-G distribution.

A second formula for $M(t)$ can be derived from (19) as

$$M(t) = \sum_{i,k=0}^{\infty} (i+k+1)\,\gamma_{i,k}\,\rho(t, i+k), \tag{26}$$

where

$$\rho(t,b) = \int_{-\infty}^{\infty} \exp(tx)\,G(x)^b\,g(x)dx = \int_0^1 \exp\left\{t\,Q_G(u)\right\} u^b du. \tag{27}$$

We can obtain the mgf of several OL-G distributions directly from Equations (26) and (27). These equations are the main results of this section.

4. Other measures

4.1. Entropy

The entropy of a random variable X with density function $f(x)$ is a measure of variation of the uncertainty. Two popular entropy measures are due to Shannon (1951) and Rényi (1961). A large value of the entropy indicates the greater uncertainty in the data. The Rényi entropy is defined by (for $\gamma > 0$ and $\gamma \neq 1$)

$$I_R(\gamma) = \frac{1}{1-\gamma} \log \left(\int_0^\infty f^\gamma(x)dx \right).$$

The Shannon entropy is given by $E\left\{-\log[f(X)]\right\}$. It is a special case of the Rényi entropy when $\gamma \uparrow 1$.

Here, we derive expressions for the Rényi entropy for the OL-G distribution. Due to the fact that the parameter γ is not in general a natural number, it is difficult to use (19) for entropy derivation. So, we use (5), the power series for the exponential and the generalized binomial expansion to obtain the Rényi entropy of X as

$$I_R(\gamma) = \frac{1}{1-\gamma} \left\{ \gamma \log \left(\frac{a^2}{1-a} \right) + \log \left[\sum_{i,k=0}^{\infty} \frac{(-1)^k \, (a\,\gamma)^k \, \Gamma(3\gamma + i + k)}{i! \, k! \, \Gamma(3\gamma + k)} \, K(\gamma, i, k) \right] \right\}.$$

Here, $K(\gamma, i, k)$ denotes the integral

$$K(\gamma, i, k) = \int_0^1 g^{\gamma - 1}[Q_G(u)] \, u^{i+k} du,$$

to be evaluated for each OL-G model. For the OL-exponential (with parameter $\lambda > 0$), OL-Weibull (with parameters $\alpha > 0$ and $\lambda > 0$) and OL-Pareto (with parameter $\gamma > 0$) distributions, we obtain

$$\begin{aligned} K(\gamma, i, k) &= \lambda^{\gamma-1} \, B(i + k, \gamma - 1), \\ K(\gamma, i, k) &= \beta^{\gamma-1} \, \Gamma \left(\frac{(\alpha - 1)(\gamma - 1)}{\alpha} + 1 \right) \sum_{p=0}^{i+k} (-1)^p \, (p + \gamma)^{-\frac{(\alpha-1)(\gamma-1)}{\alpha} - 1} \binom{i+k}{p}, \end{aligned}$$

where $\alpha > (\gamma - 1)/\gamma$, and

$$K(\gamma, i, k) = \gamma^{\gamma-1} \, B(i + k, (1 + \gamma^{-1})(\gamma - 1)),$$

respectively.

4.2. Reliability

The measure of reliability of industrial components has many applications especially in the area of engineering. The reliability of a product (system) is the probability that the product (system) will perform its intended function for a specified time period when operating under normal (or stated) environmental conditions. The component fails at the instant that the random stress X_2 applied to it exceeds the random strength X_1, and the component will function satisfactorily whenever $X_1 > X_2$. Hence, $R = P(X_2 < X_1)$ is a measure of component reliability (see Kotz, Lai, and Xie (2003)). We derive the reliability R when X_1 and X_2 have independent OL-G$(x, a_1, \boldsymbol{\xi})$ and OL-G$(x, a_2, \boldsymbol{\xi})$ distributions with the same parameter vector $\boldsymbol{\xi}$ for the baseline G. The reliability is defined by

$$R = \int_0^\infty f_1(x) F_2(x) dx.$$

The pdf of X_1 and cdf of X_2 are obtained from Equations (19) and (20) as

$$f_1(x) = \sum_{i,j=0}^{\infty} p_{i,j}(a_1)\, g(x, \boldsymbol{\xi})\, G(x; \boldsymbol{\xi})^{i+j} \quad \text{and} \quad F_2(x) = \sum_{k,l=0}^{\infty} q_{k,l}(a_2)\, G(x; \boldsymbol{\xi})^{k+l+1},$$

where

$$p_{i,j} = \frac{(-1)^j a_1^{2+j} \Gamma(i+j+3)}{i! j! (a_1+1) \Gamma(j+3)}$$

and

$$q_{k,l} = \frac{(-1)^l a_2^{2+l} \Gamma(k+l+3)}{k! l! (a_2+1)(k+l+1) \Gamma(l+3)}.$$

Hence,

$$R = \sum_{i,j,k,l=0}^{\infty} p_{i,j}(a_1)\, q_{k,l}(a_2) \int_0^\infty g(x; \boldsymbol{\xi})\, G(x; \boldsymbol{\xi})^{i+j+k+l+1}\, dx.$$

Setting $u = G(x; \boldsymbol{\xi})$, the reliability of the OL-G distribution reduces to

$$R = \sum_{i,j,k,l=0}^{\infty} \frac{p_{i,j}(a_1)\, q_{k,l}(a_2)}{i+j+k+l+2}.$$

4.3. Order statistics

A branch of statistics known as order statistics plays a proeminent role in real-life applications involving data relating to life testing studies. These statistics are required in many fields, such as climatology, engineering and industry, among others. A comprehensive exposition of order statistics and associated inference is provided by David and Nagaraja (2003). Let $X_{i:n}$ denote the ith order statistic. The density $f_{i:n}(x)$ of the ith order statistic, for $i = 1, \ldots, n$, from independent and identically distributed random variables $X_1, \ldots, X_n$ having the OL-G distribution is given by

$$f_{i:n}(x) = M\, f(x)\, F(x)^{i-1} \left[1 - F(x)\right]^{n-i},$$

where $M = n!/[(i-1)!\,(n-i)!]$. First of all, by adding and subtracting $a\,G(x)$ in numerator of $\frac{a+\overline{G}(x)}{(1+a)\,\overline{G}(x)}$, the cdf (4) can be rewritten as

$$F(x) = 1 - \left\{1 + \frac{a}{(1+a)} \left[\frac{G(x)}{\overline{G}(x)}\right]\right\} \exp\left\{-a \frac{G(x)}{\overline{G}(x)}\right\}. \tag{28}$$

From Equations (5) and (28),

$$\begin{aligned} f_{i:n}(x) &= M \frac{a^2}{1+a} \frac{g(x)}{\overline{G}(x)} \sum_{k=0}^{i-1} (-1)^k \binom{i-1}{k} \left\{1 + \frac{a}{1+a} \frac{G(x)}{\overline{G}(x)}\right\}^{k+n-i} \\ &\times \exp\left\{-a\,(k+n-i+1) \frac{G(x)}{\overline{G}(x)}\right\}. \end{aligned}$$

The folowing equations are obtained by using the power series for the exponential function and the generalized binomial expansion:

$$\exp\left\{-a\,(k+n-i+1)\left[\frac{G(x)}{\overline{G}(x)}\right]\right\} = \sum_{m=0}^{\infty} \frac{(-1)^m\, a^m\,(n+k-i+1)^m}{m!}\left[\frac{G(x)}{\overline{G}(x)}\right]^m \tag{29}$$

and

$$[1-G(x)]^{-(j+m+1)} = \sum_{p=0}^{\infty}\binom{j+m+p}{j+m} G(x)^p. \tag{30}$$

Based on Equations (29) and (30), the density of the order statistic $X_{i:n}$ can be expressed as a mixture of Exp-G densities.

$$f_{i:n}(x) = \sum_{m,p=0}^{\infty}\sum_{j=0}^{k+n-i} \gamma_{j,m,p}\, h_{j+m+p}(x), \tag{31}$$

where

$$\gamma_{j,m,p} = \frac{M\, a^{j+m+2}}{m!\,(1+a)^{j+1}\,(j+m+p+1)}\binom{j+m+p}{j+m}\sum_{k=0}^{i-1}(-1)^{k+m}\binom{k+n-i}{j}\binom{i-1}{k}.$$

Clearly, the cdf of $X_{i:n}$ can be expressed as

$$F_{i:n}(x) = \sum_{m,p=0}^{\infty}\sum_{j=0}^{k+n-i} \gamma_{j,m,p}\, H_{j+m+p}(x). \tag{32}$$

Hence, several mathematical quantities of the OL-G order statistics such as the ordinary, incomplete and factorial moments, mgf and mean deviations can be determined from those quantities of the Exp-G distributions. For example, from Equation (31), the moments and mgf of $X_{i:n}$ are given by

$$E(X_{i:n}^s) = \sum_{m,p=0}^{\infty}\sum_{j=0}^{k+n-i} \gamma_{j,m,p}\, E(Z_{j+m+p}^s)$$

and

$$M_{i:n}(t) = \sum_{m,p=0}^{\infty}\sum_{j=0}^{k+n-i} \gamma_{j,m,p}\, E(\mathrm{e}^{t\,Z_{j+m+p}}),$$

where $Z_{j+m+p} \sim$ Exp-G$(j+m+p)$. Equations (31) and (32) are the main results of this section.

4.4. K upper record values

Chandler (1952) formulated the theory of record values as a model for successive extremes in a sequence of independently and identically random variables. Record values are found in many real life applications involving data related to economics, sports, weather and life testing problems. The statistical study of record values has now spread in various directions. Dziubdziela and Kopocinski (1976) proposed the model of k upper record values by observing successive k largest values in a sequence, where k is a positive integer.

Let $X_n^{(k)}$ denote the kth upper record value. The pdf $f_{X_n^{(k)}}(x)$ of the kth upper record value, for $k = 1, \ldots, n$, from independent and identically distributed random variables $X_1, \ldots, X_n$ from the OL-G distribution is given by

$$f_{X_n^{(k)}}(x) = \frac{k^n}{(n-1)!}\{-\log[1-F(x)]\}^{n-1}\,[1-F(x)]^{k-1}\, f(x). \tag{33}$$

By expanding the logarithm function in power series and using the binomial expansion, we have

$$f_{X_n^{(k)}}(x) = \frac{k^n}{(n-1)!} \left\{ \sum_{p=0}^{\infty} a_p \, F(x)^{p+1} \right\}^{n-1} \left\{ \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} F(x)^j \right\} f(x), \tag{34}$$

where $a_p = 1/(p+1)$.

Here, we use an equation by Gradshteyn and Ryzhik (2007) (Section 0.314) for a power series raised to a positive integer n

$$\left(\sum_{i=0}^{\infty} a_i \, u^i \right)^n = \sum_{i=0}^{\infty} c_{n,i} \, u^i, \tag{35}$$

where the coefficients $c_{n,i}$ (for $i = 1, 2, \ldots$) are determined from the recurrence equation

$$c_{n,i} = (i \, a_0)^{-1} \sum_{m=1}^{i} [m \, (n+1) - i] \, a_m \, c_{n,i-m},$$

where $c_{n,0} = a_0^n$.

Based on Equations (34) and (35), the pdf (33) can be expressed as

$$f_{X_n^{(k)}}(x) = \frac{k^n}{(n-1)!} \sum_{p=0}^{\infty} \sum_{j=0}^{k-1} (-1)^j \, c_{n-1,p} \binom{k-1}{j} F(x)^{j+n+p-1} \, f(x),$$

where $c_{n-1,p}$ can be obtained from the quantities $a_0, \ldots, a_p$ as in Equation (35).

From Equations (5) and (28), and following similar algebra of Section 4.3, we obtain

$$f_{X_n^{(k)}}(x) = \sum_{r=0}^{q} \sum_{s,t=0}^{\infty} \phi_{r,s,t} \, h_{r+s+t+1}(x), \tag{36}$$

where

$$\begin{aligned} \phi_{r,s,t} \quad &= \quad \binom{n}{k} \binom{r+s+t+2}{r+s+2} \frac{k^n \, a^{r+s+2}}{(r+s+t+1)\,(n+1)!\,(1+a)^{r+1}} \\ &\quad \times \sum_{p=0}^{\infty} \sum_{j=0}^{k-1} \sum_{q=0}^{j+n+p-1} \binom{q}{r} \binom{k-1}{j} (-1)^{j+q} \, (q+1)^s \, c_{n-1,p}. \end{aligned}$$

Equation (36) is the main result of this section. It reveals that the pdf of the OL-G k upper record values is a triple linear combination of Exp-G densities.

5. Estimation

We determine the maximum likelihood estimates (MLEs) of the parameters of the new family from complete samples only. Let $x_1, \ldots, x_n$ be observed values from the OL-G distribution with parameters a and $\boldsymbol{\xi}$. Let $\boldsymbol{\Theta} = (a, \boldsymbol{\xi})^\top$ be the $p \times 1$ parameter vector. The total log-likelihood function for $\boldsymbol{\Theta}$ is given by

$$\ell(\boldsymbol{\Theta}) \quad = \quad 2\,n \, \log(a) - n \, \log(1+a) + \sum_{i=1}^{n} \log[g(x_i; \boldsymbol{\xi})] - 3 \sum_{i=1}^{n} \log[\overline{G}(x_i; \boldsymbol{\xi})] - a \sum_{i=1}^{n} V(x_i; \boldsymbol{\xi}),$$

where $V(x; \boldsymbol{\xi}) = G(x; \boldsymbol{\xi}) / \overline{G}(x; \boldsymbol{\xi})$. We assume that the folllowing standard regularity conditions for the log-likelihood $\ell(\boldsymbol{\Theta})$ hold: i) The support of X associated to the distribution does not depend on unknown parameters; ii) The parameter space of X, say $\boldsymbol{\Psi}$ is open and $\ell(\boldsymbol{\Theta})$ has a global maximum in $\boldsymbol{\Psi}$; iii) For almost all x, the fourth-order log-likelihood derivatives with respect to the model parameters exist and are continuous in an open subset of $\boldsymbol{\Psi}$ that contains the true parameter; iv) The expected information matrix is positive definite and finite; v) The absolute values of the third-order log-likelihood derivatives with respect to the parameters are bounded by expected finite functions of X.

The components of the score function $U(\boldsymbol{\Theta}) = (U_a, U_{\boldsymbol{\xi}})^\top$ are

$$U_a = \frac{2\,n}{a} - \frac{n}{1+a} - \sum_{i=1}^{n} V(x_i;\boldsymbol{\xi})$$

and

$$U_{\boldsymbol{\xi}_k} = -a\sum_{i=1}^{n} \partial V(x_i;\boldsymbol{\xi})/\partial\boldsymbol{\xi}_k + \sum_{i=1}^{n}\frac{\partial g(x_i;\boldsymbol{\xi})/\partial\boldsymbol{\xi}_k}{g(x_i;\boldsymbol{\xi})} - 3\sum_{i=1}^{n}\frac{\partial\overline{G}(x_i;\boldsymbol{\xi})/\partial\boldsymbol{\xi}_k}{\overline{G}(x_i;\boldsymbol{\xi})}.$$

Setting U_a and $U_{\boldsymbol{\xi}}$ equal to zero and solving the equations simultaneously yields the MLE $\widehat{\boldsymbol{\Theta}} = (\widehat{a},\widehat{\boldsymbol{\xi}})^\top$ of $\boldsymbol{\Theta} = (a,\boldsymbol{\xi})^\top$. These equations cannot be solved analytically and statistical software can be used to solve them numerically using iterative methods such as the Newton-Raphson type algorithms.

For interval estimation on the model parameters, we obtain the $(p+1)\times(p+1)$ observed information matrix $J(\boldsymbol{\Theta}) = \{U_{rs}\}$ (for $r,s = a,\boldsymbol{\xi_k}$), whose elements are

$$U_{aa} = -\frac{2\,n}{a^2} + \frac{n}{(1+a)^2},\quad U_{a\boldsymbol{\xi}_k} = -\sum_{i=1}^{n}\partial V(x_i;\boldsymbol{\xi})/\partial\boldsymbol{\xi}_k$$

and

$$\begin{aligned}U_{\xi_k\xi_l} = &-a\sum_{i=1}^{n}\partial^2 V(x_i;\boldsymbol{\xi})/\partial\boldsymbol{\xi}_k\,\partial\boldsymbol{\xi}_l - \sum_{i=1}^{n}\frac{[\partial g(x_i;\boldsymbol{\xi})/\partial\boldsymbol{\xi}_k]\,[\partial g(x_i;\boldsymbol{\xi})/\partial\boldsymbol{\xi}_l]}{g(x_i;\boldsymbol{\xi})^2} + \sum_{i=1}^{n}\frac{\partial^2 g(x_i;\boldsymbol{\xi})/\partial\boldsymbol{\xi}_k\,\partial\boldsymbol{\xi}_l}{g(x_i;\boldsymbol{\xi})}\\ &+3\sum_{i=1}^{n}\frac{[\partial\overline{G}(x_i;\boldsymbol{\xi})/\partial\boldsymbol{\xi}_k]\,[\partial\overline{G}(x_i;\boldsymbol{\xi})/\partial\boldsymbol{\xi}_l]}{\overline{G}(x_i;\boldsymbol{\xi})^2} - 3\sum_{i=1}^{n}\frac{\partial^2\overline{G}(x_i;\boldsymbol{\xi})/\partial\boldsymbol{\xi}_k\,\partial\boldsymbol{\xi}_l}{\overline{G}(x_i;\boldsymbol{\xi})}.\end{aligned}$$

6. Empirical and numerical illustration

6.1. Some numerical values and simulation

The formulae derived in this paper can be easily handled in most symbolic computation software platforms such as MAPLE, MATLAB and MATHEMATICA. These platforms have currently the ability to deal with complex expressions. Table 2 provides some numerical values for the ordinary moments μ_i (with $i = 1,2,3,4.$) and quantiles $Q(u)$ of the OLW(a,α,λ) distribution calculated from MATHEMATICA. These moments were established using numerical integration. The quantile values were obtained by using Equation (14). The values of the Lambert W function is avaliable in routine `ProductLog[·, ·]`.

Table 2: The values of the first four moments and some quantiles of the OLW(a, α, λ) distribution for $\alpha = 1.0$, $\lambda = 4.0$ and different values of a

OLW for a:	0.01	0.2	1.0	1.5	2.0
μ_1	1.257	0.519	0.199	0.144	0.113
μ_2	1.620	0.303	0.053	0.030	0.019
μ_3	2.13	0.189	0.017	0.007	0.004
μ_4	2.850	0.125	0.005	0.002	<0.001
$Q(0.1)$	0.993	0.266	0.045	0.027	0.018
$Q(0.3)$	1.174	0.433	0.122	0.079	0.057
$Q(0.7)$	1.374	0.627	0.262	0.191	0.149
$Q(0.9)$	1.490	0.743	0.363	0.280	0.228
OLW for a:	0.3	0.5	0.7	0.9	1.1
μ_1	0.427	0.320	0.257	0.215	0.185
μ_2	0.212	0.126	0.085	0.062	0.047
μ_3	0.114	0.055	0.032	0.020	0.014
μ_4	0.066	0.026	0.013	0.007	0.004
$Q(0.2)$	0.276	0.178	0.127	0.096	0.076
$Q(0.5)$	0.439	0.323	0.255	0.208	0.175
$Q(0.6)$	0.483	0.366	0.294	0.244	0.209
$Q(0.8)$	0.579	0.457	0.381	0.327	0.286

We assess the performance of the MLEs of the OLW distribution with respect to sample size n. The assessment was based on a simulation study:

1. generate ten thousand samples of size n from (4)-(5). The inversion method was used to generate samples.

2. compute the MLEs for the ten thousand samples, say $\left(\widehat{a}, \widehat{\alpha}, \widehat{\lambda}\right)$ for $i = 1, 2, \ldots, 10000$.

3. compute the standard errors of the MLEs for the ten thousand samples, say $\left(s_{\widehat{a}}, s_{\widehat{\alpha}}, s_{\widehat{\lambda}}\right)$ for $i = 1, 2, \ldots, 10000$. The standard errors were computed by inverting the observed information matrices.

4. compute the biases and mean squared errors given by

$$\mathrm{bias}_{\epsilon}(n) = \frac{1}{10000} \sum_{i=1}^{10000} \left(\widehat{\epsilon}_i - \epsilon\right),$$

$$\mathrm{MSE}_{\epsilon}(n) = \frac{1}{10000} \sum_{i=1}^{10000} \left(\widehat{\epsilon}_i - \epsilon\right)^2,$$

   for $\epsilon = a, \alpha, \lambda$.

We repeated these steps for $n = 30, 31, \ldots, 100$ with $a = 1.2$, $\alpha = 2.5$ and $\lambda = 4.0$, so computing $\mathrm{bias}_{\epsilon}(n)$ and $\mathrm{MSE}_{\epsilon}(n)$.

Figure 5 shows how the four biases vary with respect to n. The biases for each parameter either decrease to zero as $n \to \infty$. The reported observations are for only one choice for (a, α, λ), namely that $(a, \alpha, \lambda) = (1.2, 2.5, 4.0)$. But the results were similar for a wide range of other choices for (a, α, λ). Figure 6 shows how the four mean squared errors vary with

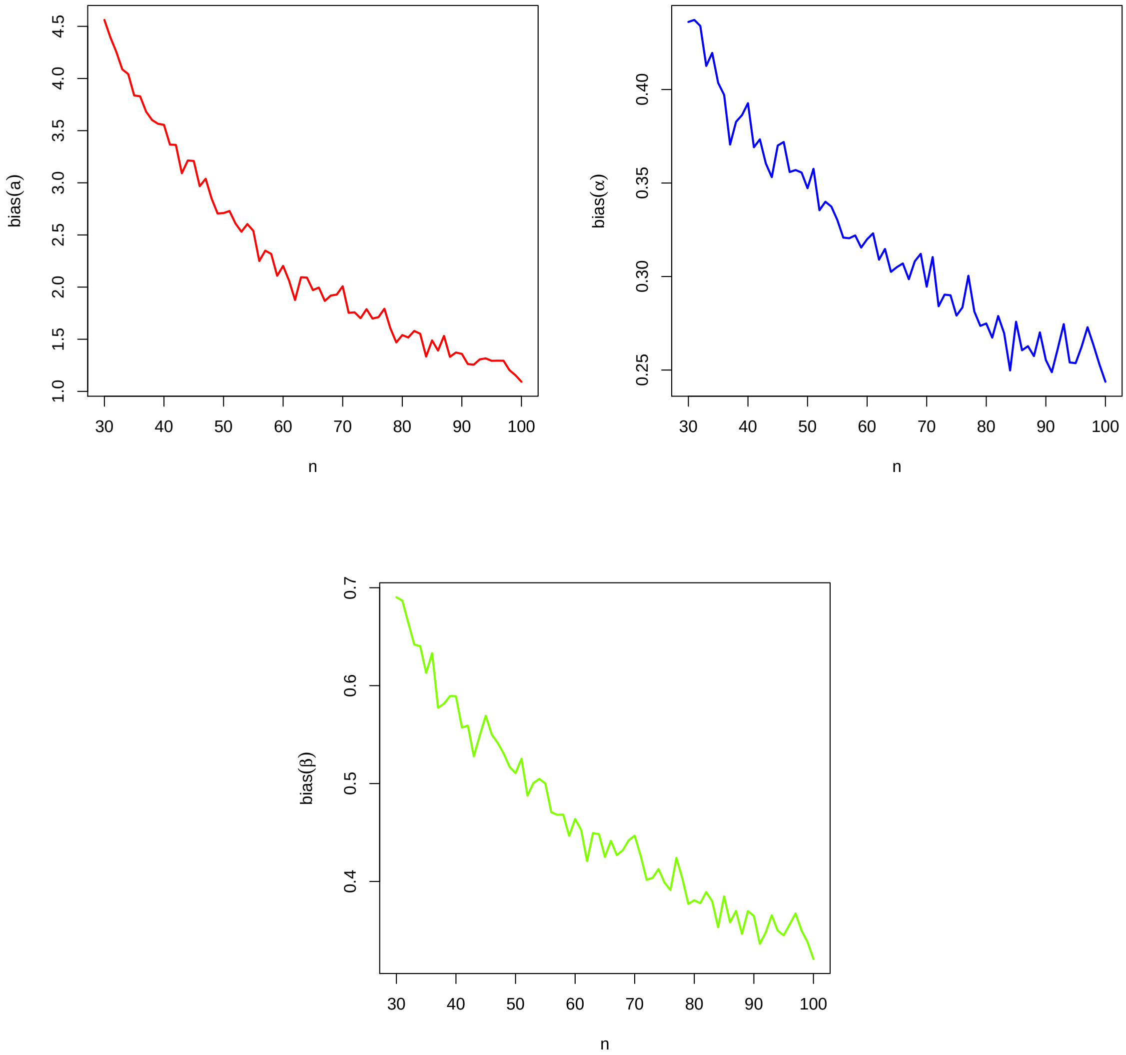


Figure 5: Biases of $\widehat{a}$, $\widehat{\alpha}$ and $\widehat{\lambda}$ versus n.

respect to n. The mean squared errors for each parameter decrease to zero as $n \to \infty$. In particular, i) the biases for each parameter either decreased to zero and appeared reasonably small at $n = 100$; ii) the mean squared errors for each parameter decreased to zero and appeared reasonably small at $n = 100$.

6.2. Application

We illustrate the flexibility of the OLW distribution by means of a real data set. Similar investigations could be performed for other OL-G distributions. We choose the Weibull as baseline because of its popularity. The computations are performed using the software R version 3.0.3 (package bbmle). The maximization follows the BFGS method with analytical derivatives. The algorithm used to estimate the model parameters converged for all current models.

The data set consists of 63 observations of the strengths of 1.5 cm glass fibres, originally obtained by workers at the UK National Physical Laboratory. Unfortunately, the units of measurement are not given in the paper. These data have also been analyzed by Bourguignon *et al.* (2014). For these data, we compare the fits of the OLW distribution defined by (6),

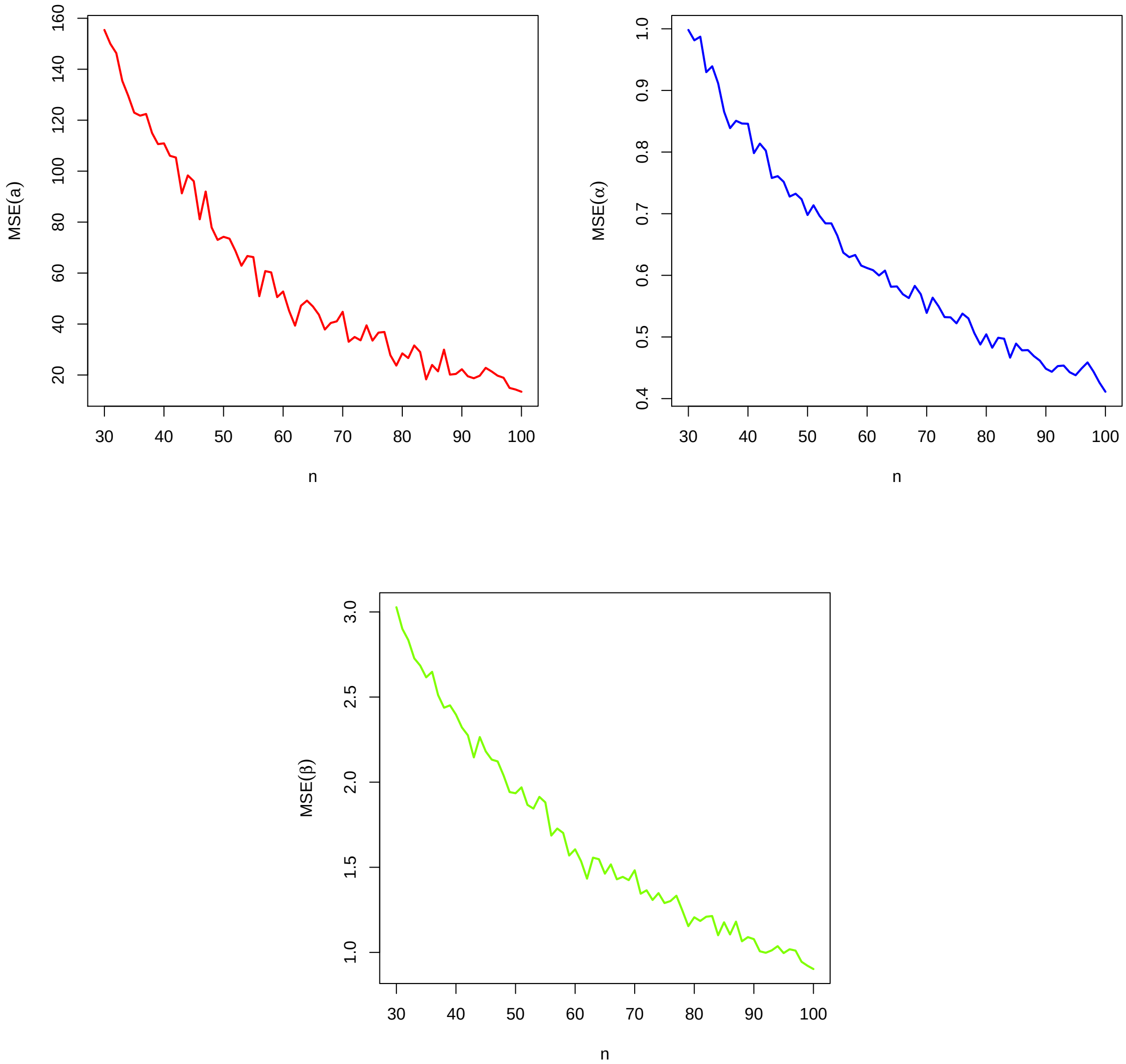


Figure 6: Mean squared errors of $\widehat{a}$, $\widehat{\alpha}$ and $\widehat{\lambda}$ versus n.

its special Weibull model (W) and of the following distributions: the exponentiated Weibull (EW) with pdf given by

$$f_{\mathrm{EW}}(x) = a\,\lambda\,\alpha^{\lambda}\,x^{\lambda-1}\,(1-\mathrm{e}^{-(\alpha\,x)^{\lambda}})^{a-1}\,\mathrm{e}^{-(\alpha\,x)^{\lambda}},$$

the beta Weibull (BW) with pdf given by

$$f_{\mathrm{BW}}(x) = \frac{1}{B(a,b)}\,\alpha^{\lambda}\,x^{\lambda-1}\,\mathrm{e}^{-(\alpha\,x)^{\lambda}\,b}\,(1-\mathrm{e}^{-(\alpha\,x)^{\lambda}})^{a-1}$$

and the Kumaraswamy Weibull (KwW) with pdf given by

$$f_{\mathrm{KwW}}(x) = a\,b\,\lambda\,\alpha^{\lambda}\,x^{\lambda-1}\,\mathrm{e}^{-(\alpha\,x)^{\lambda}}\,[1-\mathrm{e}^{-(\alpha\,x)^{\lambda}}]^{a-1}\,\left\{1-\left[1-\mathrm{e}^{-(\alpha\,x)^{\lambda}}\right]^{a}\right\}^{b-1}.$$

All parameters of these distribution are positive numbers. In Table 3, the MLEs and their standard errors (SEs) (in parentheses) of the parameters from the five fitted models and the values of the Akaike Information Criterion (AIC), Cramér-von Mises (W*) and Anderson-Darling (A*) goodness-of-fit statistics are presented. According to the lowest values of the AIC, W* and A* statistics, the OLW model could be chosen as the best model among the five fitted models.

Table 3: MLEs (SEs in parentheses) for some fitted models to the strengths data and the AIC, W* and A* values

Model	a	b	α	λ	AIC	W^*	A^*
W	-	-	5.781	1.628	34.414	0.237	1.304
	-	-	(0.576)	(0.037)			
EW	0.671	-	7.285	1.718	35.351	0.636	3.484
	(0.249)	-	(1.707)	(0.086)			
BW	0.620	10.249	7.759	2.382	37.179	0.196	1.089
	(0.248)	(95.117)	(2.023)	(2.897)			
KwW	0.606	0.214	6.908	1.337	35.252	0.161	0.908
	(0.162)	(0.029)	(0.004)	(0.003)			
OLW	0.049	-	1.102	0.492	34.387	0.153	0.870
	(0.087)	-	(0.527)	(0.494)			

The plots of the fitted OLW pdf and of the two better fitted pdfs are displayed in Figure 7. The QQ plots for the fitted models are displayed in Figure 8. These plots indicate that the OLW distribution provides a better fit to these data compared to the other models. Finally, the proposed distribution can be considered a very competitive model to the EW distribution.

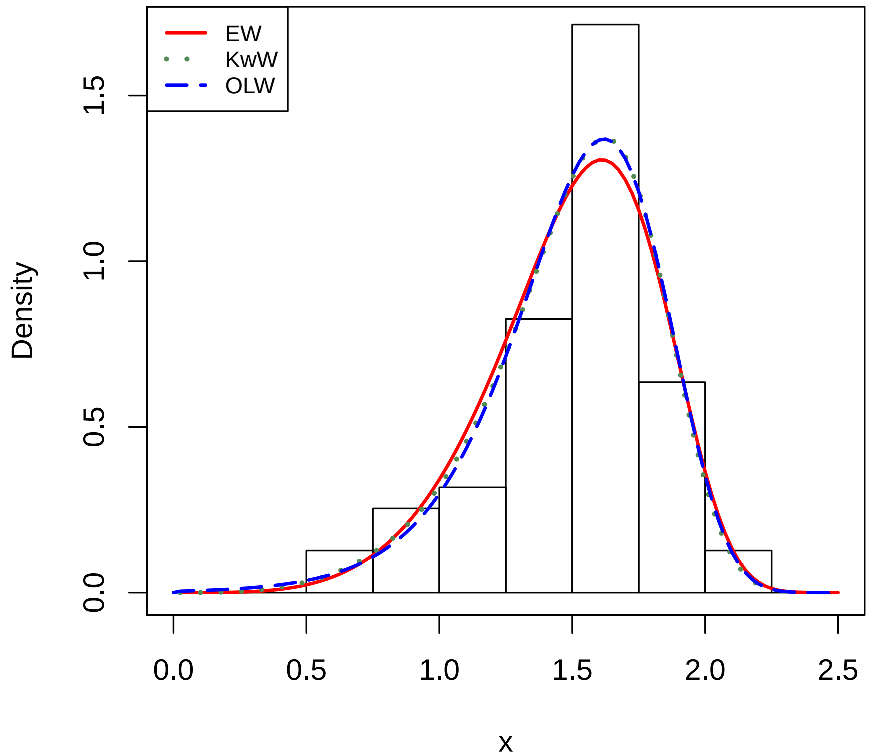


Figure 7: Fitted densities for the strengths data

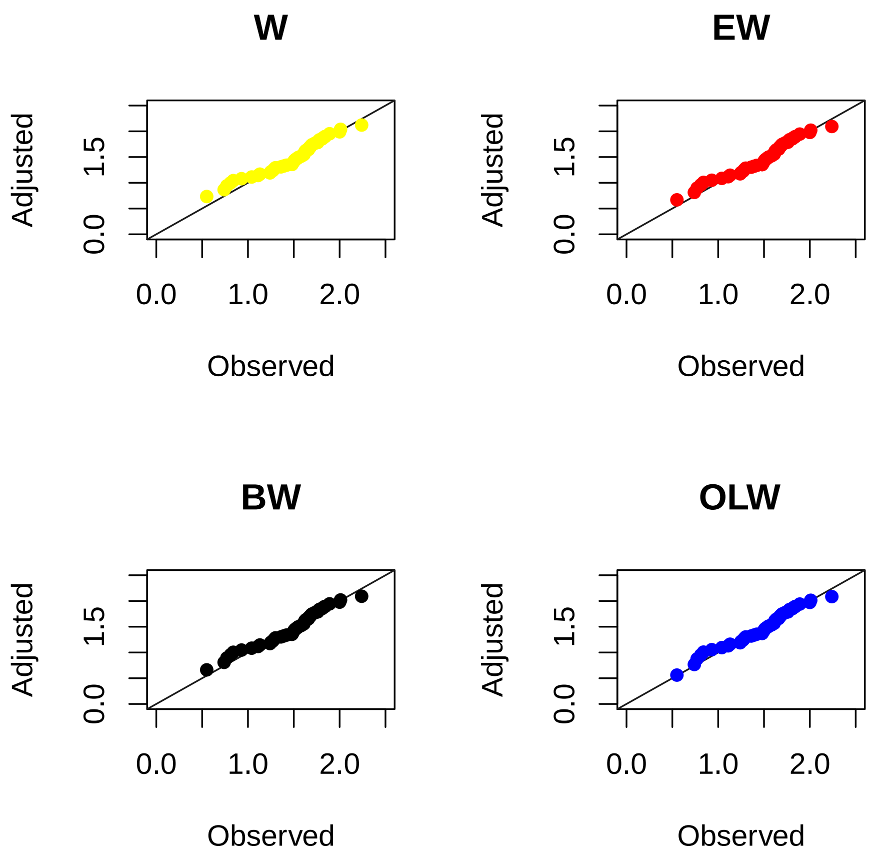


Figure 8: QQ plots for the strengths data

7. Conclusion remarks

In this paper we propose and study a new class of distributions called the *odd Lindley-G* family (OL-G). This family can extend several widely known models such as the Weibull, Kumaraswamy, half-logistic and Burr XII distributions in order to provide more flexibility. We investigate several of its structural properties such as an expansion for the density function and explicit expressions for the quantile function, ordinary and incomplete moments, generating function, Rényi entropy, reliability, order statistics and k upper record values. We estimate the parameters using maximum likelihood and determine the observed information matrix. We also discuss inference on the parameters based on Cramér-von Mises and Anderson-Darling statistics. An example to real data proves empirically the importance and potentiality of the proposed family.

References

Alexander C, Cordeiro GM, Ortega EMM, Sarabia JM (2012). "Generalized Beta-generated Distributions." *Computational Statistics and Data Analysis*, **56**, 1880–1897.

Alizadeh M, Emadi M, Doostparast M, Cordeiro GM, Ortega EMM, Pescim RR (2015). "A New Family of Distributions: The Kumaraswamy Odd Log-logistic, Properties and Applications." *Hacet. J. Math Stat*, **forthcoming**.

Alzaatreh A, Lee C, Famoye F (2013). "A New Method for Generating Families of Distributions." *Metron*, **71**, 63–79.

Alzaghal A, Felix F, Carl L (2013). "Exponentiated T-X Family of Distributions with Some Applications." *International Journal of Statistics and Probability*, **2**, 31–49.

Asgharzedah A, Bakouch HS, Esmaeli H (2013). "Pareto Poisson-Lindley Distribution with Applications." *Journal of Applied Statistics*, **40**, 1717–1734.

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

Chandler KN (1952). "The Distribution and Frequency of Record Values." *J. Roy. Statist. Soc., Ser B*, **14**, 220–228.

Cooray K (2006). "Generalization of the Weibull Distribution: The Odd Weibull Family." *Statistical Modelling*, **6**, 265–277.

Cordeiro GM, Alizadeh M, Ortega EMM (2014). "The Exponentiated Half-logistic Family of Distributions: Properties and Applications." *Journal of Probability and Statistics*, **2014**, 21 pages.

Cordeiro GM, de Castro M (2011). "A New Family of Generalized Distributions." *Journal of Statistical Computation and Simulation*, **81**, 883–893.

Cordeiro GM, Nadarajah S (2011). "Closed-form Expressions for Moments of a Class of Beta Generalized Distributions." *Brazilian Journal Probability and Statistics*, **52**, 14–33.

David HA, Nagaraja HN (2003). *Order Statistics*. John Wiley & Sons, New Jersey.

Dziubdziela W, Kopocinski B (1976). "Limiting Properties of the kâĹŠth Record Value." *Appl. Math.*, **15**, 187–190.

Eugene N, Lee C, Famoye F (2002). "Beta-normal Distribution and Its Applications." *Communications in Statistics Theory and Methods*, **31**, 497–512.

Exton H (1978). *Handbook of Hypergeometric Integrals: Theory, Applications, Tables, Computer Programs*. New York: Ellis Horwood, New York.

Ghitany ME, Atieh B, Nadarajah S (2008). "Lindley Distribution and Its Applications." *Mathematics and Computers in Simulation*, **78**, 493–506.

Gradshteyn IS, Ryzhik IM (2007). *Table of Integrals, Series, and Products*. Academic Press, New York.

Gupta P, Singh B (2012). "Parameter Estimation of Lindley Distribution with Hybrid Censored Data." *International Journal of System Assurance Engineering and Management*, **4**, 378–3751.

Gupta RC, Gupta RD, Gupta PL (1998). "Modeling Failure Time Data by Lehman Alternatives." *Communications in Statistics Theory and Methods*, **27**, 887–904.

Gupta RD, Kundu D (1999). "Generalized Exponential Distributions." *Australian and New Zealand Journal of Statistics*, **41**, 173–188.

Johnson NL, Kotz S, Balakrishnan N (1994). *Continuous Univariate Distributions*. John Wiley, New York.

Johnson NL, Kotz S, Balakrishnan N (1995). *Continuous Univariate Distributions*. John Wiley and Sons, New York.

Kotz S, Lai CD, Xie M (2003). "On the Effect of Redundancy for Systems with Dependent Components." *IIE Trans*, **35**, 1103–1110.

Lindley DV (1958). "Fiducial Distributions and Bayes' Theorem." *Journal of the Royal Statistical Society. Series B. (Methodological)*, **20**, 102–107.

Marshall AN, Olkin I (1997). "A New Method for Adding a Parameter to a Family of Distributions with Applications to the Exponential and Weibull Families." *Biometrika*, **84**, 641–552.

Mazucheli J, Achcar JA (2011). "The Lindley Distribution Applied to Competing Risks Lifetime Data." *Computer Methods and Programs in Biomedicine*, **104**, 188–192.

Mudholkar GS, Srivastava DK (1993). "Exponentiated Weibull Family for Analyzing Bathtub Failure-rate Data." *IEEE Transactions on Reliability*, **42**, 299–302.

Mudholkar GS, Srivastava DK, Freimer M (1995). "The Exponentiated Weibull Family: A Reanalysis of the Bus-motor-failure Data." *Technometrics*, **37**, 436–445.

Nadarajah Sand Bakouch HS, Tahmasbi R (2011). "A Generalized Lindley Distribution." *Sankhya B*, **73**, 331–359.

Nadarajah S (2005). "The Exponentiated Gumbel Distribution with Climate Application." *Environmetrics*, **17**, 13–23.

Nadarajah S, Gupta AK (2007). "The Exponentiated Gamma Distribution with Application to Drought Data." *Calcutta Statistical Association Bulletin*, **59**, 29–54.

Nadarajah S, Kotz S (2006). "The Exponentiated-type Distributions." *Acta Applicandae Mathematicae*, **92**, 97–111.

Rényi A (1961). *On Measures of Entropy and Information. In: Proceedings of the 4th Berkeley Symposium on Mathematical, Statistics and Probability.* University of California Press, Berkeley.

Sankaran M (1970). "The Discrete Poisson-Lindley Distribution." *Biometrics*, **26**, 145–149.

Shannon CE (1951). "Prediction and Entropy of Printed English." *The Bell System Technical Journal*, **30**, 50–64.

Shirke DT, Kakade CS (2006). "On Exponentiated Lognormal Distribution." *International Journal of Agricultural and Statistical Sciences*, **2**, 319–326.

Tahir MH, Cordeiro GM, Alzaatreh A, Mansoor M, Zubair M (2016a). "The Logistic-X Family of Distributions and Its Applications." *Commun. Stat. Theory Methods*, **(forthcoming)**.

Tahir MH, Zubair M, Mansoor M, Cordeiro GM, Alizadeh M, Hamedani GG (2016b). "A New Weibull-G Family of Distributions." *Hacet. J. Math. Stat.*, **(forthcoming)**.

Trott M (2006). *The Mathematica Guidebook for Symbolics.* Springer, New York.

Warahena-Liyanage G, Pararai M (2014). "A Generalized Power Lindley Distribution with Applications." *Asian Journal of Mathematics and Applications*, pp. 1–23.

Zakerzadeh Y, Dolati A (2009). "Generalized Lindley Distribution." *Journal of Mathematical Extension*, **3**, 13–25.

Zimmer WJ, Keats JB, Wang FK (1998). "The Burr XII Distribution in Reliability Analysis." *J. Qual. Technol*, **30**, 386–394.

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

Affiliation:

Frank Gomes-Silva
Department of Statistics and Informatics
Federal Rural University of Pernambuco
52171-900, Recife, PE, Brazil
E-mail: franksinatrags@gmail.com

Ana Percontini
Department of Exact Sciences
State University of Feira de Santana
44036-900, Feira de Santana, BA, Brazil
E-mail: anappaixao@gmail.com

Edleide de Brito
Department of Statistics
Federal University of Bahia
40110-909, Salvador, BA, Brazil
E-mail: edbnet@gmail.com

Manoel W. Ramos
Department of Mathematics
Federal Institute of Paraíba
58015-020, João Pessoa, PB, Brazil
E-mail: wallace.ifpb@gmail.com

Ronaldo Venâncio
Recife Military School
52070-080, Recife, PE, Brazil
E-mail: ronaldovenanciorvs@gmail.com

Gauss M. Cordeiro
Department of Statistics
Federal University of Pernambuco
50740-540, Recife, PE, Brazil
E-mail: gausscordeiro@gmail.com

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 46 *Submitted:* 2016-01-16
February 2017 *Accepted:* 2016-09-01

Austrian Journal of Statistics
February 2017, Volume 46, 89–90.
http://www.ajs.or.at/
doi:10.17713/ajs.v46i1.604

Reviewer: Andreas Quatember (JKU Linz)

Statistik für alle - Die 101 wichtigsten Begriffe anschaulich erklärt.

Walter Krämer
Springer Spektrum, Berlin, Germany, 2015.
ISBN 978-3-662-45030-7. 89–90 pp. EUR 15.41.
http://www.springer.com/de/book/9783662450307

Zur Vorbeugung von Missverständnissen gleich mal vorne weg: 1. Ja, Walter Krämer hat mich mit seinem in den 90er Jahren erstmals erschienen Buch „So lügt man mit Statistik“ nachhaltig in der Lehre inspiriert; 2. Sein Zugang, die statistische Methodik möglichst anschaulich erklären zu wollen, viel bei mir auf fruchtbaren Boden; und schließlich 3. Ich traue mir dennoch zu, sein neues Buch „Statistik für alle - Die 101 wichtigsten Begriffe anschaulich erklärt“, objektiv zu rezensieren.

Die 101 wichtigsten Begriffe der Statistik also! - Die Auswahl dieser Schlagworte erfolgte natürlich nach der subjektiven Einschätzung des Autors und darüber ließe sich schon mal treffend streiten: Warum sind etwa die Begriffe „Benford-Gesetz“, „Biometrischer Fingerabdruck“, „Chaos“ oder „Itemanalyse“ Bestandteil dieser Auswahl, nicht aber „Inflationsrate“, „p-Wert“, „Repräsentativität“ oder „Stichprobenumfang“? - Weil es die Auswahl des Autors ist, möchte man antworten, und als solche ist sie natürlich die richtige.

Die Begriffe sind wie in einem Wörterbuch alphabetisch geordnet von „Achsenmanipulation“ bis „Zufallszahlen“. Es querzulesen bzw. gezielt nachzuschlagen erschiene mir deshalb eigentlich als geeigneter Zugang zur Lektüre. Dabei muss festgehalten werden, dass zu diesem Zweck so manche außerhalb des Faches eher nicht so bekannten Begriffe wie beispielsweise „arithmetisches“, „geometrisches“ oder „harmonisches“ Mittel vielleicht doch besser in einem Oberbegriff wie „Mittelwert“ zusammengefasst worden wären, damit man schließlich auch auf diese Fachbegriffe, quasi en passant, stoßen kann, ohne bewusst danach suchen zu müssen.

Mein Zugang als Rezensent war es dennoch, es von vorne nach hinten zu lesen. Krämer schreibt gewohnt sachlich korrekt, dabei aber um allgemeine Verständlichkeit bemüht. Ein Beispiel dafür ist der Eintrag zur „Binomialverteilung“. Diese motiviert Krämer an Hand der Frage: „An wieviel Börsentagen in der Woche schließt der DAX im Plus? … Die Binomialverteilung sagt: wir können uns dieses Auszählen sparen.“ Und er merkt dabei auch noch an, dass die von ihm mit 1/2 angesetzte Ereigniswahrscheinlichkeit „in Wahrheit ein kleines bisschen größer“ als 1/2 ist. Ab und an wird in den Einträgen pointiert formuliert. Ein Beispiel dafür findet sich bei den Bevölkerungspyramiden unter dem Schlagwort „Histogramm“: „Man sieht sehr schön die jetzt 45-55-jährigen Babyboomer der Geburtsjahrgänge 1958-1967, die gerade in vollen Zügen ihre historische Ausnahmesituation geniessen: wenige Kinder, die man unterhalten muss, und sehr viele Geschwister, mit denen man sich den Unterhalt der Eltern teilt.“ Zuweilen sind die Einträge auch polemisch gewürzt. Unter dem Terminus „Arbeitslosenquote“ folgert Krämer beispielsweise: „Demnach sind Teilnehmer von Schul- und Umschulungsmaßnahmen, da nicht unmittelbar dem Arbeitsmarkt zur Verfügung stehend, offiziell nicht arbeitslos. … Unter anderem auch deshalb sind die jeweils regierenden und an niedrigen Arbeitslosenzahlen interessierenden Kreise so große Freunde von Langzeitstudenten und Umschulungsprogrammen aller Art.“

Das Buch wendet sich meines Erachtens vordergründig nicht an professionelle Statistiker, wenngleich auch solche durchaus den ein oder anderen interessanten Abschnitt darin finden werden. Also gab ich es nach meiner Lektüre kurzerhand an meine Frau weiter - eine interessierte, beruflich sich immer wieder mit Statistiken auseinandersetzen müssende Nichtstatistikerin. Und siehe da: Ich sah sie von Zeit zu Zeit schmunzeln, zustimmend nicken und das ein oder andere Mal diskutierte sie das eben Nachgeschlagene mit mir.

Meine Schlussfolgerung aus diesen Erfahrungen lautet: Das Lesen dieses statistischen Wörter-Büchleins wird jenen am meisten Freude bereiten und Nutzen bringen, die sich, selbst Nichtstatistiker, aus unterschiedlichen Gründen - sei es beruflich oder im Studium - mit dem Fach auseinandersetzen müssen. Für diese Leserschaft ist es absolut lesenswert, lehrreich und auch launig geschrieben - ein Krämer eben! Ob die Statistik und ihre Anwendung in unserem Leben in 101 Stichwörtern tatsächlich nur Walter Krämer kurz, prägnant und verständlich erklären kann, wie es der Verlag dem Taschenbuch auf den Rücken gedruckt hat, möchte ich dahingestellt lassen.

Reviewer:

Andreas Quatember
IFAS - Institut für Angewandte Statistik
Johannes Kepler Universität Linz
Österreich E-mail: andreas.quatember@jku.at

ISSN 1026–597X

Contents