Calidad de Revistas Científicas Españolas
FECYT | FUNDACIÓN ESPAÑOLA PARA LA CIENCIA Y LA TECNOLOGÍA | 2013

ISSN: 1696-2281
eISSN: 2013-8830
SORT 39 (1) January-June 2015

SORT

Statistics and Operations Research Transactions

Coediting institutions

Universitat Politècnica de Catalunya
Universitat de Barcelona
Universitat de Girona
Universitat Autònoma de Barcelona
Universitat Pompeu Fabra
Universitat de Lleida
Institut d'Estadística de Catalunya

Supporting institutions

Spanish Region of the International Biometric Society
Societat Catalana d'Estadística

Generalitat de Catalunya
Institut d'Estadística de Catalunya

SORT

Volume 39
Number 1
January-June 2015
ISSN: 1696-2281
eISSN: 2013-8830

Articles

SORT 39 (1) January-June 2015, 3-18

Inference on the parameters of the Weibull distribution using records

Ali Akbar Jafari*,[1] and Hojatollah Zakerzadeh[1]

Abstract

The Weibull distribution is a very applicable model for lifetime data. In this paper, we have investigated inference on the parameters of Weibull distribution based on record values. We first propose a simple and exact test and a confidence interval for the shape parameter. Then, in addition to a generalized confidence interval, a generalized test variable is derived for the scale parameter when the shape parameter is unknown. The paper presents a simple and exact joint confidence region as well. In all cases, simulation studies show that the proposed approaches are more satisfactory and reliable than previous methods. All proposed approaches are illustrated using a real example.

MSC: 62F03, 62E15, 62-04.

Keywords: Coverage probability, generalized confidence interval, generalized p-value, records, Weibull distribution.

1. Introduction

The Weibull distribution is a well-known distribution that is widely used for lifetime models. It has numerous varieties of shapes and demonstrates considerable flexibility that enables it to have increasing and decreasing failure rates. Therefore, it is used for many applications, for example in hydrology, industrial engineering, weather forecasting and insurance. The Weibull distribution with parameters α and β, denoted by $W(\alpha,\beta)$, has a cumulative distribution function (cdf)

$$F(x) = 1 - e^{-\left(\frac{x}{\alpha}\right)^{\beta}}, \quad x > 0, \quad \alpha > 0, \quad \beta > 0,$$

* Corresponding author: aajafari@yazd.ac.ir

[1] Department of Statistics, Yazd University, Yazd, Iran.

Received: September 2013
Accepted: January 2015

and probability density function (pdf)

$$f(x) = \frac{\beta}{\alpha^{\beta}} x^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^{\beta}}, \quad x > 0.$$

The Weibull distribution is a generalization of the exponential distribution and Rayleigh distribution. Also, $Y = \log(X)$ has the Gumbel distribution with parameters $b = \frac{1}{\beta}$ and $a = \log(\alpha)$, when X has a Weibull distribution with parameters α and β.

Let $X_1, X_2, \ldots$ be an infinite sequence of independent identically distributed random variables from a same population with the cdf F_θ, where θ is a parameter. An observation X_j will be called an upper record value (or simply a record) if its value exceeds that of all previous observations. Thus, X_j is a record if $X_j > X_i$ for every $i < j$. An analogous definition deals with lower record values. The record value sequence $\{R_n\}$ is defined by

$$R_n = X_{T_n}, \quad n = 0, 1, 2, \ldots.$$

where T_n is called the record time of nth record and is defined as $T_n = \min\{j : X_j > X_{T_{n-1}}\}$ with $T_0 = 1$.

Let $R_0, \ldots, R_n$ be the first $n+1$ upper record values from the cdf F_θ and the pdf f_θ. Then, the joint distribution of the first $n+1$ record values is given by

$$f_{\boldsymbol{R}}(\boldsymbol{r}) = f_\theta(r_n) \prod_{i=0}^{n-1} \frac{f_\theta(r_i)}{1 - F_\theta(r_i)}, \quad r_0 < r_1 < \cdots < r_n, \tag{1.1}$$

where $\boldsymbol{r} = (r_0, r_1, \ldots, r_n)$ and $\boldsymbol{R} = (R_0, R_1, \ldots, R_n)$ (for more details see Arnold et al., 1998).

Chandler (1952) launched a statistical study of the record values, record times and inter-record times. Record values and the associated statistics are of interest and importance in the areas of meteorology, sports and economics. Ahsanullah (1995) and Arnold et al. (1998) are two good references about records and their properties.

Some papers considered inference on the Weibull distribution based on record values: Dallas (1982) discussed some distributional results based on upper record values. Balakrishnan and Chan (1994) established some simple recurrence relations satisfied by the single and the product moments, and derived the BLUE of the scale parameter when the shape parameter is known. Chan (1998) provided a conditional method to derive exact intervals for location and scale parameters of location-scale family that can be used to derive exact intervals for the shape parameter. Wu and Tseng (2006) provided some pivotal quantities to test and establish confidence interval of the shape parameter based on the first $n+1$ observed upper record values. Soliman et al. (2006) derived the Bayes estimates based on record values for the parameters with respect to squared error loss function and LINEX loss function. Asgharzadeh and Abdi (2011b) proposed joint confidence regions for the parameters. Teimouri and Gupta (2012) computed

the coefficient of skewness of upper/lower record statistics. Teimouri and Nadarajah (2013) derived exact expressions for constructing bias corrected maximum likelihood estimators (MLEs) of the parameters for the Weibull distribution based on upper records. Gouet et al. (2012) obtained the asymptotic properties for the counting process of δ-records among the first n observations.

In this paper, we consider inference about the parameters of Weibull distribution based on record values. First, we will propose a simple and exact method for constructing confidence interval and testing the hypotheses about the shape parameter β. Then using the concepts of generalized p-value and generalized confidence interval, a generalized approach for inference about the scale parameter α will be derived. Tsui and Weerahandi (1989) introduced the concept of generalized p-value, and Weerahandi (1993) introduced the concept of generalized confidence interval. These approaches have been used successfully to address several complex problems (see Weerahandi, 1995) such as confidence interval for the common mean of several log-normal distributions (Behboodian and Jafari, 2006), confidence interval for the mean of Weibull distribution (Krishnamoorthy et al., 2009), inference about the stress-strength reliability involving two independent Weibull distributions (Krishnamoorthy and Lin, 2010), and comparing two dependent generalized variances (Jafari, 2012).

We also present an exact joint confidence region for the parameters. Our simulation studies show that the area of our joint confidence region is smaller than those provided by other existing methods.

The rest of this article is organized as follows: A simple method for inference about shape parameter and a generalized approach for inference about the scale parameter are proposed in Section 2. Furthermore, a simulation study is performed and a real example is proposed in this Section. We also present a joint confidence region for the parameters α and β in Section 3.

2. Inference on the parameters

Suppose $R_0, R_1, \ldots, R_n$ are the first $n+1$ upper record values from a Weibull distribution with parameters α and β. In this section, we consider inference on the parameters α and β. From (1.1), the joint distribution of these record values can be written as

$$f_{\boldsymbol{R}}(\boldsymbol{r}) = \frac{\beta^{n+1}}{\alpha^{\beta(n+1)}} e^{-\left(\frac{r_n}{\alpha}\right)^{\beta}} \prod_{i=0}^{n} r_i^{\beta-1} \qquad 0 < r_0 < r_1 < \cdots < r_n. \tag{2.1}$$

Therefore, $(R_n, \sum_{i=0}^{n} \log(R_i))$ are sufficient statistics for (α, β). Moreover, it can be easily shown that the MLE's of the parameters α and β are

$$\hat{\beta} = \frac{n+1}{\sum_{i=0}^{n} \log\left(\frac{R_n}{R_i}\right)}, \qquad \hat{\alpha} = \frac{R_n}{(n+1)^{\frac{1}{\hat{\beta}}}}. \tag{2.2}$$

Theorem 2.1 *Let $R_0, R_1, \ldots, R_n$ be the first $n+1$ upper record values from a Weibull distribution. Then*

i. $U = 2\beta \sum_{i=0}^{n} \log\left(\frac{R_n}{R_i}\right)$ *has a chi-square distribution with $2n$ degrees of freedom.*

ii. $V = 2\left(\frac{R_n}{\alpha}\right)^{\beta}$ *has a chi-square distribution with $2n+2$ degrees of freedom.*

iii. U and V are independent.

Proof. i. Define

$$Q_m = \frac{R_m}{R_{m-1}}, \quad m = 1, 2, \ldots, n. \tag{2.3}$$

From Arnold et al. (1998) page 20, Q_m's are independent random variables with

$$P(Q_m > q) = q^{-\beta m}, \quad q > 1,$$

and

$$2\beta m \log(Q_m) = 2\beta m \log\left(\frac{R_m}{R_{m-1}}\right) \sim \chi^2_{(2)}.$$

Therefore,

$$U = 2\beta \sum_{i=0}^{n} \log(\frac{R_n}{R_i}) = 2\beta \sum_{i=0}^{n-1} \log\left(\frac{R_n}{R_{n-1}} . \frac{R_{n-1}}{R_{n-2}} \cdots \frac{R_{i+1}}{R_i}\right)$$

$$= 2\beta \sum_{i=0}^{n-1} \sum_{m=i+1}^{n} \log\left(\frac{R_m}{R_{m-1}}\right) = 2\beta \sum_{m=1}^{n} \sum_{i=0}^{m-1} \log(Q_m) = \sum_{m=1}^{n} 2\beta m \log(Q_m),$$

has a chi-square distribution with $2n$ degrees of freedom.

ii. Define

$$Y = \left(\frac{X}{\alpha}\right)^{\beta},$$

where X has a Weibull distribution with parameters α and β. Then, Y has an exponential distribution with parameter one. Therefore, we can conclude that V has a chi-square distribution with $2n+2$ degrees of freedom (see Arnold et al., 1998, page 9).

iii. Let β be known. Then, it can be concluded from (2.1) that R_n is a complete sufficient statistic for α. Also, Q_m's in (2.3) are ancillary statistics. Therefore, R_n and Q_m's are independent, and the proof is completed. ■

2.1. Inference on the shape parameter

Here, we consider inference on the shape parameter, β from a Weibull distribution based on record values, and propose a simple and an exact method for constructing a confidence interval and testing the one-sided hypotheses

$$H_0 : \beta \leq \beta_0 \quad vs. \quad H_1 : \beta > \beta_0, \tag{2.4}$$

and the two-sided hypotheses

$$H_0 : \beta = \beta_0 \quad vs. \quad H_1 : \beta \neq \beta_0, \tag{2.5}$$

where β_0 is a specified value.

Based on Theorem 2.1, $U = 2\beta \sum_{i=0}^{n} \log\left(\frac{R_n}{R_i}\right)$ has a chi-square distribution with $2n$ degrees of freedom. Therefore, a $100\,(1-\gamma)\,\%$ confidence interval for β can be obtained as

$$\left(\frac{\chi^2_{(2n),\gamma/2}}{2\sum_{i=0}^{n} \log\left(\frac{R_n}{R_i}\right)}, \frac{\chi^2_{(2n),1-\gamma/2}}{2\sum_{i=0}^{n} \log\left(\frac{R_n}{R_i}\right)} \right), \tag{2.6}$$

where $\chi^2_{(k),\gamma}$ is the γth percentile of the chi-square distribution with k degrees of freedom. Also, for testing the hypotheses in (2.4) and (2.5), we can define the test statistic

$$U_0 = 2\beta_0 \sum_{i=0}^{n} \log\left(\frac{R_n}{R_i}\right),$$

and the null hypothesis in (2.4) is rejected at nominal level γ if

$$U_0 > \chi^2_{(2n),1-\gamma},$$

and the null hypothesis in (2.5) is rejected if

$$U_0 < \chi^2_{(2n),\gamma/2} \quad \text{or} \quad U_0 > \chi^2_{(2n),1-\gamma/2}.$$

Wu and Tseng (2006) proposed the random variable

$$W(\beta) = \frac{\sum_{i=0}^{n} R_i^{\beta}}{(n+1)(\prod_{i=0}^{n} R_i)^{\frac{\beta}{n+1}}},$$

for inference about the shape parameter, and showed that $W(\beta)$ is an increasing function with respect to β. Also, its distribution does not depend on the parameters α and β. In fact, $W(\beta)$ is distributed as

$$W^* = \frac{\sum_{i=0}^{n} R_i^*}{(n+1)(\prod_{i=0}^{n} R_i^*)^{\frac{1}{n+1}}},$$

where R_i^* is the ith record from the exponential distribution with parameter one. However, its exact distribution is very complicated, and Wu and Tseng (2006) obtained the percentiles of $W(\beta)$ using Monte Carlo simulation. The confidence limits for β are obtained by solving the following equations numerically as

$$W(\beta) = W^*_{1-\gamma/2}, \qquad W(\beta) = W^*_{\gamma/2}, \tag{2.7}$$

where W^*_δ is the δth percentile of the distribution of W^*.

2.2. Inference on the scale parameter

Here, we consider inference about the scale parameter, α for a Weibull distribution based on record values, and propose an approach for constructing a confidence interval and testing the one-sided hypotheses

$$H_0 : \alpha \leq \alpha_0 \quad \text{vs.} \quad H_1 : \alpha > \alpha_0, \tag{2.8}$$

and the two-sided hypotheses

$$H_0 : \alpha = \alpha_0 \quad \text{vs.} \quad H_1 : \alpha \neq \alpha_0, \tag{2.9}$$

where α_0 is a specified value.

We did not find any approach in literature for inference about α based on record values when the shape parameter is unknown. Here, we use the concepts of generalized p-value and generalized confidence interval introduced by Tsui and Weerahandi (1989), and Weerahandi (1993), respectively. In the Appendix, we briefly review these concepts, and refer readers to Weerahandi (1995) for more details.

Let

$$T = r_n \left(\frac{2}{V}\right)^{\frac{2C_r}{U}} = r_n \left(\frac{\alpha}{R_n}\right)^{\frac{C_r}{\sum_{i=0}^{n} \log(\frac{R_n}{R_i})}}, \tag{2.10}$$

where $C_r = \sum_{i=0}^{n} \log\left(\frac{r_n}{r_i}\right)$, and r_i, $i = 0, 1, \ldots, n$ is the observed value of R_i, $i = 0, 1, \ldots, n$, and U and V are independent random variables that are defined in Theorem 2.1. The

observed value of T is α, and distribution of T does not depend on unknown parameters α and β. Therefore, T is a generalized pivotal variable for α, and can be used to construct a generalized confidence interval for α.

Let

$$T^* = T - \alpha = r_n \left(\frac{2}{V}\right)^{\frac{2C_r}{U}} - \alpha.$$

Then, T^* is a generalized test variable for α, because i) the observed value of T^* does not depend on any parameters, ii) the distribution function of T^* is free from nuisance parameters and only depends on the parameter α, and iii) the distribution function of T^* is an increasing function with respect to the parameter α, and so, the distribution of T^* is stochastically decreasing in α. Therefore, the generalized p-value for testing the hypotheses in (2.8) is given as

$$p = P\left(T^* < 0 | H_0\right) = P\left(T < \alpha_0\right), \tag{2.11}$$

and the generalized p-value for testing the hypotheses in (2.9) is given as

$$p = 2 \min\left\{P\left(T > \alpha_0\right), P\left(T < \alpha_0\right)\right\}. \tag{2.12}$$

The generalized confidence interval for α based on T, and the generalized p-values in (2.11) and (2.12) can be computed using Monte Carlo simulation (Weerahandi, 1995; Behboodian and Jafari, 2006) based on the following algorithm:

Algorithm 2.1 *For given $r_0, r_1, \ldots, r_n$,*

1. *Generate $U \sim \chi^2_{(2n)}$ and $V \sim \chi^2_{(2n+2)}$.*
2. *Compute T in* (2.10).
3. *Repeat steps 1 and 2 for a large number times, (say $M = 10000$), and obtain the values $T_1, \ldots, T_M$.*
4. *Set $D_l = 1$ if $T_l < \alpha_0$ else $D_l = 0$, $l = 1, ..., M$.*

The $100(1-\gamma)\%$ generalized confidence interval for α is $\left[T_{(\gamma/2)}, T_{(1-\gamma/2)}\right]$, where $T_{(\delta)}$ is the δth percentile of T_l's. Also, the generalized p-value for testing the one-sided hypotheses in (2.11) *is obtained by $\frac{1}{M}\sum_{l=1}^{M} D_l$.*

2.3. Real example

Roberts (1979) gave monthly and annual maximal of one-hour mean concentration of sulfur dioxide (in pphm, parts per hundred million) from Long Beach, California, for 1956 to 1974. Chan (1998) showed that the Weibull distribution is a reasonable model

for this data set. Wu and Tseng (2006) also study this data set. The upper record values for the month of October from the data are

$$26, 27, 40, 41.$$

The 95% confidence interval for the scale parameter α based on our generalized confidence interval with $M = 10000$ is obtained as $(5.4869, 39.9734)$. The 95% confidence interval for the shape parameter β in (2.6) is obtained as $(0.6890, 8.0462)$, and based on Wu and Tseng's method in (2.7) is obtained as $(0.6352, 7.7423)$. Also, the generalized p-value is equal to 0.0227 for testing the hypotheses in (2.8) with $\alpha_0 = 5$. Therefore, the null hypothesis is rejected.

2.4. Simulation study

We performed a simulation study in order to evaluate the accuracy of the proposed methods for constructing confidence interval for the parameters of Weibull distribution. For this purpose, we generated $n+1$ record values from a Weibull distribution, and considered $\alpha = 1, 2$. For the simulation with 10000 runs and different values of the shape parameter β, the empirical coverage probabilities and expected lengths of the methods with the confidence coefficient 0.95 were obtained. The results of our generalized confidence interval for inference on α using the algorithm 2.1 with $M = 10000$ are presented in Table 1, and the results of our exact method (E) and the Wu method (W) for inference on β are given in Table 2. We can conclude that

Table 1: *Empirical coverage probabilities and expected lengths of the generalized confidence interval for the parameter α with confidence level 0.95.*

			β						
	α	n	0.5	1.0	1.2	1.5	2.0	3.0	5.0
Empirical Coverage	1.0	3	0.951	0.949	0.953	0.947	0.947	0.948	0.948
		7	0.952	0.949	0.950	0.950	0.951	0.953	0.952
		9	0.951	0.948	0.953	0.951	0.948	0.949	0.950
		14	0.945	0.949	0.950	0.950	0.954	0.952	0.952
	2.0	3	0.949	0.952	0.947	0.949	0.951	0.950	0.953
		7	0.948	0.953	0.950	0.946	0.954	0.948	0.951
		9	0.952	0.948	0.953	0.950	0.952	0.953	0.954
		14	0.950	0.946	0.949	0.951	0.951	0.952	0.955
Expected Length	1.0	3	16.740	3.581	2.804	2.155	1.653	1.211	0.847
		7	13.575	3.198	2.477	1.942	1.475	1.041	0.686
		9	13.505	3.138	2.469	1.918	1.446	1.008	0.651
		14	13.122	3.082	2.403	1.854	1.376	0.943	0.596
	2.0	3	33.516	7.187	5.579	4.341	3.323	2.427	1.704
		7	27.960	6.344	4.999	3.899	2.960	2.080	1.364
		9	27.342	6.304	4.935	3.831	2.890	2.016	1.302
		14	26.626	6.129	4.779	3.705	2.757	1.886	1.191

Table 2: *Empirical coverage probabilities and expected lengths of the methods for constructing confidence interval for the parameter β with confidence level 0.95.*

				β						
	α	n	Method	0.5	1.0	1.2	1.5	2.0	3.0	5.0
Empirical Coverage	1.0	3	W	0.950	0.952	0.952	0.949	0.949	0.953	0.947
			E	0.949	0.953	0.953	0.950	0.948	0.953	0.946
		7	W	0.951	0.949	0.950	0.948	0.949	0.949	0.953
			E	0.951	0.950	0.948	0.950	0.953	0.953	0.950
		9	W	0.949	0.949	0.945	0.949	0.947	0.949	0.950
			E	0.948	0.950	0.948	0.948	0.949	0.952	0.949
		14	W	0.946	0.949	0.951	0.951	0.953	0.949	0.951
			E	0.947	0.948	0.950	0.951	0.953	0.952	0.952
	2.0	3	W	0.954	0.955	0.948	0.950	0.950	0.952	0.949
			E	0.953	0.953	0.947	0.949	0.949	0.952	0.950
		7	W	0.950	0.955	0.947	0.948	0.952	0.947	0.948
			E	0.949	0.952	0.951	0.948	0.952	0.947	0.950
		9	W	0.953	0.948	0.956	0.950	0.953	0.951	0.952
			E	0.952	0.948	0.951	0.951	0.953	0.951	0.953
		14	W	0.948	0.947	0.949	0.950	0.951	0.951	0.952
			E	0.950	0.947	0.949	0.949	0.953	0.951	0.955
Expected Length	1.0	3	W	1.704	3.431	4.194	5.279	6.913	10.396	17.479
			E	1.630	3.285	4.013	5.041	6.611	9.937	16.722
		7	W	0.932	1.879	2.224	2.808	3.752	5.578	9.276
			E	0.853	1.716	2.038	2.574	3.437	5.115	8.499
		9	W	0.806	1.603	1.928	2.412	3.222	4.797	8.024
			E	0.730	1.450	1.748	2.185	2.928	4.352	7.267
		14	W	0.625	1.262	1.509	1.888	2.505	3.766	6.263
			E	0.558	1.125	1.343	1.685	2.236	3.353	5.590
	2.0	3	W	1.713	3.458	4.156	5.277	6.856	10.307	16.998
			E	1.638	3.306	3.967	5.053	6.560	9.859	16.266
		7	W	0.934	1.866	2.208	2.822	3.738	5.629	9.392
			E	0.854	1.710	2.026	2.589	3.419	5.151	8.581
		9	W	0.808	1.600	1.923	2.418	3.189	4.798	8.032
			E	0.733	1.451	1.743	2.193	2.890	4.347	7.276
			W	0.628	1.260	1.496	1.888	2.523	3.768	6.302
			E	0.560	1.124	1.338	1.688	2.250	3.366	5.624

i. The empirical coverage probabilities of all methods are close to the confidence level 0.95.

ii. The expected lengths of E and W increase when the parameter β increases. Additionally, the expected length of E is smaller than W especially when β is large.

iii. The expected length of our generalized confidence interval for α decreases when the parameter β increases. Moreover, it is very large when β is small.

iv. The expected lengths of all methods decrease when the number of records increases.

v. The empirical coverage probabilities and expected lengths of W and E do not change when the parameter α changes.

3. Joint confidence regions for the parameters

Suppose $R_0, R_1, \ldots, R_n$ are the first $n+1$ upper record values from a Weibull distribution with parameters α and β. In this section, we presented a joint confidence region for the parameters α and β. This is important because it can be used to find confidence bounds for any function of the parameters such as the reliability function $R(t) = \exp(-(\frac{t}{\alpha})^{\beta})$. For more references about the joint confidence region based on records, see Asgharzadeh and Abdi (2011a,b) and Asgharzadeh et al. (2011).

3.1. Asgharzadeh and Abdi method

Asgharzadeh and Abdi (2011b) present exact joint confidence regions for the parameters of Weibull distribution based on the record values using the idea presented by Wu and Tseng (2006). The following inequalities determine $100\,(1-\gamma)\,\%$ joint confidence regions for α and β:

$$A_j = \begin{cases} \dfrac{\log\left(\left(\frac{n-j+1}{j}\right)k_1+1\right)}{\log\left(\frac{R_n}{R_{j-1}}\right)} < \beta < \dfrac{\log\left(\left(\frac{n-j+1}{j}\right)k_2+1\right)}{\log\left(\frac{R_n}{R_{j-1}}\right)} \\ R_n\left(\dfrac{2}{\chi^2_{(2n+2),(1+\sqrt{1-\gamma})/2}}\right)^{\frac{1}{\beta}} < \alpha < R_n\left(\dfrac{2}{\chi^2_{(2n+2),(1-\sqrt{1-\gamma})/2}}\right)^{\frac{1}{\beta}}, \end{cases} \tag{3.1}$$

for $j = 1, \ldots, n$, where

$$k_1 = F_{(2n-2j+2,2j),(1-\sqrt{1-\gamma})/2} \quad k_2 = F_{(2n-2j+2,2j),(1+\sqrt{1-\gamma})/2},$$

and $F_{(a,b),\gamma}$ is the γth percentile of the F distribution with a and b degrees of freedom. Note that for each j, we have a joint confidence region for α and β. Asgharzadeh and Abdi (2011b) found that in most cases $A_{\lfloor\frac{n+1}{5}\rfloor}$ and $A_{\lfloor\frac{n+1}{5}+1\rfloor}$ provide the smallest confidence areas, where $\lfloor x \rfloor$ is the largest integer value smaller than x.

3.2. A new joint confidence region

From Theorem 2.1, $U = 2\beta\sum_{i=0}^{n}\log\left(\frac{R_n}{R_i}\right)$ has a chi-square distribution with $2n$ degrees of freedom and $V = 2\left(\frac{R_n}{\alpha}\right)^{\beta}$ has a chi-square distribution with $2n+2$ degrees of freedom, and U and V are independent. Therefore, an exact joint confidence region for the parameters α and β of Weibull distribution based on the record values can be given as

$$B = \begin{cases} \dfrac{\chi^2_{(2n),(1-\sqrt{1-\gamma})/2}}{2\sum_{i=0}^{n}\log\left(\frac{R_n}{R_i}\right)} < \beta < \dfrac{\chi^2_{(2n),(1+\sqrt{1-\gamma})/2}}{2\sum_{i=0}^{n}\log\left(\frac{R_n}{R_i}\right)} \\ R_n\left(\dfrac{2}{\chi^2_{(2n+2),(1+\sqrt{1-\gamma})/2}}\right)^{\frac{1}{\beta}} < \alpha < R_n\left(\dfrac{2}{\chi^2_{(2n+2),(1-\sqrt{1-\gamma})/2}}\right)^{\frac{1}{\beta}}. \end{cases} \tag{3.2}$$

Remark 3.1 *All record values are used in the proposed joint confidence region in* (3.2) *but not in the proposed joint confidence regions in* (3.1).

3.3. Real example

Here, we consider the upper record values in the example given in Section 2.3. Therefore, the 95% joint confidence regions for α and β based on Asgharzadeh and Abdi (2011b) in (3.1) are

$$A_1 = \left\{(\alpha,\beta) : 0.5826 < \beta < 11.9955,\ 41(0.1029)^{\frac{1}{\beta}} < \alpha < 41(1.1318)^{\frac{1}{\beta}}\right\}$$
$$A_2 = \left\{(\alpha,\beta) : 0.1646 < \beta < 6.4905,\ 41(0.1029)^{\frac{1}{\beta}} < \alpha < 41(1.1318)^{\frac{1}{\beta}}\right\}$$
$$A_3 = \left\{(\alpha,\beta) : 0.1720 < \beta < 58.9824,\ 41(0.1029)^{\frac{1}{\beta}} < \alpha < 41(1.1318)^{\frac{1}{\beta}}\right\}$$

and the 95% joint confidence region for α and β in (3.2) is

$$B = \left\{(\alpha,\beta) : 0.5305 < \beta < 9.0277,\ 41(0.1029)^{\frac{1}{\beta}} < \alpha < 41(1.1318)^{\frac{1}{\beta}}\right\}.$$

The plot of all joint confidence regions are given in Figure 1. Also, the area of the joint confidence regions A_1, A_2, A_3, and B are 194.9723, 166.7113, 369.7654, and 172.5757, respectively.

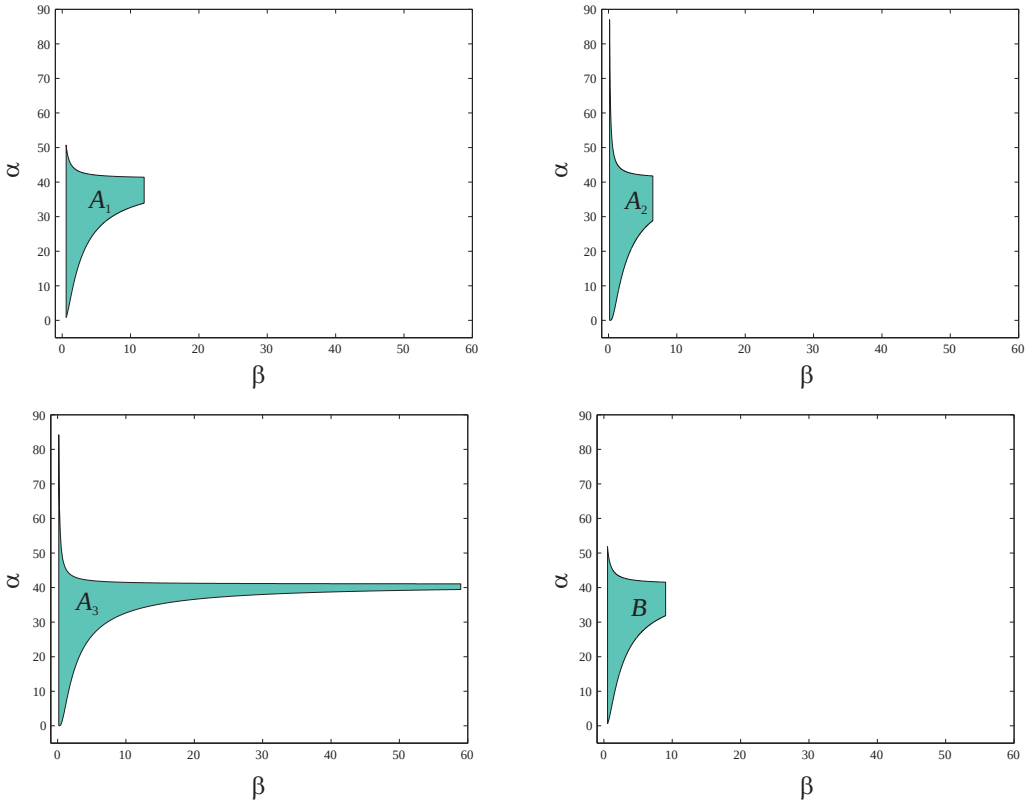


Figure 1: *The plot of the joint confidence regions A_1, A_2, A_3, and B.*

3.4. Simulation study

We performed a similar simulation given in Section 2.4 with considering $\alpha = 1$, in order to compare the joint confidence regions proposed by Asgharzadeh and Abdi (2011b) and our joint confidence region (B) in (3.2). Here, we consider the confidence areas $A_{\lfloor \frac{n+1}{5} \rfloor}$ and $A_{\lfloor \frac{n+1}{5}+1 \rfloor}$ because the coverage probabilities of all A_i's are close to the confidence coefficient and Asgharzadeh and Abdi (2011b) found that in most cases these two confidence areas provide the smallest confidence areas. The empirical coverage probabilities and expected areas of the methods for the confidence coefficient 95% are given in Table 3. We can conclude that

1. The coverage probabilities of the all methods are close to the confidence coefficient 0.95.
2. The expected area of our method is smaller than the expected areas of the proposed methods by Asgharzadeh and Abdi (2011b).
3. The expected areas of all methods decrease when the number of records increases.
4. The expected areas of all methods decrease when the parameter β increases.

Table 3: *Empirical coverage probabilities of the methods for constructing joint confidence region for the parameters α and β with $\gamma = 0.05$.*

			β						
	n	Region	0.5	1.0	1.2	1.5	2.0	3.0	5.0
Coverage Probability	4	A_1	0.950	0.949	0.949	0.951	0.954	0.946	0.950
		A_2	0.949	0.951	0.950	0.951	0.953	0.949	0.952
		B	0.949	0.950	0.949	0.952	0.954	0.949	0.950
	6	A_1	0.951	0.951	0.952	0.952	0.954	0.949	0.949
		A_2	0.952	0.948	0.950	0.948	0.953	0.948	0.952
		B	0.953	0.949	0.951	0.951	0.953	0.950	0.951
	9	A_2	0.953	0.949	0.950	0.955	0.949	0.948	0.953
		A_3	0.951	0.951	0.951	0.956	0.949	0.948	0.949
		B	0.950	0.952	0.951	0.953	0.949	0.948	0.952
	14	A_3	0.947	0.950	0.954	0.948	0.954	0.951	0.952
		A_4	0.946	0.951	0.952	0.948	0.951	0.950	0.952
		B	0.948	0.949	0.952	0.947	0.951	0.953	0.952
	29	A_6	0.951	0.953	0.950	0.950	0.948	0.948	0.950
		A_7	0.950	0.952	0.951	0.949	0.948	0.948	0.950
		B	0.953	0.955	0.951	0.953	0.950	0.948	0.952
Expected Area	4	A_1	27.787	8.548	7.339	6.331	5.725	5.330	5.203
		A_2	30.020	8.976	7.651	6.682	5.989	5.593	5.504
		B	22.985	7.371	6.388	5.596	5.099	4.792	4.713
	6	A_1	21.062	6.036	5.213	4.576	4.102	3.783	3.701
		A_2	20.035	5.824	5.046	4.436	3.985	3.701	3.648
		B	14.551	4.714	4.144	3.714	3.399	3.192	3.162
	9	A_2	14.631	4.081	3.533	3.059	2.756	2.574	2.510
		A_3	14.651	4.086	3.545	3.077	2.767	2.585	2.541
		B	9.639	3.137	2.774	2.471	2.275	2.160	2.133
	14	A_3	9.436	2.702	2.320	2.035	1.816	1.701	1.661
		A_4	9.388	2.686	2.304	2.035	1.812	1.707	1.668
		B	5.784	1.999	1.763	1.599	1.471	1.405	1.385
	29	A_6	4.244	1.298	1.124	0.988	0.905	0.849	0.828
		A_7	4.202	1.291	1.118	0.985	0.904	0.850	0.828
		B	2.380	0.932	0.838	0.761	0.719	0.689	0.678

Appendix. Generalized p-value and generalized confidence interval

Let $\boldsymbol{X}$ be a random variable whose distribution depends on a parameter of interest θ, and a nuisance parameter λ. Let $\boldsymbol{x}$ denote the observed value of $\boldsymbol{X}$. A generalized pivotal quantity for θ is a random quantity denoted by $T(\boldsymbol{X};\boldsymbol{x};\theta)$ that satisfies the following conditions:

(i) The distribution of $T(\boldsymbol{X};\boldsymbol{x};\theta)$ is free of any unknown parameters.

(ii) The value of $T(\boldsymbol{X};\boldsymbol{x};\theta)$ at $\boldsymbol{X}=\boldsymbol{x}$, i.e., $T(\boldsymbol{x};\boldsymbol{x};\theta)$ is free of the nuisance parameter λ.

Appropriate percentiles of $T(\boldsymbol{X};\boldsymbol{x};\theta)$ form a confidence interval for θ. Specifically, if $T(\boldsymbol{x};\boldsymbol{x};\theta)=\theta$, and T_δ denotes the 100δ percentage point of $T(\boldsymbol{X};\boldsymbol{x};\theta)$ then $(T_{\gamma/2},\ T_{1-\gamma/2})$ is a $1-\gamma$ generalized confidence interval for θ. The percentiles can be found because, for a given $\boldsymbol{x}$, the distribution of $T(\boldsymbol{X};\boldsymbol{x};\theta)$ does not depend on any unknown parameters.

In the above setup, suppose we are interested in testing the hypotheses

$$H_0:\theta\leq\theta_0 \qquad vs. \qquad H_1:\theta>\theta_0, \tag{A.1}$$

for a specified θ_0. The generalized test variable, denoted by $T^*(\boldsymbol{X};\boldsymbol{x};\theta)$, is defined as follows:

(i) The value of $T^*(\boldsymbol{X};\boldsymbol{x};\theta)$ at $\boldsymbol{X}=\boldsymbol{x}$ is free of any unknown parameters.
(ii) The distribution of $T^*(\boldsymbol{X};\boldsymbol{x};\theta)$ is stochastically monotone (i.e., stochastically increasing or stochastically decreasing) in θ for any fixed $\boldsymbol{x}$ and λ.
(iii) The distribution of $T^*(\boldsymbol{X};\boldsymbol{x};\theta)$ is free of any unknown parameters.

Let $t^*=T^*(\boldsymbol{x};\boldsymbol{x};\theta_0)$, the observed value of $T^*(\boldsymbol{X};\boldsymbol{x};\theta)$ at $(\boldsymbol{X};\theta)=(\boldsymbol{x};\theta_0)$. When the above conditions hold, the generalized p-value for testing the hypotheses in (A.1) is defined as

$$p=P\left(T^*(\boldsymbol{X};\boldsymbol{x};\theta_0)\leq t^*\right) \tag{A.2}$$

where $T^*(\boldsymbol{X};\boldsymbol{x};\theta)$ is stochastically decreasing in θ. The test based on the generalized p-value rejects H_0 when the generalized p-value is smaller than a nominal level γ. However, the size and power of such a test may depend on the nuisance parameters.

Acknowledgements

The authors would like to thank two referees for their helpful comments and suggestions which have contributed to improving the manuscript.

References

Ahsanullah, M. (1995). *Record Statistics*. Nova Science Publishers Commack, New York.

Arnold, B., Balakrishnan, N. and Nagaraja, H. (1998). *Records*. John Wiley & Sons Inc, New York.

Asgharzadeh, A. and Abdi, M. (2011a). Confidence intervals and joint confidence regions for the two-parameter exponential distribution based on records. *Communications of the Korean Statistical Society*, 18, 103–110.

Asgharzadeh, A. and Abdi, M. (2011b). Joint confidence regions for the parameters of the Weibull distribution based on record. *ProbStat Forum*, 4, 12–24.

Asgharzadeh, A., Abdi, M. and Kuş, C. (2011). Interval estimation for the two-parameter pareto distribution based on record values. *Selçuk Journal of Applied Mathematics*, Special Issue: 149–161.

Balakrishnan, N. and Chan, P. S. (1994). Record values from Rayleigh and Weibull distributions and associated inference. *National Institute of Standards and Technology Journal of Research, Special Publication, Proceedings of the Conference on Extreme Value Theory and Applications*, 866, 41–51.

Behboodian, J. and Jafari, A. A. (2006). Generalized inference for the common mean of several lognormal populations. *Journal of Statistical Theory and Applications*, 5, 240–259.

Chan, P. S. (1998). Interval estimation of location and scale parameters based on record values. *Statistics and Probability Letters*, 37, 49–58.

Chandler, K. N. (1952). The distribution and frequency of record values. *Journal of the Royal Statistical Society. Series B (Methodological)*, 14, 220–228.

Dallas, A. C. (1982). Some results on record values from the exponential and Weibull law. *Acta Mathematica Academy of Sciences of Hungary*, 40, 307–311.

Gouet, R., López, F. J. and Sanz, G. (2012). On δ-record observations: asymptotic rates for the counting process and elements of maximum likelihood estimation. *Test*, 21, 188–214.

Jafari, A. A. (2012). Inferences on the ratio of two generalized variances: independent and correlated cases. *Statistical Methods and Applications*, 21, 297–314.

Krishnamoorthy, K. and Lin, Y. (2010). Confidence limits for stress-strength reliability involving Weibull models. *Journal of Statistical Planning and Inference*, 140, 1754–1764.

Krishnamoorthy, K., Lin, Y. and Xia, Y. (2009). Confidence limits and prediction limits for a Weibull distribution based on the generalized variable approach. *Journal of Statistical Planning and Inference*, 139, 2675–2684.

Roberts, E. (1979). Review of statistics of extreme values with applications to air quality data: Part ii. applications. *Journal of the Air Pollution Control Association*, 29, 733–740.

Soliman, A. A., Abd Ellah, A. H. and Sultan, K. S. (2006). Comparison of estimates using record statistics from Weibull model: Bayesian and non-Bayesian approaches. *Computational Statistics and Data Analysis*, 51, 2065–2077.

Teimouri, M. and Gupta, A. K. (2012). On the Weibull record statistics and associated inferences. *Statistica*, 72, 145–162.

Teimouri, M. and Nadarajah, S. (2013). Bias corrected MLEs for the Weibull distribution based on records. *Statistical Methodology*, 13, 12–24.

Tsui, K. W. and Weerahandi, S. (1989). Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association*, 84, 602–607.

Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association*, 88, 899–905.

Weerahandi, S. (1995). *Exact Statistical Methods for Data Analysis*. Springer Verlag, New York.

Wu, J. W. and Tseng, H. C. (2006). Statistical inference about the shape parameter of the Weibull distribution by upper record values. *Statistical Papers*, 48, 95–129.

SORT 39 (1) January-June 2015, 19-34

Small area estimation of poverty indicators under partitioned area-level time models*

Domingo Morales[1], Maria Chiara Pagliarella[2] and Renato Salvatore[2]

Abstract

This paper deals with small area estimation of poverty indicators. Small area estimators of these quantities are derived from partitioned time-dependent area-level linear mixed models. The introduced models are useful for modelling the different behaviour of the target variable by sex or any other dichotomic characteristic. The mean squared errors are estimated by explicit formulas. An application to data from the Spanish Living Conditions Survey is given.

MSC: 62D05, 62J05.

Keywords: Area-level models, small area estimation, time correlation, poverty indicators.

1. Introduction

In most European countries, the estimation of poverty is done by using the Living Conditions Survey (LCS) data. The Spanish LCS (SLCS) uses a stratified two-stage design within each Autonomous Community. As most provinces have a very small sample size, the direct estimates at that level have a low accuracy. The problem is thus that domain sample sizes are too small to carry out direct estimations. This situation may be treated by using small area estimation techniques. Small Area Estimation (SAE) is a part of the statistical science that combines survey sampling and finite population inference with statistical models. See a description of this theory in the monograph of Rao (2003), or in the reviews of Ghosh and Rao (1994), Rao (1999), Pfeffermann (2002, 2012) and more recently Jiang and Lahiri (2006).

* Supported by the grants EU-FP7-SSH-2007-1, MTM2012-37077-C02-01.
[1] Centro de Investigación Operativa, Universidad Miguel Hernández de Elche, Spain. d.morales@umh.es
[2] Dipartimento di Economia Politica e Statistica, Università degli Studi di Siena, Italy. mariachiara.pagliarella@unisi.it
[3] Dipartimento di Economia e Giurisprudenza, Università di Cassino e del Lazio Meridionale, Italy. rsalvatore@unicas.it
Received: April 2014
Accepted: September 2014

This paper deals with the estimation of poverty indicators by using area-level models. For this sake, Esteban et al. (2012a,b) proposed several area-level time models. They argue that employing data from past periods produce a significant improvement of the estimation process. Marhuenda et al. (2013) introduced some more complex area-level linear mixed models that take into account for temporal and spatial correlation. The first two papers gave empirical best linear unbiased prediction (EBLUP) estimates of poverty estimators for Spanish provinces crossed by sex. The third one did not give estimates by sex. Many socio-economic indicators, such as those related with poverty and labour, behave differently in the subpopulations of men and women. This is why, we adapt some of the temporal models appearing in Esteban et al. (20121,b) and Marhuenda et al. (2013) to this situation.

In this paper we use four time-dependent area-level linear mixed models to obtain small area estimates of poverty indicators. Two of them are specified with a partition of the population in two groups. This fact allows modelling, for example, a different behaviour of the target variable by sex, as it was done by Herrador et al. (2011). This is an important modelling tool as many socioeconomic indicators behave differently for men and women. Following Esteban et al. (2012b), the first partitioned model assumes that time dependency is explained by the auxiliary variables and the second one contains a correlation parameter in the distribution of the random intercept. The estimates of the model parameters are obtained by using the residual maximum likelihood (REML) estimation method. These estimates are then used to construct empirical best linear unbiased predictors of poverty indicators by sex of the Spanish provinces. Estimation of the mean squared error (MSE) of model-based estimators is an important issue that has no easy solution. In this paper we follow Prasad and Rao (1990) and Das, Jiang and Rao (2004) to introduce an approximation of the MSE and the corresponding MSE estimator.

The rest of the paper is organized as follows. Section 2 introduces the considered area-level time models and the corresponding model-based estimators of poverty indicators. Section 3 describes the estimation problem of interest and presents an application to data from the SLCS. The target is to estimate poverty indicators by sex in the Spanish provinces. Finally, Section 4 gives a discussion on the findings of this paper.

2. The area-level partitioned time models

2.1. The models

Let us consider a population partitioned in D domains. Assume that domains are classified in two groups of sizes D_A and D_B ($D_A + D_B = D$) that behave differently with respect to some socioeconomic characteristic. For example, let us consider a country divided in provinces. Assume that a statistical agency is interested in estimating some poverty indicators of regions by sex. In that situation, they can define the domains as regions crossed by sex, so that they have $D_A = D_B$ and $D = 2D_A = 2D_B$. Another example is a

state partitioned in D_A urban-type counties and D_B rural-type counties, where the interest is the estimation of some labour indicators at the county level. In what follows, we will introduce some models adapted to these kind of situations.

Let us consider the model (model 3)

$$y_{dt} = \mathbf{x}_{dt}^{\mathsf{T}}\boldsymbol{\beta} + u_{dt} + e_{dt}, \quad d = 1,\ldots,D = D_A + D_B, \quad t = 1,\ldots,m_d, \tag{1}$$

where y_{dt} is a direct estimator of the indicator of interest for area d and time instant t, and $\mathbf{x}_{dt}^{\mathsf{T}}$ is a row vector containing the aggregated (population) values of p auxiliary variables. The index d is used for domains and the index t for time instants. We assume that the random vectors $(u_{d1},\ldots,u_{dm_d})$, $d \le D_A$, follow independent and identically distributed (i.i.d.) first order auto-regressive processes with variance and auto-correlation parameters σ_A^2 and ρ_A respectively; in short, $(u_{d1},\ldots,u_{dm_d}) \sim_{iid} AR1(\sigma_A^2,\rho_A)$, $d \le D_A$. We further assume that $(u_{d1},\ldots,u_{dm_d}) \sim_{iid} AR1(\sigma_B^2,\rho_B)$, $d > D_A$, and that the errors e_{dt}'s are independent $N(0,\sigma_{dt}^2)$ with known variances σ_{dt}^2's. Finally we assume that the $(u_{d1},\ldots,u_{dm_d})$'s and the e_{dt}'s are mutually independent.

The introduction of the partitioned model (1) is motivated by the observed different behaviour by sex of poverty indicators in Spanish data. Further, we also consider the models restricted to $\rho_A = \rho_B$ (model 2), restricted to $\rho_A = \rho_B = 0$ (model 1) and restricted to $\rho_A = \rho_B = 0$ and $\sigma_A^2 = \sigma_B^2$ (model 0). For the sake of brevity, we only present the theoretical developments for the partitioned model 3.

In matrix notation the model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where $\mathbf{y}$ can be decomposed in the form $\mathbf{y} = (\mathbf{y}_A^{\mathsf{T}}, \mathbf{y}_B^{\mathsf{T}})^{\mathsf{T}}$, with $\mathbf{y}_A = \underset{d\le D_A}{\mathrm{col}}(\mathbf{y}_d)$, $\mathbf{y}_B = \underset{d>D_A}{\mathrm{col}}(\mathbf{y}_d)$ and $\mathbf{y}_d = \underset{1\le t\le m_d}{\mathrm{col}}(y_{dt})$, and similarly for $\mathbf{u}$ and $\mathbf{e}$, $\mathbf{X}$ can be decomposed in the form $\mathbf{X} = (\mathbf{X}_A^{\mathsf{T}}, \mathbf{X}_B^{\mathsf{T}})^{\mathsf{T}}$, with $\mathbf{X}_A = \underset{d\le D_A}{\mathrm{col}}(\mathbf{X}_d)$, $\mathbf{X}_B = \underset{d>D_A}{\mathrm{col}}(\mathbf{X}_d)$ and $\mathbf{X}_d = \underset{1\le t\le m_d}{\mathrm{col}}(\mathbf{x}_{dt}^{\mathsf{T}})$, $\boldsymbol{\beta} = \boldsymbol{\beta}_{p\times 1}$, $\mathbf{Z} = \mathbf{I}_M$ and $M = \sum_{d=1}^{D} m_d$. We use the notation $\mathrm{col}(\cdots)$ to denote a column vector, or set of column vectors, composed of the elements of the argument, which can be scalars or vectors. In this notation, $\mathbf{u} \sim N(\mathbf{0},\mathbf{V}_u)$ and $\mathbf{e} \sim N(\mathbf{0},\mathbf{V}_e)$ are independent with covariance matrices

$$\mathbf{V}_u = \mathrm{var}(\mathbf{u}) = \mathrm{diag}(\sigma_A^2\Omega_A, \sigma_B^2\Omega_B), \quad \mathbf{V}_e = \mathrm{var}(\mathbf{e}) = \underset{1\le d\le D}{\mathrm{diag}}(\mathbf{V}_{ed}),$$

where $\Omega_A = \underset{d\le D_A}{\mathrm{diag}}(\Omega_d)$, $\Omega_B = \underset{d>D_A}{\mathrm{diag}}(\Omega_d)$, $\mathbf{V}_{ed} = \underset{1\le t\le m_d}{\mathrm{diag}}(\sigma_{dt}^2)$ and

$$\Omega_d = \Omega_d(\rho) = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \cdots & \rho^{m_d-2} & \rho^{m_d-1} \\ \rho & 1 & \ddots & & \rho^{m_d-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{m_d-2} & & \ddots & 1 & \rho \\ \rho^{m_d-1} & \rho^{m_d-2} & \cdots & \rho & 1 \end{pmatrix}_{m_d\times m_d}, \quad \rho = \rho_A, \rho_B.$$

The covariance matrix of vector $\mathbf{y}$ is $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta}) = \text{var}(\mathbf{y}) = \text{diag}(\mathbf{V}_A, \mathbf{V}_B)$, where $\mathbf{V}_A = \underset{d \le D_A}{\text{diag}}(\mathbf{V}_d)$, $\mathbf{V}_B = \underset{d > D_A}{\text{diag}}(\mathbf{V}_d)$, $\mathbf{V}_d = \sigma_A^2 \Omega_d + \mathbf{V}_{ed}$ if $d \le D_A$, $\mathbf{V}_d = \sigma_B^2 \Omega_d + \mathbf{V}_{ed}$ if $d > D_A$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4) = (\sigma_A^2, \rho_A, \sigma_B^2, \rho_B)$. The residual loglikelihood is

$$l_{reml} = l_{reml}(\boldsymbol{\theta}) = -\frac{M-p}{2}\log 2\pi + \frac{1}{2}\log|\mathbf{X}^\top\mathbf{X}| - \frac{1}{2}\log|\mathbf{V}_A| - \frac{1}{2}\log|\mathbf{V}_B| - \frac{1}{2}\log|\mathbf{X}_A^\top\mathbf{V}_A^{-1}\mathbf{X}_A + \mathbf{X}_B^\top\mathbf{V}_B^{-1}\mathbf{X}_B| - \frac{1}{2}\mathbf{y}^\top\mathbf{P}\mathbf{y},$$

where $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}^{-1}$. The scores and the Fisher information matrix components are

$$S_a = \frac{\partial l_{reml}}{\partial \theta_a}, \quad F_{ab} = -E\left[\frac{\partial l_{reml}^2}{\partial \theta_a \partial \theta_b}\right], \quad a,b = 1,2,3,4.$$

To calculate the residual maximum likelihood (REML) estimate, $\hat{\boldsymbol{\theta}}$, we apply the Fisher-scoring algorithm with the updating formula

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k + \mathbf{F}^{-1}(\boldsymbol{\theta}^k)\mathbf{s}(\boldsymbol{\theta}^k),$$

where $\mathbf{s}$ and $\mathbf{F}$ are the column vector of scores and the Fisher information matrix respectively. As seeds we use $\rho_A^{(0)} = \rho_B^{(0)} = 0$, and $\sigma_A^{2(0)} = \sigma_B^{2(0)} = \hat{\sigma}_{uH}^2$, where $\hat{\sigma}_{uH}^2$ is the Henderson 3 estimator under model with $\rho_A = \rho_B = 0$ and $\sigma_A^2 = \sigma_B^2$. The REML estimator of $\boldsymbol{\beta}$ and the REML empirical best linear unbiased predictor (EBLUP) of $\mathbf{u}$ are

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\hat{\mathbf{V}}^{-1}\mathbf{y}, \quad \hat{\mathbf{u}} = \hat{\mathbf{V}}_u\mathbf{Z}^\top\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}),$$

where $\hat{\mathbf{V}} = \mathbf{V}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{V}}_u = \mathbf{V}_u(\hat{\boldsymbol{\theta}})$.

2.2. Statistical inference on the model parameters

The asymptotic distributions of of the REML estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are

$$\hat{\boldsymbol{\theta}} \sim N_4(\boldsymbol{\theta}, \mathbf{F}^{-1}(\boldsymbol{\theta})), \quad \hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, (\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}).$$

Asymptotic confidence intervals at the level $1-\alpha$ for θ_a and β_j are

$$\hat{\theta}_a \pm z_{\alpha/2}\, v_{aa}^{1/2},\ a = 1,2,3,4, \quad \hat{\beta}_j \pm z_{\alpha/2}\, q_{jj}^{1/2},\ j = 1,\ldots,p,$$

where $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{\kappa}$, $\mathbf{F}^{-1}(\boldsymbol{\theta}^{\kappa}) = (v_{ab})_{a,b=1,2,3,4}$, $(\mathbf{X}^{\top}\mathbf{V}^{-1}(\boldsymbol{\theta}^{\kappa})\mathbf{X})^{-1} = (q_{ij})_{i,j=1,\ldots,p}$, κ is the final iteration of the Fisher-scoring algorithm and z_{α} is the α-quantile of the standard normal distribution $N(0,1)$. Observed $\hat{\beta}_j = \beta_0$, the p-value for testing the hypothesis $H_0 : \beta_j = 0$ is

$$p = 2P_{H_0}(\hat{\beta}_j > |\beta_0|) = 2P(N(0,1) > \beta_0/\sqrt{q_{jj}}).$$

Let $\hat{\sigma}_A^2$, $\hat{\sigma}_B^2$, $\hat{\rho}_A$ and $\hat{\rho}_B$ be the unrestricted REML estimators of σ_A^2 and σ_B^2, ρ_A and ρ_B respectively. Let $\tilde{\sigma}_A^2$, $\tilde{\sigma}_B^2$ and $\tilde{\rho}$ be the REML estimator of σ_A^2, σ_B^2 and of the common value $\rho_A = \rho_B$ under H_0 (model 2). Under model 3, the REML likelihood ratio statistic (LRS) for testing $H_0 : \rho_A = \rho_B$ is

$$\lambda = -2[l_{REML}(\tilde{\sigma}_A^2, \tilde{\sigma}_B^2, \tilde{\rho}) - l_{REML}(\hat{\sigma}_A^2, \hat{\sigma}_B^2, \hat{\rho}_A, \hat{\rho}_B)].$$

The asymptotic distribution of λ under H_0 is χ_1^2. The null hypothesis is rejected at the level α if $\lambda > \chi_{1,\alpha}^2$.

Under model 2, the REML LRS for testing $H_0 : \rho = 0$ is

$$\lambda = -2[l_{REML}(\tilde{\sigma}_A^2, \tilde{\sigma}_B^2) - l_{REML}(\hat{\sigma}_A^2, \hat{\sigma}_B^2, \hat{\rho})],$$

where $\hat{\sigma}_A^2$, $\hat{\sigma}_B^2$ and $\hat{\rho}$ are the unrestricted REML estimators of σ_A^2, σ_B^2 and ρ respectively, $\tilde{\sigma}_A^2$ and $\tilde{\sigma}_B^2$ are the REML estimator of σ_A^2 and σ_B^2 under H_0 (model 1). The asymptotic distribution of λ under H_0 is χ_1^2, so the null hypothesis is rejected at the level α if $\lambda > \chi_{1,\alpha}^2$.

2.3. The EBLUP and its mean squared error

We are interested in predicting the value of $\mu_{dt} = \mathbf{x}_{dt}^{\top}\boldsymbol{\beta} + u_{dt}$ by using the EBLUP $\hat{\mu}_{dt} = \mathbf{x}_{dt}^{\top}\hat{\boldsymbol{\beta}} + \hat{u}_{dt}$. If we do not take into account the error, e_{dt}, this is equivalent to predict $y_{dt} = \mathbf{a}^{\top}\mathbf{y}$, where $\mathbf{a} = \underset{1\le\ell\le D}{\text{col}}\,(\underset{1\le k\le m_\ell}{\text{col}}\,(\delta_{d\ell}\delta_{tk}))$ is a vector having one 1 in the position $t + \sum_{\ell=1}^{d-1} m_\ell$ and 0's in the remaining cells. To estimate $\overline{Y}_{dt}$ we use $\widehat{\overline{Y}}_{dt}^{eblup} = \hat{\mu}_{dt}$. The mean squared error of $\widehat{\overline{Y}}_{dt}^{eblup}$ can be approximated by considering the formula established by Prasad and Rao (1980) for moment-based estimators of model parameters in the Fay-Herriot model. This formula was later extended by Datta and Lahiri (2000) and Das, Jiang and Rao (2004) to a wide variety of linear mixed models when the model parameters are estimated by the ML and REML method. By adapting the mean squared error formula to model 3, we get

$$MSE(\widehat{\overline{Y}}_{dt}^{eblup}) = g_{1dt}(\boldsymbol{\theta}) + g_{2dt}(\boldsymbol{\theta}) + g_{3dt}(\boldsymbol{\theta}),$$

where $\boldsymbol{\theta} = (\sigma_A^2, \rho_A, \sigma_B^2, \rho_B)$,

$$
\begin{aligned}
g_{1dt}(\boldsymbol{\theta}) &= \mathbf{a}^\top \mathbf{Z}\mathbf{T}\mathbf{Z}^\top \mathbf{a}, \\
g_{2dt}(\boldsymbol{\theta}) &= [\mathbf{a}^\top \mathbf{X} - \mathbf{a}^\top \mathbf{Z}\mathbf{T}\mathbf{Z}^\top \mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}^\top \mathbf{a} - \mathbf{X}^\top \mathbf{V}_e^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}^\top \mathbf{a}], \\
g_{3dt}(\boldsymbol{\theta}) &\approx \operatorname{tr}\left\{ \nabla \mathbf{b}^\top \mathbf{V} \nabla \mathbf{b}\, E\left[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \right] \right\}.
\end{aligned}
$$

$\mathbf{T} = \mathbf{V}_u - \mathbf{V}_u \mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z}\mathbf{V}_u$, $\mathbf{Q} = (\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X})^{-1}$, $\mathbf{b}^\top = \mathbf{a}^\top \mathbf{Z}\mathbf{V}_u \mathbf{Z}^\top \mathbf{V}^{-1}$, $\nabla \mathbf{b}^\top = (\frac{\partial \mathbf{b}^\top}{\partial \sigma_A^2}, \frac{\partial \mathbf{b}^\top}{\partial \sigma_B^2}, \frac{\partial \mathbf{b}^\top}{\partial \rho_A}, \frac{\partial \mathbf{b}^\top}{\partial \rho_B})$.

The estimator of $MSE(\widehat{\overline{Y}}_{dt}^{eblup})$ is

$$mse_{dt}(\widehat{\overline{Y}}_{dt}^{eblup}) = g_{1dt}(\hat{\boldsymbol{\theta}}) + g_{2dt}(\hat{\boldsymbol{\theta}}) + 2g_{3dt}(\hat{\boldsymbol{\theta}}).$$

3. Estimation of poverty indicators

3.1. The indicators and the data

Let z_{dtj} be an income variable measured in all the units j of the population and let z_t be the poverty line, so that units from domain d with $z_{dtj} < z_t$ are considered as poor at time period t. Let N_t and N_{dt}, $d = 1, \ldots, D$, be the population size at time t and the population size of each domain d at time t respectively. Foster et al. (1984) introduced the family of poverty indicators

$$\overline{Y}_{\alpha,dt} = \frac{1}{N_{dt}} \sum_{j=1}^{N_{dt}} y_{\alpha,dtj}, \quad \text{where } y_{\alpha,dtj} = \left(\frac{z_t - z_{dtj}}{z_t} \right)^\alpha I(z_{dtj} < z_t), \tag{2}$$

$I(z_{dtj} < z_t) = 1$ if $z_{dtj} < z_t$ and $I(z_{dtj} < z_t) = 0$ otherwise. The proportion of units under poverty in the domain d and period t is thus $\overline{Y}_{0,dt}$ and the poverty gap is $\overline{Y}_{1,dt}$.

The Spanish Statistical Office fixes the Poverty Threshold z_t at the 60% of the median of the normalized incomes in Spanish households. The aim of normalizing the household income is to adjust for the varying size and composition of households. The definition of the total number of normalized household members uses a scale giving a weight 1.0 to the first adult, 0.5 to the second and each subsequent person aged 14 and over and 0.3 to each child aged under 14 in the household. The *normalized size* of a household is the sum of the weights assigned to each person. So for each household h in domain d and time t, the total number of normalized members is

$$H_{dth} = 1 + 0.5(H_{dth\geq 14} - 1) + 0.3 H_{dth<14},$$

where $H_{dth\geq 14}$ is the number of people aged 14 and over and $H_{dth<14}$ is the number of children aged under 14. The normalized net annual income of a household is obtained by dividing its net annual income by its normalized size. The Spanish poverty thresholds (in euros) in 2004-06 are $z_{2004} = 6098.57$, $z_{2005} = 6160.00$ and $z_{2006} = 6556.60$ respectively. These are the z_t-values used in the calculation of the direct estimates of the poverty incidence and gap.

We use data from the Spanish Living Conditions Survey (SLCS) corresponding to years 2004-2006. The SLCS started in 2004 with an annual periodicity and is the Spanish version of the European Statistics on Income and Living Conditions (EU-SILC), which is one of the statistical operations that have been harmonized for EU countries. We consider $D = 104$ domains obtained by crossing 52 provinces with 2 sexes.

The direct estimator of the total, $Y_{dt} = \sum_{j=1}^{N_{dt}} y_{dtj}$, is

$$\hat{Y}_{dt}^{dir} = \sum_{j \in S_{dt}} w_{dtj}\, y_{dtj}.$$

where S_{dt} is the domain sample at time period t and the w_{dtj}'s are the official calibrated sampling weights which take into account for non response. The estimated domain size

$$\hat{N}_{dt}^{dir} = \sum_{j \in S_{dt}} w_{dtj}.$$

Using these quantities, a direct estimator of the domain mean, $\bar{Y}_{dt}$, is $\bar{y}_{dt} = \hat{Y}_{dt}^{dir} / \hat{N}_{dt}^{dir}$. The design-based variances of these estimators can be approximated by

$$\hat{V}_{\pi}(\hat{Y}_{dt}^{dir}) = \sum_{j \in S_{dt}} w_{dtj}(w_{dtj} - 1)\,(y_{dtj} - \bar{y}_{dt})^2 \quad \text{and} \quad \hat{V}_{\pi}(\bar{y}_{dt}) = \hat{V}\left(\hat{Y}_{dt}^{dir}\right) / \hat{N}_{dt}^{2}. \tag{3}$$

The last formulas are obtained from Särndal et al. (1992), pp. 43, 185 and 391, with the simplifications $w_{dtj} = 1/\pi_{dtj}$, $\pi_{dtj,dtj} = \pi_{dtj}$ and $\pi_{dti,dtj} = \pi_{dti}\pi_{dtj}$, $i \neq j$ in the second order inclusion probabilities.

As we are interested in the cases $y_{dtj} = y_{\alpha,dtj}$, $\alpha = 0, 1$, we select the direct estimates of the poverty incidence and poverty gap at domain d and time period t (i.e. $\bar{y}_{0,dt}$ and $\bar{y}_{1,dt}$ respectively) as target variables for the time dependent area-level models.

The considered auxiliary variables are the known domain means of the category indicators of the following variables. INTERCEPT: *constant* equal to 1. AGE: Age groups are *age1-age5* for the intervals ≤ 15, $16-24$, $25-49$, $50-64$ and ≥ 65. EDUCATION: Highest level of education completed, with 4 categories denoted by *edu0* for Less than primary education level, *edu1* for Primary education level, *edu2* for Secondary education level and *edu3* for University level. LABOUR: Labour situation with 4 categories taking the values *lab0* for Below 16 years, *lab1* for Employed, *lab2* for Unemployed and *lab3* for Inactive.

3.2. The application

In this section we present an application to real data of model 3 defined in (1). We compare the obtained results with the corresponding ones under the same model restricted to $H_0 : \rho_A = \rho_B$ (model 2), $H_0 : \rho_A = \rho_B = 0$ (model 1) and $H_0 : \rho_A = \rho_B = 0, \sigma_A^2 = \sigma_B^2$ (model 0). Finally the main goal is to estimate the poverty incidence (proportion of individuals under poverty) and the poverty gap in Spanish domains for the three models.

The final selected models include only the auxiliary variables appearing in Table 1. We have included three statistically significant variables that have a relevant meaning in the socio-economic sense. We have selected the variables *age4* (age group 50-65), *edu2* (secondary education completed) and *lab2* (unemployed). Regression parameters and their corresponding p-values are also presented in Table 1 for $\alpha = 0$ and $\alpha = 1$.

By observing the signs of the regression parameters for $\alpha = 0$ (poverty proportion), we interpret that there is an inverse relation between poverty proportion and the categories *age4* and *edu2* of explanatory variables. That is, poverty incidence tends to be smaller in those domains with larger proportion of population in the subset defined by age between 50 and 64, and by secondary education level completed. On the other hand, poverty incidence tends to be larger in those domains with larger proportion of population in the subset defined by *lab2*, i.e. in the category of unemployed people. All the p-values are lower than 0.05 for all the considered auxiliary variables, except for *lab2* in model 3. By doing the same exercise with the signs of the regression parameters in the case $\alpha = 1$ (poverty gap), we can give the same interpretations as before. Again all the p-values are lower than 0.05.

The asymptotic confidence intervals (CIs) for the β's at the 90% confidence level are presented in Table 2 (top) for $\alpha = 0$ and in Table 2 (bottom) for $\alpha = 1$. The columns with labels INF and SUP contains the low and upper limits respectively. By

Table 1: *β-parameters and p-values for $\alpha = 0$ (left) and $\alpha = 1$ (right).*

	$\alpha = 0$				$\alpha = 1$			
model 3	*constant*	*age4*	*edu2*	*lab2*	*constant*	*age4*	*edu2*	*lab2*
β	0.622	−1.881	−0.272	0.260	0.215	−0.741	−0.100	0.320
p-value	0.000	0.000	0.000	0.284	0.000	0.000	0.002	0.004
model 2	*constant*	*age4*	*edu2*	*lab2*	*constant*	*age4*	*edu2*	*lab2*
β	0.778	−2.603	−0.425	0.772	0.237	−0.874	−0.115	0.413
p-value	0.000	0.000	0.000	0.026	0.000	0.000	0.002	0.002
model 1	*constant*	*age4*	*edu2*	*lab2*	*constant*	*age4*	*edu2*	*lab2*
β	0.713	−2.284	−0.445	1.264	0.232	−0.827	−0.123	0.472
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
model 0	*constant*	*age4*	*edu2*	*lab2*	*constant*	*age4*	*edu2*	*lab2*
β	0.730	−2.632	−0.411	1.829	0.198	−0.719	−0.107	0.667
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000

Table 2: *90% confidence intervals for $\alpha = 0$ (top) and for $\alpha = 1$ (bottom).*

	model 3		*model* 2		*model* 1		*model* 0	
ITEMS	INF	SUP	INF	SUP	INF	SUP	INF	SUP
constant	0.527	0.717	0.646	0.911	0.632	0.794	0.618	0.842
age4	−2.344	−1.418	−3.224	−1.982	−2.657	−1.912	−3.219	−2.045
edu2	−0.385	−0.159	−0.589	−0.262	−0.547	−0.342	−0.562	−0.260
lab2	−0.140	0.661	0.200	1.344	0.879	1.649	1.309	2.349
constant	0.173	0.257	0.188	0.286	0.199	0.264	0.154	0.242
age4	−0.941	−0.542	−1.102	−0.646	−0.978	−0.676	−0.952	−0.486
edu2	−0.152	−0.048	−0.177	−0.054	−0.166	−0.081	−0.169	−0.046
lab2	0.136	0.505	0.198	0.628	0.316	0.629	0.459	0.874

observing these confidence intervals, we conclude that all the regression parameters are significantly different from zero in both cases. The only exception is *lab2* in model 3 for $\alpha = 0$.

Table 3 presents the CIs for the variance components at the 90% confidence level, under models 3-0, for $\alpha = 0$ and $\alpha = 1$. The columns with labels INF and SUP contains the low and upper limits respectively. The column with label 0 ∈CI contains T (true) if 0 belongs to the CI and F (false) otherwise. Concerning model 3, we observe that the CIs for $\rho_A - \rho_B$ and $\sigma_A^2 - \sigma_B^2$ contain the 0. In the case of $\alpha = 0$, the observed value of the likelihood ratio statistics for testing $H_0 : \rho_A = \rho_B$ is $\lambda = 0.5738$ and the corresponding p-value is 0.4487. In the case of $\alpha = 1$, the observed value of the likelihood ratio statistics for testing $H_0 : \rho_A = \rho_B$ is $\lambda = 3.8195$ and the corresponding p-value is 0.0506. These facts suggest that model 3 is not the model fitting best to data.

Table 3: *90% confidence intervals for variances.*

		$\alpha = 0$			$\alpha = 1$		
Model	Parameter	INF	SUP	0 ∈CI	INF	SUP	0 ∈CI
3	σ_A^2	0.0002	0.0008	F	0.0003	0.0005	F
	σ_B^2	0.0005	0.0014	F	0.0002	0.0004	F
	$\sigma_A^2 - \sigma_B^2$	−0.0010	0.0001	T	−0.0000	0.0003	T
	ρ_A	0.8662	0.9957	F	0.5416	0.7484	F
	ρ_B	0.8598	0.9344	F	0.6017	0.8843	F
	$\rho_A - \rho_B$	−0.0409	0.1087	T	−0.2734	0.0774	T
2	σ_A^2	0.0101	0.0154	F	0.0014	0.0019	F
	σ_B^2	0.0023	0.0038	F	0.0004	0.0005	F
	$\sigma_A^2 - \sigma_B^2$	0.0070	0.0124	F	0.0009	0.0015	F
	ρ	0.4050	0.6108	F	0.3528	0.5756	F
1	σ_A^2	0.0025	0.0040	F	0.0004	0.0004	F
	σ_B^2	0.0028	0.0045	F	0.0006	0.0007	F
0	σ_u^2	0.0025	0.0040	F	0.0004	0.0006	F

Table 4: *Normalized Euclidean distances for* $\alpha = 0, 1$.

	Model 3		Model 2		Model 1		Model 0	
α	Men	Women	Men	Women	Men	Women	Men	Women
0	0.0194	0.0255	0.0083	0.0421	0.0285	0.0486	0.0648	0.0673
1	0.0115	0.0116	0.0121	0.0221	0.0188	0.0229	0.0290	0.0303

For models 2-0 Table 3 shows that the CIs for σ_A^2, σ_B^2 and σ_u^2 do not contain the origin 0 in any case, so the variances are significatively positive. Table 3 also presents the CIs for the difference of variances $\sigma_A^2 - \sigma_B^2$ and the CIs for ρ under model 2. The variances σ_A^2 and σ_B^2 can be considered as different at the 90% confidence level and the correlation parameter ρ is significantly greater than zero in both cases ($\alpha = 0$ and $\alpha = 1$). In the case $\alpha = 0$ the REML likelihood ratio statistic (LRS) for testing $H_0 : \sigma_A^2 = \sigma_B^2$ takes the value 1210.06 and its corresponding p-value is 0.00. In the case of $\alpha = 1$ the value of the REML LRS for testing $H_0 : \sigma_A^2 = \sigma_B^2$ is 1599.96 and the corresponding p-value is 0.00. In both cases we reject the null hypothesis of equality of variances. Therefore we can recommend model 2 for both poverty indicators.

Table 4 presents the normalized Euclidean distances between the direct and the EBLUPs estimates in both cases $\alpha = 0$ and $\alpha = 1$. We use the formula

$$D(\mathbf{y}_1, \mathbf{y}_2) = \left(\frac{1}{M} \sum_{d=1}^{D} \sum_{t=1}^{m_d} (y_{1dt} - y_{2dt})^2 \right)^{1/2}.$$

The obtained results are somehow expected. The models with more parameters present the lower normalized Euclidean distances. The extreme case would be a saturated model with as many parameters as observations, which has a perfect fit to data. As our target is explaining the data relationships, instead of looking for the best way of predicting the observed y-values, we do not modify our decision about model 3.

For being more confident about our decision of selecting model 2 as true generating model, we still give some diagnostics for models 0-2. At this stage, we drop out Model 3 from the selection procedure because of the hypotheses tested in Table 3.

Residuals $\hat{e}_{dt} = \bar{y}_{dt} - \bar{\mathbf{x}}_{dt}^{\mathsf{T}} \hat{\boldsymbol{\beta}} - \hat{u}_{dt}$ of fitted models 2, 1 and 0 are plotted against the observed values $\bar{y}_{dt}$ in the Figure 1 for $\alpha = 0$ (left) and $\alpha = 1$ (right). The dispersion graph shows a great difference in the pattern of the plots, passing from the basic model 0 to the more complex model 2. In particular, residuals of model 2 present a more flattened shape than the ones of the other two models. Figure 2 presents the boxplots of residuals of models 0-2 and also shows that partitioned models 1 and 2 fit much better to the data than model 0. This conclusion coincides with the results appearing in Table 4, where Euclidean distances decrease as moving from model 0 to model 2. So we conclude that model 2 fits better to the direct estimates and therefore we can recommend it.

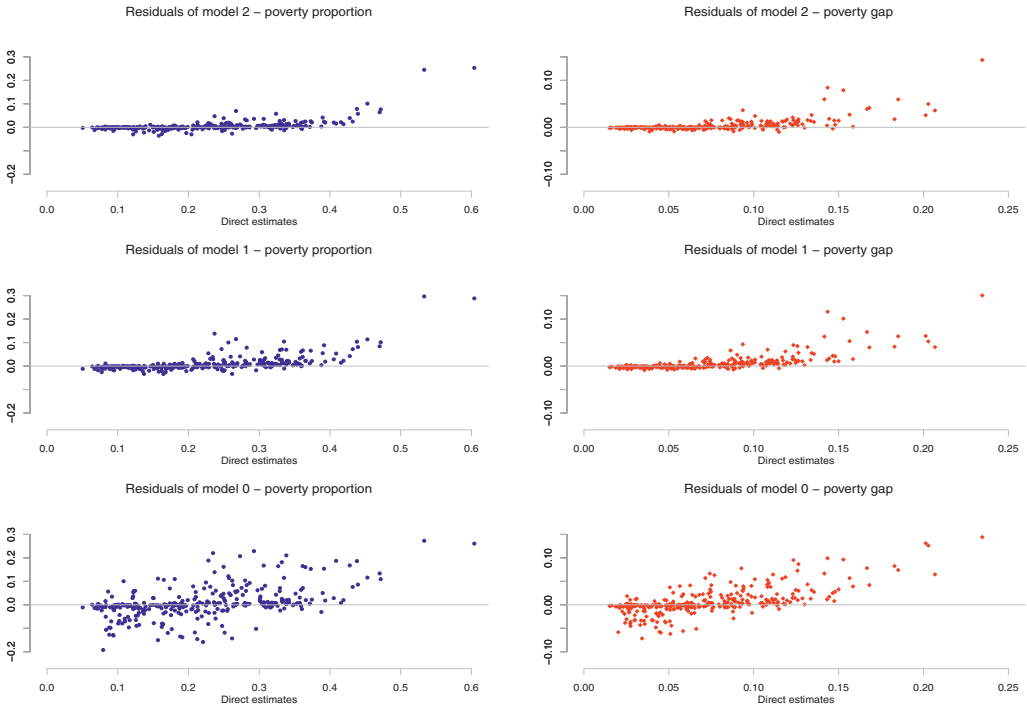


Figure 1: *Residuals versus direct estimates.*

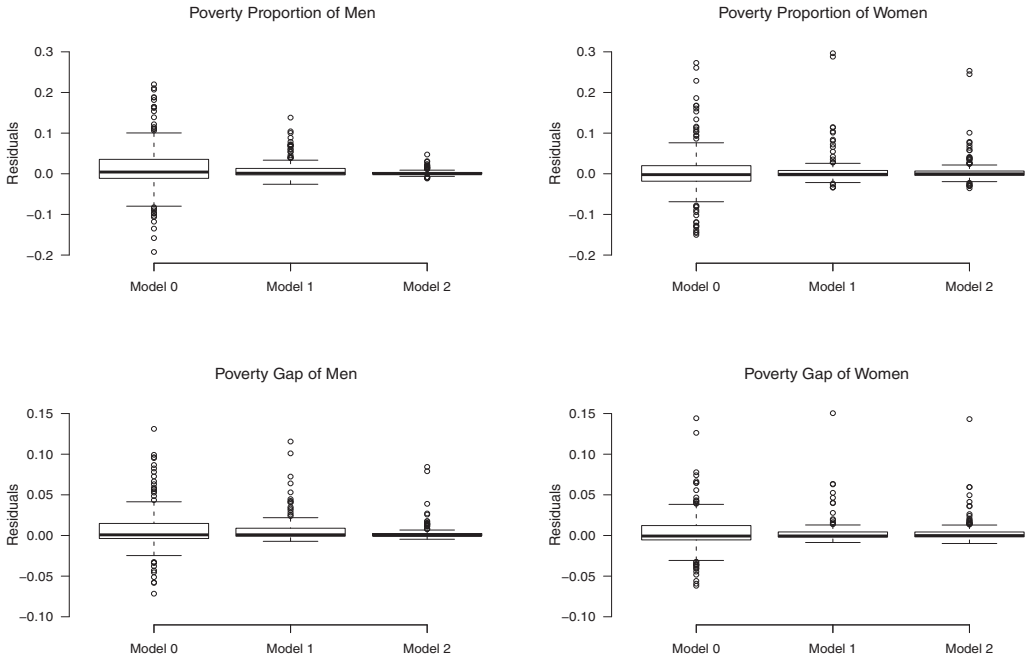


Figure 2: *Boxplots of residuals of models 0-2.*

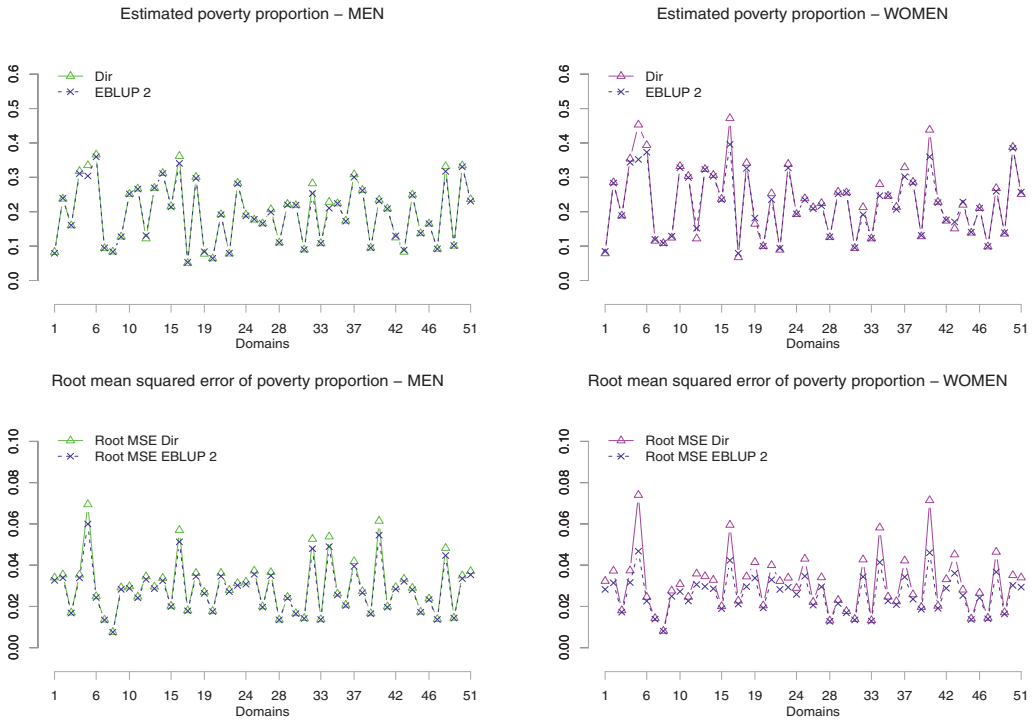


Figure 3: *Estimates of poverty proportion (top) and squared root of their estimated MSEs (bottom) respectively for men (on the left) and women (on the right) in 2006.*

The poverty proportion estimates, direct and EBLUP under model 2, are plotted in the Figure 3 with respect to the partition of the domains in men (left) and women (right). Figure 4 presents the same plots for the poverty gap. Concerning the root MSEs, these figures show that the EBLUPs under model 2 have lower MSE than the direct estimator. Therefore it is worthwhile using model-based estimators instead of the direct ones. As the estimated root MSE of the direct estimate of domain 42 is too large, Figure 3 does not plot the estimates of this domain and renumbers domains 43 to 52 as 42 to 51.

In the Figure 5 the Spanish provinces are plotted in 4 colored categories depending on the values of the EBLUP2 estimates in % of the poverty proportions and the gaps, i.e. $p_d = 100 \cdot \hat{\bar{Y}}_{0;d,2006}^{eblup2}$ and $g_d = 100 \cdot \hat{\bar{Y}}_{1;d,2006}^{eblup2}$. We observe that the Spanish regions where the proportion of the population under the poverty line is smallest are those situated in the north and east, like Cataluña, Aragón, Navarra, País Vasco, Cantabria and Baleares. On the other hand the Spanish regions with higher poverty proportion are those situated in the centre-south, like Andalucía, Extremadura, Murcia, Castilla La Mancha, Canarias, Ceuta and Melilla. In an intermediate position we can find regions that are in the centre-north of Spain, like Galicia, La Rioja, Castilla León, Asturias, Comunidad Valenciana and Madrid. If we investigate how far the annual net incomes of population under the

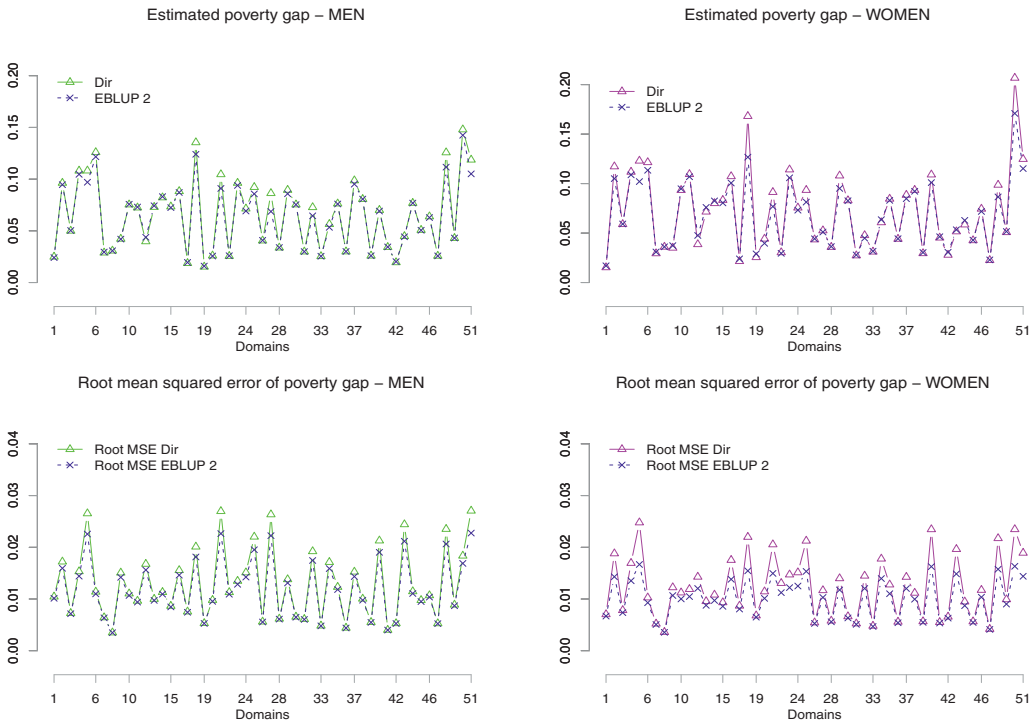


Figure 4: *Estimates of poverty gap (top) and squared root of their estimated MSEs (bottom) respectively for men (on the left) and women (on the right) in 2006.*

Table 5: *Estimated poverty proportions ($\alpha = 0$) and RMSE's in 2006.*

	Men					Women				
Province	n_d	DIR	EB_2	$RMSE_\star$	$RMSE_2$	n_d	DIR	EB_2	$RMSE_\star$	$RMSE_2$
Soria	24	0.247	0.231	0.107	0.080	18	0.604	0.351	0.126	0.057
Segovia	60	0.234	0.231	0.061	0.055	60	0.438	0.360	0.071	0.046
Palencia	73	0.228	0.210	0.054	0.049	72	0.280	0.246	0.058	0.041
Álava	98	0.083	0.079	0.034	0.033	100	0.079	0.085	0.032	0.028
Zamora	109	0.332	0.317	0.048	0.045	100	0.268	0.259	0.046	0.037
Huelva	124	0.192	0.191	0.036	0.035	124	0.253	0.235	0.040	0.033
Burgos	169	0.127	0.127	0.029	0.028	167	0.124	0.129	0.028	0.025
Albacete	173	0.237	0.239	0.035	0.034	193	0.285	0.283	0.037	0.031
Granada	189	0.301	0.297	0.036	0.035	229	0.342	0.326	0.034	0.030
Crdoba	221	0.312	0.311	0.034	0.032	233	0.307	0.303	0.033	0.029
Cáceres	261	0.252	0.252	0.030	0.029	303	0.332	0.328	0.031	0.027
Tenerife	373	0.263	0.262	0.027	0.027	397	0.286	0.283	0.026	0.024
Sevilla	473	0.209	0.209	0.020	0.020	492	0.228	0.227	0.020	0.019
Zaragoza	556	0.101	0.101	0.014	0.014	577	0.136	0.139	0.017	0.016
Barcelona	1367	0.083	0.084	0.008	0.008	1494	0.108	0.109	0.008	0.008

Table 6: *Estimated poverty gapss ($\alpha = 1$) and RMSE's for men in 2006.*

	Men					Women				
Province	n_d	DIR	EB_2	$RMSE_\star$	$RMSE_2$	n_d	DIR	EB_2	$RMSE_\star$	$RMSE_2$
Soria	24	0.153	0.074	0.088	0.038	18	0.235	0.091	0.111	0.023
Segovia	60	0.070	0.069	0.021	0.019	61	0.123	0.102	0.025	0.017
Palencia	73	0.056	0.053	0.017	0.016	78	0.052	0.053	0.020	0.015
lava	98	0.025	0.024	0.010	0.010	87	0.107	0.101	0.018	0.014
Zamora	109	0.126	0.112	0.024	0.021	100	0.099	0.087	0.022	0.016
Huelva	124	0.105	0.091	0.027	0.023	124	0.091	0.077	0.021	0.015
Burgos	169	0.042	0.042	0.015	0.014	165	0.089	0.085	0.014	0.012
Albacete	173	0.096	0.095	0.017	0.016	181	0.053	0.051	0.012	0.010
Granada	189	0.135	0.124	0.020	0.018	194	0.112	0.109	0.017	0.014
Crdoba	221	0.082	0.083	0.011	0.011	230	0.114	0.106	0.015	0.012
Cceres	261	0.075	0.076	0.011	0.011	247	0.207	0.171	0.023	0.016
Tenerife	373	0.081	0.081	0.010	0.010	397	0.093	0.092	0.011	0.010
Sevilla	473	0.034	0.034	0.004	0.004	501	0.043	0.044	0.005	0.005
Zaragoza	556	0.043	0.043	0.009	0.009	605	0.027	0.028	0.005	0.005
Barcelona	1367	0.031	0.031	0.003	0.003	1494	0.036	0.036	0.004	0.004

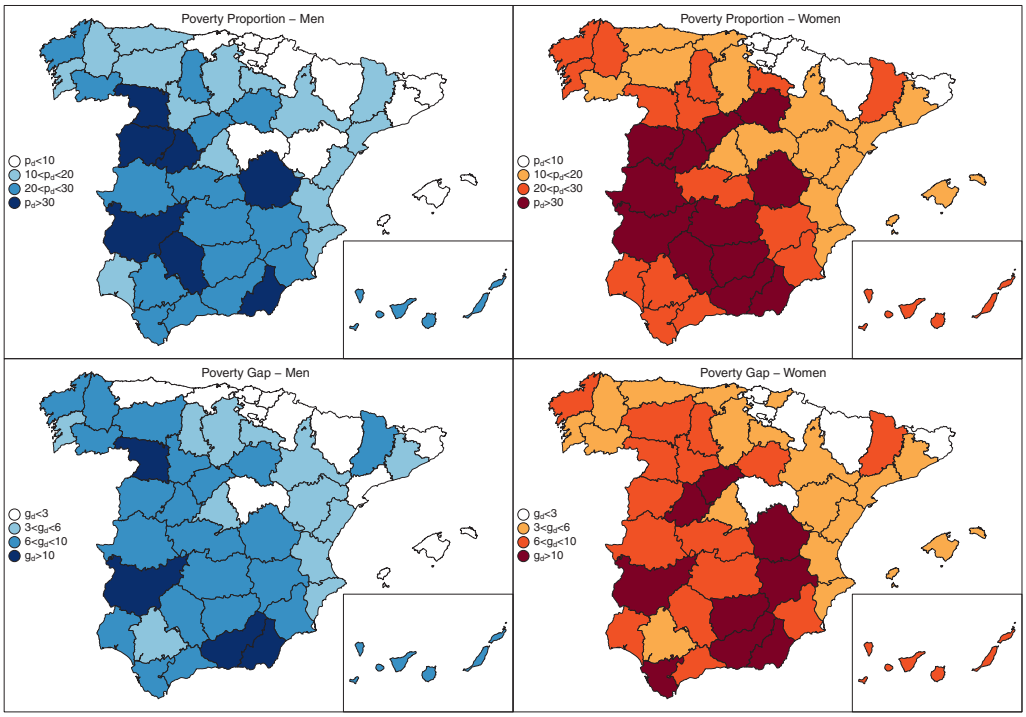


Figure 5: *EBLUP2 estimates of poverty proportions (top) and gaps (bottom) for men (left) and women (right) in 2006.*

poverty line z_{2006} are from z_{2006}, we observe that in the Spanish regions situated in the centre-north there exist a distance that is generally lower than the 6% of z_{2006}. However, the cited distance is in general greater than 6% of z_{2006} in the centre-south.

Tables 5-6 present the direct and EBLUP estimates under model 2 of poverty proportions ($\alpha = 0$) and poverty gaps ($\alpha = 1$) for some Spanish provinces. The provinces were selected accordingly with the quantiles of the set of domain sample sizes n_d. The EBLUP estimates under the model 2 are labelled by EB_2 and the direct estimates by DIR. The squared root of MSEs are labelled by $\mathrm{RMSE}_\star$ for the direct estimator and by RMSE_2 for the EBLUP under the model 2 respectively. Numerical results are sorted by sex. Regarding the reduction of the MSE when passing from direct to EBLUP estimates, we observe that model 2 performs better in domains with small sample size.

4. Discussion

As poverty indicators are nonlinear, unit-level model-based estimation approaches cannot always be used. However, their direct estimators are weighted sums that can be modelled by area-level models. Area-level models thus provide an easy-to-apply solution. These idea motivates the introduction of partitioned temporal models that borrow strength from time. The use of information from past time instants, the greater availability of auxiliary variables at the domain level and the possibility of introducing modelling differences by sex might compensate the loss of information when passing from unit-level models to area-level models. We thus considered four area-level linear mixed models and we applied the methodology to Spanish EU-SILC data.

We would also like to point out that model (1) and its particularizations have some features of interest, from a methodological point of view. It is somewhat different from the Rao-Yu model (Rao and Yu, 1994, and Rao, 2003), viewed as an extension of the Fay-Herriot area-level model in the case of time-correlated data. As we can note, the covariance matrix of the model does not contain the variance component connected with the random-effect at the domains, as clusters of time-correlated data. This fact permits to the random time-area effect to absorb completely the variation of the EBLUP due to the correlated observation, without considering any cluster-oriented random-effect components.

Another characteristic of main interest of the model (1), is that is a "partitioned" model. This means that different variance components in the covariance matrix of the random-area effects can accommodate different inputs of information, due to some relevant issues related to the specific levels of auxiliary variables. In the case of the application on the poverty indicators in Spain, the partitioning of the variance of the random-effect is significative for the gender-based class of survey domains. In fact, relevant differences in terms of the data in these classes of domains, as inputs in the fixed-effects regression, seems to drive at the same time to different variations in the related class of random-area effects.

The R programming language has been employed for doing all the computations in this paper. The deliverable D22 on software for small area estimation of the European SAMPLE project (http://www.sample-project.eu/) gives a primary version of the employed R codes.

References

Das, K., Jiang, J. and Rao, J. N. K. (2004). Mean squared error of empirical predictor. Mean squared error of empirical predictor. *The Annals of Statistics*, 32, 818–840.

Datta, G. S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 10, 613–627.

Esteban, M. D., Morales, D., Pérez, A. and Santamaría, L. (2012a). Small area estimation of poverty proportions under area-level time models. *Computational Statistics and Data Analysis*, 56, 2840–2855.

Esteban, M. D., Morales, D., Pérez, A. and Santamaría, L. (2012b). Two area-level time models for estimating small area poverty indicators. *Journal of the Indian Society of Agricultural Statistics*, 66, 75–89.

Foster, J., Greer, J. and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52, 761–766.

Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, 9, 55–93.

Herrador, M., Esteban, M. D., Hobza, T. and Morales, D. (2011). A Fay-Herriot model with different random effect variances. *Communications in Statistics (Theory and Methods)*, 10, 785–797.

Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, 15, 1–96.

Marhuenda, Y., Molina, I. and Morales, D. (2013). Small area estimation with spatio-temporal Fay-Herriot models. *Computational Statistics and Data Analysis*, 58, 308–325.

Pfeffermann, D. (2002). Small area estimation-new developments and directions. *International Statistical Review*, 70, 125–143.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, 28, 1–134.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85, 163–171.

Rao, J. N. K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, 25, 175–186.

Rao, J. N. K. (2003). *Small Area Estimation*. John Wiley.

Rao, J. N. K. and Yu, M. (1994). Small area estimation by combining time series and cross-sectional data. *Canadian Journal of Statistics*, 22, 511–528.

Särndal, C. E., Swensson, B. and Wretman, J. (1992) *Model Assisted Survey Sampling*. Springer-Verlag.

SORT 39 (1) January-June 2015, 35-50

A new class of Skew-Normal-Cauchy distribution

Jaime Arrué[1], Héctor W. Gómez[2], Hugo S. Salinas[3] and Heleno Bolfarine[4]

Abstract

In this paper we study a new class of skew-Cauchy distributions inspired on the family extended two-piece skew normal distribution. The new family of distributions encompasses three well known families of distributions, the normal, the two-piece skew-normal and the skew-normal-Cauchy distributions. Some properties of the new distribution are investigated, inference via maximum likelihood estimation is implemented and results of a real data application, which reveal good performance of the new model, are reported.

MSC: 60E05, 62F12

Keywords: Cauchy distribution, kurtosis, maximum likelihood estimation, singular information matrix, skewness, Skew-Normal-Cauchy distribution

1. Introduction

Arnold et al. (2009) introduced a random variable $X \sim ETN(\alpha,\beta)$ with probability density function given by:

$$f_{ETN}(x;\alpha,\beta) = 2c_\alpha \phi(x)\Phi(\alpha|x|)\Phi(\beta x), \quad -\infty < x < \infty, \tag{1}$$

where $\alpha,\beta \in \mathbb{R}$, $c_\alpha = 2\pi/(\pi + 2\arctan(\alpha))$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and cumulative distribution functions of the standard $N(0,1)$ distribution, respectively.

[1] Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile. jaime.arrue@uantof.cl

[2] Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile. hector.gomez@uantof.cl

[3] Departamento de Matemáticas, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile. hugo.salinas@uda.cl

[4] Departamento de Estatística, IME, Universidad de Sao Paulo, Sao Paulo, Brasil. hbolfar@ime.usp.br

Received: October 2013
Accepted: September 2014

Notice that for the particular case $\alpha = 0$ the well known skew-normal distribution (Azzalini, 1985) with density function given by

$$f_{SN}(x;\beta) = 2\phi(x)\Phi(\beta x), \quad -\infty < x < \infty, \tag{2}$$

is obtained. For $\beta = 0$, one obtains the so called two-piece skew-normal distribution given by Kim (2005), denoted by $\{TN(\alpha) : -\infty < \alpha < \infty\}$ with probability density function given by

$$f_{TN}(x;\alpha) = c_\alpha \phi(x)\Phi(\alpha|x|), \quad -\infty < x < \infty, \tag{3}$$

with c_α as the normalizing constant. Another family of models studied in Nadarajah and Kotz (2003), is generated by using the kernel of the normal distribution, that is,

$$h(x;\lambda) = 2\phi(x)G(\beta x), \quad -\infty < x < \infty, \tag{4}$$

with $\beta \in (-\infty,\infty)$ and $G(\cdot)$ is a symmetric distribution function. A particular case of this class follows by taking $G(\cdot)$ as the CDF of the Cauchy distribution, which as shown by Nadarajah and Kotz (2003), results in a model with the same range of asymmetry, but with greater kurtosis than that of the skew-normal model. The pdf for a random variable X with this distribution, which we denote by $X \sim SNC(\beta)$, can be written as

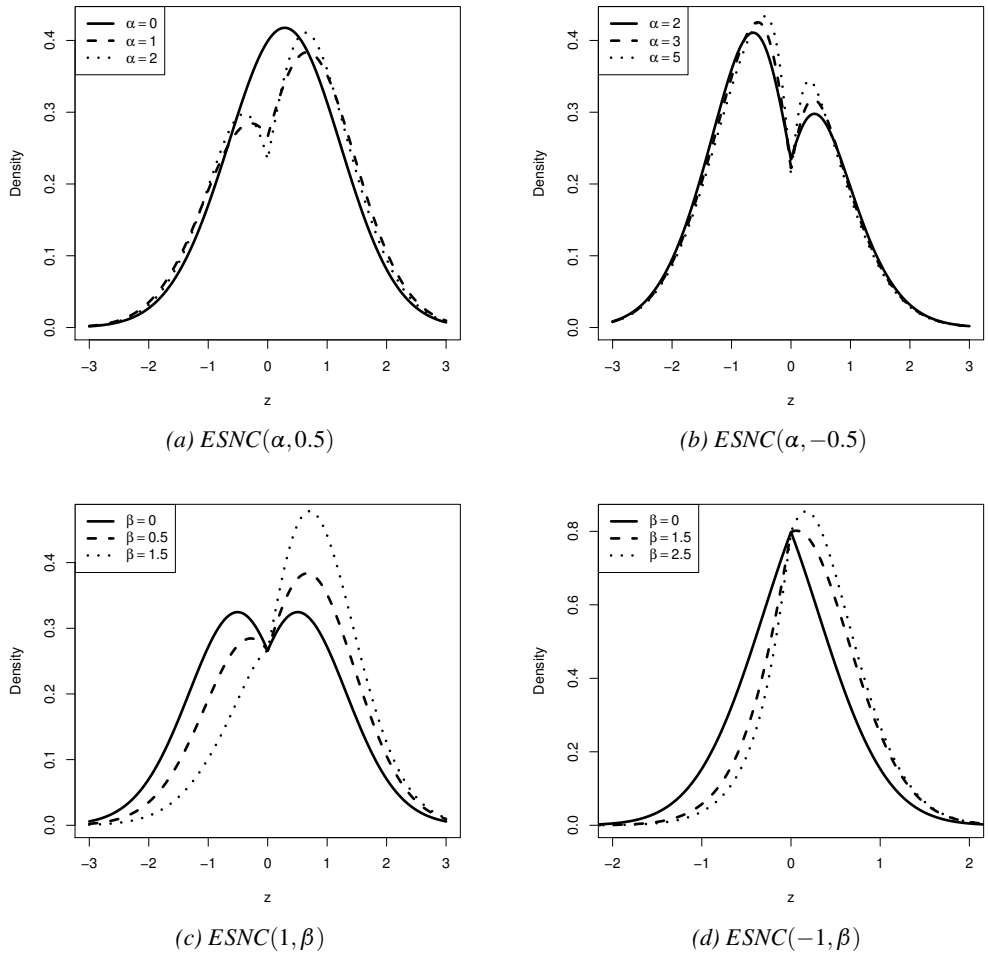
$$f_{SNC}(x;\beta) = 2\phi(x)\left\{\frac{1}{2} + \frac{1}{\pi}\arctan(\beta x)\right\}, \quad -\infty < x < \infty. \tag{5}$$

Arrué, Gómez, Varela and Bolfarine (2010) studied some properties, stochastic representation and information matrix for the model given in (5). A random variable Z has a extended skew-normal-Cauchy random variable with parameter $\alpha,\beta \in (-\infty,\infty)$, denoted $Z \sim ESNC(\alpha,\beta)$, if its probability density function is

$$f(z;\alpha,\beta) = 2c_\alpha \phi(z)\Phi(\alpha|z|)\left\{\frac{1}{2} + \frac{1}{\pi}\arctan(\beta z)\right\}, \quad -\infty < z < \infty. \tag{6}$$

For the rest of the article, Z will denote a random variable with density (6). Figures 1 depicts shapes of density function (6) for different parameter values (continuous and discontinuous lines).

This model is important because it contains strictly (not as limiting cases) the normal, SNC and TN distributions. Moreover, this distribution inherits the bimodal nature of the TN model which is controlled by parameter α, that is, when $\alpha > 0$ the model is bimodal and when $\alpha < 0$ it is unimodal. Since it contains the Cauchy distribution, greater flexibility in the kurtosis is earned and therefore could better fit data sets containing outlying observations.



(a) $ESNC(\alpha, 0.5)$ *(b)* $ESNC(\alpha, -0.5)$

(c) $ESNC(1, \beta)$ *(d)* $ESNC(-1, \beta)$

Figure 1: *Examples of the ESNC density.*

One of the main focus of the paper is to develop a stochastic representation of the ESNC model which allows moments derivation in a simpler way. We derive also the Fisher information matrix for the ESNC model and show that it is singular for $\alpha = \beta = 0$. Using the approach in Rotnitzky, Cox, Bottai and Robins (2000), an alternative parametrization is proposed which makes the Fisher information matrix nonsingular at $\alpha = \beta = 0$.

The paper is organized as follows. Section 2 presents properties of the ESNC model. Section 3 presents a stochastic representation for this model which allows a simple derivation for the moments generating function leading to simple expressions for asymmetry and kurtosis coefficients. The Fisher information matrix is derived in Section 4.2, which turns out to be singular for $\alpha = \beta = 0$. A parametrization is studied

which makes it nonsingular for $\alpha = \beta = 0$ and allow an asymptotic study of the MLE properties at this point. In Section 5 we use a data set to illustrate the flexibility of the model ESNC, for this we use the maximum likelihood approach and compare it with the TN and SNC models. The paper is concluded with a discussion section.

2. Distributional properties of the ESNC model

Clearly, density (6) is continuous at $z = 0$ for all α and β, However, it is not differentiable at $z = 0$ for $\alpha \neq 0$. In the following we present uni/bimodal properties possessed by the ESNC family. Notice that this model contains the normal, two-piece skew-normal and skew-normal-cauchy as special cases. The following properties follow immediately from the (6).

Property 1 *The $ESNC(0,0)$ density is the $N(0,1)$ density.*

Property 2 *The $ESNC(0,\beta)$ density is the $SNC(\beta)$ density.*

Property 3 *The $ESNC(\alpha,0)$ density is the $TN(\alpha)$ density.*

Property 4 *As $\alpha \to \infty$, $f(z;\alpha,\beta)$ tends to the $SNC(\beta)$ density. In contrast, as $\alpha \to -\infty$, $f(z;\alpha,\beta)$ degenerates at 0.*

Property 5 *As $\beta \to \infty$, $f(z;\alpha,\beta)$ tends to the $2c_\alpha\phi(z)\Phi(\alpha z)I(z \geq 0)$ density. In contrast, as $\beta \to -\infty$, $f(z;\alpha,\beta)$ tends to the $2c_\alpha\phi(z)\Phi(-\alpha z)I(z < 0)$ density.*

Property 6 *If $Z \sim ESNC(\alpha,\beta)$ random variable, then $-Z \sim ESNC(\alpha,-\beta)$ random variable.*

Property 7 *For $\alpha > 0$, the density (6) is bimodal, i.e. in each region of $z \in (-\infty,0]$ and $z \in [0,\infty)$, $\log f(z;\alpha,\beta)$ is a concave function of z.*

Property 8 *For $\alpha > 0$, the two modes of (6) are located at $z = z_0$ and $z = z_1$ satisfying*

$$z_0 = -\alpha\frac{\phi(\alpha z_0)}{\Phi(-\alpha z_0)} + \frac{\beta}{\pi(1+\beta^2 z_0^2)} \quad \textit{and} \quad z_1 = \alpha\frac{\phi(\alpha z_1)}{\Phi(\alpha z_1)} + \frac{\beta}{\pi(1+\beta^2 z_1^2)},$$

where $z_0 < 0$ and $z_1 > 0$.

Property 9 *For $\alpha < 0$, the single mode of (6) is located at $z = 0$, because $f'(z;\alpha,\beta) < 0$ for $z > 0$ and $f'(z;\alpha,\beta) > 0$ for $z < 0$.*

3. A stochastic representation

The main result states that if $Z \sim ESNC(\alpha, \beta)$ then the distribution of Z can be obtained as a mixture in the asymmetry parameter between the extended two-piece skew-normal and half-normal (HN) distributions. In the following $I(A)$ denotes the indicator function of the set A.

Proposition 1 *If* $Z|Y = y \sim ETN(\alpha, \beta y)$ *and* $Y \sim HN(0,1)$ *then* $Z \sim ESNC(\alpha, \beta)$.

Proof. Let $Z|Y = y \sim ETN(\alpha, \beta y)$ and $Y \sim 2\phi(y)I(y \geq 0)$, then

$$\begin{aligned} f(z;\alpha,\beta) &= \int_0^\infty 2c_\alpha \phi(z)\Phi(\alpha|z|)\Phi(\beta yz)2\phi(y)dy \\ &= 2c_\alpha \phi(z)\Phi(\alpha|z|)\int_0^\infty 2\Phi(\beta yz)\phi(y)dy \\ &= 4c_\alpha \phi(z)\Phi(\alpha|z|)\int_0^\infty \int_{-\infty}^{\beta z} \phi(t)\phi(y)dtdy \\ &= 4c_\alpha \phi(z)\Phi(\alpha|z|)\left[\int_0^\infty \int_{-\infty}^{0} \phi(t)\phi(y)dtdy + \int_0^\infty \int_0^{\beta z} \phi(t)\phi(y)dtdy\right] \end{aligned}$$

The terms $\int_0^\infty \int_{-\infty}^{0} \phi(t)\phi(y)dtdy$ and $\int_0^\infty \int_0^{\beta x} \phi(t)\phi(y)dtdy$ are the integrals of the bivariate normal distribution. Then, making changes in variables $t = r\cos u$ and $y = r\sin u$ we have

$$\begin{aligned} f(z;\alpha,\beta) &= 4c_\alpha \phi(z)\Phi(\alpha|z|)\left[\frac{1}{4} + \frac{1}{2\pi}\int_0^{\arctan(\beta z)} \int_0^{\frac{\pi}{2}} e^{-r^2/2} r dr du\right] \\ &= 2c_\alpha \phi(z)\Phi(\alpha|z|)\left\{\frac{1}{2} + \frac{1}{\pi}\arctan(\beta z)\right\}, \end{aligned}$$

which concludes the proof. ∎

3.1. Location and scale extension

For applications it is convenient to add location and scale parameters to the ESNC distribution. If $Z \sim ESNC(\alpha, \beta)$ and if $X = \mu + \sigma Z$, where $\mu \in (-\infty, \infty)$ and $\sigma > 0$, then we can write $X \sim ESNC(\mu, \sigma, \alpha, \beta)$ or, at times, $X \sim ESNC(\theta)$ where $\theta = (\mu, \sigma, \alpha, \beta)$'. This leads to the following definition.

Definition 1 *A random variable X has a distribution in the ESNC location and scale family if the density is given by*

$$f(x;\theta)=\frac{2c_\alpha}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\alpha\left|\frac{x-\mu}{\sigma}\right|\right)\left\{\frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{\beta(x-\mu)}{\sigma}\right)\right\},\quad -\infty<x<\infty. \tag{7}$$

We write $X\sim ESNC(\theta)$ *or* $X\sim ESNC(\mu,\sigma,\alpha,\beta)$.

3.2. Moments

In order to evaluate moments of the ESNC distribution, the following technical propositions will be useful. In these propositions, we use the notation

$$a_r(\alpha,\lambda):=\int_0^\infty 2c_\alpha t^r\phi(t)\Phi(\lambda t)dt, \tag{8}$$

and

$$d_r(\alpha,\beta):=\int_0^\infty 2c_\alpha t^r\phi(t)\Phi(\alpha t)\Phi(\beta t)dt, \tag{9}$$

where $\alpha,\beta,\lambda\in(-\infty,\infty)$.

We provide next the recursive formulation for computing the functions above for a random variable with density given in (6) which will be fundamental for computing moments of the random variable $X\sim ESNC(\theta)$. The proof is presented in Arnold et al. (2009).

Proposition 2 *According to (8),*

$$a_r(\alpha,\lambda)=\begin{cases}\frac{\pi+2\arctan(\lambda)}{\pi+2\arctan(\alpha)}, & r=0,\\ \frac{c_\alpha}{\sqrt{2\pi}}\left(1+\frac{\lambda}{\sqrt{1+\lambda^2}}\right), & r=1,\\ (r-1)a_{r-2}(\alpha,\lambda)+\frac{2^{r/2-1}\lambda c_\alpha}{\pi(1+\lambda^2)^{r/2}}\Gamma\left(\frac{r}{2}\right), & r\geq 2,\end{cases} \tag{10}$$

Proposition 3 *Let* $U\sim TSN(\alpha)$*, then*

$$a_r(\alpha):=E(U^r)=\begin{cases}1, & r=0,\\ \frac{c_\alpha}{\sqrt{2\pi}}\left(1+\frac{\alpha}{\sqrt{1+\alpha^2}}\right), & r=1,\\ (r-1)a_{r-2}(\alpha)+\frac{2^{r/2-1}\alpha c_\alpha}{\pi(1+\alpha^2)^{r/2}}\Gamma\left(\frac{r}{2}\right), & r\geq 2.\end{cases} \tag{11}$$

This result is obtained for $\lambda=\alpha$ in Equation (10).

Proposition 4 *According to (9),*

$$d_r(\alpha,\beta)=\begin{cases}\int_0^\infty 2c_\alpha\phi(t)\Phi(\alpha t)\Phi(\beta t)dt, & r=0,\\ \frac{c_\alpha}{\sqrt{2\pi}}\left[\frac{1}{2}+\frac{\alpha}{\sqrt{1+\alpha^2}}\Psi\left(\frac{\beta}{\sqrt{1+\alpha^2}}\right)+\frac{\beta}{\sqrt{1+\beta^2}}\Psi\left(\frac{\alpha}{\sqrt{1+\beta^2}}\right)\right], & r=1,\\ (r-1)d_{r-2}(\alpha,\beta)+\frac{\alpha}{\sqrt{2\pi}(1+\alpha^2)^{r/2}}a_{r-1}(\alpha,\lambda_1)+\frac{\beta}{\sqrt{2\pi}(1+\beta^2)^{r/2}}a_{r-1}(\alpha,\lambda_2), & r\geq 2,\end{cases} \tag{12}$$

where $\Psi(t)=1/2+\arctan(t)/\pi$ is a CDF of the standard Cauchy distribution, $\lambda_1=\beta/\sqrt{1+\alpha^2}$, $\lambda_2=\alpha/\sqrt{1+\beta^2}$ and $d_0(\alpha,\beta)$ must be evaluated numerically.

Proposition 5 *Let $Z\sim ESNC(\alpha,\beta)$, $Y\sim 2\phi(y)I(y\geq 0)$ and $X=\mu+\sigma Z\sim ESNC(\theta)$ so that, for $r=1,2,\ldots$, we have:*

$$E(Z^r)=(1-(-1)^r)E(d_r(\alpha,\beta Y))+(-1)^r a_r(\alpha)\quad and\quad E(X^r)=\sum_{k=0}^{r}\binom{r}{k}\mu^{r-k}\sigma^k E(Z^k), \tag{13}$$

where $a_r(\alpha)$ and $d_r(\alpha,\cdot)$ are given in (11) and (12), respectively.

Proof. For computing moments of the random variable $Z\sim ESNC(\alpha,\beta)$ we use conditional expectations and the stochastic representation given in Proposition 1, leading to

$$\begin{aligned}E(Z^r)=E(E(Z^r|Y)) &= \int_0^\infty[(1-(-1)^r)d_r(\alpha,\beta y)+(-1)^r a_r(\alpha)]2\phi(y)dy\\ &= (1-(-1)^r)\int_0^\infty 2d_r(\alpha,\beta y)\phi(y)dy+(-1)^r a_r(\alpha)\int_0^\infty 2\phi(y)dy\\ &= (1-(-1)^r)\int_0^\infty 2d_r(\alpha,\beta y)\phi(y)dy+(-1)^r a_r(\alpha)\\ &= (1-(-1)^r)E(d_r(\alpha,\beta Y))+(-1)^r a_r(\alpha).\end{aligned}$$

■

Corollary 1 *If $Z\sim ESNC(\alpha,\beta)$, then*

$$E(Z^r)=\begin{cases}a_r(\alpha), & r\ even,\\ 2k_r(\alpha,\beta)-a_r(\alpha), & r\ odd,\end{cases} \tag{14}$$

where $k_r(\alpha,\beta):=E(d_r(\alpha,\beta Y))$.

The even moments of the ESNC distribution coincide with the even moments of the ETN distribution given by Arnold et al. (2009).

In the following we present expressions for computing $k_r(\alpha,\beta)$ when r is odd. The proofs for the results presented next follow directly from (10) and (12).

Proposition 6 *Under the conditions in Proposition 5, we have*

$$k_r(\alpha,\beta)=\begin{cases}\frac{2c_\alpha}{\sqrt{2\pi}}\left[\frac{1}{4}+\frac{\alpha}{\sqrt{1+\alpha^2}}\int_0^\infty \phi(y)\Psi\left(\frac{\beta y}{\sqrt{1+\alpha^2}}\right)dy \right. & r=1,\\ \left.+\,\beta\int_0^\infty \frac{y\phi(y)}{\sqrt{1+\beta^2y^2}}\Psi\left(\frac{\alpha}{\sqrt{1+\beta^2y^2}}\right)dy\right], & \\ (r-1)k_{r-2}(\alpha,\beta)+\frac{\alpha}{\sqrt{2\pi}(1+\alpha^2)^{r/2}}\,g_{r-1}(\alpha,\beta)+\frac{\beta}{\sqrt{2\pi}}\,j_{r-1}(\alpha,\beta), & r=3,5,\ldots\end{cases} \tag{15}$$

where

$$g_r(\alpha,\beta)=\begin{cases}2c_\alpha\int_0^\infty \phi(y)\Psi\left(\frac{\beta y}{\sqrt{1+\alpha^2}}\right)dy, & r=0,\\ (r-1)g_{r-2}(\alpha,\beta)+\frac{\beta^{1-r}c_\alpha e^{\frac{1+\alpha^2}{2\beta^2}}}{\sqrt{2\pi^3}(1+\alpha^2)^{(1-r)/2}}\Gamma\left(\frac{r}{2}\right)\Gamma\left(1-\frac{r}{2},\frac{1+\alpha^2}{2\beta^2}\right), & r=2,4,\ldots\end{cases} \tag{16}$$

where $\Gamma(a,z)=\int_z^\infty e^{-t}t^{a-1}dt$ *is the incomplete Gamma function.*

$$j_r(\alpha,\beta)=\begin{cases}2c_\alpha\int_0^\infty \frac{y\phi(y)}{(1+\beta^2y^2)}\Psi\left(\frac{\alpha}{\sqrt{1+\beta^2y^2}}\right)dy, & r=0,\\ (r-1)j_{r-2}(\alpha,\beta)+\frac{\alpha c_\alpha\Gamma\left(\frac{r}{2}\right)}{2^{-r/2}\pi}\int_0^\infty \frac{y\phi(y)dy}{(1+\alpha^2+\beta^2y^2)^{r/2}\sqrt{1+\beta^2y^2}}, & r=2,4,\ldots\end{cases} \tag{17}$$

The terms $k_r(\alpha,\beta)$, $g_r(\alpha,\beta)$ and $j_r(\alpha,\beta)$ can be calculated using numerical integration for r, α and β. For reference we list the first four moments of the standard ESNC distribution. If $Z\sim ESNC(\alpha,\beta)$ then

$$\begin{aligned}E(Z) &= \frac{c_\alpha}{\sqrt{2\pi}}\left(-\frac{\alpha}{\sqrt{1+\alpha^2}}+\frac{4\alpha}{\sqrt{1+\alpha^2}}\int_0^\infty \phi(y)\Psi\left(\frac{\beta y}{\sqrt{1+\alpha^2}}\right)dy\right. \\ &\quad \left.+\;4\beta\int_0^\infty \frac{y\phi(y)}{\sqrt{1+\beta^2y^2}}\Psi\left(\frac{\alpha}{\sqrt{1+\beta^2y^2}}\right)dy\right), \end{aligned} \tag{18}$$

$$E(Z^2) = 1+\frac{\alpha c_\alpha}{\pi(1+\alpha^2)}, \tag{19}$$

$$E(Z^3) = 2k_3(\alpha,\beta)-\frac{c_\alpha}{\sqrt{2\pi}}\left(2+\frac{2\alpha}{\sqrt{1+\alpha^2}}+\frac{\alpha}{\sqrt{(1+\alpha^2)^3}}\right), \tag{20}$$

$$E(Z^4) = 3+\frac{\alpha c_\alpha(5+3\alpha^2)}{\pi(1+\alpha^2)^2}, \tag{21}$$

Standard expressions for kurtosis and skewness can then be obtained using Equations $(18)-(21)$.

4. ML estimation

4.1. Likelihood

Suppose that we have available a sample of size n, $X_1, X_2, \ldots, X_n$ from an $ESNC(\theta)$ distribution. In principle, the representation $X_i = \mu + \sigma Z_i$ and the four moment expressions for Z given in Equations $(18)-(21)$ could be used to obtain method of moments estimates of the four parameters. However, the approach is not pursued further. Instead, we will discuss the implementation of the maximum likelihood approach for this distribution given that it is more efficient asymptotically. The log-likelihood function of a random sample $(X_1, X_2, \ldots, X_n)$ from an $ESNC(\theta)$ distribution takes the form

$$l(\theta; X_1, X_2, \ldots, X_n) \propto n \log\left(\frac{c_\alpha}{\sigma}\right) - \frac{1}{2}\sum_{i=1}^{n} Z_i^2 + \sum_{i=1}^{n} \log \Phi(\alpha |Z_i|) + \sum_{i=1}^{n} \log \Psi(\beta Z_i), \quad (22)$$

Table 2 (see Appendix) shows the average MLEs of μ, σ, α and β for 1000 random of size n (SD: standard deviation for the 1000 estimates). We do not consider the case of $\beta < 0$, since by the reflection property 2.6, if $X \sim ESNC(0,1,\alpha,-\beta)$ then $-X \sim ESNC(0,1,\alpha,\beta)$. Several parameter values are considered and moderate and large sample sizes are used. The table shows that for large values of α and β the, MLEs tend to overestimate (if positive) the true values of α and β. This overestimation decreases as the true parameter values decrease and as sample size increases. If one wants to reduce the asymptotic bias of the MLEs one can apply the correction approach in Firth (1993), which amounts to penalize the likelihood for a MLE with less bias value.

4.2. The Fisher information matrix

4.2.1. Special cases

In the special case where $\alpha = 0$ and $\beta = 0$ the information matrix for the ESNC model (see Appendix) is singular, that is,

$$\left|I_{(\mu,\sigma,0,0)}\right| = \begin{vmatrix} \frac{1}{\sigma^2} & 0 & 0 & \frac{2}{\pi\sigma} \\ 0 & \frac{2}{\sigma^2} & \frac{2}{\pi\sigma} & 0 \\ 0 & \frac{2}{\pi\sigma} & \frac{2(\pi-2)}{\pi^2} & 0 \\ \frac{2}{\pi\sigma} & 0 & 0 & \frac{4}{\pi^2} \end{vmatrix} = 0.$$

Comparing the above information matrix with the Fisher information matrix corresponding to model $SNC(\beta)$ given in Arrué et al. (2010) we note that they differ only in the row and column corresponding to the second derivative with respect to the parameter α. The columns corresponding to the parameters μ and β are linearly dependent, so the information matrix is singular. This difficulty has been noticed and investigated in Azzalini (1985) in the context of the skew-normal distribution and was later studied in Chiogna (2005) in some other contexts. DiCiccio and Monti (2004) studied this singularity problem in the context of the skew-exponential power distribution and Salinas, Arellano-Valle and Gómez (2007) studied it in the context of the extended skew-exponential power distribution. In summary, for this special case when the parameters α and β tend to zero, we could not perform asymptotic statistical inference on these parameters, since the information matrix is singular. And to overcome this problem, we will use a reparametrization given by Rotnitzky et al. (2000), which is to transform the score function S_β in one that is linearly independent from the other score functions of score (for the other parameters). With this procedure we obtain a nonsingular information matrix.

4.2.2. Nonsingular Fisher information matrix

As considered in Arrué et al. (2010), we consider next a parameter transformation that makes the information matrix nonsingular. Indeed, after extensive algebraic manipulations, by using the approach in Rotnitzky et al. (2000), it follows that the convenient parametrization is the same as the one derived in Arrué et al. (2010) for the SNC model, namely,

$$\mu^* = \mu + \frac{2}{\pi}\sigma\beta, \quad \sigma^* = \sigma\left(1 - \frac{2}{\pi^2}\beta^2\right), \quad \alpha^* = \alpha, \quad \beta^* = \beta$$

and hence the score vector obtained is $\left(S_\mu, S_\sigma, S_\alpha, S_\beta^3/3!\right)$ where

$$S_\beta^3 = \left.\frac{\partial^3 l(\theta; X^*)}{\partial \beta^3}\right|_{\alpha=\beta=0}.$$

Therefore, to obtain the transformed Fisher information matrix, we have to compute

$$E\left(S_\mu \frac{S_\beta^3}{3!}\right) = \frac{-2}{\pi\sigma}$$
$$E\left(S_\sigma \frac{S_\beta^3}{3!}\right) = 0$$

$$E\left(S_\alpha \frac{S_\beta^3}{3!}\right) = 0$$

$$E\left(\left(\frac{S_\beta^3}{3!}\right)^2\right) = \frac{4}{9\pi^6}\left[96+\pi^2\left(\frac{1}{\sigma^2}-48\right)+15\pi^4\right]$$

leading to the nonsingular Fisher information matrix for the ESNC model

$$I_{(\mu,\sigma,0,0)}=\begin{pmatrix} \frac{1}{\sigma^2} & 0 & 0 & \frac{-2}{\pi\sigma} \\ 0 & \frac{2}{\sigma^2} & \frac{2}{\pi\sigma} & 0 \\ 0 & \frac{2}{\pi\sigma} & \frac{2(\pi-2)}{\pi^2} & 0 \\ \frac{-2}{\pi\sigma} & 0 & 0 & \frac{4}{9\pi^6}\left[96+\pi^2\left(\frac{1}{\sigma^2}-48\right)+15\pi^4\right]. \end{pmatrix}$$

Comparing this information matrix with the one in Arrué et al. (2010) for the $SNC(\beta)$, it follows that they differentiate only on the row and column corresponding to the additional parameter α. Hence, computing the inverse $(I^*)^{-1}$ we have the asymptotic variance of the maximum likelihood estimators for the parameters μ, σ, α and β, respectively.

5. Illustration

To illustrate the estimation procedure discussed in the previous section we consider the variable N-Cream available in the data base Creaminess of cream cheese (see Urlhttp://www.models.kvl.dk/Cream) which was used by Arnold et al. (2009). The corresponding descriptive statistics for this variable are given by the sample size $n = 240$, the mean $\bar{x} = 7.578$ and the variance $s^2 = 2.964$. Quantities $\sqrt{b_1} = -0.551$ and $b_2 = 3.173$ correspond to the sample asymmetry and kurtosis coefficients, respectively. In Table 1, the five models Normal (N), SNC, mixture (MIX), ETN and ESNC with additional location and scale parameters are fitted to the data. MIX is a mixture of two normal distributions represented by $f_Z(z;\mu,\sigma,\mu_1,\sigma_1,p) = p\frac{1}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right) + (1-p)\frac{1}{\sigma_1}\phi\left(\frac{z-\mu_1}{\sigma_1}\right)$. Notice that the N and SNC models are nested within the ESNC model, so that likelihood ratio tests will provide meaningful comparisons for these models.

In all cases, the parameters are estimated by maximum likelihood using the R-package `optim` (2011). The standard errors of the maximum likelihood estimates are calculated using the information matrix corresponding to each model.

The summaries provided by Table 1 illustrate a key feature of the ESNC model; its flexibility and the wide range of coefficients of skewness and kurtosis that it can adapt to, in contrast to the other models. For example, it is clear that the fit of the

normal model is inadequate because of the high degree of skewness of the data. To compare the ESNC model with the normal and SNC models, consider testing the null hypothesis of a normal or a SNC distribution against an ESNC distribution using the likelihood ratio statistics based on the ratios $\Lambda_1 = L_N(\widehat{\mu},\widehat{\sigma},\widehat{\alpha})/L_{ESNC}(\widehat{\mu},\widehat{\sigma},\widehat{\alpha},\widehat{\beta})$ and $\Lambda_2 = L_{SNC}(\widehat{\mu},\widehat{\sigma},\widehat{\alpha})/L_{ESNC}(\widehat{\mu},\widehat{\sigma},\widehat{\alpha},\widehat{\beta})$. Substituting the estimated values, we obtain $-2\log(\Lambda_1) = -2(-469.5862+461.555) = 16.062$ and $-2\log(\Lambda_2) = -2(-466.036+461.555) = 8.962$ which, when compared with the 95% critical value of the $\chi^2_1 = 3.84$, indicate that the null hypotheses are clearly rejected and there is strong indication that the ESNC distribution presents a much better fit than either the N or the SNC distribution to the data set under consideration. In particular, there are significant differences between normal and ESNC models, so not for use reparametrization Rotnitzky et al. (2000). The conclusion of these analysis is that the ESNC model appears to be more appropriate for the particular data set analyzed here. Moreover, using the AIC criterion to MIX, ETN and ESNC models, we can conclude that the ESNC distribution fits better the data. Furthermore, using the delta-method to the information matrix (see Appendix) we have calculated the population estimates of the mean and variance (and their standard deviations), given by $\widehat{E(X)} = 7.596(0.007)$ and $\widehat{V(X)} = 2.897(0.002)$. These points are illustrated in more detail in Figure 2 where the histograms and the fitted curves for the data sets are displayed.

6. Discussion

The paper introduced an extension of the SNC model in Arrué et al. (2010) based on the model defined in Arnold et al. (2009). Some properties of the model are studied and inference is implemented via the maximum likelihood approach. The Fisher information matrix is derived and it is shown to be singular in the vicinity of symmetry. A parameter transformation is presented which contours the singularity problem, and which turn out to be exactly the one derived for the model studied in Arrué et al. (2010). A data set illustration reveals the good performance of the model introduced.

7. Appendix

A. In the formula of Equation (16), use the following integral:

$$\int_0^\infty \frac{y\phi(y)dy}{(1+\alpha^2+\beta^2y^2)^{r/2}} = \frac{e^{\frac{1+\alpha^2}{2\beta^2}}}{2^{\frac{(r+1)}{2}}\sqrt{\pi}|\beta|^r}\Gamma\left(1-\frac{r}{2},\frac{1+\alpha^2}{2\beta^2}\right).$$

Proof. Using the `Integrate[ ]` of Mathematica (2008) we have the result. ■

B. $\delta_k = E\left(\text{sgn}(Z)Z^k\left(\frac{\phi(\alpha Z)}{\Phi(\alpha|Z|)}\right)\right) = \frac{4\sqrt{2}c_\alpha}{\pi^{3/2}}\int_0^\infty z^k\phi(\sqrt{1+\alpha^2}z)\arctan(\beta z)dz$

C. The score functions are given by

$$\begin{aligned}
\frac{\partial l(\theta;\underline{X})}{\partial\mu} &= \sum_{i=1}^n \frac{Z_i}{\sigma} - \frac{\alpha}{\sigma}\sum_{i=1}^n \frac{\phi(\alpha Z_i)}{\Phi(\alpha|Z_i|)}\text{sgn}(Z_i) - \frac{\beta}{\sigma}\sum_{i=1}^n \frac{\psi(\beta Z_i)}{\Psi(\beta Z_i)},\\
\frac{\partial l(\theta;\underline{X})}{\partial\sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^n Z_i^2 - \frac{\alpha}{\sigma}\sum_{i=1}^n \frac{\phi(\alpha Z_i)}{\Phi(\alpha|Z_i|)}|Z_i| - \frac{\beta}{\sigma}\sum_{i=1}^n \frac{\psi(\beta Z_i)}{\Psi(\beta Z_i)}Z_i,\\
\frac{\partial l(\theta;\underline{X})}{\partial\alpha} &= -\frac{nc_\alpha}{\pi(1+\alpha^2)} + \sum_{i=1}^n \frac{\phi(\alpha Z_i)}{\Phi(\alpha|Z_i|)}|Z_i|,\\
\frac{\partial l(\theta;\underline{X})}{\partial\beta} &= \sum_{i=1}^n \frac{\psi(\beta Z_i)}{\Psi(\beta Z_i)}Z_i.
\end{aligned}$$

where $\psi(t) = 1/(\pi(1+t^2))$ is a PDF of the standard Cauchy distribution.

D. For one observation $X \sim ESNC(\theta)$, the ij-th element of the information matrix I is given by

$$I_{\theta_i\theta_j} = -E\left[\frac{\partial^2 l(\theta;X)}{\partial\theta_i\partial\theta_j}\right], \tag{23}$$

Eventually, one obtains the following expressions for the elements of the information matrix.

$$\begin{aligned}
I_{\mu\mu} &= \frac{1}{\sigma^2} + \frac{\alpha^3 c_\alpha}{\sigma^2\pi(1+\alpha^2)} - \frac{\alpha^2}{\sigma^2}\eta_0 + \frac{\beta^2}{\sigma^2}\rho_0,\\
I_{\mu\sigma} &= \frac{2}{\sigma^2}E(Z) - \frac{1}{\sigma^2}\alpha\delta_0 + \frac{1}{\sigma^2}\alpha^3\delta_2 + \frac{\alpha^2}{\sigma^2}\eta_1 - \frac{\beta}{\sigma^2}\xi + \frac{2\pi\beta^3}{\sigma^2}\tau + \frac{\beta^2}{\sigma^2}\rho_1,\\
I_{\mu\alpha} &= \frac{1}{\sigma}\delta_0 - \frac{1}{\sigma}\alpha^2\delta_2 - \frac{\alpha}{\sigma}\eta_1,\\
I_{\mu\beta} &= \frac{\xi}{\sigma} - \frac{2\pi\beta^2}{\sigma}\tau - \frac{\beta}{\sigma}\rho_1,
\end{aligned}$$

$$\begin{aligned}
I_{\sigma\sigma} &= \frac{2}{\sigma^2} + \frac{\alpha(1+3\alpha^2)c_\alpha}{\sigma^2\pi(1+\alpha^2)^2} + \frac{\alpha^2}{\sigma^2}\eta_2 + \frac{\beta^2}{\sigma^2}\rho_2,\\
I_{\sigma\alpha} &= \frac{c_\alpha(1-\alpha^2)}{\sigma\pi(1+\alpha^2)^2} - \frac{\alpha}{\sigma}\eta_2,\\
I_{\sigma\beta} &= -\frac{\beta}{\sigma}\rho_2,\\
I_{\alpha\alpha} &= -\frac{c_\alpha^2}{\pi^2(1+\alpha^2)^2} + \eta_2,
\end{aligned}$$

$$I_{\alpha\beta} = 0,$$

$$I_{\beta\beta} = \rho_2,$$

where $\xi = E\left(\frac{\psi(\beta Z)}{\Psi(\beta Z)}\right)$, $\tau = E\left(Z^2\frac{\psi^2(\beta Z)}{\Psi(\beta Z)}\right)$, $\eta_k = E\left(Z^k\left(\frac{\phi(\alpha Z)}{\Phi(\alpha|Z|)}\right)^2\right)$, $\rho_k = = E\left(Z^k\left(\frac{\psi(\beta Z)}{\Psi(\beta Z)}\right)^2\right)$ and $\delta_k = E\left(\text{sgn}(Z)Z^k\left(\frac{\phi(\alpha Z)}{\Phi(\alpha|Z|)}\right)\right)$ must be evaluated numerically, with $Z \sim ESNC(\alpha,\beta)$.

Table 1: *Estimated parameters and log-likelihood values for the models N, SNC, MIX, ETN and ESNC for the N-Cream variable. The corresponding standard errors are in parentheses.*

MLE	N	SNC	MIX	ETN	ESNC
μ	7.577(0.110)	9.142(0.161)	6.082(1.203)	6.712(0.117)	6.717(0.104)
σ	1.712(0.078)	2.320(0.152)	1.558(0.498)	1.783(0.096)	1.781(0.094)
α	–	−4.095(1.155)	–	1.855(0.808)	1.863(0.810)
β	–	–	–	0.590(0.122)	1.062(0.267)
μ_1	–	–	8.435(0.257)	–	–
σ_1	–	–	1.097(0.138)	–	–
p	–	–	0.364(0.245)	–	–
Log-lik	−469.586	−466.036	−461.125	−463.671	−461.555
AIC	943.172	938.072	932.250	935.342	931.110

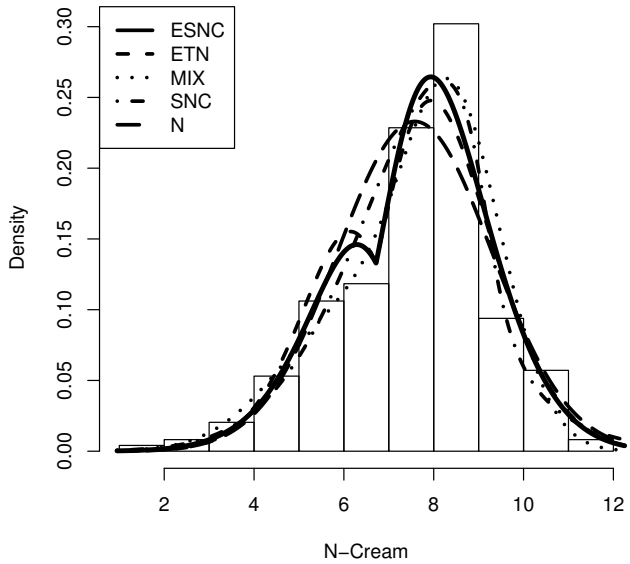


Figure 2: *Histogram for the N-Cream variable. The curves represent densities fitted by maximum likelihood.*

Table 2: *MLEs for the ESNC distribution.*

μ	σ	α	β	n	$\widehat{\mu}$(SD)	$\widehat{\sigma}$(SD)	$\widehat{\alpha}$(SD)	$\widehat{\beta}$(SD)
0	1	4	4	100	0.014 (0.007)	0.983 (0.009)	5.155 (0.507)	4.312 (0.173)
0	1	4	4	300	0.005 (0.002)	0.992 (0.003)	4.830 (0.197)	4.114 (0.053)
0	1	4	4	500	0.006 (0.001)	0.995 (0.002)	4.698 (0.120)	4.038 (0.031)
0	1	4	2	100	0.007 (0.007)	0.991 (0.009)	5.061 (0.494)	2.175 (0.082)
0	1	4	2	300	0.001 (0.002)	0.996 (0.003)	4.886 (0.196)	2.053 (0.025)
0	1	4	2	500	0.001 (0.001)	0.997 (0.002)	4.690 (0.120)	2.033 (0.015)
0	1	4	0	100	-0.011 (0.008)	1.021 (0.008)	4.625 (0.422)	0.031 (0.026)
0	1	4	0	300	-0.005 (0.003)	1.006 (0.003)	5.000 (0.195)	0.015 (0.007)
0	1	4	0	500	-0.003 (0.002)	1.003 (0.002)	4.877 (0.122)	0.009 (0.004)
0	1	2	4	100	0.044 (0.009)	0.977 (0.009)	2.801 (0.268)	4.148 (0.184)
0	1	2	4	300	0.028 (0.003)	0.988 (0.003)	2.540 (0.100)	3.937 (0.056)
0	1	2	4	500	0.021 (0.002)	0.990 (0.002)	2.440 (0.063)	3.910 (0.033)
0	1	2	2	100	0.035 (0.010)	0.985 (0.009)	2.503 (0.222)	2.043 (0.083)
0	1	2	2	300	0.020 (0.003)	0.994 (0.003)	2.366 (0.087)	1.980 (0.026)
0	1	2	2	500	0.014 (0.002)	0.995 (0.002)	2.298 (0.054)	1.978 (0.016)
0	1	2	0	100	-0.006 (0.011)	1.020 (0.008)	2.149 (0.183)	0.016 (0.027)
0	1	2	0	300	-0.003 (0.004)	1.006 (0.003)	2.175 (0.077)	0.004 (0.008)
0	1	2	0	500	-0.002 (0.002)	1.004 (0.002)	2.156 (0.050)	0.002 (0.004)
0	1	0	4	100	0.008 (0.011)	1.207 (0.098)	0.376 (0.196)	5.713 (0.526)
0	1	0	4	300	-0.003 (0.005)	1.149 (0.025)	0.143 (0.056)	4.832 (0.128)
0	1	0	4	500	-0.004 (0.004)	1.118 (0.014)	0.077 (0.033)	4.591 (0.071)
0	1	0	2	100	0.035 (0.012)	1.164 (0.084)	0.283 (0.162)	2.479 (0.228)
0	1	0	2	300	0.031 (0.006)	1.089 (0.019)	0.099 (0.044)	2.139 (0.058)
0	1	0	2	500	0.027 (0.004)	1.068 (0.009)	0.040 (0.024)	2.063 (0.031)
0	1	0	0	100	-0.008 (0.014)	1.125 (0.048)	0.469 (0.123)	0.027 (0.054)
0	1	0	0	300	0.005 (0.007)	1.087 (0.014)	0.107 (0.032)	-0.010 (0.017)
0	1	0	0	500	-0.001 (0.005)	1.062 (0.007)	0.036 (0.016)	-0.001 (0.011)
0	1	-2	0	100	0.005 (0.003)	1.234 (0.348)	-2.446 (0.805)	0.012 (0.352)
0	1	-2	0	300	0.002 (0.002)	1.258 (0.185)	-2.502 (0.422)	0.010 (0.110)
0	1	-2	0	500	0.001 (0.001)	1.248 (0.124)	-2.499 (0.285)	0.002 (0.067)
0	1	-4	0	100	0.002 (0.002)	0.902 (0.301)	-3.539 (1.270)	0.077 (0.465)
0	1	-4	0	300	0.001 (0.001)	0.925 (0.139)	-3.654 (0.587)	0.014 (0.154)
0	1	-4	0	500	0.000 (0.001)	0.938 (0.098)	-3.714 (0.414)	0.003 (0.081)
0	1	-2	4	100	-0.012 (0.003)	1.007 (0.279)	-1.937 (0.679)	5.233 (1.322)
0	1	-2	4	300	-0.010 (0.002)	0.970 (0.112)	-1.858 (0.274)	4.417 (0.467)
0	1	-2	4	500	-0.007 (0.001)	0.969 (0.076)	-1.871 (0.184)	4.171 (0.293)
0	1	-2	2	100	-0.017 (0.003)	1.116 (0.317)	-2.175 (0.750)	3.057 (0.853)
0	1	-2	2	300	-0.015 (0.002)	1.057 (0.124)	-2.036 (0.293)	2.609 (0.304)
0	1	-2	2	500	-0.013 (0.001)	1.074 (0.091)	-2.083 (0.214)	2.489 (0.203)
0	1	-4	4	100	-0.007 (0.002)	0.904 (0.358)	-3.549 (1.520)	4.468 (1.652)
0	1	-4	4	300	-0.005 (0.001)	0.921 (0.164)	-3.629 (0.690)	4.114 (0.699)
0	1	-4	4	500	-0.003 (0.001)	0.939 (0.112)	-3.712 (0.473)	4.052 (0.471)
0	1	-4	2	100	-0.007 (0.002)	0.958 (0.377)	-3.783 (1.579)	2.662 (1.134)
0	1	-4	2	300	-0.005 (0.001)	0.973 (0.167)	-3.847 (0.698)	2.331 (0.431)
0	1	-4	2	500	-0.003 (0.001)	0.965 (0.106)	-3.821 (0.444)	2.168 (0.246)

Acknowledgments

We thank two referees for comments and suggestions that substantially improved the presentation. The research of H. W. Gómez was supported by SEMILLERO UA-2014 (Chile). The research of H. Bolfarine was supported by CNPq and Fapesp (Brasil).

References

Arnold, B.C., Gómez, H.W., Salinas, H.S. (2009). On multiple constraint skewed models. *Statistics: A Journal of Theoretical and Applied Statistics*, 43, 279–293.

Arrué, J.R., Gómez, H.W., Varela H., Bolfarine, H. (2010). On the skew-normal-Cauchy distribution. *Communications in Statistics: Theory and Methods*, 40, 15–27.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171–178.

Chiogna, M. (2005). A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Statistical Methods & Applications*, 14, 331–341.

DiCiccio, T. J. and Monti, A. C. (2004). Inferential aspects of the skew exponential power distribution. *Journal of the American Statistical Association*, 99, 439–450.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80, 27–38.

Kim, H. J. (2005). On the class of two-piece skew-normal distributions. *Statistics*, 39, 537–553.

Nadarajah, S., Kotz, S. (2003). Skewed distributions generated by the normal kernel. *Statistics & Probability Letters*, 65, 269–277.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL :http://www.R-project.org.

Rotnitzky, A. Cox, D.R., Bottai, M., Robins, J (2000). Likelihood-based inference With singular information matrix. *Bernoulli*, 6, 243–284.

Salinas, H. S., Arellano-Valle, R. B., Gómez, H. W. (2007). Skew-exponential power distribution and its derivation. *Communications in Statistics: Theory and Methods*, 36, 1673–1689.

Wolfram Research, Inc., Mathematica, version 7.0, Champaign, IL (2008).

SORT 39 (1) January-June 2015, 51-70

Diagnostic plot for the identification of high leverage collinearity-influential observations

Arezoo Bagheri[1] and Habshah Midi[2]

Abstract

High leverage collinearity influential observations are those high leverage points that change the multicollinearity pattern of a data. It is imperative to identify these points as they are responsible for misleading inferences on the fitting of a regression model. Moreover, identifying these observations may help statistics practitioners to solve the problem of multicollinearity, which is caused by high leverage points. A diagnostic plot is very useful for practitioners to quickly capture abnormalities in a data. In this paper, we propose new diagnostic plots to identify high leverage collinearity influential observations. The merit of our proposed diagnostic plots is confirmed by some well-known examples and Monte Carlo simulations.

MSC: 62-09, 62G35, 62J05, 62J20

Keywords: Collinearity influential observation, diagnostic robust generalized potential, high leverage points, multicollinearity.

1. Introduction

Multicollinearity is an exact or a near linear relationship among regressors in a multiple linear regression. According to Kamruzzaman and Imon (2002), high leverage points or observations that fall far from the majority of independent variables in a data set, are a prime source of multicollinearity. Hadi (1988) pointed out that this source of multicollinearity is a special case in collinearity-influential observations, which may change the multicollinearity pattern of data. They are referred to as high leverage collinearity-enhancing observations or high leverage collinearity-reducing observations (Habshah

[1] National Population Studies & Comprehensive Management Institute, No. 3, 5th Street, Miramad Street, Motahari Street, Tehran, Iran. abagheri_ 000@yahoo.com

[2] Department of Mathematics, Faculty of Science/Institute for Mathematical Research, Universiti Putra Malaysia, 43400 Serdang , Selangor, Malaysia. habshah@upm.edu.my

Received: November 2013
Accepted: March 2015

et al., 2010; Habshah et al., 2011; Bagheri et al., 2012).With their presence, multiple linear regression models encounter serious problems (Habshah et al., 2009; Bagheri et al., 2009; Bagheri and Habshah, 2008). Hence it is very important to detect them so that appropriate steps can be taken to remedy such problems (Bagheri and Habshah, 2012-2011; Habshah et al., 2010).

Simple scatter plots are very useful in exploring the relationship between a response and a single explanatory variable as well as in detecting outliers. They are, however, ineffective in revealing the complex relationships or detecting the trend and data problems in multiple regression models. Partial plots, on the other hand, may be better substitutes for scatter plots in a multiple linear regression. This is because these plots illustrate the partial effects or the effects of a given predictor variable after adjusting for all the other predictor variables in a regression model.

There are two different kinds of partial plots, namely the partial residual and the partial regression or added variable plot (See partial plots in Myers, 1990 and also leverage plots in Sall, 1990; Leverage-Residual Plot of Gray, 1983) which are documented in the literature (Belsley et al., 1980; Cook and Weisberg, 1982). However, partial residual and partial regression plots are generally unable to detect multicollinearity. Overlaying both the partial residual and partial regression plots on the same plot, with the centered xi values on the x-axis, may in fact provide an alternative method to detect multicollinearity (Stine, 1995) by highlighting the amount of shrinkage in partial regression residuals. However, when high leverage points are the source of multicollinearity, these plots will be affected and as a result they will no longer be useful for diagnosing multicollinearity in a data set.

Unfortunately, to the best of our knowledge, we have not found any paper in the literature that establishes graphical methods for the identification of multicollinearity due to high leverage points. This gap in the literature has motivated us to propose appropriate plots that are able to classify observations according to regular observations, high leverage points, collinearity-influential observations and vertical outliers.

These plots will be examined in this paper which is organized into five sections. The next section, Section 2, reviews High Leverage Collinearity-Influential Measure (HLCIM) based on Diagnostic-Robust Generalized Potential (DRGP) which is referred to in this paper as HLCIM(DRGP). Section 3 introduces the newly proposed high leverage collinearity-influential observation regression diagnostic plots. Section 4 discusses both the performance of our proposed plots by using some real data sets and their merit according to Monte Carlo simulations. Finally, some concluding remarks are presented in Section 5.

2. Literature review

In the following section, high leverage collinearity-influential measure based on DRGP will be discussed. Firstly, the regression model can be defined as the following equation:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{Y}$ is an $(n \times 1)$ vector of response or the dependent variable, $\boldsymbol{X}$ is an $(n \times p)$ matrix of predictors $(p \times 1)$, $\boldsymbol{\beta}$ is $(p \times 1)$ vector of unknown finite parameters to be estimated and ε is an $(n \times 1)$ vector of random errors. We allow $\boldsymbol{X}_j$ to denote the j^{th} column of the $\boldsymbol{X}$ matrix; therefore, $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p]$. Additionally, we define multicollinearity in terms of the linear dependence of the columns of $\boldsymbol{X}$; thus, the vectors of $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$ are linearly dependent if there is a set of constants $t_1, t_2, \ldots, t_p$ that are not all zero, such as $\sum_{j=1}^{p} t_j \boldsymbol{X}_j = 0$. The problem of multicollinearity is said to exist when this equation holds approximately $\sum_{j=1}^{p} t_j \boldsymbol{X}_j \approx 0$.

Since multicollinearity is a problem that exists in a data set, there is no statistical test for its presence. Nonetheless, a statistical test can be substituted by a diagnostic method in order to indicate the existence and extent of multicollinearity in a data set. Belsley et al. (1980) proposed an approach for diagnosing multicollinearity based on a singular-value decomposition of a $(n \times p)$ $\boldsymbol{X}$ matrix as:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{V}\boldsymbol{D}' \tag{2}$$

where $\boldsymbol{U}$ is the $(n \times p)$ matrix in which the columns that are associated with the p non-zero eigenvalue of $(\boldsymbol{X}'\boldsymbol{X})$ is $(n \times p)$, $\boldsymbol{V}$ (the matrix of eigenvectors of $\boldsymbol{X}'\boldsymbol{X}$) is $(p \times p)$, $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{I}$, $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}$, and D is a $(p \times p)$ diagonal matrix with non-negative diagonal elements, k_j, $j = 1, 2, \ldots, p$, which is called the singular-values of $\boldsymbol{X}$. The j^{th} Condition Index (CI) of the $\boldsymbol{X}$ matrix is defined as:

$$k_j = \frac{\lambda_{max}}{\lambda_i}, \quad j = 1, 2, \ldots, p, \tag{3}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the singular values of the $\boldsymbol{X}$ matrix. The largest value of k_j is defined as the Condition Number (CN) of the $\boldsymbol{X}$ matrix. Belsley (1991) stated that an $\boldsymbol{X}$ matrix between 10 and 30 indicates a moderate to strong multicollinearity, whereas a value of more than 30 reflects severe multicollinearity.

As previously mentioned, high leverage collinearity-influential observations are those observations that may disrupt the multicollinearity pattern of a data. Unfortunately, not many studies relevant to these issues are found in the literature. Hadi(1988) noted that not all high leverage points are collinearity-influential observations, but most collinearity-influential observations are points with high leverages. He proposed a measure for the identification of high leverage collinearity-influential observations based

on the influence of the i^{th} row of $\boldsymbol{X}$ matrix on the condition index as:

$$\delta_i = \log \frac{k_{(i)} - k}{k}, \;\; i = 1, 2, \ldots, n, \tag{4}$$

where $k_{(i)}$ is the eigenvalue of $\boldsymbol{X}_{(i)}$ when the i^{th} row of $\boldsymbol{X}$ matrix has been deleted. He pointed out that a large negative value of δ_i indicates that the i^{th} observation is a collinearity-enhancing observation, while a large positive δ_i value indicates a collinearity-reducing observation. Sengupta and Behimasankaram(1997) suggested a more preferable measure to Hadi's measure (Hadi ,1988) which is defined as follows:

$$l_i = \log \frac{k_{(i)}}{k}, \;\; i = 1, 2, \ldots, n, \tag{5}$$

According to Bagheri et al. (2012), the performance of both δ_i and l_i is only good for the detection of a single high leverage collinearity influential observation. Moreover, there are some drawbacks in using δ_i or l_i because there are no given specific cutoff points to indicate which observations are collinearity-enhancing and which are collinearity-reducing. To rectify these problems, Bagheri et al. (2012) and Bagheri and Habshah (2012) proposed a high leverage collinearity-influential measure, namely HLCIM (DRGP), denoted as $\delta_i^{(D)}$ and which is defined as follows:

$$\delta_i^{(D)} = \begin{cases} \log \frac{k_{(D)}}{k_{(D-i)}} & \text{if} \quad i \in D \; and \; \neq \{D\} \neq 1 \\ \log \frac{k_{(i)}}{k} & \text{if} \quad \neq \{D\} \; and \; D = i, \; i =, 2, .., n \\ \log \frac{k_{(D+i)}}{k_{(D)}} & \text{if} \quad i \in R \end{cases} \tag{6}$$

where D is the suspected group of multiple high leverage points and R is the remaining good observations diagnosed by DRGP based on Minimum Volume Ellipsoid (MVE) (Habshah et al., 2009). The number of elements in the D group is denoted as $\neq \{D\}$. $k_{(i)}$ indicates the condition number of the $\boldsymbol{X}$ matrix without the i^{th} high leverage points. $k_{(D-i)}$ indicates the condition number of the $\boldsymbol{X}$ matrix without the entire D group minus the i_{th} high leverage points where i belongs to the suspected D group. $k_{(D+i)}$ refers to the condition number of the $\boldsymbol{X}$ matrix without the entire D group of high leverage points plus the i_{th} additional observation of the remaining group (For more information on high leverage diagnostic measures, please refer to Hadi, 1992 and Imon, 2002).

Bagheri et al. (2012) and Bagheri and Habshah(2012) proposed some cutoff points for $\theta_i, \;\; i = 1, 2, \ldots, n$:

$$\text{cut}^1(\theta) = \text{Median}(\theta_i) - c\text{Mad}(\theta_i) \tag{7}$$

$$\text{cut}^2(\theta) = \text{Median}(\theta_i) + c\text{Mad}(\theta_i) \tag{8}$$

where $\text{cut}^1(\theta)$ is the cutoff point for collinearity-enhancing measure and $\text{cut}^1(\theta)$ is the collinearity-reducing measure cutoff point. Median and Mean Absolute Deviation (MAD) stand for robust measures of central tendency and dispersion, respectively. θ_i can be δ_i , l_i , or $\delta_i^{(D)}$ and c is the chosen constant value of 3.$|\theta_i| \geq |\text{cut}^1(\theta)|$ for $\theta_i < 0$ and $\theta_i \geq \text{cut}^2(\theta)$ for $\theta_i > 0$ is an indicator that the i_{th} observation is a high leverage collinearity-enhancing or -reducing observation, respectively.

Bagheri et al. (2012) pointed out that $\delta_i^{(D)}$ values which exceed the cutoff point and belong to the D groups are called high leverage collinearity-influential observations. On the other hand, those $\delta_i^{(D)}$ which exceed the cutoff point and belong to the R group are called collinearity-influential observations. Since the existence of these points have unduly effects on the parameter estimates, it is imperative to quickly identify them by using diagnostic plots. In this regard, new diagnostic plots to separate high leverage collinearity-influential observations from collinearity-influential observations are proposed.

3. Proposed diagnostic plots

Identifying outliers and high leverage points is a fundamental step in the least squares regression model building process. The usage of graphical tools is one of the easiest ways to quickly capture abnormal points in a data set. Rousseeuw and Van Zomeren (1990) proposed the usage of diagnostic plots and referred to them as an outlier map to classify observations into four types of data points, namely regular observations, good leverage points, vertical outliers and bad leverage points. The proposed outlier map plots the standardized residual ($\frac{r_i}{\hat{\sigma}_i}$,for $i = 1,2,\ldots,n$) versus Squared Robust Mahalanobis Distance based on (MVE)($\text{RMD}^2(\text{MVE})$) or Squared Robust Mahalanobis Distance based on Minimum Covariance Determinant ($\text{RMD}^2(\text{MCD})$). The disadvantage of this plot is that it uses robust distance which has the tendency to declare more observations as high leverage points due to swamping effects (Habshah et al., 2009). Since robust distance fails to accurately identify high leverage points correctly while the DRGP is able to successfully identify their presence, in this paper we suggest the usage of DRGP in the construction of our proposed diagnostic plots.

The first proposed plot is similar to the outlier map of Rousseeuw and Van Zomeren (1990), except that the robust distance is substituted with the DRGP. As suggested by Rousseeuwand Van Zomeren (1990), the standardized Least Trimmed Squares Residuals (LTSR) residuals are plotted on the $\boldsymbol{Y}$-axis. We name the first proposed plot the LTSR-DRGP plot. First, each of the LTS residuals, r_i for $i = 1,2,..,n$, is standardized by $\hat{\sigma}$. The LTSR -DRGP plots the standardized LTS residuals against the DRGP. In the LTSR -DRGP plot, any observation which exceeds the $\boldsymbol{Y}$-axis boundaries

$(\pm\sqrt{X^2_{1,0.975}})$ is called a vertical outlier while any that exceeds the $\boldsymbol{X}$-axis boundaries $(\text{Median}(p^*_{ii}) + c\text{Mad}(p^*_{ii})$ where p^*_{ii} is the value of DRGP (Habshah et al., 2009) is called a good leverage point. When an observation exceeds both the y-axis and the x-axis boundaries, it is called a bad leverage point.

The second proposed plot is based on the newly developed diagnostic measure for the identification of multiple high leverage collinearity-influential observations, HLCIM(DRGP), denoted as $\delta_i^{(D)}$ as presented in Equation (6). We name this plot the DRGP-HLCIM plot. It plots the DRGP against the High Leverage Collinearity-influential Measure.

The third proposed plot is also based on HLCIM(DRGP). This plot is called the LTSR -HLCIM plot. In this plot, the Standardized LTS Residuals are plotted against the High Leverage Collinearity-influential Measure. Figures 1, 2 and 3 show the Venn diagram or Ballentine view of the LTSR-DRGP, the DRGP-HLCIM, and the LTSR-HLCIM plots, respectively. It is important to note that the proposed cutoff points are as follows:

$$\text{cut}^1(P^*_{ii}) = \text{Median}(P^*_{ii}) + c\text{Mad}(P^*_{ii}) \tag{9}$$

where P^*_{ii} is the DRGP. If the proposed $\delta_i^{(D)}$ in Equation 6 is employed, then $\text{cut}^1\left(\delta_i^{(D)}\right)$ and $\text{cut}^2\left(\delta_i^{(D)}\right)$ from Equations 7 and 8 are the cutoff points for detecting high leverage collinearity-enhancing and -reducing observations, respectively.

Figure 1 separates the data set into groups of regular observations, vertical (or regression) outliers, and good or bad leverage points. The figure groups the data set according to whether the observation is a high leverage point and/or a vertical outlier. Nevertheless, it does not take into consideration the multicollinearity pattern of a data set.

Figure 2 groups the data set according to whether the observation is a high leverage point or a collinearity-influential observation. Hence, it classifies the data set into groups

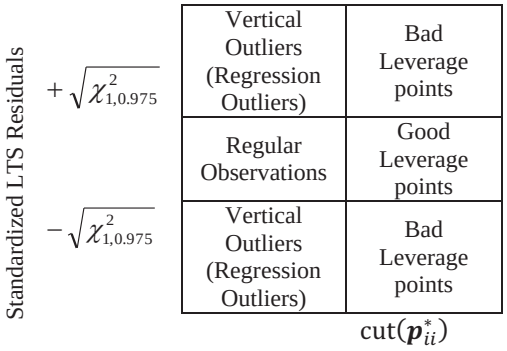


Figure 1: *The Venn Diagram or Ballentine View of LTSR-DRGP Plot.*

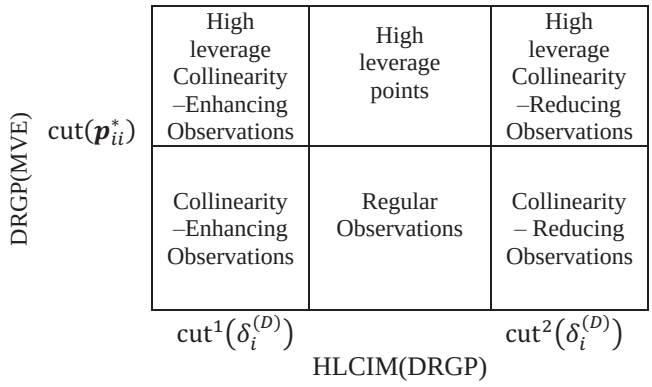


Figure 2: *The Venn Diagram or Ballentine View of DRGP-HLCIM Plot.*

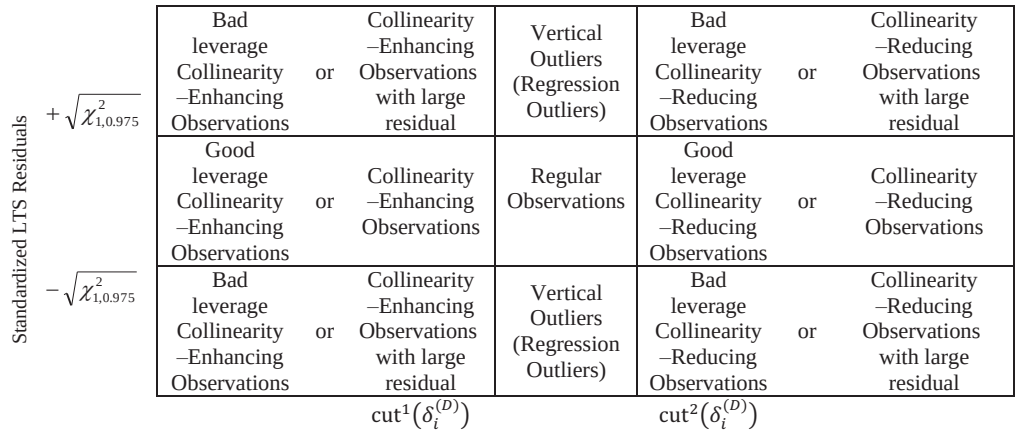


Figure 3: *The Venn Diagram or Ballentine View of LTRS-HLCIM Plot.*

of regular observations, high leverage points, high leverage collinearity-enhancing/reducing observations, and collinearity-enhancing/reducing observations.

This figure also does not take into consideration whether the observation is abnormal in the$\boldsymbol{Y}$-direction. Finally, Figure 3 classifies the data as regular observations, vertical outliers, good leverage collinearity-enhancing/reducing observations, collinearity-enhancing/reducing observations, bad leverage collinearity-enhancing/reducing observations as well as collinearity-enhancing/reducing observations with large residuals. One of the interesting features of this figure is that it takes into account the good leverage points which are also collinearity-influential observations. Most statisticians believe that good leverage points are not problematic since they are in the same fitted regression line as the other data set and they decrease the standard error of the parameter estimations because they increase the variability of $\boldsymbol{X}$ (see for instance Moller et al., 2005; Andersen, 2008). However, these points maybe collinearity-influential observations and like

bad leverage points, they may be destructive to the regression analysis. A joint DRGP-HLCIM and LTSR-HLCIM plot can give a clearer view of the outlyingness of any points in the $\boldsymbol{X}$-direction or $\boldsymbol{Y}$-direction as well as the multicollinearity pattern of a data set. In the following section, the performance of our proposed diagnostic plots is measured by applying these plots to influential cases with authentic and well-known data sets.

4. Results and discussion

Numerical and Monte Carlo simulation results will be discussed in the following sub sections.

4.1. Numerical results

In this section, the performance of the proposed diagnostic plots, namely the LTSR-DRGP, the DRGP-HLCIM, and the LTSR-HLCIM are investigated through the usage of some commonly referred data sets such as the Hawkins-Bradu-Kass data, Commercial Properties data and Body Fat data sets. The first data set is taken from Hawkins, Bradu, and Kass(1984) while the second and third are taken from Kutner et al.(2005).

The Hawkins-Bradu-Kass data set is constructed to have ten bad leverage points $(cases\ 1-10)$ and four good leverage points $(cases\ 11-14)$ (Rousseeuw and Leroy, 1987; Habshah et al., 2009; Bagheri et al., 2012). Figure 4 presents the proposed diagnostic plots for the Hawkins-Bradu-Kass data set. According to parts (a) and (c) of this figure, cases $11-14$ are not only good leverage points but are also good leverage collinearity-enhancing observations. Moreover, cases $1-10$ are bad leverage points and bad leverage collinearity-enhancing observations. It is important to mention that cases 1-14 are all high leverage collinearity-enhancing observations (Figure 4, part (b)). Also, it is worth noting that even though cases $11-14$ are good leverage points, they are collinearity-enhancing observations. Hence, more attention is needed in the estimation of their parameters.

Figure 5 presents the diagnostic plots for the Hawkins-Bradu-Kass data set without the first 14 observations. It can be observed from parts (a) and (b) of Figure 5 that this data set does not have any vertical outliers nor any high leverage points. Nonetheless, it has one collinearity-reducing observation (case 53) which was masked in the presence of the first 14 observations.

Diagnostic plots for the original and modified Commercial Properties data set are presented in Figures 6 and 7, respectively. The original data set has 19 high leverage points (observations 1, 2, 3, 6, 7, 8, 17, 21, 26, 29, 37, 45, 53, 54, 58, 61, 62, 72 and 79) with only two (cases 6 and 62) bad leverage points (Figure 6 part (a)). Moreover, cases 9, 63, 64, 65, and 68 are vertical outliers. There are no high leverage collinearity-enhancing observations in this data set (Figure 6 part (b)). Parts(b) and (c) of Figure 6

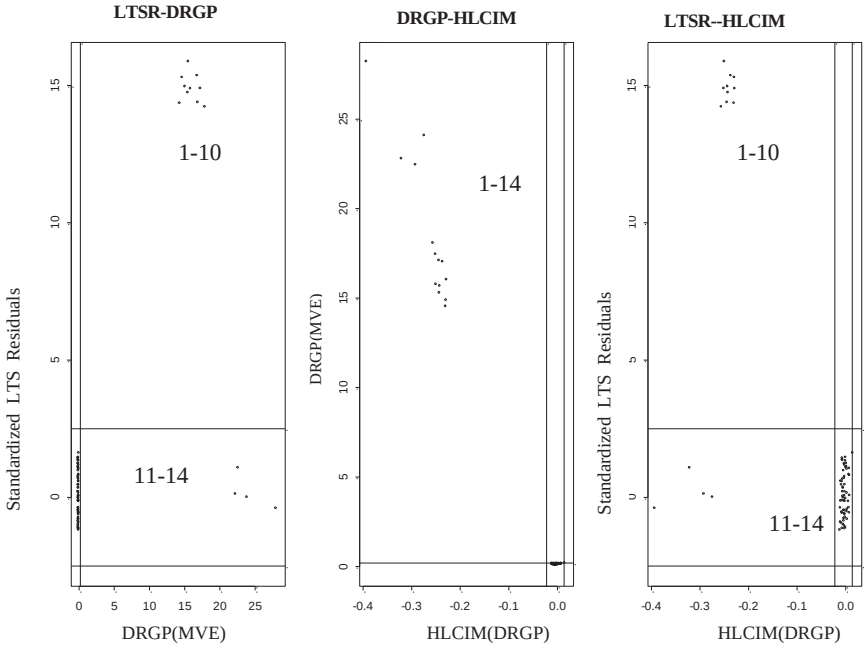


Figure 4: *Diagnostic Plots of Hawkins-Bradu-Kass Data Set.*

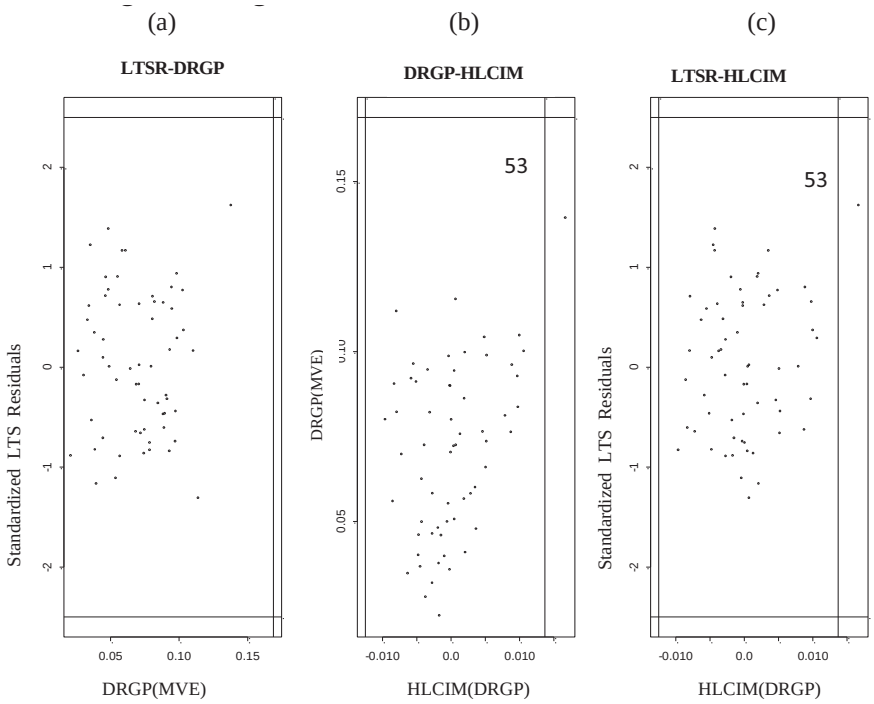


Figure 5: *Diagnostic Plots of Hawkins-Bradu-Kass Data Set Without the First 14 Observations.*

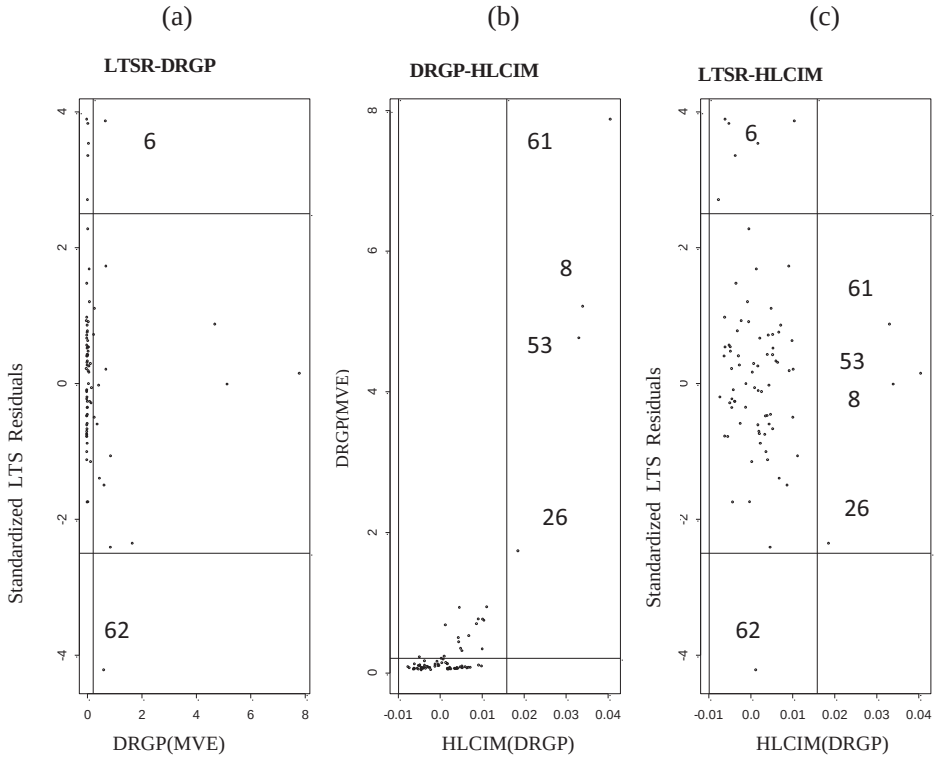


Figure 6: *Diagnostic Plots of Commercial Properties Data Set.*

reveal that cases 8, 26, 53, and 61 are high leverage collinearity-reducing observations and good leverage collinearity-reducing observations, respectively.

After modifying the Commercial Properties data set by replacing observations 1, 2, 3, 6, 7 and 8 in each of the explanatory variables by fixed values of 300, 200, 100, 300, 200, and 100, respectively, these observations became good leverage points (Figure 7 part (a)). Figure 7 part (a) also indicates that case 8 is a bad leverage point. All the modified cases of 1, 2, 3, 6, 7 and 8 are high leverage collinearity-enhancing observations (Figure 7, part (b)). According to Figure 7, part (c), case 8 is a bad leverage collinearity-enhancing observation while cases 1, 2 , 3, 6 and 7 are good leverage collinearity-enhancing observations. Hence, cases 1, 2, 3, 6, and 7 require more attention in order to prevent any misleading conclusions.

Figures 8 to 10 are diagnostic plots for the original and modified Body Fat data set. Part (a) of Figure 8 shows that the original Body Fat data set has four good leverage points (cases 5, 15, 1 and 3) and having zero vertical outliers. Only case 15 is a high leverage collinearity-reducing observation. It can be seen that case 13, a non high leverage, is also a collinearity-reducing observation (Figure 8 part (b)). Additionally, cases 15 and 13 are good leverage collinearity-reducing and collinearity-reducing observations, respectively (Figure 8 part (c)).

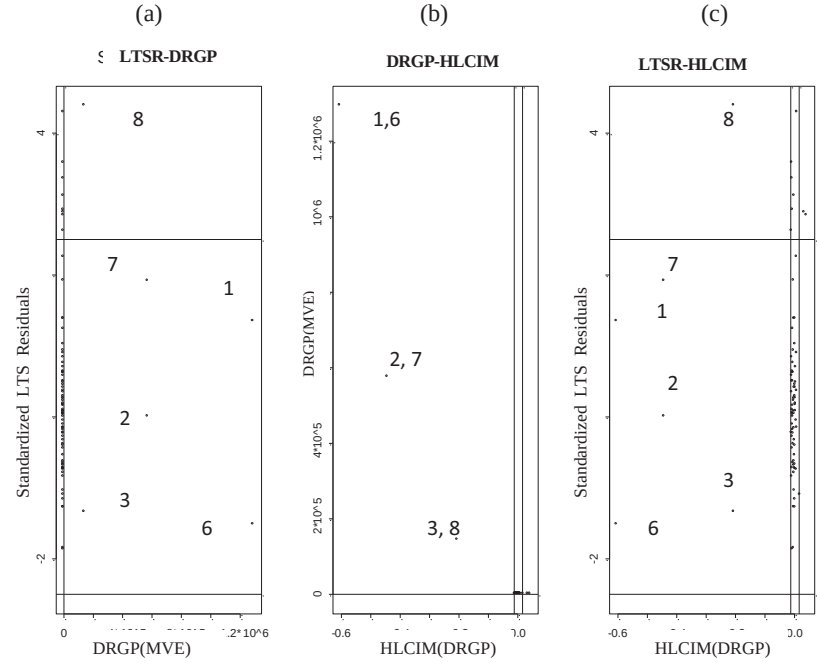


Figure 7: *Diagnostic Plots of Modified Commercial Properties Data Set.*

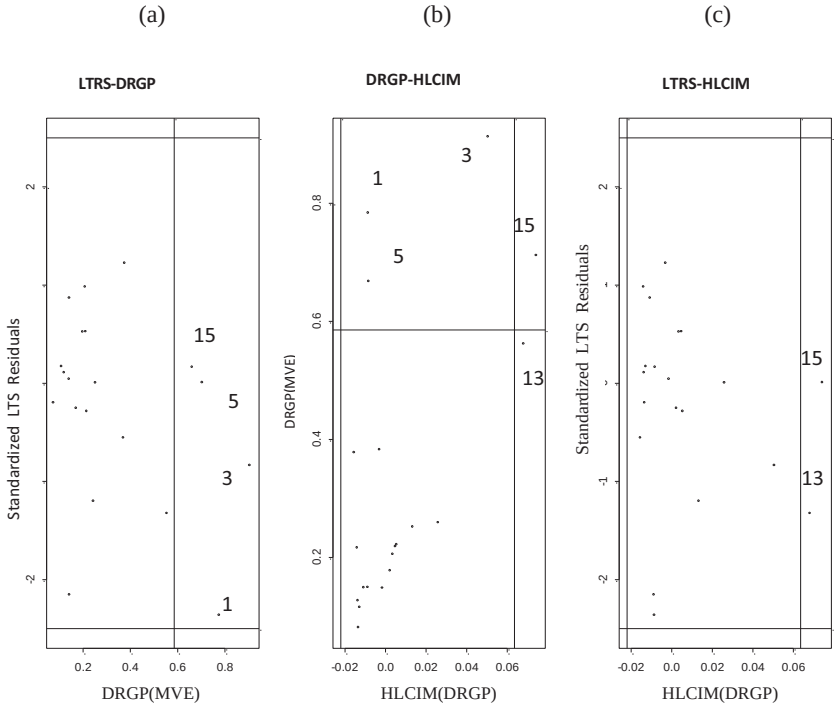


Figure 8: *Diagnostic Plots of Original Body Fat Data Set.*

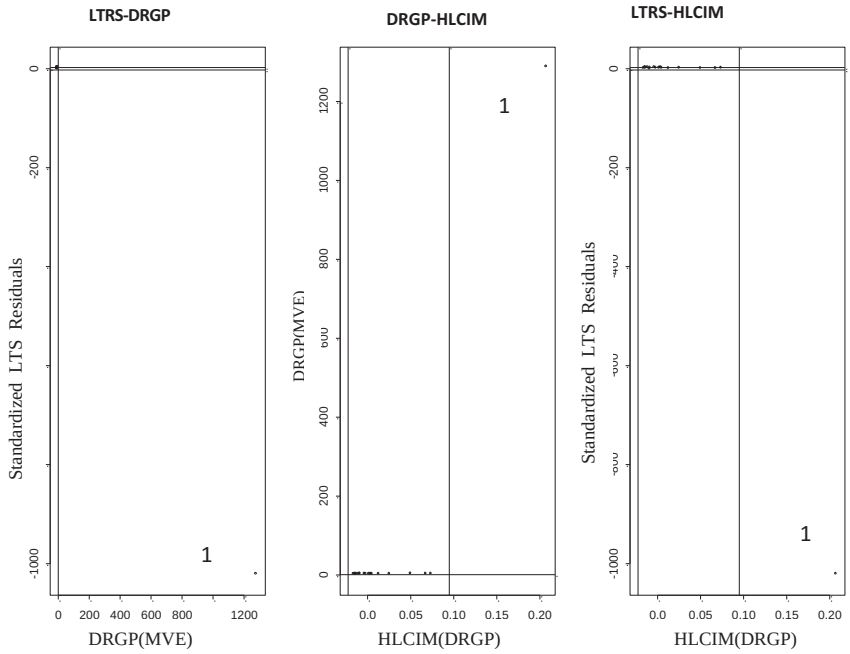


Figure 9: *Diagnostic Plots of Modified x_1 Body Fat Data Set.*

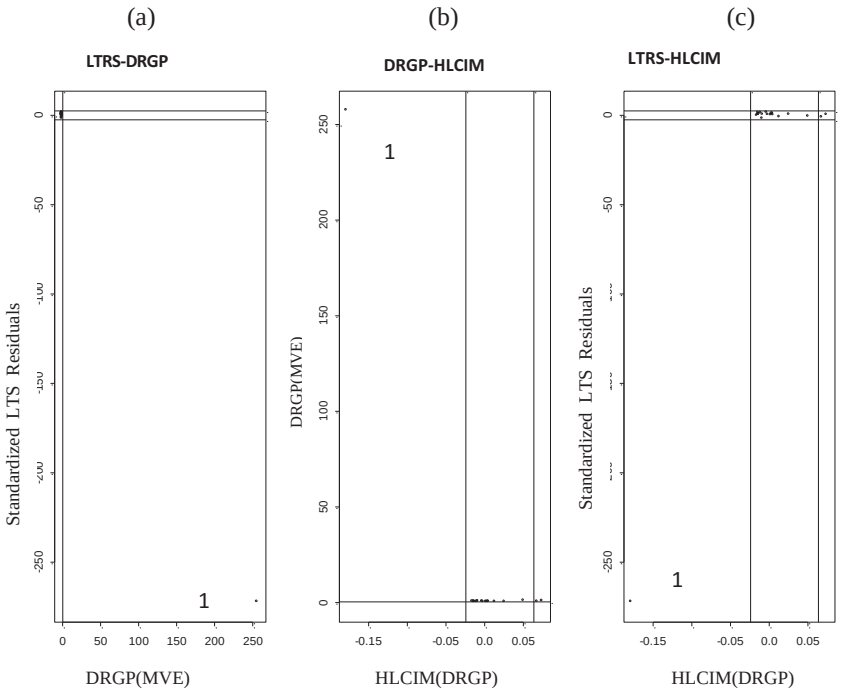


Figure 10: *Diagnostic Plots of Modified x_1 and x_2 in the Same Positions Body Fat Data Set.*

Figure 9 and 10 illustrates the modified Body Fat data set when the first observation of x_1 is fixed to 300 and when the first observation of x_1 and x_2 is fixed to 300, respectively. Figures 9 and 10, part (a), reveal that the added contaminated point is a bad leverage point. Moreover, according to Habshah et al. (2011) when the high leverage point only exists in x_1, case 1 becomes a high leverage collinearity-reducing observation (Figure 9 part (b)). Figure 10 part (b) however, shows case 1 as a high collinearity-enhancing observation when modification is for x_1 and x_2 in the same position. Furthermore, part (c) in Figure 9 shows that case 1 is a bad leverage collinearity-reducing observation while in part (c) of Figure 10 it is a bad leverage collinearity-enhancing observation.

4.2. Monte Carlo simulation study

In this section, a Monte Carlo simulation study was designed to assess the merit of our proposed diagnostic plots in terms of its ability to separate the data set according to regular observations, vertical outliers (regression outliers), collinearity-enhancing/reducing observations with large residuals, bad leverage collinearity-enhancing/reducing observations, good leverage collinearity-enhancing/reducing observations and collinearity-enhancing/reducing observations. To achieve this aim, non-collinear and collinear data sets with three regressors were generated in such a way that different scenarios were created, namely, high leverage collinearity-enhancing/reducing observations and vertical outliers. It is important to mention here that although the proposed diagnostics plots can detect collinearity-enhancing/reducing observations clearly, they were not explicitly generated. In each scenario, four samples of size 40, 60, 100, and 300 and different levels of high leverages of (the percentage of added contaminated cases) = 0.05,0.10, 0.15, 0.20 with unequal weights were considered.

In order to generate high leverage collinearity-enhancing observations, each variable was firstly generated from Uniform (0,1) to produce non-collinear data sets. This generated data is referred to as the regular observations. The last $100\%\alpha$ observations of the regular observations of each regressor were then replaced with certain percentage of high leverage points to create high leverage collinearity-enhancing observations. To generate the high leverage points as collinearity-enhancing with unequal weights in non-collinear data sets, the values corresponding to the first high leverage point were kept fixed at 10 and those of the successive values were created by multiplying the observations index, i, by 10.

As per Lawrence and Arthur (1990), high leverage collinearity-reducing observations were created by generating collinear regressors on the outset:

$$x_{ij} = (1-\rho^2)z_{ij} + \rho z_{i(t1)} \tag{10}$$

where the z_{ij}, $i = 1,\ldots,n$; $j = 1,\ldots,t+1$; t=3, are independent standard normal random numbers. The value of ρ^2 or the correlation between the two explanatory variables, was

Table 1: *The abbreviations used in Tables 2-6.*

Abbreviations	Meaning
CN	the condition number of $\boldsymbol{X}$ matrix without high leverage points
CN*	the condition number of $\boldsymbol{X}$ matrix with high leverage points
RO	the number of simulated regular observations
VO	the number of simulated vertical outliers
DCEO	the number of detected collinearity-enhancing observations

set to be equal to 0.95 which causes high collinearity between regressors. High leverage collinearity-reducing observations in collinear data sets were then created by replacing the first $100(\frac{\alpha}{2})$ percent observations of X_1 and the last $100(\frac{\alpha}{2})$ percent observations of X_2 with high leverage points. To create vertical outliers, a dependent variable from a Uniform (0, 1) was firstly generated. For each sample size, a certain percentage of outliers was generated by randomly deleting a certain percentage of 'good' observations and replacing them with 'bad' data points. The first outlier is kept fixed at 100 (10^2) and the successive values are created by multiplying the observations index, i, by 10.

The Good leverage Collinearity-Enhancing Observation (GLCEO) was created in such a way that the High leverage Collinearity-Enhancing Observation (HLCEO) is generated without any vertical outlier. On the other hand, Bad leverage Collinearity-Enhancing Observation (BLCEO) was created when both HLCEO and vertical outliers were generated. Similarly, Good leverage Collinearity-Reducing Observation (GLCRO) was created only when High leverage Collinearity-Reducing Observation (HLCRO) was generated, while the Bad leverage Collinearity-Reducing Observation (BLCRO) was created when both HLCRO and vertical outliers were generated.

Table 1 shows the notations used in Tables 2-6 (D in the entire abbreviations indicates the number of detected observations by the proposed plots). We ran 10,000 simulations. The results based on their averages are presented in Tables 2 to 6. Due to space constraints, only the results for $n = 40$ and 300 are included. The conclusions of other results were consistent.

Let us first look at Table 2 when $\alpha = 0.00$. It can be seen that when there is no vertical outliers or high leverage points in the data, the value of CN=CN* and is less than 5.0, indicating that there is no multicollinearity problem. It is also interesting to note that our proposed plots can detect almost all observations as regular observations (on the average of 96 percent). The results in Table 2 also indicate that in the presence of vertical outliers and in the absence of high leverage points, the data sets do not have multicollinearity problems $(CN < 5.0)$. The results also suggest that the number of detected vertical outliers is reasonably close to the number of generated vertical outliers.

As for the generated bad/good leverage collinearity-enhancing observations data (see Tables 3-4), all the CN* values (> 30) drastically increased in the presence of high leverage points. This indicates that high leverage points are the cause of multicollinearity.

Table 2: *The number of detected abnormal observations in the simulated data sets with vertical outliers.*

n	40					300				
α	0.00	0.05	0.10	0.15	0.20	0.20	0.00	0.05	0.1	0.20
CN	3.54	3.54	3.39	3.39	3.39	3.29	3.29	3.29	3.29	3.29
CN*	3.54	3.54	3.39	3.39	3.39	3.29	3.29	3.29	3.29	3.29
RO	**40.00**	38.00	36.00	34.00	32.00	**300.00**	285.00	270.00	255.00	240.00
DRO	**38.42**	34.59	33.24	31.95	31.27	**298.75**	283.38	268.39	253.07	238.09
VO	0.24	**2.00**	**4.00**	**6.00**	**8.00**	0.00	**15.00**	**30.00**	**45.00**	**60.00**
DVO	0.00	**1.85**	**3.87**	**5.88**	**7.89**	0.00	**14.30**	**29.47**	**44.60**	**59.74**
DCEO-VO	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.07	0.13	0.17
DBLCEO	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DGLCEO	0.05	0.19	0.25	0.25	0.25	1.06	0.00	0.00	0.00	0.00
DCEO	0.00	0.05	0.01	0.00	0.00	0.00	1.03	1.00	0.93	0.90
DCRO-VO	0.00	0.05	0.00	0.00	0.00	0.00	0.67	0.96	1.17	1.00
DBLCRO	0.76	0.05	0.12	0.13	0.00	0.10	0.00	00.00	0.00	0.00
DGLCRO	0.34	0.72	0.27	0.25	31.00	0.09	0.10	0.10	0.10	0.10
DCRO	0.00	2.5	2.24	1.54	0.28	0.00	0.49	0.01	0.00	0.00

Table 3: *The number of abnormal observations in the simulated data sets with bad leverage collinearity-enhancing observations.*

n	40					300		
α	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
CN	3.56	3.41	3.38	3.64	3.29	3.30	3.29	3.27
CN*	23.68	60.09	107.56	166.13	131.70	375.29	704.68	1107.26
RO	38.00	36.00	34.00	32.00	285.00	270.00	255.00	240.00
DRO	35.49	34.04	32.72	30.99	283.07	268.35	252.89	238.72
DVO	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DCEO-VO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BLCEO	**2.00**	**4.00**	**6.00**	**8.00**	**15.00**	**30.00**	**45.00**	**60.00**
DBLCEO	**1.74**	**3.91**	**5.95**	**8.00**	**14.87**	**30.00**	**45.00**	**60.00**
DGLCEO	0.26	0.09	0.00	0.00	0.00	0.00	0.98	0.30
DCEO	0.44	0.18	0.00	0.00	0.00	0.00	0.00	0.00
DCRO-VO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DBLCRO	0.00	0.00	0.00	0.00	0.00	0.00	00.00	0.00
DGLCRO	0.15	0.18	0.15	0.00	0.56	0.50	0.02	0.00
DCRO	1.62	1.60	1.18	1.01	1.50	1.15	1.11	0.98

On the other hand, all the CN* values (< 5.00) for the generated bad/good leverage-reducing observations (see Tables 5-6) dramatically reduced in the presence of high leverage collinearity-reducing observations, suggesting that high leverage points conceal the problem of multicollinearity. The large and small values of CN* confirm that the generated data are collinear and non-collinear data sets, respectively. It can be observed that the number of detected bad/good leverage collinearity-enhancing observations is fairly close to the simulated data. A similar conclusion can be made for the

Table 4: *The number of abnormal observations in the simulated data sets with good leverage collinearity-enhancing observations.*

n			40				300	
α	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
CN	3.46	3.50	3.13	3.64	3.29	3.30	3.29	3.27
CN*	23.69	61.92	107.60	166.05	132.15	377.64	704.52	1107.30
RO	38.00	36.00	34.00	32.00	285.00	270.00	255.00	240.00
DRO	35.62	34.36	32.65	31.00	283.16	268.53	253.91	239.01
DVO	0.15	0.00	0.08	0.00	0.00	0.00	0.00	0.00
DCEO-VO	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BLCEO	**2.00**	**4.00**	**6.00**	**8.00**	**15.00**	**30.00**	**45.00**	**60.00**
DBLCEO	**1.74**	**4.00**	**5.95**	**8.00**	**14.91**	**30.00**	**45.00**	**60.00**
DGLCEO	0.45	0.09	0.00	0.00	0.09	0.00	0.98	0.30
DCEO	0.44	0.18	0.00	0.00	0.00	0.00	0.00	0.00
DCRO-VO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DBLCRO	0.00	0.00	0.00	0.00	0.00	0.00	00.00	0.00
DGLCRO	0.15	0.09	0.10	0.00	0.55	0.31	0.02	0.00
DCRO	1.63	1.46	1.17	1.00	1.20	1.16	1.07	0.99

Table 5: *The number of abnormal observations in the simulated data sets with bad leverage collinearity-reducing observations.*

n			40				300	
α	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
CN	39.40	38.91	36.33	41.17	36.80	37.13	38.34	38.97
CN*	13.12	4.46	1.06	1.13	1.02	1.01	1.01	1.00
RO	38.00	36.00	34.00	32.00	285.00	270.00	255.00	240.00
DRO	36.45	35.48	33.67	32.00	284.02	269.78	255.00	240.00
DVO	0.15	0.03	0.00	0.00	0.00	0.00	0.00	0.00
DCEO-VO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DBLCEO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DGLCEO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DCEO	0.65	0.33	0.31	0.00	0.01	0.00	0.00	0.00
DCRO-VO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BLCRO	**2.00**	**4.00**	**6.00**	**8.00**	**15.00**	**30.00**	**45.00**	**60.00**
DBLCRO	**1.85**	**3.96**	**5.98**	**8.00**	**15.00**	**30.00**	**45.00**	**60.00**
DGLCRO	0.10	0.11	0.04	0.00	0.98	0.22	0.00	0.00
DCRO	0.80	0.10	0.00	0.00	0.00	0.00	0.00	0.00

case of detecting bad/good leverage collinearity-reducing observations. It is very important to note that as the value of alpha increases, the degree of multicollinearity also increases/decreases.

Table 6: *The number of abnormal observations in the simulated data sets with bad leverage collinearity-reducing observation.*

n	40					300		
α	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
CN	38.39	40.70	36.33	40.13	36.76	37.16	38.34	37.28
CN*	1.84	2.49	1.06	1.04	1.02	1.01	1.01	1.00
RO	38.00	36.00	34.00	32.00	285.00	270.00	255.00	240.00
DRO	36.56	35.12	33.47	31.62	282.01	269.23	255.00	240.00
DVO	0.25	0.15	0.00	0.00	0.00	0.00	0.00	0.00
DCEO-VO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DBLCEO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DGLCEO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DCEO	0.31	0.39	0.33	0.26	0.00	0.00	0.00	0.00
DCRO-VO	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BLCRO	**2.00**	**4.00**	**6.00**	**8.00**	**15.00**	**30.00**	**45.00**	**60.00**
DBLCRO	**1.85**	**3.96**	**5.98**	**8.00**	**15.00**	**30.00**	**45.00**	**60.00**
DGLCRO	0.10	0.11	0.04	0.00	0.98	0.22	0.00	0.00
DCRO	0.66	0.34	0.20	0.12	2.99	0.77	0.00	0.00

5. Conclusions

Based on Rousseeuw and Van Zomeren (1990) and Rousseeuw and Van Driessen (1999) and their development of Residual-Distance and Distance to Distance plots, three new diagnostic plots are proposed; the LTSR-DRGP, DRGP-HLCIM, and LTSR-HLCIM. The LTSR-DRGP plot was able to identify regular observations, good or bad leverage points and vertical outliers. The DRGP-HLCIM plot was able to classify the observations as regular observations, high leverage points, high leverage collinearity-enhancing or collinearity- reducing observations and collinearity-enhancing or collinearity-reducing observations. Finally, the LTSR-HLCIM plot successfully distinguishes vertical outliers, good leverage collinearity-enhancing/reducing observations, collinearity-enhancing/reducing observations and bad leverage collinearity-enhancing/reducing observations and collinearity-enhancing/re-
ducing observations with large residuals. Thus, the merits of our proposed diagnostic plots are confirmed, as reflected in their application to different authentic data sets and in the Monte Carlo simulation study.

References

Andersen, R. (2008). *Modern Methods for Robust Regression.* Sara Miller McCune: SAGE publications, USA.

Bagheri, A.,Habshah, M. and Imon, A. H. M. R. (2009). Two-step robust diagnostic method for identification of multiple high leverage points. *Journal of mathematics and Statistics*, 5, 97–106.

Bagheri, A., Habshah, M. and Imon, A. H. M.R. (2012). A novel collinearity-influentialobservation diagnostic measure based on a group deletion approach. *Communications in Statistics-Simulation and Computation*, 41, 1379–1396.

Bagheri, A., Habshah, M. (2009). Robust estimations as a remedy for multicollinearity caused by multiple high leverage points. *Journal of Mathematics and Statistics*, 5, 311–321.

Bagheri, A., Habshah, M. (2011). On the performance of robust variance inflation factors. *International Journal of Agricultural and Statistics Sciences*, 7, 31–45.

Bagheri, A., Habshah, M. (2012). On the performance of the measure for diagnosing multiple high leverage collinearity-reducing observations. *Mathematical Problems in Engineering*, Volume 2012, Article ID 531607, 16 pages.

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics: Identifying : Influential Data and Sources of Collinearity.* Wiley, New York.

Belsley, D. A. (1991). *Conditioning Diagnostics: Collinearity and Weak Data in Regression.* Wiley, New York.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression.* Chapman Hall, London.

Gray, J. B. (1983). The L-R plot: a graphical tool for assessing influence. *Proceedings of the statistical computing section. American statistical association*, 159–164.

Habshah, M. and Bagheri, A. (2013). Robust multicollinearity diagnostic measures based on minimum covariance determination approach. *Economics Computation and Economic Cybernetics Studies and Research*, 4.

Habshah, M., Norazan, M. R. and Imon, H. M. R. (2009). The performance of diagnostic-robust generalized potentials for the identification of multiple high leverage points in linear regression. *Journal of Applied Statistics*, 36, 507–520.

Habshah, M., Bagheri, A. and Imon, A. H. M. R. (2010). The application of robust multicollinearity diagnostic method based on robust coefficient determination to a non-collinear data. *Journal of Applied Sciences*, 10, 611–619.

Habshah, M., Bagheri, A. and Imon, A. H. M. R. (2011). High leverage collinearity-enhancing observation and its effect on multicollinearity pattern: Monte Carlo simulation study. *Sains Malaysiana*, 40, 1437–1447.

Hadi, A. S. (1988). Diagnosing collineariy-influential observations. *Computational Statistics and Data Analysis*, 7, 143–159.

Hadi, A. S. (1992). A new measure of overall potential influence in linear regression. *Computational Statistics and Data Analysis*, 14, 1–27.

Hawkins, D. M., Bradu, D. and Kass, V. (1984). Location of several outliers in multiple regression data using elemental sets. *Technometrics*, 26, 197–208.

Imon, A. H. M. R. (2002). Identifying multiple high leverage points in linear regression. *Journal of Statistical Studies. Special Volume in Honour of Professor Mir Masoom Ali.*, 3, 207–218.

Kamruzzaman, M. D. Imon, A. H. M. R. (2002). High leverage point: Another source of multicollinearity. *Pakistan Journal of Statistics,* 18, 435–448.

Kutner, M. H., Nachtsheim, C. J., Neter, J. and Li, W. (2005). *Applied Linear Regression Models* 5th Edition. MacGraw-Hill,New York.

Lawrence, K. D. and Arthur, J. L. (1990). *Robust Regression: Analysis and Applications*. INC: Marcel Dekker.

Moller, S. F., Frese, J. V. and Bro, R. (2005). Robust methods for multivariate data analysis. *Journal of Chemometrics*, 19, 549–563.

Myers, R. H. (1990). *Classical and Modern Regression with Applications*. 2nd Edition. CA: Duxbury Press.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* New York: Wiley.

Rousseeuw, P. J. and Van Driessen, K.(1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41, 212–223.

Rousseeuw, P. and Van Zomeren, B. (1990). Unmasking multivariate outliers and leverage points. *Journal of American Statistical Association*, 85, 633–639.

Sall, J. (1990). Leverage plots for general linear hypothesis. *The American Statistician,* 44, 308–315.

Sengupta, D. and Bhimasankaram, P. (1997). On the roles of observations in collinearity in the linear model. *Journal of American Statistical Association,* 92, 1024–1032.

Stine, R. A. (1995). Graphical interpretation of variance inflation factors. *The American Statistician,* 49, 53–56.

SORT 39 (1) January-June 2015, 71-84

Discrete alpha-skew-Laplace distribution

S. Shams Harandi and M. H. Alamatsaz

Abstract

Classical discrete distributions rarely support modelling data on the set of whole integers. In this paper, we shall introduce a flexible discrete distribution on this set, which can, in addition, cover bimodal as well as unimodal data sets. The proposed distribution can also be fitted to positive and negative skewed data. The distribution is indeed a discrete counterpart of the continuous alpha-skew-Laplace distribution recently introduced in the literature. The proposed distribution can also be viewed as a weighted version of the discrete Laplace distribution. Several distributional properties of this class such as cumulative distribution function, moment generating function, moments, modality, infinite divisibility and its truncation are studied. A simulation study is also performed. Finally, a real data set is used to show applicability of the new model comparing to several rival models, such as the discrete normal and Skellam distributions.

MSC: 60E, 62E

Keywords: Discrete Laplace distribution, discretization, maximum likelihood estimation, uni-bimodality, weighted distribution.

1. Introduction

The traditional discrete distributions (geometric, Poisson, etc.) have limited applicability in modelling certain real situations such as data on the set of integers $\mathbb{Z} = \{0, \mp 1, \mp 2, \dots\}$ or bimodal data sets. Thus, several researchers have attempted to develop new classes of discrete distributions to cover such situations. Recall that any continuous distribution on $\mathbb{R}$ with probability density function (pdf) f admits a discrete counterpart supported on the set of integers $\mathbb{Z} = \{0, \mp 1, \mp 2, \dots\}$ whose probability mass function (pmf) is defined as

$$P(X = x) = \frac{f(x)}{\sum_{y=-\infty}^{\infty} f(y)}, \quad x \in \mathbb{Z}\,. \tag{1}$$

Corresponding author: alamatho@sci.ui.ac.ir (and mh_-alamatsaz@yahoo.com)
Department of Statistics, University of Isfahan, Isfahan, Iran. salimeh.shams2013@yahoo.com
Received: December 2013
Accepted: November 2014

For instance, Roy (2003) introduced a discrete version of normal distribution to cover discrete data on the whole set of integers $\mathbb{Z} = \{0, \mp 1, \mp 2, \dots\}$ and, similarly, Inusah and Kozubowski (2006) considered a discrete analogue of Laplace (DL) distribution. Kozubowski and Inusah (2006) proposed a discrete version of the skew Laplace (skewDL) distribution as a generalization of discrete Laplace distribution which is useful for unimodal data sets. Also, Barbiero (2014) and Jayakumar and Jacob (2012) introduced other discrete distributions based on skew Laplace and wrapped skew Laplace distributions on the integers, respectively.

The aim of this paper is to propose a more flexible distribution on $\mathbb{Z}$ which can cover unimodal as well as bimodal data. The new discrete distribution can also fit both positively and negatively skewed data. In fact, using (1), we provide a discrete version of the alpha-skew-Laplace distribution which was recently introduced by Shams Harandi and Alamatsaz (2013). The probability density function of the alpha-skew-Laplace distribution is

$$f(x;\alpha,\mu,\sigma) = \frac{1}{4\sigma(1+\alpha^2)}[1+(1-\frac{\alpha}{\sigma}(x-\mu))^2]e^{\frac{-|x-\mu|}{\sigma}}, \qquad x \in \mathbb{R}, \tag{2}$$

where $\alpha \in \mathbb{R}$ is the skewness parameter and $\mu \in \mathbb{R}$ and $\sigma > 0$ are its location and scale parameters, respectively. The discrete version of (2) which is considered here can be fitted to unimodal as well as bimodal data sets having positive as well as negative skewness.

The rest of the article is organized as follows. Section 2 introduces the discrete alpha-skew-Laplace ($DASL(p,\gamma)$) distribution and discusses some of its important features and properties. In Section 3, we shall provide some distributional properties such as moment generating function and moments. Maximum likelihood estimations of parameters involved will be discussed in section 4. Section 5 describes a simulation study. In Section 6, we shall consider some interesting modification of *DASL* distribution. In Section 7, we attempt to fit the proposed model and its special cases to a real data set and compare it with several rival models such as the discrete normal, DL, skewDL and Skellam distributions.

2. The family of discrete alpha-skew-Laplace distributions

In this section, we present the pmf of our new class of discrete distributions on $\mathbb{Z}$ by discretizing alpha-skew-Laplace distribution (2). We let $\mu = 0$ and use relation (1) to obtain

$$p(x;p,\alpha) = C(p,\alpha)[1+(1+\alpha x \log p)^2]p^{|x|}, \qquad x \in \mathbb{Z}, \tag{3}$$

where $C(p,\alpha) = \frac{1}{2}\frac{1-p}{1+p}\,[1+\alpha^2(\log p)^2\frac{p}{(1-p)^2}]^{-1}$, $0 < p = e^{-\frac{1}{\sigma}} < 1$ and $\alpha \in \mathbb{R}$.

Since $\sigma = (-\log p)^{-1}$, to simplify, we let $\gamma = \frac{\alpha}{\sigma}$. Then, we have

$$p(x;p,\gamma) = C(p,\gamma)[1+(1-\gamma x)^2]p^{|x|}, \qquad x \in \mathbb{Z}, \quad p \in (0,1), \gamma \in \mathbb{R}, \tag{4}$$

where $C(p,\gamma) = \frac{1}{2}\frac{1-p}{1+p}[1+\gamma^2\frac{p}{(1-p)^2}]^{-1}$. We denote this distribution by $X \sim DASL(p,\gamma)$.

Remark 1 Recall that for a distribution with pdf (pmf) f, we can construct a new distribution with pdf (pmf)

$$g(x;\Theta_1,\Theta_2) = \frac{w(x;\Theta_1,\Theta_2)}{E_{\Theta_1}[w(x;\Theta_1,\Theta_2)]} f(x;\Theta_1),$$

where Θ_1 and Θ_2 can be two vectors of parameters and w is called a weighted distribution of f. It is worth noting that pmf (4) can also be viewed as the weighted version of the discrete Laplace distribution of Inusah and Kozubowski (2006). To see this, it is sufficient to consider the weight function $w(x;\Theta_1,\Theta_2) = (1+(1-\gamma x)^2)$ with $\Theta_1 = p$, $\Theta_2 = \gamma$ and $f(x;p) = \frac{1-p}{1+p}p^{|x|}$.

Some special cases of this new class of discrete distributions are revealed below:

1. If $\alpha = 0$ in (3), or equivalently $\gamma = 0$ in (4), we obtain the discrete Laplace (DL) distribution.

2. If $\alpha \to \infty$ in (3), or equivalently $\gamma \to \infty$ in (4), we have

$$p(x;p,\gamma) \to \frac{(1-p)^3}{2p(1+p)} x^2 p^{|x|}, \quad x \in \mathbb{Z}$$

   which is a symmetric and bimodal discrete distribution.

3. If $X \sim DASL(p,\gamma)$, then $-X \sim DASL(p,-\gamma)$.

4. By considering the continuous version of the alpha-skew-Laplace distribution of Shams and Alamatsaz (2013), we can conclude that $DASL(p,\gamma)$ is unimodal for $\log p < \gamma < -\log p$ and bimodal for $\gamma \geq -\log p$ or $\gamma \leq \log p$, respectively. Equivalently, if we consider the pmf in (3), then the distribution is unimodal for $-1 < \alpha < 1$ and bimodal for $\alpha \leq -1$ or $\alpha \geq 1$, respectively.

Figure 1 below illustrates several plots of $DASL(p,\gamma)$ distribution for selected values of the parameters p and γ which confirms our result on modality of the distribution. We note that for $\gamma < 0$, all plots are symmetric.

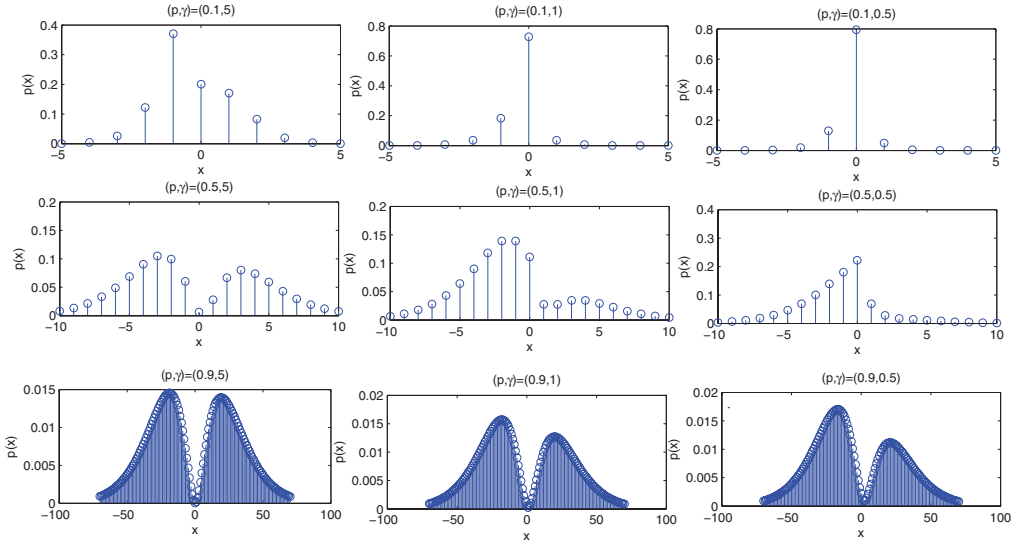


Figure 1: *Illustrations of pmf of DASL*(p,γ) *for different values of p and* γ.

The cumulative distribution function (cdf) of the random variable $X \sim DASL(p,\gamma)$ is given by

$$F(x;p,\gamma) = P(X \le x)$$
$$= C(p,\gamma) \begin{cases} p^{-[x]}\left[\frac{2}{1-p} + \gamma^2 \frac{p^2([x]+1)^2 - p(2[x]^2+2[x]-1)+[x]^2}{(1-p)^3}\right. \\ \left. -2\gamma\frac{(1-p)[x]-p}{(1-p)^2}\right], & x < 0 \\ \frac{2(1+p)}{1-p} + 2\gamma^2\frac{p(1+p)}{(1-p)^3} - p^{[x]+1}\left[\gamma^2\frac{(1-p)^2([x]+1)^2+2p(1-p)([x]+1)+p(1+p)}{(1-p)^3}\right. \\ \left. -2\gamma\frac{(1-p)([x]+1)+p}{(1-p)^2} + \frac{2}{1-p}\right], & x \ge 0. \end{cases}$$

3. Moments

The moment generating function of a random variable $X \sim DASL(p,\gamma)$ is given by

$$M_X(t) = E(e^{tX}) = C(p,\gamma)\left[\frac{2p}{e^t - p} + \frac{2}{1-pe^t} + 2\gamma\frac{pe^t}{(e^t-p)^2}\right.$$
$$\left. - 2\gamma\frac{pe^t}{(pe^t-1)^2} + \gamma^2\frac{pe^t(p+e^t)}{(e^t-p)^3} + \gamma^2\frac{pe^t(pe^t+1)}{(1-pe^t)^3}\right], \quad |t| > \log p.$$

Replacing t in $M_X(t)$ by it, $i=\sqrt{-1}$, we can easily obtain the characteristic function of $DASL(p,\gamma)$.

To find the moments, using the combinatorial identity

$$\sum_{x=1}^{\infty} x^n p^x = \sum_{x=1}^{n} S(n,x)\frac{x!p^x}{(1-p)^{x+1}},$$

(see, e.g., formula (7.46), p. 337, of Graham et al., 1989), where

$$S(n,x) = \frac{1}{x!}\sum_{k=0}^{x-1}(-1)^k\binom{x}{k}(x-k)^n$$

is the Stirling number of the second kind, we obtain the n-th moment of $X \sim DASL(p,\gamma)$ for $n \geqslant 1$ as

$$\begin{aligned}\mu_n = E(X^n) = C(p,\gamma)\sum_{x=1}^{n}\frac{x!p^x}{(1-p)^{x+1}}\Bigg[&\Big(2S(n,x)+\gamma^2 S(n+2,x)\Big)\\ &\times\Big(1+(-1)^n\Big)-2\gamma S(n+1,x)\Big(1+(-1)^{n+1}\Big)\Bigg]\\ &+C(p,\gamma)\frac{p^{n+1}(n+1)!}{(1-p)^{n+2}}\Bigg\{[\gamma^2(1+(-1)^n)(S(n+2,n+1)\\ &+p\frac{n+2}{1-p}S(n+2,n+2))-2\gamma(1+(-1)^{n+1})S(n+1,n+1)]\Bigg\}.\end{aligned} \tag{5}$$

We can easily observe that for even n,

$$\begin{aligned}\mu_n = 2C(p,\gamma)\Bigg\{&\sum_{x=1}^{n}\frac{x!p^x}{(1-p)^{x+1}}\Big[2S(n,x)+\gamma^2 S(n+2,x)\Big]\\ &+\frac{p^{n+1}(n+1)!}{(1-p)^{n+2}}\Big[\gamma^2(S(n+2,n+1)+p\frac{n+2}{1-p}S(n+2,n+2))\Big]\Bigg\}.\end{aligned}$$

and for odd n,

$$\begin{aligned}\mu_n = -4\gamma C(p,\gamma)\Bigg\{&\sum_{x=1}^{n}\frac{x!p^x}{(1-p)^{x+1}}S(n+1,x)\\ &+\frac{p^{n+1}(n+1)!}{(1-p)^{n+2}}S(n+1,n+1)\Bigg\}\end{aligned}$$

In particular, we have

$$E(X) = -2\gamma \frac{p}{(1-p)^2}[1+\gamma^2 \frac{p}{(1-p)^2}]^{-1},$$

$$E(X^2) = p\Big[\frac{2}{(1-p)^2} + \gamma^2 \frac{(p^2+10p+1)}{(1-p)^4}\Big][1+\gamma^2 \frac{p}{(1-p)^2}]^{-1},$$

$$E(X^3) = -2\gamma p \frac{p^2+10p+1}{(1-p)^4}[1+\gamma^2 \frac{p}{(1-p)^2}]^{-1},$$

$$E(X^4) = \frac{p}{(1-p)^4}\Big[2(p^2+10p+1) + \frac{\gamma^2(p^4+56p^3+246p^2+56p+1)}{(1-p)^2}\Big]$$
$$[1+\gamma^2 \frac{p}{(1-p)^2}]^{-1}$$

and thus

$$Var(X) = \frac{p}{(1-p)^2}\Big[2+\gamma^2 \frac{p^2+10p+1}{(1-p)^2} - 4\gamma^2 \frac{p}{(1-p)^2+\gamma^2 p}\Big][1+\gamma^2 \frac{p}{(1-p)^2}]^{-1}.$$

Skewness and kurtosis of our distribution can be evaluated easily. But since their formulas are too long, they are omitted and we only show their behaviour by their graphs in Figure 2.

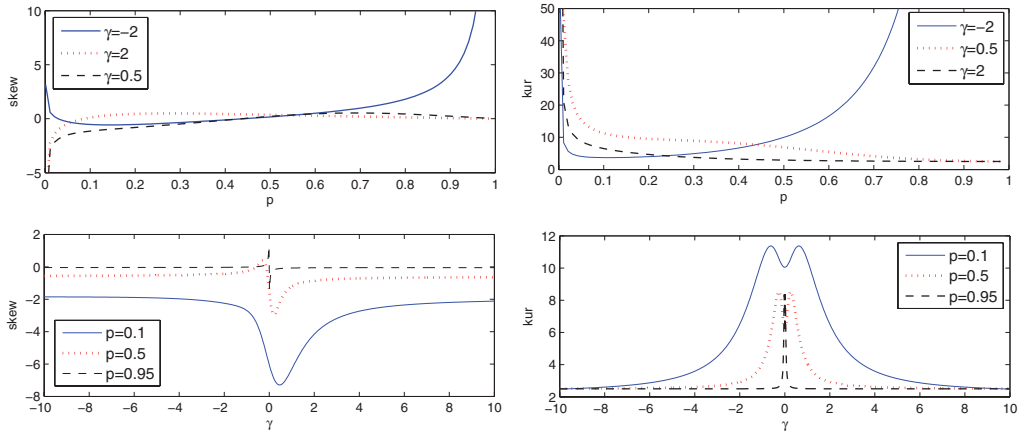


Figure 2: *Illustrations of the skewness and kurtosis as functions of p and γ.*

4. Maximum likelihood estimation

To apply maximum likelihood for estimating p and γ, assume that $x_1, x_2, \ldots, x_n$ are the observed values of a random sample of size n from a $DASL(p,\gamma)$ distribution. The log-likelihood function becomes

$$\ell(p,\gamma) = -n\log 2 + n\log(1-p) - n\log(1+p) - n\log(1+\gamma^2\frac{p}{(1-p)^2}) + \sum_{i=1}^{n}\log(1+(1-\gamma x_i)^2) + \sum_{i=1}^{n}|x_i|\log p.$$

Then, the likelihood equations for p and γ are given by

$$\frac{\partial \ell(p,\gamma)}{\partial p} = \frac{\sum_{i=1}^{n}|x_i|}{p} - \frac{2n}{1-p^2} - n\gamma^2\frac{1+p}{(1-p)^3+p\gamma^2(1-p)} = 0 \tag{6}$$

and

$$\frac{\partial \ell(p,\gamma)}{\partial \gamma} = -\frac{2n\gamma p}{(1-p)^2+p\gamma^2} - 2\sum_{i=1}^{n}x_i\frac{(1-\gamma x_i)}{1+(1-\gamma x_i)^2} = 0. \tag{7}$$

The solutions of likelihood equations (7) and (8) provide the maximum likelihood estimators (MLEs) of p and γ, which can be obtained by a numerical method such as the Newton-Raphson type procedure.

Since the MLEs of the unknown parameters (p,γ) can not be obtained in closed forms, it is not easy to derive the exact distributions of MLEs. One can show that the *DASL* family satisfies the regularity conditions which are fulfilled for parameters in the interior of the parameter space but not on the boundary (see, e.g., Ferguson, 1996, pp. 121). Hence, by using the simplest large sample approach, the MLE vector $(\hat{p},\hat{\gamma})$ is consistent and asymptotically normal, i.e.,

$$(\widehat{p}-p, \widehat{\gamma}-\gamma) \to N_2(\mathbf{0}, I^{-1}(\widehat{p},\widehat{\gamma})),$$

where I^{-1} is the variance covariance matrix of the unknown parameters (p,γ) and the covariance matrix I^{-1}, as the Fisher information matrix, can be obtained by

$$I^{-1}(p,\gamma) = \begin{bmatrix} -E\left(\frac{\partial^2\ell}{\partial p^2}\right) & -E\left(\frac{\partial^2\ell}{\partial p\partial\gamma}\right) \\ -E\left(\frac{\partial^2\ell}{\partial\gamma\partial p}\right) & -E\left(\frac{\partial^2\ell}{\partial\gamma^2}\right) \end{bmatrix}^{-1} = \begin{bmatrix} Var(\widehat{p}) & Cov(\widehat{p},\widehat{\gamma}) \\ Cov(\widehat{p},\widehat{\gamma}) & Var(\widehat{\gamma}) \end{bmatrix},$$

whose elements are evaluated by using the following expressions:

$$\frac{\partial^2 \ell}{\partial p^2} = -\frac{\sum_{i=1}^{n} |x_i|}{p^2} - 4n\frac{p}{(1-p^2)^2} - n\gamma^2 \frac{2(1-p)^2(2-p)+\gamma^2(p^2+2p-1)}{[(1-p)^3+p\gamma^2(1-p)]^2},$$

$$\frac{\partial^2 \ell}{\partial p \partial \gamma} = -2n\gamma \frac{1-p^2}{[(1-p)^2+\gamma^2 p]^2}$$

and

$$\frac{\partial^2 \ell}{\partial \gamma^2} = 2n\frac{p}{(1-p)^2+\gamma^2 p}[\frac{(1-p)^2-\gamma^2 p}{(1-p)^2+\gamma^2 p}] - 2\sum_{i=1}^{n} x_i^2 \frac{1-(1-\gamma x_i)^2}{[1+(1-\gamma x_i)^2]^2}.$$

To find expectations of the above expressions, we need to compute $E|X|$ and $E\{X^2 \frac{1-(1-\gamma X)^2}{[1+(1-\gamma X)^2]^2}\}$. The Fisher's information matrix can be computed using the approximation

$$I(\hat{p},\hat{\gamma}) = -\begin{bmatrix} \frac{\partial^2 \ell}{\partial p^2}|_{\hat{p},\hat{\gamma}} & \frac{\partial^2 \ell}{\partial p \partial \gamma}|_{\hat{p},\hat{\gamma}} \\ \frac{\partial^2 \ell}{\partial \gamma \partial p}|_{\hat{p},\hat{\gamma}} & \frac{\partial^2 \ell}{\partial \gamma^2}|_{\hat{p},\hat{\gamma}} \end{bmatrix},$$

as the observed Fisher's information matrix.

The normal approximation can then be used to construct confidence intervals for p and q to test hypothesis of the kind $H_0 : p = p_0$ and $H_0 : \gamma = \gamma_0$, respectively, as

$$(\hat{p} - z_{\alpha/2} I(\hat{p}), \hat{p} + z_{\alpha/2} I(\hat{p}))$$

and

$$(\hat{\gamma} - z_{\alpha/2} I(\hat{\gamma}), \hat{\gamma} + z_{\alpha/2} I(\hat{\gamma})).$$

where $I(\hat{p})$ and $I(\hat{\gamma})$ refer to the roots of diagonal elements of the inverse Fisher's information matrix.

5. Simulation

Here, we assess the performance of the maximum-likelihood estimate given by Equations (7) and (8) with respect to the sample size n. The simulation study assessment is based on the inversion method with 1000 iterations.

Table 1: *MLEs of p and γ in DASL(p,γ) for different values of n.*

γ	-2							
p	0.25				0.6			
n	$\hat{p}$	$\hat{\gamma}$	$\hat{Var}(\hat{p})$	$\hat{Var}(\hat{\gamma})$	$\hat{p}$	$\hat{\gamma}$	$\hat{Var}(\hat{p})$	$\hat{Var}(\hat{\gamma})$
100	0.2521	-2.0304	0.0009	0.1699	0.5994	-2.3095	0.0004	1.3711
200	0.2499	-2.0199	0.0004	0.0796	0.5997	-2.0784	0.0002	0.2564
400	0.2501	-2.0081	0.0002	0.0324	0.5999	-2.0431	0.0001	0.1072
γ	-0.5							
n	$\hat{p}$	$\hat{\gamma}$	$\hat{Var}(\hat{p})$	$\hat{Var}(\hat{\gamma})$	$\hat{p}$	$\hat{\gamma}$	$\hat{Var}(\hat{p})$	$\hat{Var}(\hat{\gamma})$
100	0.2411	-0.5797	0.0023	0.1270	0.5974	-0.5193	0.0023	0.0413
200	0.2451	-0.5363	0.0011	0.0467	0.6002	-0.5046	0.0005	0.0069
400	0.2481	-0.5126	0.0004	0.0121	0.5999	-0.5031	0.0002	0.0034
γ	0.5							
n	$\hat{p}$	$\hat{\gamma}$	$\hat{Var}(\hat{p})$	$\hat{Var}(\hat{\gamma})$	$\hat{p}$	$\hat{\gamma}$	$\hat{Var}(\hat{p})$	$\hat{Var}(\hat{\gamma})$
100	0.2434	0.5697	0.0020	0.0935	0.5994	0.5160	0.0010	0.0165
200	0.2461	0.5378	0.0011	0.0458	0.5997	0.5083	0.0004	0.0069
400	0.2475	0.5106	0.0005	0.0125	0.5998	0.5048	0.0002	0.0032
γ	1.5							
n	$\hat{p}$	$\hat{\gamma}$	$\hat{Var}(\hat{p})$	$\hat{Var}(\hat{\gamma})$	$\hat{p}$	$\hat{\gamma}$	$\hat{Var}(\hat{p})$	$\hat{Var}(\hat{\gamma})$
100	0.2523	1.5352	0.0014	0.1430	0.6002	1.6370	0.0004	0.3117
200	0.2512	1.5146	0.0006	0.0599	0.6001	1.5402	0.0002	0.0845
400	0.2507	1.5079	0.0003	0.0293	0.6001	1.5231	0.0001	0.0356

These results are presented in Table 1 accompanied by their estimated variances ($\hat{Var}$), for different values of n. Table 1 shows how the MLEs and estimated variances of parameters vary with respect to n. The difference between real and estimated values of the parameters are not too large and, thus, the method works well.

6. Some special cases

In this section, we consider the distribution of the random variable $X \sim DASL(p,\gamma)$ truncated at zero. This distribution is an important case, because it is a weighted version of geometric distribution and may be useful in fitting count or time data sets.

Let $Y = X|X \geq 0$, then the pmf of Y is given by

$$p_Y(y;p,\gamma) = C^*(p,\gamma)(1+(1-\gamma y)^2)p^y,\ y = 0,1,2,\ldots,$$

where $C^{*^{-1}}(p,\gamma) = \frac{2}{1-p} - 2\gamma\frac{p}{(1-p)^2} + \gamma^2\frac{p(1+p)}{(1-p)^3}$, $0 < p < 1$ and $\gamma \in \mathbb{R}$. This distribution is called weighted geometric distribution and is denoted by $Y \sim WGD(p,\gamma)$.

This can be used as a discrete lifetime distribution which contains geometric distribution by setting $\gamma = 0$. The survival and failure rate functions of this random variable are given by

$$R_Y(y;p,\gamma) = P(Y > y) = C^*(p,\gamma)\frac{p^{y+1}}{1-p}[2+2\gamma\frac{yp-y-1}{1-p} + \gamma^2\frac{y^2p^2+(-2y^2-2y+1)p+(y+1)^2}{(1-p)^2}],\ y = 0,1,\ldots$$

and

$$H_Y(y;p,\gamma) = (1-p)\frac{1+(1-\gamma y)^2}{2p+2p\gamma^{\frac{yp-y-1}{1-p}}+\gamma^2 p^{\frac{y^2p^2+(-2y^2-2y+1)p+(y+1)^2}{(1-p)^2}}},\ y = 0,1,\ldots,$$

respectively. The behaviour of the failure rate function of $X \sim WGD(p,\gamma)$ is described in Figure 3. As we can see the failure rate function of WGD distribution can be increasing or U-shaped. Further, we note that if $\gamma = 0$, WGD distribution will reduce to the geometric distribution with constant failure rate function which depends only on p.

Another important structural property of a distribution, both in theory and application, is its infinite divisibility. We refer, for example, to the monograph of Steutel and Van Harn (2004) for a good and complete introduction of the subject. Since most of the well-known distributions possess this property, one has to be concerned with the infinite divisibility or non-infinite divisibility property of any distribution newly introduced. Here, we note that $WGD(p,\gamma)$ distribution is not infinitely divisible. To see this, we first recall the following interesting result from the above-mentioned monograph (page 56).

Lemma 1 *If p_k , $k \in \mathbb{Z}_+$ is infinitely divisible, then we have $p_k \leq 1/e$, for all $k \in \mathbb{N}$.*

Now, we can show that $p_Y(y;p,\gamma) > 1/e$ for some values of $y \in \mathbb{N}$, p and γ. For instance, take $y = 1$, $p = 0.1$ and $\gamma = 10$. Then, we see that $p_Y(1;0.1,10) \simeq 0.5525 > 1/e \simeq 0.3679$. Thus, a $WGD(p,\gamma)$ distribution is not infinitely divisible in general. In the case $\gamma = 0$, however, we have the geometric distribution, with probability of success p, which is obviously infinitely divisible.

It is also worth noting that, we can describe the distribution of the random variable $Z = |X|$ as a new distribution on the set of non-negative integers as follows:

$$p^*(z;p,\gamma) = P(|X| = z) = 2C(p,\gamma)\begin{cases} 1, & z = 0 \\ \{2+\gamma^2 z^2\}p^z, & z = 1,2,\ldots. \end{cases}$$

where p and γ are given as before. This distribution is called a generalized geometric distribution and denoted by $Z \sim GGD(p,\gamma)$. It is worth mentioning that if $Z \sim GGD(p,\gamma)$,

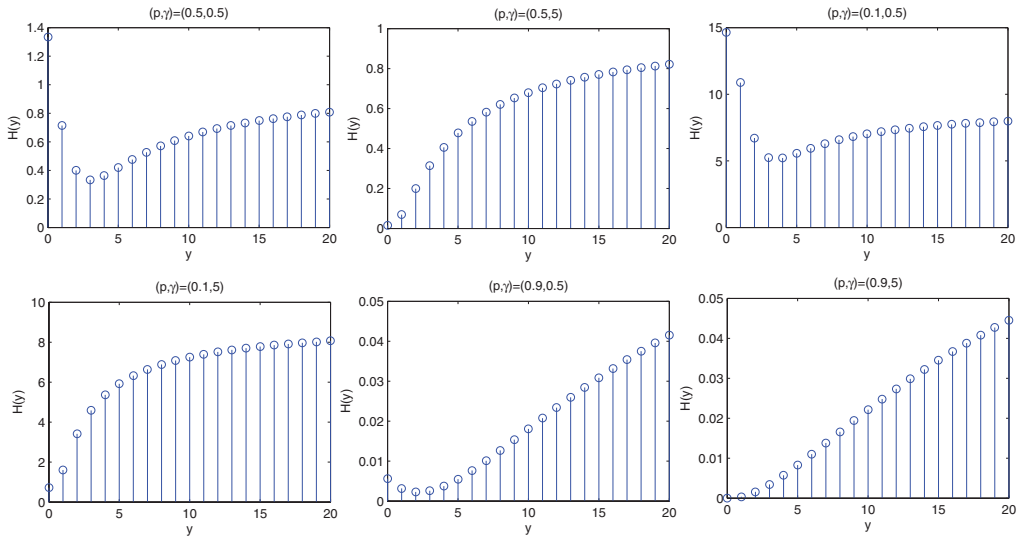


Figure 3: *Illustrations of the failure rate function of $X \sim WGD(p,\gamma)$ for some selected values of p and γ.*

then we have

1. If $\gamma = 0$, $Z \sim \frac{1-p}{1+p}\begin{cases} 1, & z=0 \\ 2p^z, & z=1,2,\ldots \end{cases}$, whose truncation at zero is a geometric distribution with parameter p and support on $\{1,2,\ldots\}$.

2. $GGD(p,\gamma) \cong GGD(p,-\gamma)$.

3. If $\gamma \longrightarrow \mp\infty$, then $p^*(z;p,\gamma) \longrightarrow \frac{(1-p)^3}{p(1+p)} z^2 p^z, \quad z=1,2,\ldots.$

7. Application and comparison

In this section, we attempt to examine application and advantage of $DASL(p,\gamma)$ and $WGD(p,\gamma)$ distributions comparing to several rival models using some real data sets.

Example. The following data set is obtained[1] based on a recent local research carried out on the extent of success of Iranian universities in transferring technology to industry and their effective factors. Out of 500 questionaries distributed, 111 were returned. The data below show the difference between the desired and the existing state values on each sample; a positive number shows the extent of positive improvement and a negative number shows the extent of negative improvement.

1. The data is part of an unpublished research by A. Rafiei, an MSc student.

$-1, 1, 3, 3, 1, 1, -20, -14, -15, -14, -5, -6, 0, 7, 4, 1, 9, 13, 5, 4, 1, 1, 1, 14, 5, -2, 4, 5, 3, 3, 12, 3, 4, 4, 5, -1, 5, 3, 4, 3, 4, 6, 6, 7, 0, 0, 0, 13, 0, 10, 15, 2, 2, 5, 11, 2, 2, -16, 2, 8, -7, -7, 2, -2, 9, 6, 11, -5, -5, 13, 13, 1, 14, 0, 8, -5, -2, 10, 2, 10, 8, -3, 10, 12, 14, 12, 11, 11, 8, 0, -2, 4, 6, 0, 0, 7, 8, 1, 9, -1, 9, 6, 11, 0, 7, 7, 10, 4, 7, 9, 28.$

Thus, it is logical to compare our distribution with some similar distributions such as *SkewDL* and *ADSLaplace* distributions. *SkewDL* distribution was introduced by Kozubowski and Inusah (2006) and has the following pmf

$$p(x;p,q) = \frac{(1-p)(1-q)}{1-pq} \begin{cases} q^{-x}, & x = \ldots, -2, -1 \\ p^x, & x = 0, 1, \ldots. \end{cases}$$

ADSLaplace distribution of Barbiero (2014) has pmf

$$p(x;p,q) = \frac{1}{\log(pq)} \begin{cases} \log(p)[q^{-(x+1)}(1-q)], & x = \ldots, -2, -1 \\ \log(q)[p^x(1-p)], & x = 0, 1, \ldots. \end{cases}$$

Furthermore, we shall also consider the Skellam (Skellam, 1946) with pmf

$$p(x;\mu_1,\mu_2) = e^{-\mu_1-\mu_2}(\mu_1/\mu_2)I_x(2\sqrt{\mu_1\mu_2}), \quad x \in \mathbb{Z}\,.$$

where $I_x(2\sqrt{\mu_1\mu_2})$ is modified Bessel function of the first kind, and discrete normal (Roy, 2003) distribution with pmf:

$$p(x;\mu,\sigma) = \Phi(x+1,\mu,\sigma) - \Phi(x,\mu,\sigma), \quad x \in \mathbb{Z}\,.$$

where $\Phi(.,\mu,\sigma)$ is the cdf of normal distribution with mean μ and variance σ^2, respectively.

The results of comparison are illustrated in Table 2. We have also obtained maximum likelihood estimates and their estimation of standard errors for the parameters involved. We note that under regularity conditions, the standard error of the parameter estimators can be asymptotically computed by root square of the diagonal elements of the inverted Fisher's matrix. The Kolomogrov-Smirnov (K-S) statistic and Akaike information criterion as AIC=$-2\log L + 2k$, where k, the number of parameters in the model, n, the sample size, and L, the maximized value of the likelihood function for the estimated model, are used to compare the estimated models.

Since $DASL(p,\gamma)$ distribution is an extension of DL distribution, in our iterative algorithm of Newton-Raphson, we have used $\gamma = 0$ and the MLE of parameters of DL distribution as initial values to find the MLEs of the parameters. As one can see from Table 2, our model is preferable comparing to other models. Also, Figure 4 shows distribution plots of the data and the models in question.

Table 2: *Comparing criterions for the rival distributions.*

Model	*Parameter estimaties*	k	$K-S$	$\log L$	AIC
DL	$\widehat{p}=0.8496$, S.E$(\widehat{p})=0.0108$	1	0.3605	-389.054	780.107
ADSLaplace	$\widehat{p}=0.8787$, S.E$(\widehat{p})=0.0085$ $\widehat{q}=0.7530$, S.E$(\widehat{q})=0.0394$	2	0.2162	-379.864	763.728
SkewDL	$\widehat{p}=0.8732$, S.E$(\widehat{p})=0.0092$ $\widehat{q}=0.7605$, S.E$(\widehat{q})=0.0368$	2	0.1973	-377.098	758.196
dnormal	$\widehat{\mu}=4.2048$, S.E$(\widehat{\mu})=0.2271$ $\widehat{\sigma}=6.9769$, S.E$(\widehat{\sigma})=0.1607$	2	0.1423	-373.189	750.378
DASL	$\widehat{p}=0.7258$, S.E$(\widehat{p})=0.0225$ $\widehat{\gamma}=-0.3120$, S.E$(\widehat{\gamma})=0.0785$	2	0.1193	-366.131	736.262
Skellam	$\widehat{\mu_1}=26.0389$, S.E$(\widehat{\mu_1})=0.3599$ $\widehat{\mu_2}=22.3355$, S.E$(\widehat{\mu_2})=0.3087$	2	0.1421	-373.102	750.204

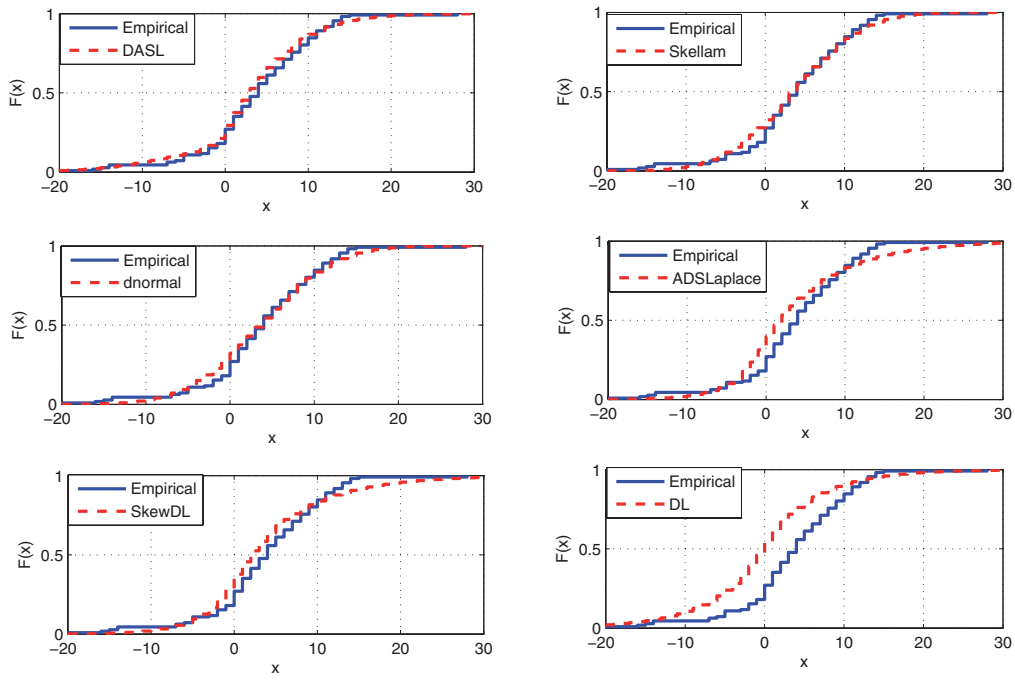


Figure 4: *Plots of empirical distribution functions for the data set and the fitted distributions.*

In addition, we can use the likelihood ratio (LR) test statistic to confirm our claim. To do this, we consider the following test of hypotheses

$$H_0 : \gamma = 0 (DL(p)) \qquad v.s \qquad H_1 : \gamma \neq 0 (DASL(p, \gamma)).$$

Observed value of the likelihood ratio (LR) test statistic is 43.845 while its tabulated value equals $\chi^2_1 = 3.84$. Thus the null hypothesis is rejected.

On the other hand, Figure 4 shows our different fitted distribution functions and the empirical distribution of the data set. From these plots, we can see that our distribution function better fits the data set.

Acknowledgments

The authors would like to sincerely thank two anonymous reviewers for their careful reading of the paper and their valuable points and comments which led to considerable improvements. The first author is also grateful to the Graduate Office of the University of Isfahan for their support.

References

Barbiero, A. (2014). An alternative discrete skew Laplace distribution. *Statistical Methodology*, 16, 47–67.

Ferguson, T. S. (1996). *A Course in Large Sample Theory*. Chapman and Hall, London.

Graham, R. L., Knuth, D. E. and Patashnik, O. (1989). *Concrete Mathematics*. Addison-Wesley Publishing Company, Reading, MA.

Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Table of Integrals, Series, and Products*, 7th ed. Academic Press, San Diego, CA.

Inusah, S. and Kozubowski, T. J. (2006). A discrete analogue of the Laplace distribution. *Journal of Statistical Planning and Inference*, 136, 1090–1102.

Jayakumar, K. and Jacob, S. (2012). Wrapped skew Laplace distribution on integers: A new probability model for circular data. *Open Journal of Statistics*, 2, 106–114.

Kozubowski, T. J. and Inusah, S. (2006). A skew Laplace distribution on integers. *Annals of the Institute of Statistical Mathematics*, 58, 555–571.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd ed. John Wiley and Sons, New York.

Molina, J. A. D and Arteaga, A. R. (2007). On the infinite divisibility of some skewed symmetric distributions, *Statistics and Probability Letters*, 77, 644–648.

Roy, D. (2003). The discrete normal distribution. *Communications in Statistics: Theory and Methods*, 32, 1871–1883.

Shams Harandi, S. and Alamatsaz, M. H. (2013). Alpha-Skew-Laplace distribution. *Statistics and Probability Letters*, 83, 774–782.

Skellam, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society, Series A*, 109, 296.

Steutel, F. W and Van Harn, K. (2004). *Infinite Divisibility of Probability Distributions on the Real Line*. Marcel Dekker, Inc, New York.

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. John Wiley and Sons, New York.

SORT 39 (1) January-June 2015, 85-108

A mathematical programming approach for different scenarios of bilateral bartering

Stefano Nasini, Jordi Castro and Pau Fonseca*

Abstract

The analysis of markets with indivisible goods and fixed exogenous prices has played an important role in economic models, especially in relation to wage rigidity and unemployment. This paper provides a novel mathematical programming based approach to study pure exchange economies where discrete amounts of commodities are exchanged at fixed prices. Barter processes, consisting in sequences of elementary reallocations of couple of commodities among couples of agents, are formalized as local searches converging to equilibrium allocations. A direct application of the analysed processes in the context of computational economics is provided, along with a Java implementation of the described approaches.

MSC: 90C27, 90C29, 90C30, 91Bxx.

Keywords: Numerical optimization, combinatorial optimization, microeconomic theory.

1. Introduction

Since the very beginning of the Economic Theory (Edgeworth, 1881; Jevons, 1888), the bargaining problem has generally be adopted as the basic mathematical framework for the study of markets of excludable and rivalrous goods. It concerns the allocation of a fixed quantity among a set of self-interested agents. The characterizing element of a bargaining problem is that many allocations might be simultaneously suitable for all the agents.

Definition 1 *Let $\mathcal{V} \subset R^n$ be the space of allocations of an n agents bargaining problem. Points in $\mathcal{V}$ can be compared by saying that $v^* \in \mathcal{V}$ strictly dominates $v \in \mathcal{V}$ if each component of v^* is not less than the corresponding component of v and at least*

* Dept. of Statistics and Operations Research. Universitat Politècnica de Catalunya.
Received: January 2014
Accepted: September 2014

one component is strictly greater, that is, $v_i \leq v_i^*$ for each i and $v_i < v_i^*$ for some i. This is written as $v \prec v^*$. Then, the Pareto frontier is the set of points of $\mathcal{V}$ that are not strictly dominated by others.

A long-standing line of research focused on axiomatic approaches for the determination of a uniquely allocation, satisfying agents' interests (Nash, 1951; Rubinstein, 1983).

More recently, an increasing attention has been devoted to the cases where the quantity to be allocated is not infinitesimally divisible. The technical difficulties associated to those markets have been often pointed (Kaneko, 1982; Quinzii, 1984; Scarf, 1994) and the equilibria of markets of indivisible goods have been characterized only under strong assumptions (Shapley and Shubik, 1972). In the general case, main focus was to address the question of existence of market clearing prices in the cases of not infinitesimally divisible allocations (Danilov, Koshevoy and Murota, 2001; Caplin and Leahy, 2010).

Another subclass of the family of bargaining problems is associated to markets with fixed prices (Dreze, 1975; Auman and Dreze, 1986), which have played an important role in macroeconomic models, especially on those models related to wage rigidities and unemployment. Dreze described price rigidity as inequality constraints on individual prices (Dreze, 1975).

Efficient algorithms to find non-dominated Pareto allocations of bargaining problems associated to markets with not infinitesimally divisible goods and fixed exogenous prices have been recently studied (Vazirani et al., 2007; Ozlen, Azizoglu and Burton, 2012). Our goal is to provide novel mathematical-programming based approaches to analyse barter processes, which are commonly used in everyday life by economic agents to solve bargaining problems associated to n-consumer-m-commodity markets of not infinitesimally divisible goods and fixed exogenous prices. These processes are based on *elementary reallocations* (ER) of two commodities among two agents, sequentially selected from the $m(m-1)n(n-1)/4$ possible combinations. Under fixed prices, markets do not clear and the imbalance between supply and demand is resolved by some kind of quantity rationing (Dreze, 1975). In our analysis this quantity rationing is implicit in the process and not explicitly taken into account.

Based on this multi-agent approach, many economical systems might be simulated (Wooldridge, 2002). For instance, some studies (Bell, 1998; Wilhite, 2001) have taken into account the effect of network structures on the performance of a barter process, for the case of endogenous prices and continuous commodity space, showing that centralized network structures, such as a stars, exhibit a faster convergence to an equilibrium allocation. Our multi-agent approach is instead devoted to the analysis of the network structures generated by the sequences of bilateral trades, namely the set of couples of agents interacting along the processes. Such a structure might be statistically analysed in term of its topological properties, as it is done in Section 5.

Section 2 illustrates the fundamental properties of the allocation space, associated to n-consumer-m-commodity markets of not infinitesimally divisible goods and fixed

exogenous prices. Section 3 provides a general mathematical programming formulation and derives an analytical expression for the Pareto frontier of the *elementary reallocation problem* (ERP). It will be shown that the *sequence of elementary reallocations* (SER) (the chain of ERP performed by agents along the interaction process) follows the algorithmic steps of a local search in the integer allocation space with exogenous prices. Section 4 introduce the case of network structures restricting agents interactions to be performed only among adjacent agents. In Section 5 the performance of these barter processes is compared with the one of a global optimization algorithm (branch and cut).

2. The integer allocation space with exogenous prices

The key characteristic of an economy is: a collection $\mathcal{A}$ of n agents, a collection $\mathcal{C}$ of m types of commodities, a commodity space X (usually represented by the nonnegative orthant in $\mathbb{R}^m$), the initial endowments $e^i_j \in X$ for $i \in \mathcal{A}$, $j \in \mathcal{C}$ (representing a budget of initial amount of commodities owned by each agent), a preference relation $\preceq_i$ on X for each agent $i \in \mathcal{A}$. It has been shown (Arrow and Debreu, 1983) that if the set $\{(x,y) \in X \times X : x \preceq_i y\}$ is closed relative to $X \times X$ the preference relation can be represented by a real-valued function $u^i : X \longmapsto \mathbb{R}$, such that, for each a and b belonging to X, $u^i(a) \leqslant u^i(b)$ if and only if $a \preceq b$.

When agents attempt to simultaneously maximize their respective utilities, conditioned to balance constraints, the resulting problems are $\max u^i(\mathbf{x})$ s.to $\sum_{i\in\mathcal{A}} x^i_j = \sum_{i\in\mathcal{A}} e^i_j$ for $j \in \mathcal{C}$, where $x^i_j \in X$, is the amount of commodity j demanded by agent i (from now on the superindex shall denote the agent and the subindex shall denote the commodity).

Under certain economic conditions (convex preferences, perfect competition and demand independence) there must be a vector of prices $\widehat{P} = (\widehat{p}_1, \widehat{p}_2, \widehat{p}_3, \ldots, \widehat{p}_m)^{\mathsf{T}}$, such that aggregate supplies will equal aggregate demands for every commodity in the economy (Arrow and Debreu, 1983).

As studied by Dreze (1975) and by Auman and Dreze (1986), when prices are regarded as fixed, markets do not clear and the imbalance between supply and demand is resolved by some kind of quantity rationing. The system of linear constraints associated to a n-consumer-m-commodity market with fixed prices exhibits a block angular structure with rank $m+n-1$:

$$\begin{bmatrix} p_1\, p_2 \cdots p_m & & & & \\ & p_1\, p_2 \cdots p_m & & & \\ & & & \ddots & \\ & & & & p_1\, p_2 \cdots p_m \\ I & I & \cdots & I & \end{bmatrix} \mathbf{x} = \begin{bmatrix} p_1 e^1_1 + \cdots + p_m e^1_m \\ p_1 e^2_1 + \cdots + p_m e^2_m \\ \vdots \\ p_1 e^n_1 + \cdots + p_m e^n_m \\ \mathbf{e}^1 + \cdots + \mathbf{e}^n \end{bmatrix}, \quad (1)$$

where $p_1,\ldots,p_m$ are relative prices between commodities, $\mathbf{e}^i = (e^i_1,\ldots,e^i_m)^\mathsf{T}$, and $\mathbf{x} = (x^1_1,\ldots,\ x^1_m,\ldots,x^n_1,\ldots,x^n_m)^\mathsf{T}$. The constraints matrix of (1) could also be written as $\begin{pmatrix} I\otimes P \\ \mathbf{1}\otimes I \end{pmatrix}$, where $P = (p_1,p_2,p_3,\ldots,p_m)$ and $\otimes$ is the Kronecker product between two matrices. Note that the linking constrains (i.e., the conservation of commodities $(\mathbf{1}\otimes I)\mathbf{x} = \mathbf{e}^1+\cdots+\mathbf{e}^n$) are implied by the balance equations of a network flow among the agents. This fact will be analysed in Section 5, where we introduced costs associated to the flow.

All the feasible allocations lay in a $(m+n-1)$ dimensional hyperplane defined by the prices (always containing at least one solution, which is represented by the vector of initial endowments $\mathbf{e}$), and restricted to the fact that agents are rational: $u^i(\mathbf{x}) \geq u^i(\mathbf{e})$, for $i \in \mathcal{A}$.

Proposition 1 below shows that an asymptotic approximation of an upper bound of the number of nonnegative solutions of (1) is $\mathcal{O}(\frac{n^{(mb)}}{b^m})$, where b is the average amount of commodities available, i.e., $b = \frac{\sum_{j=1}^m(\sum_{h=1}^n e^h_j)}{m}$.

Proposition 1 *Let Λ be the set of nonnegative solutions of (1), i.e., the allocation space of a problem of bargaining integer amounts of m commodities among n agents with fixed prices. If the allocation space satisfies the mild conditions $b_j = \sum_{h=1}^n e^h_j \geq n$ (where b_j is the overall amount of commodity j in the system, which is a fix quantity, associated to the rhs of (1)), then $|\Lambda| \in \mathcal{O}(\frac{n^{(mb)}}{b^m})$.*

Proof. The set of nonnegative solutions of (1) is a subset of the union of bounded sets, as $\Lambda \subset \bigcup_{j=1}^m\{(x^1_j\ldots x^n_j) \in \mathbb{R}^n : x^1_j+\cdots+x^n_j = e^1_j+\cdots+e^n_j; x^1_j\ldots x^n_j \geq 0\}$. Therefore, Λ is a finite set, as it is the intersection between $\mathbb{Z}$ and a bounded subset of $\mathbb{R}^{mn}$. Let Λ' be the set of nonnegative solutions of (1), without considering the price constraints, i.e., the n diagonal blocks $p_1x^h_1+p_2x^h_2+\cdots+p_mx^h_m = p_1e^h_1+p_2e^h_2+\cdots+p_me^h_m$, for $h = 1,\ldots,n$. We know that $|\Lambda'| \geq |\Lambda|$. However, $|\Lambda'|$ can be easily calculated, as the number of solutions of m independent Diophantine equations with unitary coefficients. The number of nonnegative integer solutions of any equation of the form $\sum_{h=1}^n x^h_j = b_j, j = 1,\ldots,m$, might be seen as the number of distributions of b_j balls among m boxes: $\frac{(n+b_j-1)!}{(n-1)!b_j!}$. Since we have m independent Diophantine equations of this form, then the number of possible solutions for all of them is $\prod_{j=1}^m \frac{(n+b_j-1)!}{(n-1)!b_j!}$. Thus, we know that $|\Lambda| \leq \prod_{j=1}^m \frac{(n+b_j-1)(n+b_j-2)\ldots n}{b_j!} \leq \prod_{j=1}^m \frac{(n+b_j-1)^{b_j}}{b_j!} \leq \frac{\prod_{j=1}^m(n+b_j-1)^{b_j}}{b^m}$, where the last inequality holds because $b_j \geq n \geq 2$. Finally, we conclude that

$$\frac{\prod_{j=1}^m(n+b_j-1)^{b_j}}{b^m} \leq \frac{\mathcal{O}(n)^{bm}}{b^m} \leq \mathcal{O}(\frac{n^{(mb)}}{b^m}). \qquad \blacksquare$$

In the next section we define a barter process of integer quantities of m commodities among n agents as a local search in the allocation space Λ (obtained as a sequence of elementary reallocations) and show that the Pareto frontier of the ERP might be analytically obtained without the use of any iterative procedure.

3. The sequence of elementary reallocations

As previously seen, the linear system characterizing the space of possible allocations is (1). Here the conservation of commodity (i.e., the overall amount of commodity of each type must be preserved) is generalized to include arbitrary weights in the last m rows of (1). Based on this observation consider the following multi-objective integer non-linear optimization problem (MINOP)

$$\max\{u^i(\mathbf{x}),\ i=1,\ldots,n\} \tag{2a}$$

subject to

$$\begin{bmatrix} P & & & \\ & P & & \\ & & \ddots & \\ & & & P \\ d^1 I & d^2 I & \ldots & d^n I \end{bmatrix} \mathbf{x} = \begin{bmatrix} b^1 \\ b^2 \\ \vdots \\ b^n \\ \mathbf{b}^0 \end{bmatrix} \tag{2b}$$

$$\begin{aligned} & u^i(\mathbf{x}) \geq u^i(\mathbf{e}) \quad i=1\ldots,n \\ & \mathbf{x} \in \mathbb{Z}^{mn} \geq 0, \end{aligned} \tag{2c}$$

where $u^i : \mathbb{R}^{mn} \to \mathbb{R}$, $P \in \mathbb{Q}^{1\times m}$, $d^i \in \mathbb{Q}$, $b^i \in \mathbb{Q}$, $i=1,\ldots,n$, and $\mathbf{b}^0 \in \mathbb{Q}^m$. The conditions $u^i(\mathbf{x}) \geq u^i(\mathbf{e}), \quad i=1,\ldots,n$, guarantee that no agent gets worse under a feasible reallocation, which is known in general bargaining literature as the *disagreement point*. The constraint matrix has a primal block-angular structure with n identical diagonal blocks involving m decision variables. Problem (1) is a particular case of (2) for $d_i = 1, i=1,\ldots,n$.

From a multi-objective optimization point of view, a suitable technique to generate the Pareto frontier of (2) is the ε-constraint method (Haimes et al., 1971). Recently, a general approach to generate all nondominated objective vectors has been developed (Ozlen and Azizoglu, 2009), by recursively identifying upper bounds on individual objectives using problems with fewer objectives.

3.1. The elementary reallocation problem

In everyday life, barter processes among people tend to achieve the Pareto frontier of problem (2) by a sequence of reallocations. We consider a process based on a sequence of two-commodity-two-agent reallocations, denoted as SER. Any step of this sequence requires the solution of a MINOP involving 4 variables and 4 constraints of problem (2).

Let $\mathbf{e}$ be a feasible solution of (2b) and (2c) and suppose we want to produce a feasible change of 4 variables, such that 2 of them belong to the ith and jth position of the diagonal block h and the other belong to the ith and jth position of the diagonal block k.

It can be easily shown that a feasibility condition of any affine change of these 4 variables $e_i^h+\Delta_i^h, e_i^k+\Delta_i^k, e_j^h+\Delta_j^h, e_j^k+\Delta_j^k$ is that $\Delta_i^h, \Delta_i^k, \Delta_j^h, \Delta_j^k$ must be an integer solution of the following system of equations

$$\begin{bmatrix} p_i & p_j & 0 & 0 \\ 0 & 0 & p_i & p_j \\ d^h & 0 & d^k & 0 \\ 0 & d^h & 0 & d^k \end{bmatrix} \begin{bmatrix} \Delta_i^h \\ \Delta_j^h \\ \Delta_i^k \\ \Delta_j^k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{3}$$

The solution set are the integer points in the null space of the matrix of system (3), which will be named A. A is a two-agent-two-commodity constraint matrix, and its rank is three (just note that the first column is a linear combination of the other three using coefficients $\alpha_2 = \frac{p_i}{p_j}$, $\alpha_3 = \frac{d^h}{d^k}$ and $\alpha_4 = -\frac{p_i d^h}{p_j d^k}$). Therefore the null space has dimension one, and its integer solutions are found on the line

$$\begin{bmatrix} \Delta_i^h \\ \Delta_j^h \\ \Delta_i^k \\ \Delta_j^k \end{bmatrix} = q \begin{bmatrix} p_j d^k \\ -p_i d^k \\ -p_j d^h \\ p_i d^h \end{bmatrix}, \tag{4}$$

for some $q = \alpha F(p_i, p_j, d^k, d^h)$, where $\alpha \in \mathbb{Z}$ and $F : \mathbb{Q}^4 \to \mathbb{Q}$ provides a factor which transforms the null space direction into the nonzero integer null space direction of smallest norm. We note that this factor can be computed as $F(p_i, p_j, d^k, d^h) = G(p_j d^k, p_i d^k, p_j d^h, p_i d^h)$, where

$$G(v_i = \frac{r_i}{q_i}, i = 1, \ldots, l) = \frac{\text{lcm}(q_i, i = 1, \ldots, l)}{\text{gcd}(\text{lcm}(q_i, i = 1, \ldots, l) \cdot v_i, i = 1, \ldots, l)}, \tag{5}$$

r_i and q_i being the numerator and denominator of v_i ($q_i = 1$ if v_i is integer), and lcm and gcd being, respectively, the least common multiple and greatest common divisor functions.

Hence, given a feasible point $\mathbf{e}$, one can choose 4 variables, such that 2 of them belong to the ith and jth position of a diagonal block h and the others belong to the ith and jth position of a diagonal block k, in $m(m-1)n(n-1)/4$ ways. Each of them constitutes an ERP, whose Pareto frontier is in $\mathbf{e}+\text{null}(A)$. The SER is a local search, which repeatedly explores a neighbourhood and chooses both a locally improving direction among the $m(m-1)n(n-1)/4$ possible ERPs and a feasible step length $q=\alpha F(p_i,p_j,d^k,d^h)$, $\alpha\in\mathbb{Z}$. For problems of the form of (2) the SER might be written as follows:

$$\mathbf{x}^{t+1}=\mathbf{x}^t+\alpha F(p_i,p_j,d^k,d^h)\begin{bmatrix}\vdots & \vdots\\ p_jd^k & h,i\\ \vdots & \vdots\\ -p_id^k & h,j\\ \vdots & \vdots\\ -p_jd^h & k,i\\ \vdots & \vdots\\ p_id^h & k,j\\ \vdots & \vdots\end{bmatrix}=\mathbf{x}^t+\alpha F(p_i,p_j,d^k,d^h)\Delta_{ij}^{kh}, \tag{6}$$

t being the iteration counter. In shorter notation, we write (6) as $\mathbf{x}^{t+1}=\mathbf{x}^t+\alpha S_{ij}^{kh}$, where

$$S_{ij}^{kh}=F(p_i,p_j,d^k,d^h)\Delta_{ij}^{kh} \tag{7}$$

is a direction of integer components. Since the nonnegativity of $\mathbf{x}$ has to be kept along the iterations, then we have that

$$-\frac{\max\left\{x_i^h/(p_jd^k),x_j^k/(p_id^h)\right\}}{F(p_i,p_j,d^k,d^h)}\le\alpha\le\frac{\min\left\{x_j^h/(p_id^k),x_i^k/(p_jd^h)\right\}}{F(p_i,p_j,d^k,d^h)}, \tag{8}$$

or, equivalently,

$$-\max\left\{x_i^h/(p_jd^k),x_j^k/(p_id^h)\right\}\le q\le\min\left\{x_j^h/(p_id^k),x_i^k/(p_jd^h)\right\}. \tag{9}$$

(The step length is forced to be nonnegative when the direction is both feasible and a descent direction; in our case the direction is only known to be feasible, and then negative step lengths are also considered.)

An important property of an elementary reallocation is that under the assumptions that $\dfrac{\partial u^k(\mathbf{x})}{\partial_{x_i^k}}:\mathbb{R}^{mn}\to\mathbb{R}$ is (i) non increasing, (ii) nonnegative and (iii) $\dfrac{\partial u^k(\mathbf{x})}{\partial_{x_i^j}}=0$

for $j \neq k$ (i.e., u^k only depends on $\mathbf{x}^k$), which are quite reasonable requirements for consumer utilities, then $u^k(\mathbf{x}+\alpha S_{ij}^{kh})$ is a unimodal function with respect to α, as shown by the next proposition.

Proposition 2 *Under the definition of u^k and S_{ij}^{kh}, for every feasible point $\mathbf{x} \in R^{mn}$, $u^k(\mathbf{x}+\alpha S_{ij}^{kh})$ is either a unimodal function with respect to α or locally constant beyond a certain value of α in the interval defined by (8).*

Proof. Let us define $g(\alpha) = u^k(\mathbf{x}+\alpha S_{ij}^{kh})$, differentiable with respect to α. It will be shown that for all α in the interval (8), and $0 < \tau \in \mathbb{R}$, $g'(\alpha) < 0$ implies $g'(\alpha+\tau) < 0$, which is a sufficient condition for the unimodality of $g(\alpha)$. By the chain rule, and using (6) and (7), the derivative of $g(\alpha)$ can be written as

$$\begin{aligned} g'(\alpha) &= \nabla_{\mathbf{x}} u^k(\mathbf{x}+\alpha S_{ij}^{kh}) S_{ij}^{kh} \\ &= F(p_i,p_j,d^k,d^h)\left(\frac{\partial u^k(\mathbf{x}+\alpha S_{ij}^{kh})}{\partial_{x_i^k}}(-p_j d^h) + \frac{\partial u^k(\mathbf{x}+\alpha S_{ij}^{kh})}{\partial_{x_j^k}} p_i d^h\right). \end{aligned} \tag{10}$$

If $g'(\alpha) < 0$ then, from (10) and since $F(p_i,p_j,d^k,d^h) > 0$, we have that

$$\frac{\partial u^k(\mathbf{x}+\alpha S_{ij}^{kh})}{\partial_{x_i^k}} p_j d^h > \frac{\partial u^k(\mathbf{x}+\alpha S_{ij}^{kh})}{\partial_{x_j^k}} p_i d^h. \tag{11}$$

Since from (6) the component (k,i) of S_{ij}^{kh} is $F(p_i,p_j,d^k,d^h)(-p_j d^h) < 0$, and $\dfrac{\partial u^k(\mathbf{x})}{\partial_{x_i^k}}$ is non increasing, we have that for $\tau > 0$

$$\frac{\partial u^k(\mathbf{x}+(\alpha+\tau)S_{ij}^{kh})}{\partial_{x_i^k}} \geq \frac{\partial u^k(\mathbf{x}+\alpha S_{ij}^{kh})}{\partial_{x_i^k}}. \tag{12}$$

Similarly, since the component (k,j) of S_{ij}^{kh} is $F(p_i,p_j,d^k,d^h)(p_i d^h) > 0$, we have

$$\frac{\partial u^k(\mathbf{x}+\alpha S_{ij}^{kh})}{\partial_{x_j^k}} \geq \frac{\partial u^k(\mathbf{x}+(\alpha+\tau)S_{ij}^{kh})}{\partial_{x_j^k}}. \tag{13}$$

Multiplying both sides of (12) and (13) by, respectively, $p_j d^h$ and $p_i d^h$, and connecting the resulting inequalities with (11) we have that

$$\frac{\partial u^k(\mathbf{x}+(\alpha+\tau)S_{ij}^{kh})}{\partial_{x_i^k}} p_j d^h > \frac{\partial u^k(\mathbf{x}+(\alpha+\tau)S_{ij}^{kh})}{\partial_{x_j^k}} p_i d^h,$$

which proofs that $g'(\alpha+\tau) < 0$. ■

Using Proposition 2 and the characterization of the space of integer solutions of (3), we are able to derive a closed expression of the Pareto frontier of the ERP, based on the behaviour of $u(\mathbf{x}+\alpha S_{ij}^{kh})$ (see Corollary 1 below), as it is shown in this example:

Example 1 *Consider the following ERP with initial endowments* [40, 188, 142, 66].

$$\max\{2-e^{-0.051x_1^1}-e^{-0.011x_2^1}, 2-e^{-0.1x_1^2}-e^{-0.031x_2^2}\}$$

subject to

$$\begin{array}{l} 5x_1^1+10x_2^1 = 2080 \\ 5x_1^2+10x_2^2 = 1370 \end{array}$$

$$\begin{array}{l} 5x_1^1+6x_1^2 = 1052 \\ 5x_2^1+6x_2^2 = 1336 \end{array} \qquad (14)$$

$$\begin{array}{l} 2-e^{-0.05x_1^1}-e^{-0.01x_2^1} \geq 1.68 \\ 2-e^{-0.1x_1^2}-e^{-0.031x_2^2} \geq 1.50 \end{array}$$

$$x_j^i \geq 0 \ \in Z \ i=1,2; \ j=1,2;$$

The utility functions $g^1(\alpha)=u^1(\mathbf{x}+\alpha S_{12}^{12})$ *and* $g^2(\alpha)=u^2(\mathbf{x}+\alpha S_{12}^{12})$ *are*

$$\begin{aligned} g^1(\alpha)=u^1(\mathbf{x}+\alpha S_{12}^{12}) &= u^1\left(\begin{bmatrix}40\\188\\142\\66\end{bmatrix}+\alpha\begin{bmatrix}12\\-6\\-10\\5\end{bmatrix}\right)= \\ &= 2-e^{-0.051(40+12\alpha)}-e^{-0.011(188-6\alpha)} \end{aligned}$$

$$\begin{aligned} g^2(\alpha)=u^2(\mathbf{x}+\alpha S_{12}^{12}) &= u^2\left(\begin{bmatrix}40\\188\\142\\66\end{bmatrix}+\alpha\begin{bmatrix}12\\-6\\-10\\5\end{bmatrix}\right)= \\ &= 2-e^{-0.1(142-10\alpha)}-e^{-0.031(66+5\alpha)}, \end{aligned}$$

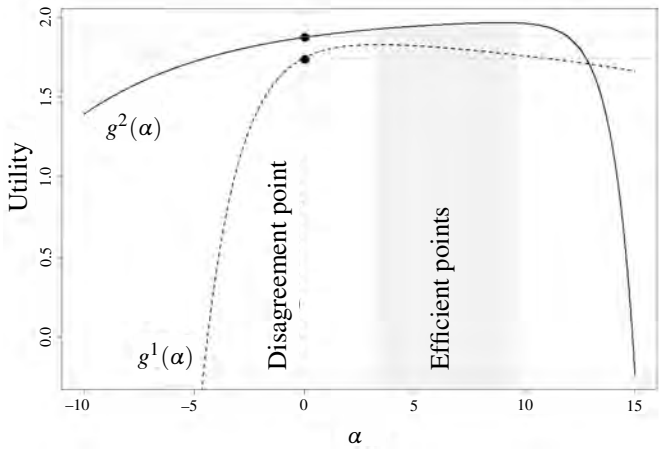


Figure 1: Plots of $g^1(\alpha)$ and $g^2(\alpha)$, and interval of α associated to the Pareto frontier. The disagreement point corresponds to $g^1(0)$ and $g^2(0)$, the utilities in the current iterate.

which are plotted in Fig. 1. The continuous optimal step lengths for the two respective agents are $\operatorname{argmax}\ g^1(\alpha) = 3.33$ *and* $\operatorname{argmax}\ g^2(\alpha) = 8.94$. *Due to the unimodality of* $u^k(\mathbf{x}+\alpha S_{ij}^{hk})$, *all efficient solutions of (14) are given by integer step lengths* $\alpha \in [3.33, 8.94]$ *(see Fig. 1), i.e., for* $\alpha \in \{4,5,6,7,8\}$ *we have*

$$g^1(4) = 1.82412 \quad g^1(5) = 1.81803 \quad g^1(6) = 1.80882 \quad g^1(7) = 1.79752 \quad g^1(8) = 1.78465,$$
$$g^2(4) = 1.93043 \quad g^2(5) = 1.94035 \quad g^2(6) = 1.94873 \quad g^2(7) = 1.95558 \quad g^2(8) = 1.96057.$$

Due to the unimodality of both utility functions with respect to α, *no efficient solution exists for an* α *outside the segment* $[3.33, 8.94]$.

The above example illustrates a case where the segment between $\operatorname{argmax}\ u^h(\mathbf{x}+\alpha S_{ij}^{kh})$ and $\operatorname{argmax}\ u^k(\mathbf{x}+S_{ij}^{kh})$ contains five integer points, associated with the feasible step lengths.

The following statements give a constructive characterization of the Pareto frontier of an ERP for the case of a concave utility function and linear utility functions respectively.

Corollary 1 *Let* Γ *be the set of integer points in the interval* $[a^{down}, a^{up}]$, *where* $a^{down} = \min\{\operatorname{argmax}_\alpha u^k(\mathbf{x}+\alpha S_{ij}^{kh}), \operatorname{argmax}_\alpha u^h(\mathbf{x}+\alpha S_{ij}^{kh})\}$ *and* $a^{up} = \max\{\operatorname{argmax}_\alpha u^k(\mathbf{x}+\alpha S_{ij}^{kh}), \operatorname{argmax}_\alpha u^h(\mathbf{x}+\alpha S_{ij}^{kh})\}$, *and let* $[\alpha^{down}, \alpha^{up}]$ *be the interval of feasible step lengths defined in (8). Then, due to Proposition 2, the set* $\mathcal{V}^*$ *of Pareto efficient solutions of an ERP can be obtained as follows:*

1. $\mathcal{V}^* = \{[u^h(\mathbf{x}+\alpha S_{ij}^{kh}), u^k(\mathbf{x}+\alpha S_{ij}^{kh})] : \alpha \in \Gamma\}$ *if* $\Gamma \subseteq [\alpha^{down}, \alpha^{up}]$ *is not empty and does not contain the zero.*

2. *If* Γ *is empty and there exists an integer point between* 0 *and* a^{down} *but no integer point between* a^{up} *and* α^{up} *then* $\mathcal{V}^*$ *contains the unique point given by* $[u^h(\mathbf{x}+\alpha S_{ij}^{kh},\ u^k(\mathbf{x}+\alpha S_{ij}^{kh})]$*such that* α *is the greatest integer between 0 and* a^{down}.
3. *If* Γ *is empty and there exists an integer point between* a^{up} *and* α^{up} *but no integer point between* 0 *and* a^{down} *then* $\mathcal{V}^*$ *contains either the unique point given by* $[u^h(\mathbf{x}+\alpha S_{ij}^{kh},\ u^k(\mathbf{x}+\alpha S_{ij}^{kh})]$ *such that* α *is the smallest integer between* a^{up} *and* α^{up}*, or* $\alpha = 0$*, or both of them if they do not dominate each other. (In this case the three possibilities must be checked, since if for only one of the utilities —let it be h, for instance—* $u^h(\mathbf{x}) > u^h(\mathbf{x}+\bar{\alpha}S_{ij}^{kh})$*,* $\bar{\alpha}$ *being the smallest integer between* a^{up} *and* α^{up}*, then both values* 0 *and* $\bar{\alpha}$ *are Pareto efficient.)*
4. *If* Γ *is empty and there are integer points both between* a^{up} *and* α^{up} *and between* 0 *and* a^{down} *then* V^* *contains the points given by* $[u^h(\mathbf{x}+\alpha S_{ij}^{kh},\ u^k(\mathbf{x}+\alpha S_{ij}^{kh})]$ *such that* α *is either the smallest integer between* a^{up} *and* α^{up}*, or the greatest integer between 0 and* a^{down}*, or both points if they do not dominate each other.*
5. *In the case that* Γ *contains the zero, then no point dominates the initial endowment* $\mathbf{x}$*, so that the only point in the Pareto frontier is* $\mathbf{x}$.

Corollary 2 *Consider the case of an economy where agents have linear utility functions with gradients* $\mathbf{c}^1,\ldots,\mathbf{c}^n$ *and let again* Γ *be the set of integer points in the interval* $[a^{down},a^{up}]$*, where* $a^{down} = \min\{\text{argmax}_\alpha \alpha \mathbf{c}^k S_{ij}^{kh},\ \text{argmax}_\alpha \alpha \mathbf{c}^h S_{ij}^{kh}\}$ *and* $a^{up} = \max\{\text{argmax}_\alpha \alpha \mathbf{c}^k S_{ij}^{kh},\ \text{argmax}_\alpha \alpha \mathbf{c}^h S_{ij}^{kh}\}$*, and let* $[\alpha^{down},\alpha^{up}]$ *be the interval of feasible step lengths defined in (8). It might be easily seen that either* $\Gamma = \mathbb{Q}$ *or* $\Gamma = \varnothing$*. The set* $\Gamma = \mathbb{Q}$ *in the case* $(c_i^h p_j d^k - c_j^h p_i d^k)$ *and* $(c_j^k p_i d^h - c_i^k p_j d^h)$ *have opposite signs, whereas* $\Gamma = \varnothing$ *if* $(c_i^h p_j d^k - c_j^h p_i d^k)$ *and* $(c_j^k p_i d^h - c_i^k p_j d^h)$ *have the same sign. Then, due to Proposition 2, the set* $\mathcal{V}^*$ *of Pareto efficient solutions of an ERP may contain at most one point:*

1. *if there is at least one non-null integer between* $-\max\{x_i^h/(p_j d^k),\ x_j^k/(p_i d^h)\}/F(p_i,p_j,d^k,d^h)$ *and* $\min\{x_j^h/(p_i d^k),\ x_i^k/(p_j d^h)\}/F(p_i,p_j,d^k,d^h)$ *and* $\Gamma = \varnothing$*, then* $\mathcal{V}^*$ *only contains the unique point corresponding to the allocation* $\mathbf{x}^{t+1} = \mathbf{x}^t + \alpha S_{ij}^{kh}$ *for a step-length* α *which is either equal to* $-\max\{x_i^h/(p_j d^k),\ x_j^k/(p_i d^h)\}/F(p_i,p_j,d^k,d^h)$ *(if* $(c_i^h p_j d^k$ - $c_j^h p_i d^k)$ *and* $(c_j^k p_i d^h$ - $c_i^k p_j d^h)$ *are negative) or equal to* $\min\{x_j^h/(p_i d^k),\ x_i^k/(p_j d^h)\}/F(p_i,p_j,d^k,d^h)$ *(if* $(c_i^h p_j d^k$ - $c_j^h p_i d^k)$ *and* $(c_j^k p_i d^h$ - $c_i^k p_j d^h)$ *are positive).*
2. $\mathcal{V}^*$ *only contains the disagreement point in the opposite case.*

Having a characterization of the Pareto frontier for any ERP in the sequence allows not just a higher efficiency in simulating the process but also the possibility of measuring the number of non dominated endowments of each of the $m(m-1)n(n-1)/4$ ERPs, which might be used as a measure of uncertainty of the process. Indeed, the uncertainty

of a barter process of this type might come from different sides: i) how to choose the couple of agents and commodities in each step? ii) which Pareto efficient solution of each ERP to use to update the endowments of the system? In the next subsection we shall study different criteria for answering these two questions.

Note that the set of non-dominated solutions of the ERP, obtained by the local search movement (6) might give rise to imbalances between supply and demand (Dreze, 1975). To resolve this imbalance Dreze introduced a quantity rationing, which can be also extended to the ERP.

Consider a rationing scheme for the ERP as a pair of vectors $l \in \mathbb{Z}^m$, $L \in \mathbb{Z}^m$, with $L \geq 0 \geq l$, such that the t^{th} and $(t+1)^{th}$ ER verifies $l_i \leq \mathbf{x}_i^{t+1} - \mathbf{x}_i^t \leq L_i$, for $i = 1, \ldots, n$, where l_i and L_i are the i^{th} components of l and L respectively. Thus, for two given agents h and k and two given commodities i and j we have

$$l_i \leq \alpha F(p_i, p_j, d^k, d^h) \begin{bmatrix} \vdots \\ p_j d^k \\ \vdots \\ -p_j d^h \\ \vdots \end{bmatrix} \leq L_i, \text{ and } l_j \leq \alpha F(p_i, p_j, d^k, d^h) \begin{bmatrix} \vdots \\ -p_i d^k \\ \vdots \\ p_i d^h \\ \vdots \end{bmatrix} \leq L_j.$$

An open problem, which is not investigated in this paper, is the formulation of equilibrium conditions for this rationing scheme. One possibility might be the construction of two intervals for l and L which minimize the overall imbalances, under the conditions that (3.1) is verified in each ERP, as long as l and L are inside the respected intervals. The integrality of the allocation space Λ forbids a straightforward application of the equilibrium criteria proposed by Dreze to the markets we are considering in this work.

3.2. Taking a unique direction of movement

The sequence of elementary reallocations formalized in (3) requires the iterative choice of couples of agents (h,k) and couples of commodities (i,j), i.e., directions of movement among the $m(m-1)n(n-1)/4$ in the neighbourhood of the current solution. If this choice is based on a welfare function (summarizing the utility functions of all the agents), the selection of couples of agents and couples of commodities can be made mainly in two different ways: first improving and best improving directions of movement.

Noting that each direction of movement in the current neighbourhood constitutes a particular ERP, a welfare criterion might be a norm of the objective vector (e.g., Euclidean, L_1 or L_∞ norms).

The best improving direction requires an exhaustive exploration of the neighbourhood, whereas the first improving direction stops the exploration of when an improving ERP has been found.

If at iteration t an improving direction exists the respective endowments are updated in accordance with the solution of the selected ERP: for each couple of commodities (i, j) and each couple of agents (h,k), agent k gives $\alpha F(p_i, p_j, d^k, d^h) p_j d^k$ units of i to agent h and in return he/she gets $\alpha F(p_i, p_j, d^k, d^h) p_i d^k$ units of j, for some $\alpha \in \mathbb{Z}$. At iteration $t+1$, a second couple of commodities and agents is considered in accordance with the defined criterion. If we use a first improving criterion, the process stops when the endowments keep in *status quo* continuously during $m(m-1)n(n-1)/4$ explorations, i.e., when no improving direction is found in the current neighbourhood.

3.3. Observing the paths of all improving directions of movement

When simulating social systems it might be interesting to enumerate all possible stories which are likely to be obtained starting from the known initial point. In this subsection we introduce a method to enumerate possible paths exclusively based on the Pareto efficiency of each elementary reallocation.

The idea is to solve $m(m-1)n(n-1)/4$ ERPs and keep all the efficient solutions generated. If in a given iteration we have r non-dominated solutions, and observe $l_i \leq m(m-1)n(n-1)/4$, for $i = 1, \ldots, r$, Pareto improving directions, with $f_{i,j}$ for $j = 1 \ldots l_i$ efficient solutions for each of them, we would expect some of the $r + \sum_{i=1}^{r} \sum_{j=1}^{l_i} f_{i,j}$ solutions to be non-dominated by some others and the incumbent should be updated by adding to the r previous solutions those which are non-dominated and removing those which are dominated by some other. From the point of view of a local search, the incumbent solution of this process is not a unique point in the allocation space but a collection of points which Pareto-dominate the initial endowment and do not dominate each other.

This procedure requires a method to find Pareto-optimal vectors each time $m(m-1)n(n-1)/4$ ERPs are solved. An efficient algorithms to find the set $\mathcal{V}^*$ of Pareto vectors among r given vectors $\mathcal{V} = \{v_1, v_2, \ldots, v_r\}$, where $v_i = (v_{i1}, v_{i2}, \ldots, v_{in}) \in \mathbb{R}^n$, $i = 1, 2, \ldots, r$, has been described (Corley and Moon, 1985; Sastry and Mohideen, 1999). In our implementation of the the best-improving barter process, we use the modified Corley and Moon algorithm, shown below.

Step 0. Set $\mathcal{V}^* = \emptyset$.

Step 1. Set $i = 1$, $j = 2$.

Step 2. If $i = r-1$, goto Step 6. For $k = 1, 2, \ldots, n$, if $v_{jk} \geq v_{ik}$ for some k, then go to Step 3; else, if $v_{ik} \geq v_{jk}$ for all k, then go to Step 4; otherwise, go to Step 5.

Step 3. Set $i = i+1$, $j = i+1$; go to Step 2.

Step 4. If $j = r$, set $\mathcal{V}^* = \mathcal{V}^* \cup \{v_i\}$ and $v_j = \{\infty, \infty, \ldots, \infty\}$ go to Step 3; otherwise, set $v_{jk} = v_{rk}$, where $k = 1, 2, \ldots, n$. Set $r = r - 1$ and go to Step 2.

Step 5. If $j = r$, set $\mathcal{V}^* = \mathcal{V}^* \cup \{v_i\}$ go to Step 3; otherwise, set $j = j + 1$ and go to Step 2.

Step 6. For $k = 1, 2, \ldots, n$, if $v_{jk} \geq v_{ik}$, then set $\mathcal{V}^* = \mathcal{V}^* \cup \{v_j\}$ and stop; else, if $v_{ik} \geq v_{jk}$, then set $\mathcal{V}^* = \mathcal{V}^* \cup \{v_i\}$ and stop. Otherwise, let $\mathcal{V}^* = \mathcal{V}^* \cup \{v_i, v_j\}$. Return $\mathcal{V}^*$.

The nice property of the modified Corley and Moon algorithm is that it doesn't necessarily compare each of the $r(r-1)/2$ couples of vectors for each of the n components. This is actually what the algorithm does in the worst case, so that the complexity could be written as $\mathcal{O}(nr^2)$, which is linear with respect of the dimension of the vectors and quadratic with respect to the number of vectors. For the case of linear utilities, the next subsection provides a small numerical example and the pseudo-code of the procedure used to enumerate the paths of all possible stories.

3.4. Linear utilities

In microeconomic theory the utility functions are rarely linear, however the case of linear objectives appears particularly suitable from an optimization point of view and allows a remarkable reduction of operations, as the ERPs cannot have more than one Pareto-efficient solution (see Corollary 1).

Consider a given direction of movement S_{ij}^{kh}. We know that a feasible step length α belongs to the interval defined by (8). Since in the case of one linear objective the gradient is constant, for any direction of movement (i, j, k, h) the best Pareto improvement (if there exists one) must happen in the endpoints of the feasible range of α (let $\alpha^{down}(i, j, k, h)$ and $\alpha^{up}(i, j, k, h)$ denote the left and right endpoints of the feasible range of α, when the direction of movement is (i, j, k, h)). Therefore, the line search reduces to decide either $\alpha^{down}(i, j, k, h)$, $\alpha^{up}(i, j, k, h)$ or none of them. Then for every given point $\mathbf{x}$, we have a neighbourhood of at most $m(m-1)n(n-1)/2$ candidate solutions. The pseudo-code to generate all sequences of elementary reallocations for n linear agents, keeping the Pareto-improvement in each interaction, is shown in Algorithm 1.

The function *CorleyMoon*() applies the modified Corley and Moon algorithm to a set of utility vectors and allocation vectors and returns the Pareto-efficient utility vectors with the associated allocations.

Despite the idea behind the SER of a process among self-interested agents, which are by definition local optimizers, this algorithm could also be applied to any integer linear programming problem of the form of (2) with one linear objective: $u(\mathbf{x}) = c^{\mathsf{T}}\mathbf{x}$. In this case however the branch and cut algorithm is much more efficient even for big instances, as we will show in the next section.

Algorithm 1 Generating paths of all improving directions of movement

Initialize the endowments $E = <\mathbf{e}^1,\ldots,\mathbf{e}^n>$ and utilities $U = <u^1,\ldots,u^n>$.
Initialize the incumbent allocations $\widetilde{E}^t = \{E\}$ and the incumbent utilities $\widetilde{U}^t = \{U\}$.
repeat
  for $\mathbf{x} \in \widetilde{E}^t$ **do**
    Let $< S_{\mathbf{x}}, G_{\mathbf{x}} >$ be the set of movements and utilities $\{(\mathbf{x}+\alpha S_{ij}^{kh},\, c^{\mathsf{T}}(\mathbf{x}+\alpha S_{ij}^{kh}))\}$ for each couple of commodities and agents (i,j,k,h) and $\alpha \in \{\alpha^{down}(i,j,k,h), \alpha^{up}(i,j,k,h)\}$
  end for
  Let $< S, G >= \bigcup_{x\in\widetilde{E}} < S_{\mathbf{x}}, U_{\mathbf{x}} >$
  Let $< S, G >= CorleyMoon(< S, G >)$
  Let $\widetilde{E}^{t+1} = \widetilde{E}^t \cup S$ and $\widetilde{U}^{t+1} = \widetilde{U}^t \cup G$
until $\widetilde{E}^t = \widetilde{E}^{t-1}$

If a first-improve method is applied, an order of commodities and agents is required when exploring the neighbourhood and the equilibrium allocation might be highly affected by this order (path-dependence). The pseudocode of algorithm 2 describes the first improve search of the barter algorithm applied to the case of one linear objective function.

Note that if the nonnegativity constraints are not taken into account, problem (2) is unbounded for linear utility functions. This corresponds to the fact that without lower bounds the linear version of this problem would make people infinitely get into debt. As a consequence, the only possible stopping criterion, when the objective function is linear, is the fulfillment of nonnegativity constraints, i.e. a given point $\mathbf{x}$ is a final endowment (an equilibrium of the barter process) if we have that for any direction of movement and for any given integer α if $c^{\mathsf{T}}(\mathbf{x}+\alpha S_{ij}^{kh}) > c^{\mathsf{T}}\mathbf{x}$ then $\mathbf{x}+\alpha S_{ij}^{kh}$ has some negative component. In some sense the optimality condition is now only based on feasibility.

Algorithm 2 First-improve SER with linear utility function

Initialize the endowments $E = <\mathbf{e}^1,\ldots,\mathbf{e}^n>$ and utilities $U = <u^1,\ldots,u^n>$.
Let $t = 0$;
Let (i,j,k,h) be the t^{th} direction in the order set of directions;
if $c^{\mathsf{T}}(x+\alpha^{down}(i,j,k,h)S_{ij}^{kh}) > c^{\mathsf{T}}(x+\alpha^{up}(i,j,k,h)S_{ij}^{kh})$ and $c^{\mathsf{T}}(x+\alpha^{down}(i,j,k,h)S_{ij}^{kh}) > c^{\mathsf{T}}(x)$ **then**
  Update the incumbent $x = x+\alpha^{down}(i,j,k,h)S_{ij}^{kh}$ and GOTO 3;
else if $c^{\mathsf{T}}(x+\alpha^{up}(i,j,k,h)S_{ij}^{kh}) > c^{\mathsf{T}}(x+\alpha^{down}(i,j,k,h)S_{ij}^{kh})$ and $c^{\mathsf{T}}(x+\alpha^{up}(i,j,k,h)S_{ij}^{kh}) > c^{\mathsf{T}}(x)$ **then**
  Update the incumbent $x = x+\alpha^{up}(i,j,k,h)S_{ij}^{kh})$ and GOTO 3;
else
  $t = t+1$;
  if $t < m(m-1)n(n-1)$ **then**
    GOTO 4;
  else
    RETURN
  end if
end if

3.5. The final allocation

For the case of a continuous commodity space and exogenous prices, pairwise optimality implies global optimality, as long as all agents are initially endowed with some positive amount of a commodity (Feldman, 1973). Unfortunately, the SER described in this paper does not necessarily lead to Pareto efficient endowments. Let $T_{\mathbf{x}}(\alpha) = \mathbf{x} + \sum_{k \neq h} \sum_{i \neq j} \alpha(i,j,k,h) S_{ij}^{kh}$, representing a simultaneous reallocation of m commodities among n agents, with step length α_{ij}^{kh} for each couple of commodities ij and agents hk, starting from $\mathbf{x} \in \Lambda$. Whereas a SER is required to keep feasibility along the process, a simultaneous reallocation $T_{\mathbf{x}}(\alpha)$ of m commodities among n agents does not consider the particular path and any feasibility condition on the paths leading from $\mathbf{x}$ to $T_{\mathbf{x}}(\alpha)$. Hence, remembering that all SERs described in this section stop when no improving elementary reallocation exists in the current neighbourhood, we can conclude that the non-existence of a feasible improving ER does not entail the non-existence of an improving simultaneous reallocation of m commodities among n agents. In this sense a SER provides a lower bound of any sequence of reallocations of more than two commodities and two agents at a time.

4. Bartering on networks

An important extension of the problem of bargaining integer amounts of m commodities among n agents with fixed prices is to define a network structure such that trades among agents are allowed only for some couples of agents who are linked in this network. In this case the conservation of commodities $d^1\mathbf{x}^1 + d^2\mathbf{x}^2 + \cdots + d^n\mathbf{x}^n = d^1\mathbf{e}^1 + d^2\mathbf{e}^2 + \cdots + d^n\mathbf{e}^n$ is replaced by balance equations on a network, so that the final allocation of commodity i must verify $A\mathbf{y}_i = D(\mathbf{x}_i - \mathbf{e}_i)$, where $\mathbf{y}_i$ is the flow of commodity i in the system, A is the incidence matrix, and D is a $n \times n$ diagonal matrix containing the weights of the conservation of commodity i, that is $D = \mathrm{diag}(d^1 \ldots d^n)$ (for more details on network flows problems see Ahuja, Magnanti and Orlin (1991)).

It is also possible for the final allocation to have a given maximum capacity, that is, an upper bound of the amount of commodity i that agent h may hold: $x_i^h \leq \bar{x}_i^h$.

The variables of the problem are now x_i^h, which again represent the amount of commodity i held by agent h, s_i^h which are the slack variables for the upper bounds, and $y_i^{h,k}$ which denote the flow of commodity i from agent h to agent k.

The objective functions $\tilde{u}^i(\mathbf{x},\mathbf{y}), i = 1 \ldots n$, might depend on both the final allocation $\mathbf{x}$ and the interactions $\mathbf{y}$, since the network topology could represent a structure of geographical proximity and reachability.

The resulting mathematical programming formulation of the problem of bargaining integer commodities with fixed prices among agents on a network with upper bounds on the final allocations is as follows:

$$\max\{\tilde{u}^i(\mathbf{x},\mathbf{y}), i=1,\ldots,n\} \tag{15a}$$

subject to

$$\left[\begin{array}{ccc|ccc|ccc}
P & & & & & & & & \\
& \ddots & & & & & & & \\
& & P & & & & & & \\
\hline
I & & & I & & & & & \\
& \ddots & & & \ddots & & & & \\
& & I & & & I & & & \\
\hline
 & & & & & & A & & \\
 & \mathbb{D} & & & & & & \ddots & \\
 & & & & & & & & A
\end{array}\right]
\left[\begin{array}{c}\mathbf{x}\\ \mathbf{s}\\ \mathbf{y}\end{array}\right] =
\left[\begin{array}{c} b^1 \\ \vdots \\ b^n \\ \bar{x}^1 \\ \vdots \\ \bar{x}^n \\ \mathbf{b}^0 \end{array}\right] \tag{15b}$$

$$\begin{aligned} & u^i(\mathbf{x},\mathbf{y}) \geq u^i(\mathbf{e},\mathbf{0}) \quad i=1\ldots,n \\ & \mathbf{x}\in\mathbb{Z}^{mn}\geq 0, \quad \mathbf{y}\in\mathbb{Z}^{mn(n-1)}\geq 0, \end{aligned} \tag{15c}$$

where $\tilde{u}^i : \mathbb{R}^{mn} \to \mathbb{R}$, $P \in \mathbb{Q}^{1\times m}$, $\mathbb{D} \in \mathbb{Q}^{mn\times mn}$, $b^i \in \mathbb{Q}$, $i=1,\ldots,n$, $A \in \mathbb{Q}^{n\times n(n-1)}$, and $\mathbf{b}^0 \in \mathbb{Q}^{nm}$. Matrix $\mathbb{D}$ is an appropriate permutation of the diagonal matrix made of m copies of the matrix D with the weights of the conservation of commodity and $\tilde{u}^i(\mathbf{e},\mathbf{0})$ is the utility function of agent i evaluated in the initial endowments $\mathbf{e}$ with null flow.

Problem (2) had mn variables and $m+n$ constraints, whereas problem (15) has $mn(n+1)$ variables and $n(1+2m)$ constraints. When a SER is applied, the definition of a network structure and the application of upper bounds to the final allocation reduce the number of feasible directions of movement in each iteration and the bound of the interval of feasible step length, as for any incumbent allocation $\mathbf{x}$, the step length α must be such that $0 \leq \mathbf{x} + \alpha S_{ij}^{kh} \leq \widetilde{\mathbf{x}}$.

An application of this problem is the transfer of workers among plants of the same franchising company or chain store. When a change of demand requires a reorganization of the production, laying workers off and contracting new workers might be costly both for the company (severance pays and taxes) and for the workers (finding a new job and experiencing a possible period of unemployment). Suppose that each plant is independent and led by a different director, whose interest is to maximize the utility of his/her particular plant and suppose the price per hour is fixed by law or the collective labour agreement for each category of worker. In this case prices are exogenous and each plant is interested in maximizing its benefit separately. The objective functions $\tilde{u}^i(\mathbf{x},\mathbf{y}), i=1\ldots n$, might depend on both the final allocation $\mathbf{x}$ and the interactions $\mathbf{y}$,

since the network structure could represent a structure of geographic proximity and plants could wish to minimize the distance of displacement of their workers. The upper bounds on the final endowments might be used to model the maximum capacity that each plant has to accommodate and to employ a given type of worker. Also in this particular application, the bargaining nature of the problem lays on the assumption that the commodities we are considering are private goods, as the labour of one worker is excludable and rivalrous.

The formulation (15) also allows the definition of arbitrary network structures, whose topology is given in A. However, despite the absence of any prior definition of A, any sequence of bilateral trades intrinsically gives rise to a network structure generated by the set of couples of agents interacting along the process. Such a structure might be statistically analysed in term of its topological properties, as it is done in the next section with a battery of problems of different sizes. We shall study the assortativity of networks generated by the set of couples of agents interacting along the SER. The assortativity is the preference for an agent to interact with others that are similar or different in some way, it is often operationalized as a correlation between adjacent node's properties. Two kinds of assortativities emerges in the best-improve barter algorithm: 1) couples of agents with highly different marginal utilities are more often commercial partners, 2) and also agents who are more sociable (trade more often) interact frequently with agents who are not sociable. These results suggest that when the interactions are restricted to be performed only among adjacent agents on a network, highly dissortative structure would allow better performance of the process.

The effect of network structures on the performance of a barter process has been already studied (Bell, 1998; Wilhite, 2001), for the case of endogenous prices and continuous commodity space. In this case the process takes into account how agents update prices each time they perform a bilateral trade. Reasonably, prices should be updated based either on the current state of the only two interacting agents or on the state of the overall population or also on the history of the system, such as previous prices. Bell showed that centralized network structures, such as a stars, exhibit a faster convergence to an equilibrium allocation.

5. Computational results

We have already seen that a SER can also be applied to any integer linear programming problem of the form (2), where the individual utilities are aggregated in a single welfare function. If this aggregated welfare is defined as a linear function of the endowments of the form $u(\mathbf{x}) = c^{\mathsf{T}}\mathbf{x}$, the comparison of the SERs with the standard branch and cut algorithm is easily carried out. Considering the ERP as the basic operation of a SER and the simplex iteration as the basic operation of the branch and cut algorithm, the comparison between the two methods is numerically shown in Table 1 for three replications of 11 problems with the same number of agents and commodities, which

Table 1: *Numerical results of the SER and Branch and Cut for different instances of problem (2). The first column shows the number of agents and commodities of the problem. Columns 'ERPs' provide the number of elementary reallocations and column 'neighbourhood' shows the proportion of neighbourhood which has been explored. Columns 'solution' give the maximum total utility found. Column 'simplex' gives the number of simplex iterations performed by branch and cut.*

size	initial welfare	first-improve			best-improve		branch and cut	
		neighbourhood	ERPs	solution	ERPs	solution	simplex	solution
10	75.134	0.66	267	353.269	91	365.126	87	394.630
10	147.958	0.84	271	763.188	91	767.371	12	769.861
10	1.205.972	0.77	375	3.925.921	74	3.844.165	70	4.060.685
15	297.713	0.70	1.343	1.455.839	215	1.471.387	49	1.488.149
15	326.996	0.71	1.090	2.544.271	237	2.554.755	63	2.614.435
15	625.800	0.71	806	2.640.317	224	2.644.008	76	2.684.016
20	183.573	0.67	2.759	3.432.832	378	3.425.665	110	3.525.421
20	1.064.023	0.81	1.582	4.197.757	361	4.194.187	94	4.331.940
20	201.377	0.78	2.629	1.017.906	351	1.089.860	80	1.180.977
25	228.365	0.89	4.358	2.221.790	648	2.226.152	237	2.271.552
25	687.492	0.65	2.806	3.416.982	572	3.403.937	113	3.462.043
25	323.495	0.61	4.706	2.262.657	666	2.245.817	50	2.474.429
30	973.955	0.79	6.648	5.428.473	975	5.427.207	101	5.377.843
30	1.811.905	0.82	13.126	8.945.605	1.084	8.953.611	127	9.080.651
30	1.302.404	0.85	12.089	7.583.841	957	7.573.400	132	7.605.525
35	653.739	0.87	13.201	3.456.918	1.310	3.458.570	112	3.474.126
35	564.905	0.80	8.772	3.579.713	1.308	3.585.815	77	3.599.639
35	753.056	0.83	14.199	5.132.226	1.290	5.107.933	67	5.333.123
40	482.570	0.87	16.307	2.429.707	1.608	2.428.731	145	2.446.953
40	430.174	0.68	7.885	5.281.060	1.640	5.229.740	90	5.279.631
40	2.795.862	0.79	14.240	19.175.278	1.578	14.503.963	186	19.276.444
45	3.392.010	0.98	62.398	22.681.229	2.300	22.664.443	162	22.728.195
45	842.645	0.92	12.900	6.606.875	2.137	6.642.397	204	6.755.016
45	1.909.859	0.97	48.688	15.979.841	2.173	15.865.744	180	16.071.407
50	839.559	0.93	20.615	4.822.082	2.105	4.859.830	137	4.895.655
50	718.282	0.97	20.744	3.586.560	2.459	3.588.633	160	3.610.194
50	1.570.652	0.99	58.165	18.872.864	2.530	19.018.519	180	19.069.868
55	351.051	0.98	20.344	2.761.203	2.935	2.748.862	1.242	2.799.187
55	413.656	0.96	26.780	4.566.394	2.922	4.569.975	336	4.585.475
55	551.355	0.99	32.053	5.136.295	3.139	5.135.647	253	5.157.444
60	468.575	0.99	27.208	1.941.409	3.568	1.949.786	271	1.995.930
60	501.366	0.99	34.323	5.051.429	3.521	5.051.836	313	5.067.154
60	575.950	0.98	43.227	4.751.072	3.589	4.747.097	273	4.801.179

amounts to 33 instances. The branch and cut implementation of the state-of-the-art optimization solver Cplex was used.

In the special case of a unique linear utility function a system of many local optimizers (agents) could be highly inefficient if compared with a global optimizer, who acts for the "goodness" of the system, as in the case of branch and cut. Also the increase of elementary operations of the barter process is much higher than the one of

the branch and cut, particularly when the direction of movement is selected in a best-improve way, as it is shown in Table 1. Each point is averaged over the three instances for each size (m,n). The economical interpretation suggests that if the time taken to reach an equilibrium is too long, it is possible that this equilibrium is eventually never achieved since in the meanwhile many perturbing events might happen.

5.1. Application in computational economics

From the point of view of computational economics, sequences of k-lateral trades of fractional amounts of commodities with local Walrasian prices have also been studied (Axtell, 2005), along with the convergence rate to the equilibrium. Some studies have also taken into account the performance of the process under a variety of network structures restricting the interactions to be performed only among adjacent agents (Bell, 1998; Wilhite, 2001). Populations of Cobb-Douglas' agents trading continuous amount of two commodities with local Walrasian prices have been considered, with the aim of analysing the speed of convergence to an equilibrium price and allocation: more centralized networks converged with fewer trades and had less residual price variation than more diverse networks.

An important question when sequences of elementary reallocations in markets with fixed prices are studied is to find factors which affect the number of non dominated allocations related to improving paths of algorithm 1 and the number of neighbourhoods explored. We consider a theoretical case where 2 agents with linear utility functions have to trade 9 commodities. The following three factors are taken into account:

- $Fact_1$: the variability of prices;
- $Fact_2$: association between the initial endowment and the marginal utility of the same agent;
- $Fact_3$: association between the initial endowment and the marginal utility of the other agent.

The aforementioned factors are measured at three levels and four randomized replicates have been simulated for each combination of factors. A multivariate analysis of variance (MANOVA) is performed, considering the two following response variables

- $Resp_1$: the number of non dominated allocations related to improving paths of algorithm 1;
- $Resp_2$: the number of neighbourhoods explored.

The MANOVA table in Table 2 illustrates the effects and the significance of two factors to the bivariate response: $Fact_1$ and $Fact_3$. The interaction between $Fact_2$ and $Fact_3$ is significant, suggesting a higher increase in the response variables when they

Table 2: *MANOVA analysis of the paths of all improving directions*

	Pillai	F	p-value
$Fact_1$	**0.135084**	**2.9336**	**0.022435**
$Fact_2$	0.050411	1.0472	0.384650
$Fact_3$	**0.162063**	**3.5712**	**0.008055**
$Fact_1 \times Fact_2$	0.057542	0.5999	0.777027
$Fact_1 \times Fact_3$	0.063816	0.6674	0.719612
$Fact_2 \times Fact_3$	**0.195534**	**2.1943**	**0.030408**
$Fact_1 \times Fact_2 \times Fact_3$	**0.291627**	**1.7284**	**0.046013**

are both low. The correlation between the amounts of the initial endowments and the coefficients of the objective function of the same agent does not appear by itself to have a significant effect on the response variables.

The results of the MANOVA should be interpreted in accordance with the analysis of the assortativity behaviour of the economical interaction network. Any SER intrinsically gives rise to a network structure generated by the set of couples of agents interacting along the process. Such a structure might be statistically analysed in term of its topological properties. We consider two kind of assortativity measure (the preference for an agent to interact with others that are similar or different in some way, often operationalized as a correlation between adjacent node's properties):

- $Type_1$: couples of agents with highly different marginal utilities are more often commercial partners – Pearson correlation between the Euclidean distance of marginal utilities and the number of interactions of each couple of agents, $cor(dist(c^h, c^k), interactions^{(h,k)})$;
- $Type_2$: agents who are more sociable (trade more often) interact frequently with agents who are not sociable –Pearson correlation between the Euclidean distance of couples of agents with respect to their number of interactions and the number of joint interactions of each couple, $cor(dist(degree^h, degree^k), degree^{(h,k)})$;
- $Type_3$: the more two agents are different with respect to their marginal utilities, the more they are different with respect to their number of interactions –Pearson correlation between the Euclidean distance of marginal utilities and the Euclidean distance of the number of interactions of each couple of agents, $cor(dist(c^h, c^k), dist(degree^h, degree^k))$.

The numerical values in Table 3 corresponds to the aforementioned assortativities, associated to the same instances of Table 1.

The significant effect of $Fact_3$ (the association between the initial endowment and the marginal utility of the other agent) in the MANOVA of Table 2 seems coherent with the $Type_1$ and $Type_3$ assortativity reported in Table 3, in the vague sense that assortativity between nodes relates with the number of interactions.

Table 3: *Three types of network assortativity.*

Size	First-improve			Best-improve		
	$Type_1$	$Type_2$	$Type_3$	$Type_1$	$Type_2$	$Type_3$
10	0.40	0.48	0.63	0.70	0.67	0.74
10	0.46	0.66	0.61	0.85	0.63	0.74
10	0.60	0.48	0.75	0.71	0.70	0.75
15	0.49	0.28	0.64	0.74	0.48	0.56
15	0.44	0.52	0.59	0.58	0.44	0.67
15	0.37	0.60	0.54	0.56	0.74	0.66
20	0.22	0.36	0.49	0.39	0.62	0.54
20	0.33	0.53	0.48	0.54	0.48	0.55
20	0.053	0.30	0.47	0.48	0.45	0.42
25	0.37	0.53	0.56	0.55	0.66	0.53
25	0.33	0.63	0.40	0.65	0.56	0.66
25	0.32	0.43	0.56	0.48	0.70	0.49
30	0.10	0.28	0.33	0.42	0.55	0.53
30	0.07	0.30	0.39	0.56	0.62	0.68
30	0.33	0.42	0.59	0.61	0.63	0.65
35	0.27	0.41	0.43	0.44	0.59	0.43
35	0.26	0.33	0.56	0.46	0.55	0.48
35	0.13	0.40	0.47	0.46	0.58	0.53
40	0.20	0.37	0.36	0.44	0.64	0.38
40	0.48	0.48	0.52	0.68	0.52	0.64
40	0.40	0.44	0.59	0.64	0.64	0.60
45	0.10	0.13	0.45	0.62	0.60	0.54
45	0.29	0.45	0.51	0.57	0.59	0.58
45	0.20	0.23	0.52	0.58	0.57	0.68
50	0.17	0.26	0.37	0.35	0.55	0.32
50	0.21	0.28	0.50	0.45	0.62	0.42
50	0.15	0.30	0.42	0.51	0.50	0.65
55	0.14	0.53	0.17	0.39	0.52	0.38
55	0.17	0.33	0.38	0.29	0.53	0.44
55	0.19	0.37	0.38	0.47	0.56	0.43
60	0.35	0.45	0.60	0.54	0.57	0.62
60	0.20	0.30	0.43	0.34	0.50	0.52
60	0.16	0.38	0.29	0.39	0.51	0.48

What clearly emerges from these results is an interaction pattern which is far from random. In the case the SER is forced to be performed only among agents adjacent in a network, it suggests that highly dissortative structure match pretty well with the best-improve directions of movement, so that no improving direction is penalized by the presence of a predefined network structure.

6. Summary and future directions

We studied the use of barter processes for solving problems of bargaining on a discrete set, representing markets with indivisible goods and fixed exogenous prices. We showed that the allocation space is characterized by a block diagonal system of linear constraints, whose structural properties might be exploited in the construction and analysis of barter processes. Using Proposition 2 and the characterization of the space of integer solutions of the ERP, we were able to derive a constructive procedure to approximate its Pareto frontier, as shown by Corollary 1 and Corollary 2.

Further research on this topic should include the characterization of the integer points in the null space of a general reallocation problem with fixed prices to obtain a closed form solution of a general problem of reallocating integer amounts of m commodities among n agents with fixed prices.

An open problem, which has not been investigated in this paper, is the formulation of equilibrium conditions for this rationing scheme proposed in Section 3, as suggested by Dreze (1975) for the case of continuous allocation space.

In Section 4 we proposed a mathematical programming model for the problem of reallocating integer amounts of m commodities among n agents with fixed prices on a sparse network structure with nodal capacities. Further research on this issue should include the study of mathematical properties of a SER in dealing with markets with sparsely connected agents, as formulated in (15).

References

Ahuja, R.K., Magnanti, T.L. and Orlin, J.B. (1991). *Network Flows: Theory, Algorithms, and Applications*. (1st ed.). Englewood Cliffs, Prentice-Hall.

Arrow, K. J. and Debreu, G. (1983). Existence of an equilibrium for a competitive economy. *Econometrica*, 22, 265–290.

Auman R. and Dreze, J. (1986). Values of Markets with Satiation or fixed prices. *Econometrica*, 54, 1271–1318.

Axtell, R. (2005). The complexity of exchange. *In Working Notes: Artificial Societies and Computational Markets. Autonomous Agents 98 Workshop, Minneapolis/St. Paul (May)*.

Bell, A.M. (1998). Bilateral trading on network: a simulation study. *In Working Notes: Artificial Societies and Computational Markets. Autonomous Agents 98 Workshop, Minneapolis/St. Paul (May)*, 1998.

Caplin A. and Leahy, J. (2010). A graph theoretic approach to markets for indivisible goods. *Mimeo, New York University*, NBER Working Paper 16284.

Corley, H.W. and Moon, I.D. (1985). Shortest path in network with vector weights. *Journal of Optimization Theory and Applications*, 46, 79–86.

Danilov, V., Koshevoy, G. and Murota, K. (2001). Discrete convexity and equilibria in economies with indivisible goods and money. *Journal of Mathematical Social Sciences*, 41, 3, 251–273.

Dreze, J. (1975). Existence of an exchange equilibrium under price rigidities. *International Economic Review*, 16, 2, 301–320.

Edgeworth, F.Y. (1932). Mathematical psychics, an essay on the application of mathematics to the moral sciences, (3th ed.). *L.S.E. Series of Reprints of Scarce Tracts in Economics and Political Sciences*.

Feldman, A. (1973). Bilateral trading processes, pairwise optimality, and Pareto optimality. *Review of Economic Studies*, XL(4) 463–473.

Haimes, Y.Y., Lasdon, L.S. and Wismer, D.A. (1971). On a bicriterion formulation of the problems of integrated system identification and system optimization. *IEEE Transactions on Systems, Man, and Cybernetics*, 1(3), 296–297.

Jevons, W.S. (1888). *The Theory of Political Economy*, (3rd ed.). London, Macmillan.

Kaneko, M. (1982). The central assignment game and the assignment markets. *Journal of Mathematical Economics*, 10, 205–232.

Nash, J.F. (1951). The bargaining problem. *Econometrica* 18, 155–162.

Ozlen, M. and Azizoglu, M. (2009). Multi-objective integer programming: a general approach for generating all non-dominating solutions. *European Journal of Operational Research*, 199(1), 25–35.

Ozlen, M., Azizoglu, M. and Burton, B.A. (Accepted 2012). Optimising a nonlinear utility function inmulti-objective integer programming. *Journal of Global Optimization*, to appear.

Quinzii, M. (1984). Core and competitive equilibria with indivisibilities. *International Journal of Game Theory*, 13, 41–60.

Rubinstein, A. (1983). Perfect equilibrium in a bargaining model. *Econometrica*, 50, 97–109.

Sastry, V.N. and Mohideen, S.I. (1999). Modified algorithm to compute Pareto-optimal vectors. *Journal of Optimization Theory and Applications*, 103, 241–244.

Scarf, H. (1994). The allocation of resources in the presence of indivisibilitie. *Journal of Economic Perspectives*, 8, 111–128.

Shapley, L. and Shubik, M. (1972). The assignment game I: the core. *International Journal of Game Theory*, 1, 111–130.

Uzawa, H. (1962). On the stability of edgeworth's barter process. *International Economic Review*, 3(2), 218–232.

Vazirani, V.V., Nisan, N., Roughgarden, T. and Tardos, E. (2007). *Algorithmic Game Theory*, (1st ed.). Cambridge, Cambridge University Press.

Wilhite, A. (2001). Bilateral trade and small-worl-networks. *Computational Economics*, 18, 49–44.

Wooldridge, M. (2002). *An Introduction to MultiAgent Systems*, (1st ed.). Chichester, UK, John Wiley and Sons Ltd.

SORT 39 (1) January-June 2015, 109-126

A comparison of computational approaches for maximum likelihood estimation of the Dirichlet parameters on high-dimensional data

Marco Giordan and Ron Wehrens[1]

Abstract

Likelihood estimates of the Dirichlet distribution parameters can be obtained only through numerical algorithms. Such algorithms can provide estimates outside the correct range for the parameters and/or can require a large amount of iterations to reach convergence. These problems can be aggravated if good starting values are not provided. In this paper we discuss several approaches that can partially avoid these problems providing a good trade off between efficiency and stability. The performances of these approaches are compared on high-dimensional real and simulated data.

MSC: 62F10, 65C60

Keywords: Levenberg-Marquardt algorithm, re-parametrization, starting values, metabolomics data.

1. Introduction

The Dirichlet distribution has multiple applications. It is well known for being the conjugate prior of the multinomial distribution and can be therefore used to get Bayesian estimates of the multinomial parameters. It is the basis for complicated models such as Dirichlet Processes and mixture distributions based on the Dirichlet distribution. In addition, it is interesting in its own right. It can be used to analyse positive continuous data that sum up to one, i.e. compositional data. Such kinds of data can arise in many situations. For example, when the data in each unit are represented by an intensity signal it can be of interest to normalize them through the total intensity of that unit. In this way

[1] Biostatistics and Data Management. IASMA Research and Innovation Centre. marco.giordan@fmach.it, ron.wehrens@wur.nl
Received: April 2014
Accepted: February 2015

the data from different units are comparable and the final data can be analysed using the Dirichlet distribution. Another possible application is in the analysis of taxonomic assignements. For each unit the percentages of the microbial composition of the unit are assigned to the specific taxonomies. These data can be easily produced with Next Generation Sequencing (NGS) technologies. Many other examples of the use of the Dirichlet distribution are provided in the paper of Wicker et al. (2008).

The use of the Dirichlet distribution has been criticized due to the strong independence properties associated to this distribution (Aitchison, 1986). In the literature some generalizations have been proposed to overcome these limits, see for example Connor and Mosiman (1969) and Rayens and Srinivasan (1994). In this paper we only marginally discuss this point giving a case where the application of the Dirichlet distribution to high-dimensional compositional data leads to reliable conclusions. The focus of paper is related to the comparison of the computational performances of different methods to get maximum likelihood estimates of the Dirichlet distribution. There is no closed form solution of the maximum likelihood equations, therefore numerical methods must be employed. At the moment, commonly adopted methods are rather unstable. Final estimates can be outside the correct range for the parameters and the algorithms can fail to reach convergence in a reasonable amount of time. Wicker et al. (2008) reported many convergence failures in their simulation studies. Strategies to improve stability have been studied by many authors. Many proposals are focused on the choice of good starting values for the optimization algorithms. Useful references for these problems and the study of Dirichlet maximum likelihood estimation are the papers of Dishon and Weiss (1980), Ronning (1989), Narayanan (1991a), and Narayanan (1991b).

In this work we compare eight different algorithms and four different initialisation methods on real and simulated data. As an application, we consider the analysis of metabolomics data. We consider the Newton-Raphson algorithm as the reference algorithm. A fixed-point algorithm, shown in literature to have very good performance, is taken into account as well. Moreover, a novel and more stable algorithm based on Levenberg-Marquardt ideas (Levenberg, 1944; Marquardt, 1963) will be employed to get the final maximum likelihood estimates. In the appendix we give a proof of the convergence to the optimum for this algorithm. Finally, to avoid the problem of estimates outside the admissible parameter space, a simple re-parametrization of the Dirichlet parameters and an algorithm with box constraints are considered. The re-parametrization will be used together with the Newton-Raphson algorithm and the Levenberg-Marquardt algorithm, but not with the FPI algorithm because this does not suffer from the problem of a constrained parameter space (see Huang, 2005). The re-parametrization and the algorithm with box-constraints are straightforward but have not been considered in the literature before.

2. Dirichlet likelihood

In this section we introduce notation and we summarize useful results from literature (see Minka, 2000). If $\mathbf{y} = (y_1, \ldots, y_K)^\mathsf{T}$ is Dirichlet distributed with vector parameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^\mathsf{T}$ then its density is

$$\frac{\Gamma(\sum_k \alpha_k)}{\Pi_k \Gamma(\alpha_k)} \Pi_k y_k^{\alpha_k - 1}$$

where $\alpha_k > 0$, $y_k > 0$, $k = 1, \ldots, K$ and $\sum_k y_k = 1$.

The log-likelihood for N independent observations can be written as

$$f(\boldsymbol{\alpha}) = N \ln \Gamma \left(\sum_k \alpha_k \right) - N \sum_k \ln \Gamma (\alpha_k) + N \sum_k (\alpha_k - 1) \frac{1}{N} \sum_i \ln y_{ik}. \tag{1}$$

The gradient of the log-likelihood with respect to one α_k is:

$$[\nabla f(\boldsymbol{\alpha})]_k = N \left(\Psi(\sum_k \alpha_k) - \Psi(\alpha_k) + \frac{1}{N} \sum_i \ln y_{ik} \right) \tag{2}$$

where Ψ denotes the digamma function. In what follows the arguments of a function (e.g. the parameters $\boldsymbol{\alpha}$ for the function $f(\cdot)$ in equation (2)) can be suppressed in the notation when this does not generate confusion. The Hessian can be written in matrix form as

$$\mathbf{H} = \mathbf{Q} + \mathbf{1}\mathbf{1}^\mathsf{T} z \tag{3}$$

$$q_{jk} = -N \Psi'(\alpha_k) \delta(j - k) \tag{4}$$

$$z = N \Psi'(\sum_k \alpha_k) \tag{5}$$

where Ψ' denotes the trigamma function and δ is the Dirac function (zero on the real line, except at the origin where it is one). Let us note that the diagonal form of $\mathbf{Q}$ assures the existence of its inverse when the diagonal elements are different from zero. This is exactly the present case because the trigamma function is positive for positive real arguments.

2.1. Some preliminary considerations

The number of available algorithms to maximize a function is huge and its impossible to summarize all of them in a meaningful way. In this work we have focused our

attention on algorithms suggested by the literature about the Dirichlet distribution and on a modification of the Levenberg-Marquardt algorithm for which we are able to provide a theoretical result related to the properties of the Dirichlet distribution.

In the literature the algorithms studied and suggested are mainly two: the Newton-Raphson algorithm and the Fixed Point Iteration algorithm (see Minka, 2000; Huang, 2005). We include them in our comparison and we describe them briefly in the following sections. Other algorithms, like BFGS (Ronning, 1989) or Gradient Ascent (Huang, 2005), had a poor performance in the literature and are therefore disregarded here. The Levenberg-Marquardt algorithm is an algorithm to find least-squares estimates. In the appendix we give a theoretical result of convergence for a modification of the Levenberg-Marquardt algorithm with a fixed damping parameter when we apply its adaptation to find maximum likelihood estimates. We study its performance in the paper through simulations and real data. Finally, to avoid the problem of estimates outside the allowed space, we consider a re-parametrization of the Dirichlet distribution and an implementation of the BFGS algorithm with bounding box constraints, L-BFGS-B (see Byrd et al., 1995).

In this work the efficiency of the algorithms will be compared essentially by the number of iterations required to reach convergence. This number does not depend on the implementation of the code and in this sense it is objective. We will see that the Levenberg-Marquardt approach can be thought as a penalized version of the Newton-Raphson algorithm and therefore it is expected to be slightly slower. However in general the *iteration time* for different algorithms can vary substantially and the algorithm with fewest iterations for convergence can require the largest amount of time. This can be the case when we want to compare Levenberg-Marquard, FPI and L-BFGS-B, therefore for these algorithms we provide also a comparison on time.

2.2. Newton-Raphson algorithm

The Newton-Raphson (NR) algorithm is used to solve the maximum likelihood equations $[\nabla f(\boldsymbol{\alpha})] = \mathbf{0}$. It can be summarized by the following equations:

$$\boldsymbol{\alpha}^{\text{new}} = \boldsymbol{\alpha}^{\text{old}} - \mathbf{H}^{-1}\nabla f(\boldsymbol{\alpha}^{\text{old}}) \tag{6}$$

$$\mathbf{H}^{-1} = \mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1}\mathbf{1}\mathbf{1}^{\mathsf{T}}\mathbf{Q}^{-1}}{\frac{1}{z} + \mathbf{1}^{\mathsf{T}}\mathbf{Q}^{-1}\mathbf{1}} \tag{7}$$

$$\left[\mathbf{H}^{-1}\nabla f(\boldsymbol{\alpha})\right]_k = \frac{[\nabla f(\boldsymbol{\alpha})]_k - b}{q_{kk}} \tag{8}$$

$$b = \frac{\mathbf{1}^{\mathsf{T}}\mathbf{Q}^{-1}\nabla f(\boldsymbol{\alpha})}{\frac{1}{z} + \mathbf{1}^{\mathsf{T}}\mathbf{Q}^{-1}\mathbf{1}} = \frac{\sum_j [\nabla f(\boldsymbol{\alpha})]_j / q_{jj}}{1/z + \sum_j 1/q_{jj}}. \tag{9}$$

When **Q** is invertible the inversion of the matrix **H** is always guaranteed by the use of the Sherman-Morrison formula. However the equations (8) and (9) highlight that the algorithm does not require the storage and the computation of the inverse of **H**. This is a great advantage, especially when the number of variables is high.

The Newton-Raphson algorithm is expected to converge to the global optimum because the Dirichlet distribution belongs to the exponential family and is therefore concave. However, the convergence can be very slow and the final estimates can be outside the admissible range for the parameters. Good starting values for the algorithm can partially avoid these problems.

Finally let us note that using the relationship (7) is easy to build marginal confidence intervals based on the observed Fisher information at the maximum-likelihood estimate.

2.3. Fixed Point Iteration algorithm

A fixed point iteration (FPI) scheme was considered initially by Minka (2000) and later by Huang (2005) to get the maximum likelihood estimates of the Dirichlet distribution. It is based upon minorize maximize (MM) algorithms (see Lange, 2010) and in our case the minorizing function of the log-likelihood employs an inequality of the gamma function. Specifically the log-likelihood can be bounded as follows

$$\frac{1}{N} f\left(\boldsymbol{\alpha}\right) \geq (\sum_k \alpha_k) \Psi\left(\sum_k \alpha_k^{\text{old}}\right) - \sum_k \ln\Gamma(\alpha_k) + \sum_k (\alpha_k - 1)\frac{1}{N}\sum_i \ln y_{ik} + C$$

where C is a constant. This leads to the following equation that must be solved:

$$\Psi(\alpha_k^{\text{new}}) = \Psi(\sum_k \alpha_k^{\text{old}}) + \frac{1}{N}\sum_i \ln y_{ik}. \tag{10}$$

To get α_k^{new} we need to invert the digamma function and this is done using another iterative algorithm; therefore the whole procedure can be slow.

2.4. Starting values

The Newton-Raphson method is based upon a Taylor approximation and therefore good convergence properties are guaranteed only if the initial starting value is in a neighbourhood of the true parameter. In the literature there are many suggestions to find good starting values. We review four of them that will be used throughout the paper. We use the following notation: $\bar{y}_k = \frac{1}{N}\sum_{i=1}^N y_{ik}$, $\bar{y}_k^{(2)} = \frac{1}{N}\sum_{i=1}^N y_{ik}^2$ and $s_k^2 = \bar{y}_k^2 - (\bar{y}_k)^2$. The four initialisations will be indicated as:

moments. Matching the first two moments of the Dirichlet with the empirical moments provides the useful initialisation:

$$[\boldsymbol{\alpha}^{\text{start}}]_k = \bar{y}_k \frac{\bar{y}_k - \bar{y}_k^{(2)}}{s_k^2}. \tag{11}$$

Such initialisation employs only the marginal distributions and therefore its use of the information can be inefficient.

Ronning. Ronning (1989) proposed an initialisation to guarantee parameters in the correct range after the first iteration of the Newton-Raphson algorithm. There is however no warranty that the final estimates are in the correct range. Such initialisation gives the same value to all the parameters and unlike the moments method uses all the available data for initialising each parameter:

$$[\boldsymbol{\alpha}^{\text{start}}]_k = \min_{i\in 1,\ldots,N,\ k\in 1,\ldots,K} y_{ik}. \tag{12}$$

Dishon. Following a suggestion of Dishon and Weiss (1980), Ronning (1989) proposed a modification to the method of moments using information from all the marginals. Each parameter is estimated through:

$$[\boldsymbol{\alpha}^{\text{start}}]_k = \hat{\alpha}_0 \bar{y}_k \tag{13}$$

$$\hat{\alpha}_0 = \left\{ \Pi_{k=1}^{K-1} \left(\frac{\bar{y}_k(1-\bar{y}_k)}{s_k^2} - 1 \right) \right\}^{1/(K-1)}. \tag{14}$$

This initialisation can give parameters outside the admissible region.

Wicker. Recently Wicker et al. (2008) proposed an initialisation based on an asymptotic approximation of the likelihood. This approximation uses the limiting behaviour of the digamma function when its real argument goes to zero or infinity. Such situations are met when the number of parameters goes to infinity, i.e. for high-dimensional data. However in their simulation study Wicker et al. (2008) considered only a five-dimensional setting. Their initialisation is given by

$$[\boldsymbol{\alpha}^{\text{start}}]_k = \hat{\alpha}_0 \bar{y}_k \tag{15}$$

$$\hat{\alpha}_0 = \frac{N(K-1)\Psi(1)}{N\sum_{k=1}^K \bar{y}_k \ln \bar{y}_k - \sum_{k=1}^K \bar{y}_k \sum_{i=1}^N \ln y_{ik}}. \tag{16}$$

3. A re-parametrization

An algorithm producing estimates outside the correct range is useless. In the case of the Dirichlet distribution this problem can be avoided using a simple re-parametrization. The idea is to see the parameters $\boldsymbol{\alpha}$ as functions of other parameters free to vary on the real line: in the log-likelihood we replace α_k with $\exp(\beta_k)$. In what follows we will indicate with expNR the use of the NR algorithm applied to this re-parametrization. This way, the range of β_k is the real line and $\exp(\beta_k)$ is in the correct range for α_k. With these replacements the log-likelihood $f(\boldsymbol{\beta})$ is $f(\boldsymbol{\alpha})$ with $\boldsymbol{\alpha} = \exp(\boldsymbol{\beta}) = (\exp(\beta_1), \ldots, \exp(\beta_K))^\mathsf{T}$. The gradient can now be expressed as:

$$[\nabla f(\boldsymbol{\beta})]_k = [\nabla f(\boldsymbol{\alpha})]_k \Big|_{\exp(\boldsymbol{\beta})} \exp(\beta_k).$$

The Hessian has a form similar to the original one:

$$\mathbf{H} = \mathbf{Q} + \exp(\boldsymbol{\beta}) \exp(\boldsymbol{\beta})^\mathsf{T} z \tag{17}$$

$$q_{jk} = \left([\nabla f(\boldsymbol{\beta})]_k - \exp(\beta_k + \beta_j)\Psi'(\exp(\beta_k))\right) \delta(j-k)N \tag{18}$$

$$z = N\Psi' \left(\sum_k \exp(\beta_k) \right) \tag{19}$$

As before, the diagonal form of the square matrix $\mathbf{Q}$ and the Sherman-Morrison formula are sufficient to guarantee that the inverse of $\mathbf{H}$ exists if the diagonal elements of $\mathbf{Q}$ are non zero. However, after the re-parametrization the problem is not necessarily concave in the new parameters. Therefore we cannot say that the inequalities of the previous section hold true also now. For the re-parametrization the elements in the diagonal of $\mathbf{Q}$ are strictly negative in a neighbourhood of the point of maximum. Indeed at the maximum the gradient is zero and therefore $q_{kk} = -\exp(2\beta_k)\Psi'(\exp(\beta_k))N < 0$. Within such neighbourhood we have:

$$\mathbf{H}^{-1} = \mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1} \exp(\boldsymbol{\beta}) \exp(\boldsymbol{\beta})^\mathsf{T} \mathbf{Q}^{-1}}{\frac{1}{z} + \exp(\boldsymbol{\beta})^\mathsf{T} \mathbf{Q}^{-1} \exp(\boldsymbol{\beta})} \tag{20}$$

$$\left[\mathbf{H}^{-1} \nabla f(\boldsymbol{\beta})\right]_k = \frac{1}{q_{kk}} \left([\nabla f(\boldsymbol{\beta})]_k - \exp(\beta_k) b\right) \tag{21}$$

$$b = \frac{\exp(\boldsymbol{\beta})^\mathsf{T} \mathbf{Q}^{-1} \nabla f(\boldsymbol{\beta})}{\frac{1}{z} + \exp(\boldsymbol{\beta})^\mathsf{T} \mathbf{Q}^{-1} \exp(\boldsymbol{\beta})} \tag{22}$$

$$= \frac{\sum_j \exp(\beta_j) [\nabla f(\boldsymbol{\beta})]_j / q_{jj}}{1/z + \sum_j \exp(2\beta_j)/q_{jj}}. \tag{23}$$

Equations (21), (22) and (23) assure an easy way to implement the Newton-Raphson algorithm avoiding explicit matrix inversion. With these new quantities the Newton-Raphson iteration is again:

$$\boldsymbol{\beta}^{\text{new}} = \boldsymbol{\beta}^{\text{old}} - \mathbf{H}^{-1}\nabla f(\boldsymbol{\beta}^{\text{old}}).$$

As starting values for the algorithm we can consider the logarithms of the starting values previously described.

4. A stable algorithm

The Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) was originally proposed to solve non-linear least-squares minimization problems. The idea is to use a function of $\mathbf{H}$ instead of $\mathbf{H}$ itself. With an appropriate choice the iterations of the algorithm can take into account the curvature of the function being optimized. For the optimization of the Dirichlet log-likelihood we propose an iteration algorithm similar to the Levenberg-Marquardt one (LM):

$$\boldsymbol{\alpha}^{\text{new}} = \boldsymbol{\alpha}^{\text{old}} - \{\mathbf{H} + \gamma \text{diag}\mathbf{H}\}^{-1}\nabla f(\boldsymbol{\alpha}^{\text{old}}) \tag{24}$$

where γ is a positive constant or a positive function not depending on the parameters. The effect of the damping parameter γ is that of shortening the steps of NR algorithm, providing in this way prudent steps in the iterations. The same algorithm can be applied to the re-parametrization (expLM):

$$\boldsymbol{\beta}^{\text{new}} = \boldsymbol{\beta}^{\text{old}} - \{\mathbf{H} + \gamma \text{diag}\mathbf{H}\}^{-1}\nabla f(\boldsymbol{\beta}^{\text{old}}). \tag{25}$$

Let us note that a similar approach is backtracking. In backtracking the matrix approximating the Hessian is multiplied by a damping parameter that is eventually shrunken to assure an ascent step. The damping parameter influences the step length. The rational of this approach is related to the Taylor expansion of the gradient calculated at the new parameter. We prefer instead the Levenberg-Marquardt approach because the damping parameter can influence both the direction and the size of the step (Madsen et al., 2004). Let us denote with $\mathbf{x}$ the parameters of interest ($\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ in the previous cases); working on the iteration map $\mathbf{M}(\mathbf{x})$ defined by

$$\mathbf{M}(\mathbf{x}) = \mathbf{x} - \{\mathbf{H}(\mathbf{x}) + \gamma \text{diag}\mathbf{H}(\mathbf{x})\}^{-1}\nabla f(\mathbf{x}) \tag{26}$$

we show in the appendix that both algorithms (24) and (25) converge to the maximum.

Similarly to what we have seen in the previous sections for the Newton-Raphson algorithm, we show how to rearrange the quantities involved in this version of the Levenberg-Marquardt algorithm using expressions without an explicit use of inverse matrices. Equations (24) and (25) can be rewritten with the following quantities

$$\{\mathbf{H}(\mathbf{x})+\gamma \text{diag}\mathbf{H}(\mathbf{x})\}^{-1} = \mathbf{D}-\mathbf{L} \tag{27}$$

$$\mathbf{D} = \{\mathbf{Q}(\mathbf{x})+\gamma \text{diag}\mathbf{H}(\mathbf{x})\}^{-1} \tag{28}$$

where for the original parametrization

$$\mathbf{L} = \frac{\mathbf{D}\mathbf{1}\mathbf{1}^{\mathsf{T}}\mathbf{D}}{\frac{1}{z}+\mathbf{1}^{\mathsf{T}}\mathbf{D}\mathbf{1}} \tag{29}$$

$$[(\mathbf{D}-\mathbf{L})\nabla f(\boldsymbol{\alpha})]_k = \frac{[\nabla f(\boldsymbol{\alpha})]_k - b}{q_{kk}(1+\gamma)+\gamma z} \tag{30}$$

$$b = \frac{\mathbf{1}^{\mathsf{T}}\mathbf{D}\nabla f(\boldsymbol{\alpha})}{\frac{1}{z}+\mathbf{1}^{\mathsf{T}}\mathbf{D}\mathbf{1}} \tag{31}$$

$$= \frac{\sum_k [\nabla f(\boldsymbol{\alpha})]_k/(q_{kk}(1+\gamma)+\gamma z)}{1/z+\sum_k 1/(q_{kk}(1+\gamma)+\gamma z)} \tag{32}$$

while for the re-parametrization

$$\mathbf{L} = \frac{\mathbf{D}\exp(\boldsymbol{\beta})\exp(\boldsymbol{\beta})^{\mathsf{T}}\mathbf{D}}{\frac{1}{z}+\exp(\boldsymbol{\beta})^{\mathsf{T}}\mathbf{D}\exp(\boldsymbol{\beta})} \tag{33}$$

$$[(\mathbf{D}-\mathbf{L})\nabla f(\boldsymbol{\beta})]_k = \frac{[\nabla f(\boldsymbol{\beta})]_k - \exp(\beta_k)b}{q_{kk}(1+\gamma)+\gamma z\exp(2\beta_k)} \tag{34}$$

$$b = \frac{\exp(\boldsymbol{\beta})^{\mathsf{T}}\mathbf{D}\nabla f(\boldsymbol{\beta})}{\frac{1}{z}+\exp(\boldsymbol{\beta})^{\mathsf{T}}\mathbf{D}\exp(\boldsymbol{\beta})} \tag{35}$$

$$= \frac{\sum_k \frac{\exp(\beta_k)[\nabla f(\boldsymbol{\beta})]_k}{(q_{kk}(1+\gamma)+\gamma z\exp(2\beta_k))}}{\frac{1}{z}+\sum_k \frac{\exp(2\beta_k)}{(q_{kk}(1+\gamma)+\gamma z\exp(2\beta_k))}}. \tag{36}$$

All the equalities are valid when we can apply the Sherman-Morrison formula, see Sections 2 and 3. In such cases, as for the described Newton-Raphson algorithms, there is no need to store and invert the Hessian matrix.

4.1. Damping parameter and stopping criteria

The damping parameter γ influences the step size in the LM iterations. A very small value of γ leads to an algorithm that is very close to the NR algorithm. This behaviour is good when we are close to the maximum because we are close to quadratic convergence. However, larger values of γ can be good if the actual value is far from the optimum. In this case, a larger γ produces shorter steps in the iterations. Nielsen (1998) proposed to use values very close to zero or related to the diagonal element of the matrix, approximating the Hessian in the Levenberg-Marquardt original algorithm. Similarly, we propose to use $1/K$ as a value close to zero, or 1 which is the diagonal element of the Hessian when rescaled by its diagonal elements, $[\text{diag}(\mathbf{D})]^{-1}\mathbf{D}$ (this form is similar to that of the Levenberg-Marquardt algorithm). Contrary to the Levenberg-Marquardt algorithm our damping parameter is not adaptive. However, we prove in the appendix a convergence property of our algorithm and we show its performance in simulations and on real data.

We stop the algorithm as soon as one of these three criteria is satisfied: if the norm of the gradient is very close to zero: $\| \nabla f \| < \epsilon_1$; if the relative changes of the parameters are very small: $\| \mathbf{x}_{\text{new}} - \mathbf{x} \| < \epsilon_2(\| \mathbf{x} \| + \epsilon_2)$; if the number of iterations is greater than a pre-established threshold.

5. Simulated data

To compare the different proposals in an high-dimensional setting we have implemented a simulation with 1000 variables and 20 units. With a huge number of parameters there is an high chance that an element of a simulated unity is so close to zero to be considered zero due to machine precision. This problem almost disappears when all the parameters have values far from zero. To investigate the consequences of such choices we consider a range of values for the sum of the parameters $\sum \alpha_k$ from 10000 to 50000 with step size of 2000. Each parameter is drawn from a uniform distribution between $\sum \alpha_k/K - 2$ and $\sum \alpha_k/K + 2$ where $K = 1000$. Let us remark that the final sum of the simulated parameters is not necessarily equal to the pre-established values in the sequence.

We consider that a method has reached convergence when it is stopped before the number of iterations reaches 1000 and the estimated vector is in the correct range. The tolerance parameters ϵ_1 and ϵ_2 are both set to 10^{-8} and for the damping parameter γ we use the values considered in the previous section. The number of simulations is 2000.

The results are reported in Figure 1. In the upper panel we report the convergence rate for each combination of starting values/methods. In the lower panel the mean number of iterations required for convergence is shown. FPI and LM methods with $\gamma = 1$ have a similar performance, reaching convergence very often and for every starting value. L-BFGS-B shows a good range of convergence when coupled with Wicker or the method of moments but not with the other two initialisation methods. The starting values of

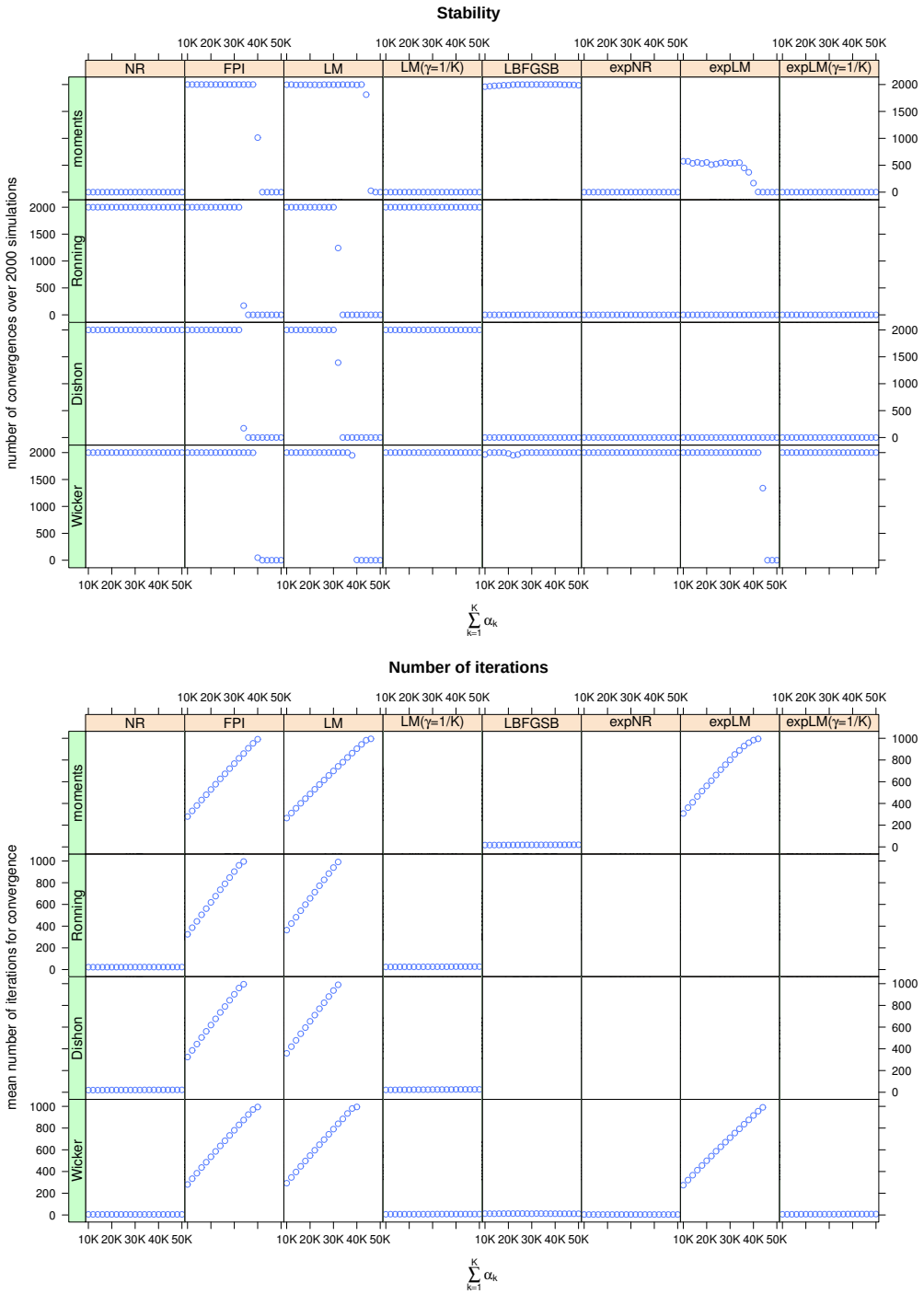


Figure 1: *Results from the simulation study. In the upper panel we show how many times we reached convergence for each combination of starting value and algorithm. Below we show instead the number of iterations used to reach convergence in the upper panel.* $\sum_{k=1}^{K} \alpha_k$ *indicates approximatively the sum of the parameters to be estimated.*

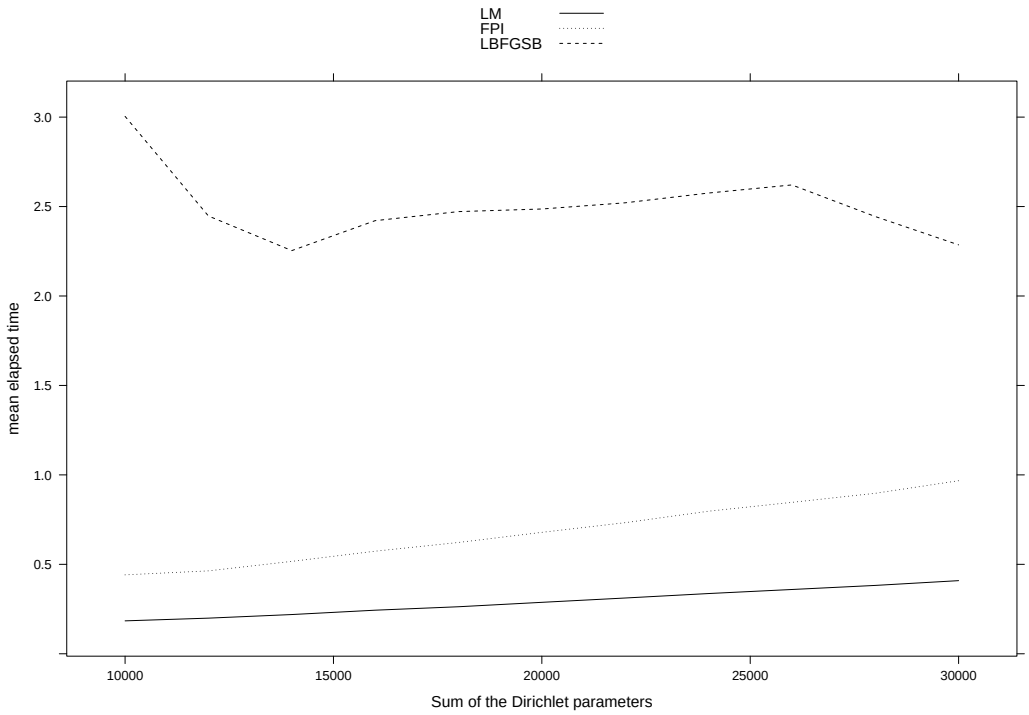


Figure 2: *Time comparison of the algorithms LM, FPI and L-BFGS-B over 100 simulations. Time is expressed in seconds.*

Wicker et al. (2008) are the only ones able to guarantee convergence for all the methods. The algorithms using the re-parametrization show poor performance, probably due to the possible lack of concavity and the fact that only the starting values of Wicker et al. (2008) seem to be often in a neighbourhood of the maximum. However, the price to pay for the highest stability is the high number of iterations required to reach convergence for both FPI and LM with $\gamma = 1$. NR was instead the fastest method. As expected, see Section 4.1, LM with $\gamma = 1/K$ and expLM with with $\gamma = 1/K$ have a performance close to those of NR and expNR, respectively.

The efficiency of the algorithms LM, FPI and L-BFGS-B cannot be compared only looking for differences in the number of iterations because their corresponding iteration times can be totally different. For these algorithms we therefore implemented also a comparison on the total time required to reach convergence. The settings for this simulation were similar to the ones used above with a range for the sum of the parameters going from 10000 to 30000 with step size of 2000 and a number of simulations equal to 2000. The results are reported in Figure 2. On average LM requires only half of the time employed by FPI. L-BFGS-B is clearly much slower than the other two competitors.

6. Apple data set

We have seen in the introduction that the Dirichlet distribution has strong independence properties that are often unrealistic for real data. However, for high-dimensional data once we focus on a single variable we expect that this is correlated only with a limited number of the other variables and uncorrelated with the rest. In this case it can be of interest to see the fit of a simple model as the Dirichlet distribution, that can be thought of as a raw approximation of the reality. We are going to consider a spike-in experiment that has been already investigated in the literature with the appropriate statistical tools. If we make the corresponding data compositional and we investigate them through the Dirichlet distribution we can judge if there is a discrepancy or an agreement between what is already known and what we can learn with the Dirichlet distribution. Since a spike-in experiment is a controlled experiment where we know the thruth, this comparison can indirectly tell us if the Dirichlet distribution can be used for high-dimensional data. This is not intended to be a proof to say that the Dirichlet distribution can safely be used for high-dimensional compositional data. For a stronger result, comparisons must be done with other distributions. Here we focus on the computational performance of different algorithms to get the maximum likelihood estimates and we offer a brief look at the possibilities of using the Dirichlet distribution to analyse high-dimensional data.

Specifically we apply the previous algorithms to a data set from the field of metabolomics, available in the R package BioMark (Wehrens and Franceschi, 2012). The data set consists of mass spectrometric measurements on apples and is fully described in Franceschi et al. (2012). We consider the positive ionization data for the 10 control samples, and the first group of 10 spiked-in samples. There are 1632 variables in total. We delete the variables with missing values and normalize the data to give 1 as the sum of the elements of each unit. This corresponds to have a total intensity for each unit that is redistributed through the variables. The final data set has 1602 variables. The results are summarised in Table 1.

For these data, the initialisation of Wicker and co-workers is able to give convergence in the correct range of the parameters for five methods. In two cases (LM with $\gamma = 1/K$, NR) the result is outside the correct range. The other initialisations fail for expLM ($\gamma = 1$), expLM ($\gamma = 1/K$) and expNR. The initialisation based on the method of moments fails also for LM ($\gamma = 1$). L-BFGS-B is not able to reach convergence with any initialisation method. The only method that is always able to reach convergence in the correct range is FPI. LM ($\gamma = 1$) and LM ($\gamma = 1/K$) reach convergence in three out of four cases, while NR converges in two out of four cases. Using the exponential parametrization the other methods reach convergence only with the Wicker initialisation, but in these cases very few iterations are needed.

Table 1: *Number of iterations required for convergence in the Apple data set. We report with a bar the cases where convergence is not reached or the result is outside the correct range for the parameters.*

	NR	FPI	LM	LM($\gamma = \frac{1}{K}$)	L-BFGS-B	expNR	expLM	expLM($\gamma = \frac{1}{K}$)
moments	—	543	—	—	—	—	—	—
Ronning	30	495	543	32	—	—	—	—
Dishon	21	494	529	23	—	—	—	—
Wicker	—	434	447	—	—	7	411	8

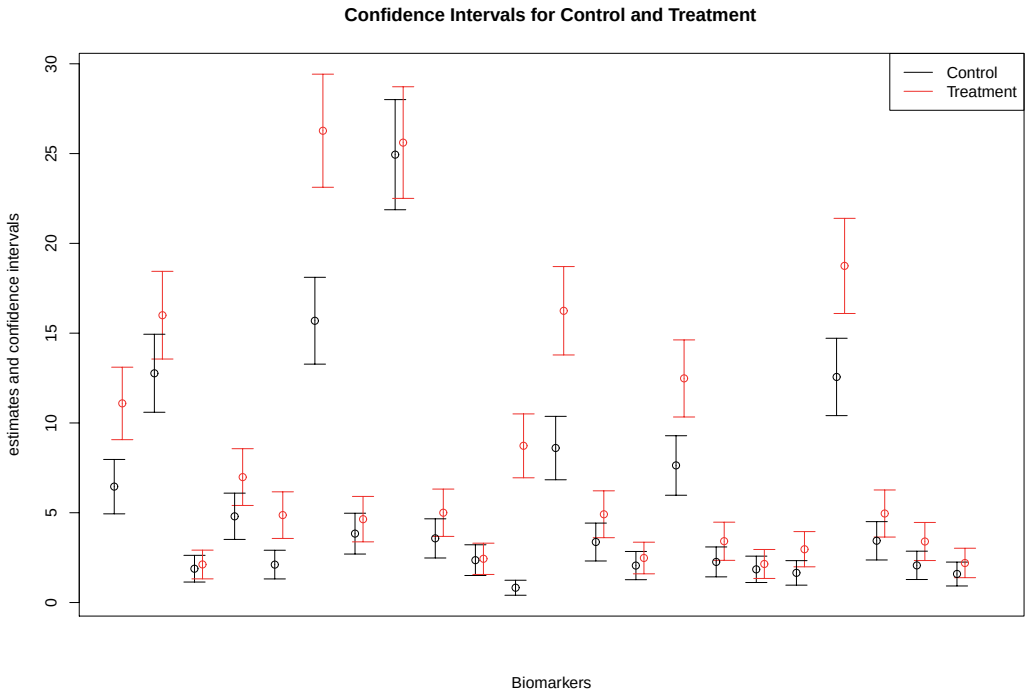


Figure 3: *Each pair in the figure represents the confidence intervals for control and treatment respectively, for 22 biomarkers.*

While FPI is very stable, it also requires a large number of iterations to reach convergence with each initialisation (mean for the four initialisations $= 491.5$). The same is true for LM ($\gamma = 1$) where the mean for three initialisations that reach convergence is 506.3 and for the initialisation with the method of moments we do not have convergence because we reach the maximum number of iteration allowed (1000 in our setting). However FPI can require a longer time than LM because FPI requires for each iteration a second inner iterative algorithm. For example, comparing FPI and LM ($\gamma = 1$) using the Wicker initialisation we have an elapsed time of 2.259 and 0.678 seconds respectively.

With Wicker initialisation expLM ($\gamma = 1$) requires 411 iterations to reach convergence while expLM ($\gamma = 1/K$) requires only 8 steps; similarly expNR requires only

7 steps. LM ($\gamma = 1/K$) and NR also required a low number of iterations when they reached convergence.

Fitting different Dirichlet distributions to control and treatment data it is possible to compare them. We report the confidence intervals for 22 biomarkers in Figure 3. These biomarkers are known to correspond to spike-in compounds (see Franceschi et al., 2012). Since these confidence intervals are not simultaneous we cannot use them for identifying 'directly' statistically significant biomarkers. We use them for ranking these biomarkers, visualizing the order of magnitude of the differences between control and treatment. There are several cases where the differences are clear. This is in agreement with the previous findings in Franceschi et al. (2012).

7. Conclusions

In this paper we have compared the computational performance of eight different algorithms and four different starting value strategies to estimate the Dirichlet distribution through maximum likelihood. Such a comparison provides indications about the methods to use in order to analyse high-dimensional compositional data with the Dirichlet distribution.

The Newton-Raphson algorithm is very fast, but can lead to estimated parameters outside the allowed region. On the other hand, the FPI algorithm has a slow convergence but is very stable. The other algorithms have a performance between these two extremes.

To have parameters always in the correct range we considered a re-parametrization and a box-constraints algorithm, L-BFGS-B. The re-parametrization allows us to have parameters always in the correct range but possibly loses the characteristic of being a concave function. This means that good convergence is assured only in a neighbourhood of the maximum and that convergence cannot be guaranteed. In practice from our study we can see that the re-parametrization is useful only if coupled with the initialisation method of Wicker et al. (2008). L-BFGS-B is more stable than the re-parametrization but less than FPI and moreover its iteration time is huge.

The proposed modifications to the Levenberg-Marquardt algorithm consider a prudent step compared to the Newton-Raphson algorithm and therefore can offer a good trade-off between speed and stability. Newton-Raphson and the proposed algorithms have local convergence characteristics and therefore the starting values are very important even if the function to be optimized is concave. These features are particularly relevant in a high-dimensional setting where the number of parameters largely exceed the number of units. From the simulations and the real study only the Wicker et al. (2008) approach seems able to provide convergence for high-dimensional data.

Considering both simulations and the real data example the combination of the Levenberg-Marquardt methods or fixed point iteration method with the starting values of Wicker appear to be the most promising. However, the Levenberg-Marquardt methods leave room for improvements. In this paper we have been able to prove convergence

properties for a fixed damping parameter γ. If this parameter can be made adaptive, as in the original Levenberg-Marquardt algorithm, it is not unreasonable to expect a higher stability and a lower mean number of iterations for convergence.

Appendix

In what follows the notation refers to the quantities previously introduced in the paper.

Lemma 1 *At the point of maximum, $\hat{\mathbf{x}}$, the differential of the iteration map (26) is given by*

$$d\mathbf{M}(\hat{\mathbf{x}}) = \mathbf{I} - \{\mathbf{H}(\hat{\mathbf{x}}) + \gamma \text{diag}\mathbf{H}(\hat{\mathbf{x}})\}^{-1}\mathbf{H}(\hat{\mathbf{x}}).$$

Proof. We treat γ as a positive constant, but the proof holds even if γ is a positive function that does not depend on $\mathbf{x}$.

Let us denote by $\mathbf{G}(\mathbf{x})$ the matrix $\{\mathbf{H}(\mathbf{x}) + \gamma \text{diag}\mathbf{H}(\mathbf{x})\}^{-1}$. We know from Section 4 that $\mathbf{G}(\mathbf{x})$ is well defined for every $\mathbf{x}$ in the original parametrization and at least in a neighbourhood of the maximum for the re-parametrization. For such cases the elements of $\mathbf{G}(\mathbf{x})\nabla f(\mathbf{x})$ can be written as $\sum_j g_{ij}(\mathbf{x})\nabla f_j(\mathbf{x})$. To prove the lemma we have to show that the partial derivatives of this expression are well defined. Using the product rule it is patent that the difficult part is to prove that the partial derivatives of the elements $g_{ij}(\mathbf{x})$ are well defined. If we are able to prove that the l_{ij} and the elements of the diagonal matrix $\mathbf{D}$ are derivable it follows that also the g_{ij} are derivable and therefore the lemma easily follows. Using standard rules for derivation we see that these terms are derivable if the trigamma function is derivable. For positive reals the trigamma function can be expressed as a positive series dominated by $\sum n^{-2}$. Therefore by the Weierstrass M-test there is uniform convergence and the trigamma function is derivable. Moreover, the series form assures that the trigamma function is strictly decreasing for positive reals.

We can summarize the results in matrix form. At the point of maximum $\nabla f(\hat{\mathbf{x}}) = \mathbf{0}$ and therefore we get $d\mathbf{M}(\hat{\mathbf{x}}) = \mathbf{I} - \mathbf{G}(\hat{\mathbf{x}})\mathbf{H}(\hat{\mathbf{x}}) = \mathbf{I} - \{\mathbf{H}(\hat{\mathbf{x}}) + \gamma \text{diag}\mathbf{H}(\hat{\mathbf{x}})\}^{-1}\mathbf{H}(\hat{\mathbf{x}})$. ■

Theorem 1 *The proposed Levenberg-Marquardt algorithms based upon equations (24) and (25) are locally attracted to the maximum $\hat{\mathbf{x}}$ at a linear rate equal to the spectral radius of*

$$\mathbf{I} - \{\mathbf{H}(\hat{\mathbf{x}}) + \gamma \text{diag}\mathbf{H}(\hat{\mathbf{x}})\}^{-1}\mathbf{H}(\hat{\mathbf{x}})$$

or at a better rate.

Proof. The point of maximum for $f(\mathbf{x})$ is a fixed point for $\mathbf{M}(\mathbf{x})$. According to Proposition 15.3.1 in Lange (2010), it suffices to show that all eigenvalues of the differential

$d\mathbf{M}(\hat{\mathbf{x}})$ lie on the open interval $(0,1)$. By Lemma 1 the following equalities hold:

$$\begin{aligned} d\mathbf{M}(\hat{\mathbf{x}}) &= \mathbf{I} - \{\mathbf{H}(\hat{\mathbf{x}}) + \gamma \text{diag}\mathbf{H}(\hat{\mathbf{x}})\}^{-1} \mathbf{H}(\hat{\mathbf{x}}) \\ &= \{\mathbf{H}(\hat{\mathbf{x}}) + \gamma \text{diag}\mathbf{H}(\hat{\mathbf{x}})\}^{-1} \{\mathbf{H}(\hat{\mathbf{x}}) + \gamma \text{diag}\mathbf{H}(\hat{\mathbf{x}}) - \mathbf{H}(\hat{\mathbf{x}})\} \\ &= \{\mathbf{H}(\hat{\mathbf{x}}) + \gamma \text{diag}\mathbf{H}(\hat{\mathbf{x}})\}^{-1} \{\gamma \text{diag}\mathbf{H}(\hat{\mathbf{x}})\}. \end{aligned}$$

The maximum and minimum eigenvalues of $d\mathbf{M}(\hat{\mathbf{x}})$ are determined by the maximum and minimum values of the Rayleigh quotient ($\mathbf{v} \neq \mathbf{0}$):

$$\begin{aligned} R(\mathbf{v}) &= \frac{\mathbf{v}^\mathsf{T} [\gamma \text{diag}\mathbf{H}(\hat{\mathbf{x}})] \mathbf{v}}{\mathbf{v}^\mathsf{T} [\mathbf{H}(\hat{\mathbf{x}}) + \gamma \text{diag}\mathbf{H}(\hat{\mathbf{x}})] \mathbf{v}} \\ &= 1 - \frac{\mathbf{v}^\mathsf{T} \mathbf{H}(\hat{\mathbf{x}}) \mathbf{v}}{\mathbf{v}^\mathsf{T} [\mathbf{H}(\hat{\mathbf{x}}) + \gamma \text{diag}\mathbf{H}(\hat{\mathbf{x}})] \mathbf{v}}. \end{aligned}$$

If the quantities $\mathbf{H}(\hat{\mathbf{x}})$ and $\gamma \text{diag}\mathbf{H}(\hat{\mathbf{x}})$ are definite negative then also $\mathbf{H}(\hat{\mathbf{x}}) + \gamma \text{diag}\mathbf{H}(\hat{\mathbf{x}})$ is definite negative and it follows that $0 < R(\mathbf{v}) < 1$.

For the original parametrization we can show that $\gamma \text{diag}\mathbf{H}(\mathbf{x})$ is always definite negative. We use the parametric form of $\mathbf{H}$ for the Dirichlet distribution. We have $\mathbf{v}^\mathsf{T} \gamma \text{diag}\mathbf{H}(\mathbf{x}) \mathbf{v} = \gamma \sum v_i^2 h_{ii}$ where $h_{ii} = N(\Psi'(\sum \alpha_j) - \Psi'(\alpha_i))$.We have seen that for positive reals the trigamma function is strictly decreasing and therefore $h_{ii} < 0$. Therefore $\mathbf{v}^\mathsf{T} \gamma \text{diag}\mathbf{H}(\mathbf{x}) \mathbf{v} < 0$. Being $f(\cdot)$ concave $\mathbf{H}$ is semi-definite negative, but $\mathbf{H}$ is also invertible and therefore is definite negative.

For the re-parametrization we cannot assure that the Hessian matrix is negative definite for every $\mathbf{x}$. However, to apply Proposition 15.3.1 in Lange (2010) we need only to prove that this is true at $\hat{\mathbf{x}}$. For the diagonal matrix we observe that:

$$[\lambda \text{diag}\mathbf{H}(\hat{\mathbf{x}})]_{kk} = \lambda q_{kk} + \lambda \exp(2\beta_k) N \Psi' \left(\sum_k \exp(\beta_k) \right) \tag{37}$$

$$= \lambda N [\nabla f(\boldsymbol{\beta})]_k \tag{38}$$

$$+ \lambda \exp(2\beta_k) N \left[\Psi'(\sum_k \exp(\beta_k)) - \Psi'(\exp(\beta_k)) \right]. \tag{39}$$

By the properties of the trigamma function $\Psi'(\sum_k \exp(\beta_k)) - \Psi'(\exp(\beta_k)) < 0$ and therefore at $\hat{\mathbf{x}}$ the matrix $\lambda \text{diag}\mathbf{H}$ is negative definite. Moreover at $\hat{\mathbf{x}}$ the Hessian of the re-parametrization is semi-definite negative and invertible and therefore is definite negative. ■

Acknowledgement

The authors thank Federico Vaggi for useful comments on an earlier version of the manuscript.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Monographs on statistics and applied probability. Chapman and Hall.

Byrd, R. H., Lu, P., Nocedal, J. and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. Scientific Computing*, 16, 1190–1208.

Connor, R. J. and Mosiman, J. E. (1969). Concepts of independence for proportions with a generalization of the Dirichlet distribution. *Journal of the American Statistical Association*, 64, 194–206.

Dishon, M. and Weiss, G. (1980). Small sample comparison of estimation methods for the beta distribution. *Journal of Statistics Computation and Simulation*, 11, 1–11.

Franceschi, P., Masuero, D., Vrhovsek, U., Mattivi, F. and Wehrens, R. (2012). A benchmark spike-in data set for biomarker selection in metabolomics. *Journal of Chemometrics*, 26, 16–24.

Huang, J. (2005). Maximum likelihood estimation of Dirichlet distribution parameters. http://www.stanford.edu/ jhuang11/research/dirichlet/, Robotics Institute, Carnegie Mellon University.

Lange, K. (2010). *Numerical Analysis for Statisticians*. Springer, second edition.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2, 164–168.

Madsen, K., Nielsen, H. B. and Tingleff, O. (2004). *Methods for Non-Linear Least Squares Problems* (2nd ed.). Informatics and Mathematical Modelling, Technical University of Denmark, DTU.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11, 431–441.

Minka, T. P. (2000). Estimating a Dirichlet distribution. http://research.microsoft.com/ en-us/um/people/minka/papers/, M.I.T.

Narayanan, A. (1991a). Algorithm AS266: maximum likelihood estimation of parameters of the Dirichlet distribution. *Applied Statistics*, 40, 365–374.

Narayanan, A. (1991b). Small sample properties of parameter estimation in Dirichlet distribution. *Communications in Statistics Simulations and Computation*, 20, 647–666.

Nielsen, H. B. (1998). *Damping Parameter in Marquardt's Method*. Technical Report IMM-REP-1999-05. Department of Mathematical Modelling, DTU.

Rayens, W. S. and Srinivasan, C. (1994). Dependence properties of generalized Liouville distributions on the Simplex. *Journal of the American Statistical Association*, 89, 1465–1470.

Ronning, G. (1989). Maximum likelihood estimation of Dirichlet distributions. *Journal of Statistics Computation and Simulation*, 32, 215–221.

Wehrens, R. and Franceschi, P. (2012). Meta-statistics for variable selection: the R package BioMark. *Journal of Statistical Software*, 51, 1–18.

Wicker, N., Muller, J., Kalathur, R. K. R. and Poch, O. (2008). A maximum likelihood approximation method for Dirichlet's parameter estimation. *Computational Statistics and Data Analysis*, 52, 1315–1322.

SORT 39 (1) January-June 2015, 127-146

The exponentiated discrete Weibull distribution

Vahid Nekoukhou[1] and Hamid Bidram[2,*]

Abstract

In this paper, the exponentiated discrete Weibull distribution is introduced. This new generalization of the discrete Weibull distribution can also be considered as a discrete analog of the exponentiated Weibull distribution. A special case of this exponentiated discrete Weibull distribution defines a new generalization of the discrete Rayleigh distribution for the first time in the literature. In addition, discrete generalized exponential and geometric distributions are some special sub-models of the new distribution. Here, some basic distributional properties, moments, and order statistics of this new discrete distribution are studied. We will see that the hazard rate function can be increasing, decreasing, bathtub, and upside-down bathtub shaped. Estimation of the parameters is illustrated using the maximum likelihood method. The model with a real data set is also examined.

MSC: 60E05, 62E10

Keywords: Discrete generalized exponential distribution, exponentiated discrete Weibull distribution, exponentiated Weibull distribution, geometric distribution, infinite divisibility, order statistics, resilience parameter family, stress-strength parameter.

1. Introduction

It is sometimes impossible or inconvenient to measure the life length of a device on a continuous scale. In practice, we come across situations where lifetimes are recorded on a discrete scale. For example, on/off switching devices, bulb of photocopier machine, to and fro motion of spring devices, etc. (cf. Krishna and Singh, 2009) are some typical situations.

The failure rate function of an object, when the failures are reported on a discrete scale, may be bathtub-shaped or unimodal. Jiang (2010) investigated some discrete dis-

* Corresponding author: h.bidram@sci.ui.ac.ir (and hamid_bidram@yahoo.com)

[1] Department of Statistics, University of Isfahan, Khansar Unit, Isfahan, Iran.

[2] Department of Statistics, University of Isfahan, Isfahan, Iran.

Received: July 2014
Accepted: February 2015

tributions and used the exponentiated Poisson distribution and the two-fold competing risk model exhibiting bathtub-shaped or increasing failure rate functions to introduce a model for bus-motor failure data. Nooghabi et al. (2011) introduced the discrete modified Weibull distribution with increasing and bathtub-shaped failure rate function. However, in application areas, the absence of a suitable discrete model whose hazard rate function covers and contains different possible shapes, i.e., bathtub-shaped, upside-down bathtub and monotonically increasing and decreasing, is perceived. On the other hand, the traditional discrete distributions have limited applicability as models for reliability, failure times, counts, etc.

In the last two decades some papers dealing with discrete distributions obtained by discretizing a continuous distribution have appeared in the literature. Lisman and van Zuylen (1972) proposed and Kemp (1997) studied the discrete normal distribution which is characterized by maximum entropy for specified mean and variance; see also Dasgupta (1993) and Szablowski (2001). Roy (2003) introduced another discrete analog of normal distribution. Kemp (2008) also considered the discrete half-normal distribution as a maximum entropy distribution for given mean and variance. Inusah and Kozubowski (2006) and Kozubowski and Inusah (2006) introduced Laplace and skew-Laplace distributions on the lattice of integers, respectively. Barbiero (2014) considered an alternative discrete skew Laplace distribution. Krishna and Pundir (2007) introduced the discrete Maxwell distribution. Krishna and Pundir (2009) introduced the discrete Burr distribution and studied a special case of the distribution which led to perform the discrete Pareto distribution. Jazi et al. (2010) studied the discrete inverse Weibull distribution and proposed some important properties of their discrete model. Gómez-Déniz (2010) obtained a generalization of the geometric distribution from a member of the Marshall and Olkin (1997) family of distributions. Gómez-Déniz and Calderin-Ojeda (2011) considered the discrete Lindley distribution and investigated some properties and applications of the model. Chakraborty and Chakravarty (2012) studied discrete gamma distributions and discussed estimation of the parameters. In addition, Chakraborty (2013) introduced a new discrete distribution related to generalized gamma distribution. Chakraborty and Chakravarty (2013) introduced a new discrete probability distribution on the lattice of integers. Nekoukhou et al. (2013a) studied the discrete beta exponential (DBE) distribution and illustrated that the hazard rate function of this discrete analogue of the beta exponential distribution of Nadarajah and Kotz (2006) is monotone. Moreover, Hussain and Ahmad (2014) and Chakraborty and Chakravarty (2014) introduced the discrete inverse Rayleigh and discrete Gumbel distributions, respectively.

Recently, Nekoukhou et al. (2012) and (2013b) introduced two different discrete counterparts of the well-known two-parameter generalized exponential (GE) distribution of Gupta and Kundu (1999, 2001 and 2007). The probability mass functions (pmfs) of these distributions are

$$p_x = f(x;p,\gamma) = cp^{x-1}(1-p^x)^{\gamma-1}, \quad x \in \mathbb{N} = \{1,2,3,\dots\}, \tag{1}$$

where c is the norming constant, and

$$p_x = f(x;p,\gamma) = (1-p^{x+1})^{\gamma} - (1-p^{x})^{\gamma}, \quad x \in \mathbb{N}_0 = \{0,1,2,\dots\}, \tag{2}$$

respectively. Nekoukhou et al. (2012) and (2013b) introduced these discrete analogues by using relations

$$p_x = \frac{f(x)}{\sum_{x=1}^{\infty} f(x)}, \quad x \in \mathbb{N} \tag{3}$$

and

$$p_x = S(x) - S(x+1), \quad x \in \mathbb{N}_0, \tag{4}$$

respectively, where $f(.)$ and $S(.)$ are the probability density function (pdf) and survival function of the GE distribution. Discrete generalized exponential (DGE) distribution and discrete generalized exponential distribution of a second type (DGE_2) are introduced in the literature via Eq.'s (1) and (2), respectively. The last authors denoted these two-parameter discrete distributions by $DGE(\gamma,p)$ and $DGE_2(\gamma,p)$. Eq. (2) yields that the cumulative distribution function (cdf) of the $DGE_2(\gamma,p)$ distribution is given by

$$F(x;p,\gamma) = (1-p^{[x]+1})^{\gamma}, \quad x \geq 0. \tag{5}$$

It is interesting to note that the above cdf coincides with the exponentiated geometric distribution which was mentioned in Jiang (2010), and investigated by Chakraborty and Gupta (2012).

In this paper we will introduce the exponentiated discrete Weibull (EDW) distribution, which is really a generalization of the discrete Weibull (DW) distribution of Nakagawa and Osaki (1975) and also DGE_2 distribution, and illustrate its important features and properties. The failure rate function of the new model is found to be bathtub-shaped, unimodal and also increasing and decreasing. In the application section we will see that the new model provides a satisfactory fit and that is competitive with traditional and also newly developed discrete models. The new discrete distribution also contains a generalization of the discrete Rayleigh distribution of Roy (2004) which has not been introduced in the literature yet.

The paper is organized as follows. Section 2 introduces the three-parameter EDW distribution and discusses some of its important features and properties such as cumulative distribution and hazard rate functions, moments, infinite divisibility and the order statistics. In Section 3, the researchers will consider the maximum likelihood method to estimate the parameters of EDW distribution. In addition, in this section, estimation of the stress-strength parameter is discussed. Section 4 describes fitting of the proposed model to a real data set. Finally, in Section 5 some concluding remarks are given.

2. Three-parameter EDW distribution

When the cdf of the DW distribution, denoted by $DW(p,\alpha)$, of Nakagawa and Osaki (1975), i.e.,

$$G(x;p,\alpha) = 1 - p^{([x]+1)^\alpha}, \quad x \geq 0, \tag{6}$$

where $0 < p < 1$ and $\alpha > 0$ are the model parameters, is inserted into the *resilience parameter family* of distributions, the cdf of the resulting discrete distribution is given by

$$F(x;p,\alpha,\gamma) = \{1 - p^{([x]+1)^\alpha}\}^\gamma, \quad x \geq 0 \tag{7}$$

in which $\gamma > 0$ is the *resilience parameter*.

We call such a random variable X, with cdf (7), an *exponentiated discrete Weibull distribution* with parameters $0 < p < 1$, $\alpha > 0$ and $\gamma > 0$ and denote it by $EDW(p,\alpha,\gamma)$.

It is evident that when $\gamma > 0$ is an integer value, the cdf given by (7) agrees with the cdf of the maximum of γ independent and identical $DW(p,\alpha)$ random variables.

2.1. Probability mass, survival and hazard rate functions

The corresponding pmf of a random variable X following an $EDW(p,\alpha,\gamma)$ distribution for $x \in \mathbb{N}_0$ is given by

$$p_x = P(X = x) = f(x;p,\alpha,\gamma) = \{1 - p^{(x+1)^\alpha}\}^\gamma - \{1 - p^{x^\alpha}\}^\gamma \tag{8}$$

$$= \sum_{j=1}^{\infty} (-1)^{j+1} \binom{\gamma}{j} \{p^{jx^\alpha} - p^{j(x+1)^\alpha}\}, \tag{9}$$

where $\binom{\gamma}{j} = \frac{\Gamma(\gamma+1)}{\Gamma(\gamma+1-j)j!}$. For integer $\gamma > 0$, the sum in Eq. (9) stops at γ.

Nekoukhou et al. (2013b) indicated that $\sum_{j=1}^{\infty}(-1)^{j+1}\binom{\gamma}{j} = 1$. Hence, if $0 < \gamma < 1$ the pmf (9) can be viewed as an infinite mixture of $DW(p^j,\alpha)$ distributions, $j = 1,2,\ldots$

It is interesting to note that the EDW distribution with pmf (8) or (9) may also be viewed as a discrete analog of the exponentiated Weibull (EW) distribution of Mudholkar and Srivastava (1993) via Eq. (4) and doing reparametrization $0 < e^{-\beta^\alpha} = p < 1$ in the structure of EW distribution.

Nassar and Eissa (2003) obtained expressions for the mode of the EW pdf. They stated that EW distribution is monotone decreasing for $\alpha\gamma \leq 1$ and for $\alpha\gamma > 1$, it is unimodal. Naturally, it follows that $EDW(p,\alpha,\gamma)$ distributions are also unimodal for all values of parameters. Figure 1 illustrates the pmf of an $EDW(p,\alpha,\gamma)$ distribution for different values of parameters.

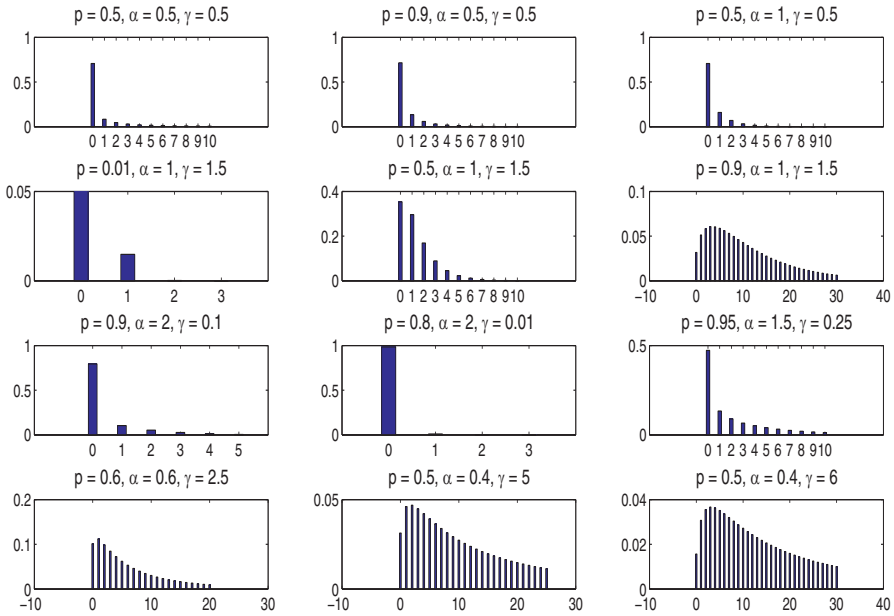


Figure 1: *Illustrations of the pmf of* $EDW(p,\alpha,\gamma)$ *for possible values of* p, α *and* γ.

The survival and hazard rate functions of $EDW(p,\alpha,\gamma)$ distribution are given by

$$S(x;p,\alpha,\gamma) = 1-\{1-p^{([x]+1)^{\alpha}}\}^{\gamma}, \quad x \geq 0 \tag{10}$$

and

$$h(x;p,\alpha,\gamma) = \frac{\{1-p^{(x+1)^{\alpha}}\}^{\gamma}-\{1-p^{x^{\alpha}}\}^{\gamma}}{1-\{1-p^{(x+1)^{\alpha}}\}^{\gamma}}, \quad x \in \mathbb{N}_0, \tag{11}$$

respectively.

Discrete hazard rates arise in several common situations in reliability theory where clock time is not the best scale on which to describe lifetime. For example, in weapons reliability, the number of rounds fired until failure is more important than age in failure. This is the case also when a piece of equipment operates in cycles and the observation is the number of cycles successfully completed prior to failure. In other situations a device is monitored only once per time period and the observation then is the number of time periods successfully completed prior to the failure of the device (cf. Shaked et al., 1995).

Figure 2 illustrates the hazard rate function of $EDW(p,\alpha,\gamma)$ distribution for different values of p, α and γ. As we see from the figure, a characteristic of the EDW distribution is that its hazard rate function can be decreasing, increasing, bathtub-shaped, and upside-down bathtub depending on its parameters values. Hence, EDW distributions are more flexible than other discrete distributions such as the geometric, DGE, DGE_2 and DBE distributions, whose hazard rate functions are constant and monotone.

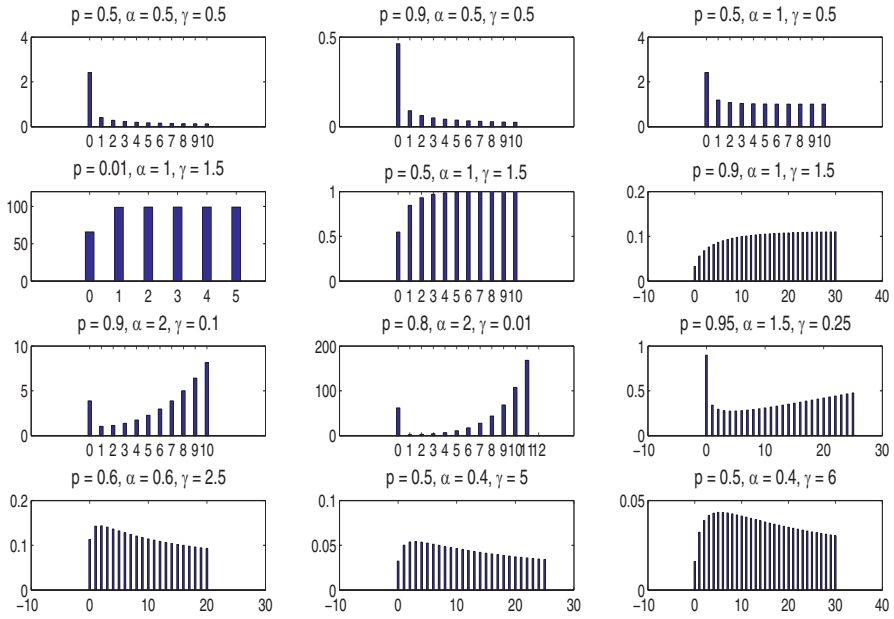


Figure 2: *Illustrations of the hazard rate function of* $EDW(p,\alpha,\gamma)$ *for possible values of p,* α *and* γ.

2.2. Special sub-models

Some special discrete distributions are achieved from EDW distribution as follows:

(1) Discrete Weibull distribution of Nakagawa and Osaki (1975), with pmf

$$p_x = (1-p^{(x+1)^\alpha}) - (1-p^{x^\alpha}), \tag{12}$$

is obtained when $\gamma = 1$. If, in addition, $\alpha = 1$, the geometric distribution is achieved. The discrete Weibull distribution is used for estimation of replicative senescence via population dynamics models (Wein and Wu, 2001), stress-strength reliability (Roy, 2002), evaluation of reliability of complex systems (Roy, 2002), wafer probe operation in semiconductor manufacturing (e.g., Wang, 2009), minimal availability variation design of repairable systems (e.g., Wang et al., 2010) and microbial counts in water (Englehardt and Li, 2011). Since the EDW distribution is an extension of DW distribution, one may expect from EDW model to be more flexible in such application areas.

(2) If $\alpha = 1$, then the discrete generalized exponential distribution of a second type ($DGE_2(\gamma,p)$) of Nekoukhou et al. (2013b) with pmf given by Eq. (2) is obtained. If, in addition, $\gamma = 1$, the geometric distribution will be obtained again from a different point of view.

(3) If $\alpha = 2$, then the pmf of $EDW(p,\alpha,\gamma)$ distribution reduces to

$$p_x = f(x;p,\gamma) = \{1-p^{(x+1)^2}\}^{\gamma} - \{1-p^{x^2}\}^{\gamma} \tag{13}$$

$$= \sum_{j=1}^{\infty}(-1)^{j+1}\binom{\gamma}{j}p^{jx^2}(1-p^{j(2x+1)}), \tag{14}$$

which defines a generalized discrete Rayleigh distribution $GDR(\gamma,p)$ for the first time in the literature. Moreover, for $\gamma = 1$ in Eq. (13) the discrete Rayleigh (DR) distribution of Roy (2004) is obtained.

2.3. Quantiles, mean and variance

The m-th quantile of an EDW distribution is obtained by solving the equation

$$F(q_m;p,\alpha,\gamma) = m,$$

where $F(.)$ is the cdf of an $EDW(p,\alpha,\gamma)$ distribution and q_m denotes the corresponding quantile function which is given by

$$q_m = \left\{\frac{\log(1-m)^{1/\gamma}}{\log p}\right\}^{1/\alpha} - 1. \tag{15}$$

Particularly, the median is immediately achieved by setting $m = 0.5$ in the above equation.

The mean and variance of a random variable X following an $EDW(p,\alpha,\gamma)$ distribution are given, respectively, by

$$E(X) = \sum_{j=1}^{\infty}\binom{\gamma}{j}(-1)^{j+1}\frac{p^{\alpha j}}{1-p^{\alpha j}} \tag{16}$$

and

$$Var(X) = 2\sum_{j=1}^{\infty}\binom{\gamma}{j}(-1)^{j+1}\frac{p^{\alpha j}}{(1-p^{\alpha j})^2} + E(X) - \{E(X)\}^2. \tag{17}$$

Remark 2.1 For an integer value of $\gamma > 0$, $\sum_{j=1}^{\infty}$ should be replaced by $\sum_{j=1}^{\gamma}$ in the above equations.

Remark 2.2 For $\alpha = 1$, Eq. (16) reduces to

$$E(X) = \sum_{j=1}^{\infty} \binom{\gamma}{j} (-1)^{j+1} \frac{p^j}{1-p^j}, \tag{18}$$

which is the mean of the $DGE_2(\gamma, p)$ distribution obtained by Nekoukhou et al. (2013b). In addition, in this case, it is easy to show that the variance of an EDW distribution reduces to the variance of a DGE_2 distribution.

The mean and variance of an $EDW(p, \alpha, \gamma)$ distribution for different values of p, α and γ, using Eq.'s (16) and (17), are calculated in Table 1 below. It appears that depending on the values of the parameters, the mean of the distribution can be smaller or greater than its variance. Hence, EDW models are appropriate for modeling both over and under dispersed data since, in these models, the variance can be larger or smaller than the mean which is not the case with some standard classical discrete distributions.

Table 1: *Mean (Variance) of $EDW(p, \alpha, \gamma)$ for different values of p, α and γ.*

		$\gamma = 0.50$	
α/p	0.25	0.5	0.75
0.50	0.3938 (2.6970)	2.0105 (46.7602)	12.7288 (1517.5424)
0.75	0.2253 (0.5441)	0.8322 (3.9644)	3.2749 (44.2522)
1.00	0.1761 (0.2573)	0.5546 (1.2777)	1.7437 (8.2081)
2.00	0.1359 (0.1213)	0.3256 (0.2856)	0.7168 (0.7357)
3.50	0.1339 (0.1160)	0.2930 (0.2075)	0.5194 (0.2885)
		$\gamma = 1.00$	
α/p	0.25	0.5	0.75
0.50	0.7598 (5.0596)	3.7882 (85.6990)	23.5837 (2743.1543)
0.75	0.4296 (0.9364)	1.5272 (6.6530)	5.8068 (71.5589)
1.00	0.3333 (0.4444)	1.0000 (1.9999)	2.9999 (11.9999)
2.00	0.2539 (0.1972)	0.5644 (0.3787)	1.1522 (0.8241)
3.50	0.2500 (0.1875)	0.5003 (0.2507)	0.7885 (0.2439)
		$\gamma = 3.00$	
α/p	0.25	0.5	0.75
0.50	2.0009 (12.1856)	9.3360 (197.0330)	56.1430 (6147.2083)
0.75	1.0828 (1.8981)	3.4610 (11.9090)	12.2932 (123.5352)
1.00	0.8158 (0.7907)	2.1428 (2.9251)	5.8725 (16.5319)
2.00	0.5898 (0.2653)	1.0569 (0.3155)	1.9058 (0.6676)
3.50	0.5781 (0.2438)	0.8761 (0.1108)	1.0957 (0.1178)

Remark 2.3 Remember that a random variable X with cdf G is stochastically smaller than Y with cdf F, denoted by $X \leq_{st} Y$, if for all x, $G(x) \geq F(x)$. This is the most basic and oldest stochastic order in Probability and Statistics. In this case, if G is simpler than F, $G(x)$ may provide a useful lower bound for $F(x)$ (see, e.g., Shaked and Shanthikumar

(2007) for more details). Now, let G and F denote the cdfs of the DW and EDW distributions which are defined via Eq.'s (6) and (7), respectively. It is obvious that for $\gamma > 1$, we have $X \leq_{st} Y$ because $[G(x)]^{\gamma} \leq G(x)$ and if $0 < \gamma < 1$, it follows that $X \geq_{st} Y$. Hence, For $\gamma \geq 1$ it follows that $E(X) \leq E(Y)$ and corresponding result holds if X is stochastically larger than Y. One can consider the results of Table 1 again.

2.4. Infinite divisibility

The researchers here make the following note in regards to the famous structural property of infinite divisibility of the distribution in question. Such a characteristic has a close relation to the Central Limit Theorem and waiting time distributions. Thus, it is a desirable question in modeling to know whether a given distribution is infinitely divisible or not. To settle this question, we recall that according to Steutel and van Harn (2004, pp. 56), if p_x, $x \in \mathbb{N}_0$, is infinitely divisible, then $p_x \leq e^{-1}$ for all $x \in \mathbb{N}$. However, e.g., in an $EDW(0.9,3,1)$ distribution we see that $p_2 = 0.372 > e^{-1} = 0.367$. Therefore, in general, $EDW(p,\alpha,\gamma)$ distributions are not infinitely divisible. In addition, since the classes of self-decomposable and stable distributions, in their discrete concepts, are subclasses of infinitely divisible distributions, we conclude that an EDW distribution can be neither self-decomposable nor stable in general.

2.5. Order statistics

Order statistics are among the most fundamental tools in non-parametric statistics and inference. They enter the problems of estimation and hypothesis testing in a variety of ways. The aim of the present section is to establish some general relations regarding the EDW distributions. More precisely, let $F_i(x;p,\alpha,\gamma)$ and $f_i(x;p,\alpha,\gamma)$ be the cdf and pmf of the i-th order statistic of a random sample of size n from $EDW(p,\alpha,\gamma)$ distribution.

Since,

$$F_i(x;p,\alpha,\gamma) = \sum_{k=i}^{n} \binom{n}{k} [F(x;p,\alpha,\gamma)]^k [1-F(x;p,\alpha,\gamma)]^{n-k}, \tag{19}$$

using the binomial expansion for $[1-F(x;p,\alpha,\gamma)]^{n-k}$, we obtain the following result:

$$\begin{aligned} F_i(x;p,\alpha,\gamma) &= \sum_{k=i}^{n}\sum_{j=0}^{n-k} \binom{n}{k}\binom{n-k}{j}(-1)^j [F(x;p,\alpha,\gamma)]^{k+j} \\ &= \sum_{k=i}^{n}\sum_{j=0}^{n-k} \binom{n}{k}\binom{n-k}{j}(-1)^j [\{1-p^{([x]+1)^{\alpha}}\}^{\gamma}]^{k+j} \\ &= \sum_{k=i}^{n}\sum_{j=0}^{n-k} \binom{n}{k}\binom{n-k}{j}(-1)^j F_{EDW}(x;p,\alpha,\gamma(k+j)), \end{aligned} \tag{20}$$

where F_{EDW} denotes the cdf of an EDW distribution. The corresponding pmf of the i-th order statistic, $f_i(x;p,\alpha,\gamma) = F_i(x;p,\alpha,\gamma) - F_i(x-1;p,\alpha,\gamma)$ for an integer value of x, then is given by

$$f_i(x;p,\alpha,\gamma) = \sum_{k=i}^{n}\sum_{j=0}^{n-k}\binom{n}{k}\binom{n-k}{j}(-1)^j f_{EDW}(x;p,\alpha,\gamma(k+j)), \tag{21}$$

where f_{EDW} denotes the pmf of an EDW distribution.

Remark 2.4 In view of the fact that $f_i(x;p,\alpha,\gamma)$ is a linear combination of a finite number of $EDW(p,\alpha,\gamma(k+j))$ distributions, we may obtain some properties of order statistics, such as their moments, from the corresponding EDW distribution. For example, the mean of the i-th order statistic is given by

$$\mu_{i:n} = \sum_{t=1}^{\infty}\sum_{k=i}^{n}\sum_{j=0}^{n-k}\binom{n}{k}\binom{n-k}{j}\binom{\gamma(k+j)}{t}(-1)^{j+t+1}\frac{p^{\alpha t}}{1-p^{\alpha t}}. \tag{22}$$

3. Estimation

To apply the method of maximum likelihood for estimating the parameter vector $\boldsymbol{\theta} = (p,\alpha,\gamma)^{\mathsf{T}}$ of EDW distribution, assume that $x = (x_1,x_2,...,x_n)^{\mathsf{T}}$ is a random sample of size n from an $EDW(p,\alpha,\gamma)$ distribution. The log-likelihood function becomes

$$\ell = \sum_{i=1}^{n}\log[(1-p^{(x_i+1)^\alpha})^\gamma - (1-p^{x_i^\alpha})^\gamma]. \tag{23}$$

Hence, the likelihood equations are

$$\frac{\partial \ell}{\partial p} = \sum_{i=1}^{n}\frac{v_{\alpha,\gamma}(x_i+1) - v_{\alpha,\gamma}(x_i)}{m_{\alpha,\gamma}(x_i)}, \tag{24}$$

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{n}\frac{\gamma\log p[u_{\alpha,\gamma}(x_i)\log x_i - u_{\alpha,\gamma}(x_i+1)\log(x_i+1)]}{m_{\alpha,\gamma}(x_i)} \tag{25}$$

and

$$\frac{\partial \ell}{\partial \gamma} = \sum_{i=1}^{n}\frac{\gamma[u_{\alpha,\gamma}(x_i) - u_{\alpha,\gamma}(x_i+1)]}{pm_{\alpha,\gamma}(x_i)}, \tag{26}$$

where

$$m_{\alpha,\gamma}(x) = \{1-p^{(x+1)^\alpha}\}^\gamma - \{1-p^{x^\alpha}\}^\gamma,$$

$$v_{\alpha,\gamma}(x) = (1-p^{x^\alpha})^\gamma \log(1-p^{x^\alpha})$$

and

$$u_{\alpha,\gamma}(x) = (1-p^{x^\alpha})^{\gamma-1} p^{x^\alpha} x^\alpha.$$

The solutions of likelihood equations (24)-(26) provide the maximum likelihood estimators (MLEs) of $\boldsymbol{\theta} = (p,\alpha,\gamma)^\mathsf{T}$, say $\hat{\boldsymbol{\theta}} = (\hat{p},\hat{\alpha},\hat{\gamma})^\mathsf{T}$, which can be obtained by a numerical method such as the three variable Newton-Raphson type procedure.

For interval estimation and hypothesis tests on the model parameters, we require the information matrix. The 3×3 observed information matrix is

$$I_n(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} -\frac{\partial^2\ell}{\partial p^2} & -\frac{\partial^2\ell}{\partial p\partial\alpha} & -\frac{\partial^2\ell}{\partial p\partial\gamma} \\ -\frac{\partial^2\ell}{\partial\alpha\partial p} & -\frac{\partial^2\ell}{\partial\alpha^2} & -\frac{\partial^2\ell}{\partial\alpha\partial\gamma} \\ -\frac{\partial^2\ell}{\partial\gamma\partial p} & -\frac{\partial^2\ell}{\partial\gamma\partial\alpha} & -\frac{\partial^2\ell}{\partial\gamma^2} \end{bmatrix}, \tag{27}$$

whose elements are given in the Appendix.

One can show that the EDW family satisfies the regularity conditions which are fulfilled for parameters in the interior of the parameter space but not on the boundary (see, e.g., Cox and Hinkley, 1974). Hence, the MLE vector $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically normal. That is, $I_n^{\frac{1}{2}}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ converges in distribution to trivariate normal with the (vector) mean zero and the identity covariance matrix.

One can use the normal distribution of $\hat{\boldsymbol{\theta}}$ to construct approximate confidence regions for some parameters. Indeed, an asymptotic $100(1-\xi)$ confidence interval for each parameter θ_i, is given by

$$(\hat{\theta}_i - z_{\xi/2}\sqrt{\hat{J}_{ii}}, \hat{\theta}_i + z_{\xi/2}\sqrt{\hat{J}_{ii}}), \qquad i=1,2,3,$$

where $\hat{J}_{ii}$ denotes the (i,i) diagonal element of $I_n^{-1}(\hat{\boldsymbol{\theta}})$ and $z_{\xi/2}$ is the $(1-\xi/2)$-th quantile of the standard normal distribution.

3.1. Simulation study

Let X be a random variable that follows an EW distribution with cdf

$$F(x;\alpha,\beta,\gamma) = \{1 - e^{-(\beta x)^{\alpha}}\}^{\gamma}, \quad x > 0,$$

where $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ (two shapes and one scale) are the model parameters. It is easy to show that $[X]$ has an $EDW(p,\alpha,\gamma)$ distribution in which $0 < p = e^{-\beta^{\alpha}} < 1$. Therefore, we can simulate an EDW random variable from the corresponding continuous EW distribution. Table 2 below presents the maximum likelihood estimates of $\boldsymbol{\theta} = (p,\alpha,\gamma)^{\mathsf{T}}$ of an EDW distribution and also contains their standard errors for different values of n as a simulation study. Standard errors are attained by means of the asymptotic covariance matrix of the MLEs of EDW parameters when the Newton-Raphson procedure converges in, e.g., MATLAB software.

Table 2: *MLEs of EDW parameters for different values of n.*

n	$\hat{\alpha}(\hat{SE}(\hat{\alpha}))$	$\hat{\gamma}(\hat{SE}(\hat{\gamma}))$	$\hat{p}(\hat{SE}(\hat{p}))$	$\hat{\alpha}(\hat{SE}(\hat{\alpha}))$	$\hat{\gamma}(\hat{SE}(\hat{\gamma}))$	$\hat{p}(\hat{SE}(\hat{p}))$
(α,γ)		$(0.5,0.75)$			$(0.75,0.5)$	
p		0.25			0.75	
40	0.525(0.563)	1.113(2.987)	0.221(0.657)	0.822(0.427)	0.591(0.431)	0.784(0.322)
100	0.492(0.477)	0.984(2.764)	0.213(0.527)	0.812(0.396)	0.442(0.393)	0.753(0.297)
200	0.511(0.323)	0.788(1.347)	0.288(0.410)	0.730(0.268)	0.551(0.379)	0.719(0.237)
500	0.501(0.242)	0.792(1.099)	0.217(0.264)	0.745(0.158)	0.526(0.204)	0.751(0.129)
1000	0.568(0.175)	0.799(0.675)	0.257(0.185)	0.743(0.108)	0.534(0.144)	0.745(0.090)
(α,γ)		$(2,3)$			$(3,2)$	
p		0.5			0.9	
40	2.197(1.083)	2.536(2.931)	0.542(0.410)	2.857(1.453)	1.903(1.879)	0.927(0.194)
100	2.077(0.951)	2.656(2.652)	0.564(0.349)	2.912(1.197)	1.872(1.542)	0.897(0.156)
200	1.904(0.663)	2.941(2.352)	0.494(0.289)	2.937(0.818)	2.022(1.163)	0.888(0.113)
500	1.915(0.465)	3.290(1.880)	0.462(0.187)	3.153(0.605)	1.980(0.781)	0.914(0.065)
1000	2.004(0.321)	2.950(1.068)	0.511(0.124)	2.918(0.306)	1.981(0.427)	0.895(0.041)
(α,γ)		$(1,1)$			$(1.5,0.5)$	
p		0.5			0.95	
40	1.202(0.996)	1.183(1.210)	0.717(0.712)	1.139(0.611)	0.733(0.693)	0.896(0.206)
100	1.278(0.723)	0.864(1.005)	0.601(0.416)	1.257(0.436)	0.808(0.494)	0.867(0.150)
200	0.933(0.363)	0.974(0.850)	0.488(0.293)	1.443(0.393)	0.471(0.198)	0.947(0.060)
500	0.982(0.230)	1.043(0.553)	0.484(0.177)	1.521(0.233)	0.522(0.125)	0.957(0.023)
1000	1.058(0.172)	0.909(0.318)	0.542(0.122)	1.507(0.177)	0.481(0.087)	0.955(0.012)

3.2. Stress-strength parameter

The stress-strength parameter $R = P(X > Y)$ is a measure of component reliability and its estimation problem when X and Y are independent and follow a specified common distribution has been discussed widely in the literature. Suppose that the random variable X is the strength of a component which is subjected to a random stress Y. Estimation of R when X and Y are independent and identically distributed following a well-known distribution has been considered in the literature. Many applications of the stress-strength model, for its own nature, are related to engineering or military problems. There are also natural applications in Medicine or Psychology, which involve the comparison of two random variables, representing for example the effect of a specific drug or treatment administered to two groups, control and test. Almost all of these studies consider continuous distributions for X and Y, because many practical applications of the stress-strength model in engineering fields presuppose continuous quantitative data. A complete review is available in Kotz et al. (2003). However, in this regard, a relatively small amount of work is devoted to discrete or categorical data. Data may be discrete by nature. For example, the stress pattern in a step-stress accelerated life test can be treated as a discrete random variable of which the possible values can be obtained from all stress levels, and the corresponding probabilities can be obtained from the acting times of each stress levels. Moreover, the stress state of a component can be categorized based on the characteristic of external loads. For instance, the stress state of a mechanical component can be simply classified as state 1, state 2 and state 3, which correspond to low load, moderate load and heavy load, respectively. More generally, according to the change of external loads, the stress of a component can be categorized into arbitrary finite state: state 1, state 2, ..., state m.

The stress-strength parameter, in discrete case, is defined as

$$R = P(X > Y) = \sum_{x=0}^{\infty} f_X(x) F_Y(x),$$

where f_X and F_Y denote the pmf and cdf of the independent discrete random variables X and Y, respectively. Now, let $X \sim EDW(\boldsymbol{\theta_1})$ and $Y \sim EDW(\boldsymbol{\theta_2})$, where $\boldsymbol{\theta_1} = (p_1, \alpha_1, \gamma_1)^{\mathsf{T}}$ and $\boldsymbol{\theta_2} = (p_2, \alpha_2, \gamma_2)^{\mathsf{T}}$. Using Equations (7) and (8), we obtain

$$R = \sum_{x=0}^{\infty} [\{1 - p_1^{(x+1)^{\alpha_1}}\}^{\gamma_1} - \{1 - p_1^{x^{\alpha_1}}\}^{\gamma_1}] \{1 - p_2^{(x+1)^{\alpha_2}}\}^{\gamma_2}.$$

Using the binomial expansion, it is easy to show that

$$R = \sum_{j=1}^{\infty} \sum_{t=1}^{\infty} \sum_{x=0}^{\infty} (-1)^{j+t+1} \binom{\gamma_1}{j} \binom{\gamma_2}{t} p_2^{t(x+1)^{\alpha_2}} \{p_1^{jx^{\alpha_1}} - p_1^{j(x+1)^{\alpha_1}}\}. \tag{28}$$

Now, assume that $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$ are independent observations from $X \sim EDW(\boldsymbol{\theta_1})$ and $Y \sim EDW(\boldsymbol{\theta_2})$, respectively. The total likelihood function is $\ell_R(\boldsymbol{\theta}^*) = \ell_n(\boldsymbol{\theta_1})\ell_m(\boldsymbol{\theta_2})$, where $\boldsymbol{\theta}^* = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$. The score vector is given by

$$U_R(\boldsymbol{\theta}^*) = (\partial\ell_R/\partial p_1, \partial\ell_R/\partial\alpha_1, \partial\ell_R/\partial\gamma_1, \partial\ell_R/\partial p_2, \partial\ell_R/\partial\alpha_2, \partial\ell_R/\partial\gamma_2),$$

and the MLE of $\boldsymbol{\theta}^*$, say $\hat{\boldsymbol{\theta}}^*$, may be attained from the nonlinear equation $U_R(\hat{\boldsymbol{\theta}}^*) = \mathbf{0}$. Thus, by inserting the MLEs in equation (28) the stress-strength parameter R will be estimated.

4. Application

In this section, the EDW model will be examined for a real data set which is given by Karlis and Xekalaki (2001) on the numbers of fires in Greece for the period from 1 July 1998 to 31 August of the same year. This data set consists of 123 observation and are presented in Table 3. Only fires in forest districts are considered. Bakouch et al. (2014) considered these data to indicate the potentiality of discrete Lindley (DL) distribution in data modeling and compared it with Poisson, geometric and discrete gamma (DG) distributions. The pmf of the DG distribution, which has been used first by Yang (1994) and recently considered by Chakraborty and Chakravarty (2012), for $x \in \mathbb{N}_0$, is given by

$$p_x = \frac{\gamma(\alpha, \beta(x+1)) - \gamma(\alpha, \beta x)}{\Gamma(\alpha)}, \quad \alpha > 0,\ \beta > 0,$$

where $\gamma(a,x) = \int_0^x t^{a-1}e^{-t}dt$ denotes the incomplete gamma function. Additionally, the pmf of the DL distribution for $x \in \mathbb{N}_0$ is given by

$$p_x = \frac{p^x}{1+\theta}\{\theta(1-2p) + (1-p)(1+\theta x)\}, \quad 0 < p < 1,\ \theta > 0.$$

Here, we compare the EDW and GDR models with these discrete distributions. In addition, because of the over dispersion phenomena in the data set, $\bar{x} = 5.3984$ and $s^2 = 30.0449$, the negative binomial (NB) distribution is also compared with the others. Maximum likelihood method is used to obtain the estimates of the parameters of the proposed new distributions (EDW and GDR). Comparing the EDW model with its rival models is performed by using the Akaike information criterion (AIC) and Kolmogrov-Smirnov (K-S) test statistic. Table 4 indicates the MLEs, AICs and the values of the K-S test statistics determined by the fitted models. The MLEs and K-S test statistic values of the DL and DG distributions, given in this table, are directly reported from Table 7 of Bakouch et al. (2014).

Table 3: *Numbers of fires in Greece.*

Numbers	0	1	2	3	4	5	6	7	8	9	10	11	12	15	16	20	43
Frequency	16	13	14	9	11	13	8	4	9	6	3	4	6	4	1	1	1

Table 4: *Summary.*

Models	MLEs	AIC	K-S statistic
EDW	$(\hat{\alpha},\hat{\gamma},\hat{p}) = (1.0809, 1.0923, 0.8599)$	685.5859	0.1254
GDR	$(\hat{\gamma},\hat{p}) = (0.3934, 0.9924)$	694.6178	0.1467
DGE_2	$(\hat{\gamma},\hat{p}) = (1.2548, 0.8225)$	683.7049	0.1301
NB	$(\hat{r},\hat{p}) = (1.3360, 0.1984)$	683.2989	0.3350
DL	$(\hat{\theta},\hat{p}) = (0.3090, 0.7343)$	685.8067	0.1122
DG	$(\hat{\alpha},\hat{\beta}) = (0.7525, 0.1543)$	749.7162	0.2683

According to the values of the K-S test statistics and AICs in Table 4, it seems that EDW model gives a satisfactory fit to this real data set.

To construct approximate confidence intervals for the parameters of EDW model and also for evaluating accuracy of the estimated parameters, we use the corresponding estimated standard errors. For instance, 95% asymptotic confidence intervals for EDW parameters are obtained as $\alpha \in (1.081 \mp 0.4531)$, $\gamma \in (1.0923 \mp 0.8554)$ and $p \in (0.8599 \mp 0.1895)$.

5. Conclusions and comments

In this paper, a new three-parameter generalization of the discrete Weibull distribution is proposed, so-called exponentiated discrete Weibull (EDW) distribution which is, indeed, a member of resilience parameter family of distributions. The hazard rate function of the new model can be increasing, decreasing, upside-down bathtub and also bathtub-shaped and hence presents a very flexible behavior. Fitting the EDW model to a real data set indicates the flexibility and capacity of the proposed distribution in data modeling. In addition, a special sub-model of EDW distribution, i.e., the generalized discrete Rayleigh distribution is introduced for the first time in the literature.

Appendix

The elements of the 3×3 information matrix in Eq. (27) are given by

$$\frac{\partial^2 \ell}{\partial p^2} = \sum_{i=1}^{n} \{ \frac{\gamma}{p^2 m_{\alpha,\gamma}(x_i)} \{ \ (\gamma-1) u_{\alpha,\gamma-1}(x_i+1) p^{(x_i+1)^\alpha} (x_i+1)^\alpha$$
$$- u_{\alpha,\gamma}(x_i+1)[(x_i+1)^\alpha - 1]$$
$$- (\gamma-1) u_{\alpha,\gamma-1}(x_i) p^{x_i^\alpha} x_i^\alpha + u_{\alpha,\gamma}(x_i)(x_i^\alpha - 1) \ \} - \frac{\gamma^2 [u_{\alpha,\gamma}(x_i) - u_{\alpha,\gamma}(x_i+1)]^2}{p^2 m_{\alpha,\gamma}^2(x_i)} \},$$

$$\frac{\partial^2 \ell}{\partial \alpha^2} = \sum_{i=1}^{n} \{ \frac{\gamma}{m_{\alpha,\gamma}(x_i)} \{ \ (\gamma-1) u_{\alpha,\gamma-1}(x_i+1) p^{(x_i+1)^\alpha} (x_i+1)^\alpha \log^2(x_i+1) \log^2 p$$
$$- (\gamma-1) u_{\alpha,\gamma-1}(x_i) p^{x_i^\alpha} x_i^\alpha \log^2 x_i \log^2 p + u_{\alpha,\gamma}(x_i) \log^2 x_i \log p [x_i^\alpha \log p + 1]$$
$$- u_{\alpha,\gamma}(x_i+1) \log^2(x_i+1) \log p [(x_i+1)^\alpha \log p + 1] \ \}$$
$$- \frac{\{\gamma \log p [u_{\alpha,\gamma}(x_i) \log x_i - u_{\alpha,\gamma}(x_i+1) \log(x_i+1)]\}^2}{m_{\alpha,\gamma}^2(x_i)} \},$$

$$\frac{\partial^2 \ell}{\partial \gamma^2} = \sum_{i=1}^{n} \{ \frac{v_{\alpha,\gamma}(x_i+1) \log(1 - p^{(x_i+1)^\alpha}) - v_{\alpha,\gamma}(x_i) \log(1 - p^{x_i^\alpha})}{m_{\alpha,\gamma}(x_i)}$$
$$- \frac{\{v_{\alpha,\gamma}(x_i+1) - v_{\alpha,\gamma}(x_i)\}^2}{m_{\alpha,\gamma}^2(x_i)} \},$$

$$\frac{\partial^2 \ell}{\partial p \partial \alpha} = \sum_{i=1}^{n} \{ \frac{\gamma}{p m_{\alpha,\gamma}(x_i)} \{ \ (\gamma-1) u_{\alpha,\gamma-1}(x_i+1) p^{(x_i+1)^\alpha} (x_i+1)^\alpha \log(x_i+1) \log p$$
$$- (\gamma-1) u_{\alpha,\gamma-1}(x_i) p^{x_i^\alpha} x_i^\alpha \log x_i \log p + u_{\alpha,\gamma}(x_i) \log x_i [x_i^\alpha \log p + 1]$$
$$- u_{\alpha,\gamma}(x_i+1) \log(x_i+1) [(x_i+1)^\alpha \log p + 1] \ \}$$
$$- \frac{\gamma^2 \log p}{p m_{\alpha,\gamma}^2(x_i)} \{ [u_{\alpha,\gamma}(x_i) \log x_i - u_{\alpha,\gamma}(x_i) \log(x_i+1)][u_{\alpha,\gamma}(x_i) - u_{\alpha,\gamma}(x_i+1)] \} \},$$

$$\frac{\partial^2 \ell}{\partial p \partial \gamma} = \sum_{i=1}^{n} \{ \frac{u_{\alpha,\gamma}(x_i)[\gamma \log(1 - p^{x_i^\alpha}) + 1] - u_{\alpha,\gamma}(x_i+1)[\gamma \log(1 - p^{(x_i+1)^\alpha}) + 1]}{p m_{\alpha,\gamma}(x_i)}$$
$$- \frac{\gamma [u_{\alpha,\gamma}(x_i) - u_{\alpha,\gamma}(x_i+1)][v_{\alpha,\gamma}(x_i+1) - v_{\alpha,\gamma}(x_i)]}{p m_{\alpha,\gamma}^2(x_i)} \}$$

and

$$\frac{\partial^2 \ell}{\partial \alpha \partial \gamma} = \sum_{i=1}^{n} \{ \frac{\log p}{m_{\alpha,\gamma}(x_i)} \{ u_{\alpha,\gamma}(x_i) \log x_i [\gamma \log(1-p^{x_i^\alpha}) + 1] - u_{\alpha,\gamma}(x_i+1) \log(x_i+1) [\gamma \log(1-p^{(x_i+1)^\alpha}) + 1] \} - \frac{\gamma \log p \{ u_{\alpha,\gamma}(x_i) \log x_i - u_{\alpha,\gamma}(x_i+1) \log(x_i+1) \} \{ v_{\alpha,\gamma}(x_i+1) - v_{\alpha,\gamma}(x_i) \}}{m^2_{\alpha,\gamma}(x_i)} \}.$$

Acknowledgments

The authors sincerely thank two anonymous referees for their valuable comments and also special thanks to the Editor-in-Chief for all his considerations in this paper.

References

Bakouch, H.S., Jazi, M.A. and Nadarajah, S. (2014). A new discrete distribution. *Statistics*, 48, 200–240.

Barbiero, A. (2014). An alternative discrete skew Laplace distribution. *Statistical Methodology,* 16, 47–67.

Chakraborty, S. and Gupta, R.D. (2012). Exponentiated geometric distribution: another generalization of geometric distribution. *Communications in Statistics-Theory and Methods*, DOI: 10.1080/03610926.2012.763090.

Chakraborty, S. and Chakravarty, D. (2012). Discrete gamma distributions: properties and parameter estimation. *Communications in Statistics-Theory and Methods*, 41, 3301–3324.

Chakraborty, S. (2013). A new discrete distribution related to generalized gamma distribution. Published on line August, 2013 in *Communications in Statistics-Theory and Methods*.

Chakraborty, S. and Chakravarty, D. (2013). A new discrete probability distribution with integer support on $(-\infty, \infty)$. Published on line in July, 2013, in *Communications in Statistics-Theory and Methods*.

Chakraborty, S. and Chakravarty, D. (2014). *A Discrete Gumbel Distribution*. arXiv:1410.7568.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

Dasgupta, R. (1993). Cauchy equation on discrete domain and some characterization theorems. *Theory of Probability & Its Applications*, 38, 520–524.

Englehardt, J.D. and Li, R.C. (2011). The discrete Weibull distribution: an alternative for correlated counts with confirmation for microbial counts in water. *Risk Analysis*, 31, 370–381.

Gómez-Déniz, E. (2010). Another generalization of the geometric distribution. *Test*, 19, 399–415.

Gómez-Déniz E. and Calderin-Ojeda, E. (2011). The discrete Lindley distribution: properties and applications. *Journal of Statistical Computation and Simulation*, 81, 1405–1416.

Gupta, R.D. and Kundu, D. (1999). Generalized exponential distributions. *Australian & New Zealand Journal of Statistics*, 41, 173–188.

Gupta, R.D. and Kundu, D. (2001). Exponentiated exponential family: an alternative to gamma and Weibull distributions. *Biometrical Journal*, 43, 117–130.

Gupta, R.D. and Kundu, D. (2007). Generalized exponential distribution: existing results and some recent developments. *Journal of Statistical Planning and Inference*, 137, 3537–3547.

Hussain, T. and Ahmad, M. (2014). Discrete inverse Rayleigh distribution. *Pakistan Journal of Statistics*, 30, 203–222.

Inusah, S. and Kozubowski, T.J. (2006). A discrete analogue of the Laplace distribution. *Journal of Statistical Planning and Inference*, 136, 1090–1102.

Jazi, M.A., Lai, C.D. and Alamatsaz, M.H. (2010). A discrete inverse Weibull distribution and estimation of its parameters. *Statistical Methodology*, 7, 121–132.

Jiang, R. (2010). Discrete competing risk model with application to modeling bus-motor failure data. *Reliability Engineering & System Safety*, 95, 981–988.

Karlis, D. and Xekalaki, E. (2001). On some discrete valued time series models based on mixtures and thinning, in *Proceedings of the Fifth Hellenic-European Conference on Computer Mathematics and Its Applications*, E.A. Lipitakis, ed., 872–877.

Kemp, A.W. (1997). Characterizations of discrete normal distribution. *Journal of Statistical Planning and Inference*, 63, 223–229.

Kemp, A.W. (2008). The discrete half-normal distribution. *International Conference on Mathematical and Statistical Modeling in honor to Enrique Castillo*, June 28-30.

Kotz, S., Lumelskii, M. and Pensky, M. (2003). *The Stress-Strength Model and its Generalizations: Theory and Applications*. Work Scientific, New-York.

Kozubowski, T.J. and Inusah, S. (2006). A skew Laplace distribution on integers. *AISM*, 58, 555–571.

Krishna, H. and Pundir, P.S. (2007). Discrete maxwell distribution. *InterStat*, November.

Krishna, H. and Singh, P. (2009). A bivariate geometric distribution with applications to reliability. *Communications in Statistics-Theory and Methods*, 38, 1079–1093.

Krishna, H. and Pundir, P.S. (2009). Discrete Burr and discrete Pareto distributions. *Statistical Methodology*, 6, 177–188.

Lisman, J.H.C. and van Zuylen, M.C.A. (1972). Note on the generation of most probable frequency distributions. *Statistica Neerlandica*, 26, 19–23.

Marshall, A.W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*, 84, 641–652.

Mudholkar, G.S. and Srivastava, D.K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, 42, 299–302.

Mudholkar, G.S, Srivastava, D.K. and Freimer, M. (1995). The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. *Technometrics*, 37, 436–445.

Nadarajah, S. and Kotz, S. (2006). The beta exponential distribution. *Reliability Engineering & System Safety*, 91, 689–697.

Nakagawa, T. and Osaki, S. (1975). The discrete Weibull distribution. *IEEE Transactions on Reliability*, 24, 300–301.

Nassar, M.M. and Eissa, F.H. (2003). On the exponentiated Weibull distribution.*Communications in Statistics-Theory and Methods*, 32, 1317–1336.

Nekoukhou, V., Alamatsaz, M.H. and Bidram, H. (2012). A discrete analog of the generalized exponential distribution. *Communications in Statistics-Theory and Methods*, 41, 2000–2013.

Nekoukhou, V., Alamatsaz, M.H., Bidram, H. and Aghajani A.M. (2013a). Discrete beta-exponential distribution. *Communications in Statistics-Theory and Methods*, DOI: 10.1080/03610926.2013.773348.

Nekoukhou, V., Alamatsaz, M.H. and Bidram, H. (2013b). Discrete generalized exponential distribution of a second type. *Statistics*, 47, 876–887.

Nooghabi, M.S., Roknabadi, A.H.R. and Borzadaran, G.M. (2011). Discrete modified Weibull distribution. *Metron*, LXIX, 207–222.

Pollard, R., Benjamin, P. and Reep, C. (1977). Sport and negative binomial distribution. In S. P. Ladany and R. E. Machol (eds.) *Optimal Strategies in Sports.* New York: North-Holland, 188-195.

Roy, D. (2002). Discretization of continuous distributions with an application to stress-strength reliability. *Calcutta Statistical Association Bulletin*, 52, 297–313.

Roy, D. (2003). The discrete normal distribution. *Communications in Statistics-Theory and Methods*, 32, 1871–1883.

Roy, D. (2004). Discrete Rayleigh distribution. *IEEE Transactions on Reliability*, 53, 255–260.

Shaked, M. and Shanthikumar, J.G. and Valdez-Torres, J.B. (1995). Discrete hazard rate functions. *Computers & Operations Research*, 22, 391–402.

Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*. Springer, New York.

Steutel, F.W. and van Harn, K. (2004). *Infinite Divisibility of Probability Distributions on the Real Line*. New York: Marcel Dekker.

Szablowski, P.J. (2001). Discret normal distribution and its relationship with Jacobi Thete functions. *Statistics & Probability Letters*, 52, 289–299.

Wang, C.H. (2009). Determining the optimal probing lot size for the wafer probe operation in semiconductor manufacturing. *European Journal of Operational Research*, 197, 126–133.

Wang, L.-C., Yang, Y., Yu, Y.-L. and Zou, Y. (2010). Undulation analysis of instantaneous availability under discrete Weibull distributions. *Journal of Systems Engineering*, 25, 277–283.

Wein, L.M. and Wu, J.T. (2001). Estimation of replicative senescence via a population dynamics model of cells in culture. *Experimental Gerontology*, 36, 79–88.

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution*, 39, 306–314.

Information for authors and subscribers

Author Guidelines

SORT accepts for publication only original articles that have not been submitted simultaneously to any other journal in the areas of statistics, operations research, official statistics or biometrics.

Articles should be preferably applied and may include computational or educational elements. Publication will be exclusively in English. All articles will be forwarded for systematic peer review by independent specialists and/or members of the Editorial Board, except for those articles specifically invited by the journal or reprinted with permission. Reviewer comments will be sent to the first corresponding author if changes are requested in form or content.

To submit an article, the author must upload it in **PDF format**.

The article should be prepared in double-spaced **format**, using a 12-point typeface.

The **title page** must contain the following items: title, name of the author(s), professional affiliation and complete address, and an abstract (75100 words) followed by the keywords and MSC2010 classification of the American Mathematical Society.

Before submitting an article, the author(s) would be well advised to ensure that the text uses **correct English**.

Bibliographic references within the text must follow this format: author surname followed by the year of publication in parentheses [i.e., Mahalanobis (1936), Rao (1982b)]. The complete reference citations should be listed alphabetically at the end of the article, with multiple publications by a single author listed chronologically. Examples of reference formats are as follows:

- Article: Casella, G. and Robert, C. (1998). Post-processing accept-reject samples: recycling and rescaling. Journal of Computational and Graphical Statistics, 7, 139–157.
- Book: Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2003). Bayesian Data Analysis, 2nd Ed. Chapman & Hall / CRC, New York.
- Chapter in book: Engelmann, B. (2006). Measures of a rating's discriminative power-applications and limitations. In: Engelmann, B. and Rauhmeier, R. (eds), The Basel Risk Parameters: Estimation, Validation, and Stress Testing. Springer, New York.
- Online article (put issue or page numbers and last accessed date): Marek, M. and Lesaffre, E. (2011). Hierarchical Generalized Linear Models: The R Package HGLMMM. Journal of Statistical Software, 39 (13). http://www.jstatsoft.org/v39/i13. Last accessed 28 March 2011.

Explanatory footnotes should be numbered sequentially and placed at the bottom of the corresponding page. **Tables and figures** should also be numbered sequentially.

Once the article has been accepted, the journal editorial office will **contact the author** with instructions about this final version, which should be submitted using **LaTeX**. The journal secretary will provide authors with LaTeX templates and appropriate references to the MSC2010 classification of the American Mathematical Society. All the submissions should be handled by RACO (Revistes Catalanes en Accs Obert) website.

New Authors: please register at: http://www.raco.cat/index.php/SORT/user/register. Upon successful registration you will be sent an e-mail with instructions to verify your registration.

Submission Preparation Checklist

As part of the submission process, authors are required to check off their submission's compliance with all of the following items, and submissions may be returned to authors that do not adhere to these guidelines.

1. The submitted manuscript follows the guidelines to authors published by SORT
2. Published articles are under a Creative Commons License BY-NC-ND
3. Font size is 12 point
4. Text is double-spaced
5. Title page includes title, name(s) of author(s), professional affiliation(s), complete address of corresponding author
6. Abstract is 75100 words and contains no notation, no references and no abbreviations
7. Keywords and MSC2010 classification have been provided
8. Bibliographic references are according to SORT's prescribed format
9. English spelling and grammar have been checked
10. Manuscript is submitted in PDF format

Copyright notice and author opinions

SORT Statistics and Operations Research Transactions
Institut d'Estadstica de Catalunya (Idescat)
Via Laietana, 58 08003 Barcelona. SPAIN
Tel. +34-93.557.30.76 – Fax +34-93.557.30.01
sort@idescat.cat

How to cite articles published in SORT

Commenges, D. (2003). Likelihood for interval-censored observations from multi-state models. *SORT*, 27 (1), 1-12.

Subscription form
SORT *(Statistics and Operations Research Transactions)*

Name _____
_____
Organisation _____
_____
Street Address _____
_____
Zip/Postal code _____ City _____
State/Country _____ Tel. _____
Fax _____ NIF/VAT Registration Number _____
E-mail _____
Date _____

Signature

I wish to subscribe to **SORT *(Statistics and Operations Research Transactions)*** for the year 2015 (volume 39)

Annual subscription rates:

– Spain: €40 (4% VAT included)
– Other countries: €44 (4% VAT included)

Price for individual issues (current and back issues):

– Spain: €15/issue (4% VAT included)
– Other countries: €17/issue (4% VAT included)

Method of payment:

☐ Bank transfer to account number 2100 5000 50 0200051156
IBAN NUM. ES61 2100 5000 5002 0005 1156

Please send this subscription form (or a photocopy) to:

SORT *(Statistics and Operations Research Transactions)*
Institut d´Estadística de Catalunya (Idescat)
Via Laietana, 58
08003 Barcelona
SPAIN
Fax: +34-93-557 30 01